DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
221

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
221

KATRIN MÄNNIK

Exploring the genomics of cognitive impairment: whole-genome SNP genotyping experience in Estonian patients and general population

TARTU UNIVERSITY PRESS

This study was carried out at the Institute of Molecular and Cell Biology, University of Tartu, Estonia, and at the Center for Integrative Genomics, University of Lausanne, Switzerland.

The Dissertation is accepted for the commencement of the degree of Doctor of Philosophy in gene technology on 20.06.2012 by the Council of the Institute of Molecular and Cell Biology, University of Tartu.

Supervisor: Prof. Ants Kurg, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

Opponent: PD Mag. Dr. Reinhard Ullmann, Max Planck Institute for Molecular Genetics, Berlin, Germany

Commencement: Room No 105, Institute of Molecular and Cell Biology, University of Tartu; 23b Riia St, Tartu, on September 14th, 2012, at 10:00.

Publication of this dissertation is granted by the Institute of Molecular and Cell Biology, University of Tartu and by the Graduate School in Biomedicine and Biotechnology created under the auspices of European Social Fund.

ISSN 1024–6479
ISBN 978–9949–32–085–1 (trükis)
ISBN 978–9949–32–086–8 (PDF)

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr 396

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

I. **Männik K**, Parkel S, Palta P, Žilina 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 the general population in Estonia" **Eur J Med Genet 2011** Mar–Apr;54(2):136–43

II. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, Martinet D, Shen Y, Valsesia A, Beckmann ND, Thorleifsson G, Belfiore M, Bouquillon S, Campion D, de Leeuw N, de Vries BB, Esko T, Fernandez BA, Fernández-Aranda F, Fernández-Real JM, Gratacòs M, Guilmatre A, Hoyer J, Jarvelin MR, Kooy RF, Kurg A, Le Caignec C, **Männik K**, Platt OS, Sanlaville D, Van Haelst MM, Villatoro Gomez S, Walha F, Wu BL, Yu Y, Aboura A, Addor MC, Alembik Y, Antonarakis SE, Arveiler B, Barth M, Bednarek N, Béna F, Bergmann S, Beri M, Bernardini L, Blaumeiser B, Bonneau D, Bottani A, Boute O, Brunner HG, Cailley D, Callier P, Chiesa J, Chrast J, Coin L, Coutton C, Cuisset JM, Cuvellier JC, David A, de Freminville B, Delobel B, Delrue MA, Demeer B, Descamps D, Didelot G, Dieterich K, Disciglio V, Doco-Fenzy M, Drunat S, Duban-Bedu B, Dubourg C, El-Sayed Moustafa JS, Elliott P, Faas BH, Faivre L, Faudet A, Fellmann F, Ferrarini A, Fisher R, Flori E, Forer L, Gaillard D, Gerard M, Gieger C, Gimelli S, Gimelli G, Grabe HJ, Guichet A, Guillin O, Hartikainen AL, Heron D, Hippolyte L, Holder M, Homuth G, Isidor B, Jaillard S, Jaros Z, Jiménez-Murcia S, Helas GJ, Jonveaux P, Kaksonen S, Keren B, Kloss-Brandstätter A, Knoers NV, Koolen DA, Kroisel PM, Kronenberg F, Labalme A, Landais E, Lapi E, Layet V, Legallic S, Leheup B, Leube B, Lewis S, Lucas J, MacDermot KD, Magnusson P, Marshall C, Mathieu-Dramard M, McCarthy MI, Meitinger T, Mencarelli MA, Merla G, Moerman A, Mooser V, Morice-Picard F, Mucciolo M, Nauck M, Ndiaye NC, Nordgren A, Pasquier L, Petit F, Pfundt R, Plessis G, Rajcan-Separovic E, Ramelli GP, Rauch A, Ravazzolo R, Reis A, Renieri A, Richart C, Ried JS, Rieubland C, Roberts W, Roetzer KM, Rooryck C, Rossi M, Saemundsen E, Satre V, Schurmann C, Sigurdsson E, Stavropoulos DJ, Stefansson H, Tengström C, Thorsteinsdóttir U, Tinahones FJ, Touraine R, Vallée L, van Binsbergen E, Van der Aa N, Vincent-Delorme C, Visvikis-Siest S, Vollenweider P, Völzke H, Vulto-van Silfhout AT, Waeber G, Wallgren-Pettersson C, Witwicki RM, Zwolinksi S, Andrieux J, Estivill X, Gusella JF, Gustafsson O, Metspalu A, Scherer SW, Stefansson K, Blakemore AI, Beckmann JS, Froguel P "Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus" **Nature 2011** Aug 31;478(7367):97–102

III. 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, Blakemore 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

IV. Van der Aa N, Rooms L, Vandeweyer G, van den Ende J, Reyniers E, Fichera M, Romano C, Delle Chiaie B, Mortier G, Menten B, Destrée A, Maystadt I, **Männik K**, Kurg A, Reimand T, McMullan D, Oley C, Brueton L, Bongers EM, van Bon BW, Pfund R, Jacquemont S, Ferrarini A, Martinet D, Schrander-Stumpel C, Stegmann AP, Frints SG, de Vries BB, Ceulemans B, Kooy RF "Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome" **Eur J Med Genet 2009** Mar-Jun;52(2–3):94–100

This thesis is based on the original publications referred by Roman numerals I – IV, and previously unpublished data. My contribution to the listed original articles is following:

Publication I: Design of the study; performing of the experiments; data analysis; writing of the manuscript.

Publication II: Recruitment and analysis of Estonian samples; coordinator and leading author from Estonia.

Publication III: Recruitment and analysis of Estonian samples; coordinator and leading author from Estonia.

Publication IV: Recruitment and analysis of Estonian samples; coordinator and leading author from Estonia.

ABBREVIATIONS

ADHD	attention-deficit hyperactivity disorder
ASD	autism spectrum disorder
BAF	B allele frequency
BMI	body mass index
BP	rearrangement break-point
CA	congenital anomalies
CHARGE	coloboma, heart anomaly, choanal atresia, retardation, genital, and ear anomalies syndrome
CNV	DNA copy-number variation
DD	developmental delay
DECIPHER	Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources
DGV	Database of Genomic Variants
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders, 4th Edition
ECARUCA	European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations
EGC UT	Estonian Genome Centre, at University of Tartu, Estonian general population based biobank
EGC	cohort of Estonian general population individuals
EID	Estonian patients with idiopathic intellectual disability
FISH	fluorescence *in situ* hybridization
FoSTeS	fork stalling and template switching mechanism
GWAS	genome-wide association study
ID	intellectual disability
IEG	immediate early gene
Indel	insertions and deletions smaller than 1 kb in size
IQ	intelligence quotient
ISCA	International Standards for Cytogenomic Arrays Consortium
kb	kilo base pairs i.e. thousand base pairs
LCL	lymphoblastoid cell line
LCR	low copy repeat sequence
LRR	log R ratio
NAHR	non-allelic homologous recombination
NCBI	National Center for Biotechnology Information
NHEJ	non-homologous end joining mechanism
MAF	minor allele frequency
Mb	mega base pairs i.e. million base pairs
MCA	multiple congenital anomalies
MMBIR	microhomology-mediated break-induced replication
MRI	magnetic resonance imaging
OFC	occipitofrontal circumference

OMIM	Online Mendelian Inheritance in Man database
p	short arm of human chromosome
PAR1	pseudoautosomal region, homologous sequence on the X and Y chromosomes
PCR	polymerase chain reaction
q	long arm of human chromosome
qPCR	quantitative PCR
RT-qPCR	reverse transcription qPCR
SD	standard deviation
SNP	single nucleotide polymorphism
TAR	thrombocytopenia-absent radius syndrome
UPD	uniparental disomy
WHO	World Health Organization
WHO ICD-10	WHO International Statistical Classification of Diseases and Related Health Problems, 10th Revision
XLID	chromosome X-linked intellectual disability
XLMR	chromosome X-linked mental retardation

Symbols of genes that are used within the text, are not listed in the above list. For genes and syndromes OMIM reference numbers are provided in the text.

1. INTRODUCTION

A recent technological revolution in human genomics has enabled a conceptual shift in approaches to study the genetic backround of diseases, leading from single gene specific to a genome-wide emphasis. The use of high-resolution genomic microarrays has revealed widespread presence of DNA copy number variations (CNVs) in the human genome. CNVs are defined as segments of DNA ranging in size from thousands to millions of base pairs and altered in dosage of genomic copies compared to the reference genome. Depending on the genomic context, these variants can be harmless polymorphisms, act as susceptibility factors for common traits and diseases or play an important role in the pathogenesis of developmental disorders and congenital anomalies. Chromosomal imbalances contribute as major players in neuropsychiatric disorders and several distinctive microdeletion and microduplication syndromes have been defined during recent years. However, studies have revealed that numerous variants initially detected in patients with brain-related disorders also occur with lower frequency in apparently normal individuals. Assessing the clinical significance of these CNVs, and thus providing proper genetic counselling, is further challenged by intra-individual diversity within patient cohorts. Although investigation of inheritance patterns may offer additional information, it is often difficult to attribute pathogenic significance based on whether the CNV was inherited from a parent or occurs as a *de novo* event only. Therefore, characterizing rare genomic variants using a family-based approach, as well as cohort-exceeding strategies is essential for reliable assessment of the phenotypic consequences.

In this study, single nucleotide polymorphism (SNP) based whole-genome screening was used to investigate genomic variants in Estonian families with idiopathic intellectual disability. In addition, genotype and phenotype data from Estonian general population individuals was used for accurate interpretation of rare structural aberrations of uncertain relevance. This study is the first comprehensive effort to investigate genomic causes of cognitive impairment and offer state-of-the-art diagnostic possibilities in Estonian patients. It presents the benefits and opportunities provided by well-characterized comparative cohorts and SNP genotyping data in the diagnostics of developmental disorders and complex traits.

2. REVIEW OF THE LITERATURE

2.1. Structural variants and their consequence on human health

2.1.1. DNA copy-number variants as a cause of genetic variability and human disease

The extensive use of genomic profiling by array-based platforms and next-generation sequencing over the last years has extended our understanding of the genetic diversity of the human genome, and revealed DNA copy-number variation (CNV) as an essential contributor to inter-individual variability and a major driving force in human evolution [1–9]. Copy-number variants are a form of structural variation, defined as genomic segments >1 kb in size that vary in their number of genomic copies compared to the representative reference genome. These stretches of altered copy-number DNA sometimes encompass scores of protein coding genes or regulatory elements. Depending on the genomic content, they can be harmless polymorphisms or have an impact on a carrier's risk to develop a disease. CNV formation can rise meiotically as well as somatically, and accumulating data demonstrates that variations in different tissues contribute also to the individual's somatic mosaicism [10–13]. This supports the hypothesis that age-accumulated CNVs might have a role also in diseases that develop symptoms later in life [14, 15]. Although precise estimation of CNV mutation rates at the genome-wide level is still elusive and the rates have been expected to vary among loci by several orders of magnitude, different studies have found an average per-generation per-nucleotide rate of CNV formation in the range between 2×10^{-6} and 3×10^{-2} per-nucleotide per-haploid genome, i.e. several orders of magnitude higher than the single base substitution rate [2, 16–19]. Altogether, CNVs have been shown to occur in a high portion (approximately 35%) of the human genome, and to be common in normal population without major phenotypic effect (http://projects.tcag.ca/variation). However, about 14% of genes in the OMIM morbid map have been estimated to be subject to copy-number variation [5], and during the recent years CNVs have been identified as one of the common causes of human disease [20]. Pathologic conditions caused by these structural rearrangements are collectively termed as genomic disorders [21].

2.1.2. Landscape and prevalence of human copy-number variation

Structural variation in the human genome is not randomly distributed, but complex patterns of inverted and directly oriented low copy repeat sequences (LCRs; known also as segmental duplications) have made some chromosomal regions prone to the rearrangements. According to their occurrence, CNVs can be divided into two broad categories: (i) recurrent and non-recurrent; (ii) common and rare variants [22] (**Figure 1**).

Recurrent, non-recurrent and atypical CNVs

Recurrent CNVs are flanked by directly oriented blocks of highly homologous LCRs. Misalignment of these LCRs during meiosis acts as a substrate for a process called non-allelic homologous recombination (NAHR) and gives rise to CNVs with common size and nearly identical boundaries in carriers. Recurrent CNVs mediated by NAHR are usually larger in size and several of these have been linked to distinctive genomic syndromes [23, 24]. However, the majority of pathogenic microdeletions and microduplications (i.e. aberrations that are too small to be detected under the microscope) consists of rare non-recurrent CNVs scattered throughout the genome [25]. Non-recurrent CNVs can be generated by NAHR between repetitive SINE and LINE sequences, but occur more often via other molecular mechanisms such as non-homologous end joining (NHEJ) [22], or perturbations of DNA replication and repair (e.g. Fork Stalling and Template Switching (FoSTeS) [26] and microhomology-mediated break-induced replication (MMBIR) [27]). Although break-points of these CNVs do not cluster within segmental duplications, they still tend to occur in the vicinity of LCRs and to be stimulated by complex genome architecture [28, 29]. Non-recurrent CNVs are of variable size in patients, but carriers may share a critical region whose copy-number change results in common clinical features [25].

CNVs that overlap with recurrent disease regions but appear with break-points mediated either by different LCRs or a non-homologous mechanism are termed as atypical deletions and duplications. These imbalances (when shorter in size) might provide evidence for underlying monogenic factors or allow one to refine the critical interval of the recurrent syndrome [30–33].

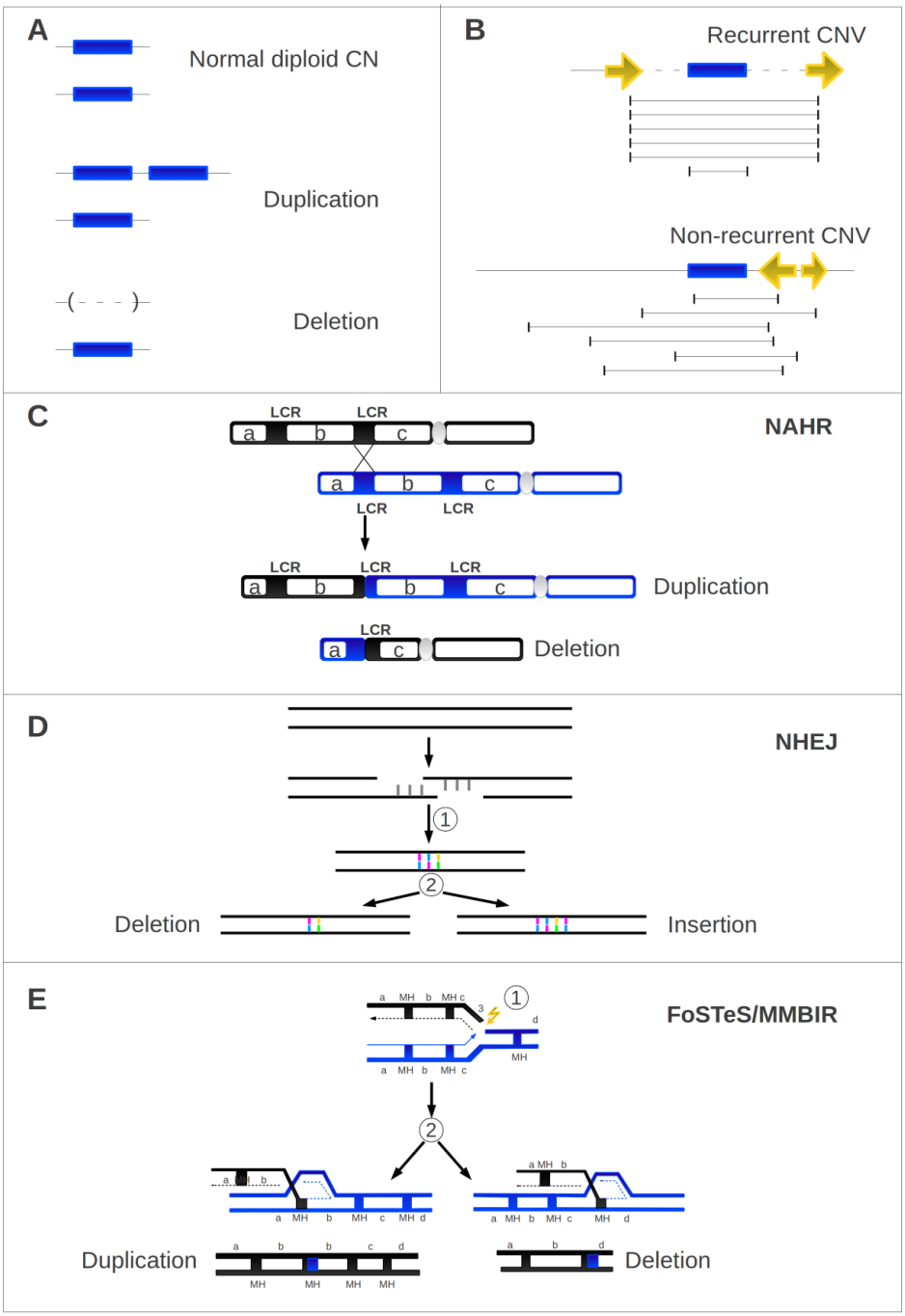

Figure 1. Schematic representation of copy-number variations in the human genome and mechanisms of their formation**.** CNV is defined as a segment of DNA that has decreased (deletion) or increased (duplication) number of genomic copies compared to

the reference genome (**A**). The recurrent deletions and duplications result mostly from NAHR, have common size and nearly identical breakpoints that cluster within the directly oriented LCRs. The CNVs with different break-points in the recurrent rearrangement region are called atypical. The majority of non-recurrent CNVs result from FoSTeS/MMBIR mechanisms of formation, they are variable in size and have scattered break-points (**B**). Non-allelic homologous recombination (NAHR) occurs when directly oriented highly identical LCRs lead to misalignment of alleles and result in unequal crossing-over mediated production of deletions or reciprocal duplications (**C**). Non-homologous end-joining (NHEJ) is initiated by a double-stranded DNA break (1), followed by bridging, processing and ligation of broken DNA ends (2). The products of NHEJ repair include deletions and insertions that often contain some additional nucleotides at the DNA end junction, called a "molecular scar" (**D**). FoSTeS/MMBIR (fork stalling and template switching/microhomology mediated break-induced replication) is a DNA repair mechanism that utilizes nucleotide microhomology (MH) at the rearrangement breakpoints. After stalling of the replication fork due to single strand break (1), the lagging strand disengages, anneals to another fork and replication starts at a different place by the 3' end invasion of lagging strand via regions of microhomology (2). Since serial FoSTeS cycles may occur, the mechanism plays especially important role in the formation of disease-associated nonrecurrent rearrangements with a complex structure (**E**). Adapted from [43, 44].

Common and rare CNVs

The overall population-specific allele frequency of CNVs has been shown to resemble that of the SNPs, with most variants having a low to rare frequency (minor allele frequency, MAF = 0.05–5% and MAF < 0.05%, respectively), while common CNVs (MAF ≥ 5%) account for the majority of the heterozygosity [34–36]. Although associations with complex diseases have been established for a few common CNVs [37–42], similar to SNP association studies, common variants collectively seem to make only a small contribution to the heritable disease risk [45]. This observation has challenged the popular „common disease – common variant" hypothesis and risen interest in rare genetic variants with strong effect on complex disease and traits [45–47]. The latter is supported by the knowledge that variants with clinical consequences segregate in the population at lower frequencies and most of the deleterious variants in humans have been held at low frequency by purifying selection [46, 48, 49]. Rare CNVs have already been defined as risk factors for several common disorders, including obesity [50–53], Alzheimer's disease [54], pancreatitis [55] and epilepsy [56–60]. However, the most remarkable finding has been the identification of rare CNVs with major clinical effect, which holds particularly true for developmental and neurobehavioural disorders [20, 30, 61–65]. Altogether, different studies have screened thousands of human genomes over the last years and conclude that a wide spectrum of disease-susceptibility variants exist, and that most of these are rare with a frequency below 0.1% and of variable and minuscule effect. Thus, due to the widespread presence of CNVs in the general population [1–5], the main challenge ahead is to assess

whether each particular CNV has any clinical significance. To date, only for a minor fraction of variants has their disease-causative role been determined. Furthermore, in the case of several novel genomic disorders, a broad spectrum of phenotypic consequences has been described and some CNVs initially considered pathogenic have also been observed in apparently normal individuals. To clarify the pathogenic importance of rare variants, large numbers of high-resolution studies of different clinical cohorts, as well as comparative analyses of the general population are necessary.

The review of the literature in the current thesis is focused only on rare unbalanced structural variants in the human genome and their effect on the health. The potential phenotypic impact of common genomic variants was out of the scope of this study.

2.2. Intellectual disability as a frontline phenotype for studying the clinical impact of genomic variants

2.2.1. Definition of intellectual disability

Intellectual disability (ID), previously referred to as mental retardation, is according to the Diagnostic and Statistical Manual of Mental Disorders (DSM – IV) defined as a condition of significantly subaverage intellectual function with limitations in adaptive behaviour such as conceptual, practical, and social adaptive skills that must be diagnosed before the age of 18 years. ID is a clinically diverse condition with variable degrees of cognitive impairment and may exist in isolation (nonsyndromic ID) or to be accompanied by additional congenital anomalies and other clinical features (syndromic ID). The World Health Organization International Statistical Classification of Diseases and Related Health Problems 10th Revision (WHO ICD – 10) divides ID into four categories – (i) mild (approximate IQ range of 50 to 69, which in adults corresponds to mental age 9–12 years); (ii) moderate (IQ of 35 to 49, mental age 6–9 years); (iii) severe (IQ of 20 to 34, mental age 3–6 years); (iv) profound (IQ below 20, mental age under 3 years).

The prevalence of ID is estimated to be 1–3% of the general population in developed countries (The World Health Organization, 2002; [66]) which makes it a common cause of lifelong disability contributing to high socio-economic costs [67–69].

2.2.2. Cytogenetics, genomic rearrangements and intellectual disability

Aetiology of intellectual disability is extremely heterogeneous and the impairment in cognitive functioning can result from genetic, epigenetic as well as environmental causes, solely or in their interaction. In case of severe ID, which occurs in 0.3% of the world`s population, genetics is thought to play a role in

approximately half of cases [70]. Due to its burden on families and society, considerable effort has been invested in the identification of aetiological factors and understanding the molecular basis of human cognition. Amongst genetic causes, Down syndrome (Trisomy 21; OMIM #190685) has remained the most important single cause of ID despite widespread availability of prenatal diagnostics (reviewed in [71]). Other chromosomal aneuploidies and cytogenetically visible rearrangements together have been found to be causative in approximately 10–15% of ID patients [72–74], and hundreds of genes responsible for monogenic forms of ID have been mapped to date (http://www.ncbi.nlm.nih.gov/omim; http://xlmr.interfree.it/home.htm; http://www.lovd.nl). However, challenged by the extreme genetic and phenotypic heterogeneity, the underlying factors in about half of the individuals with ID have still remained unknown. Hindering genetic counselling of the families and clinical management of the patients, this has sustained intellectual disability as one of the most important problems to solve in health care [75, 76].

Significant progress regarding the genetic causes of cognitive impairment has been made during the past decade when technological advances made it affordable to investigate entire human genomes. The evolution of molecular and cytogenetic methods commonly used for identifying chromosomal rearrangements is summarized in **Table 1**.

Since 2003, when Vissers *et al*. first introduced the array-based application for genome-wide identification of submicroscopic imbalances in patients with ID [78], a variety of genomic arrays with constantly improving probe design and density to capture CNVs have been available for diagnostics and research. To date, tens of novel distinctive microdeletion and microduplication syndromes have been described, and numerous genomic regions have been linked with susceptibility to neuropsychiatric diseases. The small size of several CNVs has made systematic screening and molecular characterization of encompassed genes a successful approach also for the identification of disease genes. Moreover, the localization of several ID genes has been determined by the mutation analysis of potential candidates in cryptic aberrant intervals and the investigation of their breakpoints [33, 79–81].

Table 1. Overview of molecular and cytogenetic methods commonly used for identifying chromosomal rearrangements. Adapted and supplemented from [77].

Platform	Rearrangement detection			Resolution	Sensitivity	# of loci[1]
	Unbalanced	Balanced	UPD			
Chromosome-based methods (from early 1970 to 1990s)						
G-banded karyotyping	×	×	-	Low (>5Mb)	Low	High
Chromosomal CGH	×	-	-	Low(>3Mb)	High	High
FISH	×	×	-	High (<100kb)	High	Low
SKY/M-FISH	×	×	-	Low (several Mb)	High	High
Amplification based methods (from 1990s to 2000s)						
qPCR	×	-	-	High (~100bp)	Very high	Low
MLPA	×	-	-	High (~100bp)	High	Medium
Array-based methods (from early 2000s to 2010s)						
BAC array-CGH	×	-	-	Medium (>1Mb)	High	High
Tiling-path array-CGH	×	-	-	High (>50kb)	High	Very high
Oligonucl. array-CGH	×	-	-	High (up to kb)	High	Very high
SNP genotyping arrays	×	-	×	High (>5kb)	High	Very high
Ultra high throughput sequencing-based methods (from 2010s)						
UHT-sequencing	×	×	×	Very high (bp)	Very high	Very high

[1]Number of analyzed genomic loci per single assay; UPD – uniparental disomy; CGH – comparative genomic hybridization; FISH – fluorescence *in situ* hybridization; SKY – spectral karyotyping; M-FISH – multiplex FISH; qPCR – quantitative polymerase chain reaction; MLPA – multiplex ligation-dependent probe amplification; UHT – ultra high-throughput sequencing

2.2.3. Research and diagnostics of neurodevelopmental disorders in the era of genomics

More than hundred studies that have applied genomic arrays in different genetic centres for CNV profiling of individuals with unexplained ID have been referred in the PubMed database. The average CNV burden based on experimental and literature surveys has been estimated to be 10–15% of idiopathic ID patients [30, 74, 82–86]. Two comprehensive studies by Cooper and Kaminsky [20, 30] used high-resolution case-control data to investigate the role of rare CNVs in a large sample size of paediatric neurological diseases, and consistently confirmed significant enrichment of large CNVs among patients. This excess was evident at the 250 kb level and became more pronounced with increased size of the aberration. In addition, a strong correlation between *de novo* rate and increased CNV size was observed, with 50% of events at 1Mb reported as being inherited [20, 30]. The study also confirmed an elevated significance of CNVs in severe phenotypes associated with other congenital anomalies, especially craniofacial and cardiovascular defects [30]. Different phenotypic features, frequently accompanied by cognitive impairment, have required more widespread analytic approaches, and have made investigation of ID patients a nearly comprehensive showcase of strategies for genome-wide discovery of disease-related genetic factors. For now, genomic arrays are also successfully utilized for testing of patients with other brain related diseases (e.g. autism, epilepsy, schizophrenia) [62, 87, 88], congenital heart defects [89, 90] and other complex phenotypes. General analysis pipelines to shed light on the aetiology of neurodevelopmental phenotypes have also facilitated the genetic diagnosis of autism, epilepsy and behavioral problems, which frequently co-exist in intellectual disability patients and have now been shown to have alterations in the same genes or related pathways.

The ability to detect genetic variants with high diagnostic yield, proved in scientific research, has made genomic microarrays attractive also for routine clinical diagnostics. After evaluating technical advantages and limitations, the International Standard Cytogenomic Array (ISCA) consortium (https://www.iscaconsortium.org), which unites clinical and molecular cytogenetic laboratories worldwide, has strongly suggested high-density array-based analysis as the first-tier diagnostic test for patients with intellectual disability, autism spectrum disorders and multiple congenital anomalies [86]. This is already a case in several countries, including the Netherlands, Belgium, and as of 2011, Estonia. Biotechnology companies have launched standardized cytogenetic array formats and complementary software packages designed to target diagnostic needs and facilitate data interpretation, e.g. CytoSure^{TM} ISCA Arrays by Oxford Gene Technology (http://www.ogt.co.uk) or HumanCytoSNP BeadChip by Illumina Inc. (http://www.illumina.com). Thus in the so called (post-)genomics era, throughput and technical sensitivity of rare small variants determination is no longer a bottleneck. Instead, uniform validation requirements enabling low false positive and negative rates and guidelines for data

processing have been raised during the last years. Also several ethical questions have been raised related to consent when performing analyses that provide extensive genetic information unrelated to the disorder being tested, and which might reveal unforseen risk factors, medical and psychological consequences for patients and their families. Regarding clinical utility, consensus workflows have been suggested [85, 86, 91] which would help clinicians handle practical challenges in interpreting genomic reports containing many variants of unknown diagnostic relevance [92]. General criteria for assessing the phenotypic relevance of individual CNVs, adapted from the consensus report by Miller *et al*. [86], are provided in **Table 2**. However, distinguishing variants of pathogenic relevance from functionally neutral ones and understanding the true phenotypic effect requires large and diverse cohorts to be studied [93].

Table 2. Consensus criteria for assessing pathogenicity of a CNV in clinical testing of patients with unexplained ID by genomic arrays. Each criteria indicates respectively whether the impact of CNV is probably pathogenic or neutral. Adapted from [86].

Major criteria		**Pathogenic**	**Neutral**
1.	Identical CNV inherited from a healthy parent[a]		×
	Expanded or altered CNV inherited from a parent	×	
	Identical CNV inherited from an affected parent	×	
2.	CNV has been identified in one or more heaelthy members of the family		×
	CNV has been identified in affected member(s) of the family	×	
3.	CNV overlaps entirely with an imbalance detected by a high-resolution technology in reference populations or in a database of healthy individuals		×
	CNV overlaps with an imbalance detected by a high-resolution technology in a CNV database for patients with ID or other congenital anomalies	×	
4.	CNV overlaps with a known deletion or duplication syndrome region	×	
5.	CNV encompasses morbid OMIM genes[b]	×	
6.	CNV is gene rich	×	
	CNV is gene poor		×
Minor criteria[c]		**Pathogenic**	**Neutral**
1.	CNV is a homo- or heterozygous deletion	×	
	CNV is a duplication (that does not encompass any known dosage-sensitive genes)		×
	CNV is an amplification (gain of more than one genomic copy)	×	
2.	CNV is devoid of known regulatory elements		×

[a]A deletion inherited from unaffected parent could unmask recessive pathogenic point mutation on the trans allele inherited from the other parent. [b]CNV should produce the same type of mutation that is causative for OMIM disease and the produced phenotype should be that expected for the OMIM disease. [c]Exceptions to each criteria have been demonstrated.

More recent than the revolution of array-based technologies, large-scale sequencing of X-chromosome coding exons in mental retardation patients, used to discover rare disease-causing sequence variants by Tarpey and colleagues, exposed nine XLMR-associated genes and highlighted the analytical benefits and challenges of large-scale sequencing of rare variants [94]. This work was a pioneering effort prior to a wave of whole-exome deep-sequencing studies which have identified several novel genes harboring mutations responsible for ID syndromes [95–99], and have greatly impacted the speed of new disease gene mapping and revised the clinical diagnosis of rare diseases in general (reviewed in [92, 100, 101]).

2.2.4. *De novo* mutations and the genetic heterogeneity of intellectual diability

Nevertheless, the aetiology of cognitive impairment has remained unsolved in a significant fraction of patients and accumulative evidence favors very rare or even unique short-lived mutations to explain the aetiology of ID, instead of major „blockbuster“ factors. To explain this extreme genetic heterogeneity, Vissers *et al*. [97] tested in families of patients with sporadic unexplained ID the so-called „*de novo* mutation“ hypothesis, clarifying paradox of the wide-spread presence of neurodevelopmental disorders despite the fact that severely reduced fitness and fertility of the patients should lead to the „genetic lethality“ of the mutations responsible for such condition. The authors found most likely pathogenic *de novo* variants (all in different genes) for as much as seven out of ten investigated patients in their study. This findings strongly support the hypothesis that high rate of novel spontaneous mutations might „compensate“ strong negative selection and keep the rare variants associated with neuro-psychiatric diseases in the genetic pool. The result suggest that the majority of sporadic ID cases in the population could indeed be explained by *de novo* CNVs and single-nucleotide variants of strong effect [97]. As further elaborated by Prof. James Lupski, these new mutations which influence the development and function of the central nervous system could be the price we have to pay, as a species, for better adaption of our brain to the constantly changing environment [102].

