DISSERTATIONES
MEDICINAE
UNIVERSITATIS
TARTUENSIS
379

PRIIT PALUOJA

Computational methods for NIPT-based aneuploidy and microdeletion screening

DISSERTATIONES MEDICINAE UNIVERSITATIS TARTUENSIS
379

DISSERTATIONES MEDICINAE UNIVERSITATIS TARTUENSIS
379

PRIIT PALUOJA

Computational methods for NIPT-based aneuploidy and microdeletion screening

UNIVERSITY OF TARTU
Press
1632

Institute of Clinical Medicine, Faculty of Medicine, University of Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (Medicine) on May 21st, 2025, by the Council of the Faculty of Medicine, University of Tartu, Estonia.

Supervisors	Professor Priit Palta (PhD), Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia
	Professor Andres Salumets (PhD), Institute of Clinical Medicine, University of Tartu, Tartu, Estonia
	Kaarel Krjutškov (PhD), Celvia CC AS, Tartu, Estonia
Reviewed	Senior Doctor Ele Hanson (MD, PhD), Department of Obstetrics and Gynecology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia
	Associate Professor Sander Pajusalu (MD, PhD), Department of Genetics and Personalized Medicine, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia
Opponent	Professor Silvia Bottini (PhD), Institut national de recherche pour l'agriculture, l'alimentation et l'environnement, Université Côte d'Azur, CNRS, Institut Sophia Agrobiotech, Sophia-Antipolis, France
Commencement	August 25, 2025

ISSN 1024-395X (print)
ISBN 978-9916-27-904-5 (print)
ISSN 2806-240X (pdf)
ISBN 978-9916-27-905-2 (pdf)

Copyright: Priit Paluoja, 2025

University of Tartu Press
www.tyk.ee

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

Paper I Teder, H., **Paluoja, P.**, Rekker, K., Salumets, A., Krjutškov, K., Palta, P. (2019). Computational framework for targeted high-coverage sequencing based NIPT. *PLOS ONE* 14(7): e0209139. https://doi.org/10.1371/journal.pone.0209139

Paper II **Paluoja, P.**, Teder, H., Ardeshirdavani, A., Bayindir, B., Vermeesch, J., Salumets, A., Krjutškov, K., Palta, P. (2021). Systematic evaluation of NIPT aneuploidy detection software tools with clinically validated NIPT samples. *PLoS computational biology*, 17(12), e1009684. https://doi.org/10.1371/journal.pcbi.1009684

Paper III **Paluoja, P**., Jatsenko, T., Teder, H., Krjutškov, K., Vermeesch, J.R., Salumets, A. and Palta, P. (2025), BinDel: Detecting Clinically Relevant Fetal Genomic Microdeletions Using Low-Coverage Whole-Genome Sequencing-Based NIPT. *Prenatal Diagnosis*. https://doi.org/10.1002/pd.6758

Contribution of the author to the original publications:

Paper I Co-developed the computational data analysis methods and framework. Contributed to improving the hidden Markov model's (HMM) applicability in detecting trisomy risk. Participated in conceptualising and implementing the secondary computational analysis layer applied to the HMM output. Devised and implemented usage of a support vector machine (SVM) in the second layer. Participated in study design, manuscript draft editing and reviewing.

Paper II Conceptualised the idea of comparing different computational NIPT aneuploidy detection tools, curated sequencing data, developed formal analysis framework for interpreting Z-scores of different computational NIPT tools, developed and implemented methodology, developed and used computational pipeline, developed methodology to simulate data, simulated data, validated and visualised results, wrote the original draft and participated in reviewing and editing. Composed answers to reviewers questions.

Paper III Developed methodology and implemented the BinDel software package, containerised BinDel, performed data analysis, developed methodology to simulate data, simulated data, validated and visualised results, wrote the original draft and participated in reviewing and editing.

ABBREVIATIONS

AC	amniocentesis
AR	allelic ratio
BAM	Binary Alignment Map
cfDNA	cell-free DNA
cffDNA	cell-free fetal DNA
CVS	chorionic villus sampling
FN	false negative
FP	false positive
FF	fetal fraction
FCT	first-trimester combined test
GATK	Genome Analysis Toolkit
HMM	hidden Markov model
β-hCG	human β subunit of chorionic gonadotropin
LCR	low-copy repeats
MD	microdeletion
NF1	neurofibromatosis 1
NIPT	non-invasive prenatal genetic testing
PCA	principal component analysis
PPV	positive predictive value
PAPP-A	protein-A
RC	read count
RCAR	read count and allelic ratio
RPS	reads per sample
SD	standard deviation
SNP	single nucleotide polymorphism
SVM	support vector machine
TP	true positive
T13	trisomy 13
T18	trisomy 18
T21	trisomy 21
WGS	whole-genome sequencing

1. INTRODUCTION

Non-invasive prenatal genetic testing (NIPT) is a screening test that relies on circulating cell-free DNA (cfDNA) and sequencing techniques to detect the risk of fetal aneuploidies. The presence of fetal (placenta) DNA in maternal blood samples was initially reported in 1997, and subsequently, in the years 2011–2012, large-scale NIPT validation studies were conducted and published based on sequencing of cell-free fetal DNA (Bianchi et al., 2012; Lo et al., 1998; Palomaki et al., 2011). Since NIPT relies on cell-free DNA, it not only allows screening for common aneuploidies such as trisomy 21 but also offers the capability to analyse any part of the genome, including areas associated with fetal microdeletions. For example, one frequent microdeletion is the 22q11.2 heterozygous deletion, causing the DiGeorge syndrome, characterised by adverse outcomes that can include schizophrenia, intellectual disabilities, cardiovascular anomalies or hearing deficits (Bassett et al., 2005; Dar et al., 2022). Even though the 22q11.2 microdeletion can lead to a range of phenotypes, including DiGeorge syndrome, velocardiofacial syndrome, or Sedlackova syndrome, the term DiGeorge syndrome is used throughout this thesis to avoid ambiguity (McDonald-McGinn et al., 1993).

NIPT can be based on whole-genome sequencing (WGS) or targeted analysis, which relies on computational sequencing data analysis to assess fetal aneuploidy or microdeletion risk. The analysis of both targeted and WGS-based NIPT data necessitates the utilisation of dedicated algorithms and software tools. Multiple general computational tools exist that are suitable for pre-processing the sequencing data, but the NIPT-specific analysis tools that follow the pre-processing and output the fetal aneuploidy or microdeletion risk are relatively limited. Furthermore, there is a scarcity of information regarding the accuracy of computational NIPT software and algorithms under diverse sample conditions, including variations in sequencing coverage and fetal fraction (FF) and the availability of clinical validation for these tools.

The accurate computational analysis in NIPT depends on a multitude of factors, including the computational tool, sequencing coverage and fetal DNA fraction. While low-coverage WGS-based NIPT assays do not distinguish whether sequencing reads are of fetal or maternal origin, a sufficient proportion of FF and sequencing coverage is crucial to obtain a true result, the latter affecting both the cost, availability and the accuracy of NIPT.

Although NIPT has been studied, the influence of computational tools on its screening accuracy has not been systematically evaluated, and the use of NIPT for assessing the risk of pathogenic microdeletions remains insufficiently explored.

The aim of the doctoral thesis was to create a computational framework for targeted high-coverage sequencing-based NIPT, clinically validate five commonly used and published WGS-NIPT software tools, evaluate the published computational NIPT tools' performance and accuracy considering various sequencing depths and the proportion of cell-free fetal DNA, develop and clinically validate a computational tool for detecting fetal microdeletions risk in WGS-NIPT.

2. REVIEW OF THE LITERATURE

2.1 Fetal aneuploidies

During fertilisation, the chromosomes from both the egg and sperm combine to form the embryo. Each parent is expected to contribute one copy of each chromosome, resulting in a complete set. However, when an incorrect number of chromosomes is contributed by the sperm or egg, a condition known as aneuploidy occurs (**Figure 1**) (Charalambous et al., 2023). Aneuploidies can result in a viable embryo, but the child may be born with congenital conditions, such as Down syndrome, which is caused by an extra copy of chromosome 21 and also referred to as trisomy 21 (T21) (Charalambous et al., 2023).

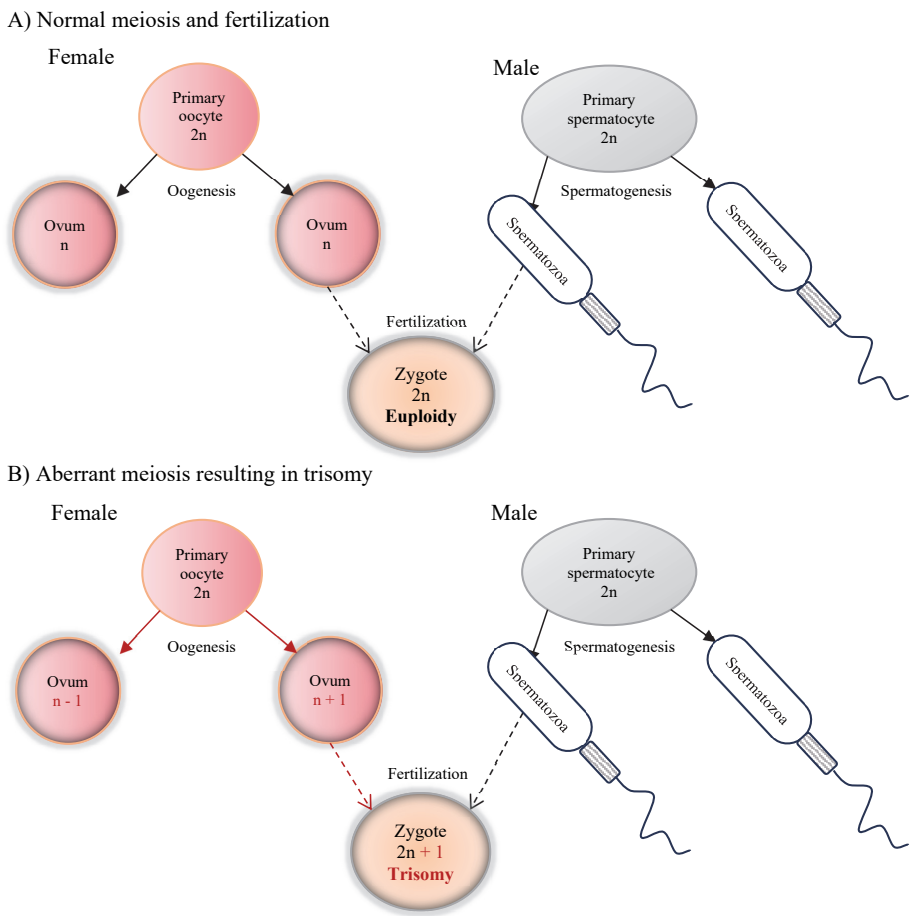


Figure 1. Panel A depicts the process of meiosis and the subsequent fertilisation. Panel B illustrates an instance where a maternal pair of chromosomes fails to separate evenly during meiosis II, leading to a zygote with trisomy after fertilisation. Trisomy can also arise from paternal inheritance or chromosomal translocation, which is not addressed in this figure.

Down syndrome, occurring in approximately 1 of 800 births worldwide, represents the most prevalent form of intellectual disability caused by dosage imbalance of genes due to the triplicate (trisomic) state of all or a critical portion of chromosome 21 (Asim et al., 2015; Bull, 2020; Reimand et al., 2006). Individuals with Down syndrome are predisposed to a spectrum of health issues, including congenital heart disease, Alzheimer's disease or leukaemia (Asim et al., 2015). There is considerable variability between individuals with Down syndrome. For example, cognitive impairment is present in all individuals with Down syndrome, but the extent of it varies widely (Antonarakis et al., 2004). Simultaneously, congenital heart defect occurs only in ∼40% of individuals with Down syndrome (Antonarakis et al., 2004). The median approximate life expectancy of people with Down syndrome in developed countries is approximately 60 years (Bittles & Glasson, 2004).

The second most common autosomal aneuploidy is trisomy 18, causing Edwards syndrome, with prevalence ranging from 1:6,000 to 1:8,000 live births (Cereda & Carey, 2012). The overall prevalence is higher (~1:2,500) due to fetal loss and pregnancy termination after prenatal diagnosis (Cereda & Carey, 2012). Trisomy 18 is associated with multisystem anomalies, with a predilection for affecting the musculoskeletal and cardiac systems (Roberts et al., 2016). Frequently observed musculoskeletal anomalies encompass clenched fists, rocker-bottom feet, and low-set or malformed ears. In the cardiac system, malformations predominantly manifest as ventricular septal defects (Roberts et al., 2016). In one study, the median survival time was 14.5 days for people with trisomy 18 (Rasmussen et al., 2003).

Trisomy 13, causing the Patau syndrome, is the third most common trisomy in live-born infants, with a prevalence of 1 in 16,000 newborns (Heft et al., 2019; Patau et al., 1960). Patau syndrome can cause cryptorchidism, abnormal auricles, polydactyly of hands and/or feet, microphthalmia, micrognathia, low-set ears, aplasia cutis/scalp defects and microcephaly (Petry et al., 2013). The median survival time in one study was seven days for people with trisomy 13 (Rasmussen et al., 2003). With trisomy 13, 18 and 21, the frequency increases with maternal age (Heft et al., 2019).

2.2 Fetal microdeletions

Chromosomal microdeletions represent a subset of copy-number variants (CNV). CNVs are defined as the gain or loss of a DNA segment compared to the reference human genome (Watson et al., 2014). Microdeletions (MDs), which involve the loss of a DNA segment (**Figure 2**), typically result from a process characterised by nonallelic homologous recombination during meiosis between duplicated blocks of sequences known as low-copy repeats (LCRs), which flank the microdeletion event (Watson et al., 2014).

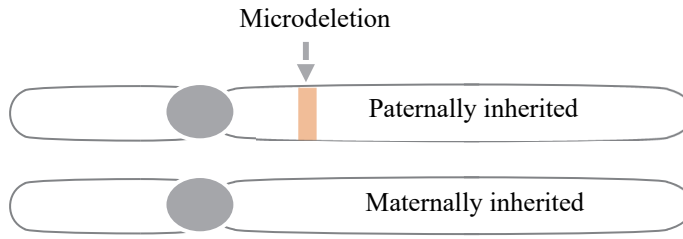
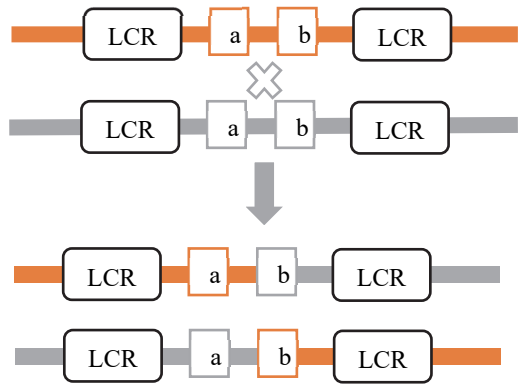


Figure 2. The figure illustrates heterozygous microdeletion on the long arm of chromosome 22, associated with DiGeorge syndrome. One chromosome is inherited maternally, and the other is paternally inherited. Microdeletion can originate from either parent.

Genomic regions prone to microdeletions due to the homologous LCRs can misalign during sperm or egg meiosis so that when crossing-over occurs, the exchange of genetic material is not the same, with one copy of the chromosome getting an extra segment (microduplication) and the other copy losing a segment (microdeletion) (**Figure 3**) (Antshel et al., 2005). Microdeletions can also be inherited or occur early in fetal development and are not connected to maternal age (Heft et al., 2019). Certain genomic regions are more prone to microdeletions and are associated with genomic disorders and neurodevelopmental traits (**Table 1**) (Smajlagić et al., 2021; Stankiewicz & Lupski, 2002).

A) Meiosis I: crossing over



B) Meiosis I: misalignment resulting in unequal crossing cover

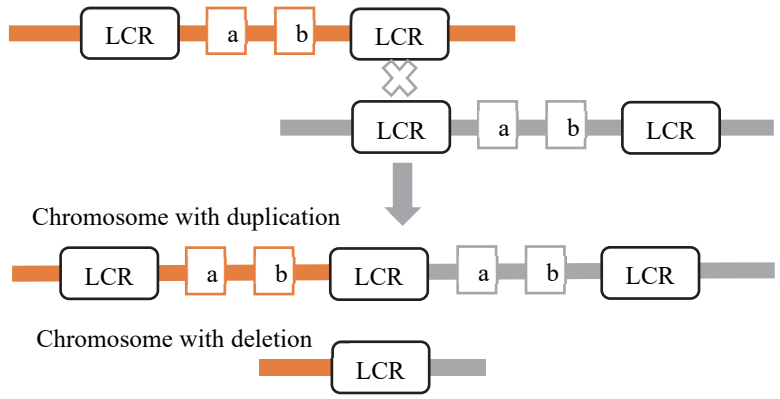


Figure 3. The figure depicts crossing over during meiosis I, wherein homologous chromosomes exchange genetic material. Specifically, the figure portrays the arrangement of two low-copy repeats (LCRs) flanking a genomic locus containing genes designated as 'a' and 'b'. Panel A illustrates the conventional process of crossing over in meiosis I, wherein homologous chromosomes undergo successful recombination without errors. Conversely, panel B delineates a scenario where misalignment occurs between homologous chromosomes facilitated by the presence of LCRs, resulting in an illegitimate crossing over during meiosis. Consequently, one of the chromosome copies undergoes duplication, manifesting as two instances of genes 'a' and 'b', and the other copy undergoes a deletion, leading to the complete loss of genes 'a' and 'b'. Adopted and modified from Watson et al., 2014.

Table 1. A selection of microdeletion syndromes alongside their corresponding frequencies, clinical symptoms, and associated microdeletion genomic regions.

Syndrome	Microdeletion cytogenetic location	Frequency	Clinical symptoms [a]
Angelman syndrome	15q11.2	1:12,000–20,000 (Heft et al., 2019)	Severe cognitive disability, motor dysfunction, speech impairment, hyperactivity, and frequent seizures (Margolis et al., 2015).
Chromosome 1p36 deletion syndrome	1p36	1:5,000–10,000 (Heft et al., 2019)	Intellectual disability, seizures, vision and hearing problems, orofacial clefting, and heart defects (Jordan et al., 2015).
Cri-du-chat syndrome	5p	1:20,000–50,000 (Heft et al., 2019)	High-pitched monochromatic cry, severe psychomotor and mental retardation (Cerruti Mainardi, 2006).
DiGeorge syndrome	22q11.2	1:4,000 (Heft et al., 2019)	Intellectual disability, immunodeficiency, cardiac issues, skeletal defects, and developmental delay (McDonald-Mcginn & Sullivan, 2011).
Jacobsen syndrome	11q23.3	1:100,000 births (Mattina et al., 2009)	Growth and psychomotor retardation, skull deformities, heart, kidney, and gastrointestinal tract malformations (Mattina et al., 2009).
Neurofibromatosis type 1 (NF1)	17q11.2	1:2,500–3,000 (Williams et al., 2009)	Susceptibility to the development of both benign and malignant tumours (Ferner & Gutmann, 2013).
Prader-Willi syndrome	15q11.2	1:10,000–30,000 (Heft et al., 2019)	Severe infantile hypotonia, hypogonadism, obesity, developmental delay and mild intellectual disability (Cassidy et al., 2012).
Smith-Magenis syndrome	17p11.2	1:15,000–1:25,000 (Rinaldi et al., 2022)	Broad face, abnormally shaped ears, congenital heart defects, genitourinary malformations, hearing loss, scoliosis, and impaired production of antibodies (Rinaldi et al., 2022).
Williams-Beuren syndrome	7q11.23	1:7,500 (Strømme et al., 2002)	Cardiovascular anomalies, hypercalcemia, diabetes mellitus, mild-to-moderate intellectual disability (Pober, 2010)
Wolf-Hirschhorn syndrome	4p16.3	1:20,000–50,000 (Battaglia et al., 2015)	Widely spaced eyes, postnatal growth deficiency, intellectual disability, seizures, and antibody deficiency (Battaglia et al., 2015)

[a] The table provides a selection of clinical symptoms, but the list is not exhaustive.

2.2.1 22q11.2 Deletion Syndrome

Chromosome 22q11.2 deletion can result in different overlapping syndromes like DiGeorge syndrome (DGS), velocardiofacial syndrome (VCFS), conotruncal anomaly face syndrome (CTAF), some cases of autosomal dominant Opitz G/BBB syndrome, and Cayler cardiofacial syndrome (asymmetric crying facies) (McDonald-McGinn et al., 1993). Among these, the DiGeorge syndrome occurs in approximately 1 in 4,000 births (overall prevalence of 1:1,524) (Dar et al., 2022; McDonald-Mcginn & Sullivan, 2011).

The majority of 22q11 deletions result from de novo meiotic unequal crossing-over events between homologous chromosomes, with fewer than 10% inherited from an affected parent (**Figure 3**) (Baumer et al., 1998; McDonald-McGinn et al., 2001; Saitta et al., 2004).

DiGeorge syndrome is characterised by adverse infant and childhood outcomes that can include immune deficiency, cardiac, renal and eye anomalies, hypoparathyroidism, skeletal defects, and developmental delay (McDonald-Mcginn & Sullivan, 2011). A Canadian study has shown that the individuals with DiGeorge syndrome who survived to adulthood were at elevated risk of premature death at an average age of 41.5 compared to the Canadian general population expectations of age 80.4 (Bassett et al., 2009).

This DiGeorge microdeletion region is abundant in LCRs, which facilitate misalignment and unequal crossing over during meiosis, resulting in deletions of varying sizes. The most common deletion observed in patients, known as the A-D deletion, accounts for 84%-90% of cases (**Figure 4, Table 2)** (Campbell et al., 2018; Guris et al., 2006). This deletion spans approximately 3 Mb and results in the loss of around 30 genes (Saitta et al., 2004; Yagi et al., 2003). The haploinsufficiency of TBX1 and CRKL is a significant contributor to the phenotype (Guris et al., 2006; Yagi et al., 2003). Another common deletion, the A-B deletion, is approximately 1.5 Mb in length (Saitta et al., 2004). Although less common, the 22q11 region can also harbour other deletions of different sizes (**Figure 4, Table 2**) (Campbell et al., 2018). Mapping unique sequencing reads in the DiGeorge microdeletion region is particularly challenging due to the presence of neurofibromatosis 1 (NF1) microdeletion pseudogenes (Jett & Friedman, 2010; Luijten et al., 2000, 2001). These pseudogenes are present not only in the 22q11 microdeletion region but also in Prader-Willi and Angelman microdeletion region (Jett & Friedman, 2010; Luijten et al., 2000, 2001). This overlap complicates accurately identifying and characterising the deletions specific to DiGeorge syndrome.

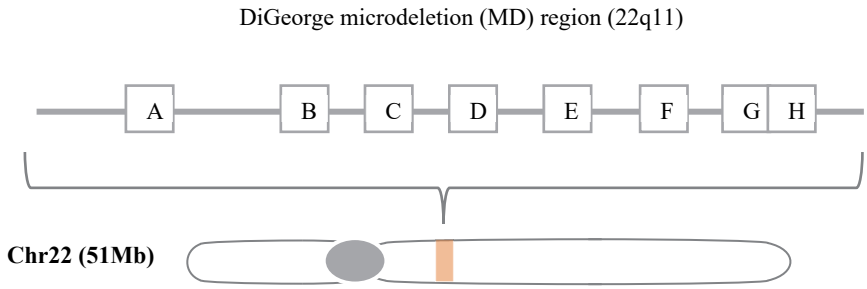


Figure 4. The DiGeorge microdeletion region on chromosome 22 is characterised by multiple low-copy repeats (LCRs), designated as A through H. These LCRs create a genomic landscape susceptible to recombination events, allowing various combinations of microdeletions of varying sizes, including A-D, A-B, or C-D deletions. The deletions can vary in size and encompass different genes within the region, contributing to the heterogeneity observed in individuals affected by DiGeorge syndrome.

Table 2. Microdeletions within the DiGeorge syndrome-related region on chromosome 22 exhibit significant size variability among affected individuals.

Microdeletion	Percentage of affected individuals
A-D	84 %
A-B	5 %
A-C	2 %
B-D	4 %
C-D	1%
C-E, D-E, D-F, D-H, E-F	2%
Non-LCR mediated	2 %

LCR, low-copy repeat. Adapted from Campbell et al., 2018.

Microdeletions collectively exhibit a notable prevalence. Even the most common 22q11.2 microdeletion solely has a more frequent prevalence of 1:1,524 than trisomy 18, causing the Edwards syndrome (~1:2,500) (Cereda & Carey, 2012; Dar et al., 2022). Other common microdeletion syndromes include monosomy 1p36 (1 in 5,000–10,000), Prader-Willi syndrome (1 in 10,000–30,000), Angelman syndrome (1 in 12,000–20,000) and cri-du-chat syndrome (1 in 20,000–50,000) (Heft et al., 2019).

Despite microdeletions' rarity on an individual basis, the cumulative impact of microdeletions is substantial, particularly concerning adverse outcomes during infancy and childhood and no association with increased maternal age (Dar et al., 2022). The high clinical importance of microdeletions necessitates reliable first-trimester fetal screening methods to enable informed decision-making processes and facilitate appropriate medical interventions.

2.3 First-trimester prenatal testing

Given the clinical outcomes of aneuploidies and microdeletions, including viable embryos, early detection during first-trimester screening is essential. Several screening strategies are available, with non-invasive prenatal testing (NIPT) emerging as a recommended method over traditional screening methods due to improved accuracy over conventional screening methods by the American College for Medical Geneticists (Dungan et al., 2023). NIPT uses cell-free DNA (cfDNA) from maternal blood samples to screen for fetal aneuploidies and microdeletions (**Figure 5**) (Mayes et al., 2016). For example, in Belgium and the Netherlands, NIPT is offered as a first-tier screening test to all pregnant women as part of the national prenatal screening program (Lannoo et al., 2023). In Estonia, NIPT serves as a second-tier screening.

In Estonia, for the first-tier screening, the more traditional first-trimester combined test (FCT, also known as One Stop Clinic for Assessment of Risk – OSCAR) is used, which combines the maternal age, ultrasound measurement of fetal nuchal translucency thickness, maternal concentrations of human β subunit of chorionic gonadotropin (β-hCG), and pregnancy-associated plasma protein-A (PAPP-A) (Spencer et al., 1999). FCT is non-invasive and can screen for trisomies 13, 18, and 21 (Geipel et al., 2010). FCT does not screen for microdeletions, but ultrasound sonography can sporadically identify associated findings specific to certain microdeletion syndromes (Lin et al., 2021).

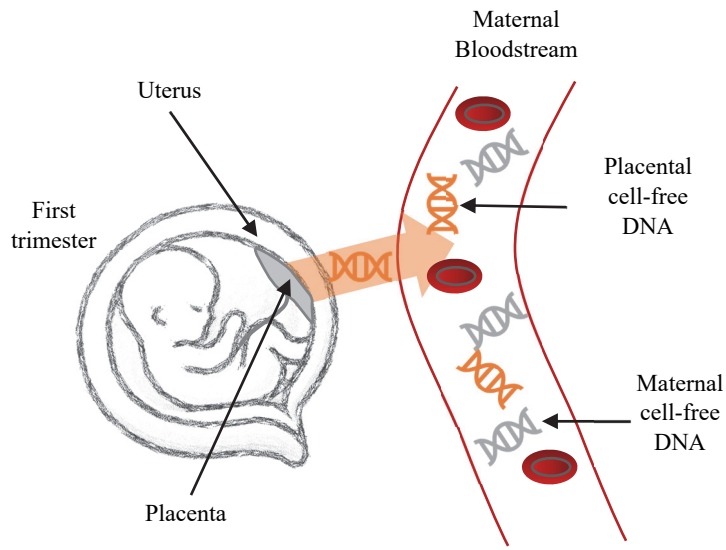


Figure 5. Within the maternal bloodstream, a mixture of genetic material comprises maternal cell-free DNA alongside placental DNA. Placental DNA primarily stems from apoptotic processes within the trophoblastic cells, which are subsequently released into the maternal bloodstream for utilisation. NIPT analysis relies on maternal peripheral blood sample analysis, where the cell-free DNA is extracted, sequenced using high-throughput sequencing, and computationally analysed.

In case of a high-risk finding in NIPT or FCT, a diagnostic invasive confirmation with chorionic villus sampling (CVS) or amniocentesis (AC) is needed. CVS is an invasive procedure where a needle goes through the abdominal wall and uterus or a cannula through the cervix, and placental tissue is collected and karyotyped (Alldred et al., 2017; De Jong et al., 2011). CVS has a risk of a miscarriage of 0.22% (Akolekar et al., 2015). AC is an invasive procedure where a needle is inserted through the abdominal wall into the uterus to sample and karyotype amniotic fluid surrounding the fetus (Alldred et al., 2017; De Jong et al., 2011). The procedure-related risk of miscarriage is 0.11% (Akolekar et al., 2015).

2.4 Non-invasive prenatal testing

Cell-free DNA, which comprises short DNA fragments freely circulating in the bloodstream, was first discovered in 1948 (Mandel & Metais, 1948). A seminal development occurred in 1997 when Lo et al. successfully demonstrated the presence of male fetal DNA in the blood sample of pregnant women (Lo et al., 1997). This pivotal discovery of fetal-derived DNA within maternal blood has catalysed the emergence of a novel generation of prenatal screening methodologies that rely on the analysis of cell-free fetal DNA (cffDNA). Maternal plasma cfDNA contains both maternal and fetal sources of cfDNA (**Figure 5**) (Hui & Bianchi, 2020). The source of fetal DNA is the trophoblast (making the NIPT screening test, not a diagnostic test) (Faas et al., 2012). In DNA sequencing, the proportion of fetal DNA fraction (FF) is the percentage of total maternal plasma cfDNA that is of fetoplacental origin (Hui & Bianchi, 2020). FF is a critical sample-level quality control determinant, as with too low FF, the computational aneuploidy/microdeletion calling nor euploidy confirmation could not be performed with confidence.

NIPT can be based on targeted or whole-genome sequencing and relies on computational sequencing data analysis to assess fetal aneuploidy or microdeletion risk (**Figure 6**). In targeted sequencing-based NIPT, specific genomic regions of interest are selectively amplified or captured before sequencing (Liao et al., 2011). This allows for detecting fetal alleles and the inference of possible fetal aneuploidy (Liao et al., 2012). The targeted strategy reduces the total number of sequencing reads analysed, reducing overall costs (Norwitz & Levy, 2013). At the same time, not all genome is covered, only loci of interest. Therefore, it is only possible to analyse the regions covered during the sequencing in the data analysis. Departing from targeted NIPT, in whole-genome sequencing (WGS) based NIPT, the entire DNA sequence of the genome is interrogated. Instead of targeting specific loci and investigating the count and identity of each allele from each sequence read, the WGS-based NIPT counts sequencing reads across the entire genome and compares these counts to those from a group of healthy, euploid patients.

