Realia et naturalia

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
282

PRIIT PALTA

Computational methods for DNA copy number detection

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
282

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
282

PRIIT PALTA

Computational methods for DNA copy number detection

UNIVERSITY OF TARTU
PRESS

Institute of Molecular and Cell Biology, University of Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy in bioinformatics on 25.08.2015 by the Council of the Institute of Molecular and Cell Biology, University of Tartu.

Supervisor:
Prof. Maido Remm, Ph.D.
University of Tartu
Tartu, Estonia

Opponent:
Assoc. Prof. Lars Feuk, Ph.D.
University of Uppsala
Uppsala, Sweden

Commencement:
Room No 105, 23B Riia St., Tartu, on October 16th 2015, at 10.15.

The publication of this thesis is granted by the Institute of Molecular and Cell Biology, University of Tartu.

ISSN 1024-6479
ISBN 978-9949-32-926-7 (print)
ISBN 978-9949-32-927-4 (pdf)

Copyright: Priit Palta, 2015

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS 6

LIST OF ABBREVIATIONS 7

INTRODUCTION 8

REVIEW OF LITERATURE 9
- 1. Copy number variation in the human genome 9
  - 1.1 Mechanisms of CNV formation 11
- 2. Microarray-based CNV detection 12
  - 2.1 Nature and principles of SNP genotyping microarray data 13
  - 2.2 Computational methods for CNV detection 15
- 3. CNVs in health and disease 16
  - 3.1 Genomic impact of CNVs 16
  - 3.2 Associated trait and disease phenotypes 17
- 4. Population genetics of CNVs 18
- 5. Haplotype and allelic variability in CNV regions 19

AIMS OF THE PRESENT STUDY 21

RESULTS AND DISCUSSION 22
- 1. Statistical method for detecting copy number variants from a chromosome X-specific microarray (Ref. I) 22
- 2. Computational pipeline for CNV calling from SNP genotyping microarray data (Ref. II–IV) 24
- 3. Computational method for phasing normal and variant-carrying haplotypes within CNV regions (Ref. II and IV) 25
- 4. Transmission and allelic variability of CNV-carrying haplotypes (Ref. II and IV) 27
  - 4.1 Transmission of normal and CNV-carrying haplotypes within CNV regions in nuclear families (Ref. IV) 27
  - 4.2 Putative *de novo* copy number variants (Ref. IV) 28
  - 4.2 Allelic variability in duplication CNV regions (Ref. IV) 29

SUMMARY AND CONCLUSIONS 31

REFERENCES 32

SUMMARY IN ESTONIAN 39

ACKNOWLEDGEMENTS 41

PUBLICATIONS 43

CURRICULUM VITAE 106

ELULOOKIRJELDUS 111

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications which will be referred to in the text by their Roman numerals:

I Patsalis PC, Kousoulidou L, Männik K, Sismani C, Zilina O, Parkel S, Puusepp H, Tõnisson N, Palta P, Remm M, Kurg A. **Detection of small genomic imbalances using microarray-based multiplex amplifiable probe hybridization.** *Eur J Hum Genet.* 2007 Feb;15(2):162-72.

II Männik K, Parkel S, Palta P, Zilina O, Puusepp H, Esko T, Mägi R, Nõukas M, Veidenberg A, Nelis M, Metspalu A, Remm M, Ounap K, Kurg A. **A parallel SNP array study of genomic aberrations associated with mental retardation in patients and general population in Estonia.** *Eur J Med Genet.* 2011 Mar-Apr;54(2):136-43.

III Nagirnaja L, Palta P, Kasak L, Rull K, Christiansen OB, Nielsen HS, Steffensen R, Esko T, Remm M, Laan M. **Structural genomic variation as risk factor for idiopathic recurrent miscarriage.** *Hum Mutat.* 2014 Aug;35(8):972-82.

IV Palta P, Kaplinski L, Nagirnaja L, Veidenberg A, Möls M, Nelis M, Esko T, Metspalu A, Laan M, Remm M. **Haplotype phasing and inheritance of copy number variants in nuclear families.** *PLoS One.* 2015;10(4):e0122713.

Author's contributions:

Ref. I: Implemented hybridisation probe design software, performed *in silico* design of the microarray probes. Developed the statistical method for CNV calling and contributed to manuscript preparation.

Ref. II: Performed microarray data QC, CNV calling, filtering and analyses. Developed and implemented software for parent-of-origin analysis, allelic composition determination and performed corresponding analyses in the MR cohort. Contributed to manuscript preparation.

Ref. III: Contributed to microarray data QC, CNV calling and filtering. Performed gene set enrichment and pathway analyses for Estonian population and RM cohort case/control samples. Participated in writing of the first draft of the manuscript.

Ref. IV: Planned and conducted the study, performed Estonian sample selection, microarray data QC, CNV calling and filtering. Developed and partly implemented CNV phasing methodology, phased and analysed Estonian and HapMap family datasets, and wrote the first draft of the manuscript.

All articles are reprinted with the permission of the copyright owners.

LIST OF ABBREVIATIONS

array-CGH	microarray-based comparative genomic hybridisation
BAF	B allele frequency
BISRS	break-induced serial replication slippage
bp	base pair(s)
CGH	comparative genomic hybridization
CNP	copy number polymorphism
CNV	copy number variant
DNA	deoxyribonucleic acid
FoSTeS	fork stalling and template switching
FSI	fluorescent signal intensity
HMM	Hidden Markov model
k	kilo, thousand (1,000)
kb	1,000 base pairs
LBF	log Bayes Factor
LoF	loss-of-function
LOH	loss-of-heterozygosity
LRR	Log R ratio
MAF	minor allele frequency
MAPH	multiplex amplifiable probe hybridisation
Mb	1,000,000 base pairs
MMBIR	microhomology-mediated break-induced replication
MMRDR	microhomology-mediated replication-dependent recombination
MR	mental retardation
NAHR	non-allelic homologous recombination
NHEJ	non-homologous end joining
QC	quality control
RM	recurrent miscarriage
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
TI	tolerance interval(s)
WES	whole exome sequencing
WGS	whole genome sequencing

INTRODUCTION

Advances in sequencing and microarray technology have enabled the progress of systematic and thorough genome-wide analyses to assess the localisation and frequency of different variants in the human genome. For more than a decade, single nucleotide polymorphisms (SNPs) have been a focus of interest for human geneticists. As a result, a large number of studies have associated these single nucleotide variants with variability in different human phenotypes[1]. In addition to SNPs, short (one to a few tens of base pairs) insertions and deletions of genomic DNA sequences (commonly referred to as indels) are also present in the human genome[2–4]. During the last decade, knowledge concerning even longer unbalanced gains and losses of genomic DNA sequences (conventionally called copy number variants or CNVs) has massively increased and has also excited the interest of human geneticists[5–8]. DNA copy number variation is a type of genetic variation that may increase the copy number of a particular DNA segment from the normal two copies per diploid genome to more than two copies (e.g., duplication or triplication) or decrease it to less than two copies (deletion).

Although the extensive presence of CNVs in the human genome has been studied for the last decade[5,6], the reported extent of copy-number variation in different species and individual genomes differs greatly[7,9]. CNV lengths in the human genome typically range from a few kilo-bases to several mega-bases and are estimated to cumulatively cover at least 5% of the human genome[7,9–11]. In many instances, CNVs have been shown to play an important role in different human phenotypic traits, and their association with disease susceptibility has been increasingly recognised.

Alterations in DNA copy-number are the most likely cause of many genetic disorders (hereditary and *de novo*) and play an important role in tumourigenesis and susceptibility to common diseases. The identification of such disease-associated altered regions in the DNA provides valuable information about the genes involved in the disease and presents one step towards understanding the underlying molecular mechanisms. However, computational methods that can help investigators find and prioritise variants obtained from different experiments are needed.

In the review of literature of the current thesis, I will provide a short overview of copy number variants in the human genome, including how CNVs are discovered with current microarray-based methods, their known effects on human phenotypic traits or disease susceptibility and what is known about their population genetics.

In the results and discussion I will describe and discuss a few methods developed for CNV calling and phasing. I will also describe and discuss the inheritance of copy number variants and their allelic variability.

REVIEW OF LITERATURE

1. Copy number variation in the human genome

DNA copy number variation is a type of genetic variation in which case the number of allelic copies of a particular region of a genome is altered from its normal 'state'. In the non-repetitive portion of the human genome, the normal haploid copy number is one – one copy of each sequence per chromosome. Accordingly, the normal diploid copy number in humans is two – one copy inherited from both parents. A copy number variant (CNV) can result from either a loss of copies (most often called a deletion) or gain of copies (called a duplication or amplification). Different deletion and duplication variants are illustrated in Figure 1.

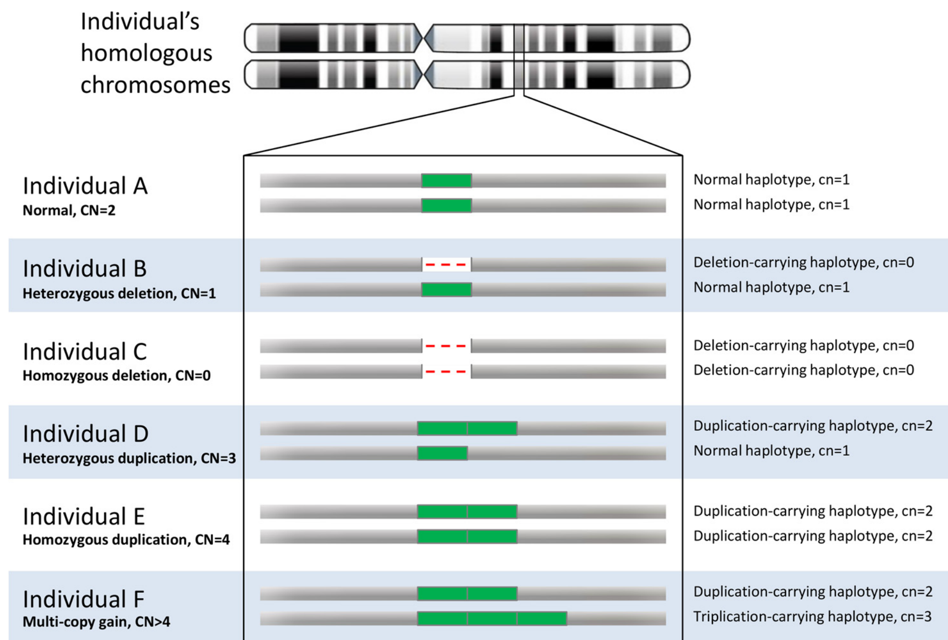

Figure 1. Copy number variants in a diploid genome. Close-up picture of a chromosomal region that contains a CNV (the green region).Whereas the grey regions are present in normal copy numbers (haploid copy number of one: cn=1; diploid copy number of two: CN=2) in individuals A-F, the green region has different haploid and diploid copy numbers in different individuals.

CNVs in the human genome range from tens of bases to several mega-bases[12,13]. One of the largest sequencing-based studies estimated that the median CNV size in the human genome was 729 bp, with mean of 8 kb[11]. Although the abundant presence of CNVs in healthy individuals has been known for many years[5,7,14], the reported number and extent of CNVs has been wide-ranging and has varied from a few to up to a few thousand per diploid genome[9,15]. CNVs are cumulatively estimated to cover 1–18% of the DNA sequence in the human

diploid genome[7,10,15–17]. Often, these discrepancies are due to differences in the technology and the experimental or computational methods used; the broad consensus is that CNVs cover approximately 5–10% of human genomic DNA sequences[9,10,18,19]. Thus, CNVs cover approximately 0.3–0.8% of the genome in every individual[9,20], corresponding to 9–24 Mb of DNA sequence.

Although copy number variants occur virtually everywhere in the human genome, there are many CNV hotspots where CNVs occur more frequently. This is well illustrated by an autosomal map of long (>100 kb) CNVs from approximately 2500 individuals from ethnically diverse populations (Figure 2).

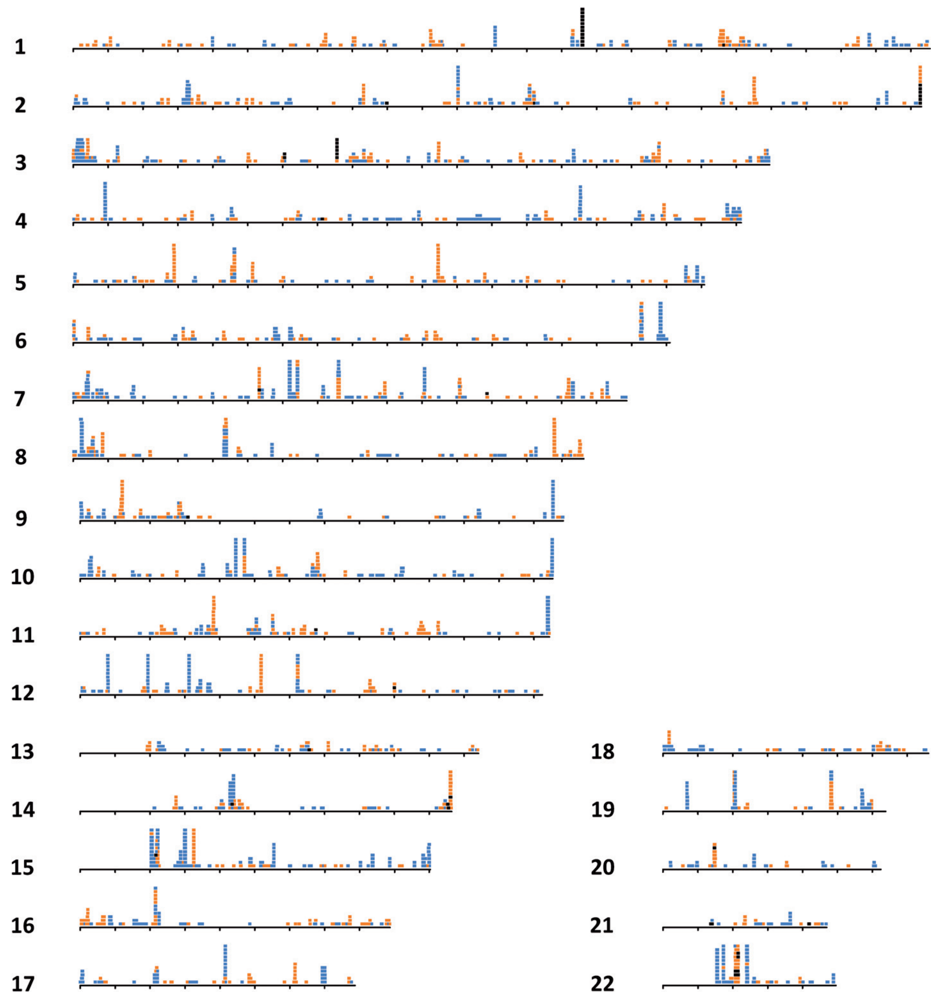

Figure 2. Autosomal landscape of large CNVs. Large (>100 kbp) copy number variants in the human genome based on analysis of 2493 individuals. Observed duplications are depicted in blue, heterozygous deletions in orange and homozygous deletions with black rectangles. Chromosomes are drawn to their size scale (outside tick marks indicate 10 Mb), and CNVs observed more frequently than ten times for a given locus are cropped (based on Itsara *et al.*, 2009).

Many CNVs in different individuals have identical recurrent breakpoints. However, some CNVs have slightly different breakpoints, often depending on the mechanism responsible for the introduction of the CNV[20].

1.1 Mechanisms of CNV formation

There are three primary classes of known mechanisms that create CNVs. A large number of copy number variants that often have recurrent breakpoints in the same locus arise by non-allelic homologous recombination (NAHR) between segmental duplication or low-copy repeats[16,21]. NAHR can result in deletions and duplications as well as inversions and translocations depending on the location and orientation of the interacting homologous sequences[22,23]. Inter-chromosomal and intra-chromosomal (inter-chromatidal) NAHR events can facilitate both deletions and duplications, whereas intra-chromatidal events result in only deletions[24,25]. Non-homologous end joining (NHEJ) is the most common DNA repair mechanism in mammalian cells; this mechanism involves ligating together any two broken DNA ends and does not require homologous sequences, although it has been shown that the presence of terminal micro-homologies also facilitates NHEJ[26]. NHEJ can result in small scale non-recurrent deletions and translocation rearrangements with minimal to no junction homology[21]. Microhomology-mediated mechanisms (collectively referred to as MMRDR) include microhomology-mediated break-induced replication (MMBIR), break-induced serial replication slippage (BISRS), and fork stalling and template switching (FoSTeS)[22,27–29]. These mechanisms are consistent with a common hypothesis that the leading/lagging strand primer/ polymerase disengages from its template upon replication fork stalling, translocates and then re-associates (probably using short microhomology at the 3′ end) to another available template/replication fork in physical proximity and resumes replication[27,30]. This phenomenon can result in non-recurrent deletions, duplications or complex rearrangements, including triplications and inversions[28].

Considering these most prevalent mutational mechanisms, it would be intuitive to propose that deletions and duplications are not generated and do not occur at equal rate in the human genome because some mechanisms give rise to only deletions. It has been estimated that deletions and duplications are likely to be generated in proportions of approximately 2:1 to 3:1, with the ratio varying widely from region to region[17]. Based on empirical genome-wide studies we know that large deletions and duplications occur with an approximately 2:1 ratio[23].

2. Microarray-based CNV detection

Despite the decreasing costs of whole-genome sequencing (WGS) and whole-exome sequencing (WES), microarrays are still extensively used in genetic studies of disease susceptibility and clinical diagnostics. Although sequencing-based applications have great deal of potential and enable investigators to interrogate almost every base pair in the studied genomes, both experimental and computational methods to infer copy number variants from sequencing data are still being developed and standardised[31,32]. Microarrays still represent the gold standard for both genotyping and CNV calling, especially in case of rare variants. Furthermore, increased sample sizes and improved analytical tools have stimulated what could be called a 'microarray-based genotyping renaissance'.

DNA copy number detection using microarray techniques is an indirect method to estimate diploid copy numbers of the studied DNA at pre-defined genomic loci by interrogating the sample with specifically designed or selected microarray probes. Briefly, the genomic DNA of interest is amplified and hybridised to a solid chip matrix carrying locus-specific capture probes. These probes most often are 25–60 bp in length and include *in situ* synthesised or spotted oligonucleotide molecules that are complementary to a specific unique sequence in the human genome assembly. In the case of array-CGH (microarray-based comparative genomic hybridisation), an additional reference DNA (often a cytogenetically controlled DNA or pool of DNAs) is co-hybridised to the same array. After the hybridisation process in which amplified and fluorescent dye-labelled genomic sequences are specifically bound to capture probes, the arrays are washed to remove any non-specifically bound material and the quantity of the specifically hybridised DNA is estimated by measuring the fluorescent signal intensity (FSI) for each capture probe by scanning the microarray chip.