General understanding of the molecular causes of cognitive impairment is however far from complete. New genetic factors are identified regularly, and whether there is a diversity of mechanisms behind these or they are converging into a limited number of common pathways, is not yet clear.

2.3. The effectiveness of the "genotyping first" approach in revealing novel genomic syndromes

Traditionally the determination of cytogenetic bases of genomic disorders has been dependent on the obtention of patients with established clinical phenotype, and the characterization of new syndromes has required finding the key features consistently appearing in collections of individuals with similar abnormalities. Because of the relative rarity of genomic disorders it has been difficult to draw reliable conclusions about patterns of concurrent clinical traits, and the cytogenetic causes of these syndromes has often remained unknown [103]. The means by which novel syndromes are identified have been completely altered by the cost-effective analysis of entire genomes. The growing availability of large genotyping data-sets have made it possible to use an opposite approach called „reverse phenotyping" or „genotyping first". In this case, patients are first discriminated by identical (or overlapping) genomic imbalances, and as sufficient numbers of patients are collected, characteristic features of a clinical entity can be delineated. While phenotypes of an individual are inherently dynamic, vulnerable to masking by other factors and difficult to evaluate objectively, genotypes are relatively straightforward to determine and stay stable over an individual's lifetime. Even when phenotypic features are reliably established, the underlying genetic backround is often not homogeneous, since multiple genes and alterations may contribute to the same pathway and therby to a similar final phenotype [104]. „Reverse phenotyping" has considerably accelerated the pace of identifying novel syndromic imbalances in patients with ID and accompanying (often apparently nonspecific) features. In the last five years, nearly 20 new recurrent CNV-causative clinical syndromes have been defined (for comprehensive review see for example [83, 105]).

2.3.1. Monosomy 17q21.31 exemplifies how a distinctive intellectual disability syndrome can be identified by large-scale genome screening

Monosomy 17q21.31 (OMIM #610443) is amongst examples of novel ID-associated diseases where initial identification of the microdeletions has led to a consistent and well recognizable clinical entity. This recurrent microdeletion syndrome was one of the first genomic disorders identified by microarrays in 2006 when initial patients with recurrent approximately 500 kb heterozygous deletions in 17q21.31 and distinctive clinical presentations were reported by three groups [106–108]. The subsequent characterization of the syndrome in larger cohorts has shown the early presence of hypotonia with poor feeding, epilepsy, global developmental delay of variable degree accompanied by an amiable and cooperative disposition, and a facial phenotype including as common features abnormal hair colour, a long face with a high broad forehead, upwards slanting palpebral fissures, ptosis, large ears and a tubular nose with

bulbous nasal tip [109–111]. The estimated population prevalence of around 1/16,000 by Koolen *et al*. indicates that monosomy 17q21.31 has been highly underdiagnosed and is one of the most common new ID syndromes which could count for 0.64% of unexplained patients [110]. The canonical deletion is mediated by NAHR between directly oriented LCRs and encompasses at least 6 genes. Typically to contiguous gene syndrome it is currently not known whether haploinsufficiency of one or several genes might underly clinical features. Amongst potential candidates, the regulator of chromatin modification *KANSL1* (OMIM #612452) has recently been identified as causative for the core phenotype [33, 81]. Also the microtubule-associated protein tau gene *MAPT* (OMIM #157140) has been of particular interest because of it's high expression in the brain and involvement in the aetiology of several neurodegenerative diseases [112]. This locus in 17q21.31 chromosome region is one of the most complex and evolutionarily dynamic intervals in the genome. It harbors a common 900 kb inversion polymorphism that can occur as two divergent haplotypes termed H1 and H2 in humans [113, 114]. Interestingly, both of these haplotypes have different functional impacts. While degenerative diseases of the nervous system have been linked with the H1 haplotype [115, 116], the H2, due to the inversion, results in a local architecture of directly oriented LCRs, that predispose the region to 17q21.31 microdeletion syndrome. The H2 lineage is nearly absent in Africans and Asians, but has been under positive selection in European populations, where it has been found with a frequency of 20% [113]. The latter could also explain why one of the common causes of ID has almost exclusive presence in subjects of European ancestry [117]. Almost all cases of 17q21.31 syndrome have resulted from a *de novo* deletion, and although the inverted H2 has been found in at least one parental genome of 17q21.31 patients, most affected individuals are single occurences in the family and the recurrence risk for future pregnancies is low [118]. In a recent study, carriers of an atypical smaller deletion embedded in the 17q21.31 monosomy region, and a strikingly similar phenotype, were identified. This discovery narrowed the critical dosage-sensitive interval down to only three genes, including *MAPT* [30].

Few ID patients carrying reciprocal duplications have been reported to date. The associated clinical presentations seem to be variable and somewhat milder than those linked with the monosomy but behavioural disorder and poor social interaction observed in all of these patients [119–121] might suggest a contrasting impact of this region on the clinical phenotype.

In addition to defining novel recognizable syndromes, array-based screening has been a powerful strategy for finding the causative defects underlying previously known syndromic disorders, for example CHARGE syndrome (OMIM #214800; identified by microdeletions in 8q12 that encompass the *CHD7*; OMIM #608892) [31], Peters plus syndrome (OMIM #261540; caused by the *B3GALTL*; OMIM #610308 in chromosome locus 13q12.3) [122], or thrombo-

cytopenia-absent radius (TAR) syndrome (OMIM #274000; associated with deletions on chromosome 1q21.1) [123].

2.3.2. The further definition of rare genomic disorders relies upon international sharing and coordinated collaborations

As identification of patients sharing a genomic variant and having phenotypic features in common leads to greater certainty in the pathogenic impact of CNV and is the prerequisite for defining new syndromes, data sharing and collaboration between clinical and research centres is crucial. For handling rapidly expanding data sets and providing reliable information to geneticists worldwide about extremely rare cases (of which a considerable percentage remains unpublished), open access databases for cytogenetic and clinical data of rare genomic aberrations have been established. The most comprehensive catalogue of novel potentially pathogenic copy number changes and patient reports is stored in the DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources) database hosted by the DECIPHER Consortium and the Wellcome Trust Sanger Institute. As of June 2012, the database includes 15,957 patient reports from 243 participating centres, as well as descriptions of 59 distinctive syndromes (http://decipher.sanger.ac.uk). With the purpose of further facilitating the interpretation of submicroscopic chromosomal rearrangements, DECIPHER utilises the human genome via the Ensembl Genome Browser (http://www.ensembl.org) and incorporates a suite of tools for annotation of aberrant regions [124] which has made it one of the most applicable resources for deciphering the phenotypic significance of rare CNVs. Another effort with a similar purpose to encourage information exchange and collaboration between genetic centres in the field of rare chromosomal disorders is the ECARUCA (European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations) database, coordinated by the European Cytogeneticists Association and Radboud Hospital at the University of Nijmegen, which contains over 4700 case reports with more than 6000 aberrations (http://www.ecaruca.net).

2.4. Variable boundaries of "normality" and clinical phenotypes

2.4.1. Phenotypic variability in genomic disorders

While most syndromic CNVs have been shown to arise *de novo* and occur as sporadic events, a subset of clinically relevant rare variants are often inherited and has been identified showing variable expressivity and incomplete penetrance. As proposed by Girirajan and colleagues, two general types of genomic disorders may be distinguished – (i) syndromic forms where the phenotypic

features are largely invariant (e.g. abovementioned 17q21.31 microdeletion syndrome); and those (ii) where the same genetic lesion is associated with a diverse set of morbidities and severities [105]. The accumulating number of newly described recurrent CNVs associated with extremely variable clinical features [105, 125] has posed significant demands for interpretation of their impact, especially for geneticists who have used to handle genetic disorders mainly as Mendelian traits. Also, the non-specific spectrum of pathogenity has revived the necessity to define clear boundaries of clinical entities and recruit large numbers of subjects who fit with the description of the category. The definition of phenotypes *per se* has been a challenging endeavour in genetic studies, that often includes problems in distinguishing norm from disease, establishing diagnostic criteria, and determining their reliability with no consistent standards for reproducibility and validity [104]. For a diagnosis of ID, an IQ score that is two standard deviations below the general population mean has been widely used as a criterion to quantify "significant limitations in intellectual functioning" [126]. However, standardized values of cognitive capacities are not always available, and establishment of a person`s membership within a certain diagnostic class can be further complicated by co-occurrence of other psychopathologies (e.g. behavioural problems, speech delay etc.). Relative to severe cognitive dysfunction, milder deficits in intellectual capacity and especially the borderline intellectual functioning has received much less attention. Defined by an IQ between 70 and 84 (i.e. between -2 and -1 SD), these individuals can be considered as being in the lower range of normal population variation (reviewed in [127]) that further blurrs the borders between „normality" and clinical diagnosis.

Some well-known examples of CNVs predisposing to neuropsychiatric phenotypes with variable phenotypic manifestations include chromosome regions 16p11.2 (OMIM #611913) [51, 128–130], 15q13.3 (OMIM #612001) [131, 132], 22q11.2 (OMIM #188400) [133, 134], 16p12.1 (OMIM #136570) [135]. All of these are associated with decreased cognitive functioning, with the formal diagnosis of ID or major neurodevelopmental disturbances in only a fraction of patients.

2.4.2. Modifying factors to explain clinical heterogeneity and incomplete penetrance of genomic disorders

The exact mechanisms underlying phenotypic heterogeneity and incomplete penetrance of seemingly identical aberrations are not known and most probably vary according to the characteristics of a particular genomic region. This could be explained by only emerging current knowledge about molecular pathways and compensatory mechanisms involved in the neurodevelopmental processes which might influence inter-individual susceptibility. Depending on the genetic and environmental context the interaction may result in diverse neurological

conditions. In other words, the same genetic payhways may, for example lead to autism, ID, or epilepsy [136].

Distinctive from other tissues, the transcription patterns in the central nervous system seems to be particularly sensitive to the parental origin-dependent regulation of gene expression [137–139]. Up to hundreds of brain-specific transcripts have been demonstrated to be imprinted [140–142], and as demonstrated in the cerebral cortex of mice carrying a heterozygous deletion of *KIDINS220*, a downstream signal transducer of neurotrophins and essential modulator of nervous system development, gender might indeed be an important determinant of central nervous system vulnerability [143]. Still, practically nothing is currently known about the gender-specific influence on CNVs. Most CNV studies have made no distinction whether the maternal or the paternal copy of a chromosome is altered in patients, and associations with the parent-of-origin effect on the neurodevelopmental phenotype have been established until now for a few CNVs overlapping known uniparental disomy regions [144–146].

Also the segmental duplication architecture itself in regions where recurrent aberrations tend to occur poses a challenge to characterize embedded genes and refine the mapping of breakpoint positions. Until recently these complex regions have remained difficult to study despite being known to be gene-rich and to act as a primary source of evolutionary innovation in the human lineage. Concordantly, recent efforts have revealed that several duplication genes of previously unknown function or completely missing from the current version of reference genome are human-specific and implicated in neurodevelopmental processes [147–149]. Thus, only subtle differences in LCR structure and aberration breakpoints might determine the differences in clinical outcome between individuals.

2.4.3. Multivariant contribution in neuropsychiatric and other complex phenotypes

Described initially by Prof. James Lupski, the concept that some genomic disorders might result from a combination of two or more variations, where each of these alone do not provide a genetic burden that is great enough to cause disease [150] has gained support and popularity over the last few years. Emerging data on oligogenic diseases, especially human ciliopathies – a group of diseases with strikingly variable penetrance and expressivity, have shown that the manifestation of a causative mutation can depend upon other genetic variants in the human genome and that these epistatic interactions between causal and second-site modifying alleles are prevalent mechanisms underlying the variable clinical spectrum of the disease [151, 152]. In 2007, Klopocki *et al*. found that clinically well-described TAR syndrome is associated with a common 200-kb deletion, but for developing the phenotype, the existence of one or more yet unknown modifier alleles, called „mTAR“, is required [123]. It is believed that similar interactions between multiple rare structural variants

could contribute to the overall CNV burden that creates differently sensitized backgrounds during human development and eventually leads to different outcomes of phenotypic features [105].

Girirajan and colleagues used the chromosome locus 16p12.1 to test the so called „second hit" model in genomic disorders [135]. Similarly to the well-defined syndromic 17q21.31 region, the locus of 16p12.1 harbors two structurally different haplotypes [153]. Of these, haploblock S2 is more common and predisposes the interval to 520 kb deletions, which in meta-analyses has found to be significantly more prevalent in patients with developmental delay, autism and schizophrenia when compared to controls [30, 154]. However, low prevalence (about 1/15,000) and inconsistent segregation with clinical features have reduced the power of genome-wide studies to definitely identify a disease association and to delineate the phenotypic consequences of the microdeletion. The targeted characterization of the 16p12.1 deletion by Girirajan *et al*. in large ID and control cohorts supported a two-hit model and suggested that although the deletion is necessary to reach a threshold to induce DD, more severe ID phenotype and comorbidities can be manifested only with the addition of another large genomic alteration. To test whether the model might serve more broadly among genomic disorders, the authors expanded the analysis to other recurrent microdeletions with both syndromic and variable phenotypes. The results indicated inverse correlation between the proportion of *de novo* cases reported for a given CNV and the prevalence of double hits in carriers. In comparision with canonical syndromes (e.g. microdeletions of 7q11.23, 17q21.31 and 17p11.2), clear clustering of double-hit CNVs was observed in patients diagnosed with disorders that present variable expressivity and relatively low rates of *de novo* occurence (e.g. microdeletions of 16p11.2, 1q21.1, 15q13.3 and 22q11.2). Thus, the model of a certain single event as a predisposing factor for neuropsychiatric phenotypes, and which may exacerbate the disorder only when co-occuring with other large deletions or duplications might be more generally applicable than previously thought to neuropsychiatric disease [135]. Although formal replication of the double-hit enrichment and epistatic impact of the second hits on the severity of phenotype have been hampered by the lack of large homogeneous sample cohorts [129], the tendency torward high prevalence of second genetic „hits" in syndromes with variable expressivity that distinguishes the patients with more severe clinical manifestations is notable [155].

More globally, the „general genome ecology" concept is supported by the genetic association studies of common diseases. Evidence from the investigations of epilepsy, type 2 diabetes and obesity indicate that at least some of these phenotypes may be inherited in a complex manner cumulating the effect of numerous rare genetic variants that differ from person to person, modify genomic landscape, and when combined have a strong influence on which diseases an indvidual will get and when [156–158].

Thus, in the case of many individual alterations, the driving or modifying effect on the disease has remained vague due to their rarity, heterogeneous com-

binations and thus the need for extraordinarily large sample size. Furthermore, it may be naïve to expect that individuals carrying causative CNVs display the uniform phenotype regardless of their ethnicity or environmental and geographical background. Since the majority of population groups within Europe and worldwide have not been assessed for population specific variants, there is so far no information on potential origin-specific modifiers. Therefore, the ability to reveal true disease associations for rare structural variants depends on obtaining data from cohorts of sufficient size, not biased by pre-determined ascertainment criteria, and including appropriate ethnicity-matched controls in CNV association studies.

2.5. The functional effect of structural variations on gene expression and clinical phenotype

2.5.1. Genes in CNV regions show more variability in their expression

Although genotype-phenotype correlations have been established for hundreds of CNV loci, the exact impact through which CNVs lead to altered expression of genes and result in the ultimate clinical features have remained largely unknown.

It has been demonstrated that both balanced and unbalanced structural variants may have a profound and dramatic effect on the expression levels of genes located within the rearranged region, influence genes in their vicinity, and affect global „genome regulation“ [159–166]. On a whole-transcriptome level, about 5% of human genes are altered in dosage by CNVs [2], and copy-number variation has been estimated to explain almost 20% of the variation in gene expression [167]. The latter impact might be underestimated due to the bias towards large CNVs on current maps of genomic variation. Less is known about smaller CNVs which are more abundant, likely to affect individual functional units, and are expected to have more specific effect on transcription [168].

Assessment of the global impact caused by CNVs on tissue transcriptomes using model organisms has confirmed enrichment of altered transcripts among genes mapped within variable region, and positive correlation between relative expression levels and gene dosage. However, a subset of about 10% of genes within CNVs shows negative correlation between gene copies and expression levels [160, 169]. Although the exact underlying mechanism is not known, two models have been suggested that may explain this inverse effect of genomic gain. First, the higher concentration of a CNV-gene, termed as immediate early gene (IEG), and thus also proportionally higher amount of the gene product induces a repressor that subsequently downregulates the expression of the IEG. Secondly, tandemly located extra copies of a gene physically affect its transcription via impaired access to the transcription factory [159, 170]. It has also been demonstrated that CNV-genes have specific properties with respect to

their spatial expression. For example, they have a lower and restricted tissue-specific transcription pattern, and show more inter-individual differences in temporal patterns of expression. This indicates that altered number of genomic copies may affect not only the expression level of the gene, but may also modify the timing of its expression [160, 162]. Interestingly, although genes with brain-specific expression are more tightly regulated, corroborating the general vulnerability of the nervous system [160, 163], transcriptional control over the CNV-genes is looser during the early period in development, when neurons are subject to pruning and competing for growth factors. This raises the question whether reduced control is due to the lack of regulatory proteins at this time-point or strict regulation is harmful in specific stages of neurodevelopment. Thus, relaxing the expression control of genes within copy-number variable regions during a critical period may somehow favor neuronal outgrowth, differentiation and formation of synapses [162].

2.5.2. Local mechanisms by which CNVs may impact gene expression

One of the key features of structural variants with regard to functional impact is their large size, allowing CNVs to encompass thousands of basepairs and affect large functional units. The diversity in physical extent in combination with the genomic architecture of rearranged region provides many ways for gene expression to be altered by CNVs. Comprehensive reviews by Zhang *et al.* [171] and Harewood *et al.* [159] have been used as a base for following classification.

If the aberrant region contains dosage-sensitive genes (i.e. genes of which only two copies produce the normal phenotype), change in the number of functional copies can lead to disease. A classical example of copy number variant-driven dosage-sensitivity includes the *peripherial myelin protein 22* (*PMP22*; OMIM #601097) in the 17p12 region. While increased levels of the peripherial myelin protein contribute to the Charcot-Marie-Tooth disease Type1A (OMIM #118220), reduced production of the same protein results in hereditary neuropathy with liability to pressure palsies (OMIM #162500). Concordant with the description of novel microdeletion and -duplication syndromes, dosage-sensitivity has been progressively linked with neurological disorders [172–175]. Intriguingly, so called „mirror-phenotypes“ have been observed for some recurrent CNV regions. Examples of these include opposite effects of reciprocal deletions and duplications on height, body weight and head circumference in the 16p11.2 [51, 129, 176], the 5q35.2q35.3 [175, 177, 178] and the 17p11.2 syndrome regions [179], as well as social and language aspects in patients with Williams-Beuren and 7q11.23 duplication syndrome [180, 181]. Since the presence of deletions is twice as common and shows greater penetrance compared to duplications in patients with severe paediatric diseases, it has been suggested that amongst alterations in gene dosage, haploinsufficiency is less tolerated and more common than triplosensitivity [30].

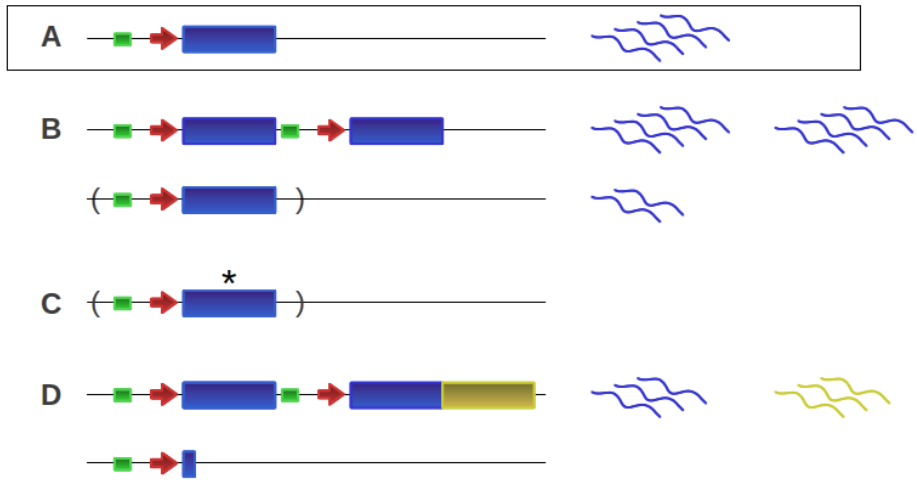

Figure 2. Schematic presentation of local mechanisms by which CNVs may impact gene expression. The coding region in a locus is indicated by blue box, promoter by red arrow and enchancer by green box. Encoded transcript levels are indicated by blue wavy lines. Deleted are is marked by parentheses and deleterious mutation by asterix. Additional affected gene and it's product are showed by yellow box and wavy lines, respectively. No CNV is present and gene expression is not affected (**A**). Gene-dosage is altered due to genomic gain or loss (**B**). Unmasking of recessive allele by loss of heterozygosity (**C**). Dysregulation of expression due to gene fusion and interuption (**D**). Modified from [170].

When breakpoints of the CNV map within a gene, the rearrangement can cause it`s inactivation by disrupting the transcript structure, or result in gain of function by fusing different genes or their regulatory elements. Loss of heterozygosity by deletion may cause unmasking of a detrimental recessive point mutation or functional polymorphism that might contribute to particular features in the clinical phenotype. For instance, congenital malformations of the vertebral column, as well as epilepsy and paroxysmal dyskinesia have been seen recurrently, though only in subset of patients with the 16p11.2 deletions. Considering the T-box protein gene *TBX6* (OMIM #602427) [182–184] and the proline-rich transmembrane protein 2 gene *PRRT2* (OMIM # 614386) [185–191] in the imbalanced interval, these features are likely determined by hemizygous expression of the mutant allele.

The functional mechanisms driven by structural rearrangements in a genomic locus are schematically presented on **Figure 2**.

2.5.3. Structural rearrangements may modulate genome-wide expression

Current knowledge is rather hypothetical about the *cis*- and *trans*-position effect of CNVs on genes outside of the imbalanced region. Considering the dense presence of CNVs in the human genome, yet unknown compensatory mechanisms might also exist which reduce the functional impact of genomic variations. For example, it has been proposed that in case of a dominant loss-of-function mutation the phenotype could be rescued by the gain of gene copies resulting in a „balanced" transcript [192, 193]. Consistent with this hypothesis, the rescue of the phenotype of the 22q11.2 deletion, usually leading to DiGeorge syndrome (OMIM #188400) and velocardiofacial syndrome (OMIM #192430) has been demonstrated to be due to balancing reciprocal duplication on the other copy of chromosome 22 [193]. Although this is a first known example of genetic compensation in a human genomic disorder, a similar compensatory effect has been shown in the mouse model for human chromosomal region 22q11.2 [194]. Contrary to the latter, in compound heterozygous mice, investigations of the deletion and duplication in the Smith-Magenis/Potocki-Lupski syndrome region at 17p11.2 have revealed that restoration of normal genomic copy number in *cis*-configuration does not restore some neurobehavioural traits. Thus, regardless of gene dosage, at certain positions in the human genome, rearrangements *per se* can perturb certain pathways and generate clinical phenotypes [163]. Examples of how structural change can disturb gene functionality include (i) physical dissociation of the transcription unit from its *cis*-acting regulators, (ii) placing a gene under the influence of a foreign promoter, (iii) modification of transcription control through altered chromatin structure, loops and position within the nucleus, (iv) disrupting a regulatory interactions between homologous chromosomes, or (v) altering normal spatial organization of the nucleus and thus placing genes into an anomalous chromatin context [159, 160, 164, 195].

Consequently, emerging data in this field suggest that different mechanisms of transcriptional variation might be driven by structural rearrangements. Given that gene expression is fundamental to cellular function and transcript differences could serve as a proxy for other levels of phenotypic variability, CNVs play a crucial role with respect to risk and development of neurodevelopmental disorders, as well as other complex diseases [160, 167].

3. AIMS OF THE STUDY

The aims of the current study were as follows:

1. To perform the first comprehensive investigation for identifying clinically relevant genomic rearrangements in Estonian families with unexplained intellectual disability and to establish the workflow for array-based genomic profiling for improving the diagnostic possibilities of patients with neuro-developmental disorders.
2. To perform the first investigation of rare structural variants and associated phenotypic traits in individuals from the Estonian general population.
3. To investigate how rare potentially pathogenic CNVs impact phenotypes by using the data across two abovementioned cohorts, and to shed light on the phenotypic variability of these CNVs.

4. MATERIALS AND METHODS

4.1. Clinical collection of Estonian ID patients (EID)

DNA samples from 77 Estonian families with idiopathic ID, ranging from mild to severe, or developmental delay (DD) were analysed in the current study. In addition to index patients all family members available for investigation were analysed to determine accurately the segregation of variants with the disease phenotype. The number of individuals investigated per family ranged from 1 to 10, making a total of 257 samples of which 165 were affected and 92 unaffected. Throughout the study, every family was considered as a single separate ID case.

All patients were assessed by a clinical geneticist at the Department of Genetics, United Laboratories, Tartu University Hospital. Clinical evaluation of this study was leaded by Prof. Katrin Õunap. In most patients, ID was accompanied by dysmorphisms and/or other congenital anomalies (CA). No consanguinity was reported before the study, but was later confirmed in one family according to the genotyping results. Standard G-banded karyotypes on a 550-band level showed no obvious aberrations in all cases. Routine metabolic analysis and test for fragile X syndrome was carried out for all patients. Tests for Prader-Willi/Angelman syndrome or other specific ID disorders were carried out in case of clinical indications.

Genomic DNA was extracted from peripheral blood leukocytes according to a standard protocol. DNA concentrations were measured and quality was assessed by agarose gel electrophoresis and a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA).

The study was approved by the Ethics Review Committee on Human Research of the University of Tartu, Tartu, Estonia. Informed consent was obtained from all families included in the study.

4.1.1. Patient EID-6

The proband of the family EID-6 was born as a second child to non-consanguineous parents of Estonian-Russian origin. No data about the delivery and his birth antropomethry is available, but since he was allowed to leave the hospital on the third day after birth, it is assumed to be uneventful.

At the age of 4 years and 10 months the patient was diagnosed with global DD accompanied by severe speech delay. Tests to assess his intellectual abilities were not administered due to the level of his cognitive disability, and the exact degree of ID was not possible to determine. However according to the psychiatrist's opinion his cognitive functioning corresponded to moderate to severe ID. The patient showed attention-deficit hyperactivity disorder (ADHD) and exhibited severe aggression towards his mother and siblings. Autistic features and stereotypic movements were also noticed. The neurological

examination showed no pathological reactions, but the brain MRI revealed leukoencephalopathy. His growth parameters were in the normal range for his age, weight of 18.4 kg (50th percentile), height of 102 cm (10th percentile), and OFC of 51 cm (50th percentile). The patient's dysmorphic facial features include a broad nose, protruding, dysmorpic ears, deep-set eyes, hypertelorism, strabismus on the right side, slight synophrysis, short philtrum and thin upper lip (**Figure 3**). Additional dysmorphisms are sandal gaps, a broad first toe, and hirsutism at the back and extremities. The spinal X-ray revealed a hypoplastic 12^{th} pair of ribs. Patent ductus arteriosus was diagnosed by the ultrasound investigation. It was also known that the patient has been hypotonic and had cryptorchidism which was resolved with no intervention before the age of 2 years.

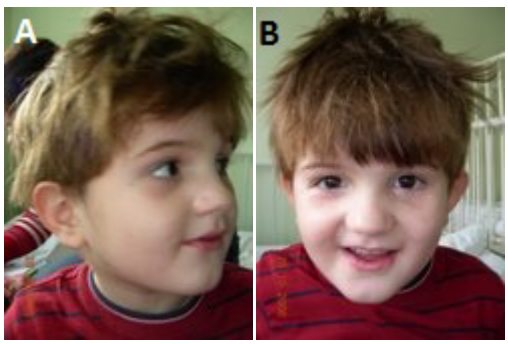

Figure 3. Profile and frontal view of the index patient EID-6 at the age of 4 years and 10 months. Note protruding ears (**A**), thin upper lip and a high broad nose (**B**). Written permission to publish the photos of this patient was obtained from the family.

The father (I:2 at **Figure 4**) of the proband did not complete his special education program in his youth. Since he refused testing, there is no official data about his current intellectual status, but cognitive disability was obvious to the clinical geneticist at the time of evaluation. In addition, nervous, aggressive behavior and speech impairment was recorded. He has mildly dysmorphic facial features including hypertelorism, a broad nose, deep-set eyes, a low frontal hairline and a short philtrum. At the age of 7, the proband's older brother (II:1) showed developmental and speech delay, stereotypic movements, hypotonia and mildly dysmorphic features. The younger brother (II:4) 6 months old at the time of evaluation, presents global DD, hypotonia and dysmoprhisms: a broad nose with a flat nasal bridge, deep-set eyes, hypertelorism, and epichanthal folds. His growth parameters were normal, weight of 9100 g (85^{th} percentile), height of 69 cm (50^{th} percentile), OFC of 44,5 cm (85^{th} percentile). The daughter in this family (II:3) has normal cognitive development and is healthy.

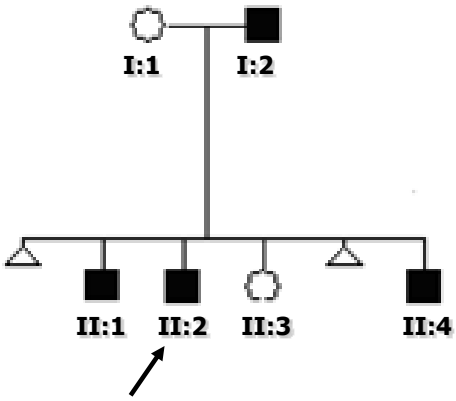

Figure 4. Pedigree of the family EID-6. In the diagram, members of the family are represented by standard symbols – circles indicate females, squares males and the bottom line shows the children of couple above. Affected members are indicated by black and unaffected by white symbols. Proband (II:2), his two affected brothers (II:1, II:4) and father (I:2) as well as healthy mother (I:1) and sister (II:3) were analyzed in the current study.

4.2. Estonian general population cohort (EGC)

Genotype and phenotype information from Estonian general population individuals provided by the Estonian Genome Centre at the University of Tartu (EGC UT) was used as the comparative data-set for CNV analysis. The EGC UT maintains a general population-based biobank, described in greater detail in [196]. The EGC UT is run according to the Estonian Gene Research Act. Written informed consent was obtained from all voluntary participants, and the study was approved by the Ethics Review Committee on Human Research of the University of Tartu.

First, 1058 randomly selected unrelated samples were genotyped. Based on the data from 998 samples that passed the quality control standards for CNV analysis, population specific list of common CNV regions (frequency ≥1%) was generated. Secondly, this information was used to identify the presence of rare genomic imbalances of potential clinical significance and to estimate their phenotypic consequence.

For follow-up analysis of the phenotypic effect of the CNVs in genomic regions 7p21.2-p21.1, 7q11.23, 15q13.2-q13.3, 16p11.2, Xp22.31 and Xq28, an additional set of mixed GWAS cases and controls for common traits (n=6901) was used. 6628 of the samples passed the quality control standards for CNV analysis that was performed using the algorithms and workflow described below.

4.3. CNV analysis by whole-genome SNP genotyping

Genomic rearrangements in the EID and initial EGC cohort were screened by the Infinium® II whole-genome genotyping assay with the HumanCNV370 BeadChips (Illumina Inc.). The HumanCNV370 BeadChip covers the entire human genome with an average spacing of 5 kb, allowing an average effective resolution of 50 kb (i.e. 10 consecutive markers). The genotyping data in the follow-up EGC cohort was obtained from the Infinium® II whole-genome genotyping assay analyzed with different BeadChips (Illumina Inc., San Diego, CA, USA) was used. Samples were processed and the assay performed according to a routine protocol provided by the manufacturer. Cohorts and genotyping platforms analyzed in the current study are summarized in **Table 3**.