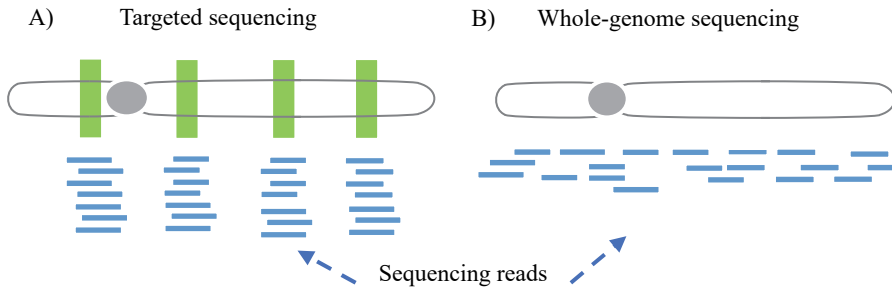


Figure 6. Targeted sequencing involves the selective capture and amplification of specific genomic regions of interest before sequencing (A). This approach allows for the focused analysis of particular regions, such as known disease-associated genes or regions of interest. In contrast, whole-genome sequencing encompasses examining the entire DNA sequence without specifically targeting particular loci (B). By sequencing the entire genome whole-genome sequencing offers a comprehensive interrogation of the entire genome.

2.5 Bioinformatics in routine NIPT screening

In WGS-NIPT, the bioinformatics pipeline involves several key steps to accurately screen fetal aneuploidy and microdeletion risk (**Figure 7**). First, cfDNA is extracted from the maternal blood sample, which contains a mixture of maternal and fetal (placental) DNA. This cfDNA is then sequenced using high-throughput sequencing, generating large volumes of raw sequencing reads. Next, the sequencing reads are aligned to a reference human genome to determine the origin of each read. Following alignment, the bioinformatics pipeline quantifies fetal DNA signals from maternal background noise and performs screening for fetal genetic variations, including aneuploidies and microdeletions. Finally, the results produced by the bioinformatics pipeline are interpreted and reported to the clinician.

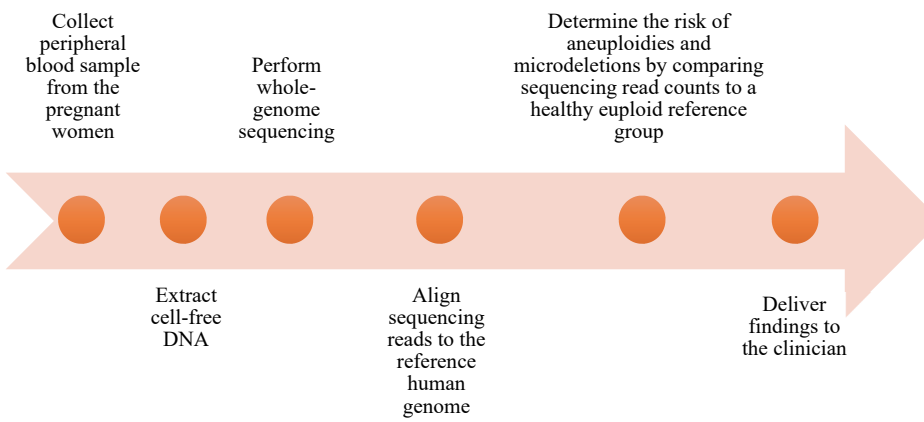


Figure 7. Overview of the whole-genome sequencing-based non-invasive prenatal testing workflow from blood collection to results reporting.

While aneuploidy screening in WGS-based NIPT is relatively straightforward due to the long length of chromosomes, the screening for clinically important microdeletions poses further challenges. The ability to accurately detect microdeletion (MD) risk in NIPT is affected by four main factors: FF, the size of the MD region, sequencing coverage, and the variability in bin read counts in the MD region (Zhao et al., 2015). The higher the FF, the greater the likelihood of detecting potential fetal MDs despite the presence of a significant maternal DNA 'background'. The same principle applies to the MD region length: the longer the MD region, the greater the anticipated coverage by sequencing reads, thereby enhancing the likelihood of detecting any deficiency in sequencing reads (Kucharik et al., 2020; Zhao et al., 2015). MD region length affecting MD screening accuracy has also shown in a study where the detection rate for short microdeletion causing Williams-Beuren syndrome was the lowest compared to the longer microdeletions causing DiGeorge, renal cysts and diabetes, Angelman/Prader-Willi, 1p36 microdeletion, cri-du-chat, and chromosome 9p deletion syndrome (Tian et al., 2023).

Furthermore, MD regions possess different GC% (guanine-cytosine content) compositions and frequently reside proximal to homologous LCRs, which do not facilitate unambiguously unique sequencing read mapping, influencing the standard deviation of the bin read counts, further challenging the accurate microdeletion screening (Watson et al., 2014; Zhao et al., 2015). Last, the low incidence rates require balance in minimising false-positive, false-negative, and maximising true-positive microdeletion calls. Given NIPT's computational intensity and sequencing data complexity, the development of robust algorithms and sophisticated software tools is needed.

2.5.1 Estimating fetal fraction in NIPT

The fetal fraction indicates the signal of interest and is a critical sample-level quality control determinant as different computational tools have different sensitivities for fetal aneuploidy and microdeletion calling. Fetal fraction is influenced by several factors, including gestational age, maternal age, maternal body mass index, and certain medications (Deng et al., 2023; Kuhlmann-Capek et al., 2019). Numerous computational methodologies have been devised to estimate the fetal DNA fraction. One direct approach involves a single nucleotide variation-based method, where the fetal DNA fraction is computed by assessing the ratio of fetal-specific alleles to the total alleles (Peng & Jiang, 2017). To identify fetal-specific alleles, additional parental genotyping or high-depth targeted sequencing as high as ~120× is needed (Jiang et al., 2012; Peng & Jiang, 2017; Sparks et al., 2012).

The available options become more constrained in the context of low-coverage sequencing-based NIPT. It has been shown that size distributions of maternal and fetal cfDNA in maternal plasma differ due to the different fragmentation patterns, fetal DNA being generally shorter than maternal DNA (Chan et al., 2004). Using size distribution differences, Yu et al. developed a method to estimate fetal fraction based on the ratio between the count of fragments ranging from 100 to

150 bp and from 163 to 169 bp (Yu et al., 2014). However, the associated assay requires relatively long sequencing reads, thereby increasing the cost of routine NIPT.

In the context of low-coverage short-read sequencing-based NIPT, estimating fetal fraction tends to be less straightforward. For male fetuses, the FF estimation is relatively straightforward as maternal cfDNA does not contain Y chromosome material, but the male fetal cfDNA contains. Therefore, the Y chromosome ratio to autosomes can be calculated. However, chromosome Y is not completely unique; even in a female fetus pregnancy, some reads align to some parts of the Y chromosome (van Beek et al., 2017). Several computational tools are available to estimate fetal fraction in male fetus pregnancies using reads mapped to the Y chromosome. One such tool is DEFRAG (van Beek et al., 2017). DEFRAG calculates the fetal fraction by taking the fraction of sequencing reads mapped to the Y chromosome and subtracting the average number of reads mapped to the Y chromosome in previously seen female fetus pregnancies (van Beek et al., 2017). This result is then divided by the average fraction of reads in the Y chromosome previously seen from male control samples (van Beek et al., 2017).

In 2015, Kim et al. published a SeqFF, a pre-trained machine learning-based computational tool that estimates fetal fraction using sequence read counts (Kim et al., 2015). In contrast to earlier approaches, SeqFF enables the estimation of fetal fraction for pregnancies with both female and male fetuses without needing prior training (Kim et al., 2015). As a result, SeqFF facilitates immediate and training-data-free utilisation (Kim et al., 2015). In short, SeqFF divides sequenced reads into genomic bins and, with pre-trained coefficients for genomic bins, estimates the fetal fraction (Kim et al., 2015). Elastic net and reduced-rank regression were used to estimate coefficients during the training (Kim et al., 2015). The SeqFF authors suggest that SeqFF leverages the differential representation of the regional sequence of read counts of maternal/fetal ccfDNA, suggesting that fetal ccfDNA may be more likely to originate from genomic regions with increased euchromatic DNA structure, resulting in observed differential fragment length distribution and nonuniformity in coverage relative to maternal ccfDNA (Kim et al., 2015).

Although SeqFF can be readily employed without any additional configuration, it is designed under the assumption that the end-user follows a laboratory protocol and data processing procedure similar to that used during the model's training, as the end-user cannot retrain the model. Deviating from the initial laboratory procedures can reduce accuracy in estimating fetal fraction. In 2019, Raman et al. published a trainable computational fetal fraction estimator PREFACE, relying on neural networks and principal component analysis (PCA), applying biological principles similar to SeqFF (Raman et al., 2019). In contrast to SeqFF, PREFACE necessitates pre-training before its initial use. While PREFACE is a powerful method, it mandates a substantial number of training samples. For example, the authors employed a training set of approximately 5,000 samples in the study (Raman et al., 2019).

2.5.2 Computational approaches for aneuploidy and microdeletion screening in NIPT

The aneuploidy or microdeletion screening of targeted and WGS-based NIPT data necessitates using dedicated algorithms and software tools. For example, in targeted NIPT, there exists software such as FORTE (Fetal Fraction Optimized Risk of Trisomy Evaluation) or NATUS (Next-Generation Aneuploidy Test Using SNPs)(Norwitz & Levy, 2013).

In WGS-based NIPT, multiple analytical aneuploidy risk detection tools have emerged, including RAPIDR, NIPTmer, NIPTeR, and sub-chromosomal aberration (including microdeletion or -duplication) risk detection tools such as Wisecondor or WisecondorX (Johansson et al., 2018; Lo et al., 2014; Raman et al., 2019; Sauk et al., 2018; Straver et al., 2014). In the context of WGS-based NIPT, computational methods for aneuploidy or microdeletion screening generally follow a sequence of steps. Initially, sequencing reads are partitioned into genomic bins to calculate read counts per bin. These read counts undergo normalisation to mitigate various biases inherent in genomic sequencing data. Subsequently, the normalised bins are aggregated to create segments representing entire or specific chromosome segments. The next step involves comparing these normalised values to a reference group of euploid samples, identifying differences that are indicative of aneuploidies or subchromosomal changes in the sample under analysis (**Figure 8**).

The specific algorithms and additional steps in this process may vary between tools. For instance, RAPIDR employs genomic binning, read counting, GC% normalisation to address GC-nucleotide-rich regions, summarisation by chromosome, and computes chromosomal differences relative to a euploid reference (Lo et al., 2014). NIPTeR, in comparison, introduces additional features like chi-squared-based variation reduction to reduce the impact of regions with unexpected variability and alternative methods for quantifying deviations from the euploid reference group (Johansson et al., 2018). WisecondorX employs a Circular Binary Segmentation algorithm to analyse aneuploidies and screen for copy number variations (**Figure 8**)(Olshen et al., 2004; Raman et al., 2019). NIPTmer, however, relies on counting pre-defined sets of unique k-mers on a per-chromosome basis from raw sequencing data, followed by a linear regression model to compare predicted and observed k-mer counts, marking a distinct departure from the methodologies previously described (Sauk et al., 2018).

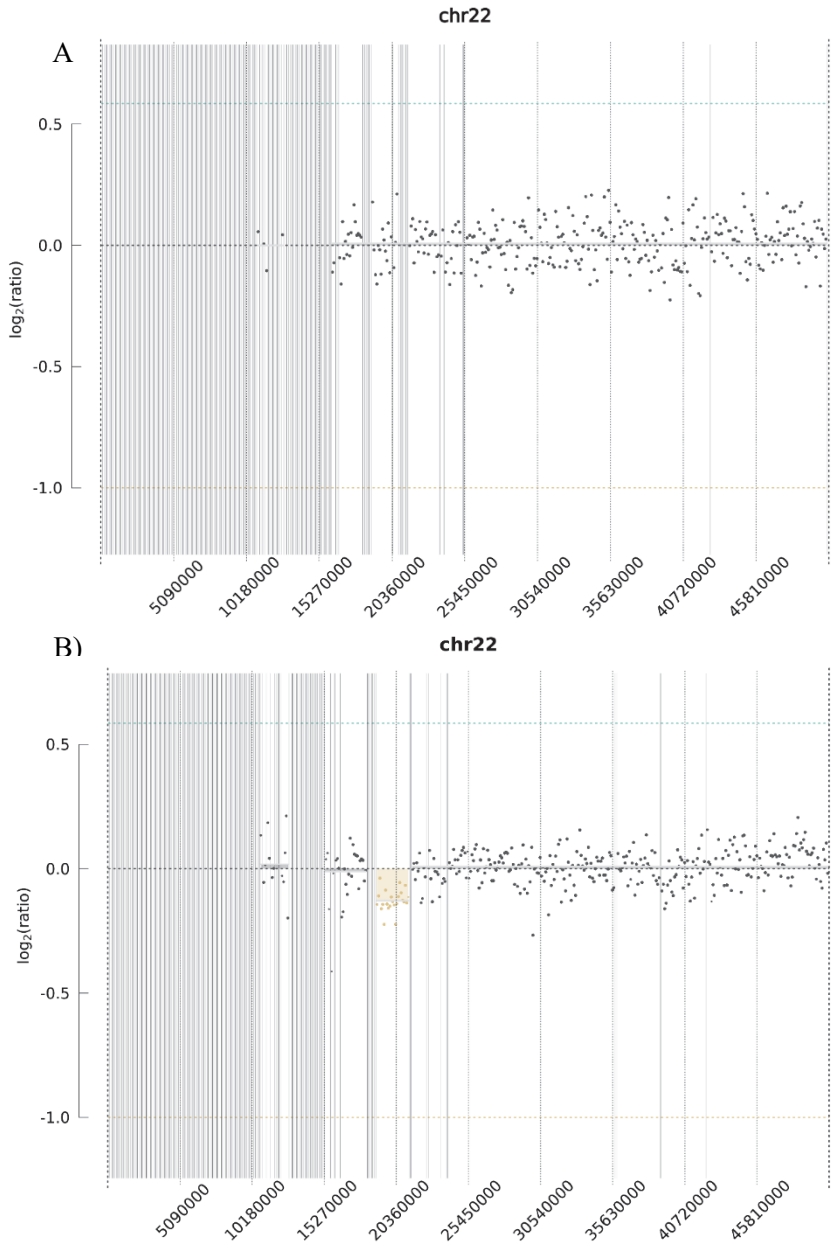


Figure 8. Visualisation of WisecondorX analysis for chromosome 22. The figure illustrates genomic bins and their ratio between observed and expected copy numbers. WisecondorX has not detected any aberration (A). WisecondorX called and highlighted aberration in the DiGeorge microdeletion region for SeraCare Life Sciences Inc circulating cell-free DNA like a mixture of human genomic DNA that consists of matched maternal and fetus DNA with DiGeorge syndrome (lot #10571706) (B). According to the SeraCare Life Sciences Inc specification, 22q11.21 loss is at least 2.8Mb in size from 18,706,046 to 21,540,288 base pairs and fetal fraction 8.50%.

2.6 Rationale for studies

NIPT has emerged as a reliable and safe method for screening common fetal chromosomal abnormalities, such as trisomies, using cell-free DNA obtained from maternal blood. NIPT relies heavily on computational analysis of this DNA. The targeted NIPT approach reduces the number of sequencing reads required, thereby lowering costs and making the method more accessible to a broader patient population. However, the number of available algorithms for analysing targeted sequencing data and calculating trisomy risk remains limited.

WGS-based NIPT is increasingly used and benefits from a range of existing computational tools for trisomy risk estimation, the combined effect of variables such as fetal fraction, sequencing depth, and algorithm choice on screening accuracy has not been systematically evaluated using the same clinically validated sample set.

Finally, while pathogenic microdeletions can lead to phenotypes as severe as those caused by whole-chromosome trisomies, their risk estimation with WGS-based NIPT remains challenging and the availability of appropriate computational tools remains limited.

3. AIMS OF THE STUDY

1. Develop computational methods required for targeted NIPT-based fetal trisomy risk detection and systematically evaluate fetal trisomy risk calling in simulated targeted NIPT sequencing data.
2. Assess the accuracy of fetal trisomy risk assessment using low-coverage WGS-based NIPT data. Determine how sequencing depth and estimated fetal fraction affect trisomy risk calling in low-coverage WGS-based NIPT.
3. Examine the feasibility of accurate fetal microdeletion risk estimation from low-coverage WGS NIPT data. Develop the necessary methodologies and a software tool that implements them to accurately assess fetal microdeletion risk and validate its performance on clinically confirmed microdeletion samples.

4. MATERIALS AND METHODS

4.1 Study subjects and samples

4.1.1 Clinical patients (Papers II–III)

For the evaluation of computational trisomy screening tools, 1092 samples were used, of which 669 were used as reference sets for evaluated computational tools. For these 669 samples, euploidy had been confirmed by the NIPTIFY screening test (Celvia CC) and postnatal evaluation. The remaining 423 samples were used for the validation and were from a previously published study by Žilina et al. (Žilina et al., 2019). Of 423, 258 were high-risk pregnancies that had undergone diagnostic invasive prenatal analysis (Žilina et al., 2019). There were 19 samples with confirmed fetal trisomy 21 (T21), eight with trisomy 18 (T18) and three with trisomy 13 (Žilina et al., 2019). Samples were sequenced with the Illumina NextSeq 500 platform, producing 85 bp single-end reads with an average per-sample coverage of 0.32× at the University of Tartu, Institute of Genomics Core Facility as described previously (Žilina et al., 2019).

We used 300 samples reported previously as euploid fetus pregnancies by NIPTIFY screening test and postnatal evaluation for BinDel fetal microdeletion simulation analysis. Of these 300 samples, 200 were retained for the BinDel reference set, and 100 were used to simulate fetal microdeletions. Samples were processed with modifications similar to previously published guidelines from Katholieke Universiteit Leuven (Bayindir et al., 2015). Peripheral blood samples were collected in Streck cell-free DNA BCT tubes, and plasma was separated with standard dual centrifugation. Cell-free DNA was extracted from 3 ml of plasma using the ThermoFisher Scientific MagMAX Cell-Free DNA Isolation Kit. Subsequently, whole-genome libraries were prepared using the Fragmented DNA Compact Sequencing Assay NIPT protocol with 12 cycles for the final PCR enrichment step. After library quantification, 36 samples were equimolarly pooled, and the quality and quantity of the pool were evaluated using the Agilent Technologies 2200 TapeStation. WGS was carried out on the Illumina Inc. NextSeq 550 instrument with 85 bp single-end reads (average coverage of 0.32×), resulting in the read count between 8M to 18.5M reads per sample (RPS), with a median of 12.7M RPS.

In the BinDel and WisecondorX analysis, Katholieke Universiteit Leuven samples received as single-end FASTQ files were used. We used 200 known fetal euploid samples for the BinDel and WisecondorX (v1.2.4) reference set. For initial validation, 84 samples containing 50 euploids and 34 samples with clinically confirmed MDs were analysed. Samples with clinically confirmed MDs originated from 22 unique patients, from which some cfDNAs were sequenced 2–3 times in different libraries (**Table 3**). Microdeletion origin, region, sample fetal fraction and gestational age are presented in **Table 3**. We applied BinDel and WisecondorX on 84 samples, reported high-risk calls to Katholieke Universiteit Leuven, and received the results.

Table 3. BinDel validation sample set information.

State / Syndrome	Patient	Origin	FF (%) [a]	Gestational age when genomic aberration was proven by microarray analysis [b]
DiGeorge (A-D)	1	Fetal	10.0	32 weeks
DiGeorge (A-D)	1	Fetal	8.8	32 weeks
Angelman	2	Fetal	16.2	Postnatal
Angelman	2	Fetal	8.7	Postnatal
DiGeorge (A-B)	3	Fetal	8.6	Postnatal
DiGeorge (A-D)	4	Maternal and fetal	11.9	16 weeks (NIPT reported maternal)
DiGeorge (A-D)	4	Maternal and fetal	10.0	16 weeks (NIPT reported maternal)
Williams-Beuren	5	Fetal	8.9	Postnatal
DiGeorge (A-D)	6	Fetal	7.8	Postnatal
DiGeorge (A-D)	6	Fetal	9.8	Postnatal
NF1	7	Maternal and fetal	7.9	Postnatal
DiGeorge (A-D)	8	Fetal	12.1	21 weeks
DiGeorge (A-D)	8	Fetal	8.1	21 weeks
Smith-Magenis	9	Fetal	16.4	15 weeks (NIPT reported)
Prader-Willi	10	Fetal	12.2	15 weeks (NIPT reported)
Prader-Willi	10	Fetal	10.9	15 weeks (NIPT reported)
Angelman	11	Fetal	10.1	Postnatal
DiGeorge (A-D)	12	Fetal	7.4	32 weeks
DiGeorge (A-D)	12	Fetal	6.2	32 weeks
DiGeorge (A-D)	13	Fetal	6.1	17 weeks

State / Syndrome	Patient	Origin	FF (%) [a]	Gestational age when genomic aberration was proven by microarray analysis [b]
Williams-Beuren	14	Fetal	7.4	postnatal
DiGeorge (A-D)	15	Fetal	5.7	Stillbirth (31 weeks)
DiGeorge (A-D)	16	Maternal and fetal	5.8	15 weeks (NIPT reported maternal)
DiGeorge (A-D)	16	Maternal and fetal	9.1	15 weeks (NIPT reported maternal)
Smith-Magenis	17	Fetal	12.0	Postnatal
Smith-Magenis	17	Fetal	10.3	Postnatal
DiGeorge (A-B)	18	Maternal	12.2	16 weeks (NIPT reported maternal)
DiGeorge (A-D)	19	Fetal	16.5	17 weeks (NIPT reported)
DiGeorge (A-B)	19	Fetal	15.8	17 weeks (NIPT reported)
DiGeorge (A-B)	19	Fetal	13.4	17 weeks (NIPT reported)
DiGeorge (A-D)	20	Fetal	5.9	Postnatal
DiGeorge (A-D)	20	Fetal	7.8	Postnatal
Williams-Beuren	21	Fetal	6.4	Postnatal
DiGeorge (A-D)	22	Fetal	7.4	Postnatal

[a] The fetal fraction (FF) values provided in the table correspond to the first trimester of pregnancy.
[b] Microdeletions were detected during screening only if reported as “NIPT reported”.
Adapted from Paluoja et al., 2025.

4.1.2 Simulated NIPT samples (Paper I-III)

A total of 1,800 datasets, each containing 100 training and 10,000 test samples simulating pregnant women's cfDNA targeted sequencing data, were used for the development of the computational framework for targeted sequencing-based NIPT. Each sample incorporated 1,000 targeted loci per chromosome of interest. Datasets varied in fetal condition (euploidy, maternally or paternally originated trisomy), sequencing depth (500–15,000 reads per locus) and FF (1–20%).

To measure the impact of sequencing depth on the accuracy of WGS-based NIPT computational tools, we systematically subsampled raw sequencing read data from the average of 20M reads to 1.25M reads per sample. Specifically, each sample was subsampled into six different groups of subsamples, including 20M, 15M, 10M, 5M, 2.5M, and 1.25M (**Table 4**). Subsampling to 5–15M was achieved by leaving the appropriate number of NextSeq 500 output lanes out. For lower coverages, one lane (5M) was taken and subsampled with samtools view (Danecek et al., 2021). Next, we applied a computational tool to infer Z-scores. As all computational NIPT tools required the creation of a euploid reference set to which the sample under analysis is later compared, we also created a corresponding reference set for each analysed sequencing coverage. To illustrate, with a 5M RPS coverage, both validation and reference samples were subsampled to 5M RPS and analysed as a 5M RPS group. The fetal DNA fraction and Z-score estimates for chromosomes 13, 18, and 21 for each sample on each coverage were calculated. If the tool failed to operate on certain coverages, then analysis of that coverage for the failed tool was left out.

Table 4. Analysed sample groups.

M RPS [a]	Sample Group [b]	SD [c]	Min [d]	Average [e]	Max [f]	N [g]
20	T21 (clinically validated)	4,277,096	14,347,516	24,163,449	32,626,785	19
15	T21 (clinically validated)	3,148,835	10,937,064	18,083,151	24,334,564	19
10	T21 (clinically validated)	2,081,671	7,209,906	11,958,958	16,127,678	19
5	T21 (clinically validated)	1,048,024	3,670,427	6,023,952	8,173,000	19
2.5	T21 (clinically validated)	523,966	1,833,823	3,011,415	4,086,277	19
1.25	T21 (clinically validated)	262,260	915,450	1,505,824	2,042,374	19
20	T18 (clinically validated)	2,791,633	18,450,798	23,124,624	27,336,184	8
15	T18 (clinically validated)	2,071,817	14,024,066	17,350,621	20,485,890	8
10	T18 (clinically validated)	1,395,351	9,235,616	11,476,668	13,625,081	8
5	T18 (clinically validated)	697,871	4,685,815	5,794,650	6,891,991	8
2.5	T18 (clinically validated)	348,455	2,344,985	2,898,169	3,446,589	8
1.25	T18 (clinically validated)	174,390	1,172,245	1,449,377	1,722,745	8
20	T13 (clinically validated)	2,980,546	19,042,280	22,183,555	24,971,957	3
15	T13 (clinically validated)	2,257,362	14,500,538	16,725,422	19,013,917	3
10	T13 (clinically validated)	1,469,534	9,567,867	11,039,287	12,506,927	3
5	T13 (clinically validated)	742,824	4,870,521	5,591,835	6,354,446	3
2.5	T13 (clinically validated)	370,783	2,435,984	2,796,437	3,176,752	3
1.25	T13 (clinically validated)	184,533	1,219,133	1,398,883	1,587,854	3
20	Validation Low Risk	3,594,972	13,643,477	23,747,773	32,396,033	393
15	Validation Low Risk	2,688,192	10,345,586	17,800,264	24,374,332	393

M RPS [a]	Sample Group [b]	SD [c]	Min [d]	Average [e]	Max [f]	N [g]
10	Validation Low Risk	1,791,357	6,839,734	11,799,169	16,181,461	393
5	Validation Low Risk	896,917	3,453,849	5,944,443	8,061,718	393
2.5	Validation Low Risk	448,456	1,726,568	2,972,178	4,028,665	393
1.25	Validation Low Risk	224,239	863,244	1,486,192	2,014,776	393
20	Reference	6,032,848	11,534,450	23,506,647	55,026,706	669
15	Reference	4,531,992	8,656,925	17,654,442	41,383,148	669
10	Reference	3,013,960	5,736,648	11,695,518	27,432,704	669
5	Reference	1,516,731	2,919,971	5,942,981	13,920,933	669
2.5	Reference	758,325	1,458,860	2,971,523	6,957,681	669
1.25	Reference	379,135	729,274	1,485,737	3,476,940	669

T21, trisomy 21; T18, trisomy 18; T13, trisomy 13. The table presents the average read count in millions of reads coverage group[a], condition[b], coverage standard deviation[c], minimum[d], average[e] and maximum[f] read count and the number of samples[g]. Adapted from Paluoja et al., 2021.

To observe low sequencing coverage-driven uncertainty in Z-score inference, we used ten high FF (10.23–18.57%) low-risk (NIPTeR chromosome 21 Z-score -0.0996–0.0942) WGS-NIPT (22M–30M) samples. We generated a single pooled low-risk sample with 247M reads by concatenating and sorting these ten samples. Next, the pooled sample was randomly subsampled 2,000 times to groups of 2.5, 5, 10 and 20M RPS, and for each subsample, we applied NIPTeR to infer Z-scores. Using the inferred Z-scores, we calculated deviations by subtracting the average group Z-score from the calculated Z-scores. Finally, deviations were added to the original validation sample Z-scores, leading to 2,358,000 simulated low-risk Z-scores and the number of false positive and true negative trisomy calls were counted. The same methodology was also applied to eight T21 samples (20M–33M, FF 10.43%–14.21%, Z-score 13.77–23.64), resulting in the pooled sample with 203M reads. For T21 analysis, we also created 6, 7, 8 and 9M RPS reference groups.

We analysed sequencing depth's impact on the accuracy of fetal fraction estimation. First, we estimated FF for the 20M RPS group. As the number of samples exhibiting low FF estimations (0–4%) was limited to 12 samples, we divided samples by FF into groups of 0–5% (n = 28) and 5–15% (n = 302). This stratification allows for more accurate correlation estimates on samples with low FF. Next, we subsampled raw sequencing read data from the average of 20M reads to 1.25M reads per sample and compared the original 20M FF estimates with reduced coverage estimates by calculating Pearson correlation.

We simulated heterozygous fetal MDs in WGS-NIPT using GATK (Genome Analysis Toolkit (Van der Auwera et al., 2020)) for BinDel development and evaluation. First, we extracted and downsampled the reads within the MD region to a temporary BAM (Binary Alignment Map) file, replicating heterozygous deletions while considering the target fetal fraction. Next, we extracted the reads outside the MD region from the original file and transferred them to another temporary (BAM) file. Finally, these temporary BAM files were then merged. We applied this simulation process for every combination of 12 FF, 12 MD, and 100 euploid fetus samples.

4.2 Computational methods

4.2.1 Sample pre-processing (Paper II–III)

For alignment-based computational NIPT tools, all samples were aligned against human reference assembly GRCh37 (RAPIDR) or GRCh38 (BinDel, NIPTeR, WisecondorX). Sequencing reads originating from identical DNA fragments were flagged as duplicates and sorted. K-mer-based NIPTmer required no special preprocessing.

4.2.2 Support vector machine as a framework for HMM output interpretation (Paper I)

Support vector machine (SVM) is a supervised learning model that finds a hyperplane that separates samples by category during a training phase. In the case of the developed computational framework, the hyperplane is found based on the hidden Markov model (HMM) state proportions, read count and FF such that it maximally separates samples by their state (euploid, maternal trisomy or paternal trisomy). Once trained, SVM can predict the state of an unseen sample by observing which side of the hyperplane the sample falls based on the read count, FF, and HMM state proportions.

4.2.3 Uniform framework for defining trisomy risk (Paper II)

While all evaluated fetal trisomy detection tools outputted a per-chromosome Z-score for assessing the trisomy risk, each algorithm also had specific differences. For example, RAPIDR also outputs trisomy calls, WisecondorX calculates in addition to the Z-score the $\log_2$ ratio of the observed number and expected number of reads, NIPTmer specifies a trisomy risk calling cut-off, GIPseq provides decision rules for interpretation of the output, and NIPTeR avoids defining cut-off threshold (Bayindir et al., 2015; Johansson et al., 2018; Lo et al., 2014; Raman et al., 2019; Sauk et al., 2018). To ensure uniform scoring between different computational tools, we used Z-scores as all the computational NIPT tools supported outputting it in the analysis.

However, a trisomy risk-calling Z-score cut-off threshold is needed to decide trisomy risk. We defined a generally usable framework, relying on a percent point (quantile) function. The framework allows choosing a probability such that, presuming a euploid sample, sets a cut-off value such that the Z-score of a euploid sample would be less than or equal to the cut-off with the chosen probability. The cut-off threshold can be calculated by presuming the Z-score distribution follows a standard normal distribution, or it can be calculated by empirically observing the mean and SD of the euploid sample reference group. The first cut-off is referenced as a theoretical cut-off Z_t and the second as the empirical cut-off Z_e. A probability of 0.9999 was used in all calculations: $P_{99.99}(P[Z_{sample} \leq z_{cut\ off\ threshold}]) = 0.9999$.

4.2.4 Computational tools (Papers II–III)

The most relevant analysis aspects for each evaluated computational tool are described below.