FSIs from microarray probes exhibit dosage-dependent responses to alterations in the DNA copy number up to a certain level. Thus, the more copies of a certain sequences there are in the studied sample, the higher the corresponding probe's FSI level will be. In the case of array-CGH, two different fluorescent dyes are used for labelling – one to label the studied DNA and one to label the reference DNA. In the case of SNP genotyping microarrays, two different dyes are used to label different alleles of the bi-allelic SNP variants interrogated[33,34]. The most commonly used dyes for labelling are fluorescent cyanine-3 and cyanine-5 (abbreviated Cy3 and Cy5, respectively). These two fluorescent dyes are excited at different wavelengths and are quantified by a two-channel microarray scanner[33].

Because only pre-defined regions of the genome can be investigated with this method, CNV detection relies on the array design (i.e. how many probes are used and how these probes are distributed across the chromosomes). Therefore, as a rule of thumb (and quite intuitively), the use of more probes to cover all chromosomes uniformly results in a more accurate and higher resolution CNV

profile. It has been shown that the differences between platforms are more pronounced for short CNVs, whereas the differences resulting from genotyping platform largely disappear for larger (>500kb) CNVs[16]. In their high resolution study, Conrad *et al.* used custom-designed array-CGH arrays carrying 42 million oligonucleotide probes that covered the assayable portion of the genome with median spacing of 56 bp[9]. Currently (*Anno Domini* 2015), top-notch genotyping microarrays have been produced with as many as 4.3 million probes (Illumina HumanOmni5Exome, www.illumina.com) and often additionally carry non-polymorphic probes to uniformly interrogate CNVs in the human genome. For example, the Affymetrix Genome-Wide Human SNP Array 6.0 includes approximately 946 thousand CNV probes, with 200 thousand of such probes targeting common CNP loci and an additional 744 thousand probes evenly spaced along all of the chromosomes (www.affymetrix.com).

In addition to high density and coverage, one obvious advantage of SNP genotyping arrays over array-CGH is the identification of allele-specific information[34,35]. In addition to SNP genotype calling, the simultaneous measurement of total and allele-specific FSI makes it possible to detect both copy number changes and copy number neutral events, such as loss-of-heterozygosity (LOH).

Known problems with genotyping arrays are their higher intrinsic variability caused by the PCR-based approach used to amplify the studied genomic material and the optimisation of these platforms for SNP variant calling (i.e., oligonucleotide design and the hybridisation chemistry/conditions used) rather than for CNV detection[33,36]. Consequently, these confounding factors can increase the proportion of false positive (region incorrectly called a variant) and false negative (undetected true variant) CNV calls if not properly taken into account. A relatively high rate of false positive and false negative findings has raised important concerns regarding the suitability of microarray-based copy number detection for clinical diagnostics applications. Although microarrays are suitable for CNV screening, reliable assays that provide clear and high quality results of measurable significance are required in a clinical diagnostics set-up. Moreover, false findings should be minimal and their rate should be accurately predictable[37,38].

2.1 Nature and principles of SNP genotyping microarray data

Aside from the experimental microarray assay set-up, copy number detection for the studied sample can be quite complicated and cannot be regarded as an easy task. The reason for this is that even though the underlying biological nature of DNA copy numbers is always discrete (i.e., in one cell, genomic DNA copy number in a certain genomic region always corresponds to a certain non-negative integer), the measured FSIs of the studied and reference DNA material from the microarray experiment most often come from a mixture of cells. Thus, these FSI values are continuous variables with a considerable amount of non-biological variance and comprehensive and precise computational and statistical

methods are required to estimate the DNA copy number for different loci and to confidently detect CNVs[34].

As the first step in *in silico* copy number analysis, FSIs for all probes are organised by their genomic coordinates and normalised. For arrayCGH, logarithmic FSI ratios of the studied and reference DNA are used to infer CNVs[9,20,39]. For SNP genotyping arrays, FSIs from both channels (denoted X and Y values and representing different SNP alleles usually indicated as alleles A and B) for the studied DNA are normalised against a panel of reference samples[34,40,41]. For the ith probe on the microarray, the total normalised logarithmic FSI (aka the Log R ratio or LRR) is calculated as follows:

$$LRR_i = \log_2\left(\frac{R_i observed}{R_i expected}\right)$$

where $R_i observed = X_i + Y_i$ measures the total combined signal intensity of two FSI channels and $R_i expected$ is measured from a set of 'normal' reference samples.

Several adjacent positive or negative LRR values that significantly deviate from the expected value of zero corresponding to the most frequent diploid copy number of two observed in most loci for most individuals can be indicative of copy number gain (i.e., duplication and triplication) or loss (i.e., hetero- or homozygous deletion).

Another relevant measure – the B allele frequency (BAF) is the normalised measurement of the relative FSI ratio corresponding to the B and A alleles:

$$BAF_i = \begin{cases} 0, \text{if } \theta_i < \theta_{i\,\mathrm{AA}} \\ 0.5(\theta_i - \theta_{i\,\mathrm{AA}})/(\theta_{i\,\mathrm{AB}} - \theta_{i\,\mathrm{AA}}), \text{if } \theta_{i\,\mathrm{AA}} \le \theta_i < \theta_{i\,\mathrm{AB}} \\ 0.5 + 0.5(\theta_i - \theta_{i\,\mathrm{AB}})/(\theta_{i\,\mathrm{BB}} - \theta_{i\,\mathrm{AB}}), \text{if } \theta_{i\,\mathrm{AB}} \le \theta_i < \theta_{i\,\mathrm{BB}} \\ 1, \text{if } \theta_i \ge \theta_{i\,\mathrm{BB}} \end{cases}$$

where the ith SNP-specific theta-value $\theta_i = \arctan(X_i/Y_i)/(\pi/2)$ refers to the relative allelic signal intensity ratio and where $\theta_{i\,\mathrm{AA}}$, $\theta_{i\,\mathrm{AB}}$, and $\theta_{i\,\mathrm{BB}}$ are the theta values for three canonical two-letter genotype clusters generated from a large set of reference samples[42]. This transformation from theta to BAF values adjusts for different characteristics of each probe so that the BAF values for different SNPs are comparable to each other[41]. A few examples of LRR and BAF values for normal, deletion and duplication CNVs are depicted in Figure 3.

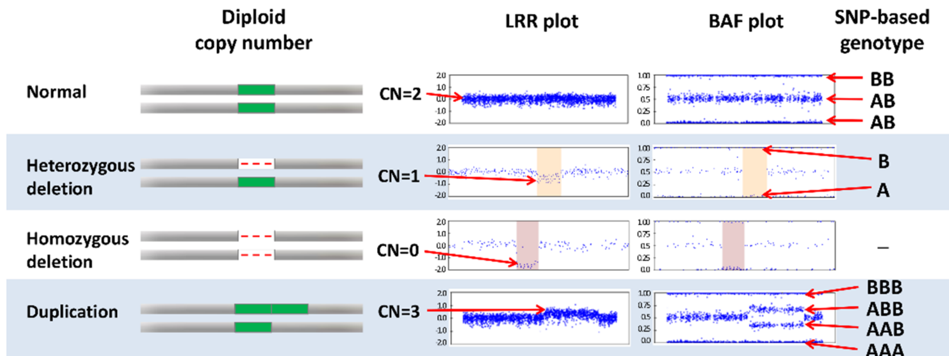

Figure 3. Examples of plots with different LRR and BAF values (and corresponding SNP genotypes) for normal (diploid copy number of two – CN=2), heterozygous deletion (CN=1), homozygous deletion (CN=0 with 'null' alleles) and duplication (CN=3) copy number variants.

2.2 Computational methods for CNV detection

A naïve method for copy number variant detection is to identify genomic regions where the consecutive microarray probes demonstrate LRR values that significantly deviate from the expected LRR value (corresponding to a diploid copy number of two). This approach is often called the 'fixed threshold method' in the pertinent literature[43–46]. The problem with this simple method is that threshold-based methods have limited specificity (i.e., many false CNV calls are made) and sensitivity (i.e., many true CNVs are missed) because microarray FSIs exhibit considerable variance due to technical reasons that are often specific for each assay and probe, resulting in a high rate of false positive and negative findings[47].

More advanced methods were developed to overcome these problems by considering the intrinsic variance of microarray assays. Some methods consider the variance across all microarray probes ('information-lending' methods)[48,49] or take advantage of a permutation procedure to find a better balance between the false positive and false negative findings within each assay[47,50,51]. If reference data are available (e.g., healthy individuals typed on the same platform), it is possible to consider the estimated FSI distribution for each microarray probe and to separately calculate a specific threshold for each probe[33]. These methods are one step forward compared to the naïve method and allow for the estimation of the proportion of false positive and negative CNV regions to decrease their proportion in the called CNVs.

Additionally, several more sophisticated methods for CNV calling have been developed to date[41]. Hidden Markov model (HMM) and Bayesian methods are frequently used for CNV calling from high-density microarray data. Most of these methods use LRR values for CNV detection. A few algorithms use LRR and BAF values together and some also include user-specified or automatically

estimated experiment-related parameters. Hidden states in HMM represent the actual underlying copy number that is estimated by considering FSI and often also reflect other relevant measures. For example QuantiSNP – one of the first tools developed for CNV calling from SNP genotyping microarray data – uses an Objective-Bayes-HMM approach and simultaneously incorporates the LRR and BAF as well as a (fixed) value for the expected frequency of the B allele for all microarray probes in a HMM framework[52]. Another popular algorithm (PennCNV) uses the total diploid copy number specific and probe distance-dependent transition probabilities in its HMM state-transition matrix, which is a more realistic model for the state transition between different diploid copy numbers in genomic DNA. PennCNV incorporates additional sources of relevant data to achieve finer breakpoint mapping and improve a posteriori CNV validation, including optionally population-specific values for the expected frequency of the B allele for each SNP and the distance between adjacent probes[42].

To diminish the rate of false positive findings, additional filtering steps are often applied to raw CNV calls. For example, very short predicted CNVs that are especially enriched in false positives or CNVs with low confidence metrics are often excluded[33,34]. Additionally, considering only CNVs detected by more than one algorithm can effectively minimise the proportion of false positive calls[53–55].

3. CNVs in health and disease

The widespread distribution of CNVs in the human genome suggests that they play an important role in different human phenotypic traits, including disease susceptibility[8,56]. As many as 13% of the RefSeq genes cumulatively overlap with CNVs[9], and CNVs have been demonstrated to have a direct impact on the gene expression of the encompassed or nearby genes[57,58]. CNVs can have functional and phenotypic consequences by changing gene expression patterns or altering the coding sequences of these genes.

3.1 Genomic impact of CNVs

Both deletion and duplication variants can act as loss-of-function (LoF) variants by simply disrupting coding sequences[9]. Heterozygous deletion variants (i.e., if one haplotype has been deleted, but the other haplotype is still present) can unmask other functional variants or even recessive deleterious mutations in the remaining haplotype[59]. Shorter intragenic CNVs can also generate novel genes and transcripts by fusing coding regions from different genes or by deleting exons of existing genes[11,29].

Another obvious mechanism through which CNVs can effect on phenotypic traits is gene dosage[60]. Copy number variants encompassing dosage-sensitive genes or their regulatory regions can directly or through positional effects result in over- or under-expressed genes and consequent disparity in the corresponding phenotypes[41]. CNVs can also induce changes in DNA methylation with functional and phenotypic consequences[61].

Copy number variants can cumulatively ('genomic burden') or in combination with other variants ('two-hit' model) alter human phenotypes and cause sporadic or Mendelian and complex disease[62]. In contrast, it has been shown that there are approximately 100 genes in humans that can be completely or partially deleted without producing any apparent trait or disease consequences[2,13].

3.2 Associated trait and disease phenotypes

The following three categories of genes are often enriched for copy number variants in healthy individuals: a) genes involved in extracellular biological processes and immunity (e.g., cell adhesion, cell recognition, signalling and immune response); b) genes involved in environmental response functions (e.g., sensory perception) and c) retrovirus- and transposition-related protein coding genes[9,18,62]. Additional known examples of normal phenotypical variability include CNVs associated with salivary amylase production[5,7,63], red-green colour vision[64–66], testosterone excretion level[67] and rhesus blood group sensitivity[68].

CNVs have also been associated with several Mendelian or common disease phenotypes and infectious diseases[69–74]. Furthermore, it has been shown that newly arisen *de novo* copy number variants (germline mutations present in the offspring but not in the parents) can cause a variety of different diseases[19,75]. Examples of known CNV-driven disease associations include HIV infection susceptibility[76,77], haemophilia A[78], Parkinson's disease[79,80], Alzheimer's disease[81–83], obesity[84–86], schizophrenia[87–90], susceptibility to autism spectrum disorders[91–95], systemic lupus erythaematosus[96], psoriasis[97,98] and other immune system-related diseases[99,100]. As with dominant *de novo* and recessive disease-causing SNP variants, *de novo* copy number alterations alone and in combination with inherited CNVs can be responsible for more severe clinical phenotypes[101] or cause sporadic disease[75,102].

The penetrance is nearly complete for some copy number variants[84,89,95,103–106]; however, this is not the case for most CNV loci (especially for duplications)[89,107–110]. Possible reasons include epigenetic modifications, other genetic variants in the vicinity, modifier genes, regulatory elements (reviewed by Cooper *et al.*)[111] and potentially alternative allelic copies present within the CNV regions.

4. Population genetics of CNVs

Deleterious CNVs that predispose individuals to severe phenotypes are to some extent most likely to occur under purifying selective pressure[18,112]. This reasoning is supported by the observation that CNVs overlap with genes and other functional elements (e.g., enhancer regions and ultra-conserved elements such as exons in genes involved in RNA processing and introns or nearby genes involved in the regulation of transcription and development) less frequently in the human genome than could be expected to occur by chance[7,113,114]. In their high resolution study of HapMap individuals, Conrad *et al.* reported that the strongest negative selection was observed for exonic CNVs, followed by intronic and then intergenic CNVs[9].

Considering the frequency distribution of copy number variants in outbred populations, CNVs have clear tendency towards a lower frequency of variant-carrying haplotypes[20]. Approximately 50% of all observed CNVs are singletons or are observed at very low frequency[9,16,115]. Indeed, CNV frequency has been shown to be strongly negatively correlated with CNV gene density and size (i.e., CNVs larger than 100 kb are rare, whereas common CNVs tend to be relatively small (<10 kb))[10,16]. Similarly, common CNVs usually harbour fewer genes than rare CNVs which can be relatively gene-rich. On the population level approximately 65–80% of individuals carry at least one CNV that is at least 100 kb in size and variants larger than 500 kb are present in 5–10% of individuals. In contrast, very large copy number variants (>1 Mb) are relatively rare and observed in only 1–2% of individuals on the population level[16,62].

Most CNV loci in the human genome can be explained by normal and simple deletion- and duplication-carrying haplotypes. As expected, deletion-carrying haplotypes have been observed 1.3–2.2 -fold more frequently than duplication-carrying haplotypes in both microarray- and sequencing-based CNV studies[11,20]. An analysis of 849 whole genomes sequenced by the 1000 Genomes Project showed that there were simple deletion-carrying haplotypes in 55% of all detected CNVs; simple duplication-carrying haplotypes were found in 29% of CNVs and multi-allelic CNVs with three or more different segregating haplotypes (i.e., normal, deletion-carrying, and duplication-carrying haplotypes) were found in 16% of CNVs, thereby demonstrating that a considerable proportion of such multi-allelic loci exist in the human genome[116]. It should be noted that the term ‘multi-allelic’ in the context of CNVs can be confusing because it is commonly used to note the fact that different types of copy number variants (i.e., deletion, duplication, triplication, etc.) segregate in the same region at the population level.

Compared to frequency-matched SNP variants, ‘bi-allelic’ CNVs with normal and deletion- or normal and duplication-carrying haplotypes exhibit greater population stratification[39]. Additionally, bi-allelic copy number polymorphisms (CNPs – CNVs with a frequency ≥1%) show a strong correlation with flanking bi-allelic SNP genotypes. Furthermore, deletion variants can be

tagged more efficiently than duplications. In a study by Handsaker *et al.*, duplication and multi-copy CNPs with a MAF >10% showed an almost uniform distribution of imputation correlation (r^2) values between 0–1, whereas in another study, only 40% of duplication CNVs showed a high correlation (r^2>0.8) with neighbouring SNPs[39,116]. This discrepancy could be due to allelic heterogeneity in proximity to the duplicated CNV or the fact that the transposed duplications are actually located farther away from the SNPs being tested for LD[9,39,116].

5. Haplotype and allelic variability in CNV regions

It has been somewhat disregarded that in addition to variable copy numbers, the allelic composition (i.e., the actual DNA sequence) can differ slightly from copy to copy (Figure 4). In CNV studies published to date, only dichotomous conditions (i.e., gain or loss[7], estimated total number of copies on two homologous chromosomes (0, 1, 2, 3, 4, etc.)[73,74,108,117] or the continuous normalised microarray intensity data[57,115,118]) have been considered.

A few studies have demonstrated that in addition to CNV detection it is also possible to infer and output the total allelic composition for each SNP marker (called the 'CNV genotype' or 'CNV-based SNP genotype call') within individuals' CNV regions using SNP genotyping arrays[52,119,120]. Similar to two-letter genotypes (e.g., 'AA' or 'AB' or 'BB'), these null, mono-, tri- and tetraploid genotypes (e.g., '–', 'A', 'AAB' and 'AABB', respectively) are inferred from the B allele frequency (BAF) data for each SNP variant and represent the total allelic composition from both homologous chromosomes in the studied individual[52,121,122]. When these SNP-based CNV genotypes (i.e., the exact allelic composition within the CNV region) are considered, it is possible to infer the haplotypic phase of the allelic copies within each copy number variable region – a problem that was well defined for trisomic chromosomes by Clark and colleagues[123].

The correct phase information has been shown to be important for SNP variants[124–126]. Therefore, it is important to know, how exactly the allelic copies of a DNA sequence are distributed on different haplotypes on two sets of chromosomes within a CNV region[8,127]. The ability to phase and differentiate between normal haplotypes and CNV-carrying haplotypes with different allelic composition and to determine their parental origin would enable new, possibly more powerful association analyses with human disease traits[128,129].