Table 3. Estonian samples analyzed in the current study.

Individuals	Sample size	QC passed	Recruitment	Genotyping platform
Estonian families with intellectual disability (EID)				
Probands	77	77	ID (and CA)	Illumina HumanCNV370 BeadChips
Affected members	88	88	ID (and CA)	Illumina HumanCNV370 BeadChips
Unaffected members	92	92	ID families	Illumina HumanCNV370 BeadChips
Estonian general population (EGC)				
Initial cohort	1058	998	General population	Illumina HumanCNV370 BeadChips
Follow-up cohort	6901	6628	Mixed common traits	Illumina HumanCNV370, Human610, OmniExpress and custom BeadChips

Genotypes were called by BeadStudio software GT module v3.1 (Illumina Inc.). Log R Ratio (LRR) and B Allele Frequency (BAF) values produced by the BeadStudio software were formatted for further CNV analysis and break-point mapping with Hidden Markov Model-based softwares QuantiSNP (ver. 1.1 and 2.1) [197] and PennCNV (ver. 2009aug27) [198]. In addition to LRR and BAF values, SNP marker allele frequency data from the Estonian general population was used as the reference in the PennCNV software. Parameters suggested by the software authors were used in both QuantiSNP and PennCNV. Only samples with a call rate greater than 98% that passed QuantiSNP quality control parameters were analyzed. To minimize the number of false positive findings, CNVs >50 kb in size, detected by both algorithms and visually confirmed in BeadStudio GenomeViewer were selected for further interpretation. In families where both parents and offspring were available for investigation, the parental origin of variants and exact CNV haplotypes were determined *in silico*. Details

of the allelic composition determination algorithm will be described elsewhere (Palta *et al*., manuscript in preparation).

To exclude neutral variations, inheritance of CNVs detected in ID patients was determined in the corresponding family. Only CNVs that arose *de novo* or segregated in the family with clinical phenotype were selected as potentially relevant. These CNVs were further compared with those recurrently present in the Database of Genomic Variants (DGV) and in the Estonian general population. The potential clinical significance of CNVs not present in normal individuals was evaluated using OMIM and DECIPHER databases and peer-reviewed literature searches in the PubMed database. The genomic context of aberrant regions was studied using the Ensembl database version 54 (based on NCBI build 36).

The presence or absence of genomic aberrations of potential clinical relevance was confirmed by quantitative PCR in all investigated family members. FISH analysis was performed according to standard cytogenetic protocol in most cases of individuals carrying duplications and in which unbalanced translocation was suspected. The workflow of CNV analysis and interpretation is provided in **Figure 5**.

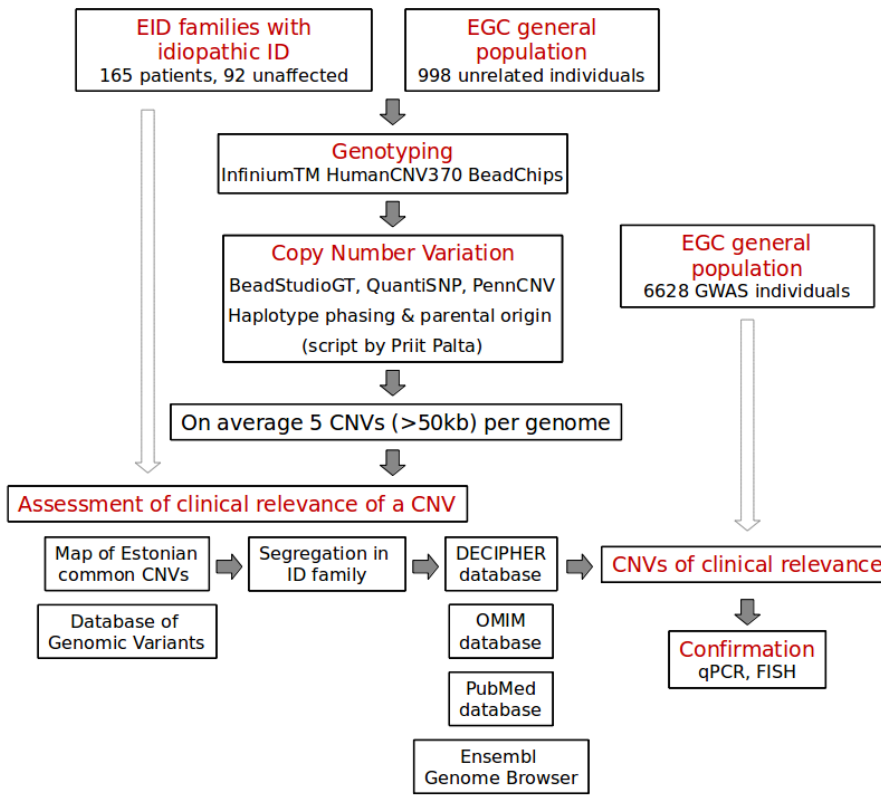

Figure 5. The flowchart of CNV analysis and interpretation used in the current study.

4.4. Gene expression analysis by RT-qPCR

Real-time reverse transcription-qPCR (RT-qPCR) was applied to investigate the expression status of the candidate genes mapping to the deleted regions in patient EID-3.

Total RNA was extracted from whole-blood using the Tempus™ Spin RNA Isolation Kit (Applied Biosystems, Carlsbad, CA, USA). For replication experiments, a lymphoblastoid cell line (LCL) was established from the proband's peripheral blood and total RNA was extracted using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA). RNA samples were treated with the TURBO DNA-free™ Kit (Applied Biosystems/Ambion) and used as templates for synthesis of complementary DNA (cDNA) with oligo(dT) primers and the First Strand cDNA Synthesis Kit (Thermo Scientific, Vilnius, Lithuania).

Assays for target (*MEOX2*, *SOSTDC1*, *POU1F1*, *CHMP2B*, *BZW2*, *CGGBP1*, *C3orf38*, *TWIST1*), reference (*HMGB2* OMIM #160938; *PTPN1* OMIM #176885, *RGS9* OMIM #604067), and normalization (*ACTB* OMIM #102630) genes were designed using CloneManager software (Sci-Ed Software, Cary, NC, USA) and the web-based program GeTprime (http://updepla1srv1.epfl.ch/getprime) using default parameters. The list of transcripts and validated assays is given in **Table 4**.

RT-qPCR experiments were performed on the 7900HT Real-Time PCR system (Applied Biosystems) using ready-to-use HOT FIREPol® EvaGreen® qPCR MixPlus (Solis BioDyne, Tartu, Estonia). The following amplification conditions were applied: denaturation at 95ºC for 15 min, quantitation step by 40 cycles of denaturing at 95ºC for 15 s, and combined annealing and extension at 60ºC for 1 min. The threshold cycle values were obtained and processed using SDS v2.4 software (Applied Biosystems), and the absolute quantification values were further analyzed using qBase^{PLUS} software (Biogazelle, Ghent, Belgium).

Table 4. Genes and primer sequences used for RT-qPCR.

Gene	RefSeq ID	Forward primer	Reverse primer	Amplicon (bp)
Target genes				
CHMP2B	NM_014043.3	CGGTTCTGATGACGAAGAAG	AGTAGAGGCAGATGGTAAGC	127
BZW2	NM_014038.2	GGCTGCTTGAACTCTTTCC	CTTAAGACCTGCGTCAGTG	77
CGGBP1	NM_001195308.1	TCCCTTATTCCCTTGCTCCT	ACCAGAGACGCATCAAATCC	235
C3orf38	NM_173824.3	CACAATCTTATTCAGCATGC	TTATCTTCTTTCACCTGCTG	127
Reference genes				
HMGB2	NM_001130688	TACGCCTTCTTCGTGCAGACC	CTGTCATAGCGAGCTTTGTC	173
PTPN1	NM_002827.2	AGACCCAGGAGGATAAAGAC	CCCGACTTCTAACTTCAGTG	111
RGS9	NM_001081955.2	CAACGATGCCATCATGTCAG	CGCATCTTGGTTGGGATTTC	108
Normalization gene				
ACTB	NM_001101.3	CTGGAACGGTGAAGGTGACA	CGGCCACATTGTGAACTTTG	65

In addition to the patient, expression levels were measured in six unrelated control individuals. All samples were analyzed in triplicates, with mean values used for calculations. Throughout the workflow, quality of the samples was assessed and procedures were validated according to the requirements advocated by Nolan *et al.* [199].

4.5. Chromosome X inactivation analysis

Individuals with rearrangements on the X chromosome and their female relatives, were subjected to X-inactivation assay based on the analysis of the polymorphic CAG repeat sequence in the human androgen receptor (*AR*; OMIM #313700) gene.

To evaluate the pattern of X-inactivation, genomic DNA was digested overnight at 37ºC with methylation-sensitive restriction enzyme *Hpa*II followed by 20 min of inactivation at 80ºC. The reference reaction with restriction enzyme *Rsa*I (both from Thermo Scientific) was performed for each sample using the same conditions. Digested DNA was amplified by using the *AR*-specific 6-FAM™ labelled foward primer 5'-GTCTACCCTCGGCCGCCGTC, reverse primer 5'-GTAGCCTGTGGGGCCTCTACG (Metabion AG, Martinsried, Germany), and applying the following PCR conditions: denaturation at 95ºC for 10 min; 33 cycles of denaturing at 95ºC for 30 s, annealing at 60ºC for 30 s, extension at 72ºC for 20 s; and final extension step at 72ºC for 5 min. Amplicon size and inactivation ratios were determined by fragment analysis on the ABI 3130 Genetic Analyzer (Applied Biosystems) under standard conditions, and analyzed using GeneMapper® 4.0 software (Applied Biosystems). The peaks corresponding to the two chromosome X alleles in both *Hpa*II and *Rsa*I digested samples were identified. The area under the peak curve obtained by visualization of fluorescently labelled PCR products were used for X-inactivation calculations. The peak areas from the reference reactions were used for normalization, followed by the comparision of the peak areas for two alleles upon digestion for detecting the presence of a skewed or a random inactivation pattern. The formula 4.1, provided by Kiedrowski *et al*. [200] was used to simultaneously perform both normalization and calculation of the proportion of allele 1 on the active X chromosome.

$$1/A = D1_{Hpa}/D2_{Hpa} * D1_{Rsa}/D2_{Rsa} + 1$$

A – proportion of allele 1 on the active X
$D1_{Hpa}$ – *Hpa*II digested peak area 1
$D2_{Hpa}$ – *Hpa*II digested peak area 2
$D1_{Rsa}$ – *Rsa*I digested peak area 1
$D2_{Rsa}$ – *Rsa*I digested peak area 2
Skewed inactivation was flagged whenever the ratio between two alleles was over 75%:25%.

5. RESULTS AND DISCUSSION

5.1. Structural rearrangements in Estonian patients with intellectual disability and general population indviduals (Publication I)

In the current study, genotyping information provided by high-resolution SNP arrays was used to investigate Estonian families with unexplained ID. In parallel, genomic data, medical history, and information regarding educational level and daily life of Estonian general population individuals was analyzed in order to better determine the relationship between phenotype and genomic re-arrangements.

CNV detection in two Estonian cohorts

During the first stage of the study, DNA samples from 77 probands with idiopathic ID, 88 other affected family members, 92 unaffected family members and 998 unrelated reference individuals were analyzed. An average of 5 CNVs were detected per investigated genome with a size range above the resolution limit of the platform, i.e. 0.05 Mb, up to 8.3 Mb. CNVs with the general population frequency ≥1% in Estonian reference samples and/or regions with more than a single record in the Database of Genomic Variants comprised the majority of aberrations and were excluded as likely benign polymorphic variants. Distinction of polymorphic CNVs by using the Database of Genomic Variants only was hampered by its (i) heterogeneous content of platform resolutions, some of which overestimate the size of CNVs; (ii) lack of information about the population frequencies of the variants, and (iii) lack of phenotypic background of the sample cohorts. In addition to the uniform definition of population-specific common CNVs, the majority of the alterations detected in our clinical samples were completely encompassed by common CNVs in our general population. Since there are no generally accepted guidelines for reconciling overlap between CNVs in patients and in control cohorts [85], utilization of the same array platform in the current study proved to overcome this commonly encountered problem and facilitated the initial filtering of rare CNVs.

As shown previously, SNP allele frequencies and linkage disequilibrium patterns of the Estonian population are similar to the other populations having European ancestry [196]. Concordantly, common CNV regions detected in the current reference group were at least partially covered by the Database of Genomic Variants. As of November 2010, only 20 variants with frequencies from 0.1 to 0.5% and a size range from 0.06 to 0.26 Mb not present in the Database of Genomic Variants were detected in the Estonian general population cohort [201].

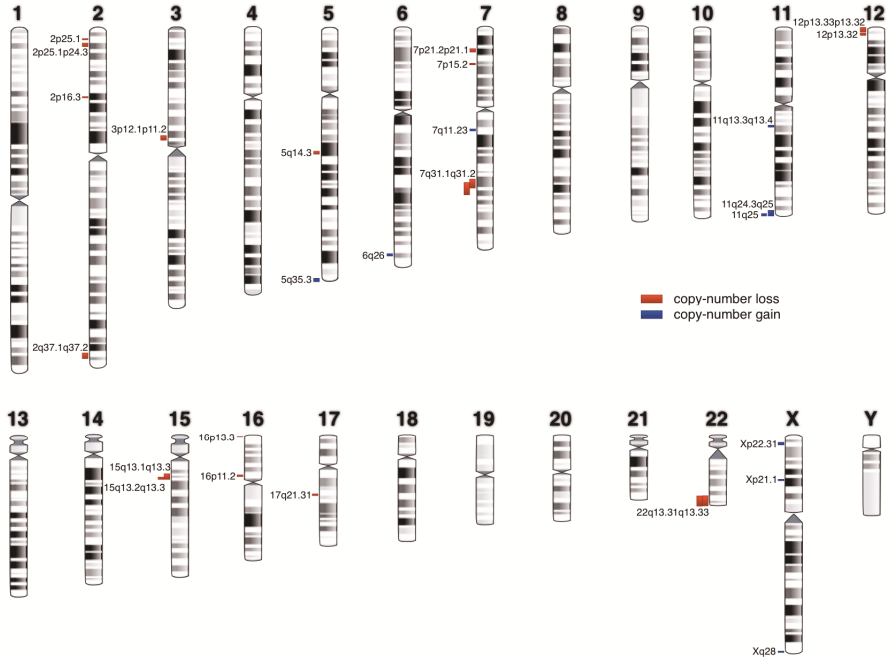

Figure 6. Overview of clinically significant aberrations in patients with idiopathic ID. In total, 18 aberrations of clinical relevance and 5 aberrations of uncertain significance were identified by SNP genotyping arrays. Copy-number losses are indicated by the red bars and copy-number gains by the blue bars [201].

Clinically relevant CNVs in the Estonian intellectual disability cohort.

Clinically relevant rearrangements were identified in 18 out of the 77 investigated ID families by determining their inheritance patterns, comparing them with reports in the DECIPHER database and peer-reviewed literature and annotating genomic intervals. The diagnostic yield of 23% in our clinical cohort is comparable to the results from other similar reports (reviewed in [82, 85, 125, 202]). Determined pathogenic rearrangements included 13 deletions, 3 duplications and 3 apparently unbalanced translocations. One patient had 2 seemingly independent deletions. Five additional rare genomic variants found in ID families were classified as of uncertain clinical significance. The genomic locations of these imbalances is given in **Figure 6**. Phenotype information and molecular data from all ID cohort probands carrying potentially relevant CNVs is summarized in **Tables 5** and **7**.

Rare genomic imbalances of clinical significance in the Estonian general population

In addition to filtering out population-specific common variants, the cohort of 998 randomly selected unrelated EGC UT individuals was used to identify the

presence of rare genomic imbalances of clinical significance in the general population, and to further estimate their phenotypic consequence. The selection of genomic regions was based on the findings in the clinical cohort of the current study, DECIPHER syndromes, and those listed by Girirajan and Eichler [105]. This resulted in 19 general population carriers in total. As a follow-up analysis of the phenotypic effect, an additional sample set of 6628 individuals was screened for the CNVs of special interest in genomic regions 7p21.2-p21.1, 7q11.23, 15q13.2-q13.3, 16p11.2, Xp22.31 and Xq28, which identified 10 more carriers of these imbalances. All CNVs of potential clinical relevance detected in the Estonian general population cohort and associated information about education and neuropsychiatric phenotype for these individuals is shown in **Tables 5** and **6**. Results from both investigated cohorts are organized and discussed according to the genomic regions in sections 5.1.1. to 5.1.2.

5.1.1. Recurrent genomic rearrangements of clinical relevance

Chromosome regions that have directly oriented LCR-rich architecture act as hot-spots for NAHR-mediated recurrent deletions and duplications [21, 157]. As suggested by Girirajan and Eichler [105], genomic syndromes associated with these loci may be divided into two types based on their clinical consistency: (i) specific syndromes with relatively straightforward clinical phenotypes (Grade 1 by the DECIPHER database), and (ii) genomic lesions with diverse phenotypic expressivity and incomplete penetrance (Grades 2 and 3). Amongst Type I disorders, one proband (EID-14, **Table 5**) was diagnosed with 17q21.3 recurrent microdeletion syndrome and one family (EID-6, **Table 5**) with 7q11.23 duplication syndrome in the current clinical cohort. Concordantly, no carrierers of these CNVs were detected in the general population. The latter region is discussed in paragraph 5.1.1.1. as an example of a syndromic disorder.

Findings from disorder regions with less clear phenotypic outcome and the usefulness of well-phenotyped reference data for interpretation of Type II rare variants that were associated, but not limited, to ID phenotype is covered in more detail in sections 5.1.1.2 to 5.1.1.5.

Table 5. Summary of clinical and molecular data of ID patients showing aberrations with potential clinical significance.

Case ID	SNP array result	Length (Mb)	Origin	Ensembl genes	OMIM genes	Gender, Age	Phenotype
EID-1	arr 2q37.1q37.2(231,585,451–235,470,992)x1	3,9	De novo (mat)	73	26	M, 15y	Severe ID, speech delay, autistic features, macrocephaly, Chiari anomaly, severe growth retardation
EID-2	arr 2p25.1(8,786,625–9,505,542)x1	0,7	Mat (aff)	7	3	F, 12y	Severe ID, behavioural problems, speech delay, hypotonia, epilepsy, dysmorphisms
	arr 2p25.1p24.3(12,560,894–14,924,411)x1	2,3	Mat (aff)	5	2		
EID-3	arr 3p12.1p11.2(85,807,759–88,750,837)x1	2,9	Not determined	18	6	M, 7y	Moderate ID, speech delay, severe growth retardation, cranial and skeletal deformations, dysmorphisms, irregular and crowed teeth, behavioral problems, hyperopia
	arr 7p21.2p21.1(14,693,604–16,792,988)x1	2,1	Not determined	14	3		
EID-4	arr 5q14.3(89,093,506–90,644,765)x1	1,6	De novo (pat)	8	2	M, 5y	ID, speech delay, autistic features, hypotonia
EID-5	arr 5q35.3(178,547,299–180,623,543)x3	2,1	Mat (aff)	48	18	M, 6y	Mild ID, speech delay, behavioural problems (hyperactivity), dysmorphisms
	arr 16p13.3(37,354–479,141)x1	0,4	Mat (aff)	29	16		
EID-6[1]	arr 7q11.23(72,388,281–73,777,987)x3	1,4	Pat (aff)	28	18	M, 7y	Severe ID, speech delay, attention deficit/hyperactivity disorder, dysmorphisms
EID-7[2]	arr 7q31.1q31.2(108,290,244–114,759,023)x1	6,5	Not determined	29	7	F, 6y	Moderate ID, severe speech delay, hypotonia, dysmorphisms
EID-8[3]	arr 7q31.1q31.2(111,801,620–120,139,346)x1	8,3	Mat (aff)	52	18	F, 5y	DD, severe speech delay, dysmorphisms, kidney anomaly
EID-9	arr 12p13.33p13.32(52,602–3,996,049)x1	3,9	Pat [6]	45	18	F, 9y	ID, speech delay, autistic features, behavioural problems, dysmorphisms
	arr 11q24.3q25(129,951,233–134,435,899)x3	4,5		23	8		
EID-10	arr 12p13.32(3,019,125–4,836,475)x1	1,8	Mat (aff)	24	12	M, 13y	Mild ID, severe speech delay, growth retardation, dysmorphisms
EID-11[4]	arr 15q13.1q13.3(26,772,437–30,676,740)x1	3,9	De novo (mat)	65	8	M, 18y	Mild ID, dysarthria, growth retardation, obesity, microcephaly, Prader-Willi like, behavioural problems
EID-12	arr 15q13.2q13.3(28,849,136–30,302,218)x1	1,5	Mat (aff)	14	4	F, 4y	Moderate ID, hypotonia, dysmorphisms, Prader-Willi like phenotype
EID-13[5]	arr 16p11.2(29,502,984–30,085,308)x1	0,6	De novo (mat)	33	15	M, 5y	Mild ID, speech delay, behavioural problems (hyperactivity), dysmorphisms, asthma
EID-14	arr 17q21.31(40,941,921–41,560,151)x1	0,6	De novo (mat)	14	3	M, 5y	ID, agenesis corpus callosum, macrocephaly, hypotonia, dysmorphisms
EID-15	arr 22q13.31q13.33(42,861,412–49,562,479)x1	6,7	De novo (mat)	104	41	F, 15y	ID, severe speech delay, hypotonia, ataxia, mild obesity, dysmorphisms

Case ID	SNP array result	Length (Mb)	Origin	Ensembl genes	OMIM genes	Gender, Age	Phenotype
EID-16	arr 22q13.31q13.33(43,288,781–49,524,956)x1	6,2	Mat [7]	93	38	F, 9y	Severe ID, severe speech delay, macrocephaly, epilepsy, hypotonia, dysmorphisms
	arr 11q25(133,219,732–134,435,899)x3	1,2		14	6		
EID-17	arr Xp21.1(31,665,779–32,083,817)x2	0,4	Mat (unaff)	1	1	M, 7y	Moderate ID, severe speech delay, autistic feat., overgrowth, dysmorphisms, thumb anomaly
EID-18	arr Xq28(153,777,927–154,218,134)x2	0,4	Not determined	16	6	M, 16y	Mild ID, dysarthria, dysmorphisms

Aberration breakpoints are reported as detected using the QuantiSNP algorithm. Only index cases are listed in the table. The following patients have been described in more detail: [1]as patient 8 in Publication IV [181]; [2]as proband 2 and [3]as proband 1 in [203]; [4]as patient 17 in [132]; [5]as case 1 in Publication III [51]. [6]Father is a carrier of balanced translocation with 46,XY.ish t(11;12)(qter;pter)(11qtel38-,12ptel27+;12ptel27-,11qtel38+). [7]Mother is a carrier of balanced translocation 46,XX.ish t(11;22)(q25;q13.3)(N85A3-,D11S4437+;D11S4437-,N85A3+).

Table 6. Summary of phenotype and molecular data of general population individuals showing aberrations in the loci of genomic syndromes*

Case ID	SNP array result	Length (Mb)	Ensembl genes	OMIM genes	Gender, Age	Education	BMI	Neuropsychiatric, congenital anomalies and features characteristic to the syndrome (if known)
EGC-1	arr 1q21.1(144,106,312–144,487,950)x1	0,4	49	21	M, 21y	Secondary	19.9	Not reported
EGC-2	arr 1q21.1(144,863,910–146,325,557)x3	1,5	185	54	M, 56y	Basic	30.1	Self-report of moderate pain/discomfort
EGC-3	arr 3q29(197,162,292–198,830,963)x3	1,7	34	11	M, 29y	Elementary	24.2	Self-assessment of moderate anxiety/depression and pain/discomfort
EGC-4	arr 15q13.2q13.3(28,570,072–30,302,218)x3	1,7	26	4	M, 20y	Basic	24.8	Not reported
EGC-5§	arr 15q13.2q13.3(28,712,787–30,297,184)x1	1,6	18	4	M, 67y	Secondary	34.5	F98.6 (cluttering); self-assessment of moderate anxiety/depression and pain/discomfort
EGC-6§	arr 15q13.2q13.3(28,712,787–30,297,184)x1	1,6	18	4	F, 76y	Elementary	25.4	Self-assessment of moderate pain/discomfort and problems with daily duties
EGC-7	arr 16p13.11(15,000,279–16,215,852)x3	1,2	18	5	M, 28y	Professional secondary	26.8	Not reported
EGC-8	arr 16p13.11(15,000,279–16,370,751)x3	1,4	22	6	M, 64y	Elementary	15.8	Self-report of moderate pain/discomfort and dificulties with daily duties
EGC-9	arr 16p13.11(15,000,279–16,370,751)x3	1,4	22	6	M, 35y	Professional secondary	31.0	Not reported
EGC-10	arr 16p12.1(21,702,011–22,343,312)x1	0,6	13	3	F, 65y	Basic	23.6	G40 (epilepsy); F33 (recurrent depressive disorder); G30+ F00 (dementia in Alzheimer disease)
EGC-11	arr 16p12.1(21,746,841–22,343,312)x1	0,6	12	3	F, 28y	Basic	19.3	F48.9 (specified neurotic disorders)
EGC-12[1]	arr 16p11.2(29,502,984–30,085,308)x1	0,6	33	15	M, 23y	Basic	35.9	F32 (depressive episode); G25.2 (specified forms of tremor); G40.9 (unspecified epilepsy); self-assessment of moderate anxiety/depression
EGC-13§	arr 16p11.2(29,563,365–30,085,308)x1	0,6	30	15	F, 50y	Professional secondary	43.8	Self-assessment of severe anxiety/depression and problems with daily duties
EGC-14[2]	arr 16p11.2(29,502,984–30,085,308)x3	0,6	33	15	F, 24y	Basic	19.1	Self-assessment of severe anxiety/depression and problems with daily duties
EGC-15[2]	arr 16p11.2(29,502,984–30,195,224)x3	0,7	38	17	F, 20y	Basic	17.6	G40 (epilepsy); Q21 (congenital malformations of cardiac septa); self-report of problems with daily duties
EGC-16§	arr 16p11.2(29,563,365–30,037,994)x3	0,5	30	15	M,49y	Higher	22.2	Not reported
EGC-17	arr 17p11.2(14,041,963–15,390,352)x3	1,4	16	3	M, 33y	Professional secondary	23.9	G60.0 (hereditary motor and sensory neuropathy); self-report of severe pain/discomfort and problems with daily duties

Case ID	SNP array result	Length (Mb)	Ensembl genes	OMIM genes	Gender, Age	Education	BMI	Neuropsychiatric, congenital anomalies and features characteristic to the syndrome (if known)
EGC-18	arr 22q11(17,118,296–19,792,353)x3	2,7	100	37	M, 45y	Basic	22.4	Self-assessment of moderate pain/discomfort and problems with daily duties
EGC-19§	arr Xq28(154,192,808–154,582,606)x3	0,4	12	5	F,38y	Secondary	23.1	Q74 (congenital malformations of limb)

*The selection is based on the findings in clinical cohort of the current study, DECIPHER syndromes and those listed by Girirajan and Eichler [105].
Aberration breakpoints are reported as detected using the QuantiSNP algorithm. Diagnoses are given according to World Health Organization International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10). §Carriers identified as a result of the region-targeted analysis in GWAS samples. The following individuals have been described in more detail: [1]as case 32 in Publication III [51]; [2]in Publication II [129].

Table 7. Summary of clinical and molecular data of ID patients and general population individuals showing rare variants with unknown clinical significance.

Case ID	SNP array result	Length (Mb)	Origin	Ensembl genes	OMIM genes	Gender, Age	Phenotype
EID-19	arr 2p16.3(50,735,657–50,799,203)x1	0,06	Mat (aff)	2	1	M, 13y	Moderate ID, speech delay, hypotonia, cerebral atrophy, dysmorphic features
EID-20	arr 6q26(161,291,611–161,481,246)x3	0,19	Mat (aff)	2	1	M, 9y	Moderate ID, speech delay, dysmorphic features, cryptorhid testis
EID-21	arr 7p15.2(27,101,839–27,254,061)x1	0,15	Not determined	14	12	F, 36y	Mild ID, speech delay, growth retardation, dysmorphic features, conductive hearing loss
EID-22	arr 11q13.3q13.4(70,631,298–70,746,443)x3	0,15	Not determined	0	0	M, 9y	ID, behavioural problems (hyperactivity), epilepsy, dysmorphic features
EGC-20§	arr Xp22.31(6,198,923–7,980,930)x4	1,8	Not determined	10	4	F, 47y	Higher education; self-assessment of moderate anxiety/depression and pain/discomfort
EGC-21§	arr Xp22.31(6,410,623–8,104,888)x3	1,7	Not determined	9	4	F, 20y	Secondary education; G35 (multiple sclerosis)
EID-23	arr Xp22.31(6,456,825–8,095,053)x2	1,6	Mat (aff)	8	4	M, 10y	Mild ID
EGC-22§	arr Xp22.31(6,456,825–8,095,053)x3	1,6	Not determined	8	4	F, 42y	Professional secondary education; F32.2 (severe depressive episode without psychotic symptoms); self-assessment of moderate anxiety/depression
EGC-23§	arr Xp22.31(6,456,825–8,095,053)x3	1,6	Not determined	8	4	F, 21y	Secondary education; neuropsychiatric or congenital anomalies not reported
EGC-24	arr Xp22.31(6,468,166–8,095,053)x2	1,6	Not determined	7	3	M, 31y	Secondary education; neuropsychiatric or congenital anomalies not reported
EGC-25	arr Xp22.31(6,468,166–8,095,053)x2	1,6	Not determined	7	3	M, 47y	Basic education; neuropsychiatric or congenital anomalies not reported
EGC-26	arr Xp22.31(6,490,342–8,071,437)x4	1,6	Not determined	7	3	F, 55y	Secondary education; F32 (depressive episode); self-assessment of severe anxiety/depression
EGC-27	arr Xp22.31(6,468,166–8,095,053)x3	1,6	Not determined	7	3	F, 27y	Secondary education; G40 (epilepsy); self-assessment of moderate anxiety/depression
EGC-28§	arr Xp22.31(6,517,158–8,095,053)x3	1,6	Not determined	7	3	F, 83y	Higher education; neuropsychiatric or congenital anomalies not reported

Aberration breakpoints are reported as detected using the QuantiSNP algorithm. Diagnoses are given according to World Health Organization International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10). Except family EID-21, only index patients are listed in the table. §Carriers identified as a result of the region-targeted analysis in GWAS sample

5.1.1.1. Duplications in 7q11.23 are causative for a novel distinctive syndrome (Publication IV)

State of the art.

The 1.5 Mb microduplication of 7q11.23 encompasses 26 genes and is reciprocal to the recurrent deletions responsible for Williams-Beuren syndrome. Williams-Beuren syndrome (OMIM #194050; reviewed in [204]) is a multi-system developmental disorder which is esitimated to occur with approximate population frequency of 1 in 10,000. The disease is a distinctive contiguous-gene syndrome that was first described already in the nineteen sixties by J.C. Williams [205], A.J. Beuren [206] and is mainly characterized by congenital cardiovascular malformations (most frequently supra valvular aortic stenosis), hypercalcaemia and a characteristic „elfin-like" profile of facial dysmorphisms in childhood. Although ID is common, verbal skills of Williams-Beuren patients are well-preserved. Accompanied by hypersocial behaviour, overfriendliness and easy interaction with other people, these traits display a very specific well-recognizable cognitive and behavioral profile. The Williams-Beuren syndrome deletion is flanked by highly similar sequences of duplicated DNA and arises through unequal meiotic recombination [207, 208]. Because of the genomic architecture, this interval on the long arm of chromosome 7 is prone to other rearrangements and reciprocal duplications mediated by the same blocks of LCRs have been anticipated to exist, but until recently had not emerged as a recognizable syndrome by phenotype-based diagnostics. Only in 2005, Somerville and colleagues described the first patient with *de novo* duplication of the Williams-Beuren region [209], which was quickly followed by a few additional cases detected as a result of whole-genome CNV screening of cohorts with ID and other neuropsychiatric diseases [210–213]. Although speech and language impairment was a common manifestation in majority of these patients, the sparse data and incomplete penetrance on these initial case reports did not allow one to define the exact clinical consequences of this chromosomal imbalance.