NIPTeR v1.0.2 was used. Each sample was binned and GC% corrected with bin_bam_sample (separate_strands set to TRUE) and gc_correct (method ‘bin’, ref_genome ‘hg38’), respectively. A reference group was created with the as_control_group. Each sample under analysis was further chi-corrected with chi_correct, and the Z-score was calculated with calculate_z_score. We also cal-

culated using prepare_ncv (max_elements set to 9) and calculate_ncv_score normalised chromosome values (denoted NIPTeR NCV), which minimises variation between sequencing runs (Johansson et al., 2018; Sehnert et al., 2011).

WisecondorX v1.1.5 was evaluated. Each sample was converted to .npz format with the convert command. The reference group was created with newref -nipt -refsize 669 -binsize 100000. Z-scores were calculated using the predict command. We also tested Wisecondor (Paco_0.1) but left it out due to the existence of newer WisecondorX and due to Wisecondor not providing Z-scores for the entire chromosomes.

NIPTmer was obtained from the University of Tartu Department of Bioinformatics webpage but was updated to the latest version from GitHub due to software issues. Pre-built lists based on the GRCh37 reference genome were used as NIPTmer is not an alignment-based method, and GRCh37 lists are known to work (Žilina et al., 2019).

GIPseq was run by KU Leuven (Bayindir et al., 2015). We uploaded samples to the KU Leuven Google Cloud bucket for both the reference set creation and analysis and received the raw output of the pipeline. Since the GIPseq NIPT pipeline provided sophisticated output not limited to euploid and trisomy states, the pipeline was not directly comparable with other evaluated tools. GIPseq author's interpretations (euploid or trisomy) for each sample were used to assess the number of false-negative and false-positive trisomy cases. The read count effect on the Z-scores was analysed using the theoretical Z-score cut-off for comparability between different computational tools.

We also planned to include DASAF R in the analysis, but the corresponding web links to the software in the publication were inoperable, and the software was not included (Liu et al., 2016).

RAPIDR v0.1.1 was used. GRCh37 human reference genome was used as RAPIDR does not support GRCh38. Function makeGCContentData was used for calculating the GC% content before binning reference samples with makeBinnedCountsFile (using a default bin size of 20000). Finally, the R object representing the reference set was created with createReferenceSetFromCounts (gcCorrect and filterBin set to TRUE). Although RAPIDR provides trisomy calls (based on a fixed Z-score cut-off of three) alongside Z-scores, it was not used for uniform computational tool comparison (Lo et al., 2014).

SeqFF was used to estimate FF (Kim et al., 2015). Before applying SeqFF, samples were aligned against the human reference assembly GRCh37.

BinDel, developed as part of this doctoral thesis, is a computational tool for estimating fetal microdeletion risk in whole-genome sequencing-based NIPT. The tool is publicly available at: https://github.com/seqinfo/BinDel. Microdeletion high risk was called if the BinDel risk probability was 50% with PCA99% for shorter MD regions (NF1, Williams-Beuren) or 80% with PCA95% for longer target MD regions. PCAX% denotes that the number of PCA components are chosen to capture X% of the cumulative variance based on BinDel the reference group. BinDel supports varying genomic bin sizes. After analysing different bin sizes

and considering random sequencing read placement effects, differences in sequencing read placement patterns across various microdeletion regions, and the number of bins per MD region, 300kb bin size was used (Paluoja et al., 2021). However, shorter bin sizes were used in regions where 300kb led to too few bins (**Table 5**). The coordinates were mainly derived from the DECIPHER, ClinGen and OMIM databases (Amberger et al., 2015; Firth et al., 2009; Rehm et al., 2015).

Table 5. A priori chosen target microdeletion regions, their coordinates and bin sizes used in the analysis.

Coordinate (chromosome:start-end)	Microdeletion region and database used	Genomic bin size (kb)
1:1–27 600 000	1p36 OMIM (Amberger et al., 2015)	300
1:10 001–12 780 116	1p36 DECIPHER (Firth et al., 2009)	300
1:898 703–6 229 913	1p36 terminal region, ClinGen dosage ID: ISCA-37434 (Firth et al., 2009; Rehm et al., 2015)	300
3:192 600 000–198 295 559	Chromosome 3q29 (Amberger et al., 2015)	300
4:1–4 500 000	Wolf-Hirschhorn OMIM (Amberger et al., 2015)	300
4:337 779–2 009 235	Wolf-Hirschhorn terminal region, ClinGen dosage ID: ISCA-37429 (Firth et al., 2009; Rehm et al., 2015)	300
4:1 567 470–2 108 509	Wolf-Hirschhorn DECIPHER (Firth et al., 2009)	100
5:10 001–12 533 192	Cri-du-chat DECIPHER (Firth et al., 2009)	300
7:72 700 001–77 900 000	Williams-Beuren OMIM (Amberger et al., 2015)	300
7:73 330 452–74 728 334	Williams-Beuren DECIPHER (Firth et al., 2009)	300
8:116 700 000–126 300 000	Langer-Giedion OMIM (Amberger et al., 2015)	300
11:114 600 000–135 086 622	Jacobsen OMIM (Amberger et al., 2015)	300
15:22 677 345–28 134 728	Prader-Willi/Angelman Type I. Start coordinate DECIPHER CNV syndrome, end coordinate ClinGen (dosage ID ISCA-37478) (Firth et al., 2009; Rehm et al., 2015)	300
15:23 374 765–28 134 728	Prader-Willi/Angelman Type II. Start coordinate DECIPHER CNV syndrome, end coordinate ClinGen (dosage ID ISCA-37404) (Firth et al., 2009; Rehm et al., 2015)	300
17:16 869 758–20 318 836	Smith-Magenis DECIPHER (Firth et al., 2009)	300
17:30 780 079–31 936 302	NF1 DECIPHER (Firth et al., 2009)	200
20:80 106–1 311 812	20p13. Coordinates from a custom-ordered DNA Mix (lot #10560229) specification from SeraCare Life Sciences Inc with fetus DNA having a pathogenic loss of the terminal region of 20p13 and a pathogenic 3q29 duplication.	300
22:18 924 718–20 299 685	DiGeorge A-B, ClinGen dosage ID: ISCA-37433 (Firth et al., 2009; Rehm et al., 2015)	300
22:18 924 718–21 111 383	DiGeorge A-D, ClinGen dosage ID: ISCA-37446 (Firth et al., 2009; Rehm et al., 2015)	300
22:20 377 696–21 111 383	DiGeorge B/C-D, ClinGen dosage ID: ISCA-37516 (Firth et al., 2009; Rehm et al., 2015)	100

Adapted from Paluoja et al., 2025.

4.3 Ethics

This study was performed with the written informed consent from the participants and with the approval of the Research Ethics Committee of the University of Tartu (#315/T-13, #352/M-12) and the Ethics Committee Research UZ / KU Leuven (S66817).

5. RESULTS

5.1 Paper I: Computational fetal trisomy risk detection with targeted sequencing-based NIPT

Our objective was to integrate a targeted NIPT strategy aimed at screening common fetal trisomies. Targeted NIPT strategy lowers overall costs but also facilitates increased sample throughput. However, the majority of available computational methods for sequencing-based NIPT rely on low-coverage WGS and are unsuitable for targeted NIPT. We developed a novel computational framework for a targeted high-coverage sequencing-based NIPT to screen fetal trisomy.

5.1.1 Analytical framework for HMM output interpretation (Paper I)

The developed HMM-based approaches aimed at detecting fetal chromosomal trisomies within targeted sequencing assays exhibited limitations in sensitivity under specific scenarios. For example, the combined read count and allelic ratio model (RCAR) failed to detect any paternally inherited trisomies across varying fetal fractions (**Table 6**). Conversely, within a fetal fraction range of 1–5%, the sensitivity for maternally inherited trisomies was also only 0.47 (**Table 6**). Although the read count (RC) model showed a high sensitivity of 1.0 across all fetal fractions, its robustness relied on the accurate FF estimation (**Table 7**). Since determining fetal fraction based solely on the read count in targeted NIPT is unfeasible, instances may arise when FF cannot be estimated. In such cases, assuming a presumed FF of 10% while the actual FF falls within the 1–5% range, the RC model sensitivity dropped from 1.0 to 0.11 (**Table 7**).

Table 6. Fetal trisomy classification accuracy with read count and allelic ratio model.

Fetal fraction	Specificity	Sensitivity (maternal trisomy)	Sensitivity (paternal trisomy)
1–5%	1.00	0.47	0.00
6–10%	1.00	0.96	0.00
11–15%	1.00	1.00	0.00
16–20%	1.00	1.00	0.00

Each value represents an average classification accuracy over 1,200,000 simulated cell-free DNA samples with fetal (maternally or paternally inherited) trisomy at the different sequencing depth intervals (500–15,000 reads). Adopted and modified from Teder et al., 2019.

Table 7. Fetal trisomy classification accuracy with read count computational model.

Fetal fraction	Specificity		Sensitivity	
	Fetal fraction is known.	The model presumes a fetal fraction of 10%.	Fetal fraction is known.	The model presumes a fetal fraction of 10%.
1–5%	1.00	1.00	1.00	0.11
6–10%	1.00	1.00	1.00	1.00
11–15%	1.00	1.00	1.00	1.00
16–20%	1.00	1.00	1.00	1.00

Each value is calculated over 1,200,000 simulated cell-free DNA samples with fetal trisomy at the different sequencing depth intervals (uniformly 500 - 15,000 reads). Adopted and modified from Teder et al., 2019.

The underlying reason for the notably reduced sensitivity observed in the HMM-based solution is caused by the interpretation of HMM output scores. Sample trisomy is determined by the highest frequency of locus states (e.g., trisomy or euploid). In this approach, the sample is designated as trisomy if the count of trisomic loci surpasses that of euploid loci. However, this methodology presents certain limitations. For instance, when the trisomic and euploid loci counts are nearly equivalent, the final sample state is susceptible to random fluctuations.

Moreover, the inability to detect any paternally inherited trisomies relates to the overlapping emission distributions of maternally and paternally originated trisomies within the HMM model (**Table 6**). The HMM model was designed to favour maternally originated trisomies due to their relatively higher prevalence in the population. As a result of this bias, combined with the decision-making process based on the most prevalent state of loci, no paternally inherited trisomies were identified (**Table 6**).

Having recognised the necessity for a more advanced decision-making system to classify the sample state, we developed the concept of integrating an additional classification layer upon the HMM output scores that would also consider the distribution of loci counts associated with euploid, maternally, and paternally inherited trisomies when determining the sample state. The additional layer determines the decision boundary distinguishing samples with euploid, maternally, and paternally inherited trisomies during the training phase that could later be applied to unseen samples. For implementation, we used a support vector machine (SVM) as it finds a decision boundary (hyperplane) such that the distance (margin) between different states (in the feature space) is maximised, replacing previous highest frequency based classification system with SVM outputted state (based on the decision boundary).

5.1.2 Evaluation of the analytical framework of HMM output interpretation (Paper I)

Supplemental SVM considerably increased trisomy detection sensitivity across all three models and marginally decreased specificity in some cases (**Table 8, Table 9, Table 10**). Notably, the SVM increased paternally inherited trisomy detection sensitivity from zero to one across all FF for the allelic ratio (AR) and RCAR models (**Table 8, Table 9**). For the AR and RCAR models, the SVM also considerably improved the detection of maternally inherited trisomy, albeit, for RCAR, there was a decrease of 0.01 specificity on FF 1–5% and for the AR model, a decrease between 0.02–0.03 on all tested FF (**Table 8, Table 9**).

Table 8. Fetal trisomy classification accuracy with allelic ratio computational model and with supplemental SVM.

Fetal fraction	Specificity		Sensitivity (maternal trisomy)		Sensitivity (paternal trisomy)	
	HMM	SVM	HMM	SVM	HMM	SVM
1–5%	1.00	0.98	0.02	0.10	0.00	1.00
6–10%	1.00	0.97	0.63	0.78	0.00	1.00
11–15%	1.00	0.97	0.89	0.97	0.00	1.00
16–20%	1.00	0.98	0.96	1.00	0.00	1.00

SVM, support vector machine; HMM, hidden Markov model. Each value represents an average classification accuracy over 1,200,000 simulated cell-free DNA samples with fetal (maternally or paternally inherited) trisomy at the different sequencing depth intervals (uniformly 500–15,000 reads). Adopted and modified from Teder et al., 2019.

Table 9. Fetal trisomy classification accuracy with RCAR computational model and supplemental SVM.

Fetal fraction	Specificity		Sensitivity (maternal trisomy)		Sensitivity (paternal trisomy)	
	HMM	SVM	HMM	SVM	HMM	SVM
1–5%	1.00	0.99	0.47	0.73	0.00	1.00
6–10%	1.00	1.00	0.96	1.00	0.00	1.00
11–15%	1.00	1.00	1.00	1.00	0.00	1.00
16–20%	1.00	1.00	1.00	1.00	0.00	1.00

RCAR, combined read count and allelic ratio model; SVM, support vector machine; HMM, hidden Markov model. Each value represents an average classification accuracy over 1,200,000 simulated cell-free DNA samples with fetal (maternally or paternally inherited) trisomy at the different sequencing depth intervals (500–15,000 reads). Adopted and modified from (Teder et al., 2019).

Table 10. Fetal trisomy classification accuracy with read count computational model and supplemental SVM.

Fetal fraction	Specificity				Sensitivity			
	Fetal fraction is calculated beforehand and provided as input		The model presumes a fetal fraction of 10%		Fetal fraction is calculated beforehand and provided as input		The model presumes a fetal fraction of 10%	
	HMM	SVM	HMM	SVM	HMM	SVM	HMM	SVM
1–5%	1.00	1.00	1.00	0.98	1.00	1.00	0.11	0.92
6–10%	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00
11–15%	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
16–20%	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

SVM, support vector machine; HMM, hidden Markov model. Each value is calculated over 1,200,000 simulated cell-free DNA samples with fetal trisomy at the different sequencing depth intervals (uniformly 500 - 15,000 reads). Adopted and modified from Teder et al., 2019.

RC model benefits from the SVM only if the FF is not known. If FF is known, then SVM does not affect sensitivity or specificity. If FF is unknown, then using SVM on low FF of 1–5%, sensitivity increases by 8-fold, and specificity decreases between 0.01 and 0.02 across all tested FF (**Table 10**).

The targeted strategy with HMM combined with supplemental SVM would work well, but the need for pre-sequencing loci selection does not allow straightforward NIPT panel expansion. Consequently, we transitioned to using WGS-based NIPT. Although WGS-based NIPT, when paired with publicly available computational analysis tools, enables comprehensive genome analysis, determining cost-effective sequencing depth while keeping clinically acceptable trisomy detection accuracy must be determined.

5.2 Paper II: Fetal trisomy risk screening in low-coverage WGS-based NIPT

While WGS-based NIPT is a powerful screening method with several published computational NIPT analysis tools available, there is no comprehensive accuracy comparison of these tools on the same clinically validated data. We applied computational NIPT aneuploidy analysis tools on the same clinically validated samples across various sequencing coverages and fetal DNA fractions to determine clinically usable sequencing depth(s) and computational tool(s).

5.2.1 Sequencing depth effect on trisomy detection (Paper II)

One central parameter in WGS is sequencing depth. To assess the sequencing depth effect on fetal trisomy risk assessment accuracy, we compared GIPseq, NIPTeR, NIPTmer, RAPIDR and WisecondorX on the same clinically validated NIPT samples.

Our findings indicated predominantly accurate results when sequencing coverages exceeded 5M reads per sample (RPS) (**Table 11**). In contrast, differentiation trends among compared algorithms became apparent at sequencing depths lower than 5M RPS, particularly with increased false-negative trisomy calls driven by systematically more conservatively estimated Z-scores (**Figure 9)**. For example, in one instance, on very low read coverage of 1.25M RPS, GIPseq detected 13 more cases of trisomy than RAPIDR (**Figure 10**).

Table 11. Number of undetected trisomy 13, 18 and 21 samples at 5M reads per sample per computational NIPT tool.

Number of cases	GIPseq	NIPTeR	NIPTeR NCV	NIPTmer	RAPIDR	WisecondorX
T21: 19	0	0	0	2	4	0
T18: 8	1	1	0	1	8	0
T13: 3	0	0	0	0	2	0

Combined from Paluoja et al., 2021.

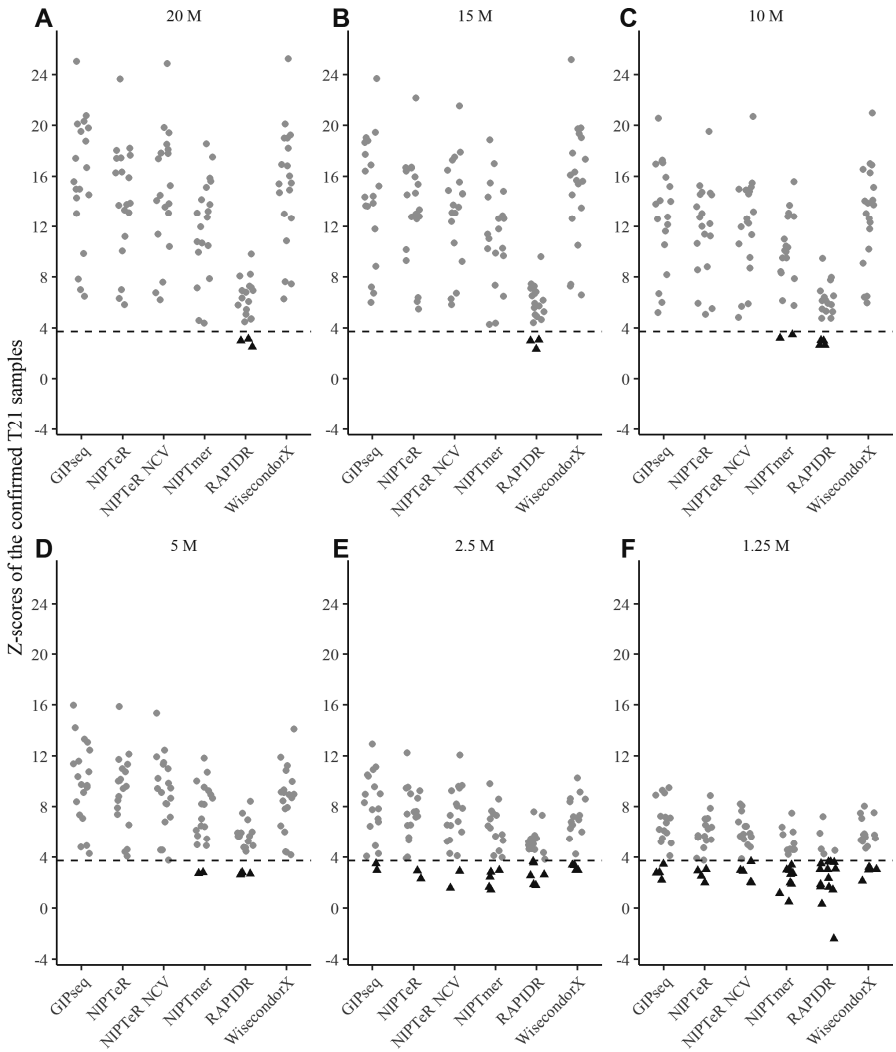


Figure 9. Z-scores of clinically validated T21 (trisomy 21) samples across a range of sequencing depths. Z-scores of known trisomy samples at sequencing depths of 20M RPS (A), 15M RPS (B), 10M RPS (C), 5M RPS (D), 2.5M RPS (E), and 1.25M RPS (F) are shown. Undetected (false negative) trisomies falling below the Zt cut-off thresholds (black dashed line in each graph) are represented as black triangles. Adopted from Paluoja et al., 2021.

Although systematically decreased Z-scores contribute to the rise in false-negative trisomy calls due to the distancing from the trisomy cut-off threshold, our observations also highlight that low sequencing coverages increase false-positive trisomy calls (**Figure 10B**). We demonstrated that lower trisomy detection accuracy at lower sequencing depths is influenced by the naturally occurring arbitrary placement of sequencing reads, introducing uncertainty in the Z-score estimation of the studied sample (**Figure 11)**.

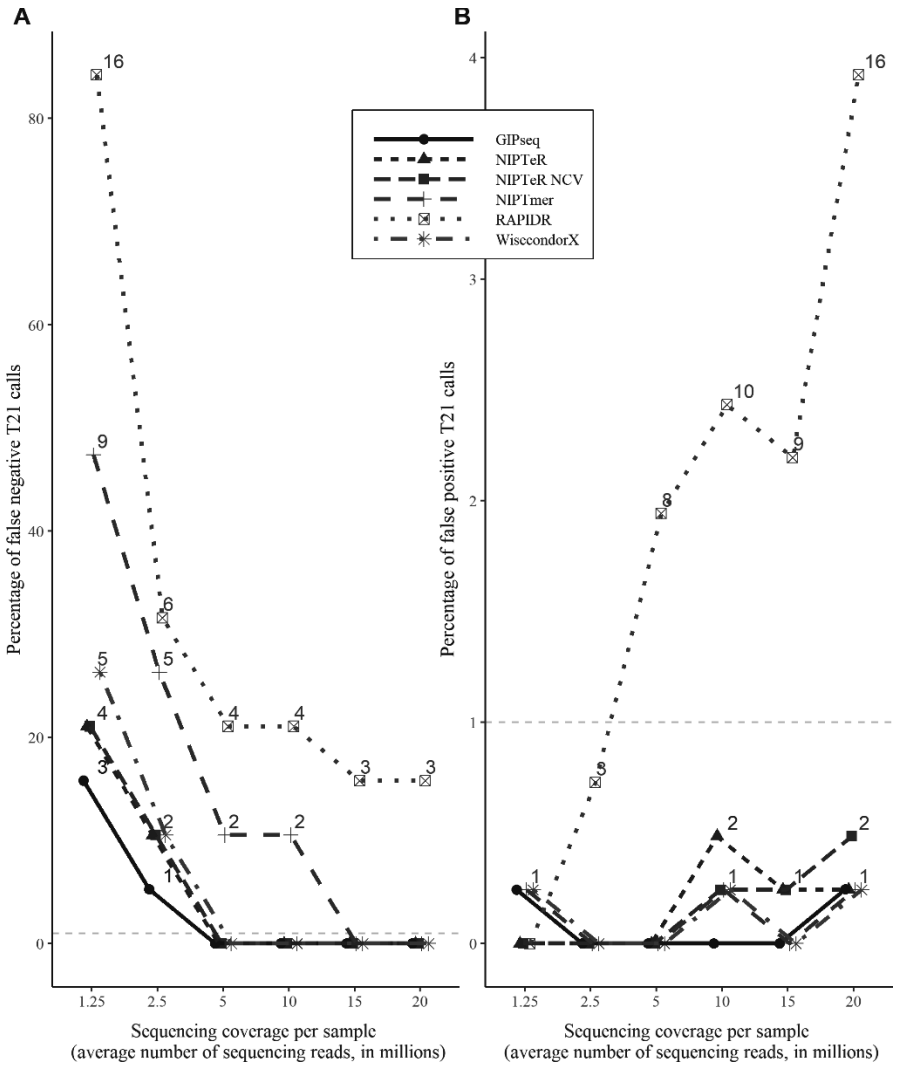


Figure 10. Trisomy detection accuracy of tested NIPT software tools across different sequencing depths. (A) Percentages and absolute numbers of undetected trisomy cases among known trisomies. (B) False-positive T21 results among known euploid samples. The horizontal dashed line in each graph marks the 1% cut-off level often used in clinical screening. Adopted from Paluoja et al., 2021.

Furthermore, we noted that when certain computational tools were applied to their own euploid reference group across different sequencing depths, the Z-score distribution deviated from the anticipated theoretical distribution, translating into a variance in the empirical cut-off threshold (**Figure 12**). For instance, without accounting for this variance, WisecondorX introduced two additional undetected trisomies at a sequencing depth of 2.5M RPS compared to the cut-off threshold considering this variance (**Figure 9E, Figure 10**).

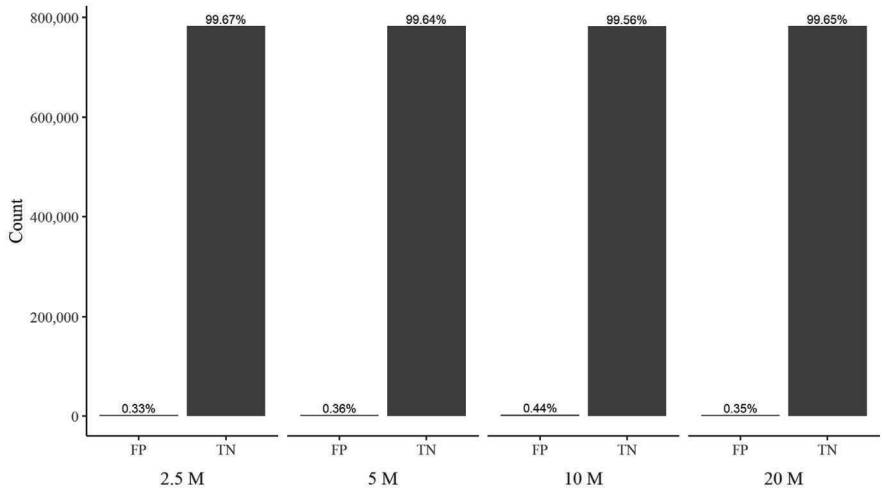


Figure 11. Effect of the natural sequencing read placement uncertainty on euploid sample group number of false-positive (FP) trisomy 21 cases. 0.33%–0.44% of Z-scores depending on the subsample group were detected as FP trisomy 21 cases. Adopted from Paluoja et al., 2021.

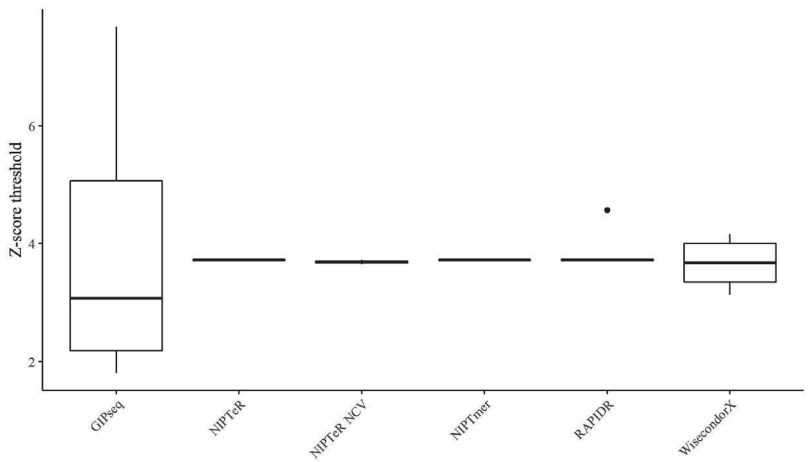


Figure 12. Chromosome 21 Z_e cut-off threshold variability**.** Adopted from Paluoja et al., 2021.

Finally, we also noted variations in accuracy specific to chromosomes and algorithms. For example, RAPIDR did not detect any of the trisomy 18 samples or NIPTeR and NIPTeR-NCV detected trisomy 21 cases equally accurately, but NCV showed better trisomy 18 detection accuracy compared to the NIPTeR in lower sequencing coverage conditions (**Table 11, Figure 13**).

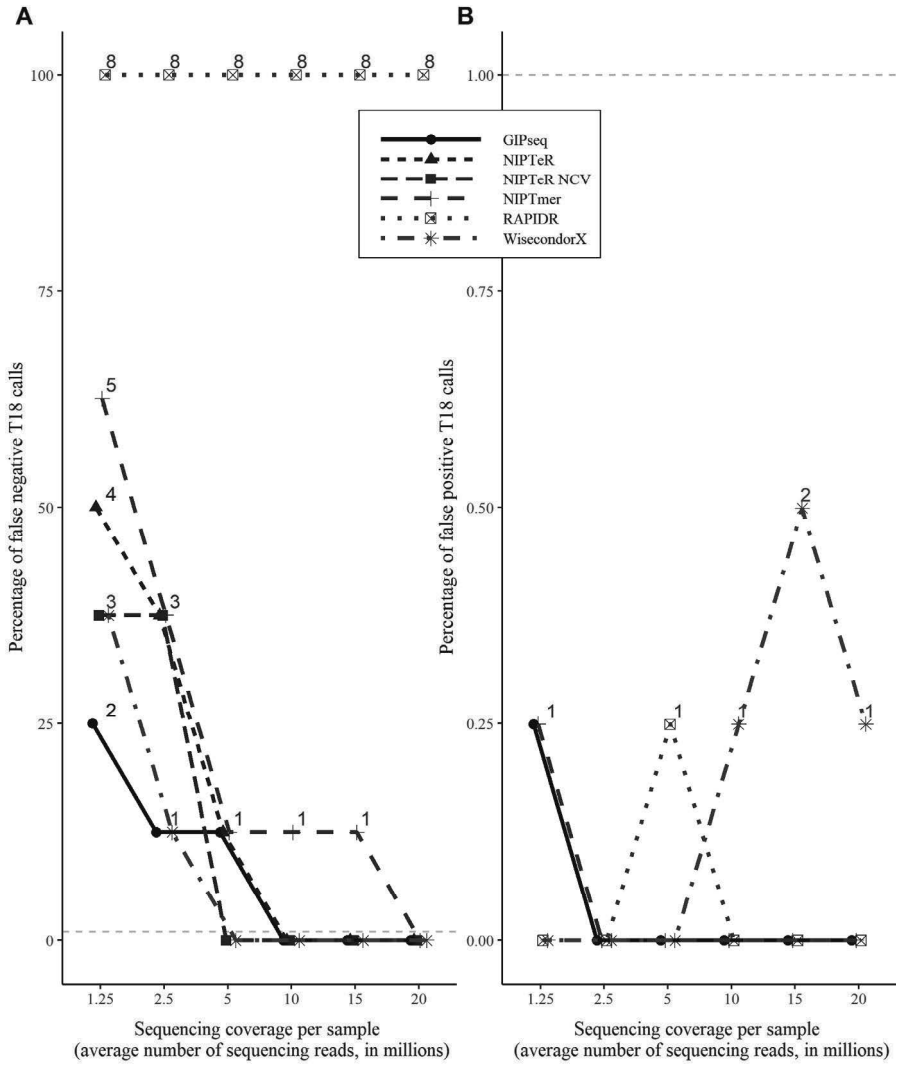


Figure 13. The percentage of false-negative (A) and false-positive (B) cases of trisomy 18 on different emulated sequencing coverages. (A) depicts percentages and the absolute number of false negative trisomy cases out of all known trisomy cases and (B) illustrates false-positive trisomy 18 calls out of all samples obtained with each NIPT software tool in case of various sequencing coverages. The horizontal dashed line marks the 1% cut-off, often used in case of clinical screening tests. Adopted from Paluoja et al., 2021.

5.2.2 Fetal DNA fraction effect on trisomy detection (Paper II)

As expected, we observed that fetal DNA fraction had a significant effect on trisomy detection accuracy, especially in the case of the lowest sequencing depth of 1.25M RPS, where all trisomy 21 samples with an FF under 7% remained

undetected (**Figure 14**). NIPTeR NCV and WisecondorX were the only tested tools with no false negative calls at extremely low FF of 3.65% at 5M RPS sequencing depth (**Figure 14, Table 11**). At 10M RPS with an FF of 3.65%, GIPseq and NIPTeR had no false negative trisomy calls. NIPTmer achieved a no-false-negative outcome with a 3.65% FF only at a sequencing depth of 20M RPS.