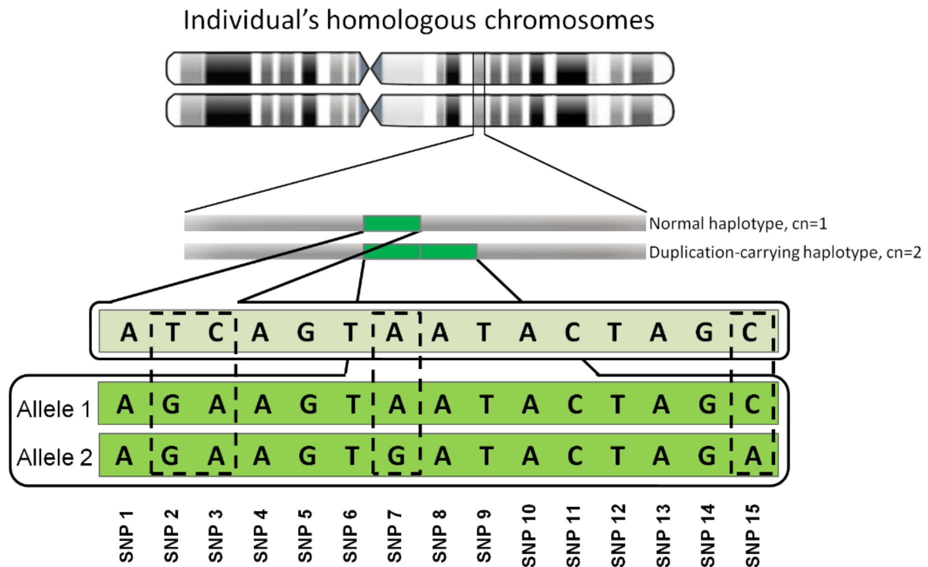

Figure 4. Haplotypes with slightly different allelic copies in a duplication CNV region. Haplotype-informative SNP genotypes in a duplicated region that are polymorphic between normal or duplication-carrying haplotypes and that can be used for phasing of given normal and CNV-carrying haplotypes are indicated with dashed rectangles.

AIMS OF THE PRESENT STUDY

In the era of high-throughput genetics and genomics, bioinformatics and statistical tools are required for the critical interpretation of thousands or even millions of parallel experiments and to help draw the correct conclusion(s) from these experiments. The present thesis aimed to find and develop statistical and computational tools and methods for DNA copy number variant calling and analysis from DNA microarray data.

The main research aims of the thesis were as follows:

1. To develop a statistical framework for CNV detection from human chromosome X-specific microarrays (Ref. I)

2. To develop a computational pipeline for CNV calling from whole-genome SNP genotyping microarray data (Ref. II–IV)

3. To develop a computational method for phasing of normal and variant-carrying haplotypes within CNV regions in families (Ref. II and IV)

4. To study and describe the transmission and allelic variability of CNV haplotypes (Ref. II and IV)

RESULTS AND DISCUSSION

1. Statistical method for detecting copy number variants from a chromosome X-specific microarray (Ref. I)

Prior to any analysis of interesting copy number variants (i.e., association analysis, experimental CNV validation, breakpoint refinement, and functional consequences), the variants have to be detected from the raw data. For this, special statistical tools and methods are required.

For the first practical application, we developed a computational workflow to detect copy number variants from a chromosome X-specific microarray containing 558 probes spanning almost the entire chromosome (with median spacing of 238 kb) and 111 control probes representing all autosomal chromosomes and chromosome Y. This single-channel array platform was developed in-house to screen patients with idiopathic mental retardation (MR). For the first step in CNV calling, microarray signals were transformed to the logarithmic scale and normalised between the arrays with respect to the median of: a) all fluorescent signal intensity (FSI) values or b) autosomal control probe-specific FSIs from microarrays included in the analysis. This was necessary because the absolute FSI values from different microarrays often varied several fold, and between-array normalisation enabled comparisons between FSI values from different microarrays. Next, the average and 90% tolerance interval values (TI90%) were calculated for each microarray probe using the data from the control panel containing FSIs from 5–15 cytogenetically controlled and phenotypically normal individuals. This step was performed separately for the male and female control panels.

To discover possible CNVs, FSI values for the studied individuals were compared to the calculated 90% upper and lower tolerance interval values. Regions with two or more adjacent probes deviating in the same direction were marked as putative copy number variants and visually inspected from the FSI plots (Figure 5c and d). Theoretically, such criteria would have resulted in a false positive proportion of 0.5%. Empirically, we calculated the false positive rate of 1% on average and 3% in the worst case scenario when comparing normal females against the panel of normal females (and male to males).

In practice, if no more than 10–20 probes deviated from the expected FSI values in a particular array-MAPH experiment, single probe alterations that might otherwise be interpreted as false positives were often further investigated, allowing smaller deletions or duplications to be detected or rejected.

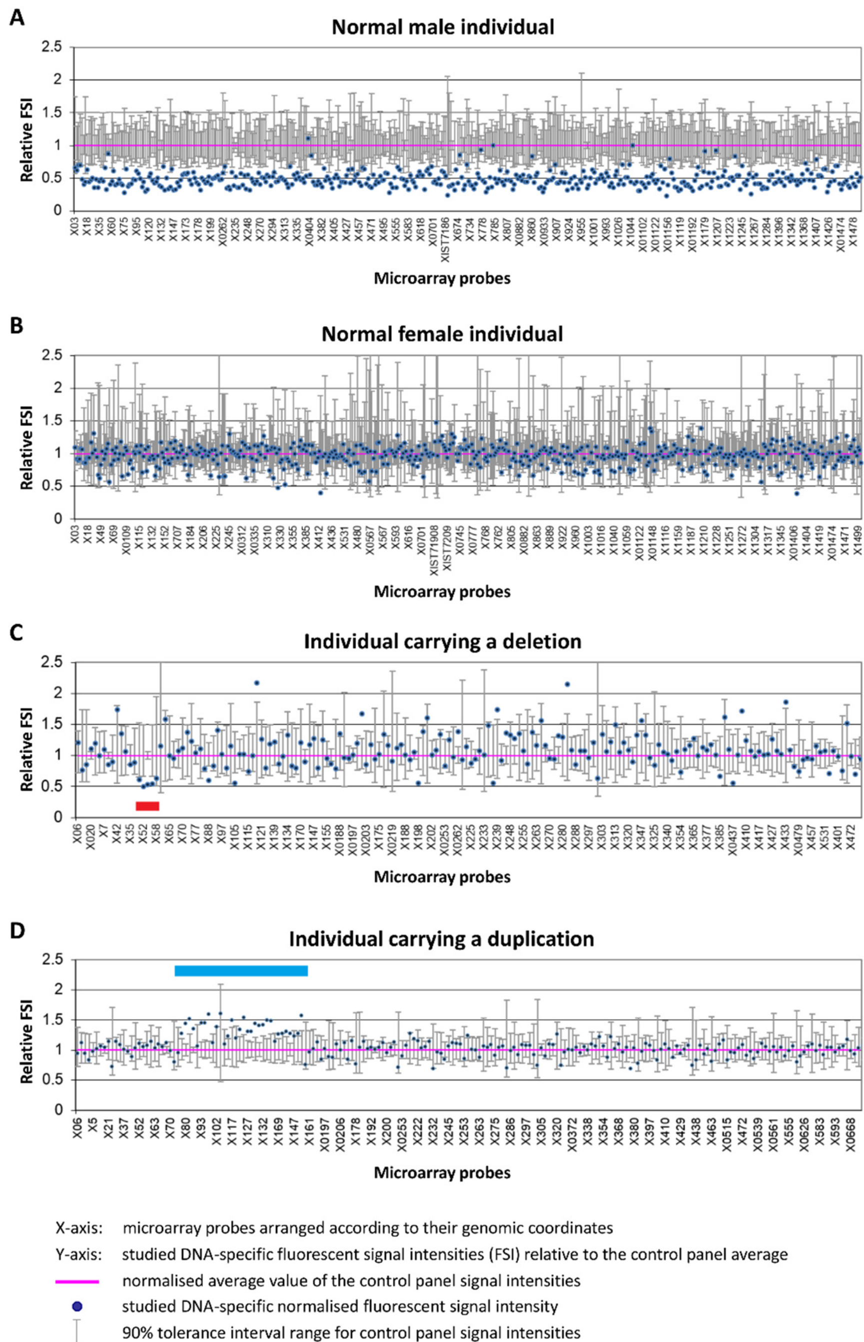

Figure 5. Different samples analysed on a chromosome X-specific single-channel microarray. Normal male individual (A) and normal female individual (B) compared to a panel of healthy female controls illustrate how one copy of chromosome X (male, XY) compares to two copies of the chromosome (female, XX). The examples show individuals carrying a deletion (C) and duplication (D) CNV detected by the array-MAPH method.

Although the chromosome X-specific array platform and related methods functioned quite robustly, the production and usage of this array was discontinued. New emerging computational methods developed for microarray-based genome-wide SNP genotyping platforms provided much higher resolution and more sensitive detection of CNVs.

2. Computational pipeline for CNV calling from SNP genotyping microarray data (Ref. II–IV)

Maintaining a good balance between false positive and false negative findings is nearly always a challenging task. Different scientific questions and analyses might prioritise one or the other, often with a resulting trade-off between specificity and sensitivity.

Because SNP genotyping arrays were first developed and optimised for SNP genotyping rather than genome-wide CNV detection and calling, CNVs detected from the SNP genotyping arrays include a considerable amount of false positive findings. To control the proportion of such false positive CNVs and to avoid possible false associations and/or interpretations in any downstream analyses, we developed a CNV calling pipeline for this application that combined two different CNV calling algorithms and a few additional QC steps in addition to standard quality control and filtering steps. As CNV calls from only one algorithm can include a large number of false positive calls, considering CNVs called by more than one independent algorithm will help decrease the number of false positive CNVs.

For each studied individual, CNV calling was performed in parallel with the QuantiSNP and PennCNV programs. These initial raw CNV calls were further merged as intersections; only CNVs that were called by both algorithms for the same individual in the same genomic loci were considered for most analyses. To achieve unambiguous results in downstream analyses, only CNVs that were similarly called (same type of overlapping copy number change – gain or loss) were considered in Ref. IV. We used population-specific B allele frequency reference files (PFB-file) for different datasets in PennCNV, because this approach seemed to provide higher sensitivity in CNV calling and more accurate CNV start and end coordinates. Separate raw calls in relevant CNV regions from both algorithms were studied either manually (Ref. II and Ref. III) or programmatically (Ref. IV) to avoid false negative findings in individuals without 'called-by-both-programs' CNVs (either in studied cases and healthy controls or family members) resulting from our conservative calling methodology. Additionally, we used breakpoints inferred by one single CNV calling program rather than the QuantiSNP-PennCNV intersection coordinates for experimental validations and other analyses that required CNV breakpoint coordinates because our CNV merging procedure systematically underestimated CNV size and therefore could possibly have provided incorrect starting and

ending positions. We used QuantiSNP coordinates (as described in Ref. II and Ref. III) because, in our hands, this algorithm gave more accurate breakpoint estimates than PennCNV. Discrete CNV regions (CNVRs) were defined by merging overlapping (≥1 bp) CNVs across all individuals in a study group as described previously[7,130].

For additional filtering steps, we filtered out very short putative CNVs and CNVs that contained only a few microarray probes. CNVs that had low log Bayes Factor (LBF) values (representing QuantiSNP confidence estimates for CNV calls) were also filtered out. If raw data were available, putative CNV regions were confirmed or omitted by visually inspecting the LRR and BAF plots for the CNV regions (as described in Ref. II and Ref. IV). We also filtered out CNV regions where any member of the corresponding family had any unclear raw CNV calls ('called-by-one-program') when working with family data (Ref. IV).

3. Computational method for phasing normal and variant-carrying haplotypes within CNV regions (Ref. II and IV)

Considering only CNVs called by two algorithms (and excluding CNV regions with inconclusive raw calls) provided good quality data that we further used to study how exactly new CNVs emerged and how existing CNVs were transmitted from parents to offspring. To accomplish this objective, we developed an algorithm that could computationally differentiate and phase normal and variant-carrying haplotypes within CNV regions in nuclear families based on genotyping microarray data.

Our algorithm works by examining adjacent regular two-letter and null, mono-, tri- or tetraploid CNV genotypes (described earlier) in each family member in a CNV region present in any member of the corresponding family (Figure 6). Our algorithm uses copy number variant calls, user-defined family structure and informative genotypes that are polymorphic between the parents to phase the allelic composition within each copy number variable region in the studied families by deterministically testing all possible molecular haplotypes and their transmission according to the Mendelian inheritance scenarios in the studied CNV locus in a given family. If there is more than one child in the studied family, all children will be considered simultaneously in this step. Finally, our algorithm will select these normal and/or CNV-carrying haplotypes and transmission scenarios that can unambiguously explain the allelic content for every member of the corresponding family for each CNV locus. If it is not possible to explain the allelic composition in the offspring by Mendelian transmission scenarios of parental haplotypes, non-Mendelian scenarios –

A Data collection and removal of uninformative and low-confidence SNP genotype calls

	QuantiSNP CNV genotypes			
Marker ID	**Father**	**Mother**	**Child 1**	**Child 2**
rs589559	AA	BB	B -	AB
rs2867167	AB	BB	B -	BB
rs666536	AB	AA	A -	AB
rs694861	BB	AB	A -	BB
rs12959303	AB	BB	B -	BB
rs608406	AB	AA	A -	AB
rs2124297	AA	AB	B -	AA
rs8083190	AB	AA	A -	AA
rs650464	BB	AB	B -	AB

B Haplotype phasing

	Phased haplotypes							
Marker ID	**Father**		**Mother**		**Child 1**		**Child 2**	
rs589559	A	A	B	B	B	-	B	A
rs2867167	A	B	B	B	B	-	B	B
rs666536	A	B	A	A	A	-	A	B
rs694861	B	B	A	B	A	-	B	B
rs12959303	A	B	B	B	B	-	B	B
rs608406	A	B	A	A	A	-	A	B
rs2124297	A	A	B	A	B	-	A	A
rs8083190	B	A	A	A	A	-	A	A
rs650464	B	B	B	A	B	-	A	B

C Illustrated haplotype segregation

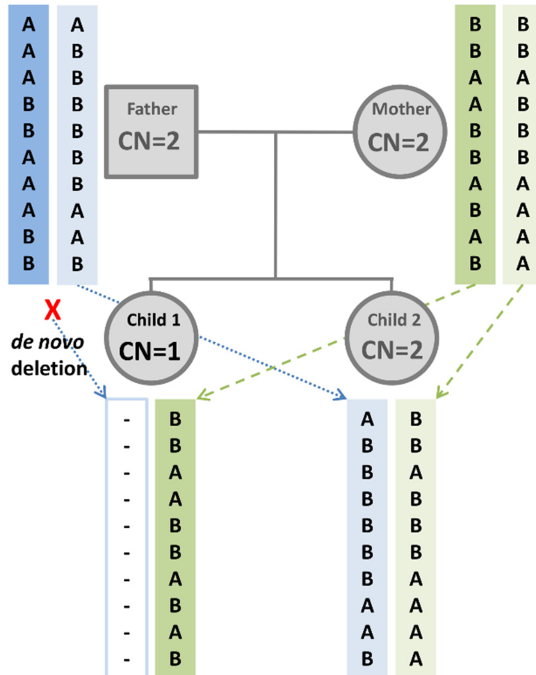

Figure 6. Computational phasing of normal and CNV-carrying haplotypes. A putative *de novo* CNV region where Child 1 has inherited the normal haplotype from the mother and a deletion-carrying haplotype from the father. The other child (Child 2) has inherited normal haplotypes from both parents. In Child 1, both normal and deletion-carrying haplotypes are already inherently 'phased' (A). By considering the haplotypes and genotypes observed in the mother, Child 2 and the father, finding the exact haplotypes transmitted from both parents is intuitively and computationally very straightforward (B and C).

de novo deletion or duplication events and uniparental isodisomy/heterodisomy – are considered. In the presence of informative genotypes between parental haplotypes, our algorithm can also determine on which parental chromosome the mutation occurred (e.g., inter- or intra-chromosomal duplication). If there are no haplotype-informative markers within a CNV region and the simplest haplotypes and Mendelian inheritance scenarios do not explain the allelic composition in all family members, it is not possible to determine a single set of correct haplotypes and inheritance scenarios and several equally possible haplotypes and Mendelian and/or non-Mendelian transmissions are suggested as the result.

4. Transmission and allelic variability of CNV-carrying haplotypes (Ref. II and IV)

It is often important to identify which parent transmitted a certain variant (e.g., a putatively disease-causing CNV). Additionally, it is useful to understand how the variant of interest segregates within the corresponding family and whether the other affected or unaffected siblings have inherited the same variant and alleles (as described in Ref. II). In the case of *de novo* variants that are not directly inherited but have often emerged in parental germline cells, it might be important to elucidate whether the mutation occurred on the maternal or paternal chromosomes.

To solve these questions, we used our phasing algorithm to resolve the normal and CNV-carrying haplotypes in trios or larger nuclear families based on genotype and copy number estimates from SNP microarray data. By studying 34 Estonian families (22 mother-father-child trios and 12 families with multiple siblings) and 30 HapMap Yoruban (YRI) trios, we discovered that it was computationally possible to unambiguously determine and phase the underlying normal and CNV-carrying parental haplotypes in vast majority (94%) of CNV regions and follow their transmission in offspring (Ref. IV). We observed that haplotype phasing was more efficient in CNV regions where only one CNV-carrying haplotype was present in one of the parents, resulting in a single or very few equally possible Mendelian and/or non-Mendelian transmission scenarios. In comparison, when *de novo* CNVs or multiple CNV-carrying haplotypes were present in the parents (especially duplication-carrying haplotypes), two or more equally possible haplotype combinations/transmission scenarios were often plausible (Ref. IV).

4.1 Transmission of normal and CNV-carrying haplotypes within CNV regions in nuclear families (Ref. IV)

We further analysed CNV regions that could be unambiguously phased in Estonian and HapMap YRI families and counted the number of transmission

events of normal or CNV-carrying haplotypes from the parents carrying a CNV. CNV-carrying haplotypes were observed in the offspring slightly less frequently than expected by random Mendelian inheritance. To further investigate this phenomenon, we analysed the transmission of deletion and duplication-carrying haplotypes in the combined dataset while also considering CNV length. Although this analysis revealed small deviations from the expected Mendelian transmission rate of 50% for both deletions and duplication in nearly all CNV length intervals, the previously observed differences were mainly driven by the under-transmission of short (<10 kb) deletion-carrying haplotypes in the HapMap dataset.

Although these CNVs in the HapMap dataset were validated by other independent means, it is possible that these short under-transmitted CNVs contained proportionally more false positive CNV calls (either cell-line artefacts of actually incorrect CNV calls) that could not be transmitted and therefore appeared to be under-transmitted. However, a similar tendency was observed for longer (>10 kb) deletion-carrying haplotypes in both studied cohorts. Such bias might be expected in the case of some high penetrance CNVs associated with severe disease phenotypes. Moreover, it has been suggested that this effect could be more pronounced for larger deletion variants interrupting genes of vital importance; consequently this effect is more likely under stronger (prenatal) selection. The opposite effect that was observed for a slightly increased transmission rate for longer (>100 kb) duplications (although this did not reach statistical significance) could possibly be explained by the contribution of duplications in providing the means for functional redundancy and facilitating exon shuffling, gene fusion and gene duplication. By generating new functional genes, duplication events may be an important mechanism for long-term evolutionary changes in humans and thus occur under positive selection. Although similar results were also observed in several earlier studies, only a very cautious interpretation of both findings should be considered. Larger studies with high-resolution platforms and validation techniques are merited to confirm and further investigate these interesting phenomena.