Findings in the current study

In the intellectual disability cohort, a proband of the family EID-6 (**Figure 3**; **Table 5**) who exhibited global DD accompanied by severe speech delay, autistic features, ADHD and episodes of severe aggression was found to carry a 1.4 Mb duplication in the chromosome region 7q11.23. The duplication was inherited from his father (I:2, **Figure 4**), who was recorded to have cognitive and language impairment, as well as problems with aggressive behaviour, and segregated also to the proband's younger brother (II:4) with global DD at the age of 6 months.

No carriers of the duplication were found amongst 7626 investigated general population individuals.

To describe the novel duplication syndrome associated with the 7q11.23 duplication, collaborative effort by different cytogenetic centres in Europe collected and evaluated in a standardized way 12 probands, two siblings and

seven parents carrying identical 1.5 Mb duplications, exactly reciprocal to the Williams-Beuren syndrome critical region. Patients were analyzed using various genome-wide screening platforms and the presence of other potentially pathogenic CNVs was excluded in all cases. The proband EID-6 and his younger brother are referred as patients 1 and 8, respectively, in this paper by Van der Aa *et al.* [181]. In total, thirteen unrelated carriers were found amongst 5130 ID patients, suggesting that the 7q11.23 duplication can explain 0.25% cases of idiopathic ID. Due to the heterogeneous ethnic background of the samples, it is not possible to provide a reliable estimate of the general population frequency in the current study. However, assuming the prevalence of ID to be 2–3% of the general population in developed countries, a reasonable estimate of the frequency of the duplication is 1/13,000 – 20,000 [181]. The duplication was not detected in the Estonian general population cohort, which further confirms very low population prevalence and likely pathogenic effect.

The evaluation of data in the current patient cohort suggested that the clinical phenotype of the 7q11.23 duplication syndrome (OMIM #609757) is milder, less distinct and more variable than that of the Williams-Beuren syndrome. In agreement with previous reports, severe language delay (either expressive, receptive or both) was seen in all patients and presented the most characteristic feature of the syndrome. Also, a deficit in cognitive and/or social abilities was a predominant trait – 11 out of 14 patients met the criteria for ID, and 6 out of 14 were diagnosed with autism or ASD. This indicates that contrary to Williams Beuren syndrome, language and social skills are the most severely affected aspects of cognitive functioning in reciprocal duplication patients. Additional recurrent findings associated with the duplication include neonatal period hypotonia (8/14), joint laxity (3/14), epilepsy (2/14), abnormal findings in the brain MRI (5/7) and increased incidence of other congenital malformations. For the first time, our study described a facial phenotype associated with this duplication, including a high broad nose, straight eyebrows, a thin upper lip, deep-set eyes, a short philtrum and a prominent forehead. Intriguingly, some of these dysmorphic features are in direct contrast to facial dysmorphisms seen in Williams-Beuren syndrome patients. Moreover, the dysmorphic profile was retrospectively recognizable in previously published 7q11.23 duplication patients. Photographs of patients with 7q11.23 duplication presenting characteristic dysmorphisms is provided as Figure 2 in Publication IV [181].

Discussion

Amongst genes within an imbalanced locus, *Elastin* (OMIM #130160) has been considered to be responsible for supra valvular aortic stenosis in Williams-Beuren patients [214], and might contribute to the joint hyperflexibility in duplication patients [181], despite that preliminary analyses in human skin fibroblasts have shown no direct haploinsufficiency-caused change in its expression levels [161]. The *GTF2IRD1* (OMIM #604318) and *GTF2I* (OMIM

#601679), members from the general transcription factor family, near the distal breakpoint, have been shown to contribute most to the cognitive deficit and craniofacial features of Williams-Beuren patients [215–218]. The contrasting traits in language development, behavioural, and facial profile seen in patients with reciprocal imbalances indicate the presence of dosage-sensitive genes within the 7q11.23 chromosome interval. No definite candidates establishing these mirror effects have yet been found in humans, and in case of some genes in the region, dosage compensation mechanisms have been demonstrated [219]. However, social interactions are increased in mice hemizygous to the *GTF2I* [220], while the duplications of this gene have recently been associated with autism and anxiety disorder [221, 222].

In the EID-6 family, interestingly, the 7q11.23 duplication was not present in the oldest, similarly affected son. Determination of haplotypes in the aberrant region also excluded the possibility of copy-neutral structural rearrangement in this patient. To exclude the possibility of tissue mosaicism, the analysis was repeated with DNA extracted from a skin biopsy specimen, which gave the same outcome. Although all three duplication carriers presented the characteristic phenotype of the syndrome, the cognitive impairment of the proband was more severe than commonly seen in patients with the single copy gain of 7q11.23 [181, 223]. Thus, additional yet undetermined single-gene mutation or other factor might contribute to the ultimate clinical phenotype in this family.

Conclusive statement.
Although duplication syndromes in general have been considered milder, and more challenging to diagnose than the corresponding microdeletion syndromes due to more heterogeneous clinical outcome, our study by Van der Aa *et al*. demonstrates the power of the "genotype-first" approach to characterize of recurrent genomic disorders, allowing the authors to suggest a novel clinically recognizable duplication syndrome [181].

5.1.1.2. Gene dosage at the chromosome locus 16p11.2 is associated with neuropsychiatric disorders and mirror phenotypes on BMI (Publications II and III)

State of the art
The short arm of chromosome 16 has been one of the most actively duplicated regions in human autosomes. In the course of recent hominoid evolution, rapid integration of segmental duplications have generated complex genomic structure enriched in highly homologous and repetitive sequence blocks in 16p. These act as a substrate for intrachromosomal NAHR and predispose the region to recurrent structural rearrangements [224–226]. Five regions on the proximal short arm of chromosome 16 have been defined as „hotspots" to genomic imbalances of clinical relevance and associated with neuropsychiatric phenotypes. All below-mentioned imbalances are mediated by different LCRs

and should be considered as different syndromic entities in the light of current knowledge. Distinctive breakpoints that mediate these recurrent rearrangments are indicated on **Figure 7** and numbered from telomere to centromere as breakpoints BP1 to BP5.

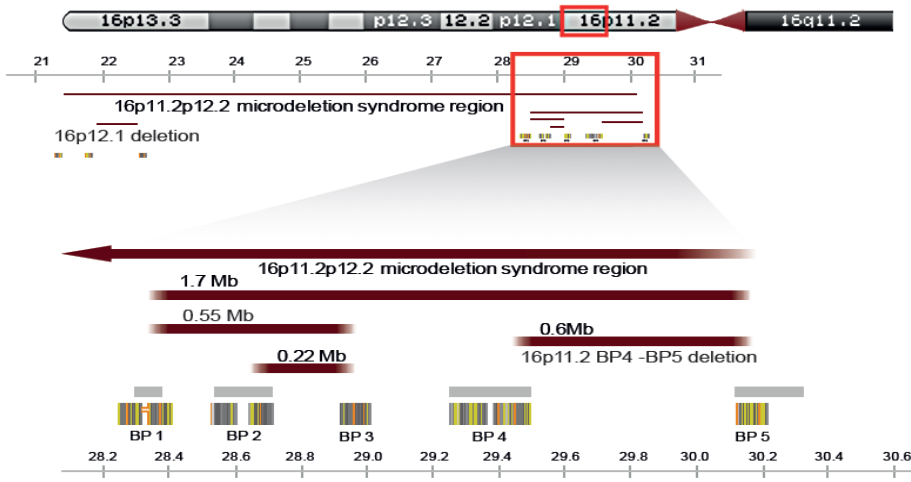

Figure 7. Genomic locus of the 16p11.2. The extent of neurodevelopment-associated genomic losses and gains in 16p11.2 are schematically pinpointed with bordeaux bars, while grey bars and striped blocks indicate intervals of recurrent polymorphisms reported in the Database of Genomic Variants and stretches of LCRs, respectively. Recombination hotspots that act as mediators of clinically relevant CNVs are termed from telomere to centromere as breakpoints BP1 to BP5.

(i) The distalmost recurrent 1.5-Mb microdeletions and –duplications were first reported in 16p13.11 as predisposing factors to autism and ID [227]. Followed by the comprehensive evaluation of their contribution to cognitive impairment [228, 229] and idiopathic epilepsies [56], both genomic gain and loss of 16p13.11 are currently considered as a susceptibility factor for neurocognitive disorder rather than sufficient in itself to cause clinical phenotype. (ii) A 500 kb recurrent microdeletion on 16p12.1 has been suggested to act as a risk factor for neurodevelopmental phenotypes, although the precise clinical impact of this deletion has remained vague [30, 230]. (iii) Large deletions that encompass the 16p11.2 interval have been reported between telomeric LCR at the position 21.4 Mb and variable proximal breakpoints (BP4 or BP5). Referred to as 16p11.2-p12.1 microdeletion syndrome this distinctive disorder is characterized by subtle facial dysmorphisms accompanied by ID, delayed speech development, feeding difficulties and recurrent ear infections [231–233]. (iv) Deletions of 220 or 550 kb containing the *SH2B1* (OMIM #608937) gene in the distal part

of 16p11.2 are mediated by BP1 and, respectively, BP2 or BP3, and have been reported in a few individuals with severe early-onset obesity and variable degrees of developmental delay [50, 234]. (v) Proximal 600 kb recurrent microdeletions and -duplications that are defined by BP4 and BP5 and encompass 28 genes in the 16p11.2 were initially reported to be markedly frequent in cohorts of autism spectrum disorders and evaluated to explain as many as 1% of autism cases [128]. The extent in which deletions and duplications in 16p11.2 account for the total burden of idiopathic autism spectrum disorders has since been reduced to 0.5% and 0.3%, respectively [235]. However, with the population prevalence of about 1/2000 [129], the locus is one of the most frequent known causes of neurodevelopmental disorders, and the 16p11.2 phenotypic spectrum has been extended by several other clinical traits [51, 60, 129, 130, 236–247]. The majority of these publications are based on the clinical data of a limited number of patients and do not always provide formal associations with the rearrangement. The great diversity of the described features, together with consistent reports of asymptomatic transmitting parents and apparently „normal“ control individuals have led the clinical community to question the essential phenotypic impact and penetrance of the recurrent 600 kb deletion and duplication in 16p11.2.

Findings in the current study

In the Estonian clinical cohort, a 5-year old patient EID-13 (**Table 5**) was detected carrying a 0.6 Mb BP4-BP5 deletion in 16p11.2. One identical deletion and two reciprocal duplications (EGC-12,14,15, **Table 6**) were detected in the initial, and one deletion and one duplication (EGC-13, 16, **Table 6**) in a follow-up cohort of the general population. Contrary to the paediatric patient EID-13, with mild ID, speech delay, behavioural problems and normal growth parameters, the adult deletion carriers presented BMI of 35.9 and 43.8 kg/m^2, which according to the WHO are classified as class II and class III obesity (http://apps.who.int/bmi). Inversely, one duplication carrier was underweight (BMI 17.6), one on a lower normal weight level (BMI 19.1) and a third with a normal BMI (22.2). Three out of 5 individuals with reciprocal 16p11.2 imbalances had only elementary education, 3 of the 5 reported problems with daily living and 3 had depression according to the EGC UT standard questionnaire. EGC-12 and EGC-15 were also diagnosed with epilepsy.

During the last years, the clinical phenotype of heterozygous deletions and duplications in 16p11.2 has been further studied in large cohorts by ourselves and others. To obtain the initial association of 16p11.2 deletions, patients with ID/DD and congenital malformations from eight cytogenetic centres in France, Switzerland and Estonia (n=3947) were analyzed in parallel with small cohorts of obese patients with cognitive deficit and/or MCA from France and the United Kingdom (n=312). As a surprising result, the frequency of the deletion was found to be significantly higher in the obese cohort (9 carriers, 2.9%) than in ID cohort (22 unrelated carriers, 0.6%; $p=2.2\times10^{-4}$, Fisher`s exact test), while the

latter frequency was consistent with those previously published in similar cohorts [128, 130, 248, 249]. Moreover, regardless of initial ascertainment of carrier, the clinical data revealed in all instances an age-dependent penetrance of adiposity – the obesity phenotype was strongly expressed in adults, and showed more variable espressivity in childhood (**Figure 8**).

All clinical cohorts taken together, the data revealed a possible direct association of 16p11.2 deletions with obesity, which is independent of individuals' cognitive functioning. To further specify the relationship between obesity, neurodevelopmental phenotypes and 16p11.2 deletion with high confidence, data from Swiss [250], Finnish [251], Estonian [196] general populations (in total 11 856 individuals), and five different extreme obesity cohorts (n=3844) [252–254] were combined in an overall case-control association analysis. The 16p11.2 deletion was absent in healthy non-obese European individuals, but supplemented the analysis with 19 carriers which strongly associated this heterozygous deletion with obesity ($p=5.8\times10^{-7}$, Fisher's exact test; OR=29.8, CI95%=[3.9,225]), as well as morbid obesity ($p=6.4\times10^{-8}$; OR=43.0, CI95%=[5.6,329]).

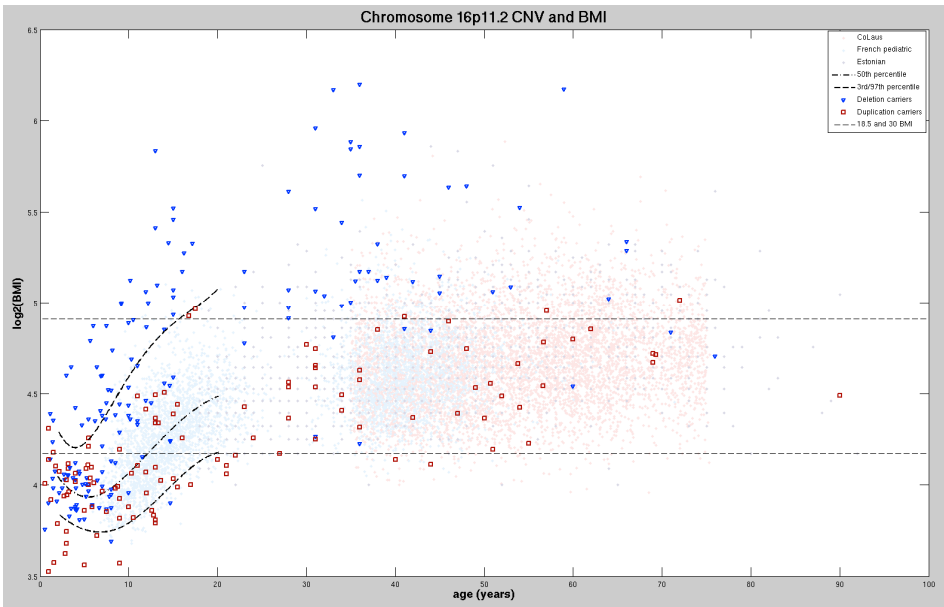

Figure 8. Dependence of BMI on age in patients with reciprocal 16p11.2 imbalances and corresponding general population individuals. Broken bolder lines denote 3rd, 50th and 97th BMI percentiles, finer lines correspont to cut-off thresholds for underweight and obesity in adults (BMI 18.5 and 30, respectevely). Red squares represent 16p11.2 duplication carriers and blue triangles deletion carriers. BMI data from three reference populations have marked with pink (Swiss), blue (French) and violet (Estonian) dots. Modified from [51].

Investigation of the segregation in 16p11.2 families showed that *de novo* incidence of the deletion is consistent with previous reports of cohorts with developmental delay and congenital anomalies [130]. In case of parental transmission, both maternal (n=11) and paternal (n=4) inheritance of the deletion was observed and all first-degree relatives carrying a deletion were also obese [51].

In search of converse clinical manifestations, a similar strategy was used to detect reciprocal duplications in the chromosome region 16p11.2. As a result, a total of 138 duplication carriers were identified from different population-based (n=58,635), neurodevelopmental (n=31,424), psychiatric (n=1080) and obesity/underweight (n=3544) cohorts. Consistent with the previous association with psychiatric conditions [128, 246], comparing their frequency in European general populations [196, 250, 251] showed significantly higher prevalence in DD/ID ($p=4.23\times10^{-13}$; OR=4.4; CI95%=[2.9,6.9]), and in schizophrenia/bipolar disorder cohorts ($p=3.6\times10^{-3}$; OR=7.0; CI95%=[1.8,19.9]). To assess whether the gene dosage effects accountable for obesity in deletion carriers, may in an opposite manner influence the body weight of duplication carriers, we compared the BMI distribution in all carriers of the duplication for whom anthropometric measures were available (n=105). The testing also took into account gender, age and ethnic background, as influencing factors, and resulted in a strong association of the 16p11.2 duplication with lower postnatal weight (mean Z-score –0.56; $p=4.4\times10^{-4}$) and BMI (mean Z-score –0.47; $p=2.0\times10^{-3}$). Thus, the risk of being clinically underweight was 8.3-fold higher for adult carriers of the 16p11.2 duplication (BMI <18.5; $p = 1.53\text{x}10^{-10}$) [129].

Separate analyses in neuropsychiatric and non-medical cohorts showed significantly lower weight with a similar effect size in both groups, although the proportion of individuals meeting criteria for being underweight (BMI ≤–2SD) was higher amongst neurocognitive patients (p=0.017). Interestingly, stratification by age and gender showed that these factors play a role in the expressivity of the phenotype. When all cohorts were combined, the relative risk of underweight was as high as 23.2 for adult males ($p=4.6\times10^{-11}$; CI95%=[9.1,59.3]), while only 4.7 for females ($p=9.9\times10^{-4}$; CI95%=[1.9,11.8]; gender difference p=0.0168). Moreover, the overrepresentation of males in clinical cohorts, as well as stronger impact on body weight in male patients with ID/DD suggests that men are more likely to present severe phenotype caused by the 16p11.2 duplication. By contrast, the higher representation of females compared to males was noticed in the general population (p=0.035), and amongst transmitting parents ($p=5.53\times10^{-4}$), that further confirms the reduced number of male duplication carriers in non-medically ascertained cohorts [129].

In their study two years ago, Shinawi and colleagues observed the link between autism and macrocephaly in 16p11.2 deletion patients, whereas duplication carriers presented microcephaly and an elevated risk of psychotic conditions [236]. Notably, the association of opposite alterations in head size with genomic gain (mean Z-score –0.89; $p=7.8\times10^{-6}$) and loss (mean Z-score +0.57; $p=1.79\times10^{-5}$) was validated in our study (**Figure 9**). The positive

correlation between OFC and BMI in both duplication (ρ=0.37; $p=2.65\times10^{-3}$) and deletion (ρ=0.42; $p=1.9\times10^{-5}$) carriers indicates that a related underlying molecular mechanism may exist between these traits and allow one to hypothesize that these associated conditions may represent opposite states at different ends of the same neurobehavioural continuum [129].

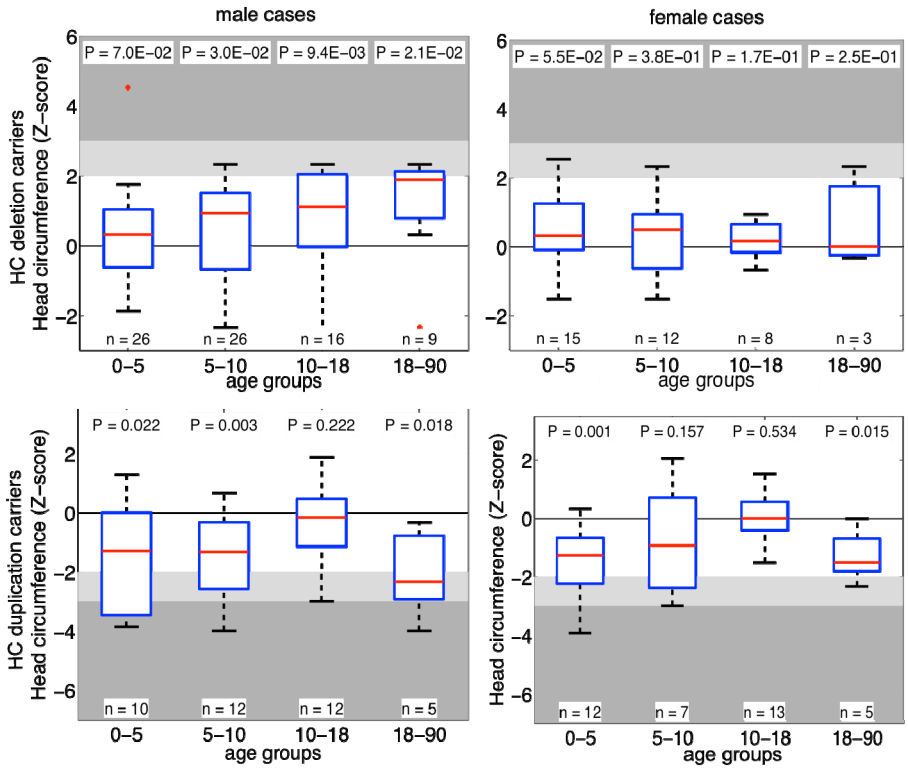

Figure 9. Effect of the 16p11.2 deletion and duplication on head circumference. Z-score values of head circumference deletion (top panels) and duplication (bottom panels) carriers stratified by age group (in years). Boxplots represent the 5th, 25th, median, 75th and 95th percentile for each age group. The light and dark grey backgrounds represent the 2nd and 3rd standard deviation, respectively [129].

Discussion

Pathological fluctuations in body weight are considered a major issue because of their severe consequences on the health. The epidemic increase of adiposity in the modern „obesogenic“ environment has made studying inherited variants and associated regulatory mechanisms of energy control a high priority. One factor that is likely to modulate susceptibility to the hedonic effect of food and thus contributes to weight variation has been suggested to be a genetically

determined difference in appetite and satiety [255]. As a confirmation of this hypothesis, family-based studies have shown that genetic factors account for 40–70% of the variation in BMI (summarized in [256, 257]), and monogenic forms of obesity described so far disrupt satiety mechanisms and regulation of appetite in the brain [258, 259]. Also the correlation between obesity and impairment in cognitive functioning is well known [260–263] and several obesity-linked genomic loci have been first characterized in individuals with impaired cognitive functioning [50], suggesting that related molecular pathways might be involved in these conditions. Although these observations indicate that human adiposity might be an inherited neurobehavioural disorder, the molecular factors and central regulators of eating behaviour are poorly understood. Even less is known on the abnormalities underlying anorexia and the clinical manifestation of being underweight. During the last years, GWAS studies have revealed tens of common genetic variants as predisposition factors to obesity [264–268], yet all of these have a small effect size, poor predictive power and explain altogether less than 2% of BMI variation in humans [269]. This further challenges the popular „common disease, common variant" hypothesis. Although the exact extent to which CNVs might contribute to the etiology of obesity is not clear, recent population-studies have demonstrated that common CNVs are unlikely to contribute greatly to the genetic basis of common human diseases [2, 270]. Alternatively, cohorts with extreme manifestations of common traits may have a higher frequency of rare variants with strong effect and provide valuable improvement in initial power for identifying loci responsible for missing heritability in obesity and other complex diseases [252, 271]. Our studies by Walters *et al*. and Jacquemont et al. of the 16p11.2 genomic interval examplify the power of a two-step strategy in the association of rare variants with complex traits, and show how the initial discovery stage in small well-phenotyped cohorts combined with a targeted follow-up association analysis in large case-control and population cohorts may improve the likelihood of discovering new variants and identifying phenotypes that are not biased by pre-existing ascertainment criteria. These large-scale analysis results demonstrate the burden of rare variants that exert strong effects in complex diseases [51, 129].

The higher frequency of 16p11.2 deletions in the current cohort recruited for both obesity and ID (2.9%), compared with cohorts ascertained for either phenotype alone (0.4% and 0.6%, respectively), confirms its involvement in both etiologies, and adds further evidence to the strong correlation observed between these two phenotypes. Possible explanations for this relationship include the involvement of related neural circuits, or different outcomes of the same set of neurobehavioural disorders with complex pleiotropic effects [51]. Moreover, although the evidence was not sufficient for formal associations, low food intake or selective and restrictive eating was recurrently reported by clinicians in 16p11.2 duplication carriers, and the opposite behavior, hyperphagia, in deletion carriers. This further indicates that dysregulation of control mechanisms involved in eating behavior might be responsible for reciprocal

extreme BMI phenotypes [129]. Preliminary gene expression analyses using lymphoblastoid cell lines and adipocytes showed that transcript levels in 16p11.2 deletion and duplication patients correlate positively with genomic dosage for genes mapping within the imbalanced interval, but not in flanking regions. This observation suggests that opposite phenotypes are likely caused by the genes within the 16p11.2 region that have an impact on pathways involved in dosage-sensitive regulation of energy balance. An altered copy number of these genes might result in the opposite effect on head size and through dys-regulation of central satiety and food intake control give rise to obesity or underweight [129]. As functional evidence, the reciprocal effect of genomic loss and gain on OFC and brain architecture has recently been showed on mice harboring deletion or duplication of the chromosomal region corresponding to the human 16p11.2 [272]. Also recent study in zebrafish embryos have revealed that overexpression and supression of the human *KCTD13* (OMIM #608947) within the 16p11.2 imbalanced interval cause micro- and macrocephaly [176]. In combination with gene-specific deletion of the *KCTD13* detected in a single autistic patient, these findings allow the authors to suggest that this gene might be a major driver for the neurodevelopmental phenotypes associated with the CNVs at 16p11.2. Moreover, two other transcripts, *MAPK3* (OMIM #601795) and the *MVP* (OMIM #605088) significantly increased the expressivity of the phenotype in both directions, thus pinpointing a likely epistatic contribution by different genes within the 16p11.2 locus [176].

In summary, the causal link of the 600 kb region at human 16p11.2 with a highly penetrant form of obesity and pathological leanness, as well as a variety of neuropsychiatric conditions, provides a unique opportunity to explore the molecular pathways underlying the central regulation of energy balance and its relationship with brain disturbances. Three different types of CNV-driven mechanisms might play a role in the etiology of 16p11.2 syndromes: (i) the region likely contains dosage-sensitive regulators of energy balance, for which the altered copy number results in opposite manifestations on body weight, eating behavior and head circumference; (ii) unmasking of recessive mutations or functional polymorphisms in certain genes by hemizygosity could explain the presence of some reproducible features, e.g. vertebral malformations, epilepsy or paroxysmal dyskinesia, only in a portion of 16p11.2 deletion carriers (discussed in paragraph 2.5.2); (iii) the presence of „double-hit“ CNVs or other modifying factors might be responsible for the eventual inter-individual phenotypic variability amongst 16p11.2 patients.

Conclusive statement

Our studies of the 16p11.2 genomic interval have demonstrated the potential importance of rare variants with strong effect in complex neurobehavioral disease, and highlighted successful strategies for discovering formal phenotypic associations of rare structural variants [51, 129].

5.1.1.3. A variable spectrum of phenotypes is associated with deletions and duplications at the chromosome locus 15q13.3 (Publication I and unpublished data)

State of the art.

Another highly unstable locus in the human genome maps to the centromeric long arm of chromosome 15, where complex sets of LCRs (known as BP1 to BP6) give rise to several types of recurrent rearrangements. An approximately 1.6 Mb recurrent deletion mediated by breakpoints BP4 and BP5 has been considered as a susceptibility factor for different forms of epilepsy, ID and autism, but also linked to a wide range of other neurocognitive phenotypes [58, 59, 131, 132, 273–277]. Analogous to the aforementioned interval 16p11.2, the 15q13.3 microdeletion syndrome (OMIM #612001) is relatively frequent and characterized by incomplete penetrance and remarkable variability in phenotypic expression. The imbalanced interval encompass eight genes, amongst which the altered dosage of the *CHRNA7* (OMIM #118511) has been considered as causative for the neurodevelopmental features in the 15q13.3 deletion syndrome [278–280]. Reciprocal duplications in 15q13.3 might pose a risk for autism and expressive language impairment, but have so far not been considered as clearly pathogenic variants [132, 280, 281].

Findings in the current study

In the current cohort of ID patients, a deletion overlapping with the 15q13.3 microdeletion syndrome region was detected in a sporadic male patient (EID-11, **Table 5**; described greater detail in [132]), and was also found to segregate with ID phenotype in another family (EID-12, **Table 5**). Except for the polymorphic duplications with approximate genomic coordinates of 29.8 and 30.4 Mb (recently termed also as „small microduplications of *CHRNA7*" by Szafranski and colleagues) [280, 282], the rearrangements in chromosome band 15q13 were extremely rare in the Estonian general population. Only two individuals were found to carry syndromic BP4-BP5 microdeletion in the 15q13.3 (EGC-5, 6, **Table 6**) amongst 7626 analyzed samples, and one carrier was harboring a reciprocal duplication (EGC-4, **Table 6**). The sole duplication carrier found in the current study is insufficient for drawing any conclusions, but since neuropsychiatric problems were not reported, our finding further supports the idea that duplications in this region do not result in clinical consequences or are not fully penetrant. Surprisingly, the 15q13.3 deletion carriers both presented overweight or obesity phenotype and were exclusively identified from a sub-group initially recruited as cases for the GWAS study of metabolic traits (n=880).

Discussion

Although the sample size is small, potential prevalence of 0.2% in individuals with metabolic diseases in the current study would be comparable to that of ID (0.3%), autism (0.2%) or schizophrenia (0.2%), while an order of magnitude higher than the proposed population frequency of 15q13.3 deletions [131, 283].

Whereas attention has mostly been paid to the role of the 15q13.3 interval in epilepsy and psychiatric diseases, alterations in growth parameters have so far not associated with this CNV. However, retrospective observation of clinical information in a few pre-reviewed articles where data about weight, height and head circumference was available showed repeating reports of weight gain and OFC above the 90th percentile in 15q13.3 deletion carriers [131, 273, 284]. Considering the close proximity of the Prader-Willi syndrome locus to 15q13.3, it is notable that both deletion carriers from Estonian ID cohort were decribed by clinicians prior to CNV analysis to have obese Prader-Willi syndrome like phenotype. Whereas most of the patients referred to the screening for genomic disorders are paediatric, three out of four deletion carriers in the current study were adolescents or adults. As examplified by the rearrangements in 16p11.2, the age-dependent differences in penetrance might be overlooked in case of some traits and the current knowledge is vague about the phenotypic dynamics of genomic disorders in adulthood. To the best of our knowledge, no phenotype data of the adult general population have yet been reported in association with the 15q13.3 rearrangements.

Conclusive statement
Although additional information is necessary to understand whether the alterations in body composition are associated with the genomic variants in 15q13.3, our data indicates that the recruitment of investigated individuals should not be biased by the phenotype of interest or age-restricted cohorts. Instead, large cross-population association studies with the recruitment on the basis of being a carrier of the variant under study should be pursued to correctly elucidate the clinical outcome of the CNVs.

5.1.1.4. A novel syndromic microduplication in Xq28 including the *Rab39B* (Publication I and unpublished data)

Findings in the current study
A novel likely pathogenic duplication in a complex gene rich region of Xq28 was detected in a male patient EID-18 (**Table 5**) who has mild ID, dysarthria, difficulties with socializing and mildly dysmorphic facial features. From the family history it was known that his mother had an early menopause. Unfortunately, she was not available for further investigation and it was not possible to determine whether the patient's rearrangement was inherited or arose *de novo*.