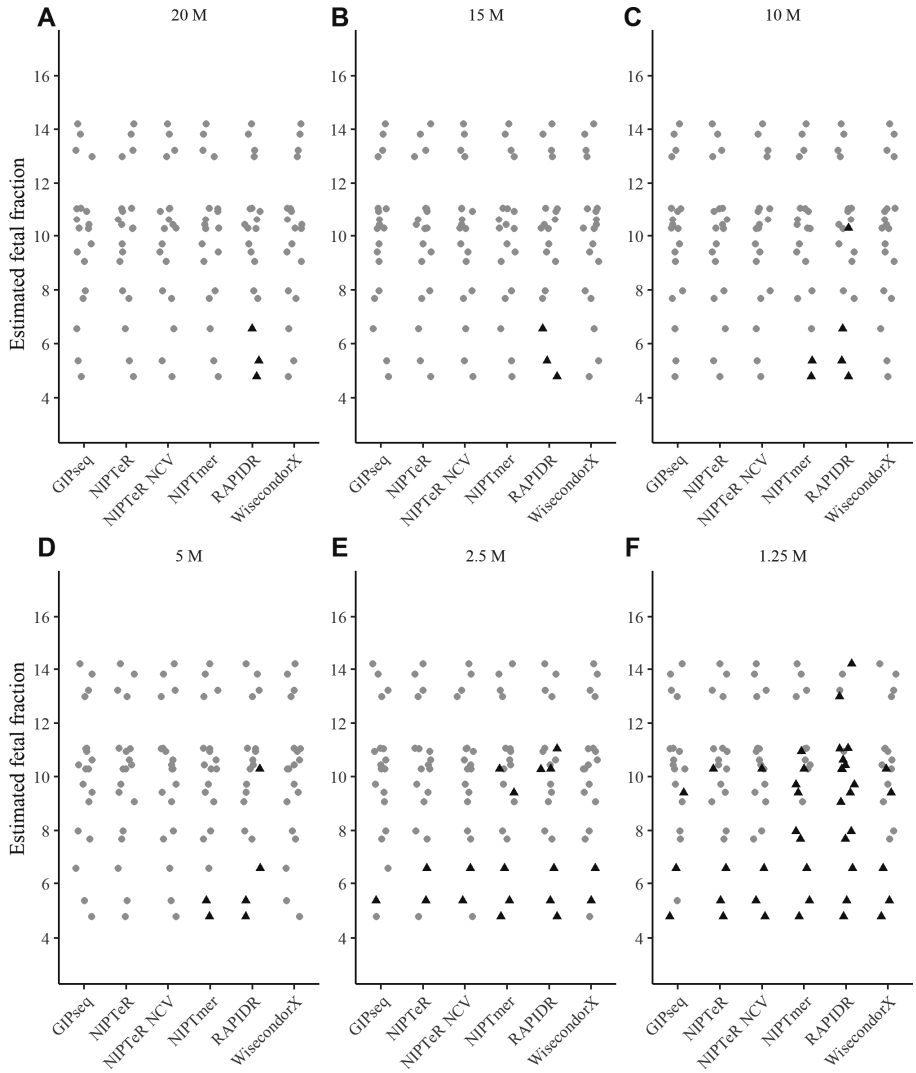


Figure 14. The effect of fetal DNA fraction on the detection of trisomy 21 across different sequencing coverages**.** Z_e cut-off was used for identifying the presence of the trisomy (internal classification in the case of GIPseq). Black triangles represent undetected trisomy cases. The 20M RPS group served as the standard for FF calculations. Data obtained with sequencing depths of 20M RPS (A), 15M RPS (B), 10M RPS (C), 5M RPS (D), 2.5M RPS (E), and 1.25M RPS (F) are shown. Adopted from (Paluoja et al., 2021).

We noted inconsistencies in SeqFF FF estimates at lower sequencing depths than estimates obtained in the reference coverage condition (~20M RPS) (**Figure 15**). For FFs ranging from 0% to 5%, the Pearson correlation value between 1.25M RPS FF estimates and corresponding 20M RPS FF estimates was only 0.217 (**Figure 15E**). At higher FF values (true FF, 5%–15%), the influence of sequencing depth was more subtle. For example, we obtained a correlation value of 0.959 between 10M RPS and 20M RPS but only 0.636 between 1.25M RPS and 20M RPS fetal fraction estimates. The observed correlation patterns between FF estimates at low sequencing depths and those at higher sequencing depths suggest a systematic tendency for FF to be overestimated under low-sequencing-depth conditions.

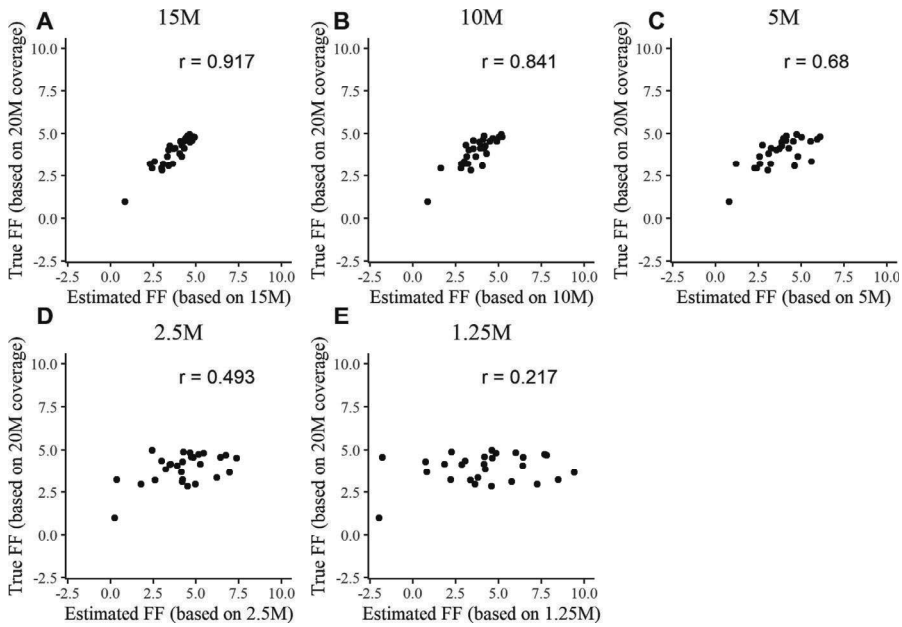


Figure 15. Correlation between the 'true' (based on 20M RPS data) and low sequencing depth based fetal fraction (FF) estimates. Pearson correlation data are shown for 20M RPS FF 0–5% estimates and estimates obtained at sequencing depths of 15M RPS (A), 10M RPS (B), 5M RPS (C), 2.5M RPS (D), and 1.25M RPS (E). Adopted from Paluoja et al., 2021.

WGS-based NIPT facilitates analysis beyond screening whole-chromosome aberrations. WGS allows examining any genomic region, including those spanning only a few million base pairs or less and recognised to be at risk of having pathogenic (micro)deletions that can lead to phenotypic severity comparable to trisomies. Unlike trisomy screening, with the availability of multiple computational tools, the options for microdeletion screening are scarce, and existing computational tool(s) are not validated on the same clinically confirmed NIPT fetal microdeletion samples under various fetal fraction and sequencing depth conditions.

5.3 Paper III: Fetal microdeletion risk screening in low-coverage WGS-based NIPT

Despite WGS-based NIPT covering the entire genome, only a fraction of the sequencing data originates from the fetal/placental source. Combining a limited amount of signal originating from the fetal/placental source with relatively short microdeletion regions and with the lack of microdeletion-screening-specific computational tools, we examined the feasibility of accurately estimating fetal microdeletion risk from low-coverage WGS NIPT data. We developed the necessary methodologies and a software tool, BinDel, with a microdeletion region-specific normalisation to assess and validate fetal microdeletion risk screening and compared it with a universal alternative computational tool.

5.3.1 Methodology and a software tool BinDel to estimate fetal microdeletion risk (Paper III)

In the clinical context, not all microdeletions are pathogenic. We concentrated solely on a predefined list of 12 microdeletion regions with known pathogenic clinical significance, as reporting variants of uncertain significance would lead to unnecessary invasive diagnostic procedures during pregnancy (**Figure 16**). Since clinically relevant pathogenic MDs typically have well-defined genomic coordinates, we designed the microdeletion risk detection algorithm to use these coordinates in risk detection, unlike heuristic/exploratory algorithms that do not consider and use external known biological information. Exploratory algorithms may not estimate MD risk if they interpret the MD as noise. However, the coordinate-based approach guarantees that the target MD region always receives an MD risk estimate.

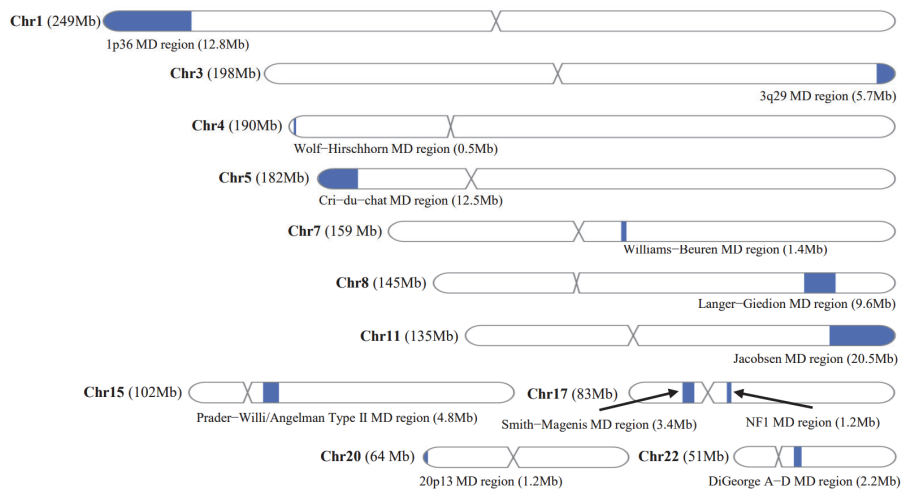


Figure 16. A priori chosen clinically relevant microdeletions (MDs) and their locations on the genome. Microdeletion regions are indicated with blue sections on the chromosomes. Adapted from Paluoja et al., 2025.

As we designed the BinDel for WGS data, which in NIPT relies on dividing sequencing reads to genomic bin counts, we made the genomic bins individually configurable as microdeletions are often in locations that are challenging to analyse. After the division of aligned sequencing reads into a priori selected and named genomic bins (certain bins denote MD regions, others non-MD regions), the genomic bins undergo GC% correction and are further normalised based on the total sequencing read count and bin lengths. The normalised values are divided by PCA dimensionality-reduced versions. Next, a normalised Z-score for each genomic bin using the euploid fetus reference set is calculated and aggregated per target region (see **Materials and Methods**). The Mahalanobis distance for each MD region from the euploid fetus reference group is calculated using the aggregated features, and for uniform interpretation between different genomic regions, the distances are transformed into chi-squared distribution p-values and subsequently -log10 transformed. This step restricts the estimated MD risk output to a range between 0% and 100%, ensuring a consistent interpretation of MD risk levels.

5.3.2 Calibrating BinDel (Paper III)

When developing BinDel, we observed that the number of PCA components used in the normalisation process affects the sensitivity and specificity of MD risk detection. To quantify the PCA effect on specificity and sensitivity, we used euploid and simulated fetal microdeletions as outlined in the **Materials and Methods** section. We observed that a higher number of PCA components (PCA99%) increased MD detection sensitivity, particularly in short MD regions like NF1 (1.2Mb) (**Figure 17A1**). However, this does not apply universally across all MD regions. For example, in the Williams-Beuren MD region, the overall sensitivity and specificity were either equal or slightly lower compared to the lower number of components setting (PCA95%) (**Figure 17B1, B2**). It is also important to note that using a higher number of PCA components also increased false-positive MD risk calls in certain regions, such as DiGeorge A-D (**Figure 17A2, C2**).

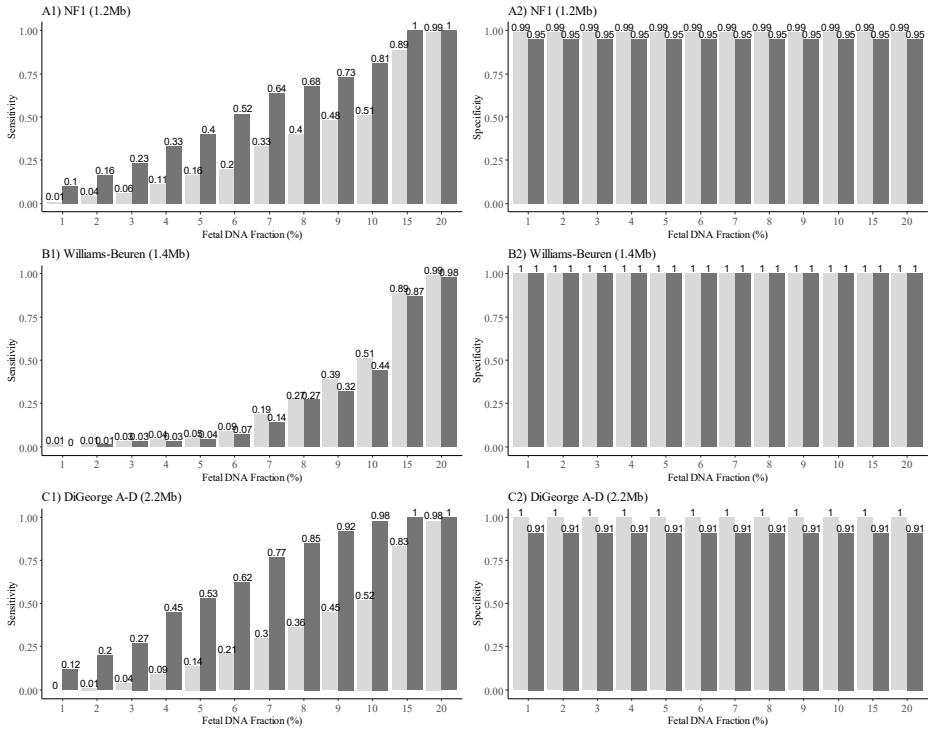


Figure 17. BinDel detection sensitivity and specificity with *in silico* simulated microdeletions using different numbers of PCA components. Sensitivity estimates are indicated by grey and dark grey bars, denoting PCA95% and PCA99%, respectively. The analysis focuses on three MD regions: NF1 (A1, A2), Williams-Beuren (B1, B2), and DiGeorge A-D (C1, C2) associated microdeletion regions. Adapted from Paluoja et al., 2025.

5.3.3 Fetal fraction and region length effect on BinDel performance using computationally simulated fetal microdeletions (Paper III)

We investigated the impact of FF on MD calling accuracy with computationally simulated fetal microdeletions in 1p36, 20p13del, 3q29, Cri-du-chat, DiGeorge, Jacobsen, Langer-Giedion, NF1, Prader-Willi/Angelman, Smith-Magenis, Williams-Beuren and Wolf-Hirschhorn MD regions.

As anticipated, higher FF considerably increased the MD detection sensitivity (**Figure 18A-D, relates to Figure 17**). However, notably, the observed patterns of increased sensitivity exhibited variations among distinct MD regions. For example, when comparing the Williams-Beuren or NF1 MD region to the NF1 MD region, they exhibited different MD risk-calling across all FF levels despite having a similar MD length (**Figure 18A-D)**.

Furthermore, longer MD regions exhibited higher MD risk detection sensitivity, particularly in cases with low FF (**Figure 18A**). However, this trend did not linearly persist across all MD regions. For instance, upon comparing the Langer-Giedion and 1p36 MD regions, the longer 1p36 MD region exhibited lower sensitivity than the shorter Langer-Giedion MD region (**Figure 18A**).

Based on our microdeletion simulation detection results, as a general guideline, screening accuracy notably decreased when the fetal DNA fraction fell below 10% and the microdeletion region length was under 5 Mb (**Figure 18C**). In a clinical setting, these microdeletion screening limits would be influenced by various factors, including sequencing depth, usage of fetal fraction enrichment, and the strategies used to balance false-positive and false-negative microdeletion calls.

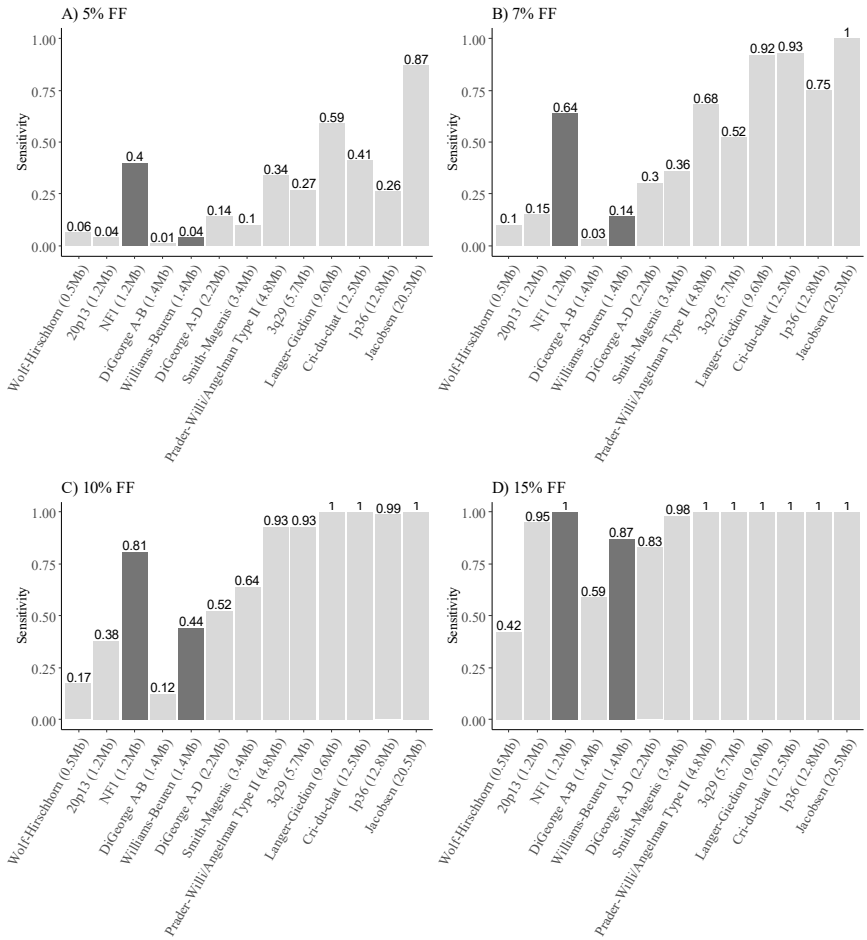


Figure 18. Microdeletion associated region length effect on the BinDel sensitivity of detecting microdeletion risk. Fetal DNA fractions of 5% (A), 7% (B), 10% (C) and 15% (D) were used for *in silico* simulated microdeletions. Sensitivity estimates are indicated by grey and dark grey bars, denoting PCA95% and PCA99%, respectively. Adapted from Paluoja et al., 2025.

5.3.4 Evaluation of BinDel with clinically validated microdeletion samples (Paper III)

We conducted an analysis with BinDel for 84 samples containing 34 clinically confirmed MD and 50 confirmed euploid NIPT samples. These 34 samples had microdeletions in DiGeorge, Prader-Willi/Angelman, Smith-Magenis, Williams-Beuren and NF1 MD regions. BinDel detected 73.5% (25 out of 34) clinically validated microdeletions (**Table 12**).

However, as also in clinical practice, microdeletions are always not fetal-only, as is also the case with the validation sample set. When excluding maternal and fetal or one solely maternal deletion and considering samples only with fetal microdeletions, BinDel correctly detected 69% (19 out of 28) of fetal MD samples (**Table 12**).

We observed that samples missed with the microdeletions were either in the Williams-Beuren microdeletion region or had an average lower FF (8.9%) compared to those where fetal MDs were correctly called (10.4%).

BinDel had 15 false-positive microdeletion calls, whereas the majority of them were concentrated in a few challenging regions, like the NF1 microdeletion region (**Table 12**).

To assess the BinDel results in context, we also inferred MD calls with WisecondorX, a universal WGS copy number detection tool intended for broader genomic applications (Raman et al., 2019). WisecondorX detected 47% (16 out of 34) MD samples or 35.7% (10 out of 28) fetal-only MDs (**Table 12**). WisecondorX had no false-positive MD calls.

Table 12. The number of true positive (TP), false positive (FP), false negative (FN), and positive predictive values (PPV) of microdeletion high-risk calls of BinDel and WisecondorX software packages on clinically confirmed 34 microdeletion and 50 euploid NIPT samples.

Microdeletion region, origin and number of samples (n)	BinDel				WisecondorX			
	TP	FP	FN	PPV	TP	FP	FN	PPV
DiGeorge, fetal, n = 17	12	0	5	1	4	0	13	1
DiGeorge, maternal and fetal, n = 4	4	0	0	1	4	0	0	1
DiGeorge, maternal, n = 1	1	0	0	1	1	0	0	1
Prader-Willi/Angelman, fetal, n = 5	5	2	0	0.71	4	0	1	1
Smith-Magenis, fetal, n = 3	2	0	1	1	2	0	1	1
Williams-Beuren, fetal, n = 3	0	3	3	0	0	0	3	0
NF1, maternal and fetal, n = 1	1	10	0	0.09	1	0	0	1

NF1, neurofibromatosis 1; Adapted from Paluoja et al., 2025.

5.3.5 Improving BinDel and reanalysing with clinically validated microdeletion samples (Paper III)

The analysis revealed two significant concerns. First, there was an unsuitable number of false-positive MD calls in NF1 and Williams-Beuren MD region for clinical usage, where we used PCA99% combined with a lower cut-off threshold (see **Materials and Methods**) (**Table 12**). Secondly, no true-positive MD calls were found in the Williams-Beuren MD region (**Table 12**).

Based on these insights, we systematically re-evaluated the BinDel data normalisation and MD calling process to improve region-specific microdeletion risk detection. First, we established if the high-risk MD cut-off threshold was optimal and if conservative PCA95% would yield higher specificity without compromising sensitivity. Our investigation revealed that elevating the cut-off threshold for calling MD across all microdeletion regions to 90% and using PCA95% reduced the total number of false-positive MD calls from 15 to 1 (false positive rate of 0.02) without affecting the number of true-positive MD calls (**Table 13**). The single false-positive MD call was in the Prader-Willi/Angelman MD region.

Table 13. Microdeletion cut-off threshold effect on BinDel accuracy.

Microdeletion high-risk cut-off threshold (%)	TP	FP	FN	PPV
10	27	19	7	0.59
20	26	9	8	0.74
30	25	5	9	0.83
40	25	5	9	0.83
50	25	4	9	0.86
60	25	3	9	0.89
70	25	3	9	0.89
80	25	3	9	0.89
90	25	1	9	0.96

The number of true-positive (TP), false-positive (FP), false-negative (FN), and positive predictive values (PPV) of microdeletion high-risk calls by BinDel with a PCA95% on different microdeletion high-risk cut-off thresholds. Adapted from Paluoja et al., 2025.

However, while using PCA95% and a uniform cut-off threshold of 90% across all MD regions lowered the false positive rate to 0.02, it did not resolve the absence of correct microdeletion detections in the Williams-Beuren MD region. As also indicated by the simulations and witnessed by Tian *et al.*, the Williams-Beuren MD region shows a lower microdeletion detection rate than other microdeletion regions with similar lengths (**Figure 18)** (Tian et al., 2023).

Observing the non-uniform MD detection sensitivity across MD regions of comparable sizes, we implemented a region-specific novel methodology (**Table 12, Figure 18**). Building upon our prior observation that PCA normalisation across all genomic bins increased MD detection sensitivity, we subsequently re-implemented the same normalisation technique. However, instead of uniformly applying it to all genomic bins, after the initial PCA normalisation, we applied PCA normalisation once more, but on a per MD region basis, factoring in the cumulative variance within each specific MD region (**Figure 19**). We also considered MD region risk score distribution to set the per-MD region cut-off threshold. We measured BinDel accuracy on validation sample set with combinations of PCA95%, PCA99%, cut-offs 1%-100% and cumulative regional PCA of 5%, 10%, 15%, 25% and 50% (only for PCA95%).

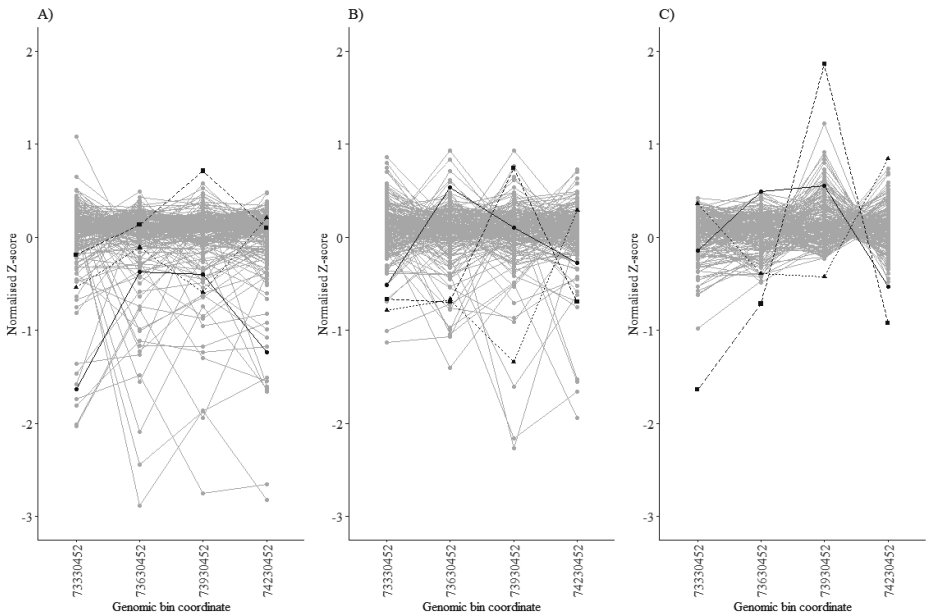


Figure 19. BinDel normalisation effects on genomic bin values in Williams-Beuren microdeletion region. (A) depicts normalised Z-scores (see **Methods**) without utilising the PCA normalisation method. (B) illustrates the impact of PCA95% normalisation on normalised Z-scores. (C) displays the combined effects of PCA95% and regional PCA (cumulative variance of 50%) normalised Z-scores. Each line in the figure represents a sample, where grey lines denote euploid reference samples, and black lines represent validation samples with fetal genomic microdeletion in the Williams-Beuren syndrome genomic region. Adapted from Paluoja et al., 2025.

The novel MD-region-specific approach detected 2 out of 3 samples with a fetal deletion in the Williams-Beuren MD region. Furthermore, this approach overall detected 30 out of 34 (88.2%) MD samples with a total of only three false positive MD calls, marking a 14.7% increase in detection sensitivity compared to the initial validation analysis (**Table 12, Table 14)**. Notably, the two false-positive MD calls emerged in the Williams-Beuren MD region (**Table 14)**. In a clinical setting, error-prone microdeletion regions with high false-positive rates, such as the Williams-Beuren MD region, can be excluded from the analysis. Instead, the focus could be on more common microdeletions, like those associated with the DiGeorge and NF1 syndromes, where BinDel demonstrated no false-positive calls (**Table 12**, **Table 14**). We observed that the average FF for four samples not correctly identified as MD samples was 8% compared to 10% for 30 correctly identified MD samples.

Table 14. The number of true-positive (TP), false-positive (FP) and positive predictive values (PPV) of microdeletion calls by BinDel with a region-specific approach.

Microdeletion, origin and number of samples (n)	Microdeletion high-risk cut-off threshold (%)	Cumula-tive PCA (%)	Cumulative regional PCA (%)	TP	FP	PPV
DiGeorge fetal, fetal and maternal, maternal n = 22	9	95	Not used	19[a]	0	1
Prader-Willi/ Angelman fetal n = 5	87	95	Not used	5	1[d]	0.8
Smith-Magenis fetal n = 3	4	95	Not used	3[b]	0	1
Williams-Beuren fetal n = 3	10	95	50	2[c]	2[f]	0.5
NF1 maternal and fetal n = 1	87	95	Not used	1	0[e]	1

The number of true-positive high-risk microdeletion calls increased compared to the initial analysis in DiGeorge[a], Smith-Magenis[b] and Williams-Beuren[c] microdeletion regions. False-positive microdeletion calls are reduced in Prader-Willi/Angelman[d], NF1[e] and Williams-Beuren[f] microdeletion regions. Adapted from Paluoja et al., 2025.

6. DISCUSSION

In computational NIPT, algorithms necessitating training (distinct from reference set requirements) are less commonly used due to the need to obtain a well-balanced set of positive and negative trisomy controls necessary for unbiased training. This challenge is further compounded by the high dimensionality of model features resulting from the multitude of genomic loci involved, requiring a substantial-sized training set. However, the trainability aspect enables the estimation of cut-off points and the determination of sample states based on learned data distributions, eliminating the need to devise trisomy-calling rules manually.

Interestingly, there is a scarcity of readily available, pre-packaged computational NIPT tools in the public domain for targeted NIPT. While some, such as FORTE or NATUS, are described, they are not readily available or pre-packaged for use in the public domain (Norwitz & Levy, 2013). Due to this scarcity, developing and implementing a computational framework for targeted NIPT became necessary. An optimal methodology should ideally be without excessive complexity while maintaining high accuracy. While the HMM approach displayed promise (**Results section Analytical framework for HMM output interpretation (Paper I)**) and alone did not have excessive complexity and was not affected by the dimensionality of the input, it did not have high accuracy in all cases due to the criterion used to interpret HMM output, which relied solely on the locus state with the highest frequency (trisomy or euploid). Introducing other available classification algorithms, such as random forests instead of HMM, was not considered due to their possible susceptibility to high dimensionality (the number of loci involved). In the clinical setting, prioritising simplicity over accuracy is not tenable, so we introduced a second classification layer. Since the two-level framework supported training with simulated data and the HMM reduced the dimensionality to only locus counts associated with euploid, maternally, and paternally inherited trisomies, we opted for SVM as it is known to work well when there is a clear distinction between different states. However, other machine learning-based classifiers, such as random forests or nearest neighbours, could also work well due to different states' low dimensionality and clear distinction. Combining HMM with supplementary SVM showed potential (**Results section Evaluation of the analytical framework of HMM output interpretation (Paper I)),** but the need for pre-sequencing loci selection hindered the straightforward expansion of the NIPT panel. Consequently, we shifted our focus to utilising WGS-based NIPT as a more viable alternative.