4.2 Putative *de novo* copy number variants (Ref. IV)

The process of directly deleterious variants being under negative selection and thus systematically under-transmitted from generation to generation is balanced by germline mutation events that generate *de novo* variants that are observable in new generations.

In our combined dataset we detected 27 CNVs that were identified as putative *de novo* copy number alterations in the offspring. Out of these 27 putative *de novo* CNVs, 20 were deletions and 7 were duplication events. As determined by our algorithm and confirmed by manual inspection, five out of nine unambiguously phased putative *de novo* CNVs appeared on maternally inherited chromosomes and four on a paternally inherited chromosome. We use

the term 'putative' *de novo* CNVs because such variants might not be true *de novo* mutations even if these CNVs are unambiguously phased and validated by other experimental methods. Alternatively, these can be somatically derived cloned mutations or artefacts often observed when DNA from cell-lines is analysed. *De novo* CNVs might also appear due to complex haplotype compositions of a studied family in a given locus (e.g., in CNP loci where haplotypes with 0 and 2 copies are combined in one parent, leading to incorrect calling of *de novo* variants from unphased CNV data, thereby highlighting the relevance of phasing of the exact parental haplotypes).

4.2 Allelic variability in duplication CNV regions (Ref. IV)

Deterministic phasing within CNV regions allowed us to differentiate between normal and CNV-carrying haplotypes with different allelic compositions and to more extensively study the allelic variability within the discovered copy number gain-carrying haplotypes. Because alternative allelic copies within CNV regions may modify the severity of a phenotype (including disease susceptibility), we aimed to determine the occurrence of alternative allelic copies within these multi-copy haplotypes.

We studied these CNV regions in a combined dataset where one or both parents of the same family had duplication or triplication-carrying haplotypes in the same regions. We determined that informative polymorphic genotypes were present within the multi-copy haplotypes in 93% of such CNV regions. Therefore, it was possible to phase and differentiate between normal and multi-copy haplotypes (as shown in Figure 3). Furthermore, heterozygous genotypes were also present within the multi-copy haplotypes in 67% of these CNV regions (Figure 7), thereby allowing us to define alternative allelic copies within these copy number gain-carrying haplotypes and demonstrating extensive and to-date unmeasured allelic variability in multi-copy CNV regions in the human genome.

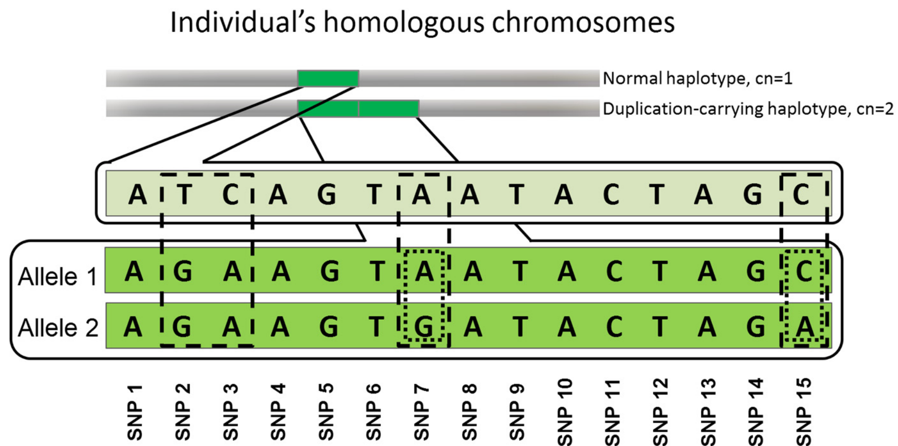

Figure 7. Allelic variability in CNV regions and within duplication CNVs. SNP-based genotypes within duplicated region that are polymorphic between normal (haploid copy number cn=1) and duplication-carrying haplotypes (cn=2) are indicated with dashed rectangles. The duplication-carrying haplotype is composed of two allelic copies distinguished by single nucleotide variants (indicated as SNP7 and SNP15) that are polymorphic within the duplication-carrying haplotype (depicted by a dotted rectangle).

SUMMARY AND CONCLUSIONS

Using statistical and computational methods, we studied copy number variants (CNVs) and different CNV-related aspects of the human genome. The most important results of this thesis are as follows:

1. We developed computational methods and software for phasing haplotypes within CNV regions in nuclear families based on SNP genotyping microarray data. Using two different datasets, we determined that it was possible to unambiguously phase all parental haplotypes and follow their transmission in offspring in the vast majority of CNV regions.

2. By counting the number of transmission events of normal or CNV-carrying haplotypes in two family datasets, we found that CNV-carrying haplotypes were observed in offspring less frequently than expected by Mendelian inheritance. This results suggests the systematic under-transmission of CNVs. This trend was stronger for deletion-carrying haplotypes.

3. We determined that the copies represented different alleles in two-thirds of the studied duplication-carrying CNV regions. This result demonstrates the extensive allelic variability in multi-copy CNV regions of the human genome.

REFERENCES

1. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 42, D1001–6 (2014).
2. MacArthur, D.G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335, 823–8 (2012).
3. Ophir, R. & Graur, D. Patterns and rates of indel evolution in processed pseudogenes from humans and murids. *Gene* 205, 191–202 (1997).
4. Kondrashov, A.S. & Rogozin, I.B. Context of deletions and insertions in human coding sequences. *Hum Mutat* 23, 177–85 (2004).
5. Iafrate, A.J. *et al.* Detection of large-scale variation in the human genome. *Nat Genet* 36, 949–51 (2004).
6. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* 305, 525–8 (2004).
7. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* 444, 444–54 (2006).
8. Feuk, L., Carson, A.R. & Scherer, S.W. Structural variation in the human genome. *Nat Rev Genet* 7, 85–97 (2006).
9. Conrad, D.F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704–12 (2010).
10. McCarroll, S.A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet* 40, 1166–74 (2008).
11. Mills, R.E. *et al.* Mapping copy number variation by population-scale genome sequencing. *Nature* 470, 59–65 (2011).
12. MacDonald, J.R., Ziman, R., Yuen, R.K., Feuk, L. & Scherer, S.W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* 42, D986–92 (2014).
13. Zarrei, M., MacDonald, J.R., Merico, D. & Scherer, S.W. A copy number variation map of the human genome. *Nat Rev Genet* 16, 172–83 (2015).
14. Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nat Genet* 37, 727–32 (2005).
15. Korbel, J.O. *et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science* 318, 420–6 (2007).
16. Itsara, A. *et al.* Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* 84, 148–61 (2009).
17. Pinto, D., Marshall, C., Feuk, L. & Scherer, S.W. Copy-number variation in control population cohorts. *Hum Mol Genet* 16 Spec No. 2, R168–73 (2007).
18. Cooper, G.M., Nickerson, D.A. & Eichler, E.E. Mutational and selective effects on copy-number variants in the human genome. *Nat Genet* 39, S22–9 (2007).
19. Zhang, F., Gu, W., Hurles, M.E. & Lupski, J.R. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet* 10, 451–81 (2009).
20. Kato, M. *et al.* Population-genetic nature of copy number variations in the human genome. *Hum Mol Genet* 19, 761–73 (2010).
21. Ankala, A. *et al.* Aberrant firing of replication origins potentially explains intragenic nonrecurrent rearrangements within genes, including the human DMD gene. *Genome Res* 22, 25–34 (2012).

22. Chen, J.M., Cooper, D.N., Ferec, C., Kehrer-Sawatzki, H. & Patrinos, G.P. Genomic rearrangements in inherited disease and cancer. *Semin Cancer Biol* 20, 222–33 (2010).
23. Chen, J.-M. Genomic Rearrangements: Mutational Mechanisms. (2011).
24. Turner, D.J. *et al.* Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nat Genet* 40, 90–5 (2008).
25. Liu, P., Carvalho, C.M., Hastings, P.J. & Lupski, J.R. Mechanisms for recurrent and complex human genomic rearrangements. *Curr Opin Genet Dev* 22, 211–20 (2012).
26. Lieber, M.R. The mechanism of human nonhomologous DNA end joining. *J Biol Chem* 283, 1–5 (2008).
27. Lee, J.A., Carvalho, C.M. & Lupski, J.R. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 131, 1235–47 (2007).
28. Hastings, P.J., Ira, G. & Lupski, J.R. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet* 5, e1000327 (2009).
29. Zhang, F. *et al.* The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet* 41, 849–53 (2009).
30. Chen, J.M., Chuzhanova, N., Stenson, P.D., Ferec, C. & Cooper, D.N. Complex gene rearrangements caused by serial replication slippage. *Hum Mutat* 26, 125–34 (2005).
31. de Ligt, J. *et al.* Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum Mutat* 34, 1439–48 (2013).
32. de Ligt, J. *et al.* Platform comparison of detecting copy number variants with microarrays and whole-exome sequencing. *Genomics Data* 2, 144–146 (2014).
33. Bignell, G.R. *et al.* High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res* 14, 287–95 (2004).
34. Peiffer, D.A. *et al.* High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res* 16, 1136–48 (2006).
35. Kloth, J.N. *et al.* Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex genetic alterations in cervical cancer. *BMC Genomics* 8, 53 (2007).
36. Lockwood, W.W., Chari, R., Chi, B. & Lam, W.L. Recent advances in array comparative genomic hybridization technologies and their applications in human genetics. *Eur J Hum Genet* 14, 139–48 (2006).
37. Price, T.S. *et al.* SW-ARRAY: a dynamic programming solution for the identification of copy-number changes in genomic DNA using array comparative genome hybridization data. *Nucleic Acids Res* 33, 3455–64 (2005).
38. Yu, W. *et al.* Development of a comparative genomic hybridization microarray and demonstration of its utility with 25 well-characterized 1p36 deletions. *Hum Mol Genet* 12, 2145–52 (2003).
39. Campbell, C.D. *et al.* Population-genetic properties of differentiated human copy-number polymorphisms. *Am J Hum Genet* 88, 317–32 (2011).
40. Wang, K. & Bucan, M. Copy Number Variation Detection via High-Density SNP Genotyping. *CSH Protoc* 2008, pdb top46 (2008).
41. Li, W. & Olivier, M. Current analysis platforms and methods for detecting copy number variation. *Physiol Genomics* 45, 1–16 (2013).

42. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 17, 1665–74 (2007).
43. Menten, B. *et al.* arrayCGHbase: an analysis platform for comparative genomic hybridization microarrays. *BMC Bioinformatics* 6, 124 (2005).
44. Lingjaerde, O.C., Baumbusch, L.O., Liestol, K., Glad, I.K. & Borresen-Dale, A.L. CGH-Explorer: a program for analysis of array-CGH data. *Bioinformatics* 21, 821–2 (2005).
45. Veltman, J.A. *et al.* High-throughput analysis of subtelomeric chromosome rearrangements by use of array-based comparative genomic hybridization. *Am J Hum Genet* 70, 1269–76 (2002).
46. Kim, S.Y. *et al.* ArrayCyGHt: a web application for analysis and visualization of array-CGH data. *Bioinformatics* 21, 2554–5 (2005).
47. Diskin, S.J. *et al.* STAC: A method for testing the significance of DNA copy number aberrations across multiple array-CGH experiments. *Genome Res* 16, 1149–58 (2006).
48. Wang, J., Meza-Zepeda, L.A., Kresse, S.H. & Myklebost, O. M-CGH: analysing microarray-based CGH experiments. *BMC Bioinformatics* 5, 74 (2004).
49. Picard, F., Robin, S., Lavielle, M., Vaisse, C. & Daudin, J.J. A statistical approach for array CGH data analysis. *BMC Bioinformatics* 6, 27 (2005).
50. Myers, C.L., Chen, X. & Troyanskaya, O.G. Visualization-based discovery and analysis of genomic aberrations in microarray data. *BMC Bioinformatics* 6, 146 (2005).
51. Yang, H. & Churchill, G. Estimating p-values in small microarray experiments. *Bioinformatics* 23, 38–43 (2007).
52. Colella, S. *et al.* QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* 35, 2013–25 (2007).
53. Winchester, L., Yau, C. & Ragoussis, J. Comparing CNV detection methods for SNP arrays. *Brief Funct Genomic Proteomic* 8, 353–66 (2009).
54. Pinto, D. *et al.* Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol* 29, 512–20 (2011).
55. Kim, S.Y., Kim, J.H. & Chung, Y.J. Effect of Combining Multiple CNV Defining Algorithms on the Reliability of CNV Calls from SNP Genotyping Data. *Genomics Inform* 10, 194–9 (2012).
56. McCarroll, S.A. & Altshuler, D.M. Copy-number variation and association studies of human disease. *Nat Genet* 39, S37–42 (2007).
57. Stranger, B.E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315, 848–53 (2007).
58. Henrichsen, C.N. *et al.* Segmental copy number variation shapes tissue transcriptomes. *Nat Genet* 41, 424–9 (2009).
59. Kurotaki, N., Stankiewicz, P., Wakui, K., Niikawa, N. & Lupski, J.R. Sotos syndrome common deletion is mediated by directly oriented subunits within inverted Sos-REP low-copy repeats. *Hum Mol Genet* 14, 535–42 (2005).
60. Lupski, J.R. *et al.* Gene dosage is a mechanism for Charcot-Marie-Tooth disease type 1A. *Nat Genet* 1, 29–33 (1992).

61. Brahmachary, M. *et al.* Digital genotyping of macrosatellites and multicopy genes reveals novel biological functions associated with copy number variation of large tandem repeats. *PLoS Genet* 10, e1004418 (2014).
62. Girirajan, S., Campbell, C.D. & Eichler, E.E. Human copy number variation and complex genetic disease. *Annu Rev Genet* 45, 203–26 (2011).
63. Perry, G.H. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nat Genet* 39, 1256–60 (2007).
64. Nathans, J., Thomas, D. & Hogness, D.S. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* 232, 193–202 (1986).
65. Deeb, S.S. The molecular basis of variation in human color vision. *Clin Genet* 67, 369–77 (2005).
66. Deeb, S.S. Genetics of variation in human color vision and the retinal cone mosaic. *Curr Opin Genet Dev* 16, 301–7 (2006).
67. Jakobsson, J. *et al.* Large differences in testosterone excretion in Korean and Swedish men are strongly associated with a UDP-glucuronosyl transferase 2B17 polymorphism. *J Clin Endocrinol Metab* 91, 687–93 (2006).
68. Wagner, F.F. & Flegel, W.A. Review: the molecular basis of the Rh blood group phenotypes. *Immunohematology* 20, 23–36 (2004).
69. Singleton, A.B. *et al.* alpha-Synuclein locus triplication causes Parkinson's disease. *Science* 302, 841 (2003).
70. Estivill, X. & Armengol, L. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet* 3, 1787–99 (2007).
71. Lee, C. & Scherer, S.W. The clinical context of copy number variation in the human genome. *Expert Rev Mol Med* 12, e8 (2010).
72. Stankiewicz, P. & Lupski, J.R. Structural variation in the human genome and its role in disease. *Annu Rev Med* 61, 437–55 (2010).
73. Aitman, T.J. *et al.* Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* 439, 851–5 (2006).
74. Fellermann, K. *et al.* A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet* 79, 439–48 (2006).
75. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* (2014).
76. Gonzalez, E. *et al.* The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307, 1434–40 (2005).
77. Nakajima, T., Kaur, G., Mehra, N. & Kimura, A. HIV-1/AIDS susceptibility and copy number variation in CCL3L1, a gene encoding a natural ligand for HIV-1 co-receptor CCR5. *Cytogenet Genome Res* 123, 156–60 (2008).
78. Lakich, D., Kazazian, H.H., Jr., Antonarakis, S.E. & Gitschier, J. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nat Genet* 5, 236–41 (1993).
79. Chartier-Harlin, M.C. *et al.* Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* 364, 1167–9 (2004).
80. Ibanez, P. *et al.* Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. *Lancet* 364, 1169–71 (2004).
81. Cabrejo, L. *et al.* Phenotype associated with APP duplication in five families. *Brain* 129, 2966–76 (2006).

82. Rovelet-Lecrux, A. *et al.* APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet* 38, 24–6 (2006).
83. Theuns, J. *et al.* Promoter mutations that increase amyloid precursor-protein expression are associated with Alzheimer disease. *Am J Hum Genet* 78, 936–46 (2006).
84. Walters, R.G. *et al.* A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* 463, 671–5 (2010).
85. Willer, C.J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 41, 25–34 (2009).
86. Wheeler, E. *et al.* Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. *Nat Genet* 45, 513–7 (2013).
87. Cook, E.H., Jr. & Scherer, S.W. Copy-number variations associated with neuropsychiatric conditions. *Nature* 455, 919–23 (2008).
88. Kirov, G. *et al.* Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet* 18, 1497–503 (2009).
89. Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* 455, 232–6 (2008).
90. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* 460, 744–7 (2009).
91. Autism Genome Project, C. *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 39, 319–28 (2007).
92. Glessner, J.T. *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 459, 569–73 (2009).
93. Marshall, C.R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 82, 477–88 (2008).
94. Szatmari, P. *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 39, 319–28 (2007).
95. Weiss, L.A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 358, 667–75 (2008).
96. Yang, Y. *et al.* Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* 80, 1037–54 (2007).
97. de Cid, R. *et al.* Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat Genet* 41, 211–5 (2009).
98. Hollox, E.J. *et al.* Psoriasis is associated with increased beta-defensin genomic copy number. *Nat Genet* 40, 23–5 (2008).
99. Fanciulli, M. *et al.* FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet* 39, 721–3 (2007).
100. Willcocks, L.C. *et al.* Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *J Exp Med* 205, 1573–82 (2008).
101. Potocki, L. *et al.* DNA rearrangements on both homologues of chromosome 17 in a mildly delayed individual with a family history of autosomal dominant carpal tunnel syndrome. *Am J Hum Genet* 64, 471–8 (1999).