The fine-mapping of aberration boundaries by qPCR revealed a 500 kb duplicated region with breakpoints localized to the directly oriented highly-homologous sequences in the coagulation factor VIII genes *F8A1* (X:153, 767,829–153,769,529) and *F8A2* (X:154,264,943–154,266,643). The screening of 6628 genomes in the general population detected a single female carrier (EGC-19, **Table 6**) of a shorter duplication in the chromosome region Xq28 that overlaps only the distalmost part of the aberration identified in the ID patient.

Discussion

Although other recurrent duplications in Xq28 are known to be associated with ID phenotype [285–289], genomic imbalances in this particular interval locating distally from the recurrent copy number gain that had been identified by Vandewalle *et al*. [286], is yet sparsely described. Among seven encompassed genes (**Figure 10**), the found duplication affects two newly identified X-linked ID (XLID) genes: (i) the *CLIC2* (OMIM #300138), which missense mutation H101Q was recently predicted to be causative in a family with profound X-linked ID [290]; (ii) the *Rab39B* (OMIM #300774) that encodes a neuronal-specific small RabGTPase and has been identified as a novel XLID gene by mutation analysis [97, 291]. In addition to directly altering copies of genes affected by the aberration, the impact of CNVs on regulatory elements or position effect might also influence the gene expression. Therefore it is notable that in flanking region of less than 1 Mb, several known ID genes are present, including *MECP2* (OMIM #300005) *FLNA* (OMIM #300017), *RPL10* (OMIM #300847), *GDI1* (OMIM #300104), *IKBKG* (OMIM #300248), and *DKC1* (OMIM #300126). At least for *MECP2* and *GDI1*, dosage-sensitivity has been confirmed or considered [286, 292].

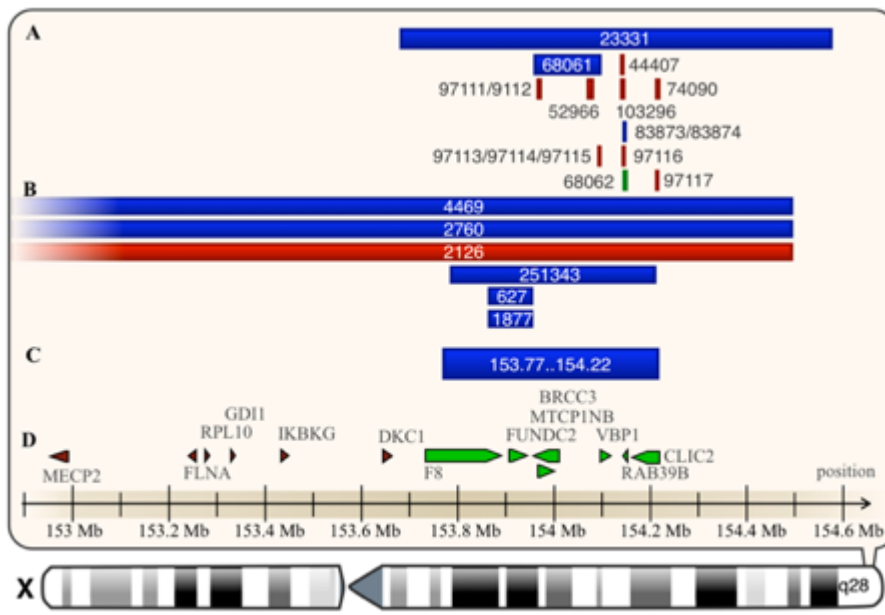

Figure 10. Schematic representation of the duplicated genomic region in Xq28 in patient EID-18. As of July 2010, reports in the Database of Genomic Variants (**A**) and DECIPHER database (**B**) overlapping with the current finding (**C**) are shown. Duplications are indicated by blue and deletions by red bars. Genes encompassed by the duplication (green), and known ID genes in the flanking region (brown) are given on the panel (**D**) [201].

The LCR *int22h-1* in intron 22 of the *F8A1* (OMIM #300841) is known to be involved in homologous recombination with more telomeric *int22h-2* [293], and most probably promoted NAHR-mediated genomic instability in our patient. The mechanism suggests that deletions and duplications in this region might be recurrent. However, at the time we initially suggested that this genomic gain is likely to be associated with ID [201], only a few cases with the current duplication were entered in the DECIPHER database. Although small non-pathogenic indels are common in Xq28, only one overlapping duplication (indicated as variation 23331 in **Figure 10**) has been reported in the Database of Genomic Variants. While the information provided by the DGV lacks gender and phenotype specifications, we analyzed this region in the Estonian general population with the purpose of better understanding the pathogenic effect of this genomic gain. A single duplication that overlaps only the distalmost, *CLIC2* encompassing, part of the ID-associated duplication was identified in a female individual who had reported no learning difficulties, nor neuropsychiatric disturbances. Since her X-inactivation pattern was assessed as random (the ratio 53%:47% between two alleles), spare evidence in favor of *CLIC2* as a potential contributor to the cognitive phenotype was challenged by this finding. The absence of *Rab39B* involving aberrations in non-ID cohort individuals, on the other hand, gives a strong reason to hypothesize that the duplications causing an overdose of this gene may present a novel region of a syndrome associated with mental disorders.

Conclusive statement
Since our report in 2010, two additional unrelated patients with similar clinical features have been found to carry identical duplications (Dr. G. Froyen, personal communication). Furthermore, our observations were recently confirmed by El-Hattab *et al*., who identified *int22h-1*/*int22h-2* mediated Xq28 duplications in three unrelated families with cognitive impairment and proposed that this duplication in Xq28 might be responsible for a novel X-linked ID syndrome [294].

5.1.1.5. A rare variant in Xp22.31 with uncertain clinical consequences (Publication I and unpublished data)

State of the art
Contrary to the above discussed genomic regions of 15q13.3 and 16p11.2, the comparative analysis of two Estonian cohorts indicated different pattern of clinical significance for recurrent duplications at Xp22.31. This PAR1 pseudoautosomal region flanking part on the short arm of human chromosome X is featured as highly unstable and interesting for several reasons. A series of historical duplication and inversion events during primate evolution has given rise to the *sulfatase*, *CD99 antigen*, *VCX*/*Y* gene clusters, and interspersed LCRs that mediate rearrangements both within Xp22.31 and between homologous

regions on chromosomes X and Y. The interval is further predisposed to rearrangements by showing the highest genome-wide concentration of the homologous recombination stimulating motif – a *cis*-acting 13-mer sequence that has been associated with approximately half of the recombination hotspots [295–297].

Investigations of genomic disorders on Xp22.31 have mainly been focused on deletions that cause steroid sulfatase deficiency and X-linked ichtyosis (OMIM #308100) [298–300], accompanied by ID, attention deficit hyper-activity disorder and social communication difficulties in some patients [301–305]. Although the reciprocal genomic gain has also been associated with cognitive disability [306–308], the frequency with which these duplications have been identified in healthy parents and population studies, leaves its impact on neuropsychiatric development unclear.

Findings in the current study

In the current study, the Xp22.31 duplication of 1.5 Mb was first identified in a male patient (EID-23, **Table 7**) and his mother, both having mild non-specific ID. The imbalance was not present in patient`s two healthy siblings, nor in any other family members. The determination of the haplotype structure in the aberrant region showed co-segregation of ID with a specific haplotype in this multi-member family. However, genomic gain of the Xp22.31 was detected also in four individuals in the initial general population cohort (EGC-26–29, **Table 7**) and in five additional carriers in the follow-up group (EGC-22–25, 30, **Table 7**).

Discussion

The detected population frequency of 0.13% is in concordance with the previously observed prevalence of the Xp22.31 duplication in healthy controls [308]. According to the EGC UT questionnaire, one male duplication carrier had completed basic education. All other individuals with Xp22.31 gain, one male and seven females, had finished at least high school. This did not lend support to the hypothesis that Xp22.31 duplication carriers in the general population might have borderline intellectual abilities. Similarly to our observations, it has recently been confirmed by two other large groups of Xp22.31 carriers [297, 309] that single-copy gain *per se* is insufficient for cognitive impairment. In the former comprehensive study, Liu *et al.* proposed that the Xp22.31 duplication may act as a predisposing factor to abnormal phenotypes, but according to the genome dosage model additional genomic alteration, either further gain of the same region or presence of another large CNV, is required for manifestating the disease [297]. Although two carriers of triplications in the Xp22.31 were identified in our study, they both were females and had no record of educational difficulties. Thus leaving open the correlation between further genomic gain and more penetrant or severe phenotype. Interestingly, both triplication and two of the duplication carriers had suffered from moderate to severe depression, more evidence that genomic gain of a

given interval may serve as a risk factor for neurobehavioral pathologies. Although no additional large CNVs were found in general population carriers of Xp22.31, nor in the family EID-23, the affected mother and her healthy daughter from the latter family demonstrated strongly skewed X-inactivation patterns (the ratios between two alleles 86%:14% and 80%:20%, respectively). Since the daughter did not carry the duplication and genes in the region of Xp22.31 escape X-inactivation due to sequence homology with the Y chromosome and the proximity of PAR1 [310, 311], this non-random inactivation may indicate the presence of a point mutation or some other secondary alteration on chromosome X, and further confirm that the duplication segment might be necessary but not sufficient to cause the phenotype. Analyzed female duplication carriers from the general population cohort showed random X-chromosome inactivation (38%:62% and 63%:37%).

Conclusive statement
In the current study, the copy number gain at Xp22.31 was considered of uncertain clinical significance due to small sample size and again clearly underlines the importance of large-scale association studies of well-characterized cases and controls to collect sufficient data for accurate phenotypic assessment of recurrent genomic gains at Xp22.31.

5.1.2. Non-recurrent rearrangements of clinical relevance

Ten ID families were found to have imbalances in genomic regions where pathogenic CNVs have been shown to be variable in size and to not share common breakpoints. Subsequent investigation of these genomic intervals in the general population revealed no overlapping CNVs that lending additional support to the univocal clinical relevance of these rearrangements.

A male patient EID-1 (**Table 5**) was identified as a carrier of a *de novo* 3.9 Mb deletion in the 2q37 monosomy region. In case EID-4 (**Table 5**), a deletion of 1.6 Mb in the 5q14.3 microdeletion syndrome region was identified. Two familial deletions involving the *FOXP2* (OMIM #605317) gene were detected in association with speech disorder (EID-7, 8, **Table 5**; decribed in detail in [203]). Two unrelated patients (EID-15, 16, **Table 5**) were diagnosed with 22q13 deletion syndrome, the latter harboring a cryptic unbalanced translocation (46,XX.ish der(22)t(11;22)(q25;q13.3) mat(N85A3-,11qter+)). A duplication of 400 kb in the *DMD* (OMIM #300377) gene, encompassing exons 45 to 51 and inherited from a healthy mother, was identified in a male patient with moderate ID (EID-17, **Table 5**). The muscular structure and function of this patient were completely normal as determined by electronmicroscopy. However, intellectual deficit of various degree is accepted as a common feature in a substantial proportion of patients with Duchenne muscular dystrophy (OMIM #310200). The cognitive impairment in these patients is likely caused

by disturbance of the expression of brain-specific products of the *DMD* [312–314], thus supporting the idea that this intragenic rearrangement might also contribute to our patient's phenotype.

Three individuals with ID were identified as carriers of cryptic imbalances in subtelomeric regions, either deletion (EID-10), or unbalanced translocations – EID-9 (46,XX.ish der(12) t(11;12)(q25;p13.3) pat(12pter-,11qter+)) and EID-5 (FISH analysis of this patient is pending). Phenotype data of all these patients is provided in **Table 5**. Since subtelomeric deletions in 12p and 16p and duplications in 5q were not listed as findings in phenotypically normal individuals (summarized in [315]), these imbalances were considered to be causative for phenotypic features in our patients.

In two EID families, aberrations were identified which were exclusively present in patients and encompass seemingly relevant, but yet sparsely described chromosomal regions.

5.1.2.1. A complex rearrangement of 2p25.1–p24.3 associated with severe ID (Publication I)

A proband EID-2 (**Table 5**) with severe ID, hypotonia, focal epilepsy, and behavioral problems, was identified as a carrier of a complex chromosomal rearrangement composed of two small deletions and separated by 3 Mb of two-copy genomic content in 2p25.1-p24.3. The aberration was inherited from his mother, who also has a moderate ID phenotype. Annotation of the aberrant region revealed among seven affected genes the *ASAP2* (OMIM #603817), a gene encoding an activator of small Arf-GTPase and the neuronal protein gene *KIDINS220* that controls neuronal development and memory formation [143, 316, 317]. Intriguingly, the region between two deleted areas is especially gene-dense and contains potential candidates, such as a neuron-specific Ca^{2+}-binding protein gene *HPCAL1* (OMIM #600207) and the neurotensin receptor 2 gene *NTSR2* (OMIM #605538) that may have an impact on the patient's clinical features. Further studies are, however, necessary to understand the exact genomic organization of this compex rearrangement and to discover whether this impacts the expression of those involved genes.

5.1.2.2. Microdeletions 3p11.2–12.1 and 7p21.1–21.2 associated with intellectual disability, short stature and clinical features suggesting Saethre-Chotzen syndrome (Publication I and unpublished data)

Findings in the current study

In the proband EID-3 (refer to **Table 5** and **Figure 11** for clinical description), two interstitial microdeletions of clinical relevance in 3p12.1-p11.2 and 7p21.2-p21.1 were detected. Neither of the imbalances was identified in his maternal

relatives. Data from the father was not available for analysis, and as the patient's existing genotypes in both deleted regions were consistent with the maternal ones, the origin of these aberrations remained unknown.

Genomic annotation revealed that both regions encompass several genes involved in processes essential for normal physical and intellectual development. A detailed overview of deleted genomic regions is given in **Figure 12**.

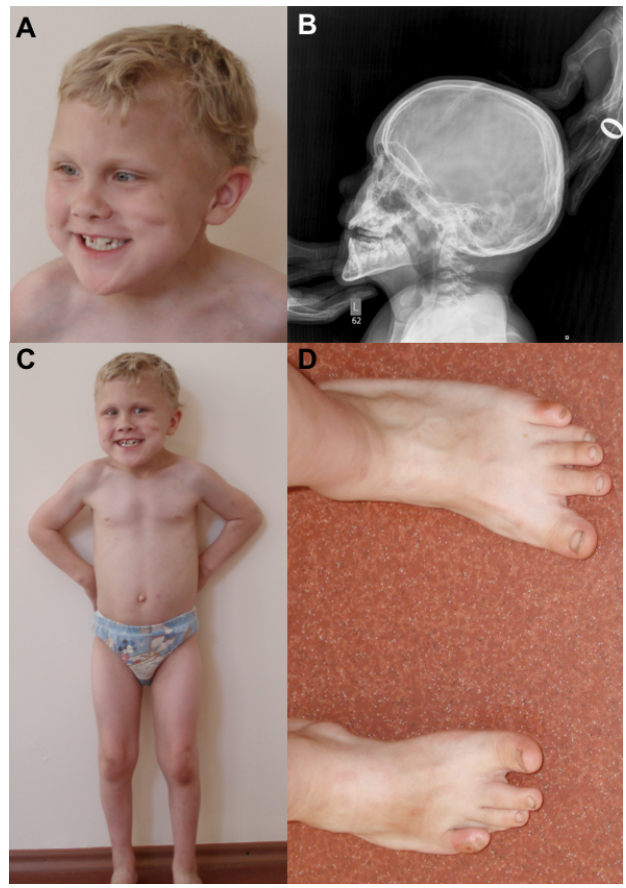

Figure 11. The patient EID-3 at the age of 7 years. A facial view of the patient; note the high forehead, prominent glabella, flat facial profile, upwardly slanting palpebral fissures, prominent and high cheeks, small and upturned nose and pointed chin (**A**). X-ray of the skull; note sclerotic sutures, flat facial skull, and asymmetric mandibula (**B**). Note wide and flat chest, pectus excavatum and protuberant abdomen (**C**). Note overriding toes, short and broad hallux, sandal gap of toes I–II and short V toes (**D**). Written permission to publish the photos of this patient was obtained from the family.

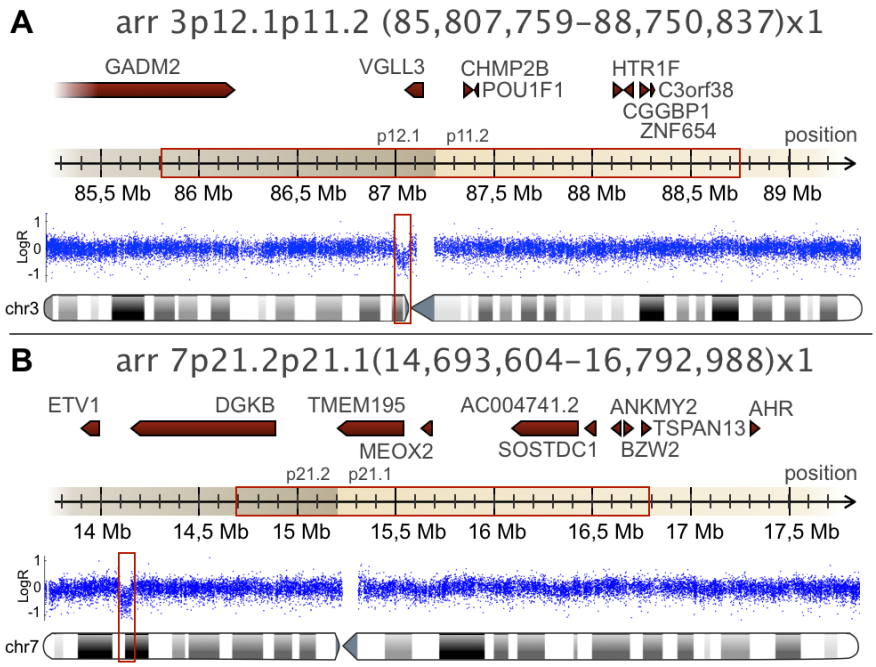

Figure 12. Schematic representation of hemizygous genomic regions detected in the patient EID-3. A 2.9 Mb deletion encompassing eight genes in chromosome region 3p12.1-p11.2 (**A**). A 2.1 Mb deletion that harbors nine genes in 7p21.2-p21.1 (**B**). Hemizygous regions are indicated by red boxes. Genes encompassed by the deletions and flanking genomic regions of 1 Mb according are shown as purple arrows.

The genomic loss in chromosome 7 has partial overlap with 7p21 microdeletion or Saethre-Chotzen syndrome (OMIM #123100) and could be considered to be an atypical short deletion of the syndrome. Amongst hemizygous genes in the patient, *MEOX2* (OMIM #600535) and *SOSTDC1* (OMIM #609675) have been linked with congenital anomalies using murine models and were considered to be the main contributors to our proband's skeletal deformations, ossification and midline defects, as well as tooth development abnormalities. In a second hemizygous region on a short arm of chromosome 3, *POU1F1* (OMIM #173110) has been shown to be associated in a dosage-dependent manner with combined pituitary hormone deficiency (OMIM #613038), growth failure, and intellectual disability [318–320].

The transcription levels of these three genes, as well as *BZW2*, *CGGBP1* (OMIM #603363), *CHMP2B* (OMIM #609512), and *C3orf38* of were analyzed in the current patient. In addition to the candidate genes mapping to the deleted regions, the expression status of the *TWIST1* (OMIM #601622) was investigated. This dosage-sensitive developmental regulator has been commonly

known as the major candidate for explaining the clinical phenotype of Saethre-Chotzen syndrome [321, 322]. Although the 7p21 microdeletion in our patient is located more distal to the critical region, leaving the *TWIST1* 2.4 Mb away from the proximal boundary of the aberration, his severe ID was accompanied by the clinical features coinciding with the Saethre-Chozen spectrum, including asymmetric, deformed skull with sclerotic sutures, dysmorphic facial phenotype, irregular crowded teeth, bifid uvula, skeletal deformations and hyperopia.

Due to the restricted tissue- and development-specific expression patterns of these genes (http://biogps.gnf.org), the mRNA levels of *MEOX2*, *POU1F1*, and *SOSTDC1* as well as *TWIST1* remained undetectable in the peripheral blood derived samples. The relative expression of other hemizygous genes was concordantly decreased in the proband's peripheral blood and lymphoblastoid cell line, whereas the mean expression levels of reference genes unaffected by deletions were not altered compared to control individuals. The relative expression levels of the investigated genes and samples is given in **Figure 13**.

Discussion

Although ID and short stature have commonly been seen in 7p21 microdeletion patients [323–325], the data suggests that our patient's severe impairment in cognitive functioning and growth failure has heterogeneous etiology and may be explained by the additive effect of two deletions. The genotype-phenotype correlation allows us to suggest two novel candidate genes, *MEOX2* and *SOSTDC1*, in the 7p21 microdeletion syndrome region. The haploinsufficiency of these genes may contribute to the patient's craniosynostosis, skeletal deformities, midline defects, and teeth abnormalities – features present in the clinical spectrum of the Saethre-Chotzen syndrome. Although the Saethre-Chotzen phenotype has been firmly associated with the haploinsufficiency of *TWIST1* [321, 322, 326], only about two-thirds of patients have been identified as carriers of *TWIST1* deletions or intragenic mutations. It has been conceived that the interruption of yet unidentified regulatory regions located 5' or 3' of the gene may lead to the syndrome by the position-effect in some patients. Also the strikingly variable expressivity of the phenotype in patients with *TWIST1* haploinsufficiency suggests the existence of additional modifying genetic factors, whose identification could further explain the mechanism of this disorder [324, 325, 327]. Of interest, the 7p21 microdeletion syndrome region includes two other genes, *MEOX2* and *SOSTDC1*, which encode proteins known as essential embryonal regulators in vertebrates, and are located in relative proximity to *TWIST1*. The dosage-dependent homeobox transcription factor encoded by *MEOX2* has been shown to function as a regulator of early mesodermal specification in the regions crucial for vertebrate head and bone development, suggesting that the human homologue may be involved in the pathogenesis of craniofacial and skeletal abnormalities [328–331]. As further evidence, Kirilenko and colleagues recently confirmed the concerted role of *Meox* transcription factors by showing the abnormal morphogenesis of branchial arches and the hypoplastic occipital bone in combined murine mutants for *Meox2* and its close homolog

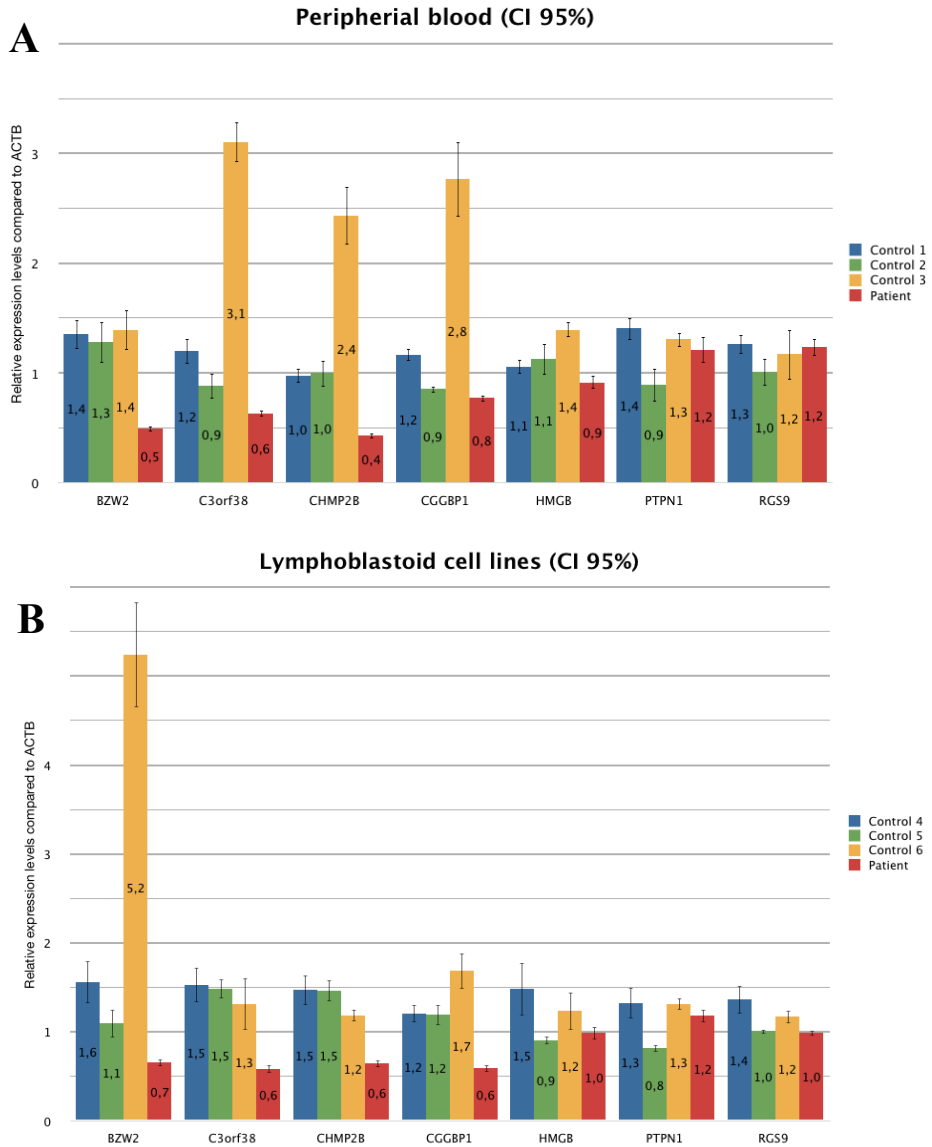

Figure 13. Gene expression analysis of hemizygous genes in peripheral blood (**A**), and lymphoblastoid cell line (**B**). The y-axis of the histogram represents relative expression levels compared with *ACTB*; the x-axis investigated candidate genes mapping to deletion regions (*BZW2*, *C3orf38*, *CHMP2*, *CGGBP1*) and reference genes (*HMGB*, *PTPN1*, *RGS9*). The height of the columns corresponds to average relative expression level, and error bars indicate the CI 95%. Non-overlapping intervals were considered to be significantly different.

Meox1 [332]. *SOSTDC1* encodes a bone morphogenetic protein antagonist, widely expressed in the epithelium and mesenchyme of the developing tooth germ. Its critical contribution to the control of tooth number and patterning via modulation of the Wnt signaling pathway has been confirmed in mice [333–337]. Although there is solid evidence that favors considering *MEOX2* and *SOSTDC1* to be the main contributors of the phenotypic features present in our patient, to the best of our knowledge there are as of yet no reports of deleterious mutations or a distinctive clinical phenotype associated with the haploinsufficiency of either of these genes in humans.

Although the transcriptional analysis of *MEOX2* and *SOSTDC1* was hampered by restricted tissue-specificity, the significantly reduced expression of other hemizygous genes in the proband uniformly indicates the down-regulating effect of imbalances on the transcriptional level. The latter is in concordance with the accepted knowledge that CNVs directly alter the mRNA levels of genes comprised by the aberration, and might also influence the expression levels of some non-hemizygous genes in the neighboring regions of about 1 Mb [160, 161, 165, 167, 169]. *TWIST1* is located outside of the flanking area of this size in our patient, yet its expression status at the transcript-level remained unknown, not permitting us to rule out the potential long-range effect of the structural rearrangement.

Conclusive statement.

We have reason to suggest that not only yet-to-be identified alterations of *TWIST1* but the haploinsufficiency of other genes in the 7p21 microdeletion syndrome region contribute to abnormalities in cranial, skeletal and dental development. Thus the targeted search for *TWIST1* alterations only might be inefficient when it comes to performing an accurate genetic diagnosis in patients presenting features resembling Saethre-Chozen syndrome.

6. SUMMARY AND CONCLUSIONS

The current study was the first comprehensive effort to investigate genomic causes of cognitive impairment and other complex phenotypes in Estonian individuals with the aim to:

(i) identify genomic rearrangements of clinical relevance
(ii) further investigate rare structural variants and associated phenotypic traits across two comparative cohorts
(iii) establish an effective workflow for state-of-the-art genetic diagnosis for Estonian patients with neurodevelopmental disorders.

The study presented benefits and opportunities provided by SNP genotyping analysis of well-characterized comparative cohorts in the diagnostics of complex disorders and reliable assessment of phenotypes associated with these rare variants.

The most important result of the study was the establishment of genetic diagnosis in 18 investigated families with idiopathic ID. The diagnostic yield of 23% in this first group of patients in Estonia is comparable with previous reports in other populations and further proves that whole-genome screening for genomic rearrangments is a reliable and effective tool in research and diagnostics. In addition to the clinical cohort, rare genetic variants with clinical impact was found in 19 Estonian general population individuals. In clinical genetics the implementation of whole-genome CNV analysis facilitates counselling of families and as of 2011, is provided as a routine diagnostic test for patients with developmental disorders by Tartu University Hospital. Our findings in the general population underline the need for more extensive genotype-phenotype correlation studies in reference individuals to establish formal genomic associations of complex traits, and emphasize the importance of adequate feedback to participants in biobanks that collect biological samples with the purpose of personalizing medical care.

In case of variants with very low population prevalence, large-scale multicenter efforts are needed for formal definition of novel genomic disorders. By participating in collaborative investigations, the core clinical phenotypes were established for two genomic loci. First, a novel distinctive duplication syndrome in genomic region 7q11.23 was described in the project led by Prof. Frank R. Kooy, University of Antwerp, Belgium. The study demonstrated the power of the "genotype-first" approach in characterization of previously unrecognized recurrent genomic disorders. Secondly, the reciprocal imbalances in the chromosome region 16p11.2 were associated with dosage-dependent mirror phenotypes in neuroanatomical traits, and this genomic interval was established as a promising model to investigate the central control of energy balance in the human body and its releation with neurobehavioural disorders. The project led by Prof. Alexandre Reymond and Prof. Jacques S. Beckmann, University of

Lausanne, Switzerland demonstrated how the initial discovery stage in small clinical cohorts combined with a targeted follow-up association analysis in large case-control and population cohorts improves the likelihood of discovering rare variants and identifying phenotype associations that are not biased by pre-existing ascertainment criteria.

In addition to regions with previously established clinical significance, four novel genomic aberrations that are likely involved in the pathogenesis of neurodevelopmental disorders were identified in Estonian individuals. One of these is a novel recurrent duplication syndrome in chromosome region Xq28, for which a further description is currently in progress. Another, an atypical short deletion in the 7p21 microdeletion syndrome region allows to suggest two novel candidate genes that might be involved in the etiology of skeletal deformities and tooth development abnormalities and act as additional modifiers in patients with the 7p21 microdeletion syndrome. A case report describing the patient has been submitted.

Arriving at reliable conclusions about the importance of rare variants with variable expressivity, incomplete penetrance and often controversial claims on their clinical significance (e.g. genomic regions 16p11.2, 15q13.3 or Xp22.31 in the current study) requires large amounts of data from cohorts with different ethnic background and phenotypic criteria. The current study included a family-wise investigated clinical cohort and large sample set of ethnically matching adult general population individuals. Genotype and phenotype association data generated on these cohorts is a valuable resource for the scientific community and available for future collaborations with the purpose of deciphering the clinical impact of rare variants and molecular mechanisms underlying genomic disorders.

In summary, the results of this study demonstrated the burden of rare variants with strong effect in cognitive disorders and other complex traits. Our experience gained by investigating different genomic loci underscores the importance of investigating large cohorts that are not age-restricted or biased by pre-existing ascertainment criteria to improve the detection of structural variation in whole-genome data and to arrive at reliable associations between rare genomic variants and clinical traits.