Interestingly, unlike in targeted NIPT, in the WGS-NIPT, numerous software options were readily accessible for WGS-based NIPT, obviating the need to develop and implement a computational tool. However, no study existed that had evaluated these available computational tools on the same clinically validated data. We conducted a comparative study to make an informed decision in choosing the most accurate computational WGS-NIPT tool. While prioritising accuracy

as the primary criterion, the tool needed to exhibit precision even at lower sequencing coverage levels, as the sequencing depth directly influences the cost-effectiveness of NIPT. We saw that the algorithm's choice significantly affects the number of false-negative trisomy calls, especially at sequencing depths lower than 5M RPS (**Results section Sequencing depth effect on trisomy detection (Paper II)**). This is somewhat expected, considering that the tested tools operate based on partitioning sequencing reads into genomic bins (or counting k-mers), followed by comparative analysis with the euploid sample (reference) group bins. In instances of reduced coverage, the discrepancy between euploid bins and reference bins diminishes, resulting in more conservative trisomy risk scores (**Results section Sequencing depth effect on trisomy detection (Paper II)**). Unexpectedly, inconsistencies in SeqFF FF estimates were also observed at lower sequencing depths (**Results section Fetal DNA fraction effect on trisomy detection (Paper II)**), particularly in scenarios where the true FF was low, despite these findings aligning with logical expectations given that SeqFF is pre-trained on sample-set with higher sequencing coverage (Kim et al., 2015). Although FF, sequencing depth, and algorithm selection are typically the primary factors influencing trisomy screening accuracy, we also observed variations in the distributions of outputted Z-scores among software tools when these tools were employed on euploid reference set samples across different sequencing depths. Variations in Z-score distributions across different sequencing coverages imply that employing a single, uniform cut-off threshold independent of the specific tool may diminish the accuracy of trisomy detection. This discrepancy was notably apparent in one instance involving the WisecondorX tool by introducing two undetected trisomies at a sequencing depth of 2.5M RPS (**Results section Fetal DNA fraction effect on trisomy detection (Paper II)**). Interestingly, variations in Z-score distributions were observed only with some tools, implying differences in the internals of the algorithms related to Z-score calculations and the steps preceding. Overall, this study allowed us to make an informed decision in choosing NIPTeR (NCV) as the optimal trisomy screening algorithm by considering the overall accuracy under different sequencing depths, fetal fractions, and software licensing, also allowing non-scientific use.

NIPTeR, however, is not designed to infer microdeletion risk scores (Johansson et al., 2018). While over 50 copy number variation tools exist, the number of tools designed for NIPT is minimal (Gabrielaite et al., 2021). For example, Wisecondor and WisecondorX were among the few readily available and prepackaged tools available at the time of writing that could infer fetal aberrations, including microdeletions from WGS-based NIPT (Raman, Dheedene, et al., 2019; Straver et al., 2014). However, these tools are not explicitly designed for fetal microdeletion screening and use a heuristic approach in screening for fetal aberrations. Moreover, existing studies examining the impact of factors like FF or MD region length on microdeletion screening, as demonstrated by Kucharik et al., are few and constrained to simulations, mixing DNA samples, or do not provide readily available computational tools for public use (Kucharik et al., 2020). Considering the lack of a dedicated low-coverage WGS-based NIPT fetal

microdeletion computational screening tool, we developed BinDel (**Results section Evaluation of BinDel with clinically validated microdeletion samples (Paper III)**).

Given the infrequency of pathogenic microdeletions, conducting an extensive study using naturally occurring fetal microdeletions (similar to our approach with trisomies) is challenging due to the rarity of such samples. For instance, in 2024, Estonia reported 9,523 live births (Eesti Meditsiiniline Sünniregister, 2025). This scarcity indicates the occurrence of only a few fetal microdeletions annually in such a population context (e.g. DiGeorge syndrome occurring in approximately 1 in 4,000 births) (McDonald-Mcginn & Sullivan, 2011). Furthermore, only NIPT samples procured and sequenced during the first trimester are suitable for computational research. This criterion introduces complexities, such as alterations in laboratory protocols that can render previously processed samples unusable for research purposes. Moreover, our sample set highlighted the diversity of microdeletions, exemplified by the presence of various forms. For instance, within our validation set, we encountered fetal-only, fetal and maternal, and a single maternal-only microdeletion or 22q11.2 (DiGeorge MD region) deletion occurring as forms of A-B and A-D. Additionally, our observations indicated differences for microdeletions in similar length. For example, the Williams-Beuren MD region exhibited a lower detection accuracy compared to other regions of similar length, such as the NF1 MD region (**Results section Evaluation of BinDel with clinically validated microdeletion samples (Paper III)**). We also observed that detection accuracy is influenced by the fetal fraction (FF). Our simulation results showed a significant decline in detection accuracy when the FF dropped below 10% and the affected region was shorter than 5 Mb. In a real-world clinical setting, increasing FF through laboratory enrichment techniques and retesting with a new sample in cases of elevated MD risk may help reduce false-positive calls and minimise the need for invasive diagnostic procedures. Furthermore, MD regions with high false-positive rates or low population prevalence may be excluded from routine NIPT analysis. Instead, clinical screening could focus on more common microdeletions, such as the DiGeorge MD region, for which BinDel showed no false positives and demonstrated significantly higher sensitivity compared to WisecondorX. These findings, supported by our validation study, underscore the feasibility of detecting MD syndrome risk through NIPT and highlight its potential to broaden the spectrum of fetal genetic conditions detectable in the first trimester.

This study was limited by the number of clinically validated samples with pathogenic microdeletions and trisomies, particularly for certain regions such as those associated with Williams-Beuren or NF1 syndromes, and for some aneuploidies like trisomy 13. Additionally, we were constrained to low sequencing coverage to maintain the cost-effectiveness of NIPT as a broadly accessible screening method. However, lower coverage reduces the sensitivity of computational tools, especially in cases with low fetal fraction. Finally, there is a lack of publicly available NIPT-specific microdeletion screening tools for comparison.

While tools like WisecondorX are available, they are not explicitly designed for detecting pathogenic microdeletion risk.

Overall, while frequently understated, the algorithms deployed significantly influence the accuracy and cost-effectiveness of NIPT.

7. CONCLUSION

Aim 1: First, we developed computational methods for targeted NIPT-based trisomy risk detection and systematically evaluated trisomy risk calling, as the majority of available computational methods for sequencing-based NIPT rely on WGS and cannot be used for targeted NIPT. The developed computational method relied on HMM incorporated with an analytical framework for HMM output interpretation using SVM. Notably, the supplemental SVM elevated paternally inherited trisomy detection sensitivity from zero to one. For some HMM-based models, the SVM also considerably improved the detection of maternally inherited trisomy, with a specificity decrease between 0.01 and 0.03 in some cases.

Aim 2: We determined how sequencing depth and estimated FF affect trisomy risk calling in low-coverage WGS-based NIPT by assessing the accuracy of WisecondorX, NIPTeR, NIPTmer, RAPIDR, and GIPseq trisomy risk calling across various sequencing depths and fetal DNA fractions. While all of the compared tools were affected by an increased number of undetected trisomy cases when sequencing depth was lowered, our findings generally indicated accurate results when sequencing depth exceeded 5M reads per sample. However, the least accurate algorithm would miss almost one-third of trisomy cases at lower sequencing depths, such as at and below 2.5M reads per sample. We also described previously undocumented aneuploidy risk uncertainty that can increase undetected trisomy cases. Finally, as expected, we observed that FF significantly affected trisomy detection accuracy. For instance, at the lowest sequencing depth of 1.25M RPS, all trisomy 21 samples with an FF under 7% remained undetected.

Aim 3: We investigated the feasibility of estimating fetal microdeletion risk from low-coverage WGS NIPT data. We developed the necessary methodologies, including BinDel, a software tool. Using simulated fetal microdeletions, we observed that higher FF considerably increased the MD detection sensitivity in 12 clinically pathogenic microdeletion regions, including 1p36 deletion, 3q29, Wolf-Hirschhorn, Cri-du-chat, Williams-Beuren, Langer-Giedion, Jacobsen, Prader-Willi/Angelman, Smith-Magenis, neurofibromatosis type 1, 20p13 and DiGeorge syndrome related microdeletion regions. While longer MD regions exhibited higher MD risk detection sensitivity, there were variations in detection sensitivity among distinct MD regions despite having a similar MD length. Finally, in the analysis of clinically confirmed 34 microdeletions and 50 euploid fetal samples, BinDel correctly identified 25 samples with microdeletions. However, 15 false-positive microdeletion calls were concentrated in a few challenging regions. The number of false-positive microdeletion calls led us to improve BinDel. After the improvement in the algorithm, 30 microdeletion samples were correctly

determined, with a total number of false-positive microdeletion calls reduced to three. The improvement confirmed BinDel's usefulness for fetal microdeletion risk screening in NIPT, opening the possibility of integrating BinDel successfully into routine NIPT protocol.

SUMMARY IN ESTONIAN

Arvutuslikud metoodikad loote trisoomiate ja mikrodeletsioonide riski tuvastamiseks mitteinvasiivses sõeluuringus

Mitteinvasiinve sünnieelne loote DNA uuring ehk NIPT (*non-invasive prenatal testing*) on sõeluuringu meetod, mis võimaldab ohutut ja täpset sünnieelset mitteinvasiivset kromosoomhaiguste riski hindamist. NIPT põhineb lapseootel naise veres ringleva loote (platsenta) rakuvaba DNA laboratoorsel ja arvutuslikul analüüsil. Kuigi loote (platsenta) DNA esinemine raseda vereproovis avastati juba 1997. aastal, siis aastatel 2011–2012 publitseeriti NIPT esimesed suuremahulised valideerimisuuringud (Bianchi et al., 2012; Lo et al., 1998; Palomaki et al., 2011).

Lisaks loote kromosoomi koopiaarvu muutustele võimaldab NIPT uuring tuvastada mikrodeletsiooni riski. Mikrodeletsiooni sündroom on lühikese kromosoomiosa kaost põhjustatud kromosoomihaigus, mille kliiniline raskusaste sõltub deleteerunud regioonist. Näiteks üks sagedane mikrodeletsioon on 22q11.2 heterosügootne deletsioon, mis põhjustab DiGeorge'i sündroomi, mida iseloomustavad mitmesugused kaasasündinud arenguhäired, nagu südamerikked või vaimupuue (McDonald-Mcginn & Sullivan, 2011).

Võtmeroll NIPT uuringus on DNA sekveneerimisandmete analüüsi arvutuslikel algoritmidel, mis otseselt väärindavad NIPT andmeid.

Uurimistöö eesmärgid

1. Töötada välja ning süstemaatiliselt hinnata arvutuslike metoodikate täpsust loote triosoomia riski määramisel suunatud NIPT sekveneerimisanalüüsi tarbeks.
2. Hinnata loote trisoomia riski määramise täpsust madala katvusega ülegenoomse NIPT sekveneerimisanalüüsi jaoks. Teha kindlaks sekveneerimissügavuse ja loote DNA fraktsiooni mõju trisoomiariski määramise täpsusele.
3. Uurida loote patogeensete mikrodeletsioonide riski määramise võimalikkust madala katvusega NIPT sekveneerimisanalüüsi tarbeks. Töötada välja vajalikud metoodikad ja neid rakendav tarkvara loote mikrodeletsiooniriski täpseks hindamiseks ja valideerida täpsus kliiniliselt kinnitunud mikrodeletsiooni proovidega.

Materjal ja metoodika

Esimeses uurimuses välja töötatud peidetud Markovi mudeli (HMM) põhinev arvutuslik metoodika loote trisoomia tuvastamiseks kõrge katvusega sekveneeritud NIPT tarbeks oli paljulubav, ent määras loote seisundi (euploidne, isa või ema päritolu trisoomia) kõige sagedasema HMM poolt klassifitseeritud lookuse oleku (euploidne, isa või ema päritolu trisoomia) järgi. Selline trisoomia riski määramine ei suutnud tuvastada isapoolse päritoluga trisoomiat ja ei arvestanud

keerukamate jaotustega (piirjuhud, kus euploidseks ja trisoomiaks klassifitseeritud lookuste arv oli peaaegu võrdne). Uurimuse raames töötati välja arvutusliku raamistik adresseerimaks HMM limitatsioone kasutades simuleeritud raseda rakuvaba DNA andmeid, mis sisaldasid nii isapoolse kui ka emapoolse päritoluga loote trisoomiat arvestades erinevad lootefraktsioone.

Teises uurimuses määrati WisecondorX, NIPTeR, NIPTmer, RAPIDR ja Leuveni Ülikooli analüütiliste töövoogude täpsus samal kliiniliselt valideeritud andmestikul teisendatult erinevale sekveneerimise sügavustele (katvustele). Kliiniliselt valideeritud kohort koosnes 423 raseda (üksikrasedus) sekveneeritud rakuvaba DNA proovist, mis oli eraldatud veeniverest. Sekveneerimisandmed sisaldasid 19 Downi, kaheksa Edwardsi ning kolme Patau sündroomiga juhtu. Teisendatud sekveneerimise katvustega NIPT andmete komplektid sisaldasid 1.25–20 miljonit lugemit patsiendi NIPT sekveneerimisandmete kohta. Ühtseks analüüsi tulemuste tõlgendamiseks defineeriti protsendipunkti funktsioonil põhineva raamistik trisoomia määramiseks. Määrati rakuvaba loote DNA hulga mõju analüütiliste töövoogude täpsusele.

Kolmandas uurimuses töötati välja bioinformaatiline analüüsitarkvara BinDel määramaks loote patogeenseid kliiniliselt olulisi mikrodeletsioone madala katvusega kogu genoomi NIPT sekveneerimisandmetel. BinDel rakendab mitmeid sekveneerimise lugemite normaliseerimise etappe, uudset algoritmilist mikrodeletsioonide signaali võimendamist, sh kaheastmelist peakomponentide analüüsil põhinevat normaliseerimist, suunatud analüüsi ning riski tõenäosuse arvutamist Z-skoori asemel. Määrasime rakuvaba loote DNA fraktsiooni ja mikrodeletsiooni pikkuse mõju mikrodeletsioonide tuvastamise täpsusele kasutades bioinformaatiliselt indutseeritud mikrodeletsioonidega NIPT proove. Koostöö Belgias asuva ülikooli Katholieke Universiteit Leuven'iga valideeriti BinDel 34 teadaoleva kliiniliselt kinnitunud patogeense mikrodeletsiooniga ja 50 euploidse NIPT proovi vastu. Kohort sisaldas DiGeorge, Angelman/Prader-Willi, neurofibromatoos tüüp 1 (NF1), Smith-Magenis ja Williams-Beureni sündroomiga proove.

Uurimistöö kokkuvõte ja järeldused

Esimese uurimuse tulemusena töötati välja kaheastmeline arvutuslik metoodika trisoomiariski tuvastamiseks suunatud NIPT sekveneerimisanalüüsi tarbeks. Väljatöötatud arvutuslik metoodika tugines peidetud Markovi mudelil (HMM), mille väljundit interpeteeriti tugivektor-masinaga (SVM). Täiendav SVM rakendatuna HMM mudelile tõstis isalt päritud trisoomia tuvastamise tundlikkuse nullilt üheni. Teatud HMM-põhiste mudelite puhul parandas SVM märkimisväärselt ka emalt päritud trisoomia tuvastamist vähese spetsiifilisuse langemisega.

Teise uurimuse tulemusena tuvastati analüütilised töövood, mille minimaalne sekveneerimise katvus ilma tuvastamata trisoomiateta oli viis miljonit lugemit proovi kohta. Lisaks näidati, et analüüs alla viie miljoni sekveneerimise lugemiga oli võimalik kõrge loote DNA fraktsiooni korral, ent kätkes tuvastamata trisoomia riski. Kõigil analüüsitud NIPT tööriistadel esines teatud sekveneerimise lugemite

piirist tuvastamata trisoomiaid. Avastamata trisoomia riski oli võimalik vähendada, kui riski määramisel arvestada euploidsete proovide võrdlushulga Z-skooride jaotusega. Veel täheldati, et loote rakuvaba DNA hulga hinnangutel esines vähenenud korrelatsioon sama vereproovi madalamatel katvustel, mis juhib ettevaatlikusele loote DNA fraktsiooni hinnangute tõlgendamisel.

Kolmanda uurimuse raames valmis bioinformaatiline analüüsitarkvara BinDel määramaks madala katvusega ülegenoomse sekveneerimise NIPT andmetest loote kliiniliselt patogeensete mikrodeletsioonide esinemise riske. Simuleeritud loote mikrodeletsioonide abil määrasime loote DNA fraktsiooni ja mikrodeletsioonipiirkonna pikkuse mõju mikrodeletsiooniriski tuvastamise täpsusele 12 kliiniliselt patogeenses mikrodeletsiooni piirkonnas, sealhulgas 1p36 deletsiooni, 3q29, Wolf-Hirschhorn, Cri-du-chat, Williams-Beuren, Langer-Giedion, Jacobsen, Prader-Willi/Angelman, Smith-Magenis, NF1, 20p13 ja DiGeorge'i sündroomiga seotud genoomsetes piirkondades. BinDeli esialgne analüüs, kuhu oli kombineeritud 34 kliiniliselt kinnitunud mikrodeletsiooniga ja 50 euploidse lootega NIPT proovi, tuvastas BinDel õigesti 25 mikrodeletsiooniga proovi. 15 terve lootega proovi määrati ekslikult kõrge riskiga proovideks. Järgnes BinDeli edasiarendus, mille tulemusel määrati korrektselt 30 mikrodeletsiooniga proovi ning valepositiivsete mikrodeletsioonide risk vähenes kolmele.

Uuringu peamine piirang oli kliiniliselt valideeritud patogeensete mikrodeletsioonidega ja trisoomiatega sekveneeritud rakuvaba DNA proovide piiratud hulk. Lisaks olid sekveneerimise andmed madala katvusega, sest sekveneerimise katvus mõjutab NIPT kulutõhusust. Madalam katvus aga vähendab arvutuslike tööriistade tundlikkust tuvastada mikrodeletsioonide või trisoomiate riski ja seda eriti madala loote fraktsiooni korral. Samuti puuduvad NIPT andmetest spetsiaalselt mikrodeletsioonide riski tuvastamiseks loodud avalikult kättesaadavad arvutuslikud tööriistad, millega tulemusi võrrelda. Kuigi tööriistad nagu WisecondorX on kättesaadavad, ei ole need spetsiaalselt loodud patogeensete mikrodeletsioonide riski tuvastamiseks.

Kokkuvõttes mõjutavad kasutatavad algoritmid märkimisväärselt NIPT täpsust ja kulutõhusust ning seeläbi NIPT kättesaadavust tervishoiuteenuse osana.

REFERENCES

Akolekar, R., Beta, J., Picciarelli, G., Ogilvie, C., & D'Antonio, F. (2015). Procedure-related risk of miscarriage following amniocentesis and chorionic villus sampling: a systematic review and meta-analysis. *Ultrasound in Obstetrics & Gynecology*, *45*(1), 16–26. https://doi.org/10.1002/UOG.14636

Alldred, S. Kate, Takwoingi, Yemisi, Guo, Boliang, Pennant, Mary, Deeks, Jonathan J., Neilson, James P., & Alfirevic, Zarko. (2017). First trimester ultrasound tests alone or in combination with first trimester serum tests for Down's syndrome screening. *Cochrane Database of Systematic Reviews*, *2017*(3). https://doi.org/10.1002/14651858.CD012600/MEDIA/CDSR/CD012600/IMAGE_N/NCD012600-TST-072.PNG

Amberger, Joanna S., Bocchini, Carol A., Schiettecatte, François, Scott, Alan F., & Hamosh, Ada. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an Online catalog of human genes and genetic disorders. *Nucleic Acids Research*, *43*(D1), D789–D798. https://doi.org/10.1093/nar/gku1205

Antonarakis, Stylianos E., Lyle, Robert, Dermitzakis, Emmanouil T., Reymond, Alexandre, & Deutsch, Samuel. (2004). Chromosome 21 and Down syndrome: From genomics to pathophysiology. *Nature Reviews Genetics*, *5*(10), 725–738. https://doi.org/10.1038/nrg1448

Antshel, Kevin M., Kates, Wendy R., Roizen, Nancy, Fremont, Wanda, & Shprintzen, Robert J. (2005). 22q11.2 Deletion Syndrome: Genetics, Neuroanatomy and Cognitive/Behavioral Features Keywords. *Child Neuropsychology*, *11*(1), 5–19. https://doi.org/10.1080/09297040590911185

Asim, Ambreen, Kumar, Ashok, Muthuswamy, Srinivasan, Jain, Shalu, & Agarwal, Sarita. (2015). "down syndrome: An insight of the disease." *Journal of Biomedical Science*, *22*(1), 1–9. https://doi.org/10.1186/s12929-015-0138-y

Bassett, A. S., Chow, E. W. C., Husted, J., Hodgkinson, K. A., Oechslin, E., Harris, L., & Silversides, C. (2009). Premature death in adults with 22q11.2 deletion syndrome. *Journal of Medical Genetics*, *46*(5), 324–330. https://doi.org/10.1136/jmg.2008. 063800

Bassett, Anne S., Chow, Eva W. C., Husted, Janice, Weksberg, Rosanna, Caluseriu, Oana, Webb, Gary D., & Gatzoulis, Michael A. (2005). Clinical features of 78 adults with 22q11 Deletion Syndrome. *American Journal of Medical Genetics. Part A*, *138*(4), 307–313. https://doi.org/10.1002/ajmg.a.30984

Battaglia, Agatino, Carey, John C., & South, Sarah T. (2015). Wolf–Hirschhorn syndrome: A review and update. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, *169*(3), 216–223. https://doi.org/10.1002/AJMG. C.31449

Baumer, Alessandra, Dutly, Fabrizio, Balmer, Damina, Riegel, Mariluce, Tükel, Turgut, Krajewska-Walasek, Malgorzata, & Schinzel, Albert A. (1998). High level of unequal meiotic crossovers at the origin of the 22q11.2 and 7q11.23 deletions. *Human Molecular Genetics*, *7*(5).

Bayindir, Baran, Dehaspe, Luc, Brison, Nathalie, Brady, Paul, Ardui, Simon, Kammoun, Molka, Van Der Veken, Lars, Lichtenbelt, Klaske, Van Den Bogaert, Kris, Van Houdt, Jeroen, Peeters, Hilde, Van Esch, Hilde, De Ravel, Thomy, Legius, Eric, Devriendt, Koen, & Vermeesch, Joris R. (2015). Noninvasive prenatal testing using a novel analysis pipeline to screen for all autosomal fetal aneuploidies improves pregnancy management. *European Journal of Human Genetics*, *23*(10), 1286–1293. https://doi.org/10.1038/ejhg.2014.282

Bianchi, Diana W., Platt, Lawrence D., Goldberg, James D., Abuhamad, Alfred Z., Sehnert, Amy J., & Rava, Richard P. (2012). Genome-wide fetal aneuploidy detection

by maternal plasma DNA sequencing. *Obstetrics and Gynecology*, *119*(5), 890–901. https://doi.org/10.1097/AOG.0b013e31824fb482

Bittles, AH, & Glasson, EJ. (2004). Clinical, social, and ethical implications of changing life expectancy in Down syndrome. *Developmental Medicine & Child Neurology*, *46*(04). https://doi.org/10.1017/s0012162204000441

Bull, Marilyn J. (2020). Down Syndrome. *Https://Doi.Org/10.1056/NEJMra1706537*, *382*(24), 2344–2352. https://doi.org/10.1056/NEJMRA1706537

Campbell, Ian M., Sheppard, Sarah E., Crowley, T. Blaine, McGinn, Daniel E., Bailey, Alice, McGinn, Michael J., Unolt, Marta, Homans, Jelle F., Chen, Erin Y., Salmons, Harold I., Gaynor, J. William, Goldmuntz, Elizabeth, Jackson, Oksana A., Katz, Lorraine E., Mascarenhas, Maria R., Deeney, Vincent F. X., Castelein, René M., Zur, Karen B., Elden, Lisa, … McDonald-McGinn, Donna M. (2018). What is new with 22q? An update from the 22q and You Center at the Children's Hospital of Philadelphia. *American Journal of Medical Genetics Part A*, *176*(10), 2058–2069. https://doi.org/10.1002/ajmg.a.40637

Cassidy, Suzanne B., Schwartz, Stuart, Miller, Jennifer L., & Driscoll, Daniel J. (2012). Prader-Willi syndrome. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *14*(1), 10–26. https://doi.org/10.1038/GIM.0B013E31822BEAD0

Cereda, Anna, & Carey, John C. (2012). The trisomy 18 syndrome. *Orphanet Journal of Rare Diseases*, *7*(1), 81. https://doi.org/10.1186/1750-1172-7-81

Cerruti Mainardi, Paola. (2006). Cri du Chat syndrome. *Orphanet Journal of Rare Diseases*, *1*(1). https://doi.org/10.1186/1750-1172-1-33

Chan, K. C. Alle., Zhang, Jun, Hui, Angela B. Y., Wong, Nathalie, Lau, Tze K., Leung, Tse N., Lo, Kwok Wai, Huang, Dolly W. S., & Lo, Y. M. Denni. (2004). Size Distributions of Maternal and Fetal DNA in Maternal Plasma. *Clinical Chemistry*, *50*(1), 88–92. https://doi.org/10.1373/clinchem.2003.024893

Charalambous, Chloe, Webster, Alexandre, & Schuh, Melina. (2023). Aneuploidy in mammalian oocytes and the impact of maternal ageing. *Nature Reviews Molecular Cell Biology*, *24*(1), 27–44. https://doi.org/10.1038/s41580-022-00517-3

Danecek, Petr, Bonfield, James K., Liddle, Jennifer, Marshall, John, Ohan, Valeriu, Pollard, Martin O., Whitwham, Andrew, Keane, Thomas, McCarthy, Shane A., & Davies, Robert M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), 1–4. https://doi.org/10.1093/gigascience/giab008

Dar, Pe'er, Jacobsson, Bo, Clifton, Rebecca, Egbert, Melissa, Malone, Fergal, Wapner, Ronald J., Roman, Ashley S., Khalil, Asma, Faro, Revital, Madankumar, Rajeevi, Edwards, Lance, Strong, Noel, Haeri, Sina, Silver, Robert, Vohra, Nidhi, Hyett, Jon, Demko, Zachary, Martin, Kimberly, Rabinowitz, Matthew, … Norton, Mary E. (2022). Cell-free DNA screening for prenatal detection of 22q11.2 deletion syndrome. *American Journal of Obstetrics and Gynecology*, *227*(1), 79.e1-79.e11. https://doi.org/10.1016/j.ajog.2022.01.002

De Jong, Antina, Dondorp, Wybo J., Timmermans, Daniëlle R. M., Van Lith, Jan M. M., & De Wert, Guido M. W. R. (2011). Rapid aneuploidy detection or karyotyping? Ethical reflection. *European Journal of Human Genetics*, *19*(10), 1020. https://doi.org/10.1038/EJHG.2011.82

Deng, Cechuan, Liu, Jianlong, Liu, Sha, Liu, Hongqian, Bai, Ting, Jing, Xiaosha, Xia, Tianyu, Liu, Yunyun, Cheng, Jing, Wei, Xiang, Xing, Lingling, Luo, Yuan, Zhou, Quanfang, Zhu, Qian, & Liu, Shanling. (2023). Maternal and fetal factors influencing

fetal fraction: A retrospective analysis of 153,306 pregnant women undergoing non-invasive prenatal screening. *Frontiers in Pediatrics*, *11*, 1066178. https://doi.org/10.3389/FPED.2023.1066178

Dungan, Jeffrey S., Klugman, Susan, Darilek, Sandra, Malinowski, Jennifer, Akkari, Yassmine M. N., Monaghan, Kristin G., Erwin, Angelika, & Best, Robert G. (2023). Noninvasive prenatal screening (NIPS) for fetal chromosome abnormalities in a general-risk population: An evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, *25*(2), 100336. https://doi.org/10.1016/J.GIM.2022.11.004

Eesti Meditsiiniline Sünniregister. (2025, March 11). *Kõik Eestis sündinud lapsed.* SR001: Sünnid, Tervise Arengu Instituut. https://statistika.tai.ee/pxweb/et/Andmebaas/Andmebaas__01Rahvastik__02Synnid/SR001.px/table/tableViewLayout2/

Faas, Brigitte Hw, De Ligt, Joep, Janssen, Irene, Eggink, Alex J., Wijnberger, Lia De, Van Vugt, John Mg, Vissers, Lisenka, & Geurts Van Kessel, Ad. (2012). Non-invasive prenatal diagnosis of fetal aneuploidies using massively parallel sequencing-by-ligation and evidence that cell-free fetal DNA in the maternal plasma originates from cytotrophoblastic cells. *Expert Opinion on Biological Therapy*, *12*(SUPPL. 1). https://doi.org/10.1517/14712598.2012.670632

Ferner, Rosalie E., & Gutmann, David H. (2013). Neurofibromatosis type 1 (NF1): diagnosis and management. *Handbook of Clinical Neurology*, *115*, 939–955. https://doi.org/10.1016/B978-0-444-52902-2.00053-9

Firth, Helen V., Richards, Shola M., Bevan, A. Paul, Clayton, Stephen, Corpas, Manuel, Rajan, Diana, Vooren, Steven Van, Moreau, Yves, Pettett, Roger M., & Carter, Nigel P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *American Journal of Human Genetics*, *84*(4), 524–533. https://doi.org/10.1016/j.ajhg.2009.03.010

Gabrielaite, Migle, Torp, Mathias Husted, Rasmussen, Malthe Sebro, Andreu-Sánchez, Sergio, Vieira, Filipe Garrett, Pedersen, Christina Bligaard, Kinalis, Savvas, Madsen, Majbritt Busk, Kodama, Miyako, Demircan, Gül Sude, Simonyan, Arman, Yde, Christina Westmose, Olsen, Lars Rønn, Marvig, Rasmus L., Østrup, Olga, Rossing, Maria, Nielsen, Finn Cilius, Winther, Ole, & Bagger, Frederik Otzen. (2021). A comparison of tools for copy-number variation detection in germline whole exome and whole genome sequencing data. *Cancers*, *13*(24), 6283. https://doi.org/10.3390/CANCERS13246283/S1

Geipel, A., Willruth, A., Vieten, J., Gembruch, U., & Berg, C. (2010). Nuchal fold thickness, nasal bone absence or hypoplasia, ductus venosus reversed flow and tricuspid valve regurgitation in screening for trisomies 21, 18 and 13 in the early second trimester. *Ultrasound in Obstetrics and Gynecology*, *35*(5), 535–539. https://doi.org/10.1002/UOG.7597

Guris, Deborah L., Duester, Gregg, Papaioannou, Virginia E., & Imamoto, Akira. (2006). Dose-dependent interaction of Tbx1 and Crkl and locally aberrant RA signaling in a model of del22q11 syndrome. *Developmental Cell*, *10*(1), 81–92. https://doi.org/10.1016/j.devcel.2005.12.002

Heft, Harrison, Soulodre, Claude, Cowan, Kara, Laing, Ana, Kaulback, Kellee, Mcdowell, Sarah, Ng, Vivian, Mitchell, Andrée, Lang, Amy, Sikich, Nancy, & Dhalla, Irfan. (2019). Noninvasive Prenatal Testing for Trisomies 21, 18, and 13, Sex Chromosome Aneuploidies, and Microdeletions: A Health Technology Assessment. *Ontario Health Technology Assessment Series*, *19*(4), 1. /pmc/articles/PMC6395059/

Hui, Lisa, & Bianchi, Diana W. (2020). Fetal fraction and noninvasive prenatal testing: What clinicians need to know. *Prenatal Diagnosis*, *40*(2), 155–163. https://doi.org/10.1002/pd.5620

Jett, Kimberly, & Friedman, Jan M. (2010). Clinical and genetic aspects of neurofibromatosis 1. *Genetics in Medicine*, *12*(1), 1–11. https://doi.org/10.1097/GIM.0b013e3181bf15e3