102. Rees, E., Moskvina, V., Owen, M.J., O'Donovan, M.C. & Kirov, G. De novo rates and selection of schizophrenia-associated copy number variants. *Biol Psychiatry* 70, 1109–14 (2011).
103. International Schizophrenia, C. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455, 237–41 (2008).
104. Cooper, G.M. *et al.* A copy number variation morbidity map of developmental delay. *Nat Genet* 43, 838–46 (2011).
105. Pinto, D. *et al.* Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* 94, 677–94 (2014).
106. Schaaf, C.P., Wiszniewska, J. & Beaudet, A.L. Copy number and SNP arrays in clinical diagnostics. *Annu Rev Genomics Hum Genet* 12, 25–51 (2011).
107. Mefford, H.C. *et al.* Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 359, 1685–99 (2008).
108. Wellcome Trust Case Control, C. *et al.* Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 464, 713–20 (2010).
109. Kirov, G. *et al.* The penetrance of copy number variations for schizophrenia and developmental delay. *Biol Psychiatry* 75, 378–85 (2014).
110. Stefansson, H. *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* 505, 361–6 (2014).
111. Cooper, D.N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet* 132, 1077–130 (2013).
112. Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E. & Pritchard, J.K. A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* 38, 75–81 (2006).
113. Derti, A., Roth, F.P., Church, G.M. & Wu, C.T. Mammalian ultraconserved elements are strongly depleted among segmental duplications and copy number variants. *Nat Genet* 38, 1216–20 (2006).
114. Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* 304, 1321–5 (2004).
115. Ionita-Laza, I. *et al.* On the analysis of copy-number variations in genome-wide association studies: a translation of the family-based association test. *Genet Epidemiol* 32, 273–84 (2008).
116. Handsaker, R.E. *et al.* Large multiallelic copy number variations in humans. *Nat Genet* 47, 296–303 (2015).
117. Barnes, C. *et al.* A robust statistical method for case-control association testing with copy number variation. *Nat Genet* 40, 1245–52 (2008).
118. Eleftherohorinou, H. *et al.* famCNV: copy number variant association for quantitative traits in families. *Bioinformatics* 27, 1873–5 (2011).
119. Wang, K. *et al.* Modeling genetic inheritance of copy number variations. *Nucleic Acids Res* 36, e138 (2008).
120. Korn, J.M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 40, 1253–60 (2008).
121. Macconaill, L.E., Aldred, M.A., Lu, X. & Laframboise, T. Toward accurate high-throughput SNP genotyping in the presence of inherited copy number variation. *BMC Genomics* 8, 211 (2007).

122. Katagiri, T. *et al.* Frequent loss of HLA alleles associated with copy number-neutral 6pLOH in acquired aplastic anemia. *Blood* 118, 6601–9 (2011).
123. Clark, A., Dermitzakis, E. & Antonarakis, S. Trisomic Phase Inference. in *Computational Methods for SNPs and Haplotype Inference*, Vol. 2983 (eds. Istrail, S., Waterman, M. & Clark, A.) 1–8 (Springer Berlin Heidelberg, 2004).
124. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* 461, 747–53 (2009).
125. Bansal, V., Tewhey, R., Topol, E.J. & Schork, N.J. The next phase in human genetics. *Nat Biotechnol* 29, 38–9 (2011).
126. Gudbjartsson, D.F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet* 47, 435–44 (2015).
127. Alkan, C., Coe, B.P. & Eichler, E.E. Genome structural variation discovery and genotyping. *Nat Rev Genet* 12, 363–76 (2011).
128. Gamazon, E.R., Cox, N.J. & Davis, L.K. Structural Architecture of SNP Effects on Complex Traits. *Am J Hum Genet* (2014).
129. Wang, W., Wang, W., Sun, W., Crowley, J.J. & Szatkiewicz, J.P. Allele-specific copy-number discovery from whole-genome and whole-exome sequencing. *Nucleic Acids Res* (2015).
130. Perry, G.H. *et al.* The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet* 82, 685–95 (2008).
131. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* 409, 860–921 (2001).
132. Pollack, J.R. *et al.* Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci U S A* 99, 12963–8 (2002).
133. Albertson, D.G. & Pinkel, D. Genomic microarrays in human genetic disease and cancer. *Hum Mol Genet* 12 Spec No 2, R145–52 (2003).

SUMMARY IN ESTONIAN

Arvutuslikud meetodid DNA koopiaarvu määramiseks

Arusaamine, et inimesed on oma geneetiliste materjali poolest erinevad, on suhteliselt vana. Ometigi sai võimalikuks ja algas selliste erinevuste ulatuslik ja süstemaatiline uurimine inimese genoomis alles suhteliselt hiljuti, uue aastatuhande alguses. Pärast inimese genoomi täieliku primaarjärjestuse esimese versiooni avaldamist[131] algas ulatuslik indiviidide ja erinevate populatsioonide vaheliste geneetiliste erinevuste (variantide) otsimine ja kaardistamine. See andis uut hoogu ka inimese haigusseoseliste geneetiliste variantide otsimisele ja kirjeldamisele. Paljude aastate jooksul keskenduti enamasti ühenukleotiidsete variantide ehk SNP-de otsimisele, sest olemasolevad tehnoloogilised vahendid ja metoodikad võimaldasid neid kõige lihtsamini suures mahus leida ja kirjeldada.

Kuigi juba aastakümneid oli tegelikult olnud teada, et erinevatel inimestel võivad olla erinevused ka pikemates DNA primaarjärjestuse lõikudes, sai DNA pikemate erinevuste, sealhulgas ka DNA koopiaarvu muutuste (ingl. *copy number variant* ehk CNV) süstemaatiline ja ulatuslik uurimine hoo sisse alles 2003. aastal. Ja kui esialgu uuriti ja seostati selliseid suuremaid muutusi DNA primaarjärjestuses tihtipeale erinevate tõsiste geneetiliste haigustega ja kasvajatega[132,133], sai peagi ühemõtteliselt selgeks, et sellised pikemad DNA ümberkorralduste variandid ei esine ainult seoses tõsiste haigustega, vaid on suhteliselt sagedased ka tervete inimeste geneetilises materjalis ja moodustavad suure osa indiviidide vahelistest erinevustest DNA tasemel[5–7].

Sellised muutused võivad olla väga erineva suurusega, alates tervete kromosoomide kordistumisest/kaotsiminekust kuni geenide ja üksikute eksonite duplitseerumiseni (koopiate kordistumine) või deleteerumiseni (koopiate kaotsiminek). Et eristada haigustpõhjustavaid CNV-sid normaalsetest ”tervete” inimeste vahelisest koopiaarvu variantidest, on tähtis aru saada sellest, kui suur on inimeste vaheline normaalne varieeruvus DNA koopiaarvu variantide tasemel. Seetõttu on oluline töötada välja uusi ja arendada edasi olemasolevaid ekperimentaalseid ja arvutuslikke meetodeid, millega detekteerida ja uurida DNA koopiaarvu muutusi inimese genoomis.

Käesoleva doktoritöö kirjanduse ülevaade keskendub DNA koopiaarvu variantide ja tekkemehhanismide kirjeldamisele. Samuti antakse ülevaade koopiaarvu variantidega senini seostatud fenotüüpidest, sealhulgas haigustest. Kirjeldatakse DNA koopiaarvu leidmiseks kasutatavate mikrokiipide tööpõhimõtteid ja olemasolevad arvutuslikke vahendeid.

Antud doktoritöö raames töötati välja erinevaid arvutuslikke meetodeid ja lahendusi DNA koopiaarvu muutuste leidmiseks ja nendega töötamiseks. Kõigepealt töötati välja statistiline raamistik ja arvutuslik metoodika CNV-de detekteerimiseks X-kromosoomi spetsiifiliselt mikrokiibilt. Kui hakkasime kasutama algselt SNP-de genotüpiseerimiseks mõeldud kogu genoomi katvaid mikrokiipe, tekkis vajadus leida ja automatiseerida meie andmetele ja kiipidele

sobivad CNV-de detekteerimise ja filtreerimise arvutuslikud lahendused tark-vara töövoogudena.

Uurides CNV-sid perekondade andmetest oli vaja täpselt määrata, milline (oletatavalt haigus-seoseline) koopiaarvu muutus milliselt vanemalt on lapsele/lastele pärandunud. Selle tarvis sai välja töötatud uudne arvutuslik metoodika, mis kasutades SNP mikrokiibi andmeid võimaldab faasida ja „järgida“ CNV-sid kandvate haplotüüpide pärandumist vanematelt lastele.

Käesoleva töö viimases osas uurisime me lisaks haigusseoseliste variantidele ka Tartu Ülikooli Eesti Geenivaramu ja rahvusvahelise HapMap projekti poolt kogutud tervetel inimestel esinevaid DNA koopiaarvu muutusi ja nende pärandumist perekondades. Selle uuringu üheks huvitavamaks tulemuseks oli deletsioonide „alapärandumine“ vanematelt lastele, st. deletsioone kandvaid haplotüüpe esines laste genoomides oluliselt vähem, kui Mendeliaalse pärandumise korral oodata oleks võinud. Teiseks huvitavaks tulemuseks oli *de novo* ehk uute DNA koopiaarvu muutuste leidmine – nimelt esines paljude laste genoomides selliseid DNA koopiaarvu variante, mida kummagi vanema genoomis ei olnud. Kolmandaks, uurides duplikatsioone perekondades faasitud CNV-de regioonides, leidsime me, et kaks kolmandikku (67%) duplikatsioonides esinevatest alleelsetest koopiatest ei olnud identsed, vaid mõnevõrra erinevad, demonstreerides seni teadmata olnud alleelse varieeruvuse määra DNA koopiaarvu korduste regioonides.

ACKNOWLEDGEMENTS

Dear Lord, do not give me light;
Let it be yours and yours alone;
But grant me a sincere desire
To strive towards it.

Desiderius Erasmus

First, I thank my supervisor Prof. Maido Remm, who throughout my years in bioinformatics has guided me with insightful thoughts and proper criticism. From Maido I learned my affection for a systematic way of thinking and working and the pursuit of accuracy. Thank you!

I am grateful to all my wonderful colleagues at University of Tartu. Thank you for your help and advice and for of all these interesting discussions and good times! My special thanks go to Lauris Kaplinski, Reidar Andreson, Reedik Mägi, Priit 'Lemps' Adler and Hedi Peterson.

I thank all of the co-authors of the papers on which this thesis is based. In my glimpse into statistics (which has proven to be extremely useful), Tõnu and Märt Möls have been my guides. Thank you all!

I am grateful to Prof. Aarno Palotie for always pushing me forward. I truly admire your energy, politeness and way of doing science!

Last but not least, I want to thank my friends and my family. Nevil, Jurgen, Lauri, Marek, Heikko, Kaarel and all others – you are the best! My dear Mirell, Mihkel and Marie – thank you for tolerating my unusual work with all of these travels and late night or weekend working hours. Thank you for giving meaning to my life and a desire to move onward and upward!

PUBLICATIONS

CURRICULUM VITAE

Name: Priit Palta
Date of birth: April 25, 1981
Address: Institute of Molecular and Cell Biology, 23 Riia St., 51010 Tartu
E-mail: priit.palta@ut.ee
Current position: Research scientist, Institute for Molecular Medicine Finland (FIMM), University of Helsinki

Education:

2007–2011 Ph.D. student in bioinformatics, Institute of Molecular and Cell Biology, University of Tartu
2005–2007 M.Sc. student in bioinformatics, Institute of Molecular and Cell Biology, University of Tartu
2000–2005 B.Sc. student in bioinformatics, Institute of Molecular and Cell Biology, University of Tartu
1994–2000 Tartu Mart Reinik's Gymnasium
1988–1994 12th High School of Tartu

Working experience:

Since 2013 Research scientist, Institute for Molecular Medicine Finland (FIMM), University of Helsinki
2011–2013 Visiting scientist, Wellcome Trust Sanger Institute
2007–2011 Research scientist, Estonian Biocentre
2005–2007 Bioinformatics specialist, Estonian Biocentre
2004–2008 Bioinformatics specialist, Institute of Molecular and Cell Biology, University of Tartu

Main fields of research:

DNA copy number variants in health and disease; rare genetic variants in migraine and schizophrenia; enriched (loss-of-function) variants in isolated populations.

Publications:

Palta P, Kaplinski L, Nagirnaja L, Veidenberg A, Möls M, Esko T, Nelis M, Metspalu A, Laan M, Remm M. **Haplotype phasing and inheritance of copy number variants in nuclear families.** *PLoS One.* 2015;10(4):e0122713. doi: 10.1371/journal.pone.0122713. PubMed PMID:25853576.

Nagirnaja L, Palta P, Kasak L, Rull K, Christiansen OB, Nielsen HS, Steffensen R, Esko T, Remm M, Laan M. **Structural genomic variation as risk factor for idiopathic recurrent miscarriage.** *Hum Mutat.* 2014 Aug;35(8):972–82. doi: 10.1002/humu.22589. Epub 2014 Jun 24. PubMed PMID: 24827138.

Lim ET, Würtz P, Havulinna AS, Palta P, Tukiainen T, Rehnström K, Esko T, Mägi R, Inouye M, Lappalainen T, Chan Y, Salem RM, Lek M, Flannick J, Sim X, Manning A, Ladenvall C, Bumpstead S, Hämäläinen E, Aalto K, Maksimow M, Salmi M, Blankenberg S, Ardissino D, Shah S, Horne B, McPherson R, Hovingh GK, Reilly MP, Watkins H, Goel A, Farrall M, Girelli D, Reiner AP, Stitziel NO, Kathiresan S, Gabriel S, Barrett JC, Lehtimäki T, Laakso M, Groop L, Kaprio J, Perola M, McCarthy MI, Boehnke M, Altshuler DM, Lindgren CM, Hirschhorn JN, Metspalu A, Freimer NB, Zeller T, Jalkanen S, Koskinen S, Raitakari O, Durbin R, MacArthur DG, Salomaa V, Ripatti S, Daly MJ, Palotie A; Sequencing Initiative Suomi (SISu) Project. **Distribution and Medical Impact of Loss-of-Function Variants in the Finnish Founder Population.** *PLoS Genet.* 2014 Jul 31;10(7):e1004494. doi: 10.1371/journal.pgen.1004494. eCollection 2014 Jul. PubMed PMID: 25078778.

Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, Georgieva L, Rees E, Palta P, Ruderfer DM, Carrera N, Humphreys I, Johnson JS, Roussos P, Barker DD, Banks E, Milanova V, Grant SG, Hannon E, Rose SA, Chambert K, Mahajan M, Scolnick EM, Moran JL, Kirov G, Palotie A, McCarroll SA, Holmans P, Sklar P, Owen MJ, Purcell SM, O'Donovan MC. **De novo mutations in schizophrenia implicate synaptic networks.** *Nature.* 2014 Feb 13;506(7487):179–84. doi: 10.1038/nature12929. Epub 2014 Jan 22. PubMed PMID: 24463507.

Zilbauer M, Rayner TF, Clark C, Coffey AJ, Joyce CJ, Palta P, Palotie A, Lyons PA, Smith KG. **Genome-wide methylation analyses of primary human leukocyte subsets identifies functionally important cell-type-specific hypomethylated regions.** *Blood.* 2013 Dec 12;122(25):e52-60. doi: 10.1182/blood-2013-05-503201. Epub 2013 Oct 24. PubMed PMID: 24159175.

Anttila V, Winsvold BS, Gormley P, Kurth T, Bettella F, McMahon G, Kallela M, Malik R, de Vries B, Terwindt G, Medland SE, Todt U, McArdle WL, Quaye L, Koiranen M, Ikram MA, Lehtimäki T, Stam AH, Ligthart L, Wedenoja J, Dunham I, Neale BM, Palta P, Hamalainen E, Schürks M, Rose LM, Buring JE, Ridker PM, Steinberg S, Stefansson H, Jakobsson F, Lawlor DA, Evans DM, Ring SM, Färkkilä M, Artto V, Kaunisto MA, Freilinger T, Schoenen J, Frants RR, Pelzer N, Weller CM, Zielman R, Heath AC, Madden PA, Montgomery GW, Martin NG, Borck G, Göbel H, Heinze A, Heinze-Kuhn K, Williams FM, Hartikainen AL, Pouta A, van den Ende J, Uitterlinden AG, Hofman A, Amin N, Hottenga JJ, Vink JM, Heikkilä K, Alexander M, Muller-Myhsok B, Schreiber S, Meitinger T, Wichmann HE, Aromaa A, Eriksson JG, Traynor BJ, Trabzuni D, Rossin E, Lage K, Jacobs SB, Gibbs JR, Birney E, Kaprio J, Penninx BW, Boomsma DI, van Duijn C, Raitakari O, Jarvelin MR, Zwart JA, Cherkas L, Strachan DP, Kubisch C, Ferrari MD, van den Maagdenberg AM, Dichgans M, Wessman M, Smith GD, Stefansson K, Daly MJ, Nyholt DR, Chasman DI, Palotie A; North American Brain Expression Consortium; UK Brain Expression Consortium;

International Headache Genetics Consortium. **Genome-wide meta-analysis identifies new susceptibility loci for migraine.** *Nat Genet.* 2013 Aug; 45(8):912–7. doi: 10.1038/ng.2676. Epub 2013 Jun 23. PubMed PMID: 23793025.

Freilinger T, Anttila V, de Vries B, Malik R, Kallela M, Terwindt GM, Pozo-Rosich P, Winsvold B, Nyholt DR, van Oosterhout WP, Artto V, Todt U, Hämäläinen E, Fernández-Morales J, Louter MA, Kaunisto MA, Schoenen J, Raitakari O, Lehtimäki T, Vila-Pueyo M, Göbel H, Wichmann E, Sintas C, Uitterlinden AG, Hofman A, Rivadeneira F, Heinze A, Tronvik E, van Duijn CM, Kaprio J, Cormand B, Wessman M, Frants RR, Meitinger T, Müller-Myhsok B, Zwart JA, Färkkilä M, Macaya A, Ferrari MD, Kubisch C, Palotie A, Dichgans M, van den Maagdenberg AM; International Headache Genetics Consortium. **Genome-wide association analysis identifies susceptibility loci for migraine without aura.** *Nat Genet.* 2012 Jun 10;44(7):777–82. doi: 10.1038/ng.2307. PubMed PMID: 22683712.

Saare M, Sõritsa D, Vaidla K, Palta P, Remm M, Laan M, Karro H, Sõritsa A, Salumets A, D'Hooghe T, Peters M. **No evidence of somatic DNA copy number alterations in eutopic and ectopic endometrial tissue in endometriosis.** *Hum Reprod.* 2012 Jun;27(6):1857–64. doi: 10.1093/humrep/des125. Epub 2012 Apr 3. PubMed PMID: 22473391.

Clark C & Palta P, Joyce CJ, Scott C, Grundberg E, Deloukas P, Palotie A, Coffey AJ. **A comparison of the whole genome approach of MeDIP-seq to the targeted approach of the Infinium HumanMethylation450 BeadChip(®) for methylome profiling.** *PLoS One.* 2012;7(11):e50233. doi: 0.1371/journal.pone.0050233. Epub 2012 Nov 29. PubMed PMID: 23209683.

Coffey AJ, Kokocinski F, Calafato MS, Scott CE, Palta P, Drury E, Joyce CJ, Leproust EM, Harrow J, Hunt S, Lehesjoki AE, Turner DJ, Hubbard TJ, Palotie A. **The GENCODE exome: sequencing the complete human exome.** *Eur J Hum Genet.* 2011 Jul;19(7):827–31. doi: 10.1038/ejhg.2011.28. Epub 2011 Mar 2. PubMed PMID: 21364695.

Männik K, Parkel S, Palta P, Zilina O, Puusepp H, Esko T, Mägi R, Nõukas M, Veidenberg A, Nelis M, Metspalu A, Remm M, Ounap K, Kurg A. **A parallel SNP array study of genomic aberrations associated with mental retardation in patients and general population in Estonia.** *Eur J Med Genet.* 2011 Mar-Apr;54(2):136–43. doi: 10.1016/j.ejmg.2010.11.005. Epub 2010 Nov 26. PubMed PMID: 21112420.