REFERENCES

1. A map of human genome variation from population-scale sequencing. *Nature* 2010, 467(7319):1061–1073.
2. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P *et al*: Origins and functional impact of copy number variation in the human genome. *Nature*, 464(7289):704–712.
3. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F *et al*: Mapping and sequencing of structural variation from eight human genomes. *Nature* 2008, 453(7191):56–64.
4. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L *et al*: Paired-end mapping reveals extensive structural variation in the human genome. *Science (New York, NY* 2007, 318(5849):420–426.
5. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W *et al*: Global variation in copy number in the human genome. *Nature* 2006, 444(7118):444–454.
6. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: Detection of large-scale variation in the human genome. *Nature genetics* 2004, 36(9):949–951.
7. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M *et al*: Large-scale copy number polymorphism in the human genome. *Science (New York, NY* 2004, 305(5683):525–528.
8. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G *et al*: The diploid genome sequence of an individual human. *PLoS Biol* 2007, 5(10):e254.
9. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT *et al*: The complete genome of an individual by massively parallel DNA sequencing. *Nature* 2008, 452(7189):872–876.
10. Piotrowski A, Bruder CE, Andersson R, Diaz de Stahl T, Menzel U, Sandgren J, Poplawski A, von Tell D, Crasto C, Bogdan A *et al*: Somatic mosaicism for copy number variation in differentiated human tissues. *Human mutation* 2008, 29(9):1118–1124.
11. Iourov IY, Vorsanova SG, Yurov YB: Chromosomal mosaicism goes global. *Mol Cytogenet* 2008, 1:26.
12. Mkrtchyan H, Gross M, Hinreiner S, Polytiko A, Manvelyan M, Mrasek K, Kosyakova N, Ewers E, Nelle H, Liehr T *et al*: Early embryonic chromosome instability results in stable mosaic pattern in human tissues. *PloS one* 2010, 5(3):e9591.
13. Bruder CE, Piotrowski A, Gijsbers AA, Andersson R, Erickson S, Diaz de Stahl T, Menzel U, Sandgren J, von Tell D, Poplawski A *et al*: Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *American journal of human genetics* 2008, 82(3):763–771.
14. Veenma D, Brosens E, de Jong E, van de Ven C, Meeussen C, Cohen-Overbeek T, Boter M, Eussen H, Douben H, Tibboel D *et al*: Copy number detection in discordant monozygotic twins of Congenital Diaphragmatic Hernia (CDH) and Esophageal Atresia (EA) cohorts. *European journal of human genetics: EJHG* 2011.

15. De S: Somatic mosaicism in healthy human tissues. *Trends in genetics: TIG* 2011, 27(6):217–223.
16. Lupski JR: Genomic rearrangements and sporadic disease. *Nature genetics* 2007, 39(7 Suppl):S43–47.
17. van Ommen GJ: Frequency of new copy number variation in humans. *Nature genetics* 2005, 37(4):333–334.
18. Itsara A, Wu H, Smith JD, Nickerson DA, Romieu I, London SJ, Eichler EE: De novo rates and selection of large copy number variation. *Genome Res* 2010, 20(11):1469–1481.
19. Fu W, Zhang F, Wang Y, Gu X, Jin L: Identification of copy number variation hotspots in human populations. *American journal of human genetics* 2010, 87(4):494–504.
20. Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, Moreno-De-Luca D, Moreno-De-Luca A, Mulle JG, Warren ST *et al*: An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genetics in medicine: official journal of the American College of Medical Genetics* 2011, 13(9):777–784.
21. Lupski JR: Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends in genetics: TIG* 1998, 14(10):417–422.
22. Gu W, Zhang F, Lupski JR: Mechanisms for human genomic rearrangements. *Pathogenetics* 2008, 1(1):4.
23. Reiter LT, Hastings PJ, Nelis E, De Jonghe P, Van Broeckhoven C, Lupski JR: Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. *American journal of human genetics* 1998, 62(5):1023–1033.
24. Stankiewicz P, Lupski JR: Genome architecture, rearrangements and genomic disorders. *Trends in genetics: TIG* 2002, 18(2):74–82.
25. van Binsbergen E: Origins and breakpoint analyses of copy number variations: up close and personal. *Cytogenet Genome Res* 2011, 135(3–4):271–276.
26. Lee JA, Carvalho CM, Lupski JR: A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 2007, 131(7):1235–1247.
27. Hastings PJ, Ira G, Lupski JR: A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet* 2009, 5(1):e1000327.
28. Stankiewicz P, Shaw CJ, Dapper JD, Wakui K, Shaffer LG, Withers M, Elizondo L, Park SS, Lupski JR: Genome architecture catalyzes nonrecurrent chromosomal rearrangements. *American journal of human genetics* 2003, 72(5):1101–1116.
29. Hastings PJ, Lupski JR, Rosenberg SM, Ira G: Mechanisms of change in gene copy number. *Nat Rev Genet* 2009, 10(8):551–564.
30. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V *et al*: A copy number variation morbidity map of developmental delay. *Nature genetics* 2011, 43(9):838–846.
31. Vissers LE, van Ravenswaaij CM, Admiraal R, Hurst JA, de Vries BB, Janssen IM, van der Vliet WA, Huys EH, de Jong PJ, Hamel BC *et al*: Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nature genetics* 2004, 36(9):955–957.

32. Kleefstra T, Smidt M, Banning MJ, Oudakker AR, Van Esch H, de Brouwer AP, Nillesen W, Sistermans EA, Hamel BC, de Bruijn D *et al*: Disruption of the gene Euchromatin Histone Methyl Transferase1 (Eu-HMTase1) is associated with the 9q34 subtelomeric deletion syndrome. *Journal of medical genetics* 2005, 42(4):299–306.
33. Koolen DA, Kramer JM, Neveling K, Nillesen WM, Moore-Barton HL, Elmslie FV, Toutain A, Amiel J, Malan V, Tsai AC *et al*: Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. *Nature genetics* 2012, 44(6):639–641.
34. Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Bonnen PE, de Bakker PI, Deloukas P, Gabriel SB *et al*: Integrating common and rare genetic variation in diverse human populations. *Nature* 2010, 467(7311):52–58.
35. Xu H, Poh WT, Sim X, Ong RT, Suo C, Tay WT, Khor CC, Seielstad M, Liu J, Aung T *et al*: SgD-CNV, a database for common and rare copy number variants in three Asian populations. *Human mutation* 2011, 32(12):1341–1349.
36. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A *et al*: Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature genetics* 2008, 40(10):1166–1174.
37. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ *et al*: The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science (New York, NY* 2005, 307(5714):1434–1440.
38. Hollox EJ, Huffmeier U, Zeeuwen PL, Palla R, Lascorz J, Rodijk-Olthuis D, van de Kerkhof PC, Traupe H, de Jongh G, den Heijer M *et al*: Psoriasis is associated with increased beta-defensin genomic copy number. *Nature genetics* 2008, 40(1):23–25.
39. de Cid R, Riveira-Munoz E, Zeeuwen PL, Robarge J, Liao W, Dannhauser EN, Giardina E, Stuart PE, Nair R, Helms C *et al*: Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nature genetics* 2009, 41(2):211–215.
40. Pollex RL, Hegele RA: Copy number variation in the human genome and its implications for cardiovascular disease. *Circulation* 2007, 115(24):3130–3138.
41. Lanktree M, Hegele RA: Copy number variation in metabolic phenotypes. *Cytogenet Genome Res* 2008, 123(1–4):169–175.
42. Jeon JP, Shim SM, Nam HY, Ryu GM, Hong EJ, Kim HL, Han BG: Copy number variation at leptin receptor gene locus associated with metabolic traits and the risk of type 2 diabetes mellitus. *BMC Genomics* 2010, 11:426.
43. Malhotra D, Sebat J: CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* 2012, 148(6):1223–1241.
44. Stankiewicz P, Lupski JR: Structural variation in the human genome and its role in disease. *Annu Rev Med* 2010, 61:437–455.
45. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A *et al*: Finding the missing heritability of complex diseases. *Nature* 2009, 461(7265):747–753.
46. Zhu Q, Ge D, Maia JM, Zhu M, Petrovski S, Dickson SP, Heinzen EL, Shianna KV, Goldstein DB: A genome-wide comparison of the functional properties of rare and common genetic variants in humans. *American journal of human genetics* 2011, 88(4):458–468.

47. Cirulli ET, Goldstein DB: Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 2010, 11(6):415–425.
48. Kryukov GV, Pennacchio LA, Sunyaev SR: Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *American journal of human genetics* 2007, 80(4):727–739.
49. Cooper GM, Goode DL, Ng SB, Sidow A, Bamshad MJ, Shendure J, Nickerson DA: Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat Methods* 2010, 7(4):250–251.
50. Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K, Saeed S, Hamilton-Shield J, Clayton-Smith J, O'Rahilly S *et al*: Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature*, 463(7281):666–670.
51. Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, Falchi M, Chen F, Andrieux J, Lobbens S *et al*: A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature*, 463(7281):671–675.
52. Wang K, Li WD, Glessner JT, Grant SF, Hakonarson H, Price RA: Large copy-number variations are enriched in cases with moderate to extreme obesity. *Diabetes* 2010, 59(10):2690–2694.
53. Glessner JT, Bradfield JP, Wang K, Takahashi N, Zhang H, Sleiman PM, Mentch FD, Kim CE, Hou C, Thomas KA *et al*: A genome-wide study reveals copy number variants exclusive to childhood obesity cases. *American journal of human genetics* 2010, 87(5):661–666.
54. Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerriere A, Vital A, Dumanchin C, Feuillette S, Brice A, Vercelletto M *et al*: APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nature genetics* 2006, 38(1):24–26.
55. Le Marechal C, Masson E, Chen JM, Morel F, Ruszniewski P, Levy P, Ferec C: Hereditary pancreatitis caused by triplication of the trypsinogen locus. *Nature genetics* 2006, 38(12):1372–1374.
56. de Kovel CG, Trucks H, Helbig I, Mefford HC, Baker C, Leu C, Kluck C, Muhle H, von Spiczak S, Ostertag P *et al*: Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain* 2010, 133(Pt 1):23–32.
57. Heinzen EL, Radtke RA, Urban TJ, Cavalleri GL, Depondt C, Need AC, Walley NM, Nicoletti P, Ge D, Catarino CB *et al*: Rare deletions at 16p13.11 predispose to a diverse spectrum of sporadic epilepsy syndromes. *American journal of human genetics* 2010, 86(5):707–718.
58. Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, Franke A, Muhle H, de Kovel C, Baker C, von Spiczak S *et al*: 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nature genetics* 2009, 41(2):160–162.
59. Dibbens LM, Mullen S, Helbig I, Mefford HC, Bayly MA, Bellows S, Leu C, Trucks H, Obermeier T, Wittig M *et al*: Familial and sporadic 15q13.3 microdeletions in idiopathic generalized epilepsy: precedent for disorders with complex inheritance. *Human molecular genetics* 2009, 18(19):3626–3631.
60. Mefford HC, Yendle SC, Hsu C, Cook J, Geraghty E, McMahon JM, Eeg-Olofsson O, Sadleir LG, Gill D, Ben-Zeev B *et al*: Rare copy number variants are an important cause of epileptic encephalopathies. *Ann Neurol* 2011, 70(6):974–985.
61. Lu XY, Phung MT, Shaw CA, Pham K, Neil SE, Patel A, Sahoo T, Bacino CA, Stankiewicz P, Kang SH *et al*: Genomic imbalances in neonates with birth defects:

high detection rates by using chromosomal microarray analysis. *Pediatrics* 2008, 122(6):1310–1318.
62. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J *et al*: Strong association of de novo copy number mutations with autism. *Science (New York, NY* 2007, 316(5823):445–449.
63. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008, 455(7210):237–241.
64. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS *et al*: Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010, 466(7304): 368–372.
65. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP *et al*: Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 2009, 459(7246):569–573.
66. Wittchen HU, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jonsson B, Olesen J, Allgulander C, Alonso J, Faravelli C *et al*: The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European neuropsychopharmacology: the journal of the European College of Neuropsychopharmacology* 2011, 21(9):655–679.
67. Polder JJ, Meerding WJ, Bonneux L, van der Maas PJ: Healthcare costs of intellectual disability in the Netherlands: a cost-of-illness perspective. *Journal of intellectual disability research: JIDR* 2002, 46(Pt 2):168–178.
68. McCandless SE, Brunger JW, Cassidy SB: The burden of genetic disease on inpatient care in a children's hospital. *American journal of human genetics* 2004, 74(1):121–127.
69. Olesen J, Gustavsson A, Svensson M, Wittchen HU, Jonsson B: The economic cost of brain disorders in Europe. *Eur J Neurol* 2012, 19(1):155–162.
70. Smith M: Mental Retardation and Developmental Delay: Genetic and Epigenetic Factors. USA: Oxford University Press; 2005.
71. Ropers HH: Genetics of intellectual disability. *Current opinion in genetics & development* 2008, 18(3):241–250.
72. Knight SJ, Flint J: The use of subtelomeric probes to study mental retardation. *Methods Cell Biol* 2004, 75:799–831.
73. de Vries BB, White SM, Knight SJ, Regan R, Homfray T, Young ID, Super M, McKeown C, Splitt M, Quarrell OW *et al*: Clinical studies on submicroscopic subtelomeric rearrangements: a checklist. *Journal of medical genetics* 2001, 38(3):145–150.
74. Zahir F, Friedman JM: The impact of array genomic hybridization on mental retardation research: a review of current technologies and their clinical utility. *Clin Genet* 2007, 72(4):271–287.
75. Leonard H, Wen X: The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev* 2002, 8(3):117–134.
76. Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, Becker C, Zenker M, Huffmeier U, Thiel C, Ruschendorf F *et al*: Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *American journal of medical genetics Part A* 2006, 140(19):2063–2074.
77. Le Scouarnec S, Gribble SM: Characterising chromosome rearrangements: recent technical advances in molecular cytogenetics. *Heredity* 2012, 108(1):75–85.

78. Vissers LE, de Vries BB, Osoegawa K, Janssen IM, Feuth T, Choy CO, Straatman H, van der Vliet W, Huys EH, van Rijk A *et al*: Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. *American journal of human genetics* 2003, 73(6):1261–1270.
79. Lo Nigro C, Chong CS, Smith AC, Dobyns WB, Carrozzo R, Ledbetter DH: Point mutations and an intragenic deletion in LIS1, the lissencephaly causative gene in isolated lissencephaly sequence and Miller-Dieker syndrome. *Human molecular genetics* 1997, 6(2):157–164.
80. Saitsu H, Kato M, Mizuguchi T, Hamada K, Osaka H, Tohyama J, Uruno K, Kumada S, Nishiyama K, Nishimura A *et al*: De novo mutations in the gene encoding STXBP1 (MUNC18–1) cause early infantile epileptic encephalopathy. *Nature genetics* 2008, 40(6):782–788.
81. Zollino M, Orteschi D, Murdolo M, Lattante S, Battaglia D, Stefanini C, Mercuri E, Chiurazzi P, Neri G, Marangi G: Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype. *Nature genetics* 2012, 44(6):636–638.
82. Hochstenbach R, van Binsbergen E, Engelen J, Nieuwint A, Polstra A, Poddighe P, Ruivenkamp C, Sikkema-Raddatz B, Smeets D, Poot M: Array analysis and karyotyping: workflow consequences based on a retrospective study of 36,325 patients with idiopathic developmental delay in the Netherlands. *European journal of medical genetics* 2009, 52(4):161–169.
83. Slavotinek AM: Novel microdeletion syndromes detected by chromosome microarrays. *Human genetics* 2008, 124(1):1–17.
84. Friedman JM, Baross A, Delaney AD, Ally A, Arbour L, Armstrong L, Asano J, Bailey DK, Barber S, Birch P *et al*: Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. *American journal of human genetics* 2006, 79(3):500–513.
85. Koolen DA, Pfundt R, de Leeuw N, Hehir-Kwa JY, Nillesen WM, Neefs I, Scheltinga I, Sistermans E, Smeets D, Brunner HG *et al*: Genomic microarrays in mental retardation: a practical workflow for diagnostic applications. *Human mutation* 2009, 30(3):283–292.
86. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, Church DM, Crolla JA, Eichler EE, Epstein CJ *et al*: Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *American journal of human genetics* 2010, 86(5):749–764.
87. Jacquemont ML, Sanlaville D, Redon R, Raoul O, Cormier-Daire V, Lyonnet S, Amiel J, Le Merrer M, Heron D, de Blois MC *et al*: Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. *Journal of medical genetics* 2006, 43(11):843–849.
88. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, Skaug JL, Thompson AP, Senman L *et al*: Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature genetics* 2007, 39(3):319–328.
89. Erdogan F, Larsen LA, Zhang L, Tumer Z, Tommerup N, Chen W, Jacobsen JR, Schubert M, Jurkatis J, Tzschach A *et al*: High frequency of submicroscopic genomic aberrations detected by tiling path array comparative genome hybridisation in patients with isolated congenital heart disease. *Journal of medical genetics* 2008, 45(11):704–709.

90. Goldmuntz E, Paluru P, Glessner J, Hakonarson H, Biegel JA, White PS, Gai X, Shaikh TH: Microdeletions and microduplications in patients with congenital heart disease and multiple congenital anomalies. *Congenit Heart Dis* 2011, 6(6):592–602.
91. Lee C, Iafrate AJ, Brothman AR: Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nature genetics* 2007, 39(7 Suppl):S48–54.
92. Topper S, Ober C, Das S: Exome sequencing and the genetics of intellectual disability. *Clin Genet* 2011, 80(2):117–126.
93. Ropers HH: New perspectives for the elucidation of genetic disorders. *American journal of human genetics* 2007, 81(2):199–207.
94. Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, Hardy C, O'Meara S, Latimer C, Dicks E, Menzies A *et al*: A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nature genetics* 2009, 41(5):535–543.
95. Krawitz PM, Schweiger MR, Rodelsperger C, Marcelis C, Kolsch U, Meisel C, Stephani F, Kinoshita T, Murakami Y, Bauer S *et al*: Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. *Nature genetics* 2010, 42(10):827–829.
96. Hoischen A, van Bon BW, Rodriguez-Santiago B, Gilissen C, Vissers LE, de Vries P, Janssen I, van Lier B, Hastings R, Smithson SF *et al*: De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nature genetics* 2011.
97. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M *et al*: A de novo paradigm for mental retardation. *Nature genetics* 2010, 42(12):1109–1112.
98. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC *et al*: Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nature genetics* 2010, 42(9):790–793.
99. Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, Hosseini M, Behjati F, Haas S, Jamali P *et al*: Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* 2011, 478(7367):57–63.
100. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J: Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011, 12(11):745–755.
101. Ku CS, Naidoo N, Pawitan Y: Revisiting Mendelian disorders through exome sequencing. *Human genetics* 2011, 129(4):351–370.
102. Lupski JR: New mutations and intellectual function. *Nature genetics* 2010, 42(12):1036–1038.
103. Shaffer LG, Theisen A, Bejjani BA, Ballif BC, Aylsworth AS, Lim C, McDonald M, Ellison JW, Kostiner D, Saitta S *et al*: The discovery of microdeletion syndromes in the post-genomic era: review of the methodology and characterization of a new 1q41q42 microdeletion syndrome. *Genetics in medicine: official journal of the American College of Medical Genetics* 2007, 9(9):607–616.
104. Schulze TG, McMahon FJ: Defining the phenotype in human genetic studies: forward genetics and reverse phenotyping. *Hum Hered* 2004, 58(3–4):131–138.
105. Girirajan S, Eichler EE: Phenotypic variability and genetic susceptibility to genomic disorders. *Human molecular genetics* 2010, 19(R2):R176–187.
106. Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, Regan R, Kooy RF, Reyniers E, Romano C, Fichera M *et al*: A new chromosome 17q21.31

microdeletion syndrome associated with a common inversion polymorphism. *Nature genetics* 2006, 38(9):999–1001.
107. Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, Gribble S, Curley R, Cumming S, Dunn C, Kalaitzopoulos D *et al*: Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nature genetics* 2006, 38(9):1032–1037.
108. Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, Stewart H, Price SM, Blair E, Hennekam RC *et al*: Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nature genetics* 2006, 38(9):1038–1042.
109. Tan TY, Aftimos S, Worgan L, Susman R, Wilson M, Ghedia S, Kirk EP, Love D, Ronan A, Darmanian A *et al*: Phenotypic expansion and further characterisation of the 17q21.31 microdeletion syndrome. *Journal of medical genetics* 2009, 46(7):480–489.
110. Koolen DA, Sharp AJ, Hurst JA, Firth HV, Knight SJ, Goldenberg A, Saugier-Veber P, Pfundt R, Vissers LE, Destree A *et al*: Clinical and molecular delineation of the 17q21.31 microdeletion syndrome. *Journal of medical genetics* 2008, 45(11):710–720.
111. Dubourg C, Sanlaville D, Doco-Fenzy M, Le Caignec C, Missirian C, Jaillard S, Schluth-Bolard C, Landais E, Boute O, Philip N *et al*: Clinical and molecular characterization of 17q21.31 microdeletion syndrome in 14 French patients with mental retardation. *European journal of medical genetics* 2011, 54(2):144–151.
112. Rademakers R, Cruts M, van Broeckhoven C: The role of tau (MAPT) in frontotemporal dementia and related tauopathies. *Human mutation* 2004, 24(4):277–295.
113. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG *et al*: A common inversion under selection in Europeans. *Nature genetics* 2005, 37(2):129–137.
114. Zody MC, Jiang Z, Fung HC, Antonacci F, Hillier LW, Cardone MF, Graves TA, Kidd JM, Cheng Z, Abouelleil A *et al*: Evolutionary toggling of the MAPT 17q21.31 inversion region. *Nature genetics* 2008, 40(9):1076–1083.
115. Myers AJ, Kaleem M, Marlowe L, Pittman AM, Lees AJ, Fung HC, Duckworth J, Leung D, Gibson A, Morris CM *et al*: The H1c haplotype at the MAPT locus is associated with Alzheimer's disease. *Human molecular genetics* 2005, 14(16):2399–2404.
116. Pittman AM, Myers AJ, Duckworth J, Bryden L, Hanson M, Abou-Sleiman P, Wood NW, Hardy J, Lees A, de Silva R: The structure of the tau haplotype in controls and in progressive supranuclear palsy. *Human molecular genetics* 2004, 13(12):1267–1274.
117. Rao PN, Li W, Vissers LE, Veltman JA, Ophoff RA: Recurrent inversion events at 17q21.31 microdeletion locus are linked to the MAPT H2 haplotype. *Cytogenet Genome Res* 2010, 129(4):275–279.
118. Koolen DA, de Vries BBA: 17q21.31 Microdeletion Syndrome. In: *GeneReviews.* Edited by Pagon RA, Bird TD, Dolan CR, Stephens K, Adam MP. Seattle (WA); 1993.
119. Grisart B, Willatt L, Destree A, Fryns JP, Rack K, de Ravel T, Rosenfeld J, Vermeesch JR, Verellen-Dumoulin C, Sandford R: 17q21.31 microduplication patients are characterised by behavioural problems and poor social interaction. *Journal of medical genetics* 2009, 46(8):524–530.

120. Kirchhoff M, Bisgaard AM, Duno M, Hansen FJ, Schwartz M: A 17q21.31 microduplication, reciprocal to the newly described 17q21.31 microdeletion, in a girl with severe psychomotor developmental delay and dysmorphic craniofacial features. *European journal of medical genetics* 2007, 50(4):256–263.
121. Kitsiou-Tzeli S, Frysira H, Giannikou K, Syrmou A, Kosma K, Kakourou G, Leze E, Sofocleous C, Kanavakis E, Tzetis M: Microdeletion and microduplication 17q21.31 plus an additional CNV, in patients with intellectual disability, identified by array-CGH. *Gene* 2012, 492(1):319–324.
122. Lesnik Oberstein SA, Kriek M, White SJ, Kalf ME, Szuhai K, den Dunnen JT, Breuning MH, Hennekam RC: Peters Plus syndrome is caused by mutations in B3GALTL, a putative glycosyltransferase. *American journal of human genetics* 2006, 79(3):562–566.
123. Klopocki E, Schulze H, Strauss G, Ott CE, Hall J, Trotier F, Fleischhauer S, Greenhalgh L, Newbury-Ecob RA, Neumann LM *et al*: Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome. *American journal of human genetics* 2007, 80(2):232–240.
124. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP: DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *American journal of human genetics* 2009, 84(4):524–533.
125. Vissers LE, de Vries BB, Veltman JA: Genomic microarrays in mental retardation: from copy number variation to gene, from research to diagnosis. *Journal of medical genetics* 2010, 47(5):289–297.
126. Schalock RL, Verdugo MA, Gomez LE: Evidence-based practices in the field of intellectual and developmental disabilities: an international consensus approach. *Eval Program Plann* 2011, 34(3):273–282.
127. Fernell E, Ek U: Borderline intellectual functioning in children and adolescents – insufficiently recognized difficulties. *Acta Paediatr* 2010, 99(5):748–753.
128. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsen E, Stefansson H, Ferreira MA, Green T *et al*: Association between microdeletion and microduplication at 16p11.2 and autism. *The New England journal of medicine* 2008, 358(7):667–675.
129. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, Martinet D, Shen Y, Valsesia A, Beckmann ND *et al*: Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* 2011, 478(7367):97–102.
130. Bijlsma EK, Gijsbers AC, Schuurs-Hoeijmakers JH, van Haeringen A, Fransen van de Putte DE, Anderlid BM, Lundin J, Lapunzina P, Perez Jurado LA, Delle Chiaie B *et al*: Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. *European journal of medical genetics* 2009, 52(2–3):77–87.
131. Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, Stevenson RE, Schroer RJ, Novara F, De Gregori M, Ciccone R *et al*: A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nature genetics* 2008, 40(3):322–328.
132. van Bon BW, Mefford HC, Menten B, Koolen DA, Sharp AJ, Nillesen WM, Innis JW, de Ravel TJ, Mercer CL, Fichera M *et al*: Further delineation of the 15q13

microdeletion and duplication syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome. *Journal of medical genetics* 2009, 46(8):511–523.
133. Fernandez L, Nevado J, Santos F, Heine-Suner D, Martinez-Glez V, Garcia-Minaur S, Palomo R, Delicado A, Pajares IL, Palomares M *et al*: A deletion and a duplication in distal 22q11.2 deletion syndrome region. Clinical implications and review. *BMC Med Genet* 2009, 10:48.
134. Derbent M, Bikmaz YE, Yilmaz Z, Tokel K: Variable phenotype and associations in chromosome 22q11.2 microdeletion. *American journal of medical genetics Part A* 2006, 140(6):659–660.
135. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A, Vives L, Walsh T, McCarthy SE, Baker C *et al*: A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nature genetics* 2010, 42(3):203–209.
136. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C *et al*: Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature genetics* 2011, 43(6):585–589.
137. Wang X, Sun Q, McGrath SD, Mardis ER, Soloway PD, Clark AG: Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PloS one* 2008, 3(12):e3839.
138. Babak T, Deveale B, Armour C, Raymond C, Cleary MA, van der Kooy D, Johnson JM, Lim LP: Global survey of genomic imprinting by transcriptome sequencing. *Curr Biol* 2008, 18(22):1735–1741.
139. Meaney MJ, Ferguson-Smith AC: Epigenetic regulation of the neural transcriptome: the meaning of the marks. *Nat Neurosci* 2010, 13(11):1313–1318.
140. Gregg C, Zhang J, Butler JE, Haig D, Dulac C: Sex-specific parent-of-origin allelic expression in the mouse brain. *Science (New York, NY* 2010, 329(5992): 682–685.
141. Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, Dulac C: High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science (New York, NY* 2010, 329(5992):643–648.
142. DeVeale B, van der Kooy D, Babak T: Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS Genet* 2012, 8(3):e1002600.
143. Duffy AM, Schaner MJ, Wu SH, Staniszewski A, Kumar A, Arevalo JC, Arancio O, Chao MV, Scharfman HE: A selective role for ARMS/Kidins220 scaffold protein in spatial memory and trophic support of entorhinal and frontal cortical neurons. *Exp Neurol* 2011, 229(2):409–420.
144. Ingason A, Kirov G, Giegling I, Hansen T, Isles AR, Jakobsen KD, Kristinsson KT, le Roux L, Gustafsson O, Craddock N *et al*: Maternally derived micro-duplications at 15q11-q13: implication of imprinted genes in psychotic illness. *Am J Psychiatry* 2011, 168(4):408–417.
145. Feuk L, Kalervo A, Lipsanen-Nyman M, Skaug J, Nakabayashi K, Finucane B, Hartung D, Innes M, Kerem B, Nowaczyk MJ *et al*: Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. *American journal of human genetics* 2006, 79(5):965–972.
146. Hogart A, Leung KN, Wang NJ, Wu DJ, Driscoll J, Vallero RO, Schanen NC, LaSalle JM: Chromosome 15q11–13 duplication syndrome brain reveals epigenetic alterations in gene expression not predicted from copy number. *Journal of medical genetics* 2009, 46(2):86–93.

147. Dai L, Bellugi U, Chen XN, Pulst-Korenberg AM, Jarvinen-Pasley A, Tirosh-Wagner T, Eis PS, Graham J, Mills D, Searcy Y *et al*: Is it Williams syndrome? GTF2IRD1 implicated in visual-spatial construction and GTF2I in sociability revealed by high resolution arrays. *American journal of medical genetics Part A* 2009, 149A(3):302–314.
148. Dennis MY, Nuttle X, Sudmant PH, Antonacci F, Graves TA, Nefedov M, Rosenfeld JA, Sajjadian S, Malig M, Kotkiewicz H *et al*: Evolution of Human-Specific Neural SRGAP2 Genes by Incomplete Segmental Duplication. *Cell* 2012, 149(4):912–922.
149. Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, Sampas N, Bruhn L, Shendure J, Eichler EE: Diversity of human copy number variation and multicopy genes. *Science (New York, NY* 2010, 330(6004):641–646.
150. Lupski JR: Structural variation in the human genome. *The New England journal of medicine* 2007, 356(11):1169–1171.
151. Davis EE, Zhang Q, Liu Q, Diplas BH, Davey LM, Hartley J, Stoetzel C, Szymanska K, Ramaswami G, Logan CV *et al*: TTC21B contributes both causal and modifying alleles across the ciliopathy spectrum. *Nature genetics* 2011, 43(3):189–196.
152. Lehner B: Molecular mechanisms of epistasis within and between genes. *Trends in genetics: TIG* 2011, 27(8):323–331.
153. Antonacci F, Kidd JM, Marques-Bonet T, Teague B, Ventura M, Girirajan S, Alkan C, Campbell CD, Vives L, Malig M *et al*: A large and complex structural polymorphism at 16p12.1 underlies microdeletion disease risk. *Nature genetics* 2010, 42(9):745–750.
154. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, Krauss RM, Myers RM, Ridker PM, Chasman DI *et al*: Population analysis of large copy number variants and hotspots of human genetic disease. *American journal of human genetics* 2009, 84(2):148–161.
155. Banka S, Fitzgibbon GJ, Gaunt L, Rankin WJ, Clayton-Smith J: A novel 800 kb microduplication of chromosome 16q22.1 resulting in learning disability and epilepsy may explain phenotypic variability in a family with 15q13 microdeletion. *American journal of medical genetics Part A* 2011, 155A(6):1453–1457.
156. Goldstein DB: Growth of genome screening needs debate. *Nature* 2011, 476(7358):27–28.
157. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA: Clan genomics and the complex architecture of human disease. *Cell* 2011, 147(1):32–43.
158. McClellan J, King MC: Genetic heterogeneity in human disease. *Cell* 2010, 141(2):210–217.
159. Harewood L, Chaignat E, Reymond A: Structural variation and its effect on expression. *Methods Mol Biol* 2012, 838:173–186.
160. Henrichsen CN, Vinckenbosch N, Zollner S, Chaignat E, Pradervand S, Schutz F, Ruedi M, Kaessmann H, Reymond A: Segmental copy number variation shapes tissue transcriptomes. *Nature genetics* 2009, 41(4):424–429.
161. Merla G, Howald C, Henrichsen CN, Lyle R, Wyss C, Zabot MT, Antonarakis SE, Reymond A: Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes. *American journal of human genetics* 2006, 79(2):332–341.