Jiang, Peiyong, Chan, K. C. Alle., Liao, Gary J. W., Zheng, Yama W. L., Leung, Tak Y., Chiu, Rossa W. K., Lo, Yuk Ming Dennis, & Sun, Hao. (2012). FetalQuant: Deducing fractional fetal DNA concentration from massively parallel sequencing of DNA in maternal plasma. *Bioinformatics*, *28*(22), 2883–2890. https://doi.org/10.1093/bioinformatics/bts549

Johansson, Lennart F., de Weerd, Hendrik A., de Boer, Eddy N., van Dijk, Freerk, te Meerman, Gerard J., Sijmons, Rolf H., Sikkema-Raddatz, Birgit, & Swertz, Morris A. (2018). NIPTeR: An R package for fast and accurate trisomy prediction in non-invasive prenatal testing. *BMC Bioinformatics*, *19*(1), 531. https://doi.org/10.1186/s12859-018-2557-8

Jordan, Valerie K., Zaveri, Hitisha P., & Scott, Daryl A. (2015). 1p36 deletion syndrome: an update. *The Application of Clinical Genetics*, *8*, 189. https://doi.org/10.2147/TACG.S65698

Kim, Sung K., Hannum, Gregory, Geis, Jennifer, Tynan, John, Hogg, Grant, Zhao, Chen, Jensen, Taylor J., Mazloom, Amin R., Oeth, Paul, Ehrich, Mathias, van den Boom, Dirk, & Deciu, Cosmin. (2015). Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts. *Prenatal Diagnosis*, *35*(8), 810–815. https://doi.org/10.1002/pd.4615

Kucharik, Marcel, Gnip, Andrej, Hyblova, Michaela, Budis, Jaroslav, Strieskova, Lucia, Harsanyova, Maria, Pös, Ondrej, Kubiritova, Zuzana, Radvanszky, Jan, Minarik, Gabriel, & Szemes, Tomas. (2020). Non-invasive prenatal testing (NIPT) by low coverage genomic sequencing: Detection limits of screened chromosomal microdeletions. *PLoS ONE*, *15*(8 August), e0238245. https://doi.org/10.1371/journal.pone.0238245

Kuhlmann-Capek, Maggie, Chiossi, Giuseppe, Singh, Prapti, Monsivais, Luis, Lozovyy, Violetta, Gallagher, Lauren, Kirsch, Nathan, Florence, Elizabeth, Petruzzi, Victoria, Chang, Jeffrey, Buenaventura, Sofia, Walden, Paul, Gardner, Benjamin, Munn, Mary, & Costantine, Maged. (2019). Effects of medication intake in early pregnancy on the fetal fraction of cell-free DNA testing. *Prenatal Diagnosis*, *39*(5), 361–368. https://doi.org/10.1002/PD.5436

Lannoo, Lore, van der Meij, Karuna R. M., Bekker, Mireille N., De Catte, Luc, Deckers, Sarah, Devriendt, Koenraad, Roggen, Nele, Galjaard, Robert Jan H., Gitsels-van der Wal, Janneke, Macville, Merryn V. E., Martin, Linda, Sistermans, Erik A., Van Calsteren, Kristel, Van Keirsbilck, Joachim, Crombag, Neeltje, & Henneman, Lidewij. (2023). A cross-country comparison of pregnant women's decision-making and perspectives when opting for non-invasive prenatal testing in the Netherlands and Belgium. *Prenatal Diagnosis*, *43*(3), 294–303. https://doi.org/10.1002/PD.6329

Liao, Gary J. W., Chan, K. C. Alle., Jiang, Peiyong, Sun, Hao, Leung, Tak Y., Chiu, Rossa W. K., & Dennis Lo, Y. M. (2012). Noninvasive prenatal diagnosis of fetal trisomy 21 by allelic ratio analysis using targeted massively parallel sequencing of maternal plasma DNA. *PLoS ONE*, *7*(5), e38154. https://doi.org/10.1371/journal.pone.0038154

Liao, Gary J. W., Lun, Fiona M. F., Zheng, Yama W. L., Chan, K. C. Alle., Leung, Tak Y., Lau, Tze K., Chiu, Rossa W. K., & Lo, Y. M. Denni. (2011). Targeted massively parallel sequencing of maternal plasma DNA permits efficient and unbiased detection of fetal alleles. *Clinical Chemistry*, *57*(1), 92–101. https://doi.org/10.1373/clinchem.2010.154336

Lin, Tzu Yi, Hsieh, T'Sang T. An., Cheng, Po Jen, Hung, Tai Ho, Chan, Kok Seong, Tsai, Chris, & Shaw, Steven W. (2021). Taiwanese Clinical Experience with Noninvasive Prenatal Testing for DiGeorge Syndrome. *Fetal Diagnosis and Therapy*, *48*(9), 672–677. https://doi.org/10.1159/000519057

Liu, Baohong, Tang, Xiaoyan, Qiu, Feng, Tao, Chunmei, Gao, Junhui, Ma, Mengmeng, Zhong, Tingyan, Cai, Jian Ping, Li, Yixue, & Ding, Guohui. (2016). DASAF: An R Package for Deep Sequencing-Based Detection of Fetal Autosomal Abnormalities from Maternal Cell-Free DNA. *BioMed Research International*, *2016*. https://doi.org/10.1155/2016/2714341

Lo, Kitty K., Boustred, Christopher, Chitty, Lyn S., & Plagnol, Vincent. (2014). RAPIDR: an analysis package for non-invasive prenatal testing of aneuploidy. *Bioinformatics (Oxford, England)*, *30*(20), 2965–2967. https://doi.org/10.1093/bioinformatics/btu419

Lo, Y. M. Dennis, Corbetta, Noemi, Chamberlain, Paul F., Rai, Vik, Sargent, Ian L., Redman, Christopher W. G., & Wainscoat, James S. (1997). Presence of fetal DNA in maternal plasma and serum. *Lancet*, *350*(9076), 485–487. https://doi.org/10.1016/S0140-6736(97)02174-0

Lo, Y. M. Dennis, Tein, Mark S. C., Lau, Tze K., Haines, Christopher J., Leung, Tse N., Poon, Priscilla M. K., Wainscoat, James S., Johnson, Philip J., Chang, Allan M. Z., & Hjelm, N. Magnus. (1998). Quantitative Analysis of Fetal DNA in Maternal Plasma and Serum: Implications for Noninvasive Prenatal Diagnosis. *The American Journal of Human Genetics*, *62*(4), 768–775. https://doi.org/10.1086/301800

Luijten, Mirjam, Redeker, Sandra, Minoshima, Shinsei, Shimizu, Nobuyoshi, Westerveld, Andries, & Hulsebos, Theo J. M. (2001). Duplication and transposition of the NF1 pseudogene regions on chromosomes 2, 14, and 22. *Human Genetics*, *109*(1), 109–116. https://doi.org/10.1007/s004390100543

Luijten, Mirjam, Wang, Yingping, Smith, Blaine T., Westerveld, Andries, Smink, Luc J., Dunham, Ian, Roe, Bruce A., & Hulsebos, Theo J. M. (2000). Mechanism of spreading of the highly related neurofibromatosis type 1 (NF1) pseudogenes on chromosomes 2, 14 and 22. *European Journal of Human Genetics : EJHG*, *8*(3), 209–214. https://doi.org/10.1038/sj.ejhg.5200434

Mandel, P., & Metais, P. (1948). Les acides nucléiques du plasma sanguin chez l'homme. *Comptes Rendus Des Séances de La Société de Biologie et de Ses Filiales*, *142*(3–4), 241–243. http://www.ncbi.nlm.nih.gov/pubmed/18875018

Margolis, Seth S., Sell, Gabrielle L., Zbinden, Mark A., & Bird, Lynne M. (2015). Angelman Syndrome. *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics*, *12*(3), 641–650. https://doi.org/10.1007/S13311-015-0361-Y

Mattina, Teresa, Perrotta, Concetta Simona, & Grossfeld, Paul. (2009). Jacobsen syndrome. *Orphanet Journal of Rare Diseases*, *4*(1), 1–10. https://doi.org/10.1186/1750-1172-4-9/FIGURES/2

Mayes, Sarah, Hashmi, Syed, Turrentine, Mark A., Darilek, Sandra, Friel, Lara A., & Czerwinski, Jennifer. (2016). Obstetrician and Gynecologist Utilization of the Noninvasive Prenatal Testing Expanded Option. *American Journal of Perinatology Reports*, *06*(01), e18–e24. https://doi.org/10.1055/S-0035-1566313

McDonald-McGinn, Donna M., Hain, Heather S., Emanuel, Beverly S., & Zackai, Elaine H. (1993). *22q11.2 Deletion Syndrome.* (Margaret P. Adam, Jerry Feldman, Ghayda M. Mirzaa, Roberta A. Pagon, Stephanie E. Wallace, & Anne Amemiya (eds.)).

McDonald-Mcginn, Donna M., & Sullivan, Kathleen E. (2011). Chromosome 22q11.2 deletion syndrome (DiGeorge syndrome/velocardiofacial syndrome). *Medicine*, *90*(1), 1–18. https://doi.org/10.1097/MD.0b013e3182060469

McDonald-McGinn, Donna M., Tonnesen, Melissa K., Laufer-Cahana, Ayala, Finucane, Brenda, Driscoll, Deborah A., Emanuel, Beverly S., & Zackai, Elaine H. (2001). Phenotype of the 22q11.2 deletion in individuals identified through an affected relative: Cast a wide FISHing net! *Genetics in Medicine*, *3*(1), 23–29. https://doi.org/10.1097/00125817-200101000-00006

Norwitz, Errol R., & Levy, Brynn. (2013). Noninvasive prenatal testing: the future is now. *Reviews in Obstetrics & Gynecology*, *6*(2), 48–62. https://doi.org/10.3909/ riog0201

Olshen, Adam B., Venkatraman, E. S., Lucito, Robert, & Wigler, Michael. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*, *5*(4), 557–572. https://doi.org/10.1093/biostatistics/kxh008

Palomaki, Glenn E., Kloza, Edward M., Lambert-Messerlian, Geralyn M., Haddow, James E., Neveux, Louis M., Ehrich, Mathias, Van Den Boom, Dirk, Bombard, Allan T., Deciu, Cosmin, Grody, Wayne W., Nelson, Stanley F., & Canick, Jacob A. (2011). DNA sequencing of maternal plasma to detect Down syndrome: An international clinical validation study. *Genetics in Medicine*, *13*(11), 913–920. https://doi.org/10.1097/GIM.0b013e3182368a0e

Paluoja, Priit, Jatsenko, Tatjana, Teder, Hindrek, Krjutškov, Kaarel, Vermeesch, Joris Robert, Salumets, Andres, & Palta, Priit. (2025). BinDel: Detecting Clinically Relevant Fetal Genomic Microdeletions Using Low-Coverage Whole-Genome Sequencing-Based NIPT. *Prenatal Diagnosis*, *45*(3), 352–361. https://doi.org/10.1002/pd.6758

Paluoja, Priit, Teder, Hindrek, Ardeshirdavani, Amin, Bayindir, Baran, Vermeesch, Joris, Salumets, Andres, Krjutškov, Kaarel, & Palta, Priit. (2021). Systematic evaluation of NIPT aneuploidy detection software tools with clinically validated NIPT samples. *PLoS Computational Biology*, *17*(12), e1009684. https://doi.org/10.1371/journal.pcbi. 1009684

Patau, Klaus, Smith, David W., Therman, Eeva, Inhorn, Stanley L., & Wagner, Hans P. (1960). Multiple Congenital Anomaly Caused By an Extra Autosome. *The Lancet*, *275*(7128), 790–793. https://doi.org/10.1016/S0140-6736(60)90676-0

Peng, Xianlu Laura, & Jiang, Peiyong. (2017). Bioinformatics approaches for fetal DNA fraction estimation in noninvasive prenatal testing. *International Journal of Molecular Sciences*, *18*(2), 453. https://doi.org/10.3390/ijms18020453

Petry, Patrícia, Polli, Janaina B., Mattos, Vinícius F., Rosa, Rosana C. M., Zen, Paulo R. G., Graziadio, Carla, Paskulin, Giorgio A., & Rosa, Rafael F. M. (2013). Clinical features and prognosis of a sample of patients with trisomy 13 (Patau syndrome) from Brazil. *American Journal of Medical Genetics, Part A*, *161*(6), 1278–1283. https://doi.org/10.1002/ajmg.a.35863

Pober, Barbara R. (2010). Williams–Beuren Syndrome. *New England Journal of Medicine*, *362*(3), 239–252. https://doi.org/10.1056/NEJMRA0903074/SUPPL_FILE/NEJM_POBER_239SA1.PDF

Raman, Lennart, Baetens, Machteld, De Smet, Matthias, Dheedene, Annelies, Van Dorpe, Jo, & Menten, Björn. (2019). PREFACE: In silico pipeline for accurate cell-free fetal DNA fraction prediction. *Prenatal Diagnosis*, *39*(10), 925–933. https://doi.org/10.1002/pd.5508

Raman, Lennart, Dheedene, Annelies, De Smet, Matthias, Van Dorpe, Jo, & Menten, Björn. (2019). WisecondorX: improved copy number detection for routine shallow whole-genome sequencing. *Nucleic Acids Research*, *47*(4), 1605–1614. https://doi.org/10.1093/nar/gky1263

Rasmussen, Sonja A., Wong, Lee Yang C., Yang, Quanhe, May, Kristin M., & Friedman, J. M. (2003). Population-Based Analyses of Mortality in Trisomy 13 and Trisomy 18. *Pediatrics*, *111*(4), 777–784. https://doi.org/10.1542/PEDS.111.4.777

Rehm, Heidi L., Berg, Jonathan S., Brooks, Lisa D., Bustamante, Carlos D., Evans, James P., Landrum, Melissa J., Ledbetter, David H., Maglott, Donna R., Martin, Christa Lese, Nussbaum, Robert L., Plon, Sharon E., Ramos, Erin M., Sherry, Stephen T., & Watson, Michael S. (2015). ClinGen — The Clinical Genome Resource. *New England Journal of Medicine*, *372*(23), 2235–2242. https://doi.org/10.1056/NEJMSR1406261/SUPPL_FILE/NEJMSR1406261_DISCLOSURES.PDF

Reimand, Tiia, Õunap, Katrin, Zordania, Riina, Ilus, Tiiu, Uibo, Oivi, Sitska, Mari, & Talvik, Tiina. (2006). Descriptive epidemiology of Down's syndrome in Estonia. *Paediatric and Perinatal Epidemiology*, *20*(6), 512–519. https://doi.org/10.1111/J.1365-3016.2006.00758.X,

Rinaldi, Berardo, Villa, Roberta, Sironi, Alessandra, Garavelli, Livia, Finelli, Palma, & Bedeschi, Maria Francesca. (2022). Smith-Magenis Syndrome—Clinical Review, Biological Background and Related Disorders. *Genes 2022, Vol. 13, Page 335*, *13*(2), 335. https://doi.org/10.3390/GENES13020335

Roberts, Wallisa, Zurada, Anna, Zurada-ZieliŃSka, Agnieszka, Gielecki, Jerzy, & Loukas, Marios. (2016). Anatomy of trisomy 18. *Clinical Anatomy*, *29*(5), 628–632. https://doi.org/10.1002/ca.22725

Saitta, Sulagna C., Harris, Stacy E., Gaeth, Ann P., Driscoll, Deborah A., McDonald-McGinn, Donna M., Maisenbacher, Melissa K., Yersak, Jill M., Chakraborty, Prabir K., Hacker, April M., Zackai, Elaine H., Ashley, Terry, & Emmanuel, Beverly S. (2004). Aberrant interchromosomal exchanges are the predominant cause of the 22q11.2 deletion. *Human Molecular Genetics*, *13*(4), 417–428. https://doi.org/10.1093/HMG/DDH041

Sauk, Martin, Žilina, Olga, Kurg, Ants, Ustav, Eva Liina, Peters, Maire, Paluoja, Priit, Roost, Anne Mari, Teder, Hindrek, Palta, Priit, Brison, Nathalie, Vermeesch, Joris R., Krjutškov, Kaarel, Salumets, Andres, & Kaplinski, Lauris. (2018). NIPTmer: Rapid k-mer-based software package for detection of fetal aneuploidies. *Scientific Reports*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-23589-8

Sehnert, Amy J., Rhees, Brian, Comstock, David, De Feo, Eileen, Heilek, Gabrielle, Burke, John, & Rava, Richard P. (2011). Optimal detection of fetal chromosomal abnormalities by massively parallel DNA sequencing of cell-free fetal DNA from maternal blood. *Clinical Chemistry*, *57*(7), 1042–1049. https://doi.org/10.1373/clinchem.2011.165910

Smajlagić, Dinka, Lavrichenko, Ksenia, Berland, Siren, Helgeland, Øyvind, Knudsen, Gun Peggy, Vaudel, Marc, Haavik, Jan, Knappskog, Per Morten, Njølstad, Pål Rasmus, Houge, Gunnar, & Johansson, Stefan. (2021). Population prevalence and inheritance pattern of recurrent CNVs associated with neurodevelopmental disorders in 12,252 newborns and their parents. *European Journal of Human Genetics*, *29*(1), 205–215. https://doi.org/10.1038/s41431-020-00707-7

Sparks, Andrew B., Struble, Craig A., Wang, Eric T., Song, Ken, & Oliphant, Arnold. (2012). Noninvasive prenatal detection and selective analysis of cell-free DNA obtained from maternal blood: Evaluation for trisomy 21 and trisomy 18. *American*

Journal of Obstetrics and Gynecology, *206*(4), 319.e1-319.e9. https://doi.org/10.1016/j.ajog.2012.01.030

Spencer, K., Souter, V., Tul, N., Snijders, R., & Nicolaides, K. H. (1999). A screening program for trisomy 21 at 10–14 weeks using fetal nuchal translucency, maternal serum free β-human chorionic gonadotropin and pregnancy-associated plasma protein-A. *Ultrasound in Obstetrics and Gynecology*, *13*(4), 231–237. https://doi.org/10.1046/j.1469-0705.1999.13040231.x

Stankiewicz, Pawel, & Lupski, James R. (2002). Genome architecture, rearrangements and genomic disorders. *Trends in Genetics*, *18*(2), 74–82. https://doi.org/10.1016/S0168-9525(02)02592-1

Straver, Roy, Sistermans, Erik A., Holstege, Henne, Visser, Allerdien, Oudejans, Cees B. M., & Reinders, Marcel J. T. (2014). WISECONDOR: Detection of fetal aberrations from shallow sequencing maternal plasma based on a within-sample comparison scheme. *Nucleic Acids Research*, *42*(5). https://doi.org/10.1093/nar/gkt992

Strømme, Petter, Bjømstad, Per G., & Ramstad, Kjersti. (2002). Prevalence estimation of Williams syndrome. *Journal of Child Neurology*, *17*(4), 269–271. https://doi.org/10.1177/088307380201700406

Teder, Hindrek, Paluoja, Priit, Rekker, Kadri, Salumets, Andres, Krjutškov, Kaarel, & Palta, Priit. (2019). Computational framework for targeted high-coverage sequencing based NIPT. *PLoS ONE*, *14*(7), e0209139. https://doi.org/10.1371/journal.pone.0209139

Tian, Weifang, Yuan, Yangyang, Yuan, Erfeng, Zhang, Linlin, Liu, Ling, Li, Ying, Guo, Jing, Cui, Xueyin, Li, Pengyun, & Cui, Shihong. (2023). Evaluation of the clinical utility of extended non-invasive prenatal testing in the detection of chromosomal aneuploidy and microdeletion/microduplication. *European Journal of Medical Research*, *28*(1). https://doi.org/10.1186/s40001-023-01285-2

van Beek, Daphne M., Straver, Roy, Weiss, Marian M., Boon, Elles M. J., Huijsdens-van Amsterdam, Karin, Oudejans, Cees B. M., Reinders, Marcel J. T., & Sistermans, Erik A. (2017). Comparing methods for fetal fraction determination and quality control of NIPT samples. *Prenatal Diagnosis*, *37*(8), 769–773. https://doi.org/10.1002/pd.5079

Van der Auwera, Geraldine, O'Connor, Brian, & Safari, an O'Reilly Media Company. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. *Genomics in the Cloud*, 300. https://www.oreilly.com/library/view/genomics-in-the/9781491975183/

Watson, Corey T., Marques-Bonet, Tomas, Sharp, Andrew J., & Mefford, Heather C. (2014). The genetics of microdeletion and microduplication syndromes: an update. *Annual Review of Genomics and Human Genetics*, *15*, 215–244. https://doi.org/10.1146/annurev-genom-091212-153408

Williams, Virginia C., Lucas, John, Babcock, Michael A., Gutmann, David H., Bruce, B., & Maria, Bernard L. (2009). Neurofibromatosis Type 1 Revisited. *Pediatrics*, *123*(1), 124–133. https://doi.org/10.1542/PEDS.2007-3204

Yagi, Hisato, Furutani, Yoshiyuki, Hamada, Hiromichi, Sasaki, Takashi, Asakawa, Shuichi, Minoshima, Shinsei, Ichida, Fukiko, Joo, Kunitaka, Kimura, Misa, Imamura, Shin ichiro, Kamatani, Naoyuki, Momma, Kazuo, Takao, Atsuyoshi, Nakazawa, Makoto, Shimizu, Nobuyoshi, & Matsuoka, Rumiko. (2003). Role of TBX1 in human del22q11.2 syndrome. *Lancet*, *362*(9393), 1366–1373. https://doi.org/10.1016/S0140-6736(03)14632-6

Yu, Stephanie C. Y., Chan, K. C. Alle., Zheng, Yama W. L., Jiang, Peiyong, Liao, Gary J. W., Sun, Hao, Akolekar, Ranjit, Leung, Tak Y., Go, Attie T. J. I., Van Vugt, John M. G., Minekawa, Ryoko, Oudejans, Cees B. M., Nicolaides, Kypros H., Chiu, Rossa W. K., & Lo, Y. M. Denni. (2014). Size-based molecular diagnostics using plasma

DNA for noninvasive prenatal testing. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8583–8588. https://doi.org/10.1073/pnas.1406103111

Zhao, Chen, Tynan, John, Ehrich, Mathias, Hannum, Gregory, McCullough, Ron, Saldivar, Juan Sebastian, Oeth, Paul, Van Den Boom, Dirk, & Deciu, Cosmin. (2015). Detection of fetal subchromosomal abnormalities by sequencing circulating cell-free DNA from maternal plasma. *Clinical Chemistry*, *61*(4), 608–616. https://doi.org/10.1373/clinchem.2014.233312

Žilina, Olga, Rekker, Kadri, Kaplinski, Lauris, Sauk, Martin, Paluoja, Priit, Teder, Hindrek, Ustav, Eva Liina, Tõnisson, Neeme, Reimand, Tiia, Ridnõi, Konstantin, Palta, Priit, Vermeesch, Joris Robert, Krjutškov, Kaarel, Kurg, Ants, & Salumets, Andres. (2019). Creating basis for introducing non-invasive prenatal testing in the Estonian public health setting. *Prenatal Diagnosis*, *39*(13), 1262–1268. https://doi.org/10.1002/pd.5578

ACKNOWLEDGMENTS

The research presented in this thesis was conducted at Celvia CC AS and the Institute of Clinical Medicine, University of Tartu.

I wish to express my sincere gratitude to all who have contributed to this thesis and supported me throughout my PhD journey. I am deeply thankful for the guidance of my supervisors, Prof. Priit Palta, Assoc. Prof. Kaarel Krjutškov, and Prof. Andres Salumets. I am especially grateful to Prof. Priit Palta, whose mentorship has been invaluable in my academic pursuits. I also extend my gratitude to my former and present colleagues at Celvia CC for their support.

I sincerely thank the referees, Senior Dr. Ele Hanson and Assoc. Prof. Sander Pajusalu, as well as the esteemed Opponent, Prof. Silvia Bottini, for their scientific and intellectual feedback and constructive criticism. I am grateful to all the co-authors and collaborators of the publications, especially Prof. Joris Vermeesch, Dr. Tatjana Jatsenko, Dr. Amin Ardeshirdavani, and Dr. Baran Bayindir.

I owe my deepest gratitude to my parents, Tiiu and Urmas, my brother Kristo, and my beloved friends for their unwavering support and belief in me throughout my university studies.

Finally, I would like to thank my beloved Marietta for her unwavering support, which has been incredibly important to me.

PUBLICATIONS

CURRICULUM VITAE

Name: Priit Paluoja
Born: September 2, 1995, Tartu
Citizenship: Estonian
Email priitpaluoja@gmail.com

Education

2020–2025	University of Tartu, Medicine, PhD
2017–2019	University of Tartu, Computer Science, MSc
2014–2017	University of Tartu, Computer Science, BSc
2012–2014	Hugo Treffner Gymnasium
2002–2012	Tartu Descartes Lyceum

Institution and position

01.10.2024–...	Celvia CC AS, Head of Data Management
01.03.2016–01.10.2024	Celvia CC AS, Bioinformatician
26.06.2017–18.08.2017	European Organization for Nuclear Research (CERN), Summer Student
20.06.2019–…	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Programmer
01.02.2023–30.08.2024	University of Tartu, Faculty of Medicine, Institute of Clinical Medicine, Junior Research Fellow in Bioinformatics

List of publications

1. **Paluoja, Priit**; Jatsenko, Tatjana; Teder, Hindrek; Krjutškov, Kaarel; Vermeesch, Joris Robert; Salumets, Andres; Palta, Priit (2025). BinDel: Detecting Clinically Relevant Fetal Genomic Microdeletions Using Low-Coverage Whole-Genome Sequencing-Based NIPT. Prenatal Diagnosis. DOI: 10.1002/pd.6758.
2. **Paluoja, Priit**; Vaher, Mihkel; Teder, Hindrek; Krjutškov, Kaarel; Salumets, Andres; Raime, Kairi (2025). Honey bulk DNA metagenomic analysis to identify honey biological composition and monitor honey bee pathogens. Npj Science of Food, 9 (1), 91. DOI: 10.1038/s41538-025-00464-1.
3. Meltsov, A.; Saare, M.; Teder, H.; **Paluoja, P**.; Arffman, R.K.; Piltonen, T.; Laudanski, P.; Wielgoś, M.; Gianaroli, L.; Koel, M.; Peters, M.; Salumets, A.; Krjutškov, K.; Palta, P. (2023). Targeted gene expression profiling for accurate endometrial receptivity testing. Scientific Reports, 13 (1), 13959. DOI: 10.1038/s41598-023-40991-z.
4. Vuoristo, Sanna; Bhagat, Shruti; Hydén-Granskog, Christel; Yoshihara, Masahito; Gawriyski, Lisa; Jouhilahti, Eeva-Mari; Ranga, Vipin; Tamirat, Mahlet; Huhtala, Mikko; Kirjanov, Ida; Nykänen, Sonja; Krjutškov, Kaarel; Damdimopoulos, Anastassius; Weltner, Jere; Hashimoto, Kosuke; Recher,

Gaëlle; Ezer, Sini; **Paluoja, Priit**; Paloviita, Pauliina; Takegami, Yujiro ... (2022). DUX4 is a multifunctional factor priming human embryonic genome activation. iScience, 104137. DOI: 10.1016/j.isci.2022.104137.

5. Paloviita, P.; Hydén-Granskog, C.; Yohannes, D.A.; **Paluoja, P**.; Kere, J.; Tapanainen, J.S.; Krjutškov, K.; Tuuri, T.; Võsa, U.; Vuoristo, S. (2021). Small RNA expression and miRNA modification dynamics in human oocytes and early embryos. Genome Research, 31 (8), 1474−1485. DOI: 10.1101/gr.268193.120.
6. **Paluoja, Priit**; Teder, Hindrek; Ardeshirdavani, Amin; Bayindir, Baran; Vermeesch, Joris; Salumets, Andres; Krjutškov, Kaarel; Palta, Priit (2021). Systematic evaluation of NIPT aneuploidy detection software tools with clinically validated NIPT samples. PLoS Computational Biology, 17 (12), e1009684. DOI: 10.1371/journal.pcbi.1009684.
7. Gordevičius, Juozas; Narmontė, Milda; Gibas, Povilas; Kvederavičiūtė, Kotryna; Tomkutė, Vita; **Paluoja, Priit**; Krjutškov, Kaarel; Salumets, Andres; Kriukienė, Edita (2020). Identification of fetal unmodified and 5-hydroxymethylated CG sites in maternal cell-free DNA for non-invasive prenatal testing. Clinical Epigenetics, 12 (1). DOI: 10.1186/s13148-020-00938-x.
8. Teder, Hindrek; **Paluoja, Priit**; Rekker, Kadri; Salumets, Andres; Krjutskov, Kaarel; Palta, Priit (2019). Computational framework for targeted high coverage sequencing based NIPT Computational framework for targeted high coverage sequencing based NIPT. PLoS ONE, 14 (7), ARTN e0209139. DOI: 10.1371/journal.pone.0209139.
9. Žilina, Olga; Rekker, Kadri; Kaplinski, Lauris; Sauk, Martin; **Paluoja, Priit**; Teder, Hindrek; Ustav, Eva-Liina; Tõnisson, Neeme; Reimand, Tiia; Ridnoi, Konstantin; Palta, Priit; Vermeesch, Joris Robert; Krjutškov, Kaarel; Kurg, Ants; Salumets, Andres (2019). Creating basis for introducing non-invasive prenatal testing in the Estonian public health setting. Prenatal Diagnosis, 39 (13), 1262−1268. DOI: 10.1002/pd.5578.
10. Saare, Merli; Laisk, Triin; Teder, Hindrek; **Paluoja, Priit**; Palta, Priit; Koel, Mariann; Kirss, Fred; Karro, Helle; Soritsa, Deniss; Salumets, Andres; Krjutskov, Kaarel; Peters, Maire (2019). A molecular tool for menstrual cycle phase dating of endometrial samples in endometriosis transcriptome studies. Biology of Reproduction, 101 (1), 1−3. DOI: 10.1093/biolre/ioz072.
11. Sauk, M.; Žilina, O.; Kurg, A.; Ustav, E-L.; Peters, M.; **Paluoja, P**.; Roost, A.M.; Teder, H.; Palta, P.; Brison, N.; Vermeesch, J.R.; Krjutškov, K.; Salumets, A; Kaplinski, L. (2018). NIPTmer: rapid k-mer-based software package for detection of fetal aneuploidies. Scientific Reports. DOI: 10.1038/s41598-018-23589-8.
12. Teder, Hindrek; Koel, Mariann; **Paluoja, Priit**; Jatsenko, Tatjana; Rekker, Kadri; Laisk-Podar, Triin; Kukuškina, Viktorija; Velthut-Meikas, Agne; Fjodorova, Olga; Peters, Maire; Kere, Juha; Salumets, Andres; Palta, Priit; Krjutškov, Kaarel (2018). TAC-seq: targeted DNA and RNA sequencing for precise biomarker molecule counting. npj Genomic Medicine, 3, ARTN 34. DOI: 10.1038/s41525-018-0072-5.