Scheler O, Kaplinski L, Glynn B, Palta P, Parkel S, Toome K, Maher M, Barry T, Remm M, Kurg A. **Detection of NASBA amplified bacterial tmRNA molecules on SLICSel designed microarray probes.** *BMC Biotechnol.* 2011 Feb 28;11:17. doi: 10.1186/1472-6750-11-17. PubMed PMID: 21356118.

Kaplinski L, Scheler O, Parkel S, Palta P, Toome K, Kurg A, Remm M. **Detection of tmRNA molecules on microarrays at low temperatures using helper oligonucleotides.** *BMC Biotechnol.* 2010 Apr 28;10:34. doi: 10.1186/1472-6750-10-34. PubMed PMID: 20426847.

Viltrop T, Krjutskov K, Palta P, Metspalu **A. Comparison of DNA extraction methods for multiplex polymerase chain reaction.** *Anal Biochem.* 2010 Mar 15;398(2):260–2. doi: 10.1016/j.ab.2009.11.026. Epub 2009 Nov 20. PubMed PMID: 19932073.

Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, Falchi M, Chen F, Andrieux J, Lobbens S, Delobel B, Stutzmann F, El-Sayed Moustafa JS, Chèvre JC, Lecoeur C, Vatin V, Bouquillon S, Buxton JL, Boute O, Holder-Espinasse M, Cuisset JM, Lemaitre MP, Ambresin AE, Brioschi A, Gaillard M, Giusti V, Fellmann F, Ferrarini A, Hadjikhani N, Campion D, Guilmatre A, Goldenberg A, Calmels N, Mandel JL, Le Caignec C, David A, Isidor B, Cordier MP, Dupuis-Girod S, Labalme A, Sanlaville D, Béri-Dexheimer M, Jonveaux P, Leheup B, Ounap K, Bochukova EG, Henning E, Keogh J, Ellis RJ, Macdermot KD, van Haelst MM, Vincent-Delorme C, Plessis G, Touraine R, Philippe A, Malan V, Mathieu-Dramard M, Chiesa J, Blaumeiser B, Kooy RF, Caiazzo R, Pigeyre M, Balkau B, Sladek R, Bergmann S, Mooser V, Waterworth D, Reymond A, Vollenweider P, Waeber G, Kurg A, Palta P, Esko T, Metspalu A, Nelis M, Elliott P, Hartikainen AL, McCarthy MI, Peltonen L, Carlsson L, Jacobson P, Sjöström L, Huang N, Hurles ME, O'Rahilly S, Farooqi IS, Männik K, Jarvelin MR, Pattou F, Meyre D, Walley AJ, Coin LJ, lakemore AI, Froguel P, Beckmann JS. **A new highly penetrant form of obesity due to deletions on chromosome 16p11.2.** *Nature.* 2010 Feb 4;463(7281):671–5. doi: 10.1038/nature08727. PubMed PMID: 20130649.

Krjutskov K, Viltrop T, Palta P, Metspalu E, Tamm E, Suvi S, Sak K, Merilo A, Sork H, Teek R, Nikopensius T, Kivisild T, Metspalu A. **Evaluation of the 124-plex SNP typing microarray for forensic testing.** *Forensic Sci Int Genet.* 2009 Dec;4(1):43–8. doi: 10.1016/j.fsigen.2009.04.007. Epub 2009 May 15. PubMed PMID: 19948333.

Scheler O, Glynn B, Parkel S, Palta P, Toome K, Kaplinski L, Remm M, Maher M, Kurg A. **Fluorescent labeling of NASBA amplified tmRNA molecules for microarray applications.** *BMC Biotechnol.* 2009 May 15;9:45. doi: 10.1186/1472-6750-9-45. PubMed PMID: 19445684.

Kousoulidou L, Männik K, Zilina O, Parkel S, Palta P, Remm M, Kurg A, Patsalis PC. **Application of two different microarray-based copy-number detection methodologies – array-comparative genomic hybridization and array-multiplex amplifiable probe hybridization – with identical amplifiable target sequences.** *Clin Chem Lab Med.* 2008;46(5):722–4. doi: 10.1515/CCLM.2008.141. PubMed PMID: 18598207.

Kousoulidou L, Männik K, Sismani C, Zilina O, Parkel S, Puusepp H, Tõnisson N, Palta P, Remm M, Kurg A, Patsalis PC. **Array-MAPH: a methodology**

for the detection of locus copy-number changes in complex genomes. *Nat Protoc.* 2008;3(5):849–65. doi: 10.1038/nprot.2008.49. PubMed PMID: 18451793.

Kousoulidou L, Parkel S, Zilina O, Palta P, Puusepp H, Remm M, Turner G, Boyle J, van Bokhoven H, de Brouwer A, Van Esch H, Froyen G, Ropers HH, Chelly J, Moraine C, Gecz J, Kurg A, Patsalis PC. **Screening of 20 patients with X-linked mental retardation using chromosome X-specific array-MAPH.** *Eur J Med Genet.* 2007 Nov-Dec;50(6):399–410. Epub 2007 Sep 29. PubMed PMID: 17980689.

Patsalis PC, Kousoulidou L, Männik K, Sismani C, Zilina O, Parkel S, Puusepp H, Tõnisson N, Palta P, Remm M, Kurg A. **Detection of small genomic imbalances using microarray-based multiplex amplifiable probe hybridization.** *Eur J Hum Genet.* 2007 Feb;15(2):162–72. Epub 2006 Nov 22. PubMed PMID: 17119536.

Scholarships and awards:

2013	Stipend of Finnish Cultural Foundation
2007	Artur Lind’s stipend, Estonian Genome Foundation
2007	Kristjan Jaak’s stipend, Archimedes Foundation

Other scientific activities:

Since 2012	Member of International Headache Society
Since 2007	Member of the Estonian Society of Human Genetics

Supervised dissertations:

2008–2009	Andres Veidenberg, M.Sc., Department of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu
2007–2008	Taavi Võsumaa (co-supervisor), M.Sc., Department of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu
2007–2008	Ksenia George, B.Sc., Department of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu

ELULOOKIRJELDUS

Nimi: Priit Palta
Sünniaeg: 25. aprill, 1981
Aadress: TÜ Molekulaar- ja Rakubioloogia Instituut, Riia 23, Tartu 51010
E-post: priit.palta@ut.ee
Praegune töökoht: Teadur, Soome Molekulaarse Meditsiini Instituut (FIMM), Helsingi Ülikool

Haridus:

2007–2011 Ph.D. õpingud bioinformaatikas, Molekulaar- ja Rakubioloogia Instituut, Tartu Ülikool
2005–2007 M.Sc. õpingud bioinformaatikas, Molekulaar- ja Rakubioloogia Instituut, Tartu Ülikool
2000–2005 B.Sc. õpingud bioinformaatikas, Molekulaar- ja Rakubioloogia Instituut, Tartu Ülikool
1994–2000 Tartu Mart Reiniku Gümnaasium
1988–1994 Tartu 12. Keskkool

Töökogemus:

Alates 2013 Teadur, Soome Molekulaarse Meditsiini Instituut (FIMM), Helsingi Ülikool
2011–2013 Külalisteadur, Wellcome Trust Sanger'i Instituut
2007–2011 Teadur, Eesti Biokeskus
2005–2007 Bioinformaatika spetsialist, Eesti Biokeskus
2004–2008 Bioinformaatika spetsialist, Molekulaar- ja Rakubioloogia Instituut, Tartu Ülikool

Peamised uurimisvaldkonnad:

DNA koopiaarvu variandid ja nende seos haigustega, migreeni ja skisofreeniaga seotud mittesagedased geneetilised variandid ja kõrgenenud alleelisagedusega geneetilised variandid isoleeritud populatsioonides.

Publikatsioonide loetelu:

Palta P, Kaplinski L, Nagirnaja L, Veidenberg A, Möls M, Esko T, Nelis M, Metspalu A, Laan M, Remm M. **Haplotype phasing and inheritance of copy number variants in nuclear families.** *PLoS One.* 2015;10(4):e0122713. doi: 10.1371/journal.pone.0122713. PubMed PMID:25853576.

Nagirnaja L, Palta P, Kasak L, Rull K, Christiansen OB, Nielsen HS, Steffensen R, Esko T, Remm M, Laan M. **Structural genomic variation as risk factor for idiopathic recurrent miscarriage.** *Hum Mutat.* 2014 Aug;35(8):972–82. doi: 10.1002/humu.22589. Epub 2014 Jun 24. PubMed PMID: 24827138.

Lim ET, Würtz P, Havulinna AS, Palta P, Tukiainen T, Rehnström K, Esko T, Mägi R, Inouye M, Lappalainen T, Chan Y, Salem RM, Lek M, Flannick J, Sim X, Manning A, Ladenvall C, Bumpstead S, Hämäläinen E, Aalto K, Maksimow M, Salmi M, Blankenberg S, Ardissino D, Shah S, Horne B, McPherson R, Hovingh GK, Reilly MP, Watkins H, Goel A, Farrall M, Girelli D, Reiner AP, Stitziel NO, Kathiresan S, Gabriel S, Barrett JC, Lehtimäki T, Laakso M, Groop L, Kaprio J, Perola M, McCarthy MI, Boehnke M, Altshuler DM, Lindgren CM, Hirschhorn JN, Metspalu A, Freimer NB, Zeller T, Jalkanen S, Koskinen S, Raitakari O, Durbin R, MacArthur DG, Salomaa V, Ripatti S, Daly MJ, Palotie A; Sequencing Initiative Suomi (SISu) Project. **Distribution and Medical Impact of Loss-of-Function Variants in the Finnish Founder Population.** *PLoS Genet.* 2014 Jul 31;10(7):e1004494. doi: 10.1371/journal.pgen.1004494. eCollection 2014 Jul. PubMed PMID: 25078778.

Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, Georgieva L, Rees E, Palta P, Ruderfer DM, Carrera N, Humphreys I, Johnson JS, Roussos P, Barker DD, Banks E, Milanova V, Grant SG, Hannon E, Rose SA, Chambert K, Mahajan M, Scolnick EM, Moran JL, Kirov G, Palotie A, McCarroll SA, Holmans P, Sklar P, Owen MJ, Purcell SM, O'Donovan MC. **De novo mutations in schizophrenia implicate synaptic networks.** *Nature.* 2014 Feb 13;506(7487):179–84. doi: 10.1038/nature12929. Epub 2014 Jan 22. PubMed PMID: 24463507.

Zilbauer M, Rayner TF, Clark C, Coffey AJ, Joyce CJ, Palta P, Palotie A, Lyons PA, Smith KG. **Genome-wide methylation analyses of primary human leukocyte subsets identifies functionally important cell-type-specific hypomethylated regions.** *Blood.* 2013 Dec 12;122(25):e52–60. doi: 10.1182/blood-2013-05-503201. Epub 2013 Oct 24. PubMed PMID: 24159175.

Anttila V, Winsvold BS, Gormley P, Kurth T, Bettella F, McMahon G, Kallela M, Malik R, de Vries B, Terwindt G, Medland SE, Todt U, McArdle WL, Quaye L, Koiranen M, Ikram MA, Lehtimäki T, Stam AH, Ligthart L, Wedenoja J, Dunham I, Neale BM, Palta P, Hamalainen E, Schürks M, Rose LM, Buring JE, Ridker PM, Steinberg S, Stefansson H, Jakobsson F, Lawlor DA, Evans DM, Ring SM, Färkkilä M, Artto V, Kaunisto MA, Freilinger T, Schoenen J, Frants RR, Pelzer N, Weller CM, Zielman R, Heath AC, Madden PA, Montgomery GW, Martin NG, Borck G, Göbel H, Heinze A, Heinze-Kuhn K, Williams FM, Hartikainen AL, Pouta A, van den Ende J, Uitterlinden AG, Hofman A, Amin N, Hottenga JJ, Vink JM, Heikkilä K, Alexander M, Muller-Myhsok B, Schreiber S, Meitinger T, Wichmann HE, Aromaa A, Eriksson JG, Traynor BJ, Trabzuni D, Rossin E, Lage K, Jacobs SB, Gibbs JR, Birney E, Kaprio J, Penninx BW, Boomsma DI, van Duijn C, Raitakari O, Jarvelin MR, Zwart JA, Cherkas L, Strachan DP, Kubisch C, Ferrari MD, van den Maagdenberg AM, Dichgans M, Wessman M, Smith GD, Stefansson K, Daly MJ, Nyholt DR, Chasman DI, Palotie A; North

American Brain Expression Consortium; UK Brain Expression Consortium; International Headache Genetics Consortium. **Genome-wide meta-analysis identifies new susceptibility loci for migraine.** *Nat Genet.* 2013 Aug;45(8):912–7. doi: 10.1038/ng.2676. Epub 2013 Jun 23. PubMed PMID: 23793025.

Freilinger T, Anttila V, de Vries B, Malik R, Kallela M, Terwindt GM, Pozo-Rosich P, Winsvold B, Nyholt DR, van Oosterhout WP, Artto V, Todt U, Hämäläinen E, Fernández-Morales J, Louter MA, Kaunisto MA, Schoenen J, Raitakari O, Lehtimäki T, Vila-Pueyo M, Göbel H, Wichmann E, Sintas C, Uitterlinden AG, Hofman A, Rivadeneira F, Heinze A, Tronvik E, van Duijn CM, Kaprio J, Cormand B, Wessman M, Frants RR, Meitinger T, Müller-Myhsok B, Zwart JA, Färkkilä M, Macaya A, Ferrari MD, Kubisch C, Palotie A, Dichgans M, van den Maagdenberg AM; International Headache Genetics Consortium. **Genome-wide association analysis identifies susceptibility loci for migraine without aura.** *Nat Genet.* 2012 Jun 10;44(7):777–82. doi: 10.1038/ng.2307. PubMed PMID: 22683712.

Saare M, Sõritsa D, Vaidla K, Palta P, Remm M, Laan M, Karro H, Sõritsa A, Salumets A, D'Hooghe T, Peters M. **No evidence of somatic DNA copy number alterations in eutopic and ectopic endometrial tissue in endometriosis.** *Hum Reprod.* 2012 Jun;27(6):1857–64. doi: 10.1093/humrep/des125. Epub 2012 Apr 3. PubMed PMID: 22473391.

Clark C & Palta P, Joyce CJ, Scott C, Grundberg E, Deloukas P, Palotie A, Coffey AJ. **A comparison of the whole genome approach of MeDIP-seq to the targeted approach of the Infinium HumanMethylation450 BeadChip(®) for methylome profiling.** *PLoS One.* 2012;7(11):e50233. doi: 0.1371/journal.pone.0050233. Epub 2012 Nov 29. PubMed PMID: 23209683.

Coffey AJ, Kokocinski F, Calafato MS, Scott CE, Palta P, Drury E, Joyce CJ, Leproust EM, Harrow J, Hunt S, Lehesjoki AE, Turner DJ, Hubbard TJ, Palotie A. **The GENCODE exome: sequencing the complete human exome.** *Eur J Hum Genet.* 2011 Jul;19(7):827–31. doi: 10.1038/ejhg.2011.28. Epub 2011 Mar 2. PubMed PMID: 21364695.

Männik K, Parkel S, Palta P, Zilina O, Puusepp H, Esko T, Mägi R, Nõukas M, Veidenberg A, Nelis M, Metspalu A, Remm M, Ounap K, Kurg A. **A parallel SNP array study of genomic aberrations associated with mental retardation in patients and general population in Estonia.** *Eur J Med Genet.* 2011 Mar-Apr;54(2):136–43. doi: 10.1016/j.ejmg.2010.11.005. Epub 2010 Nov 26. PubMed PMID: 21112420.

Scheler O, Kaplinski L, Glynn B, Palta P, Parkel S, Toome K, Maher M, Barry T, Remm M, Kurg A. **Detection of NASBA amplified bacterial tmRNA molecules on SLICSel designed microarray probes.** *BMC Biotechnol.* 2011 Feb 28;11:17. doi: 10.1186/1472-6750-11-17. PubMed PMID: 21356118.

Kaplinski L, Scheler O, Parkel S, Palta P, Toome K, Kurg A, Remm M. **Detection of tmRNA molecules on microarrays at low temperatures using helper oligonucleotides.** *BMC Biotechnol.* 2010 Apr 28;10:34. doi: 10.1186/1472-6750-10-34. PubMed PMID: 20426847.

Viltrop T, Krjutskov K, Palta P, Metspalu **A. Comparison of DNA extraction methods for multiplex polymerase chain reaction.** *Anal Biochem.* 2010 Mar 15;398(2):260–2. doi: 10.1016/j.ab.2009.11.026. Epub 2009 Nov 20. PubMed PMID: 19932073.

Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, Falchi M, Chen F, Andrieux J, Lobbens S, Delobel B, Stutzmann F, El-Sayed Moustafa JS, Chèvre JC, Lecoeur C, Vatin V, Bouquillon S, Buxton JL, Boute O, Holder-Espinasse M, Cuisset JM, Lemaitre MP, Ambresin AE, Brioschi A, Gaillard M, Giusti V, Fellmann F, Ferrarini A, Hadjikhani N, Campion D, Guilmatre A, Goldenberg A, Calmels N, Mandel JL, Le Caignec C, David A, Isidor B, Cordier MP, Dupuis-Girod S, Labalme A, Sanlaville D, Béri-Dexheimer M, Jonveaux P, Leheup B, Ounap K, Bochukova EG, Henning E, Keogh J, Ellis RJ, Macdermot KD, van Haelst MM, Vincent-Delorme C, Plessis G, Touraine R, Philippe A, Malan V, Mathieu-Dramard M, Chiesa J, Blaumeiser B, Kooy RF, Caiazzo R, Pigeyre M, Balkau B, Sladek R, Bergmann S, Mooser V, Waterworth D, Reymond A, Vollenweider P, Waeber G, Kurg A, Palta P, Esko T, Metspalu A, Nelis M, Elliott P, Hartikainen AL, McCarthy MI, Peltonen L, Carlsson L, Jacobson P, Sjöström L, Huang N, Hurles ME, O'Rahilly S, Farooqi IS, Männik K, Jarvelin MR, Pattou F, Meyre D, Walley AJ, Coin LJ, lakemore AI, Froguel P, Beckmann JS. **A new highly penetrant form of obesity due to deletions on chromosome 16p11.2.** *Nature.* 2010 Feb 4;463(7281):671–5. doi: 10.1038/nature08727. PubMed PMID: 20130649.

Krjutskov K, Viltrop T, Palta P, Metspalu E, Tamm E, Suvi S, Sak K, Merilo A, Sork H, Teek R, Nikopensius T, Kivisild T, Metspalu A. **Evaluation of the 124-plex SNP typing microarray for forensic testing.** *Forensic Sci Int Genet.* 2009 Dec;4(1):43–8. doi: 10.1016/j.fsigen.2009.04.007. Epub 2009 May 15. PubMed PMID: 19948333.