162. Chaignat E, Yahya-Graison EA, Henrichsen CN, Chrast J, Schutz F, Pradervand S, Reymond A: Copy number variation modifies expression time courses. *Genome Res* 2011, 21(1):106–113.
163. Ricard G, Molina J, Chrast J, Gu W, Gheldof N, Pradervand S, Schutz F, Young JI, Lupski JR, Reymond A *et al*: Phenotypic consequences of copy number variation: insights from Smith-Magenis and Potocki-Lupski syndrome mouse models. *PLoS Biol* 2010, 8(11):e1000543.
164. Harewood L, Schutz F, Boyle S, Perry P, Delorenzi M, Bickmore WA, Reymond A: The effect of translocation-induced nuclear reorganization on gene expression. *Genome Res* 2010, 20(5):554–564.
165. Schuster-Bockler B, Conrad D, Bateman A: Dosage sensitivity shapes the evolution of copy-number varied regions. *PloS one*, 5(3):e9474.
166. Molina J, Carmona-Mora P, Chrast J, Krall PM, Canales CP, Lupski JR, Reymond A, Walz K: Abnormal social behaviors and altered gene expression rates in a mouse model for Potocki-Lupski syndrome. *Human molecular genetics* 2008, 17(16):2486–2495.
167. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C *et al*: Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science (New York, NY* 2007, 315(5813):848–853.
168. Hurles ME, Dermitzakis ET, Tyler-Smith C: The functional impact of structural variation in humans. *Trends Genet* 2008, 24(5):238–245.
169. Guryev V, Saar K, Adamovic T, Verheul M, van Heesch SA, Cook S, Pravenec M, Aitman T, Jacob H, Shull JD *et al*: Distribution and functional impact of DNA copy number variation in the rat. *Nature genetics* 2008, 40(5):538–545.
170. Henrichsen CN, Chaignat E, Reymond A: Copy number variants, diseases and gene expression. *Human molecular genetics* 2009, 18(R1):R1–8.
171. Zhang F, Gu W, Hurles ME, Lupski JR: Copy number variation in human health, disease, and evolution. *Annual review of genomics and human genetics* 2009, 10:451–481.
172. Meechan DW, Maynard TM, Gopalakrishna D, Wu Y, LaMantia AS: When half is not enough: gene expression and dosage in the 22q11 deletion syndrome. *Gene expression* 2007, 13(6):299–310.
173. Sisodiya SM, Ragge NK, Cavalleri GL, Hever A, Lorenz B, Schneider A, Williamson KA, Stevens JM, Free SL, Thompson PJ *et al*: Role of SOX2 mutations in human hippocampal malformations and epilepsy. *Epilepsia* 2006, 47(3):534–542.
174. Elsea SH, Girirajan S: Smith-Magenis syndrome. *Eur J Hum Genet* 2008, 16(4): 412–421.
175. Tatton-Brown K, Douglas J, Coleman K, Baujat G, Cole TR, Das S, Horn D, Hughes HE, Temple IK, Faravelli F *et al*: Genotype-phenotype associations in Sotos syndrome: an analysis of 266 individuals with NSD1 aberrations. *American journal of human genetics* 2005, 77(2):193–204.
176. Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, Jacquemont S, Reymond A, Sun M, Sawa A, Gusella JF *et al*: KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* 2012, 485(7398):363–367.
177. Zhang H, Lu X, Beasley J, Mulvihill JJ, Liu R, Li S, Lee JY: Reversed clinical phenotype due to a microduplication of Sotos syndrome region detected by array

CGH: microcephaly, developmental delay and delayed bone age. *American journal of medical genetics* 2011, 155A(6):1374–1378.
178. Franco LM, de Ravel T, Graham BH, Frenkel SM, Van Driessche J, Stankiewicz P, Lupski JR, Vermeesch JR, Cheung SW: A syndrome of short stature, microcephaly and speech delay is associated with duplications reciprocal to the common Sotos syndrome deletion. *Eur J Hum Genet* 2010, 18(2):258–261.
179. Girirajan S, Patel N, Slager RE, Tokarz ME, Bucan M, Wiley JL, Elsea SH: How much is too much? Phenotypic consequences of Rai1 overexpression in mice. *Eur J Hum Genet* 2008, 16(8):941–954.
180. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA *et al*: Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. *Neuron* 2011, 70(5):863–885.
181. Van der Aa N, Rooms L, Vandeweyer G, van den Ende J, Reyniers E, Fichera M, Romano C, Delle Chiaie B, Mortier G, Menten B *et al*: Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome. *European journal of medical genetics* 2009, 52(2–3):94–100.
182. Chapman DL, Papaioannou VE: Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. *Nature* 1998, 391(6668):695–697.
183. Fei Q, Wu Z, Wang H, Zhou X, Wang N, Ding Y, Wang Y, Qiu G: The association analysis of TBX6 polymorphism with susceptibility to congenital scoliosis in a Chinese Han population. *Spine* 2010, 35(9):983–988.
184. Shen Y, Chen X, Wang L, Guo J, Shen J, An Y, Zhu H, Zhu Y, Xin R, Bao Y *et al*: Intra-family phenotypic heterogeneity of 16p11.2 deletion carriers in a three-generation Chinese family. *Am J Med Genet B Neuropsychiatr Genet* 2011, 156(2):225–232.
185. Chen WJ, Lin Y, Xiong ZQ, Wei W, Ni W, Tan GH, Guo SL, He J, Chen YF, Zhang QJ *et al*: Exome sequencing identifies truncating mutations in PRRT2 that cause paroxysmal kinesigenic dyskinesia. *Nature genetics* 2011, 43(12):1252–1255.
186. Wang JL, Cao L, Li XH, Hu ZM, Li JD, Zhang JG, Liang Y, San A, Li N, Chen SQ *et al*: Identification of PRRT2 as the causative gene of paroxysmal kinesigenic dyskinesias. *Brain* 2011, 134(Pt 12):3493–3501.
187. Schubert J, Paravidino R, Becker F, Berger A, Bebek N, Bianchi A, Brockmann K, Capovilla G, Dalla Bernardina B, Fukuyama Y *et al*: PRRT2 mutations are the major cause of benign familial infantile seizures (BFIS). *Human mutation* 2012.
188. Lee HY, Huang Y, Bruneau N, Roll P, Roberson ED, Hermann M, Quinn E, Maas J, Edwards R, Ashizawa T *et al*: Mutations in the novel protein PRRT2 cause paroxysmal kinesigenic dyskinesia with infantile convulsions. *Cell reports* 2012, 1(1):2–12.
189. Ono S, Yoshiura K, Kinoshita A, Kikuchi T, Nakane Y, Kato N, Sadamatsu M, Konishi T, Nagamitsu S, Matsuura M *et al*: Mutations in PRRT2 responsible for paroxysmal kinesigenic dyskinesias also cause benign familial infantile convulsions. *Journal of human genetics* 2012, 57(5):338–341.
190. Li J, Zhu X, Wang X, Sun W, Feng B, Du T, Sun B, Niu F, Wei H, Wu X *et al*: Targeted genomic sequencing identifies PRRT2 mutations as a cause of paroxysmal kinesigenic choreoathetosis. *Journal of medical genetics* 2012, 49(2):76–78.

191. Liu Q, Qi Z, Wan XH, Li JY, Shi L, Lu Q, Zhou XQ, Qiao L, Wu LW, Liu XQ *et al*: Mutations in PRRT2 result in paroxysmal dyskinesias with marked variability in clinical expression. *Journal of medical genetics* 2012, 49(2):79–82.
192. Beckmann JS, Estivill X, Antonarakis SE: Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet* 2007, 8(8):639–646.
193. Carelle-Calmels N, Saugier-Veber P, Girard-Lemaire F, Rudolf G, Doray B, Guerin E, Kuhn P, Arrive M, Gilch C, Schmitt E *et al*: Genetic compensation in a human genomic disorder. *The New England journal of medicine* 2009, 360(12):1211–1216.
194. Lindsay EA, Botta A, Jurecic V, Carattini-Rivera S, Cheah YC, Rosenblatt HM, Bradley A, Baldini A: Congenital heart disease in mice deficient for the DiGeorge syndrome region. *Nature* 1999, 401(6751):379–383.
195. Reymond A, Henrichsen CN, Harewood L, Merla G: Side effects of genome structural changes. *Current opinion in genetics & development* 2007, 17(5):381–386.
196. Nelis M, Esko T, Magi R, Zimprich F, Zimprich A, Toncheva D, Karachanak S, Piskackova T, Balascak I, Peltonen L *et al*: Genetic structure of Europeans: a view from the North-East. *PloS one* 2009, 4(5):e5472.
197. Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, Bassett AS, Seller A, Holmes CC, Ragoussis J: QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* 2007, 35(6):2013–2025.
198. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M: PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007, 17(11):1665–1674.
199. Nolan T, Hands RE, Bustin SA: Quantification of mRNA using real-time RT-PCR. *Nat Protoc* 2006, 1(3):1559–1582.
200. Kiedrowski LA, Raca G, Laffin JJ, Nisler BS, Leonhard K, McIntire E, Mongomery KD: DNA Methylation Assay for X-Chromosome Inactivation in Female Human iPS Cells. *Stem Cell Rev* 2011.
201. Mannik K, Parkel S, Palta P, Zilina O, Puusepp H, Esko T, Magi R, Noukas M, Veidenberg A, Nelis M *et al*: A parallel SNP array study of genomic aberrations associated with mental retardation in patients and general population in Estonia. *European journal of medical genetics* 2011, 54(2):136–143.
202. Hochstenbach R, Buizer-Voskamp JE, Vorstman JA, Ophoff RA: Genome arrays for the detection of copy number variations in idiopathic mental retardation, idiopathic generalized epilepsy and neuropsychiatric disorders: lessons for diagnostic workflow and research. *Cytogenet Genome Res* 2011, 135(3–4):174–202.
203. Zilina O, Reimand T, Zjablovskaja P, Mannik K, Mannamaa M, Traat A, Puusepp-Benazzouz H, Kurg A, Ounap K: Maternally and paternally inherited deletion of 7q31 involving the FOXP2 gene in two families. *American journal of medical genetics Part A* 2011.
204. Pober BR: Williams-Beuren syndrome. *The New England journal of medicine* 2010, 362(3):239–252.
205. Williams JC, Barratt-Boyes BG, Lowe JB: Supravalvular aortic stenosis. *Circulation* 1961, 24:1311–1318.

206. Beuren AJ, Apitz J, Harmjanz D: Supravalvular aortic stenosis in association with mental retardation and a certain facial appearance. *Circulation* 1962, 26:1235–1240.
207. Osborne LR, Mervis CB: Rearrangements of the Williams-Beuren syndrome locus: molecular basis and implications for speech and language development. *Expert reviews in molecular medicine* 2007, 9(15):1–16.
208. Valero MC, de Luis O, Cruces J, Perez Jurado LA: Fine-scale comparative mapping of the human 7q11.23 region and the orthologous region on mouse chromosome 5G: the low-copy repeats that flank the Williams-Beuren syndrome deletion arose at breakpoint sites of an evolutionary inversion(s). *Genomics* 2000, 69(1):1–13.
209. Somerville MJ, Mervis CB, Young EJ, Seo EJ, del Campo M, Bamforth S, Peregrine E, Loo W, Lilley M, Perez-Jurado LA *et al*: Severe expressive-language delay related to duplication of the Williams-Beuren locus. *The New England journal of medicine* 2005, 353(16):1694–1701.
210. Berg JS, Brunetti-Pierri N, Peters SU, Kang SH, Fong CT, Salamone J, Freedenberg D, Hannig VL, Prock LA, Miller DT *et al*: Speech delay and autism spectrum behaviors are frequently associated with duplication of the 7q11.23 Williams-Beuren syndrome region. *Genet Med* 2007, 9(7):427–441.
211. Depienne C, Heron D, Betancur C, Benyahia B, Trouillard O, Bouteiller D, Verloes A, Leguern E, Leboyer M, Brice A: Autism, language delay and mental retardation in a patient with 7q11 duplication. *BMJ case reports* 2009, 2009.
212. Torniero C, dalla Bernardina B, Novara F, Vetro A, Ricca I, Darra F, Pramparo T, Guerrini R, Zuffardi O: Cortical dysplasia of the left temporal lobe might explain severe expressive-language delay in patients with duplication of the Williams-Beuren locus. *Eur J Hum Genet* 2007, 15(1):62–67.
213. Kriek M, White SJ, Szuhai K, Knijnenburg J, van Ommen GJ, den Dunnen JT, Breuning MH: Copy number variation in regions flanked (or unflanked) by duplicons among patients with developmental delay and/or congenital malformations; detection of reciprocal and partial Williams-Beuren duplications. *Eur J Hum Genet* 2006, 14(2):180–189.
214. Tassabehji M, Metcalfe K, Donnai D, Hurst J, Reardon W, Burch M, Read AP: Elastin: genomic structure and point mutations in patients with supravalvular aortic stenosis. *Human molecular genetics* 1997, 6(7):1029–1036.
215. Schneider T, Skitt Z, Liu Y, Deacon RM, Flint J, Karmiloff-Smith A, Rawlins NJ, Tassabehji M: Anxious, hypoactive phenotype combined with motor deficits in Gtf2ird1 null mouse model relevant to Williams syndrome. *Behavioural brain research* 2012.
216. Morris CA, Mervis CB, Hobart HH, Gregg RG, Bertrand J, Ensing GJ, Sommer A, Moore CA, Hopkin RJ, Spallone PA *et al*: GTF2I hemizygosity implicated in mental retardation in Williams syndrome: genotype-phenotype analysis of five families with deletions in the Williams syndrome region. *American journal of medical genetics* 2003, 123A(1):45–59.
217. Tassabehji M, Hammond P, Karmiloff-Smith A, Thompson P, Thorgeirsson SS, Durkin ME, Popescu NC, Hutton T, Metcalfe K, Rucka A *et al*: GTF2IRD1 in craniofacial development of humans and mice. *Science (New York, NY* 2005, 310(5751):1184–1187.

218. Vandeweyer G, Van der Aa N, Reyniers E, Kooy RF: The Contribution of CLIP2 Haploinsufficiency to the Clinical Manifestations of the Williams-Beuren Syndrome. *American journal of human genetics* 2012.
219. Palmer SJ, Santucci N, Widagdo J, Bontempo SJ, Taylor KM, Tay ES, Hook J, Lemckert F, Gunning PW, Hardeman EC: Negative autoregulation of GTF2IRD1 in Williams-Beuren syndrome via a novel DNA binding mechanism. *J Biol Chem* 2010, 285(7):4715–4724.
220. Sakurai T, Dorr NP, Takahashi N, McInnes LA, Elder GA, Buxbaum JD: Haploinsufficiency of Gtf2i, a gene deleted in Williams Syndrome, leads to increases in social interactions. *Autism research: official journal of the International Society for Autism Research* 2011, 4(1):28–39.
221. Malenfant P, Liu X, Hudson ML, Qiao Y, Hrynchak M, Riendeau N, Hildebrand MJ, Cohen IL, Chudley AE, Forster-Gibson C *et al*: Association of GTF2i in the Williams-Beuren Syndrome Critical Region with Autism Spectrum Disorders. *J Autism Dev Disord* 2011.
222. Mervis CB, Dida J, Lam E, Crawford-Zelli NA, Young EJ, Henderson DR, Onay T, Morris CA, Woodruff-Borden J, Yeomans J *et al*: Duplication of GTF2I Results in Separation Anxiety in Mice and Humans. *American journal of human genetics* 2012.
223. Beunders G, van de Kamp JM, Veenhoven RH, van Hagen JM, Nieuwint AW, Sistermans EA: A triplication of the Williams-Beuren syndrome region in a patient with mental retardation, a severe expressive language delay, behavioural problems and dysmorphisms. *Journal of medical genetics* 2010, 47(4):271–275.
224. Loftus BJ, Kim UJ, Sneddon VP, Kalush F, Brandon R, Fuhrmann J, Mason T, Crosby ML, Barnstead M, Cronin L *et al*: Genome duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q. *Genomics* 1999, 60(3):295–308.
225. Johnson ME, Cheng Z, Morrison VA, Scherer S, Ventura M, Gibbs RA, Green ED, Eichler EE: Recurrent duplication-driven transposition of DNA during hominoid evolution. *Proc Natl Acad Sci U S A* 2006, 103(47):17626–17631.
226. Martin J, Han C, Gordon LA, Terry A, Prabhakar S, She X, Xie G, Hellsten U, Chan YM, Altherr M *et al*: The sequence and analysis of duplication-rich human chromosome 16. *Nature* 2004, 432(7020):988–994.
227. Ullmann R, Turner G, Kirchhoff M, Chen W, Tonge B, Rosenberg C, Field M, Vianna-Morgante AM, Christie L, Krepischi-Santos AC *et al*: Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Human mutation* 2007, 28(7):674–682.
228. Hannes FD, Sharp AJ, Mefford HC, de Ravel T, Ruivenkamp CA, Breuning MH, Fryns JP, Devriendt K, Van Buggenhout G, Vogels A *et al*: Recurrent reciprocal deletions and duplications of 16p13.11: the deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant. *Journal of medical genetics* 2009, 46(4):223–232.
229. Mefford HC, Cooper GM, Zerr T, Smith JD, Baker C, Shafer N, Thorland EC, Skinner C, Schwartz CE, Nickerson DA *et al*: A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. *Genome Res* 2009, 19(9):1579–1585.
230. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A, Vives L, Walsh T, McCarthy SE, Baker C *et al*: A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nature genetics*, 42(3):203–209.

231. Ballif BC, Hornor SA, Jenkins E, Madan-Khetarpal S, Surti U, Jackson KE, Asamoah A, Brock PL, Gowans GC, Conway RL *et al*: Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. *Nature genetics* 2007, 39(9):1071–1073.
232. Hempel M, Rivera Brugues N, Wagenstaller J, Lederer G, Weitensteiner A, Seidel H, Meitinger T, Strom TM: Microdeletion syndrome 16p11.2-p12.2: clinical and molecular characterization. *American journal of medical genetics Part A* 2009, 149A(10):2106–2112.
233. Battaglia A, Novelli A, Bernardini L, Igliozzi R, Parrini B: Further characterization of the new microdeletion syndrome of 16p11.2-p12.2. *American journal of medical genetics Part A* 2009, 149A(6):1200–1204.
234. Bachmann-Gagescu R, Mefford HC, Cowan C, Glew GM, Hing AV, Wallace S, Bader PI, Hamati A, Reitnauer PJ, Smith R *et al*: Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity. *Genetics in medicine: official journal of the American College of Medical Genetics* 2010, 12(10):641–647.
235. Walsh KM, Bracken MB: Copy number variation in the dosage-sensitive 16p11.2 interval accounts for only a small proportion of autism incidence: a systematic review and meta-analysis. *Genetics in medicine: official journal of the American College of Medical Genetics* 2011, 13(5):377–384.
236. Shinawi M, Liu P, Kang SH, Shen J, Belmont JW, Scott DA, Probst FJ, Craigen WJ, Graham BH, Pursley A *et al*: Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *Journal of medical genetics* 2010, 47(5):332–341.
237. Degenhardt F, Priebe L, Herms S, Mattheisen M, Muhleisen TW, Meier S, Moebus S, Strohmaier J, Gross M, Breuer R *et al*: Association between copy number variants in 16p11.2 and major depressive disorder in a German case-control sample. *Am J Med Genet B Neuropsychiatr Genet* 2012, 159B(3):263–273.
238. Dale RC, Grattan-Smith P, Fung VS, Peters GB: Infantile convulsions and paroxysmal kinesigenic dyskinesia with 16p11.2 microdeletion. *Neurology* 2011, 77(14):1401–1402.
239. Wat MJ, Veenma D, Hogue J, Holder AM, Yu Z, Wat JJ, Hanchard N, Shchelochkov OA, Fernandes CJ, Johnson A *et al*: Genomic alterations that contribute to the development of isolated and non-isolated congenital diaphragmatic hernia. *Journal of medical genetics* 2011, 48(5):299–307.
240. Shen Y, Chen X, Wang L, Guo J, Shen J, An Y, Zhu H, Zhu Y, Xin R, Bao Y *et al*: Intra-family phenotypic heterogeneity of 16p11.2 deletion carriers in a three-generation Chinese family. *Am J Med Genet B Neuropsychiatr Genet* 2010.
241. Nik-Zainal S, Strick R, Storer M, Huang N, Rad R, Willatt L, Fitzgerald T, Martin V, Sandford R, Carter NP *et al*: High incidence of recurrent copy number variants in patients with isolated and syndromic Mullerian aplasia. *Journal of medical genetics* 2011, 48(3):197–204.
242. Bardakjian TM, Kwok S, Slavotinek AM, Schneider AS: Clinical report of microphthalmia and optic nerve coloboma associated with a de novo microdeletion of chromosome 16p11.2. *American journal of medical genetics Part A* 2010, 152A(12):3120–3123.
243. Schaaf CP, Goin-Kochel RP, Nowell KP, Hunter JV, Aleck KA, Cox S, Patel A, Bacino CA, Shinawi M: Expanding the clinical spectrum of the 16p11.2

chromosomal rearrangements: three patients with syringomyelia. *European journal of human genetics: EJHG* 2011, 19(2):152–156.
244. Puvabanditsin S, Nagar MS, Joshi M, Lambert G, Garrow E, Brandsma E: Microdeletion of 16p11.2 associated with endocardial fibroelastosis. *American journal of medical genetics Part A* 2010, 152A(9):2383–2386.
245. Bedoyan JK, Kumar RA, Sudi J, Silverstein F, Ackley T, Iyer RK, Christian SL, Martin DM: Duplication 16p11.2 in a child with infantile seizure disorder. *American journal of medical genetics Part A* 2010, 152A(6):1567–1574.
246. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, Perkins DO, Dickel DE, Kusenda M, Krastoshevsky O *et al*: Microduplications of 16p11.2 are associated with schizophrenia. *Nature genetics* 2009, 41(11):1223–1227.
247. Shimojima K, Inoue T, Fujii Y, Ohno K, Yamamoto T: A familial 593-kb microdeletion of 16p11.2 associated with mental retardation and hemivertebrae. *European journal of medical genetics* 2009, 52(6):433–435.
248. Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, Badner JA, Gilliam TC, Nowak NJ, Cook EH, Jr., Dobyns WB *et al*: Recurrent 16p11.2 microdeletions in autism. *Human molecular genetics* 2008, 17(4):628–638.
249. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y *et al*: Structural variation of chromosomes in autism spectrum disorder. *American journal of human genetics* 2008, 82(2):477–488.
250. Firmann M, Mayor V, Vidal PM, Bochud M, Pecoud A, Hayoz D, Paccaud F, Preisig M, Song KS, Yuan X *et al*: The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC cardiovascular disorders* 2008, 8:6.
251. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J, Jones CG, Zaitlen NA, Varilo T, Kaakinen M *et al*: Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature genetics* 2009, 41(1):35–46.
252. Meyre D, Delplanque J, Chevre JC, Lecoeur C, Lobbens S, Gallina S, Durand E, Vatin V, Degraeve F, Proenca C *et al*: Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nature genetics* 2009, 41(2):157–159.
253. Balkau B, Eschwege E, Tichet J, Marre M: Proposed criteria for the diagnosis of diabetes: evidence from a French epidemiological study (D.E.S.I.R.). *Diabetes & metabolism* 1997, 23(5):428–434.
254. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S *et al*: A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007, 445(7130):881–885.
255. O'Rahilly S, Farooqi IS: Human obesity: a heritable neurobehavioral disorder that is highly sensitive to environmental conditions. *Diabetes* 2008, 57(11):2905–2910.
256. Maes HH, Neale MC, Eaves LJ: Genetic and environmental factors in relative body weight and human adiposity. *Behavior genetics* 1997, 27(4):325–351.
257. Barsh GS, Farooqi IS, O'Rahilly S: Genetics of body-weight regulation. *Nature* 2000, 404(6778):644–651.
258. Sabin MA, Werther GA, Kiess W: Genetics of obesity and overgrowth syndromes. *Best practice & research Clinical endocrinology & metabolism* 2011, 25(1):207–220.

259. Farooqi IS, O'Rahilly S: Monogenic obesity in humans. *Annu Rev Med* 2005, 56:443–458.
260. Stewart L, Van de Ven L, Katsarou V, Rentziou E, Doran M, Jackson P, Reilly JJ, Wilson D: High prevalence of obesity in ambulatory children and adolescents with intellectual disability. *J Intellect Disabil Res* 2009, 53(10):882–886.
261. Miller J, Kranzler J, Liu Y, Schmalfuss I, Theriaque DW, Shuster JJ, Hatfield A, Mueller OT, Goldstone AP, Sahoo T *et al*: Neurocognitive findings in Prader-Willi syndrome and early-onset morbid obesity. *The Journal of pediatrics* 2006, 149(2): 192–198.
262. Chen AY, Kim SE, Houtrow AJ, Newacheck PW: Prevalence of obesity among children with chronic conditions. *Obesity (Silver Spring, Md*, 18(1):210–213.
263. Boeka AG, Lokken KL: Neuropsychological performance of a clinical sample of extremely obese individuals. *Arch Clin Neuropsychol* 2008, 23(4):467–474.
264. Kilpelainen TO, Zillikens MC, Stancakova A, Finucane FM, Ried JS, Langenberg C, Zhang W, Beckmann JS, Luan J, Vandenput L *et al*: Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nature genetics* 2011, 43(8):753–760.
265. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Allen HL, Lindgren CM, Luan J, Magi R *et al*: Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics* 2010, 42(11):937–948.
266. Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, Qi L, Speliotes EK, Thorleifsson G, Willer CJ, Herrera BM *et al*: Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet* 2009, 5(6):e1000508.
267. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM, Berndt SI, Elliott AL, Jackson AU, Lamina C *et al*: Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature genetics* 2009, 41(1):25–34.
268. Zhao J, Bradfield JP, Zhang H, Sleiman PM, Kim CE, Glessner JT, Deliard S, Thomas KA, Frackelton EC, Li M *et al*: Role of BMI-Associated Loci Identified in GWAS Meta-Analyses in the Context of Common Childhood Obesity in European Americans. *Obesity* 2011.
269. Wang K, Li WD, Zhang CK, Wang Z, Glessner JT, Grant SF, Zhao H, Hakonarson H, Price RA: A genome-wide association study on obesity and obesity-related traits. *PloS one* 2011, 6(4):e18939.
270. Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D, Barnes C, Conrad DF, Giannoulatou E *et al*: Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 2010, 464(7289):713–720.
271. Froguel P, Blakemore AI: The power of the extreme in elucidating obesity. *The New England journal of medicine* 2008, 359(9):891–893.
272. Horev G, Ellegood J, Lerch JP, Son YE, Muthuswamy L, Vogel H, Krieger AM, Buja A, Henkelman RM, Wigler M *et al*: Dosage-dependent phenotypes in models of 16p11.2 lesions found in autism. *Proc Natl Acad Sci U S A* 2011, 108(41): 17076–17081.
273. Ben-Shachar S, Lanpher B, German JR, Qasaymeh M, Potocki L, Nagamani SC, Franco LM, Malphrus A, Bottenfield GW, Spence JE *et al*: Microdeletion

15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *Journal of medical genetics* 2009, 46(6):382–388.
274. Masurel-Paulet A, Andrieux J, Callier P, Cuisset JM, Le Caignec C, Holder M, Thauvin-Robinet C, Doray B, Flori E, Alex-Cordier MP *et al*: Delineation of 15q13.3 microdeletions. *Clin Genet* 2010, 78(2):149–161.
275. Muhle H, Mefford HC, Obermeier T, von Spiczak S, Eichler EE, Stephani U, Sander T, Helbig I: Absence seizures with intellectual disability as a phenotype of the 15q13.3 microdeletion syndrome. *Epilepsia* 2011, 52(12):e194–198.
276. Mulley JC, Scheffer IE, Desai T, Bayly MA, Grinton BE, Vears DF, Berkovic SF, Dibbens LM: Investigation of the 15q13.3 CNV as a genetic modifier for familial epilepsies with variable phenotypes. *Epilepsia* 2011.
277. Spielmann M, Reichelt G, Hertzberg C, Trimborn M, Mundlos S, Horn D, Klopocki E: Homozygous deletion of chromosome 15q13.3 including CHRNA7 causes severe mental retardation, seizures, muscular hypotonia, and the loss of KLF13 and TRPM1 potentially cause macrocytosis and congenital retinal dysfunction in siblings. *European journal of medical genetics* 2011, 54(4):e441–445.
278. Liao J, DeWard SJ, Madan-Khetarpal S, Surti U, Hu J: A small homozygous microdeletion of 15q13.3 including the CHRNA7 gene in a girl with a spectrum of severe neurodevelopmental features. *American journal of medical genetics* 2011, 155A(11):2795–2800.
279. Endris V, Hackmann K, Neuhann TM, Grasshoff U, Bonin M, Haug U, Hahn G, Schallner JC, Schrock E, Tinschert S *et al*: Homozygous loss of CHRNA7 on chromosome 15q13.3 causes severe encephalopathy with seizures and hypotonia. *American journal of medical genetics* 2010, 152A(11):2908–2911.
280. Shinawi M, Schaaf CP, Bhatt SS, Xia Z, Patel A, Cheung SW, Lanpher B, Nagl S, Herding HS, Nevinny-Stickel C *et al*: A small recurrent deletion within 15q13.3 is associated with a range of neurodevelopmental phenotypes. *Nature genetics* 2009, 41(12):1269–1271.
281. Miller DT, Shen Y, Weiss LA, Korn J, Anselm I, Bridgemohan C, Cox GF, Dickinson H, Gentile J, Harris DJ *et al*: Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. *Journal of medical genetics* 2009, 46(4):242–248.
282. Szafranski P, Schaaf CP, Person RE, Gibson IB, Xia Z, Mahadevan S, Wiszniewska J, Bacino CA, Lalani S, Potocki L *et al*: Structures and molecular mechanisms for common 15q13.3 microduplications involving CHRNA7: benign or pathological? *Human mutation* 2010, 31(7):840–850.
283. van Bon BWM, Mefford HC, de Vries BBA: 15q13.3 Microdeletion. In: *GeneReviews*. Edited by Pagon RA, Bird TD, Dolan CR, Stephens K, Adam MP. Seattle (WA); 1993.
284. Pagnamenta AT, Wing K, Sadighi Akha E, Knight SJ, Bolte S, Schmotzer G, Duketis E, Poustka F, Klauck SM, Poustka A *et al*: A 15q13.3 microdeletion segregating with autism. *Eur J Hum Genet* 2009, 17(5):687–692.
285. del Gaudio D, Fang P, Scaglia F, Ward PA, Craigen WJ, Glaze DG, Neul JL, Patel A, Lee JA, Irons M *et al*: Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genet Med* 2006, 8(12):784–792.
286. Vandewalle J, Van Esch H, Govaerts K, Verbeeck J, Zweier C, Madrigal I, Mila M, Pijkels E, Fernandez I, Kohlhase J *et al*: Dosage-dependent severity of the

phenotype in patients with mental retardation due to a recurrent copy-number gain at Xq28 mediated by an unusual recombination. *American journal of human genetics* 2009, 85(6):809–822.
287. Bartsch O, Gebauer K, Lechno S, van Esch H, Froyen G, Bonin M, Seidel J, Thamm-Mucke B, Horn D, Klopocki E *et al*: Four unrelated patients with Lubs X-linked mental retardation syndrome and different Xq28 duplications. *American journal of medical genetics*, 152A(2):305–312.
288. Lugtenberg D, Kleefstra T, Oudakker AR, Nillesen WM, Yntema HG, Tzschach A, Raynaud M, Rating D, Journel H, Chelly J *et al*: Structural variation in Xq28: MECP2 duplications in 1% of patients with unexplained XLMR and in 2% of male patients with severe encephalopathy. *Eur J Hum Genet* 2009, 17(4):444–453.
289. Van Esch H, Bauters M, Ignatius J, Jansen M, Raynaud M, Hollanders K, Lugtenberg D, Bienvenu T, Jensen LR, Gecz J *et al*: Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *American journal of human genetics* 2005, 77(3):442–453.
290. Witham S, Takano K, Schwartz C, Alexov E: A missense mutation in CLIC2 associated with intellectual disability is predicted by in silico modeling to affect protein stability and dynamics. *Proteins* 2011, 79(8):2444–2454.
291. Giannandrea M, Bianchi V, Mignogna ML, Sirri A, Carrabino S, D'Elia E, Vecellio M, Russo S, Cogliati F, Larizza L *et al*: Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly. *American journal of human genetics*, 86(2):185–195.
292. Collins AL, Levenson JM, Vilaythong AP, Richman R, Armstrong DL, Noebels JL, David Sweatt J, Zoghbi HY: Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Human molecular genetics* 2004, 13(21):2679–2689.
293. De Brasi CD, Bowen DJ: Molecular characteristics of the intron 22 homologs of the coagulation factor VIII gene: an update. *J Thromb Haemost* 2008, 6(10):1822–1824.
294. El-Hattab AW, Fang P, Jin W, Hughes JR, Gibson JB, Patel GS, Grange DK, Manwaring LP, Patel A, Stankiewicz P *et al*: Int22h-1/int22h-2-mediated Xq28 rearrangements: intellectual disability associated with duplications and in utero male lethality with deletions. *Journal of medical genetics* 2011, 48(12):840–850.
295. Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP *et al*: The DNA sequence of the human X chromosome. *Nature* 2005, 434(7031):325–337.
296. Myers S, Freeman C, Auton A, Donnelly P, McVean G: A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature genetics* 2008, 40(9):1124–1129.
297. Liu P, Erez A, Nagamani SC, Bi W, Carvalho CM, Simmons AD, Wiszniewska J, Fang P, Eng PA, Cooper ML *et al*: Copy number gain at Xp22.31 includes complex duplication rearrangements and recurrent triplications. *Human molecular genetics* 2011, 20(10):1975–1988.
298. Ballabio A, Bardoni B, Carrozzo R, Andria G, Bick D, Campbell L, Hamel B, Ferguson-Smith MA, Gimelli G, Fraccaro M *et al*: Contiguous gene syndromes due to deletions in the distal short arm of the human X chromosome. *Proc Natl Acad Sci U S A* 1989, 86(24):10001–10005.