ELULOOKIRJELDUS

Nimi: Priit Paluoja
Sünniaeg: 2. september 1995, Tartu
Kodakondsus: Eesti
E-post priitpaluoja@gmail.com

Hariduskäik

2020–2025	Tartu Ülikool, arstiteadus, PhD
2017–2019	Tartu Ülikool, informaatika, MSc
2014–2017	Tartu Ülikool, informaatika, BSc
2012–2014	Hugo Treffneri Gümnaasium
2002–2012	Tartu Descartes'i Lütseum

Teenistuskäik

01.10.2024–...	Celvia CC AS, Andmevaldkonna juht
01.03.2016–01.10.2024	Celvia CC AS, Bioinformaatik
26.06.2017–18.08.2017	European Organization for Nuclear Research (CERN), Suvetudeng
20.06.2019–…	Tartu Ülikool, Loodus- ja täppisteaduste valdkond, arvutiteaduse instituut, programmeerija
01.02.2023–30.08.2024	Tartu Ülikool, Meditsiiniteaduste valdkond, kliinilise meditsiini instituut, bioinformaatika nooremteadur

Publikatsioonid

1. **Paluoja, Priit**; Jatsenko, Tatjana; Teder, Hindrek; Krjutškov, Kaarel; Vermeesch, Joris Robert; Salumets, Andres; Palta, Priit (2025). BinDel: Detecting Clinically Relevant Fetal Genomic Microdeletions Using Low-Coverage Whole-Genome Sequencing-Based NIPT. Prenatal Diagnosis. DOI: 10.1002/pd.6758.
2. **Paluoja, Priit**; Vaher, Mihkel; Teder, Hindrek; Krjutškov, Kaarel; Salumets, Andres; Raime, Kairi (2025). Honey bulk DNA metagenomic analysis to identify honey biological composition and monitor honey bee pathogens. Npj Science of Food, 9 (1), 91. DOI: 10.1038/s41538-025-00464-1.
3. Meltsov, A.; Saare, M.; Teder, H.; **Paluoja, P**.; Arffman, R.K.; Piltonen, T.; Laudanski, P.; Wielgoś, M.; Gianaroli, L.; Koel, M.; Peters, M.; Salumets, A.; Krjutškov, K.; Palta, P. (2023). Targeted gene expression profiling for accurate endometrial receptivity testing. Scientific Reports, 13 (1), 13959. DOI: 10.1038/s41598-023-40991-z.
4. Vuoristo, Sanna; Bhagat, Shruti; Hydén-Granskog, Christel; Yoshihara, Masahito; Gawriyski, Lisa; Jouhilahti, Eeva-Mari; Ranga, Vipin; Tamirat, Mahlet; Huhtala, Mikko; Kirjanov, Ida; Nykänen, Sonja; Krjutškov, Kaarel; Damdimopoulos, Anastassius; Weltner, Jere; Hashimoto, Kosuke; Recher,

Gaëlle; Ezer, Sini; **Paluoja, Priit**; Paloviita, Pauliina; Takegami, Yujiro ... (2022). DUX4 is a multifunctional factor priming human embryonic genome activation. iScience, 104137. DOI: 10.1016/j.isci.2022.104137.

5. Paloviita, P.; Hydén-Granskog, C.; Yohannes, D.A.; **Paluoja, P**.; Kere, J.; Tapanainen, J.S.; Krjutškov, K.; Tuuri, T.; Võsa, U.; Vuoristo, S. (2021). Small RNA expression and miRNA modification dynamics in human oocytes and early embryos. Genome Research, 31 (8), 1474−1485. DOI: 10.1101/gr.268193.120.
6. **Paluoja, Priit**; Teder, Hindrek; Ardeshirdavani, Amin; Bayindir, Baran; Vermeesch, Joris; Salumets, Andres; Krjutškov, Kaarel; Palta, Priit (2021). Systematic evaluation of NIPT aneuploidy detection software tools with clinically validated NIPT samples. PLoS Computational Biology, 17 (12), e1009684. DOI: 10.1371/journal.pcbi.1009684.
7. Gordevičius, Juozas; Narmontė, Milda; Gibas, Povilas; Kvederavičiūtė, Kotryna; Tomkutė, Vita; **Paluoja, Priit**; Krjutškov, Kaarel; Salumets, Andres; Kriukienė, Edita (2020). Identification of fetal unmodified and 5-hydroxymethylated CG sites in maternal cell-free DNA for non-invasive prenatal testing. Clinical Epigenetics, 12 (1). DOI: 10.1186/s13148-020-00938-x.
8. Teder, Hindrek; **Paluoja, Priit**; Rekker, Kadri; Salumets, Andres; Krjutskov, Kaarel; Palta, Priit (2019). Computational framework for targeted high coverage sequencing based NIPT Computational framework for targeted high coverage sequencing based NIPT. PLoS ONE, 14 (7), ARTN e0209139. DOI: 10.1371/journal.pone.0209139.
9. Žilina, Olga; Rekker, Kadri; Kaplinski, Lauris; Sauk, Martin; **Paluoja, Priit**; Teder, Hindrek; Ustav, Eva-Liina; Tõnisson, Neeme; Reimand, Tiia; Ridnoi, Konstantin; Palta, Priit; Vermeesch, Joris Robert; Krjutškov, Kaarel; Kurg, Ants; Salumets, Andres (2019). Creating basis for introducing non-invasive prenatal testing in the Estonian public health setting. Prenatal Diagnosis, 39 (13), 1262−1268. DOI: 10.1002/pd.5578.
10. Saare, Merli; Laisk, Triin; Teder, Hindrek; **Paluoja, Priit**; Palta, Priit; Koel, Mariann; Kirss, Fred; Karro, Helle; Soritsa, Deniss; Salumets, Andres; Krjutskov, Kaarel; Peters, Maire (2019). A molecular tool for menstrual cycle phase dating of endometrial samples in endometriosis transcriptome studies. Biology of Reproduction, 101 (1), 1−3. DOI: 10.1093/biolre/ioz072.
11. Sauk, M.; Žilina, O.; Kurg, A.; Ustav, E-L.; Peters, M.; **Paluoja, P**.; Roost, A.M.; Teder, H.; Palta, P.; Brison, N.; Vermeesch, J.R.; Krjutškov, K.; Salumets, A; Kaplinski, L. (2018). NIPTmer: rapid k-mer-based software package for detection of fetal aneuploidies. Scientific Reports. DOI: 10.1038/s41598-018-23589-8.
12. Teder, Hindrek; Koel, Mariann; **Paluoja, Priit**; Jatsenko, Tatjana; Rekker, Kadri; Laisk-Podar, Triin; Kukuškina, Viktorija; Velthut-Meikas, Agne; Fjodorova, Olga; Peters, Maire; Kere, Juha; Salumets, Andres; Palta, Priit; Krjutškov, Kaarel (2018). TAC-seq: targeted DNA and RNA sequencing for precise biomarker molecule counting. npj Genomic Medicine, 3, ARTN 34. DOI: 10.1038/s41525-018-0072-5.

DISSERTATIONES MEDICINAE UNIVERSITATIS TARTUENSIS

1. **Heidi-Ingrid Maaroos**. The natural course of gastric ulcer in connection with chronic gastritis and *Helicobacter pylori*. Tartu, 1991.
2. **Mihkel Zilmer**. Na-pump in normal and tumorous brain tissues: Structural, functional and tumorigenesis aspects. Tartu, 1991.
3. **Eero Vasar**. Role of cholecystokinin receptors in the regulation of behaviour and in the action of haloperidol and diazepam. Tartu, 1992.
4. **Tiina Talvik**. Hypoxic-ischaemic brain damage in neonates (clinical, biochemical and brain computed tomographical investigation). Tartu, 1992.
5. **Ants Peetsalu**. Vagotomy in duodenal ulcer disease: A study of gastric acidity, serum pepsinogen I, gastric mucosal histology and *Helicobacter pylori*. Tartu, 1992.
6. **Marika Mikelsaar**. Evaluation of the gastrointestinal microbial ecosystem in health and disease. Tartu, 1992.
7. **Hele Everaus**. Immuno-hormonal interactions in chronic lymphocytic leukaemia and multiple myeloma. Tartu, 1993.
8. **Ruth Mikelsaar**. Etiological factors of diseases in genetically consulted children and newborn screening: dissertation for the commencement of the degree of doctor of medical sciences. Tartu, 1993.
9. **Agu Tamm**. On metabolic action of intestinal microflora: clinical aspects. Tartu, 1993.
10. **Katrin Gross**. Multiple sclerosis in South-Estonia (epidemiological and computed tomographical investigations). Tartu, 1993.
11. **Oivi Uibo**. Childhood coeliac disease in Estonia: occurrence, screening, diagnosis and clinical characterization. Tartu, 1994.
12. **Viiu Tuulik**. The functional disorders of central nervous system of chemistry workers. Tartu, 1994.
13. **Margus Viigimaa**. Primary haemostasis, antiaggregative and anticoagulant treatment of acute myocardial infarction. Tartu, 1994.
14. **Rein Kolk**. Atrial versus ventricular pacing in patients with sick sinus syndrome. Tartu, 1994.
15. **Toomas Podar**. Incidence of childhood onset type 1 diabetes mellitus in Estonia. Tartu, 1994.
16. **Kiira Subi**. The laboratory surveillance of the acute respiratory viral infections in Estonia. Tartu, 1995.
17. **Irja Lutsar**. Infections of the central nervous system in children (epidemiologic, diagnostic and therapeutic aspects, long term outcome). Tartu, 1995.
18. **Aavo Lang**. The role of dopamine, 5-hydroxytryptamine, sigma and NMDA receptors in the action of antipsychotic drugs. Tartu, 1995.
19. **Andrus Arak**. Factors influencing the survival of patients after radical surgery for gastric cancer. Tartu, 1996.

20. **Tõnis Karki**. Quantitative composition of the human lactoflora and method for its examination. Tartu, 1996.
21. **Reet Mändar**. Vaginal microflora during pregnancy and its transmission to newborn. Tartu, 1996.
22. **Triin Remmel**. Primary biliary cirrhosis in Estonia: epidemiology, clinical characterization and prognostication of the course of the disease. Tartu, 1996.
23. **Toomas Kivastik**. Mechanisms of drug addiction: focus on positive reinforcing properties of morphine. Tartu, 1996.
24. **Paavo Pokk**. Stress due to sleep deprivation: focus on $GABA_A$ receptor-chloride ionophore complex. Tartu, 1996.
25. **Kristina Allikmets**. Renin system activity in essential hypertension. Associations with atherothrombogenic cardiovascular risk factors and with the efficacy of calcium antagonist treatment. Tartu, 1996.
26. **Triin Parik**. Oxidative stress in essential hypertension: Associations with metabolic disturbances and the effects of calcium antagonist treatment. Tartu, 1996.
27. **Svetlana Päi**. Factors promoting heterogeneity of the course of rheumatoid arthritis. Tartu, 1997.
28. **Maarike Sallo**. Studies on habitual physical activity and aerobic fitness in 4 to 10 years old children. Tartu, 1997.
29. **Paul Naaber**. *Clostridium difficile* infection and intestinal microbial ecology. Tartu, 1997.
30. **Rein Pähkla**. Studies in pinoline pharmacology. Tartu, 1997.
31. **Andrus Juhan Voitk**. Outpatient laparoscopic cholecystectomy. Tartu, 1997.
32. **Joel Starkopf**. Oxidative stress and ischaemia-reperfusion of the heart. Tartu, 1997.
33. **Janika Kõrv**. Incidence, case-fatality and outcome of stroke. Tartu, 1998.
34. **Ülla Linnamägi**. Changes in local cerebral blood flow and lipid peroxidation following lead exposure in experiment. Tartu, 1998.
35. **Ave Minajeva**. Sarcoplasmic reticulum function: comparison of atrial and ventricular myocardium. Tartu, 1998.
36. **Oleg Milenin**. Reconstruction of cervical part of esophagus by revascularised ileal autografts in dogs. A new complex multistage method. Tartu, 1998.
37. **Sergei Pakriev**. Prevalence of depression, harmful use of alcohol and alcohol dependence among rural population in Udmurtia. Tartu, 1998.
38. **Allen Kaasik**. Thyroid hormone control over β-adrenergic signalling system in rat atria. Tartu, 1998.
39. **Vallo Matto**. Pharmacological studies on anxiogenic and antiaggressive properties of antidepressants. Tartu, 1998.
40. **Maire Vasar**. Allergic diseases and bronchial hyperreactivity in Estonian children in relation to environmental influences. Tartu, 1998.
41. **Kaja Julge**. Humoral immune responses to allergens in early childhood. Tartu, 1998.

42. **Heli Grünberg**. The cardiovascular risk of Estonian schoolchildren. A cross-sectional study of 9-, 12- and 15-year-old children. Tartu, 1998.
43. **Epp Sepp**. Formation of intestinal microbial ecosystem in children. Tartu, 1998.
44. **Mai Ots**. Characteristics of the progression of human and experimental glomerulopathies. Tartu, 1998.
45. **Tiina Ristimäe**. Heart rate variability in patients with coronary artery disease. Tartu, 1998.
46. **Leho Kõiv**. Reaction of the sympatho-adrenal and hypothalamo-pituitary-adrenocortical system in the acute stage of head injury. Tartu, 1998.
47. **Bela Adojaan**. Immune and genetic factors of childhood onset IDDM in Estonia. An epidemiological study. Tartu, 1999.
48. **Jakov Shlik**. Psychophysiological effects of cholecystokinin in humans. Tartu, 1999.
49. **Kai Kisand**. Autoantibodies against dehydrogenases of α-ketoacids. Tartu, 1999.
50. **Toomas Marandi**. Drug treatment of depression in Estonia. Tartu, 1999.
51. **Ants Kask**. Behavioural studies on neuropeptide Y. Tartu, 1999.
52. **Ello-Rahel Karelson**. Modulation of adenylate cyclase activity in the rat hippocampus by neuropeptide galanin and its chimeric analogs. Tartu, 1999.
53. **Tanel Laisaar**. Treatment of pleural empyema — special reference to intrapleural therapy with streptokinase and surgical treatment modalities. Tartu, 1999.
54. **Eve Pihl**. Cardiovascular risk factors in middle-aged former athletes. Tartu, 1999.
55. **Katrin Õunap**. Phenylketonuria in Estonia: incidence, newborn screening, diagnosis, clinical characterization and genotype/phenotype correlation. Tartu, 1999.
56. **Siiri Kõljalg**. *Acinetobacter* – an important nosocomial pathogen. Tartu, 1999.
57. **Helle Karro**. Reproductive health and pregnancy outcome in Estonia: association with different factors. Tartu, 1999.
58. **Heili Varendi**. Behavioral effects observed in human newborns during exposure to naturally occurring odors. Tartu, 1999.
59. **Anneli Beilmann**. Epidemiology of epilepsy in children and adolescents in Estonia. Prevalence, incidence, and clinical characteristics. Tartu, 1999.
60. **Vallo Volke**. Pharmacological and biochemical studies on nitric oxide in the regulation of behaviour. Tartu, 1999.
61. **Pilvi Ilves**. Hypoxic-ischaemic encephalopathy in asphyxiated term infants. A prospective clinical, biochemical, ultrasonographical study. Tartu, 1999.
62. **Anti Kalda**. Oxygen-glucose deprivation-induced neuronal death and its pharmacological prevention in cerebellar granule cells. Tartu, 1999.
63. **Eve-Irene Lepist**. Oral peptide prodrugs – studies on stability and absorption. Tartu, 2000.

64. **Jana Kivastik**. Lung function in Estonian schoolchildren: relationship with anthropometric indices and respiratory symptomas, reference values for dynamic spirometry. Tartu, 2000.
65. **Karin Kull**. Inflammatory bowel disease: an immunogenetic study. Tartu, 2000.
66. **Kaire Innos**. Epidemiological resources in Estonia: data sources, their quality and feasibility of cohort studies. Tartu, 2000.
67. **Tamara Vorobjova**. Immune response to *Helicobacter pylori* and its association with dynamics of chronic gastritis and epithelial cell turnover in antrum and corpus. Tartu, 2001.
68. **Ruth Kalda**. Structure and outcome of family practice quality in the changing health care system of Estonia. Tartu, 2001.
69. **Annika Krüüner**. *Mycobacterium tuberculosis* – spread and drug resistance in Estonia. Tartu, 2001.
70. **Marlit Veldi**. Obstructive Sleep Apnoea: Computerized Endopharyngeal Myotonometry of the Soft Palate and Lingual Musculature. Tartu, 2001.
71. **Anneli Uusküla**. Epidemiology of sexually transmitted diseases in Estonia in 1990–2000. Tartu, 2001.
72. **Ade Kallas**. Characterization of antibodies to coagulation factor VIII. Tartu, 2002.
73. **Heidi Annuk**. Selection of medicinal plants and intestinal lactobacilli as antimicrobil components for functional foods. Tartu, 2002.
74. **Aet Lukmann**. Early rehabilitation of patients with ischaemic heart disease after surgical revascularization of the myocardium: assessment of health-related quality of life, cardiopulmonary reserve and oxidative stress. A clinical study. Tartu, 2002.
75. **Maigi Eisen**. Pathogenesis of Contact Dermatitis: participation of Oxidative Stress. A clinical – biochemical study. Tartu, 2002.
76. **Piret Hussar**. Histology of the post-traumatic bone repair in rats. Elaboration and use of a new standardized experimental model – bicortical perforation of tibia compared to internal fracture and resection osteotomy. Tartu, 2002.
77. **Tõnu Rätsep**. Aneurysmal subarachnoid haemorrhage: Noninvasive monitoring of cerebral haemodynamics. Tartu, 2002.
78. **Marju Herodes**. Quality of life of people with epilepsy in Estonia. Tartu, 2003.
79. **Katre Maasalu**. Changes in bone quality due to age and genetic disorders and their clinical expressions in Estonia. Tartu, 2003.
80. **Toomas Sillakivi**. Perforated peptic ulcer in Estonia: epidemiology, risk factors and relations with *Helicobacter pylori*. Tartu, 2003.
81. **Leena Puksa**. Late responses in motor nerve conduction studies. F and A waves in normal subjects and patients with neuropathies. Tartu, 2003.
82. **Krista Lõivukene**. *Helicobacter pylori* in gastric microbial ecology and its antimicrobial susceptibility pattern. Tartu, 2003.

83. **Helgi Kolk**. Dyspepsia and *Helicobacter pylori* infection: the diagnostic value of symptoms, treatment and follow-up of patients referred for upper gastrointestinal endoscopy by family physicians. Tartu, 2003.
84. **Helena Soomer**. Validation of identification and age estimation methods in forensic odontology. Tartu, 2003.
85. **Kersti Oselin**. Studies on the human MDR1, MRP1, and MRP2 ABC transporters: functional relevance of the genetic polymorphisms in the *MDR1* and *MRP1* gene. Tartu, 2003.
86. **Jaan Soplepmann**. Peptic ulcer haemorrhage in Estonia: epidemiology, prognostic factors, treatment and outcome. Tartu, 2003.
87. **Margot Peetsalu**. Long-term follow-up after vagotomy in duodenal ulcer disease: recurrent ulcer, changes in the function, morphology and *Helicobacter pylori* colonisation of the gastric mucosa. Tartu, 2003.
88. **Kersti Klaamas**. Humoral immune response to *Helicobacter pylori* a study of host-dependent and microbial factors. Tartu, 2003.
89. **Pille Taba**. Epidemiology of Parkinson's disease in Tartu, Estonia. Prevalence, incidence, clinical characteristics, and pharmacoepidemiology. Tartu, 2003.
90. **Alar Veraksitš**. Characterization of behavioural and biochemical phenotype of cholecystokinin-2 receptor deficient mice: changes in the function of the dopamine and endopioidergic system. Tartu, 2003.
91. **Ingrid Kalev**. CC-chemokine receptor 5 (CCR5) gene polymorphism in Estonians and in patients with Type I and Type II diabetes mellitus. Tartu, 2003.
92. **Lumme Kadaja**. Molecular approach to the regulation of mitochondrial function in oxidative muscle cells. Tartu, 2003.
93. **Aive Liigant**. Epidemiology of primary central nervous system tumours in Estonia from 1986 to 1996. Clinical characteristics, incidence, survival and prognostic factors. Tartu, 2004.
94. **Andres, Kulla**. Molecular characteristics of mesenchymal stroma in human astrocytic gliomas. Tartu, 2004.
95. **Mari Järvelaid**. Health damaging risk behaviours in adolescence. Tartu, 2004.
96. **Ülle Pechter**. Progression prevention strategies in chronic renal failure and hypertension. An experimental and clinical study. Tartu, 2004.
97. **Gunnar Tasa**. Polymorphic glutathione S-transferases – biology and role in modifying genetic susceptibility to senile cataract and primary open angle glaucoma. Tartu, 2004.
98. **Tuuli Käämbre**. Intracellular energetic unit: structural and functional aspects. Tartu, 2004.
99. **Vitali Vassiljev**. Influence of nitric oxide syntase inhibitors on the effects of ethanol after acute and chronic ethanol administration and withdrawal. Tartu, 2004.

100. **Aune Rehema**. Assessment of nonhaem ferrous iron and glutathione redox ratio as markers of pathogeneticity of oxidative stress in different clinical groups. Tartu, 2004.
101. **Evelin Seppet**. Interaction of mitochondria and ATPases in oxidative muscle cells in normal and pathological conditions. Tartu, 2004.
102. **Eduard Maron**. Serotonin function in panic disorder: from clinical experiments to brain imaging and genetics. Tartu, 2004.
103. **Marje Oona**. *Helicobacter pylori* infection in children: epidemiological and therapeutic aspects. Tartu, 2004.
104. **Kersti Kokk**. Regulation of active and passive molecular transport in the testis. Tartu, 2005.
105. **Vladimir Järv**. Cross-sectional imaging for pretreatment evaluation and follow-up of pelvic malignant tumours. Tartu, 2005.
106. **Andre Õun**. Epidemiology of adult epilepsy in Tartu, Estonia. Incidence, prevalence and medical treatment. Tartu, 2005.
107. **Piibe Muda**. Homocysteine and hypertension: associations between homocysteine and essential hypertension in treated and untreated hypertensive patients with and without coronary artery disease. Tartu, 2005.
108. **Külli Kingo**. The interleukin-10 family cytokines gene polymorphisms in plaque psoriasis. Tartu, 2005.
109. **Mati Merila**. Anatomy and clinical relevance of the glenohumeral joint capsule and ligaments. Tartu, 2005.
110. **Epp Songisepp**. Evaluation of technological and functional properties of the new probiotic *Lactobacillus fermentum* ME-3. Tartu, 2005.
111. **Tiia Ainla**. Acute myocardial infarction in Estonia: clinical characteristics, management and outcome. Tartu, 2005.
112. **Andres Sell**. Determining the minimum local anaesthetic requirements for hip replacement surgery under spinal anaesthesia – a study employing a spinal catheter. Tartu, 2005.
113. **Tiia Tamme**. Epidemiology of odontogenic tumours in Estonia. Pathogenesis and clinical behaviour of ameloblastoma. Tartu, 2005.
114. **Triine Annus**. Allergy in Estonian schoolchildren: time trends and characteristics. Tartu, 2005.
115. **Tiia Voor**. Microorganisms in infancy and development of allergy: comparison of Estonian and Swedish children. Tartu, 2005.
116. **Priit Kasenõmm**. Indicators for tonsillectomy in adults with recurrent tonsillitis – clinical, microbiological and pathomorphological investigations. Tartu, 2005.
117. **Eva Zusinaite**. Hepatitis C virus: genotype identification and interactions between viral proteases. Tartu, 2005.
118. **Piret Kõll**. Oral lactoflora in chronic periodontitis and periodontal health. Tartu, 2006.
119. **Tiina Stelmach**. Epidemiology of cerebral palsy and unfavourable neurodevelopmental outcome in child population of Tartu city and county, Estonia Prevalence, clinical features and risk factors. Tartu, 2006.

120. **Katrin Pudersell**. Tropane alkaloid production and riboflavine excretion in the field and tissue cultures of henbane (*Hyoscyamus niger* L.). Tartu, 2006.
121. **Külli Jaako**. Studies on the role of neurogenesis in brain plasticity. Tartu, 2006.
122. **Aare Märtson**. Lower limb lengthening: experimental studies of bone regeneration and long-term clinical results. Tartu, 2006.
123. **Heli Tähepõld**. Patient consultation in family medicine. Tartu, 2006.
124. **Stanislav Liskmann**. Peri-implant disease: pathogenesis, diagnosis and treatment in view of both inflammation and oxidative stress profiling. Tartu, 2006.
125. **Ruth Rudissaar**. Neuropharmacology of atypical antipsychotics and an animal model of psychosis. Tartu, 2006.
126. **Helena Andreson**. Diversity of *Helicobacter pylori* genotypes in Estonian patients with chronic inflammatory gastric diseases. Tartu, 2006.
127. **Katrin Pruus**. Mechanism of action of antidepressants: aspects of serotoninergic system and its interaction with glutamate. Tartu, 2006.
128. **Priit Põder**. Clinical and experimental investigation: relationship of ischaemia/reperfusion injury with oxidative stress in abdominal aortic aneurysm repair and in extracranial brain artery endarterectomy and possibilities of protection against ischaemia using a glutathione analogue in a rat model of global brain ischaemia. Tartu, 2006.
129. **Marika Tammaru**. Patient-reported outcome measurement in rheumatoid arthritis. Tartu, 2006.
130. **Tiia Reimand**. Down syndrome in Estonia. Tartu, 2006.
131. **Diva Eensoo**. Risk-taking in traffic and Markers of Risk-Taking Behaviour in Schoolchildren and Car Drivers. Tartu, 2007.
132. **Riina Vibo**. The third stroke registry in Tartu, Estonia from 2001 to 2003: incidence, case-fatality, risk factors and long-term outcome. Tartu, 2007.
133. **Chris Pruunsild**. Juvenile idiopathic arthritis in children in Estonia. Tartu, 2007.
134. **Eve Õiglane-Šlik**. Angelman and Prader-Willi syndromes in Estonia. Tartu, 2007.
135. **Kadri Haller**. Antibodies to follicle stimulating hormone. Significance in female infertility. Tartu, 2007.
136. **Pille Ööpik**. Management of depression in family medicine. Tartu, 2007.
137. **Jaak Kals**. Endothelial function and arterial stiffness in patients with atherosclerosis and in healthy subjects. Tartu, 2007.
138. **Priit Kampus**. Impact of inflammation, oxidative stress and age on arterial stiffness and carotid artery intima-media thickness. Tartu, 2007.
139. **Margus Punab**. Male fertility and its risk factors in Estonia. Tartu, 2007.
140. **Alar Toom**. Heterotopic ossification after total hip arthroplasty: clinical and pathogenetic investigation. Tartu, 2007.

141. **Lea Pehme**. Epidemiology of tuberculosis in Estonia 1991–2003 with special regard to extrapulmonary tuberculosis and delay in diagnosis of pulmonary tuberculosis. Tartu, 2007.
142. **Juri Karjagin**. The pharmacokinetics of metronidazole and meropenem in septic shock. Tartu, 2007.
143. **Inga Talvik**. Inflicted traumatic brain injury shaken baby syndrome in Estonia – epidemiology and outcome. Tartu, 2007.
144. **Tarvo Rajasalu**. Autoimmune diabetes: an immunological study of type 1 diabetes in humans and in a model of experimental diabetes (in RIP-B7.1 mice). Tartu, 2007.
145. **Inga Karu**. Ischaemia-reperfusion injury of the heart during coronary surgery: a clinical study investigating the effect of hyperoxia. Tartu, 2007.
146. **Peeter Padrik**. Renal cell carcinoma: Changes in natural history and treatment of metastatic disease. Tartu, 2007.
147. **Neve Vendt**. Iron deficiency and iron deficiency anaemia in infants aged 9 to 12 months in Estonia. Tartu, 2008.
148. **Lenne-Triin Heidmets**. The effects of neurotoxins on brain plasticity: focus on neural Cell Adhesion Molecule. Tartu, 2008.
149. **Paul Korrovits**. Asymptomatic inflammatory prostatitis: prevalence, etiological factors, diagnostic tools. Tartu, 2008.
150. **Annika Reintam**. Gastrointestinal failure in intensive care patients. Tartu, 2008.
151. **Kristiina Roots**. Cationic regulation of Na-pump in the normal, Alzheimer's and CCK_2 receptor-deficient brain. Tartu, 2008.
152. **Helen Puusepp**. The genetic causes of mental retardation in Estonia: fragile X syndrome and creatine transporter defect. Tartu, 2009.
153. **Kristiina Rull**. Human chorionic gonadotropin beta genes and recurrent miscarriage: expression and variation study. Tartu, 2009.
154. **Margus Eimre**. Organization of energy transfer and feedback regulation in oxidative muscle cells. Tartu, 2009.
155. **Maire Link**. Transcription factors FoxP3 and AIRE: autoantibody associations. Tartu, 2009.
156. **Kai Haldre**. Sexual health and behaviour of young women in Estonia. Tartu, 2009.
157. **Kaur Liivak**. Classical form of congenital adrenal hyperplasia due to 21-hydroxylase deficiency in Estonia: incidence, genotype and phenotype with special attention to short-term growth and 24-hour blood pressure. Tartu, 2009.
158. **Kersti Ehrlich**. Antioxidative glutathione analogues (UPF peptides) – molecular design, structure-activity relationships and testing the protective properties. Tartu, 2009.
159. **Anneli Rätsep**. Type 2 diabetes care in family medicine. Tartu, 2009.
160. **Silver Türk**. Etiopathogenetic aspects of chronic prostatitis: role of mycoplasmas, coryneform bacteria and oxidative stress. Tartu, 2009.

161. **Kaire Heilman**. Risk markers for cardiovascular disease and low bone mineral density in children with type 1 diabetes. Tartu, 2009.
162. **Kristi Rüütel**. HIV-epidemic in Estonia: injecting drug use and quality of life of people living with HIV. Tartu, 2009.
163. **Triin Eller**. Immune markers in major depression and in antidepressive treatment. Tartu, 2009.
164. **Siim Suutre**. The role of TGF-β isoforms and osteoprogenitor cells in the pathogenesis of heterotopic ossification. An experimental and clinical study of hip arthroplasty. Tartu, 2010.
165. **Kai Kliiman**. Highly drug-resistant tuberculosis in Estonia: Risk factors and predictors of poor treatment outcome. Tartu, 2010.
166. **Inga Villa**. Cardiovascular health-related nutrition, physical activity and fitness in Estonia. Tartu, 2010.
167. **Tõnis Org**. Molecular function of the first PHD finger domain of Autoimmune Regulator protein. Tartu, 2010.
168. **Tuuli Metsvaht**. Optimal antibacterial therapy of neonates at risk of early onset sepsis. Tartu, 2010.
169. **Jaanus Kahu**. Kidney transplantation: Studies on donor risk factors and mycophenolate mofetil. Tartu, 2010.
170. **Koit Reimand**. Autoimmunity in reproductive failure: A study on associated autoantibodies and autoantigens. Tartu, 2010.
171. **Mart Kull**. Impact of vitamin D and hypolactasia on bone mineral density: a population based study in Estonia. Tartu, 2010.
172. **Rael Laugesaar**. Stroke in children – epidemiology and risk factors. Tartu, 2010.
173. **Mark Braschinsky**. Epidemiology and quality of life issues of hereditary spastic paraplegia in Estonia and implemention of genetic analysis in everyday neurologic practice. Tartu, 2010.
174. **Kadri Suija**. Major depression in family medicine: associated factors, recurrence and possible intervention. Tartu, 2010.
175. **Jarno Habicht**. Health care utilisation in Estonia: socioeconomic determinants and financial burden of out-of-pocket payments. Tartu, 2010.
176. **Kristi Abram**. The prevalence and risk factors of rosacea. Subjective disease perception of rosacea patients. Tartu, 2010.
177. **Malle Kuum**. Mitochondrial and endoplasmic reticulum cation fluxes: Novel roles in cellular physiology. Tartu, 2010.
178. **Rita Teek**. The genetic causes of early onset hearing loss in Estonian children. Tartu, 2010.
179. **Daisy Volmer**. The development of community pharmacy services in Estonia – public and professional perceptions 1993–2006. Tartu, 2010.
180. **Jelena Lissitsina**. Cytogenetic causes in male infertility. Tartu, 2011.
181. **Delia Lepik**. Comparison of gunshot injuries caused from Tokarev, Makarov and Glock 19 pistols at different firing distances. Tartu, 2011.
182. **Ene-Renate Pähkla**. Factors related to the efficiency of treatment of advanced periodontitis. Tartu, 2011.