Scheler O, Glynn B, Parkel S, Palta P, Toome K, Kaplinski L, Remm M, Maher M, Kurg A. **Fluorescent labeling of NASBA amplified tmRNA molecules for microarray applications.** *BMC Biotechnol.* 2009 May 15;9:45. doi: 10.1186/1472-6750-9-45. PubMed PMID: 19445684.

Kousoulidou L, Männik K, Zilina O, Parkel S, Palta P, Remm M, Kurg A, Patsalis PC. **Application of two different microarray-based copy-number detection methodologies – array-comparative genomic hybridization and array-multiplex amplifiable probe hybridization – with identical amplifiable target sequences.** *Clin Chem Lab Med.* 2008;46(5):722–4. doi: 10.1515/CCLM.2008.141. PubMed PMID: 18598207.

Kousoulidou L, Männik K, Sismani C, Zilina O, Parkel S, Puusepp H, Tõnisson N, Palta P, Remm M, Kurg A, Patsalis PC. **Array-MAPH: a methodology**

for the detection of locus copy-number changes in complex genomes. *Nat Protoc.* 2008;3(5):849–65. doi: 10.1038/nprot.2008.49. PubMed PMID: 18451793.

Kousoulidou L, Parkel S, Zilina O, Palta P, Puusepp H, Remm M, Turner G, Boyle J, van Bokhoven H, de Brouwer A, Van Esch H, Froyen G, Ropers HH, Chelly J, Moraine C, Gecz J, Kurg A, Patsalis PC. **Screening of 20 patients with X-linked mental retardation using chromosome X-specific array-MAPH.** *Eur J Med Genet.* 2007 Nov-Dec;50(6):399–410. Epub 2007 Sep 29. PubMed PMID: 17980689.

Patsalis PC, Kousoulidou L, Männik K, Sismani C, Zilina O, Parkel S, Puusepp H, Tõnisson N, Palta P, Remm M, Kurg A. **Detection of small genomic imbalances using microarray-based multiplex amplifiable probe hybridization.** *Eur J Hum Genet.* 2007 Feb;15(2):162–72. Epub 2006 Nov 22. PubMed PMID: 17119536.

Saadud stipendiumid:

2013	Soome Kultuuri Sihtasutuse stipendium
2007	Artur Linnu nimeline stipendium, Sihtasutus Geenikeskus
2007	Kristjan Jaagu nimeline stipendium, Sihtasutus Archimedes

Muu teaduslik organisatsiooniline ja erialane tegevus:

Alates 2012	International Headache Society liige
Alates 2007	Eesti Inimesegeneetika Ühingu liige

Juhendatud väitekirjad:

2008–2009	Andres Veidenberg, M.Sc., bioinformaatika õppetool, Molekulaar- ja Rakubioloogia Instituut, Tartu Ülikool
2007–2008	Taavi Võsumaa (co-supervisor), M.Sc., bioinformaatika õppetool, Molekulaar- ja Rakubioloogia Instituut, Tartu Ülikool
2007–2008	Ksenia George, B.Sc., bioinformaatika õppetool, Molekulaar- ja Rakubioloogia Instituut, Tartu Ülikool

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida.* Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand.** The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995, 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996, 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996, 60 p.
19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996, 150 p.

20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996, 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu, 1996, 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996, 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997, 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997, 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997, 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997, 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997, 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997, 80 p.
29. **Arvo Tuvikene**. Assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro.* Tartu, 1997, 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997, 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997, 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997, 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu, 1998, 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998, 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998, 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998, 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous grassland plant species. Tartu, 1998, 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998, 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998, 98 p.
40. **Arnold Kristjuhan.** Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998, 86 p.

41. **Sulev Ingerpuu.** Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998, 163 p.
42. **Veljo Kisand.** Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998, 118 p.
43. **Kadri Põldmaa.** Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998, 178 p.
44. **Markus Vetemaa.** Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998, 117 p.
45. **Heli Talvik.** Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998, 104 p.
46. **Katrin Heinsoo.** Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1999, 133 p.
47. **Tarmo Annilo.** Studies on mammalian ribosomal protein S7. Tartu, 1998, 77 p.
48. **Indrek Ots.** Health state indicies of reproducing great tits (*Parus major*): sources of variation and connections with life-history traits. Tartu, 1999, 117 p.
49. **Juan Jose Cantero.** Plant community diversity and habitat relationships in central Argentina grasslands. Tartu, 1999, 161 p.
50. **Rein Kalamees.** Seed bank, seed rain and community regeneration in Estonian calcareous grasslands. Tartu, 1999, 107 p.
51. **Sulev Kõks.** Cholecystokinin (CCK) — induced anxiety in rats: influence of environmental stimuli and involvement of endopioid mechanisms and serotonin. Tartu, 1999, 123 p.
52. **Ebe Sild.** Impact of increasing concentrations of O_3 and CO_2 on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva.** Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna.** Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro.** Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane.** Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
57. **Tiina Tamm.** Cocksfoot mottle virus: the genome organisation and translational strategies. Tartu, 2000, 101 p.
58. **Reet Kurg.** Structure-function relationship of the bovine papilloma virus E2 protein. Tartu, 2000, 89 p.
59. **Toomas Kivisild.** The origins of Southern and Western Eurasian populations: an mtDNA study. Tartu, 2000, 121 p.
60. **Niilo Kaldalu.** Studies of the TOL plasmid transcription factor XylS. Tartu 2000. 88 p.

61. **Dina Lepik.** Modulation of viral DNA replication by tumor suppressor protein p53. Tartu 2000. 106 p.
62. **Kai Vellak.** Influence of different factors on the diversity of the bryophyte vegetation in forest and wooded meadow communities. Tartu 2000. 122 p.
63. **Jonne Kotta.** Impact of eutrophication and biological invasionas on the structure and functions of benthic macrofauna. Tartu 2000. 160 p.
64. **Georg Martin.** Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000. 139 p.
65. **Silvia Sepp.** Morphological and genetical variation of *Alchemilla L.* in Estonia. Tartu, 2000. 124 p.
66. **Jaan Liira.** On the determinants of structure and diversity in herbaceous plant communities. Tartu, 2000. 96 p.
67. **Priit Zingel.** The role of planktonic ciliates in lake ecosystems. Tartu 2001. 111 p.
68. **Tiit Teder.** Direct and indirect effects in Host-parasitoid interactions: ecological and evolutionary consequences. Tartu 2001. 122 p.
69. **Hannes Kollist.** Leaf apoplastic ascorbate as ozone scavenger and its transport across the plasma membrane. Tartu 2001. 80 p.
70. **Reet Marits.** Role of two-component regulator system PehR-PehS and extracellular protease PrtW in virulence of *Erwinia Carotovora* subsp. *Carotovora*. Tartu 2001. 112 p.
71. **Vallo Tilgar.** Effect of calcium supplementation on reproductive performance of the pied flycatcher *Ficedula hypoleuca* and the great tit *Parus major,* breeding in Nothern temperate forests. Tartu, 2002. 126 p.
72. **Rita Hõrak.** Regulation of transposition of transposon Tn*4652* in *Pseudomonas putida*. Tartu, 2002. 108 p.
73. **Liina Eek-Piirsoo.** The effect of fertilization, mowing and additional illumination on the structure of a species-rich grassland community. Tartu, 2002. 74 p.
74. **Krõõt Aasamaa.** Shoot hydraulic conductance and stomatal conductance of six temperate deciduous tree species. Tartu, 2002. 110 p.
75. **Nele Ingerpuu.** Bryophyte diversity and vascular plants. Tartu, 2002. 112 p.
76. **Neeme Tõnisson.** Mutation detection by primer extension on oligonucleotide microarrays. Tartu, 2002. 124 p.
77. **Margus Pensa.** Variation in needle retention of Scots pine in relation to leaf morphology, nitrogen conservation and tree age. Tartu, 2003. 110 p.
78. **Asko Lõhmus.** Habitat preferences and quality for birds of prey: from principles to applications. Tartu, 2003. 168 p.
79. **Viljar Jaks.** p53 — a switch in cellular circuit. Tartu, 2003. 160 p.
80. **Jaana Männik.** Characterization and genetic studies of four ATP-binding cassette (ABC) transporters. Tartu, 2003. 140 p.
81. **Marek Sammul.** Competition and coexistence of clonal plants in relation to productivity. Tartu, 2003. 159 p

82. **Ivar Ilves.** Virus-cell interactions in the replication cycle of bovine papillomavirus type 1. Tartu, 2003. 89 p.
83. **Andres Männik.** Design and characterization of a novel vector system based on the stable replicator of bovine papillomavirus type 1. Tartu, 2003. 109 p.
84. **Ivika Ostonen.** Fine root structure, dynamics and proportion in net primary production of Norway spruce forest ecosystem in relation to site conditions. Tartu, 2003. 158 p.
85. **Gudrun Veldre.** Somatic status of 12–15-year-old Tartu schoolchildren. Tartu, 2003. 199 p.
86. **Ülo Väli.** The greater spotted eagle *Aquila clanga* and the lesser spotted eagle *A. pomarina*: taxonomy, phylogeography and ecology. Tartu, 2004. 159 p.
87. **Aare Abroi.** The determinants for the native activities of the bovine papillomavirus type 1 E2 protein are separable. Tartu, 2004. 135 p.
88. **Tiina Kahre.** Cystic fibrosis in Estonia. Tartu, 2004. 116 p.
89. **Helen Orav-Kotta.** Habitat choice and feeding activity of benthic suspension feeders and mesograzers in the northern Baltic Sea. Tartu, 2004. 117 p.
90. **Maarja Öpik.** Diversity of arbuscular mycorrhizal fungi in the roots of perennial plants and their effect on plant performance. Tartu, 2004. 175 p.
91. **Kadri Tali.** Species structure of *Neotinea ustulata*. Tartu, 2004. 109 p.
92. **Kristiina Tambets.** Towards the understanding of post-glacial spread of human mitochondrial DNA haplogroups in Europe and beyond: a phylogeographic approach. Tartu, 2004. 163 p.
93. **Arvi Jõers.** Regulation of p53-dependent transcription. Tartu, 2004. 103 p.
94. **Lilian Kadaja.** Studies on modulation of the activity of tumor suppressor protein p53. Tartu, 2004. 103 p.
95. **Jaak Truu.** Oil shale industry wastewater: impact on river microbial community and possibilities for bioremediation. Tartu, 2004. 128 p.
96. **Maire Peters.** Natural horizontal transfer of the *pheBA* operon. Tartu, 2004. 105 p.
97. **Ülo Maiväli.** Studies on the structure-function relationship of the bacterial ribosome. Tartu, 2004. 130 p.
98. **Merit Otsus.** Plant community regeneration and species diversity in dry calcareous grasslands. Tartu, 2004. 103 p.
99. **Mikk Heidemaa.** Systematic studies on sawflies of the genera *Dolerus, Empria,* and *Caliroa* (Hymenoptera: Tenthredinidae). Tartu, 2004. 167 p.
100. **Ilmar Tõnno.** The impact of nitrogen and phosphorus concentration and N/P ratio on cyanobacterial dominance and N_2 fixation in some Estonian lakes. Tartu, 2004. 111 p.
101. **Lauri Saks.** Immune function, parasites, and carotenoid-based ornaments in greenfinches. Tartu, 2004. 144 p.
102. **Siiri Rootsi.** Human Y-chromosomal variation in European populations. Tartu, 2004. 142 p.

103. **Eve Vedler.** Structure of the 2,4-dichloro-phenoxyacetic acid-degradative plasmid pEST4011. Tartu, 2005. 106 p.
104. **Andres Tover.** Regulation of transcription of the phenol degradation *pheBA* operon in *Pseudomonas putida*. Tartu, 2005. 126 p.
105. **Helen Udras.** Hexose kinases and glucose transport in the yeast *Hansenula polymorpha*. Tartu, 2005. 100 p.
106. **Ave Suija.** Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005. 162 p.
107. **Piret Lõhmus.** Forest lichens and their substrata in Estonia. Tartu, 2005. 162 p.
108. **Inga Lips.** Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005. 156 p.
109. **Kaasik, Krista.** Circadian clock genes in mammalian clockwork, metabolism and behaviour. Tartu, 2005. 121 p.
110. **Juhan Javoiš.** The effects of experience on host acceptance in ovipositing moths. Tartu, 2005. 112 p.
111. **Tiina Sedman.** Characterization of the yeast *Saccharomyces cerevisiae* mitochondrial DNA helicase Hmi1. Tartu, 2005. 103 p.
112. **Ruth Aguraiuja.** Hawaiian endemic fern lineage *Diellia* (Aspleniaceae): distribution, population structure and ecology. Tartu, 2005. 112 p.
113. **Riho Teras.** Regulation of transcription from the fusion promoters generated by transposition of Tn*4652* into the upstream region of *pheBA* operon in *Pseudomonas putida*. Tartu, 2005. 106 p.
114. **Mait Metspalu.** Through the course of prehistory in india: tracing the mtDNA trail. Tartu, 2005. 138 p.
115. **Elin Lõhmussaar.** The comparative patterns of linkage disequilibrium in European populations and its implication for genetic association studies. Tartu, 2006. 124 p.
116. **Priit Kupper.** Hydraulic and environmental limitations to leaf water relations in trees with respect to canopy position. Tartu, 2006. 126 p.
117. **Heili Ilves.** Stress-induced transposition of Tn*4652* in *Pseudomonas Putida.* Tartu, 2006. 120 p.
118. **Silja Kuusk.** Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006. 126 p.
119. **Kersti Püssa.** Forest edges on medium resolution landsat thematic mapper satellite images. Tartu, 2006. 90 p.
120. **Lea Tummeleht.** Physiological condition and immune function in great tits (*Parus major* l.): Sources of variation and trade-offs in relation to growth. Tartu, 2006. 94 p.
121. **Toomas Esperk.** Larval instar as a key element of insect growth schedules. Tartu, 2006. 186 p.
122. **Harri Valdmann.** Lynx (*Lynx lynx*) and wolf (*Canis lupus*) in the Baltic region: Diets, helminth parasites and genetic variation. Tartu, 2006. 102 p.

123. **Priit Jõers.** Studies of the mitochondrial helicase Hmi1p in *Candida albicans* and *Saccharomyces cerevisia*. Tartu, 2006. 113 p.
124. **Kersti Lilleväli.** Gata3 and Gata2 in inner ear development. Tartu, 2007. 123 p.
125. **Kai Rünk.** Comparative ecology of three fern species: *Dryopteris carthusiana* (Vill.) H.P. Fuchs, *D. expansa* (C. Presl) Fraser-Jenkins & Jermy and *D. dilatata* (Hoffm.) A. Gray (Dryopteridaceae). Tartu, 2007. 143 p.
126. **Aveliina Helm.** Formation and persistence of dry grassland diversity: role of human history and landscape structure. Tartu, 2007. 89 p.
127. **Leho Tedersoo.** Ectomycorrhizal fungi: diversity and community structure in Estonia, Seychelles and Australia. Tartu, 2007. 233 p.
128. **Marko Mägi.** The habitat-related variation of reproductive performance of great tits in a deciduous-coniferous forest mosaic: looking for causes and consequences. Tartu, 2007. 135 p.
129. **Valeria Lulla.** Replication strategies and applications of Semliki Forest virus. Tartu, 2007. 109 p.
130. **Ülle Reier**. Estonian threatened vascular plant species: causes of rarity and conservation. Tartu, 2007. 79 p.
131. **Inga Jüriado**. Diversity of lichen species in Estonia: influence of regional and local factors. Tartu, 2007. 171 p.
132. **Tatjana Krama.** Mobbing behaviour in birds: costs and reciprocity based cooperation. Tartu, 2007. 112 p.
133. **Signe Saumaa.** The role of DNA mismatch repair and oxidative DNA damage defense systems in avoidance of stationary phase mutations in *Pseudomonas putida.* Tartu, 2007. 172 p.
134. **Reedik Mägi**. The linkage disequilibrium and the selection of genetic markers for association studies in european populations. Tartu, 2007. 96 p.
135. **Priit Kilgas.** Blood parameters as indicators of physiological condition and skeletal development in great tits (*Parus major*): natural variation and application in the reproductive ecology of birds. Tartu, 2007. 129 p.
136. **Anu Albert**. The role of water salinity in structuring eastern Baltic coastal fish communities. Tartu, 2007. 95 p.
137. **Kärt Padari.** Protein transduction mechanisms of transportans. Tartu, 2008. 128 p.
138. **Siiri-Lii Sandre.** Selective forces on larval colouration in a moth. Tartu, 2008. 125 p.
139. **Ülle Jõgar.** Conservation and restoration of semi-natural floodplain meadows and their rare plant species. Tartu, 2008. 99 p.
140. **Lauri Laanisto.** Macroecological approach in vegetation science: generality of ecological relationships at the global scale. Tartu, 2008. 133 p.
141. **Reidar Andreson**. Methods and software for predicting PCR failure rate in large genomes. Tartu, 2008. 105 p.
142. **Birgot Paavel.** Bio-optical properties of turbid lakes. Tartu, 2008. 175 p.

143. **Kaire Torn.** Distribution and ecology of charophytes in the Baltic Sea. Tartu, 2008, 98 p.
144. **Vladimir Vimberg.** Peptide mediated macrolide resistance. Tartu, 2008, 190 p.
145. **Daima Örd.** Studies on the stress-inducible pseudokinase TRB3, a novel inhibitor of transcription factor ATF4. Tartu, 2008, 108 p.
146. **Lauri Saag.** Taxonomic and ecologic problems in the genus *Lepraria* (*Stereocaulaceae*, lichenised *Ascomycota*). Tartu, 2008, 175 p.
147. **Ulvi Karu.** Antioxidant protection, carotenoids and coccidians in greenfinches – assessment of the costs of immune activation and mechanisms of parasite resistance in a passerine with carotenoid-based ornaments. Tartu, 2008, 124 p.
148. **Jaanus Remm.** Tree-cavities in forests: density, characteristics and occupancy by animals. Tartu, 2008, 128 p.
149. **Epp Moks.** Tapeworm parasites *Echinococcus multilocularis* and *E. granulosus* in Estonia: phylogenetic relationships and occurrence in wild carnivores and ungulates. Tartu, 2008, 82 p.
150. **Eve Eensalu.** Acclimation of stomatal structure and function in tree canopy: effect of light and CO_2 concentration. Tartu, 2008, 108 p.
151. **Janne Pullat**. Design, functionlization and application of an *in situ* synthesized oligonucleotide microarray. Tartu, 2008, 108 p.
152. **Marta Putrinš.** Responses of *Pseudomonas putida* to phenol-induced metabolic and stress signals. Tartu, 2008, 142 p.
153. **Marina Semtšenko.** Plant root behaviour: responses to neighbours and physical obstructions. Tartu, 2008, 106 p.
154. **Marge Starast.** Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.
155. **Age Tats.** Sequence motifs influencing the efficiency of translation. Tartu, 2009, 104 p.
156. **Radi Tegova.** The role of specialized DNA polymerases in mutagenesis in *Pseudomonas putida.* Tartu, 2009, 124 p.
157. **Tsipe Aavik.** Plant species richness, composition and functional trait pattern in agricultural landscapes – the role of land use intensity and landscape structure. Tartu, 2009, 112 p.
158. **Kaja Kiiver.** Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepaciviruses. Tartu, 2009, 104 p.
159. **Meelis Kadaja.** Papillomavirus Replication Machinery Induces Genomic Instability in its Host Cell. Tartu, 2009, 126 p.
160. **Pille Hallast.** Human and chimpanzee Luteinizing hormone/Chorionic Gonadotropin beta (*LHB/CGB*) gene clusters: diversity and divergence of young duplicated genes. Tartu, 2009, 168 p.
161. **Ain Vellak.** Spatial and temporal aspects of plant species conservation. Tartu, 2009, 86 p.