183. **Maarja Krass**. L-Arginine pathways and antidepressant action. Tartu, 2011.
184. **Taavi Lai**. Population health measures to support evidence-based health policy in Estonia. Tartu, 2011.
185. **Tiit Salum**. Similarity and difference of temperature-dependence of the brain sodium pump in normal, different neuropathological, and aberrant conditions and its possible reasons. Tartu, 2011.
186. **Tõnu Vooder**. Molecular differences and similarities between histological subtypes of non-small cell lung cancer. Tartu, 2011.
187. **Jelena Štšepetova**. The characterisation of intestinal lactic acid bacteria using bacteriological, biochemical and molecular approaches. Tartu, 2011.
188. **Radko Avi**. Natural polymorphisms and transmitted drug resistance in Estonian HIV-1 CRF06_cpx and its recombinant viruses. Tartu, 2011, 116 p.
189. **Edward Laane**. Multiparameter flow cytometry in haematological malignancies. Tartu, 2011, 152 p.
190. **Triin Jagomägi**. A study of the genetic etiology of nonsyndromic cleft lip and palate. Tartu, 2011, 158 p.
191. **Ivo Laidmäe**. Fibrin glue of fish (*Salmo salar*) origin: immunological study and development of new pharmaceutical preparation. Tartu, 2012, 150 p.
192. **Ülle Parm**. Early mucosal colonisation and its role in prediction of invasive infection in neonates at risk of early onset sepsis. Tartu, 2012, 168 p.
193. **Kaupo Teesalu**. Autoantibodies against desmin and transglutaminase 2 in celiac disease: diagnostic and functional significance. Tartu, 2012, 142 p.
194. **Maksim Zagura**. Biochemical, functional and structural profiling of arterial damage in atherosclerosis. Tartu, 2012, 162 p.
195. **Vivian Kont**. Autoimmune regulator: characterization of thymic gene regulation and promoter methylation. Tartu, 2012, 134 p.
196. **Pirje Hütt**. Functional properties, persistence, safety and efficacy of potential probiotic lactobacilli. Tartu, 2012, 246 p.
197. **Innar Tõru**. Serotonergic modulation of CCK-4- induced panic. Tartu, 2012, 132 p.
198. **Sigrid Vorobjov**. Drug use, related risk behaviour and harm reduction interventions utilization among injecting drug users in Estonia: implications for drug policy. Tartu, 2012, 120 p.
199. **Martin Serg**. Therapeutic aspects of central haemodynamics, arterial stiffness and oxidative stress in hypertension. Tartu, 2012, 156 p.
200. **Jaanika Kumm**. Molecular markers of articular tissues in early knee osteoarthritis: a population-based longitudinal study in middle-aged subjects. Tartu, 2012, 159 p.
201. **Kertu Rünkorg**. Functional changes of dopamine, endopioid and endocannabinoid systems in CCK2 receptor deficient mice. Tartu, 2012, 125 p.
202. **Mai Blöndal**. Changes in the baseline characteristics, management and outcomes of acute myocardial infarction in Estonia. Tartu, 2012, 127 p.

203. **Jana Lass**. Epidemiological and clinical aspects of medicines use in children in Estonia. Tartu, 2012, 170 p.
204. **Kai Truusalu**. Probiotic lactobacilli in experimental persistent *Salmonella* infection. Tartu, 2013, 139 p.
205. **Oksana Jagur**. Temporomandibular joint diagnostic imaging in relation to pain and bone characteristics. Long-term results of arthroscopic treatment. Tartu, 2013, 126 p.
206. **Katrin Sikk**. Manganese-ephedrone intoxication – pathogenesis of neurological damage and clinical symptomatology. Tartu, 2013, 125 p.
207. **Kai Blöndal**. Tuberculosis in Estonia with special emphasis on drug-resistant tuberculosis: Notification rate, disease recurrence and mortality. Tartu, 2013, 151 p.
208. **Marju Puurand**. Oxidative phosphorylation in different diseases of gastric mucosa. Tartu, 2013, 123 p.
209. **Aili Tagoma**. Immune activation in female infertility: Significance of autoantibodies and inflammatory mediators. Tartu, 2013, 135 p.
210. **Liis Sabre**. Epidemiology of traumatic spinal cord injury in Estonia. Brain activation in the acute phase of traumatic spinal cord injury. Tartu, 2013, 135 p.
211. **Merit Lamp**. Genetic susceptibility factors in endometriosis. Tartu, 2013, 125 p.
212. **Erik Salum**. Beneficial effects of vitamin D and angiotensin II receptor blocker on arterial damage. Tartu, 2013, 167 p.
213. **Maire Karelson**. Vitiligo: clinical aspects, quality of life and the role of melanocortin system in pathogenesis. Tartu, 2013, 153 p.
214. **Kuldar Kaljurand**. Prevalence of exfoliation syndrome in Estonia and its clinical significance. Tartu, 2013, 113 p.
215. **Raido Paasma**. Clinical study of methanol poisoning: handling large outbreaks, treatment with antidotes, and long-term outcomes. Tartu, 2013, 96 p.
216. **Anne Kleinberg**. Major depression in Estonia: prevalence, associated factors, and use of health services. Tartu, 2013, 129 p.
217. **Triin Eglit**. Obesity, impaired glucose regulation, metabolic syndrome and their associations with high-molecular-weight adiponectin levels. Tartu, 2014, 115 p.
218. **Kristo Ausmees**. Reproductive function in middle-aged males: Associations with prostate, lifestyle and couple infertility status. Tartu, 2014, 125 p.
219. **Kristi Huik**. The influence of host genetic factors on the susceptibility to HIV and HCV infections among intravenous drug users. Tartu, 2014, 144 p.
220. **Liina Tserel**. Epigenetic profiles of monocytes, monocyte-derived macrophages and dendritic cells. Tartu, 2014, 143 p.
221. **Irina Kerna**. The contribution of *ADAM12* and *CILP* genes to the development of knee osteoarthritis. Tartu, 2014, 152 p.

222. **Ingrid Liiv**. Autoimmune regulator protein interaction with DNA-dependent protein kinase and its role in apoptosis. Tartu, 2014, 143 p.
223. **Liivi Maddison**. Tissue perfusion and metabolism during intra-abdominal hypertension. Tartu, 2014, 103 p.
224. **Krista Ress**. Childhood coeliac disease in Estonia, prevalence in atopic dermatitis and immunological characterisation of coexistence. Tartu, 2014, 124 p.
225. **Kai Muru**. Prenatal screening strategies, long-term outcome of children with marked changes in maternal screening tests and the most common syndromic heart anomalies in Estonia. Tartu, 2014, 189 p.
226. **Kaja Rahu**. Morbidity and mortality among Baltic Chernobyl cleanup workers: a register-based cohort study. Tartu, 2014, 155 p.
227. **Klari Noormets**. The development of diabetes mellitus, fertility and energy metabolism disturbances in a Wfs1-deficient mouse model of Wolfram syndrome. Tartu, 2014, 132 p.
228. **Liis Toome**. Very low gestational age infants in Estonia. Tartu, 2014, 183 p.
229. **Ceith Nikkolo**. Impact of different mesh parameters on chronic pain and foreign body feeling after open inguinal hernia repair. Tartu, 2014, 132 p.
230. **Vadim Brjalin**. Chronic hepatitis C: predictors of treatment response in Estonian patients. Tartu, 2014, 122 p.
231. **Vahur Metsna**. Anterior knee pain in patients following total knee arthroplasty: the prevalence, correlation with patellar cartilage impairment and aspects of patellofemoral congruence. Tartu, 2014, 130 p.
232. **Marju Kase**. Glioblastoma multiforme: possibilities to improve treatment efficacy. Tartu, 2015, 137 p.
233. **Riina Runnel**. Oral health among elementary school children and the effects of polyol candies on the prevention of dental caries. Tartu, 2015, 112 p.
234. **Made Laanpere**. Factors influencing women's sexual health and reproductive choices in Estonia. Tartu, 2015, 176 p.
235. **Andres Lust**. Water mediated solid state transformations of a polymorphic drug – effect on pharmaceutical product performance. Tartu, 2015, 134 p.
236. **Anna Klugman**. Functionality related characterization of pretreated wood lignin, cellulose and polyvinylpyrrolidone for pharmaceutical applications. Tartu, 2015, 156 p.
237. **Triin Laisk-Podar**. Genetic variation as a modulator of susceptibility to female infertility and a source for potential biomarkers. Tartu, 2015, 155 p.
238. **Mailis Tõnisson**. Clinical picture and biochemical changes in blood in children with acute alcohol intoxication. Tartu, 2015, 100 p.
239. **Kadri Tamme**. High volume haemodiafiltration in treatment of severe sepsis – impact on pharmacokinetics of antibiotics and inflammatory response. Tartu, 2015, 133 p.

240. **Kai Part**. Sexual health of young people in Estonia in a social context: the role of school-based sexuality education and youth-friendly counseling services. Tartu, 2015, 203 p.
241. **Urve Paaver**. New perspectives for the amorphization and physical stabilization of poorly water-soluble drugs and understanding their dissolution behavior. Tartu, 2015, 139 p.
242. **Aleksandr Peet**. Intrauterine and postnatal growth in children with HLA-conferred susceptibility to type 1 diabetes. Tartu. 2015, 146 p.
243. **Piret Mitt**. Healthcare-associated infections in Estonia – epidemiology and surveillance of bloodstream and surgical site infections. Tartu, 2015, 145 p.
244. **Merli Saare**. Molecular Profiling of Endometriotic Lesions and Endometria of Endometriosis Patients. Tartu, 2016, 129 p.
245. **Kaja-Triin Laisaar**. People living with HIV in Estonia: Engagement in medical care and methods of increasing adherence to antiretroviral therapy and safe sexual behavior. Tartu, 2016, 132 p.
246. **Eero Merilind**. Primary health care performance: impact of payment and practice-based characteristics. Tartu, 2016, 120 p.
247. **Jaanika Kärner**. Cytokine-specific autoantibodies in AIRE deficiency. Tartu, 2016, 182 p.
248. **Kaido Paapstel**. Metabolomic profile of arterial stiffness and early biomarkers of renal damage in atherosclerosis. Tartu, 2016, 173 p.
249. **Liidia Kiisk**. Long-term nutritional study: anthropometrical and clinico-laboratory assessments in renal replacement therapy patients after intensive nutritional counselling. Tartu, 2016, 207 p.
250. **Georgi Nellis**. The use of excipients in medicines administered to neonates in Europe. Tartu, 2017, 159 p.
251. **Aleksei Rakitin**. Metabolic effects of acute and chronic treatment with valproic acid in people with epilepsy. Tartu, 2017, 125 p.
252. **Eveli Kallas**. The influence of immunological markers to susceptibility to HIV, HBV, and HCV infections among persons who inject drugs. Tartu, 2017, 138 p.
253. **Tiina Freimann**. Musculoskeletal pain among nurses: prevalence, risk factors, and intervention. Tartu, 2017, 125 p.
254. **Evelyn Aaviksoo**. Sickness absence in Estonia: determinants and influence of the sick-pay cut reform. Tartu, 2017, 121 p.
255. **Kalev Nõupuu**. Autosomal-recessive Stargardt disease: phenotypic heterogeneity and genotype-phenotype associations. Tartu, 2017, 131 p.
256. **Ho Duy Binh**. Osteogenesis imperfecta in Vietnam. Tartu, 2017, 125 p.
257. **Uku Haljasorg**. Transcriptional mechanisms in thymic central tolerance. Tartu, 2017, 147 p.
258. **Živile Riispere**. IgA Nephropathy study according to the Oxford Classification: IgA Nephropathy clinical-morphological correlations, disease progression and the effect of renoprotective therapy. Tartu, 2017, 129 p.

259. **Hiie Soeorg**. Coagulase-negative staphylococci in gut of preterm neonates and in breast milk of their mothers. Tartu, 2017, 216 p.
260. **Anne-Mari Anton Willmore**. Silver nanoparticles for cancer research. Tartu, 2017, 132 p.
261. **Ott Laius**. Utilization of osteoporosis medicines, medication adherence and the trend in osteoporosis related hip fractures in Estonia. Tartu, 2017, 134 p.
262. **Alar Aab**. Insights into molecular mechanisms of asthma and atopic dermatitis. Tartu, 2017, 164 p.
263. **Sander Pajusalu**. Genome-wide diagnostics of Mendelian disorders: from chromosomal microarrays to next-generation sequencing. Tartu, 2017, 146 p.
264. **Mikk Jürisson**. Health and economic impact of hip fracture in Estonia. Tartu, 2017, 164 p.
265. **Kaspar Tootsi**. Cardiovascular and metabolomic profiling of osteoarthritis. Tartu, 2017, 150 p.
266. **Mario Saare**. The influence of AIRE on gene expression – studies of transcriptional regulatory mechanisms in cell culture systems. Tartu, 2017, 172 p.
267. **Piia Jõgi**. Epidemiological and clinical characteristics of pertussis in Estonia. Tartu, 2018, 168 p.
268. **Elle Põldoja**. Structure and blood supply of the superior part of the shoulder joint capsule. Tartu, 2018, 116 p.
269. **Minh Son Nguyen**. Oral health status and prevalence of temporomandibular disorders in 65–74-year-olds in Vietnam. Tartu, 2018, 182 p.
270. **Kristian Semjonov**. Development of pharmaceutical quench-cooled molten and melt-electrospun solid dispersions for poorly water-soluble indomethacin. Tartu, 2018, 125 p.
271. **Janne Tiigimäe-Saar**. Botulinum neurotoxin type A treatment for sialorrhea in central nervous system diseases. Tartu, 2018, 109 p.
272. **Veiko Vengerfeldt**. Apical periodontitis: prevalence and etiopathogenetic aspects. Tartu, 2018, 150 p.
273. **Rudolf Bichele**. TNF superfamily and AIRE at the crossroads of thymic differentiation and host protection against *Candida albicans* infection. Tartu, 2018, 153 p.
274. **Olga Tšuiko**. Unravelling Chromosomal Instability in Mammalian Preimplantation Embryos Using Single-Cell Genomics. Tartu, 2018, 169 p.
275. **Kärt Kriisa**. Profile of acylcarnitines, inflammation and oxidative stress in first-episode psychosis before and after antipsychotic treatment. Tartu, 2018, 145 p.
276. **Xuan Dung Ho**. Characterization of the genomic profile of osteosarcoma. Tartu, 2018, 144 p.
277. **Karit Reinson**. New Diagnostic Methods for Early Detection of Inborn Errors of Metabolism in Estonia. Tartu, 2018, 201 p.

278. **Mari-Anne Vals**. Congenital N-glycosylation Disorders in Estonia. Tartu, 2019, 148 p.
279. **Liis Kadastik-Eerme**. Parkinson's disease in Estonia: epidemiology, quality of life, clinical characteristics and pharmacotherapy. Tartu, 2019, 202 p.
280. **Hedi Hunt**. Precision targeting of intraperitoneal tumors with peptide-guided nanocarriers. Tartu, 2019, 179 p.
281. **Rando Porosk**. The role of oxidative stress in Wolfram syndrome 1 and hypothermia. Tartu, 2019, 123 p.
282. **Ene-Ly Jõgeda**. The influence of coinfections and host genetic factor on the susceptibility to HIV infection among people who inject drugs. Tartu, 2019, 126 p.
283. **Kristel Ehala-Aleksejev**. The associations between body composition, obesity and obesity-related health and lifestyle conditions with male reproductive function. Tartu, 2019, 138 p.
284. **Aigar Ottas**. The metabolomic profiling of psoriasis, atopic dermatitis and atherosclerosis. Tartu, 2019, 136 p.
285. **Elmira Gurbanova**. Specific characteristics of tuberculosis in low default, but high multidrug–resistance prison setting. Tartu, 2019, 129 p.
286. **Van Thai Nguyeni**. The first study of the treatment outcomes of patients with cleft lip and palate in Central Vietnam. Tartu, 2019, 144 p.
287. **Maria Yakoreva**. Imprinting Disorders in Estonia. Tartu, 2019, 187 p.
288. **Kadri Rekker**. The putative role of microRNAs in endometriosis pathogenesis and potential in diagnostics. Tartu, 2019, 140 p.
289. **Ülle Võhma**. Association between personality traits, clinical characteristics and pharmacological treatment response in panic disorder. Tartu, 2019, 121 p.
290. **Aet Saar**. Acute myocardial infarction in Estonia 2001–2014: towards risk-based prevention and management. Tartu, 2019, 124 p.
291. **Toomas Toomsoo**. Transcranial brain sonography in the Estonian cohort of Parkinson's disease. Tartu, 2019, 114 p.
292. **Lidiia Zhytnik**. Inter- and intrafamilial diversity based on genotype and phenotype correlations of Osteogenesis Imperfecta. Tartu, 2019, 224 p.
293. **Pilleriin Soodla**. Newly HIV-infected people in Estonia: estimation of incidence and transmitted drug resistance. Tartu, 2019, 194 p.
294. **Kristiina Ojamaa**. Epidemiology of gynecological cancer in Estonia. Tartu, 2020, 133 p.
295. **Marianne Saard**. Modern Cognitive and Social Intervention Techniques in Paediatric Neurorehabilitation for Children with Acquired Brain Injury. Tartu, 2020, 168 p.
296. **Julia Maslovskaja**. The importance of DNA binding and DNA breaks for AIRE-mediated transcriptional activation. Tartu, 2020, 162 p.
297. **Natalia Lobanovskaya**. The role of PSA-NCAM in the survival of retinal ganglion cells. Tartu, 2020, 105 p.

298. **Madis Rahu**. Structure and blood supply of the postero-superior part of the shoulder joint capsule with implementation of surgical treatment after anterior traumatic dislocation. Tartu, 2020, 104 p.
299. **Helen Zirnask**. Luteinizing hormone (LH) receptor expression in the penis and its possible role in pathogenesis of erectile disturbances. Tartu, 2020, 87 p.
300. **Kadri Toome**. Homing peptides for targeting of brain diseases. Tartu, 2020, 152 p.
301. **Maarja Hallik**. Pharmacokinetics and pharmacodynamics of inotropic drugs in neonates. Tartu, 2020, 172 p.
302. **Raili Müller**. Cardiometabolic risk profile and body composition in early rheumatoid arthritis. Tartu, 2020, 133 p.
303. **Sergo Kasvandik**. The role of proteomic changes in endometrial cells – from the perspective of fertility and endometriosis. Tartu, 2020, 191 p.
304. **Epp Kaleviste**. Genetic variants revealing the role of STAT1/STAT3 signaling cytokines in immune protection and pathology. Tartu, 2020, 189 p.
305. **Sten Saar**. Epidemiology of severe injuries in Estonia. Tartu, 2020, 104 p.
306. **Kati Braschinsky**. Epidemiology of primary headaches in Estonia and applicability of web-based solutions in headache epidemiology research. Tartu, 2020, 129 p.
307. **Helen Vaher**. MicroRNAs in the regulation of keratinocyte responses in *psoriasis vulgaris* and atopic dermatitis. Tartu, 2020, 242 p.
308. **Liisi Raam**. Molecular Alterations in the Pathogenesis of Two Chronic Dermatoses – Vitiligo and Psoriasis. Tartu, 2020, 164 p.
309. **Artur Vetkas**. Long-term quality of life, emotional health, and associated factors in patients after aneurysmal subarachnoid haemorrhage. Tartu, 2020, 127 p.
310. **Teele Kasepalu**. Effects of remote ischaemic preconditioning on organ damage and acylcarnitines' metabolism in vascular surgery. Tartu, 2020, 130 p.
311. **Prakash Lingasamy**. Development of multitargeted tumor penetrating peptides. Tartu, 2020, 246 p.
312. **Lille Kurvits.** Parkinson's disease as a multisystem disorder: whole transcriptome study in Parkinson's disease patients' skin and blood. Tartu, 2021, 142 p.
313. **Mariliis Põld.** Smoking, attitudes towards smoking behaviour, and nicotine dependence among physicians in Estonia: cross-sectional surveys 1982–2014. Tartu, 2021, 172 p.
314. **Triin Kikas**. Single nucleotide variants affecting placental gene expression and pregnancy outcome. Tartu, 2021, 160 p.
315. **Hedda Lippus-Metsaots**. Interpersonal violence in Estonia: prevalence, impact on health and health behaviour. Tartu, 2021, 172 p.

316. **Georgi Dzaparidze.** Quantification and evaluation of the diagnostic significance of adenocarcinoma-associated microenvironmental changes in the prostate using modern digital pathology solutions. Tartu, 2021, 132 p.
317. **Tuuli Sedman.** New avenues for GLP1 receptor agonists in the treatment of diabetes. Tartu, 2021, 118 p.
318. **Martin Padar.** Enteral nutrition, gastrointestinal dysfunction and intestinal biomarkers in critically ill patients. Tartu, 2021, 189 p.
319. **Siim Schneider.** Risk factors, etiology and long-term outcome in young ischemic stroke patients in Estonia. Tartu, 2021, 131 p.
320. **Konstantin Ridnõi.** Implementation and effectiveness of new prenatal diagnostic strategies in Estonia. Tartu, 2021, 191 p.
321. **Risto Vaikjärv.** Etiopathogenetic and clinical aspects of peritonsillar abscess. Tartu, 2021, 115 p.
322. **Liis Preem.** Design and characterization of antibacterial electrospun drug delivery systems for wound infections. Tartu, 2022, 220 p.
323. **Keerthie Dissanayake.** Preimplantation embryo-derived extracellular vesicles: potential as an embryo quality marker and their role during the embryo-maternal communication. Tartu, 2022, 203 p.
324. **Laura Viidik.** 3D printing in pharmaceutics: a new avenue for fabricating therapeutic drug delivery systems. Tartu, 2022, 139 p.
325. **Kasun Godakumara.** Extracellular vesicle mediated embryo-maternal communication – A tool for evaluating functional competency of preimplantation embryos. Tartu, 2022, 176 p.
326. **Hindrek Teder.** Developing computational methods and workflows for targeted and whole-genome sequencing based non-invasive prenatal testing. Tartu, 2022, 138 p.
327. **Jana Tuusov.** Deaths caused by alcohol, psychotropic and other substances in Estonia: evidence based on forensic autopsies. Tartu, 2022, 157 p.
328. **Heigo Reima.** Colorectal cancer care and outcomes – evaluation and possibilities for improvement in Estonia. Tartu, 2022, 146 p.
329. **Liisa Kuhi.** A contribution of biomarker collagen type II neoepitope C2C in urine to the diagnosis and prognosis of knee osteoarthritis. Tartu, 2022, 157 p.
330. **Reeli Tamme.** Associations between pubertal hormones and physical activity levels, and subsequent bone mineral characteristics: a longitudinal study of boys aged 12–18. Tartu, 2022, 118 p.
331. **Deniss Sõritsa.** The impact of endometriosis and physical activity on female reproduction. Tartu, 2022, 152 p.
332. **Mohammad Mehedi Hasan.** Characterization of follicular fluid-derived extracellular vesicles and their contribution to periconception environment. Tartu, 2022, 194 p.
333. **Priya Kulkarni.** Osteoarthritis pathogenesis: an immunological passage through synovium-synovial fluid axis. Tartu, 2022, 268 p.

334. **Nigul Ilves.** Brain plasticity and network reorganization in children with perinatal stroke: a functional magnetic resonance imaging study. Tartu, 2022, 169 p.
335. **Marko Murruste.** Short- and long-term outcomes of surgical management of chronic pancreatitis. Tartu, 2022, 180 p.
336. **Marilin Ivask.** Transcriptomic and metabolic changes in the WFS1-deficient mouse model. Tartu, 2022, 158 p.
337. **Jüri Lieberg.** Results of surgical treatment and role of biomarkers in pathogenesis and risk prediction in patients with abdominal aortic aneurysm and peripheral artery disease. Tartu, 2022, 160 p.
338. **Sanna Puusepp.** Comparison of molecular genetics and morphological findings of childhood-onset neuromuscular disorders. Tartu, 2022, 216 p.
339. **Khan Nguyen Viet.** Chemical composition and bioactivity of extracts and constituents isolated from the medicinal plants in Vietnam and their nanotechnology-based delivery systems. Tartu, 2023, 172 p.
340. **Getnet Balcha Midekessa.** Towards understanding the colloidal stability and detection of Extracellular Vesicles. Tartu, 2023, 172 p.
341. **Kristiina Sepp.** Competency-based and person-centred community pharmacy practice – development and implementation in Estonia. Tartu, 2023, 242 p.
342. **Linda Sõber.** Impact of thyroid disease and surgery on patient's quality of voice and swallowing. Tartu, 2023, 114 p.
343. **Anni Lepland.** Precision targeting of tumour-associated macrophages in triple negative breast cancer. Tartu, 2023, 160 p.
344. **Sirje Sammul.** Prevalence and risk factors of arterial hypertension and cardiovascular mortality: 13-year longitudinal study among 35- and 55-year-old adults in Estonia and Sweden. Tartu, 2023, 158 p.
345. **Maarjaliis Paavo**. Short-Wavelength and Near-Infrared Autofluorescence Imaging in Recessive Stargardt Disease, Choroideremia, *PROM1*-Macular Dystrophy and Ocular Albinism. Tartu, 2023, 202 p.
346. **Kaspar Ratnik.** Development of predictive multimarker test for preeclampsia in early and late pregnancy. Tartu, 2023, 134 p.
347. **Kärt Simre.** Development of coeliac disease in two populations with different environmental backgrounds. Tartu, 2023, 161 p.
348. **Qurat Ul Ain Reshi.** Characterization of the maternal reproductive tract and spermatozoa communication during periconception period via extracellular vesicles. Tartu, 2023, 182 p.
349. **Stanislav Tjagur.** *Mycoplasma genitalium* and other sexually transmitted infections causing urethritis – their prevalence, impact on male fertility parameters and prostate health. Tartu, 2023, 225 p.
350. **Lagle Lehes.** The first study of voice and resonance related treatment outcomes of Estonian cleft palate children. Tartu, 2023, 126 p.
351. **Liis Ilves.** Metabolomic profiling of chronic inflammatory skin diseases. Tartu, 2023, 146 p.

352. **Marina Šunina**. Flow cytometric analysis of T and B cell properties in healthy donors and subjects with vitiligo. Tartu, 2023, 164 p.
353. **Jaanus Suumann.** Gastric biomarkers and their dynamics as a less invasive method to evaluate stomach health in bariatric surgery patients. Tartu, 2023, 122 p.
354. **Ele Hanson.** Clinical and biochemical markers for the prediction and early diagnosis of pregnancy related complications. Tartu, 2023, 145 p.
355. **Priit Pauklin.** Hemodynamic and biochemical characteristics of patients with atrial fibrillation and anticoagulation of ≥65-year-old patients with atrial fibrillation in Estonia. Tartu, 2023, 144 p.
356. **Triinu Keskpaik.** Quality Indicators and Non-Ischemic Myocardial Injury in Emergency Medicine. Tartu, 2023, 121 p.
357. **Laura Roht.** Hereditary colorectal cancer syndromes in Estonia. Tartu, 2023, 178 p.
358. **Norman Ilves.** Risk factors and onset time of periventricular hemorrhagic infarction in preterm born children and periventricular venous infarction in term born children. Tartu, 2024, 177 p.
359. **Edgar Lipping.** Postoperative antibacterial therapy in complicated appendicitis and appendectomy in pregnancy. Tartu, 2024, 121 p.
360. **Celia Teresa Pozo Ramos.** Preparation and assessment of antimicrobial electrospun matrices for prospective applications in wound healing. Tartu, 2024, 203 p.
361. **Karl Kuusik.** Effects of remote ischaemic preconditioning on arterial stiffness, organ damage and metabolomic profile in patients with lower extremity artery disease. Tartu, 2024, 173 p.
362. **Kelli Somelar-Duracz.** The molecular and cellular mechanisms of brain plasticity impairing factors. Tartu, 2024, 245 p.
363. **Aleksei Baburin.** Breast cancer incidence, mortality and survival in Estonia in the context of health care system changes and screening. Tartu, 2024, 130 p.
364. **Marina Loid.** Molecular and cellular determinants of healthy receptive and aged endometrium. Tartu, 2024, 159 p.
365. **Ulvi Vaher.** Epilepsy after ischemic perinatal stroke in term born children: neuroimaging predictors, clinical course and cognitive outcome. Tartu, 2024, 160 p.
366. **Allan Tobi.** Development of Smart Nanoparticles for Experimental Treatment of Cancer. Tartu, 2024, 160 p.
367. **Leho Rips.** The influence of vitamin D on the physical performance of conscripts in the Estonian Defence Forces. Tartu, 2024, 147 p.
368. **Kati Kärberg.** Factors and markers predicting subclinical atherosclerosis in type 2 diabetes. Tartu, 2024, 161 p.
369. **Valeria Sidorenko.** Novel anthracycline-loaded nanoparticles for precision cancer therapy. Tartu, 2024, 183 p.

370. **Kadri Kõivumägi.** Acute gastroenteritis hospitalizations in Estonia after implementation of universal mass vaccination against rotavirus. Tartu, 2024, 150 p.
371. **Ingrid Oit-Wiscombe.** Genetic markers of enzymatics in the pathogenesis of chronic obstructive pulmonary disease as a systemic disease and the effects of antioxidant peptides. Tartu, 2025, 170 p.
372. **Gerli Mõts.** Ethical issues in nursing before and during the COVID-19 pandemic: a multi-method study. Tartu, 2025, 150 p.
373. **Annika Valner.** Changes in structure and function of extremities in early rheumatoid arthritis. Tartu, 2025, 129 p.
374. **Meruert Sarsenova.** Molecular and cellular landscape of endometriosis. Tartu, 2025, 157 p.
375. **Mailis Liiv.** Role of mitochondrial dynamics in Wolfram syndrome. Tartu, 2025, 180 p.
376. **Marta Velgan.** Addressing the family physician shortage: Career and migration intentions in Estonia and Europe. Tartu, 2025, 161 p.
377. **Ihor Filippov.** Single-cell data analysis in immunology: from technology to applications. Tartu, 2025, 204 p.
378. **Anna Tisler.** HPV-related cancers among people living with HIV and transition towards risk-based prevention. Tartu, 2025, 122 p.