162. **Triinu Remmel.** Body size evolution in insects with different colouration strategies: the role of predation risk. Tartu, 2009, 168 p.
163. **Jaana Salujõe.** Zooplankton as the indicator of ecological quality and fish predation in lake ecosystems. Tartu, 2009, 129 p.
164. **Ele Vahtmäe.** Mapping benthic habitat with remote sensing in optically complex coastal environments. Tartu, 2009, 109 p.
165. **Liisa Metsamaa.** Model-based assessment to improve the use of remote sensing in recognition and quantitative mapping of cyanobacteria. Tartu, 2009, 114 p.
166. **Pille Säälik.** The role of endocytosis in the protein transduction by cell-penetrating peptides. Tartu, 2009, 155 p.
167. **Lauri Peil.** Ribosome assembly factors in *Escherichia coli.* Tartu, 2009, 147 p.
168. **Lea Hallik.** Generality and specificity in light harvesting, carbon gain capacity and shade tolerance among plant functional groups. Tartu, 2009, 99 p.
169. **Mariliis Tark.** Mutagenic potential of DNA damage repair and tolerance mechanisms under starvation stress. Tartu, 2009, 191 p.
170. **Riinu Rannap.** Impacts of habitat loss and restoration on amphibian populations. Tartu, 2009, 117 p.
171. **Maarja Adojaan.** Molecular variation of HIV-1 and the use of this knowledge in vaccine development. Tartu, 2009, 95 p.
172. **Signe Altmäe.** Genomics and transcriptomics of human induced ovarian folliculogenesis. Tartu, 2010, 179 p.
173. **Triin Suvi.** Mycorrhizal fungi of native and introduced trees in the Seychelles Islands. Tartu, 2010, 107 p.
174. **Velda Lauringson.** Role of suspension feeding in a brackish-water coastal sea. Tartu, 2010, 123 p.
175. **Eero Talts.** Photosynthetic cyclic electron transport – measurement and variably proton-coupled mechanism. Tartu, 2010, 121 p.
176. **Mari Nelis.** Genetic structure of the Estonian population and genetic distance from other populations of European descent. Tartu, 2010, 97 p.
177. **Kaarel Krjutškov.** Arrayed Primer Extension-2 as a multiplex PCR-based method for nucleic acid variation analysis: method and applications. Tartu, 2010, 129 p.
178. **Egle Köster.** Morphological and genetical variation within species complexes: *Anthyllis vulneraria* s. l. and *Alchemilla vulgaris* (coll.). Tartu, 2010, 101 p.
179. **Erki Õunap.** Systematic studies on the subfamily Sterrhinae (Lepidoptera: Geometridae). Tartu, 2010, 111 p.
180. **Merike Jõesaar.** Diversity of key catabolic genes at degradation of phenol and *p*-cresol in pseudomonads. Tartu, 2010, 125 p.
181. **Kristjan Herkül.** Effects of physical disturbance and habitat-modifying species on sediment properties and benthic communities in the northern Baltic Sea. Tartu, 2010, 123 p.

182. **Arto Pulk.** Studies on bacterial ribosomes by chemical modification approaches. Tartu, 2010, 161 p.
183. **Maria Põllupüü.** Ecological relations of cladocerans in a brackish-water ecosystem. Tartu, 2010, 126 p.
184. **Toomas Silla.** Study of the segregation mechanism of the Bovine Papillomavirus Type 1. Tartu, 2010, 188 p.
185. **Gyaneshwer Chaubey.** The demographic history of India: A perspective based on genetic evidence. Tartu, 2010, 184 p.
186. **Katrin Kepp.** Genes involved in cardiovascular traits: detection of genetic variation in Estonian and Czech populations. Tartu, 2010, 164 p.
187. **Virve Sõber.** The role of biotic interactions in plant reproductive performance. Tartu, 2010, 92 p.
188. **Kersti Kangro.** The response of phytoplankton community to the changes in nutrient loading. Tartu, 2010, 144 p.
189. **Joachim M. Gerhold.** Replication and Recombination of mitochondrial DNA in Yeast. Tartu, 2010, 120 p.
190. **Helen Tammert.** Ecological role of physiological and phylogenetic diversity in aquatic bacterial communities. Tartu, 2010, 140 p.
191. **Elle Rajandu.** Factors determining plant and lichen species diversity and composition in Estonian *Calamagrostis* and *Hepatica* site type forests. Tartu, 2010, 123 p.
192. **Paula Ann Kivistik.** ColR-ColS signalling system and transposition of Tn*4652* in the adaptation of *Pseudomonas putida.* Tartu, 2010, 118 p.
193. **Siim Sõber.** Blood pressure genetics: from candidate genes to genome-wide association studies. Tartu, 2011, 120 p.
194. **Kalle Kipper.** Studies on the role of helix 69 of 23S rRNA in the factor-dependent stages of translation initiation, elongation, and termination. Tartu, 2011, 178 p.
195. **Triinu Siibak.** Effect of antibiotics on ribosome assembly is indirect. Tartu, 2011, 134 p.
196. **Tambet Tõnissoo.** Identification and molecular analysis of the role of guanine nucleotide exchange factor RIC-8 in mouse development and neural function. Tartu, 2011, 110 p.
197. **Helin Räägel.** Multiple faces of cell-penetrating peptides – their intra-cellular trafficking, stability and endosomal escape during protein transduction. Tartu, 2011, 161 p.
198. **Andres Jaanus.** Phytoplankton in Estonian coastal waters – variability, trends and response to environmental pressures. Tartu, 2011, 157 p.
199. **Tiit Nikopensius.** Genetic predisposition to nonsyndromic orofacial clefts. Tartu, 2011, 152 p.
200. **Signe Värv.** Studies on the mechanisms of RNA polymerase II-dependent transcription elongation. Tartu, 2011, 108 p.
201. **Kristjan Välk.** Gene expression profiling and genome-wide association studies of non-small cell lung cancer. Tartu, 2011, 98 p.

202. **Arno Põllumäe.** Spatio-temporal patterns of native and invasive zooplankton species under changing climate and eutrophication conditions. Tartu, 2011, 153 p.
203. **Egle Tammeleht.** Brown bear (*Ursus arctos*) population structure, demographic processes and variations in diet in northern Eurasia. Tartu, 2011, 143 p.
205. **Teele Jairus.** Species composition and host preference among ectomycorrhizal fungi in Australian and African ecosystems. Tartu, 2011, 106 p.
206. **Kessy Abarenkov.** PlutoF – cloud database and computing services supporting biological research. Tartu, 2011, 125 p.
207. **Marina Grigorova.** Fine-scale genetic variation of follicle-stimulating hormone beta-subunit coding gene (*FSHB*) and its association with reproductive health. Tartu, 2011, 184 p.
208. **Anu Tiitsaar.** The effects of predation risk and habitat history on butterfly communities. Tartu, 2011, 97 p.
209. **Elin Sild.** Oxidative defences in immunoecological context: validation and application of assays for nitric oxide production and oxidative burst in a wild passerine. Tartu, 2011, 105 p.
210. **Irja Saar**. The taxonomy and phylogeny of the genera *Cystoderma* and *Cystodermella* (Agaricales, Fungi). Tartu, 2012, 167 p.
211. **Pauli Saag.** Natural variation in plumage bacterial assemblages in two wild breeding passerines. Tartu, 2012, 113 p.
212. **Aleksei Lulla.** Alphaviral nonstructural protease and its polyprotein substrate: arrangements for the perfect marriage. Tartu, 2012, 143 p.
213. **Mari Järve.** Different genetic perspectives on human history in Europe and the Caucasus: the stories told by uniparental and autosomal markers. Tartu, 2012, 119 p.
214. **Ott Scheler**. The application of tmRNA as a marker molecule in bacterial diagnostics using microarray and biosensor technology. Tartu, 2012, 93 p.
215. **Anna Balikova**. Studies on the functions of tumor-associated mucin-like leukosialin (CD43) in human cancer cells. Tartu, 2012, 129 p.
216. **Triinu Kõressaar.** Improvement of PCR primer design for detection of prokaryotic species. Tartu, 2012, 83 p.
217. **Tuul Sepp.** Hematological health state indices of greenfinches: sources of individual variation and responses to immune system manipulation. Tartu, 2012, 117 p.
218. **Rya Ero.** Modifier view of the bacterial ribosome. Tartu, 2012, 146 p.
219. **Mohammad Bahram.** Biogeography of ectomycorrhizal fungi across different spatial scales. Tartu, 2012, 165 p.
220. **Annely Lorents.** Overcoming the plasma membrane barrier: uptake of amphipathic cell-penetrating peptides induces influx of calcium ions and downstream responses. Tartu, 2012, 113 p.

221. **Katrin Männik.** Exploring the genomics of cognitive impairment: whole-genome SNP genotyping experience in Estonian patients and general population. Tartu, 2012, 171 p.
222. **Marko Prous.** Taxonomy and phylogeny of the sawfly genus *Empria* (Hymenoptera, Tenthredinidae). Tartu, 2012, 192 p.
223. **Triinu Visnapuu.** Levansucrases encoded in the genome of *Pseudomonas syringae* pv. tomato DC3000: heterologous expression, biochemical characterization, mutational analysis and spectrum of polymerization products. Tartu, 2012, 160 p.
224. **Nele Tamberg.** Studies on Semliki Forest virus replication and pathogenesis. Tartu, 2012, 109 p.
225. **Tõnu Esko.** Novel applications of SNP array data in the analysis of the genetic structure of Europeans and in genetic association studies. Tartu, 2012, 149 p.
226. **Timo Arula.** Ecology of early life-history stages of herring *Clupea harengus membras* in the northeastern Baltic Sea. Tartu, 2012, 143 p.
227. **Inga Hiiesalu.** Belowground plant diversity and coexistence patterns in grassland ecosystems. Tartu, 2012, 130 p.
228. **Kadri Koorem.** The influence of abiotic and biotic factors on small-scale plant community patterns and regeneration in boreonemoral forest. Tartu, 2012, 114 p.
229. **Liis Andresen.** Regulation of virulence in plant-pathogenic pectobacteria. Tartu, 2012, 122 p.
230. **Kaupo Kohv.** The direct and indirect effects of management on boreal forest structure and field layer vegetation. Tartu, 2012, 124 p.
231. **Mart Jüssi.** Living on an edge: landlocked seals in changing climate. Tartu, 2012, 114 p.
232. **Riina Klais.** Phytoplankton trends in the Baltic Sea. Tartu, 2012, 136 p.
233. **Rauno Veeroja.** Effects of winter weather, population density and timing of reproduction on life-history traits and population dynamics of moose (*Alces alces*) in Estonia. Tartu, 2012, 92 p.
234. **Marju Keis.** Brown bear (*Ursus arctos*) phylogeography in northern Eurasia. Tartu, 2013, 142 p.
235. **Sergei Põlme.** Biogeography and ecology of *alnus*- associated ectomycorrhizal fungi – from regional to global scale. Tartu, 2013, 90 p.
236. **Liis Uusküla.** Placental gene expression in normal and complicated pregnancy. Tartu, 2013, 173 p.
237. **Marko Lõoke.** Studies on DNA replication initiation in *Saccharomyces cerevisiae.* Tartu, 2013, 112 p.
238. **Anne Aan.** Light- and nitrogen-use and biomass allocation along productivity gradients in multilayer plant communities. Tartu, 2013, 127 p.
239. **Heidi Tamm.** Comprehending phylogenetic diversity – case studies in three groups of ascomycetes. Tartu, 2013, 136 p.

240. **Liina Kangur.** High-Pressure Spectroscopy Study of Chromophore-Binding Hydrogen Bonds in Light-Harvesting Complexes of Photosynthetic Bacteria. Tartu, 2013, 150 p.
241. **Margus Leppik.** Substrate specificity of the multisite specific pseudouridine synthase RluD. Tartu, 2013, 111 p.
242. **Lauris Kaplinski.** The application of oligonucleotide hybridization model for PCR and microarray optimization. Tartu, 2013, 103 p.
243. **Merli Pärnoja.** Patterns of macrophyte distribution and productivity in coastal ecosystems: effect of abiotic and biotic forcing. Tartu, 2013, 155 p.
244. **Tõnu Margus.** Distribution and phylogeny of the bacterial translational GTPases and the Mqsr/YgiT regulatory system. Tartu, 2013, 126 p.
245. **Pille Mänd**. Light use capacity and carbon and nitrogen budget of plants: remote assessment and physiological determinants. Tartu, 2013, 128 p.
246. **Mario Plaas.** Animal model of Wolfram Syndrome in mice: behavioural, biochemical and psychopharmacological characterization. Tartu, 2013, 144 p.
247. **Georgi Hudjašov.** Maps of mitochondrial DNA, Y-chromosome and tyrosinase variation in Eurasian and Oceanian populations. Tartu, 2013, 115 p.
248. **Mari Lepik**. Plasticity to light in herbaceous plants and its importance for community structure and diversity. Tartu, 2013, 102 p.
249. **Ede Leppik**. Diversity of lichens in semi-natural habitats of Estonia. Tartu, 2013, 151 p.
250. **Ülle Saks.** Arbuscular mycorrhizal fungal diversity patterns in boreonemoral forest ecosystems. Tartu, 2013, 151 p.
251. **Eneli Oitmaa**. Development of arrayed primer extension microarray assays for molecular diagnostic applications. Tartu, 2013, 147 p.
252. **Jekaterina Jutkina.** The horizontal gene pool for aromatics degradation: bacterial catabolic plasmids of the Baltic Sea aquatic system. Tartu, 2013, 121 p.
253. **Helen Vellau.** Reaction norms for size and age at maturity in insects: rules and exceptions. Tartu, 2014, 132 p.
254. **Randel Kreitsberg.** Using biomarkers in assessment of environmental contamination in fish – new perspectives. Tartu, 2014, 107 p.
255. **Krista Takkis.** Changes in plant species richness and population performance in response to habitat loss and fragmentation.Tartu, 2014, 141 p.
256. **Liina Nagirnaja.** Global and fine-scale genetic determinants of recurrent pregnancy loss. Tartu, 2014, 211 p.
257. **Triin Triisberg.** Factors influencing the re-vegetation of abandoned extracted peatlands in Estonia. Tartu, 2014, 133 p.
258. **Villu Soon.** A phylogenetic revision of the *Chrysis ignita* species group (Hymenoptera: Chrysididae) with emphasis on the northern European fauna. Tartu, 2014, 211 p.

259. **Andrei Nikonov.** RNA-Dependent RNA Polymerase Activity as a Basis for the Detection of Positive-Strand RNA Viruses by Vertebrate Host Cells. Tartu, 2014, 207 p.
260. **Eele Õunapuu-Pikas**. Spatio-temporal variability of leaf hydraulic conductance in woody plants: ecophysiological consequences. Tartu, 2014, 135 p.
261. **Marju Männiste**. Physiological ecology of greenfinches: information content of feathers in relation to immune function and behavior. Tartu, 2014, 121 p.
262. **Katre Kets.** Effects of elevated concentrations of CO_2 and O_3 on leaf photosynthetic parameters in *Populus tremuloides*: diurnal, seasonal and interannual patterns. Tartu, 2014, 115 p.
263. **Külli Lokko.** Seasonal and spatial variability of zoopsammon communities in relation to environmental parameters. Tartu, 2014, 129 p.
264. **Olga Žilina**. Chromosomal microarray analysis as diagnostic tool: Estonian experience. Tartu, 2014, 152 p.
265. **Kertu Lõhmus**. Colonisation ecology of forest-dwelling vascular plants and the conservation value of rural manor parks. Tartu, 2014, 111 p.
266. **Anu Aun.** Mitochondria as integral modulators of cellular signaling. Tartu, 2014, 167 p.
267. **Chandana Basu Mallick.** Genetics of adaptive traits and gender-specific demographic processes in South Asian populations. Tartu, 2014, 160 p.
268. **Riin Tamme.** The relationship between small-scale environmental heterogeneity and plant species diversity. Tartu, 2014, 130 p.
269. **Liina Remm.** Impacts of forest drainage on biodiversity and habitat quality: implications for sustainable management and conservation. Tartu, 2015, 126 p.
270. **Tiina Talve.** Genetic diversity and taxonomy within the genus *Rhinanthus.* Tartu, 2015, 106 p.
271. **Mehis Rohtla**. Otolith sclerochronological studies on migrations, spawning habitat preferences and age of freshwater fishes inhabiting the Baltic Sea. Tartu, 2015, 137 p.
272. **Alexey Reshchikov.** The world fauna of the genus *Lathrolestes* (Hymenoptera, Ichneumonidae). Tartu, 2015, 247 p.
273. **Martin Pook.** Studies on artificial and extracellular matrix protein-rich surfaces as regulators of cell growth and differentiation. Tartu, 2015, 142 p.
274. **Mai Kukumägi.** Factors affecting soil respiration and its components in silver birch and Norway spruce stands. Tartu, 2015, 155 p.
275. **Helen Karu.** Development of ecosystems under human activity in the North-East Estonian industrial region: forests on post-mining sites and bogs. Tartu, 2015, 152 p.
276. **Hedi Peterson.** Exploiting high-throughput data for establishing relationships between genes. Tartu, 2015, 186 p.

277. **Priit Adler.** Analysis and visualisation of large scale microarray data, Tartu, 2015, 126 p.
278. **Aigar Niglas.** Effects of environmental factors on gas exchange in deciduous trees: focus on photosynthetic water-use efficiency. Tartu, 2015, 152 p.
279. **Silja Laht.** Classification and identification of conopeptides using profile hidden Markov models and position-specific scoring matrices. Tartu, 2015, 100 p.
280. **Martin Kesler.** Biological characteristics and restoration of Atlantic salmon *Salmo salar* populations in the Rivers of Northern Estonia. Tartu, 2015, 97 p.
281. **Pratyush Kumar Das.** Biochemical perspective on alphaviral nonstructural protein 2: a tale from multiple domains to enzymatic profiling. Tartu, 2015, 205 p