299. Ballabio A, Carrozzo R, Parenti G, Gil A, Zollo M, Persico MG, Gillard E, Affara N, Yates J, Ferguson-Smith MA *et al*: Molecular heterogeneity of steroid sulfatase deficiency: a multicenter study on 57 unrelated patients, at DNA and protein levels. *Genomics* 1989, 4(1):36–40.
300. Ballabio A, Sebastio G, Carrozzo R, Parenti G, Piccirillo A, Persico MG, Andria G: Deletions of the steroid sulphatase gene in "classical" X-linked ichthyosis and in X-linked ichthyosis associated with Kallmann syndrome. *Human genetics* 1987, 77(4):338–341.
301. Fukami M, Kirsch S, Schiller S, Richter A, Benes V, Franco B, Muroya K, Rao E, Merker S, Niesler B *et al*: A member of a gene family on Xp22.3, VCX-A, is deleted in patients with X-linked nonspecific mental retardation. *American journal of human genetics* 2000, 67(3):563–573.
302. Van Esch H, Hollanders K, Badisco L, Melotte C, Van Hummelen P, Vermeesch JR, Devriendt K, Fryns JP, Marynen P, Froyen G: Deletion of VCX-A due to NAHR plays a major role in the occurrence of mental retardation in patients with X-linked ichthyosis. *Human molecular genetics* 2005, 14(13):1795–1803.
303. Park SJ, Jung EH, Ryu RS, Kang HW, Ko JM, Kim HJ, Cheon CK, Hwang SH, Kang HY: Clinical implementation of whole-genome array CGH as a first-tier test in 5080 pre and postnatal cases. *Mol Cytogenet* 2011, 4:12.
304. Lonardo F, Parenti G, Luquetti DV, Annunziata I, Della Monica M, Perone L, De Gregori M, Zuffardi O, Brunetti-Pierri N, Andria G *et al*: Contiguous gene syndrome due to an interstitial deletion in Xp22.3 in a boy with ichthyosis, chondrodysplasia punctata, mental retardation and ADHD. *European journal of medical genetics* 2007, 50(4):301–308.
305. Kent L, Emerton J, Bhadravathi V, Weisblatt E, Pasco G, Willatt LR, McMahon R, Yates JR: X-linked ichthyosis (steroid sulfatase deficiency) is associated with increased risk of attention deficit hyperactivity disorder, autism and social communication deficits. *Journal of medical genetics* 2008, 45(8):519–524.
306. Wagenstaller J, Spranger S, Lorenz-Depiereux B, Kazmierczak B, Nathrath M, Wahl D, Heye B, Glaser D, Liebscher V, Meitinger T *et al*: Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation. *American journal of human genetics* 2007, 81(4):768–779.
307. Bernardini L, Alesi V, Loddo S, Novelli A, Bottillo I, Battaglia A, Digilio MC, Zampino G, Ertel A, Fortina P *et al*: High-resolution SNP arrays in mental retardation diagnostics: how much do we gain? *European journal of human genetics: EJHG* 2010, 18(2):178–185.
308. Li F, Shen Y, Kohler U, Sharkey FH, Menon D, Coulleaux L, Malan V, Rio M, McMullan DJ, Cox H *et al*: Interstitial microduplication of Xp22.31: Causative of intellectual disability or benign copy number variant? *European journal of medical genetics* 2010, 53(2):93–99.
309. Furrow A, Theisen A, Velsher L, Bawle EV, Sastry S, Mendelsohn NJ, Jarvis K, Shaffer LG, Chitayat D: Duplication of the STS region in males is a benign copy-number variant. *American journal of medical genetics Part A* 2011, 155(8):1972–1975.
310. Hernandez-Martin A, Gonzalez-Sarmiento R, De Unamuno P: X-linked ichthyosis: an update. *Br J Dermatol* 1999, 141(4):617–627.
311. Meroni G, Franco B, Archidiacono N, Messali S, Andolfi G, Rocchi M, Ballabio A: Characterization of a cluster of sulfatase genes on Xp22.3 suggests gene

duplications in an ancestral pseudoautosomal region. *Human molecular genetics* 1996, 5(4):423–431.
312. Desguerre I, Christov C, Mayer M, Zeller R, Becane HM, Bastuji-Garin S, Leturcq F, Chiron C, Chelly J, Gherardi RK: Clinical heterogeneity of duchenne muscular dystrophy (DMD): definition of sub-phenotypes and predictive criteria by long-term follow-up. *PloS one* 2009, 4(2):e4347.
313. Daoud F, Angeard N, Demerre B, Martie I, Benyaou R, Leturcq F, Cossee M, Deburgrave N, Saillour Y, Tuffery S *et al*: Analysis of Dp71 contribution in the severity of mental retardation through comparison of Duchenne and Becker patients differing by mutation consequences on Dp71 expression. *Human molecular genetics* 2009, 18(20):3779–3794.
314. Daoud F, Candelario-Martinez A, Billard JM, Avital A, Khelfaoui M, Rozenvald Y, Guegan M, Mornet D, Jaillard D, Nudel U *et al*: Role of mental retardation-associated dystrophin-gene product Dp71 in excitatory synapse organization, synaptic plasticity and behavioral functions. *PloS one* 2009, 4(8):e6574.
315. Balikova I, Menten B, de Ravel T, Le Caignec C, Thienpont B, Urbina M, Doco-Fenzy M, de Rademaeker M, Mortier G, Kooy F *et al*: Subtelomeric imbalances in phenotypically normal individuals. *Human mutation* 2007, 28(10):958–967.
316. Wu SH, Arevalo JC, Sarti F, Tessarollo L, Gan WB, Chao MV: Ankyrin Repeat-rich Membrane Spanning/Kidins220 protein regulates dendritic branching and spine stability in vivo. *Dev Neurobiol* 2009, 69(9):547–557.
317. Higuero AM, Sanchez-Ruiloba L, Doglio LE, Portillo F, Abad-Rodriguez J, Dotti CG, Iglesias T: Kidins220/ARMS modulates the activity of microtubule-regulating proteins and controls neuronal polarity and development. *J Biol Chem* 2010, 285(2):1343–1357.
318. Radovick S, Nations M, Du Y, Berg LA, Weintraub BD, Wondisford FE: A mutation in the POU-homeodomain of Pit-1 responsible for combined pituitary hormone deficiency. *Science (New York, NY* 1992, 257(5073):1115–1118.
319. Hashimoto Y, Cisternino M, Cohen LE: A novel nonsense mutation in the Pit-1 gene: evidence for a gene dosage effect. *J Clin Endocrinol Metab* 2003, 88(3):1241–1247.
320. Sun Y, Zhang F, Gao J, Gao X, Guo T, Zhang K, Shi Y, Zheng Z, Tang W, Zheng Y *et al*: Positive association between POU1F1 and mental retardation in young females in the Chinese Han population. *Human molecular genetics* 2006, 15(7):1237–1243.
321. Howard TD, Paznekas WA, Green ED, Chiang LC, Ma N, Ortiz de Luna RI, Garcia Delgado C, Gonzalez-Ramos M, Kline AD, Jabs EW: Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome. *Nature genetics* 1997, 15(1):36–41.
322. el Ghouzzi V, Le Merrer M, Perrin-Schmitt F, Lajeunie E, Benit P, Renier D, Bourgeois P, Bolcato-Bellemin AL, Munnich A, Bonaventure J: Mutations of the TWIST gene in the Saethre-Chotzen syndrome. *Nature genetics* 1997, 15(1):42–46.
323. Busche A, Graul-Neumann LM, Zweier C, Rauch A, Klopocki E, Horn D: Microdeletions of chromosome 7p21, including TWIST1, associated with significant microcephaly, facial dysmorphism, and short stature. *European journal of medical genetics* 2011, 54(3):256–261.
324. Cai J, Goodman BK, Patel AS, Mulliken JB, Van Maldergem L, Hoganson GE, Paznekas WA, Ben-Neriah Z, Sheffer R, Cunningham ML *et al*: Increased risk for

developmental delay in Saethre-Chotzen syndrome is associated with TWIST deletions: an improved strategy for TWIST mutation screening. *Human genetics* 2003, 114(1):68–76.
325. de Heer IM, de Klein A, van den Ouweland AM, Vermeij-Keers C, Wouters CH, Vaandrager JM, Hovius SE, Hoogeboom JM: Clinical and genetic analysis of patients with Saethre-Chotzen syndrome. *Plast Reconstr Surg* 2005, 115(7):1894–1902; discussion 1903–1895.
326. Johnson D, Horsley SW, Moloney DM, Oldridge M, Twigg SR, Walsh S, Barrow M, Njolstad PR, Kunz J, Ashworth GJ *et al*: A comprehensive screen for TWIST mutations in patients with craniosynostosis identifies a new microdeletion syndrome of chromosome band 7p21.1. *American journal of human genetics* 1998, 63(5):1282–1293.
327. Gripp KW, Zackai EH, Stolle CA: Mutations in the human TWIST gene. *Human mutation* 2000, 15(5):479.
328. Candia AF, Wright CV: Differential localization of Mox-1 and Mox-2 proteins indicates distinct roles during development. *Int J Dev Biol* 1996, 40(6):1179–1184.
329. Mankoo BS, Collins NS, Ashby P, Grigorieva E, Pevny LH, Candia A, Wright CV, Rigby PW, Pachnis V: Mox2 is a component of the genetic hierarchy controlling limb muscle development. *Nature* 1999, 400(6739):69–73.
330. Quinn LM, Latham SE, Kalionis B: The homeobox genes MSX2 and MOX2 are candidates for regulating epithelial-mesenchymal cell interactions in the human placenta. *Placenta* 2000, 21 Suppl A:S50–54.
331. Candia AF, Hu J, Crosby J, Lalley PA, Noden D, Nadeau JH, Wright CV: Mox-1 and Mox-2 define a novel homeobox gene subfamily and are differentially expressed during early mesodermal patterning in mouse embryos. *Development* 1992, 116(4):1123–1136.
332. Kirilenko P, He G, Mankoo BS, Mallo M, Jones R, Bobola N: Transient activation of meox1 is an early component of the gene regulatory network downstream of hoxa2. *Mol Cell Biol* 2011, 31(6):1301–1308.
333. Cobourne MT, Sharpe PT: Making up the numbers: The molecular control of mammalian dental formula. *Semin Cell Dev Biol* 2010, 21(3):314–324.
334. Kassai Y, Munne P, Hotta Y, Penttila E, Kavanagh K, Ohbayashi N, Takada S, Thesleff I, Jernvall J, Itoh N: Regulation of mammalian tooth cusp patterning by ectodin. *Science (New York, NY* 2005, 309(5743):2067–2070.
335. Itasaki N, Jones CM, Mercurio S, Rowe A, Domingos PM, Smith JC, Krumlauf R: Wise, a context-dependent activator and inhibitor of Wnt signalling. *Development* 2003, 130(18):4295–4305.
336. Laurikkala J, Kassai Y, Pakkasjarvi L, Thesleff I, Itoh N: Identification of a secreted BMP antagonist, ectodin, integrating BMP, FGF, and SHH signals from the tooth enamel knot. *Dev Biol* 2003, 264(1):91–105.
337. Yanagita M, Okuda T, Endo S, Tanaka M, Takahashi K, Sugiyama F, Kunita S, Takahashi S, Fukatsu A, Yanagisawa M *et al*: Uterine sensitization-associated gene-1 (USAG-1), a novel BMP antagonist expressed in the kidney, accelerates tubular injury. *J Clin Invest* 2006, 116(1):70–79.
338. Lazebnik Y: Can a biologist fix a radio?--Or, what I learned while studying apoptosis. *Cancer cell* 2002, 2(3):179–182.

WEB RESOURCES

BioGPS – a gene annotation portal: http://biogps.gnf.org

Database of Genomic Variants – a curated catalogue of structural variation in the human genome: http://projects.tcag.ca/variation/

DECIPHER – a database of chromosomal imbalance and phenotype in humans using Ensembl resources: http://decipher.sanger.ac.uk

ECARUCA – register of unbalanced chromosome aberrations coordinated by the European Cytogeneticists Association: http://www.ecaruca.net

Ensembl Genome Browser: http://www.ensembl.org

Estonian Genome Centre at University of Tartu – Estonian general population based biobank: http://www.geenivaramu.ee

GeTprime – Gene and Transcript-specific primer generator for real-time PCR: http://updepla1srv1.epfl.ch/getprime

Illumina Inc. – a company that manufactures platforms for the analysis of genetic variation and biological function: http://www.illumina.com

HapMap Project – an international effort to identify and catalog genetic variation in humans: http://hapmap.ncbi.nlm.nih.gov

ISCA Consortium – International Standards for Cytogenomic Arrays Consortium: https://www.iscaconsortium.org

Leiden Open Variation Database – online gene-centered collection and display of DNA variations: http://www.lovd.nl/2.0

OMIM – Online Mendelian Inheritance in Man database: http://www.ncbi.nlm.nih.gov/omim/

Orphanet – the portal for rare diseases and orphan drugs: http://www.orpha.net

Oxford Gene Technology – a company that manufactures clinical genetics and molecular diagnostic solutions: http://www.ogt.co.uk/

PubMed – a repository of peer-reviewed primary research reports in life sciences: http://www.ncbi.nlm.nih.gov/pubmed

WHO – World Health Organization: http://www.who.int

WHO Global Database on Body Mass Index: http://apps.who.int/bmi/index.jsp

WHO ICD-10 – WHO International Statistical Classification of Diseases and Related Health Problems, 10th Revision: http://apps.who.int/classifications/icd10/browse/2010

The XLMR Update site – a catalog of XLMR conditions and gene: http://xlmr.interfree.it/home.htm

SUMMARY IN ESTONIAN

Intellektipuude genoomsed põhjused: kogu-genoomi SNP genotüpiseerimisanalüüs Eesti patsientidel ja üldpopulatsioonis

Üheks kõige populaarsemaks uurimisvaldkonnaks käesoleva aja inimgeneetikas on indiviididevaheline geneetiline varieeruvus – "võti", mis aitaks mõista, millest on tingitud iga inimese unikaalsus ja risk haiguste tekkele. Viimastel aastatel teaduses laialdaselt kasutusse jõudnud kogu-genoomi analüüsimeetodid on võimaldanud uurida inimgenoomi väga väikeste muutuste suhtes. Genoomika-ajastu tulemusena on saanud selgeks, et erinevus kahe inimese genoomide vahel võib ulatuda rohkem kui 20 miljoni aluspaarini ehk ligikaudu 0.8% kogu genoomist ning revolutsiooniliselt on muutunud arusaamine geneetiliste variatsioonide rollist arenguhäirete ja komplekshaiguste põhjustamisel. Genotüübi-põhine diagnostika on toonud meditsiinigeneetikute tööpõllule hulga uusi, kliinilist ja funktsionaalset iseloomustamist vajavaid mikrodeletsiooni- ja duplikatsioonisündroome. Nende sündroomide, mida üheskoos nimetatakse genoomseteks haigusteks, aluseks on DNA koopiaarvu variatisoonid (ingl. k. *DNA copy number variation*; *CNV*). *CNV*-d on kromosoomsegmendid, mille koopiate arv on struktuurse ümberkorralduse tõttu tavapärasest suurem või väiksem ning vastavalt nimetatakse neid duplikatsioonideks või deletsioonideks. Selliste ümberkorralduste pikkus võib olla erinev ja sageli hõlmavad need geene või teisi funktsionaalseid elemente. Tänu ulatuslikele uuringutele on *CNV*-de kaardistamine olnud viimastel aastatel väga intensiivne ning tänaseks on tekkinud ettekujutus *CNV*-de ulatusest ja tähtsusest inimegenoomis. On leitud, et *CNV*-d on vastutavad ligikaudu 10–20% intellektipuude ja teiste kaasasündinud arenguhäirete eest ning võivad olla riskifaktoriks erinevate komplekshaiguste kujunemisel. Ehkki praeguseks on jõutud veendumusele, et *CNV*-d mängivad domineerivat rolli inimestevahelises geneetilises varieeruvuses, ei ole hetkel veel ülevaatlikku arusaama, kuidas täpselt *CNV*-d geenide ekspresseerumist ning seeläbi fenotüübiliste tunnuste väljakujunemist mõjutavad. Esimesed ulatuslikud tööd *CNV*-de ja geenieksprессiooni muutuste seostest kinnitavad, et *CNV*-de mõju on põhjuseks peaaegu viiendikule muutustest geenide aktiivsuses ning võib ulatuda nii geenidele muutunud koopiaarvuga regiooniga külgnevates alades, kui mõjutada regulatsiooni kogu transkriptoomi tasemel.

Käesolev töö on esimene ulatuslik uurimus genoomsete ümberkorralduste haigusseoselisest rollist Eesti inimestel. Koostöös Tartu Ülikooli Kliinikumi Meditsiinigeneetikakeskusega analüüsiti töö käigus 77 teadmata põhjusega intellektipuudega perekonda, kokku 165 patsiendi ning 92 terve pereliikmega. Kuna rea *CNV*-de puhul võib nende mõju kliiniliste tunnuste avaldumisele olla varieeruv, teostati sama uuring paralleelselt ka 6901 Tartu Ülikooli Eesti Geenivaramu vabatahtlikule geenidoonorile.

Töö tulemusena leiti 18 peres intellektipuuet põhjustav genoomne aberratsioon. Leitud ümberkorraldustest 15 paiknevad teadaolevates kliiniliselt olulistes kromosoomregioonides ning kolmel patsiendil tuvastati uus tõenäoliselt patogeenne *CNV*. Seega määrati geneetiline diagnoos 23% uuritud patsientidest, kellel seni oli nende haiguse põhjus teadmata. Ehkki ravi genoomsete haiguste vastu ei ole käesoleval ajal veel leitud, võimaldab täpne diagnoos paremat patsientide nõustamist ja kordusriskide hindamist neis perekondades. Eesti Geenivaramu geenidoonorite hulgast leiti kliinilise tähtsusega *CNV*-d 19 indiviidil. Vastavalt Eesti Geenivaramu küsimustikule on enamus neist inimestest kannatanud aastaid erinevate tervisehäirete (näiteks tugev rasvumine, epilepsia, kõnearengu häired, depressioon, teised neuroloogilised ja psühhiaatrilised probleemid) teadmata, et nende genoomis esineb vastavate haiguste riski suurendavaid variatsioone. Teaduslikust seisukohast aitas paralleelne analüüs kohordis, mis iseloomustab Eesti üldpopulatsiooni oluliselt kaasa leitud *CNV*-de kliinilise tähtsuse hindamisele ning näitas kuivõrd oluline on hästikirjeldatud fenotüübiandmetega tavapopulatsiooni indiviide uurimine *CNV*-dega seotud haigustunnuste välja selgitamisel. Samas viitavad antud töö tulemused ka geneetilise konsultatsiooni võimaldamise vajalikkusele Eesti Geenivaramus ja teistes biopankades.

Kuna genoomsete haiguste esinemissagedus on väga madal (enamasti oluliselt alla 1% populatsioonist), on tähenduslike tulemuste saamiseks vältimatu koostöö erinevate keskuste vahel ning mitmete erinevate populatsioonide ühine standardiseeritud analüüs. Käesoleva töö raames on Eesti uuringuandmed mänginud olulist rolli kahe genoomse regiooni genotüüp-fenotüüp korrelatsioonide kindlaks määramisel, mille tulemusena on jõutud genoomsete haiguste ja nendega kaasnevate kliiniliste tunnuste iseloomustamiseni vastavates piirkondades. Neist esimese, Antverpen'i Ülikooli koordineeritud koostöö, käigus kirjeldati uut kromosoomregioonis 7q11.23 esinevat duplikatsioonisündroomi, millega kaasnevad kõnearengu hilinemine, intellektipuue, autistlikud tunnused, vastsündinuea hüpotoonia ja iseloomulik düsmorfne välimus. Teise, Lausanne`i Ülikooli juhitud projekti, tulemusena, näidati, kuidas genoomse ümberkorraldusega kaasnev vastandlik geenidoosi muutus kromosoompiirkonnas 16p11.2 põhjustab patsientidel nö. peegel-efekti kehakaalule ja peaümbermõõdule ning loob suurepärased võimalused kasutada seda piirkonda inimgenoomis mudelina, mille alusel edasi uurida rasvumise ja alakaalu molekulaarseid aluseid ning nende seost toitumishäirete ja teiste psühhiaatriliste probleemidega.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor Ants Kurg and all the members of the work group, especially Olga Žilina, Margit Nõukas, Dr. Helen Puusepp-Benazzouz, Kristina Lokotar and Merle Külaots who have played an important part in the project. I wish to thank Sven Parkel for "all the little things", and bigger ones, and most importantly for always being there to help when needed.
I am sincerely thankful to Professor Katrin Õunap and all the people at the Tartu University Hospital Department of Genetics and Children's Clinic, especially Dr. Tiia Reimand and Dr. Eve Õiglane-Šlik for sharing knowledge about the clinical aspects and never letting me forget that behind these scientifically exciting cases are real people with much less luck in their lives.
Many thanks to Professor Andres Metspalu and all the people at the IMCB Department of Biotechnology and Estonian Genome Center for their help, especially Tõnu Esko, Kairit Mikkel, Eva Reinmaa, Kaarel Krjutškov, Tarmo Annilo, and Mari Nelis. I would like to thank Professor Maris Laan for critically reading the manuscript and for making useful suggestions. And, last but far from least, Viljo Soo for having the kindest heart and providing the most technical support for this study.

For their help in the *in silico* world and for having an espresso machine accompanied by good (though sometimes fiery) discussions I would like to thank Professor Maido Remm, Priit Palta, Andres Veidenberg and Reedik Mägi at the IMCB Department of Bioinformatics.

In addition to the people in Tartu, I have been very lucky to meet so many good people on my way. My warmest thanks to Prof. Frank Kooy from University of Antwerp, coordinator of the 7q11.23 duplication study, for inviting me for the first time to the international playground of genomic disorders, for our shared collaborations and discussions, and for encouraging me in January 2008 to contact Dr. Sébastien Jacquemont in Lausanne regarding 16p11.2 patients. I am sincerely grateful to Sébastien, as well as Professor Jacqui Beckmann, the "good captain" of the 16p11.2 project, and especially Professor Alexandre Reymond for providing inspiring opportunities to pursue my scientific interests and for mentoring me in my "ongoing scientific maturation". I am thankful to the people involved in the 16p11.2 study in the CHUV University Hospital of Lausanne and the Center for Integrative Genomics, as well as all the other collaborators. My special thanks to Nele Gheldof, Robert Witwicki and Gérard Didelot for creating the atmosphere so that instead of struggling in a new place, I was surrounded by helpful friends.

I am very thankful to my family and all the people who have been part of my personal life. Three of them were not only with me during the most difficult months of combining work with writing, but contributed directly to my thesis. Ott, who patiently helped and saved me time by preparing Figures 1, 2, 5 and 6, and has made me understand how lucky I am that this young man who constantly gives reasons to be proud has choosen me to be his mother. Dave,

who in a critical situation took my urgencies before his own. Moreover, instead of a quick native-speaker's proofreading, he decided to learn until understanding, converted my Estonian nuances to English equivalents, and showed what they mean in [338] by comparing biologists' and engineers' ways of expression. Mikk, the sweetest little entertainer ever, who acted here as a gender- and age-matching control and spices up our otherwise science-centered days with Estonian and French songs, (cacophonic) performances on harmonica and other shiny-cheerful moments.

My work in the University of Lausanne has been supported by the Scix-NMS^{ch} Fellowship.

PUBLICATIONS

CURRICULUM VITAE

First and last name: Katrin Männik
Date and place of birth: April 1st 1975; Kuressaare, Estonia

Contact information

Address: Center for Integrative Genomics, Génopode Building
University of Lausanne, 1015 Lausanne, Switzerland
Phone: + 41(0) 21 692 3991
Email: Katrin.Mannik@unil.ch

Education and degree information

Since 2011	PhD studies, Center for Integrative Genomics, University of Lausanne, Switzerland
Since 2009	PhD studies in gene technology, Faculty of Science and Technology, University of Tartu, Estonia
2001–2004	MSc in molecular diagnostics, Faculty of Biology and Geography, University of Tartu, Estonia
1996–2001	BSc in biomedicine and biotechnology, Faculty of Biology and Geography, University of Tartu, Estonia
1993– 1996	Faculty of Medicine, University of Tartu

Professional experience

Since 2008	Researcher, Institute of Molecular and Cell Biology, University of Tartu
2006–2008	Extraordinary Researcher, Institute of Molecular and Cell Biology, University of Tartu
2004–2006	Extraordinary Researcher, Estonian Biocentre
2001–2004	Senior Lab Assistant, Estonian Biocentre

Additional professional training

June 2012	Swiss Institute of Bioinformatics, “Analysis of differential gene expression”, Lausanne, Switzerland
Sept. 2011	CHUV University Hospital, Lausanne, cardiovascular and metabolism PhD course “Brain nutrients and hormone sensing in feeding and glycemic control”, Lausanne, Switzerland
July 2011	ICVS, University of Minho postgraduate course “Mental Retardation: from clinic to gene and back”, Braga, Portugal

2010	Faculty of Medicine, University of Tartu "Biomedical ethics"; Tartu, Estonia
Jan. 2010	Wellcome Trust Conference Centre, Clinical Genetics Society, Specialist Advisory Committee of the Royal College Physicians "Fundamentals of Clinical Genetics"; Hinxton, United Kingtom
Sept. 2008	EU FP6 "Marie Curie – Genome Architecture in Relation to Disease", Workshop "Genome bioinformatic techniques"; Braga, Portugal
Sept. 2007	EU FP6 "Marie Curie – Genome Architecture in Relation to Disease", Workshop "Array techniques to identify copy number variations"; Helsinki, Finland
Sept. 2003	European School of Genetic Medicine "5th Course in Molecular Cytogenetics and DNA Arrays"; Bertinoro, Italy

Honours and Fellowships

2011	Jérôme Lejeune Foundation; fellowship for participating in additional professional training
2011	Sciex-NMS^{ch} research fellowship for working in Prof. Alexandre Reymond's group, Center for Integrative Genomics, University of Lausanne, Switzerland
2009	European Society of Human Genetics; national fellowship for young investigator
2008	Scientific Committe of 10th Annual Conference of Estonian Society of Human Genetics; best poster presentation award
2007	Scientific Commitee of 13th International Workshop on Fragile X and X-linked Mental Retardation; travel fellowship for young investigator
2004	Estonian Academy of Sciences, MSc thesis awarded with first prize in student research

Publications

- Preiksaitiene E, **Männik K**, Dirse V, Utkus A, Ciuladaite Z, Kasnauskiene J, Kurg A, Kucinskas V „A novel de novo 1.8 Mb microdeletion of 17q21.33 associated with intellectual disability and dysmorphic features", **Eur J Med Genet**, **2012** Jul 26
- Žilina O, Reimand T, Zjabolovskaja P, **Männik K**, Männamaa M, Traat A, Puusepp- Benazzouz H, Kurg A, Õunap K "Maternally and paternally inherited deletion of 7q31 involving the FOXP2 gene in two families", **Am J Med Genet, 2011** Nov 21
- Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, Martinet D, Shen Y, Valsesia A, Beckmann ND, Thorleifsson G, Belfiore M, Bouquillon S, Campion D, de Leeuw N, de Vries BB, Esko T, Fernandez

BA, Fernández-Aranda F, Fernández-Real JM, Gratacòs M, Guilmatre A, Hoyer J, Jarvelin MR, Kooy RF, Kurg A, Le Caignec C, **Männik K**, Platt OS, Sanlaville D, Van Haelst MM, Villatoro Gomez S, Walha F, Wu BL, Yu Y, Aboura A, Addor MC, Alembik Y, Antonarakis SE, Arveiler B, Barth M, Bednarek N, Béna F, Bergmann S, Beri M, Bernardini L, Blaumeiser B, Bonneau D, Bottani A, Boute O, Brunner HG, Cailley D, Callier P, Chiesa J, Chrast J, Coin L, Coutton C, Cuisset JM, Cuvellier JC, David A, de Freminville B, Delobel B, Delrue MA, Demeer B, Descamps D, Didelot G, Dieterich K, Disciglio V, Doco-Fenzy M, Drunat S, Duban-Bedu B, Dubourg C, El-Sayed Moustafa JS, Elliott P, Faas BH, Faivre L, Faudet A, Fellmann F, Ferrarini A, Fisher R, Flori E, Forer L, Gaillard D, Gerard M, Gieger C, Gimelli S, Gimelli G, Grabe HJ, Guichet A, Guillin O, Hartikainen AL, Heron D, Hippolyte L, Holder M, Homuth G, Isidor B, Jaillard S, Jaros Z, Jiménez-Murcia S, Helas GJ, Jonveaux P, Kaksonen S, Keren B, Kloss-Brandstätter A, Knoers NV, Koolen DA, Kroisel PM, Kronenberg F, Labalme A, Landais E, Lapi E, Layet V, Legallic S, Leheup B, Leube B, Lewis S, Lucas J, MacDermot KD, Magnusson P, Marshall C, Mathieu-Dramard M, McCarthy MI, Meitinger T, Mencarelli MA, Merla G, Moerman A, Mooser V, Morice-Picard F, Mucciolo M, Nauck M, Ndiaye NC, Nordgren A, Pasquier L, Petit F, Pfundt R, Plessis G, Rajcan-Separovic E, Ramelli GP, Rauch A, Ravazzolo R, Reis A, Renieri A, Richart C, Ried JS, Rieubland C, Roberts W, Roetzer KM, Rooryck C, Rossi M, Saemundsen E, Satre V, Schurmann C, Sigurdsson E, Stavropoulos DJ, Stefansson H, Tengström C, Thorsteinsdóttir U, Tinahones FJ, Touraine R, Vallée L, van Binsbergen E, Van der Aa N, Vincent-Delorme C, Visvikis-Siest S, Vollenweider P, Völzke H, Vulto-van Silfhout AT, Waeber G, Wallgren-Pettersson C, Witwicki RM, Zwolinksi S, Andrieux J, Estivill X, Gusella JF, Gustafsson O, Metspalu A, Scherer SW, Stefansson K, Blakemore AI, Beckmann JS, Froguel P "Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus" **Nature 2011** Aug 31;478(7367):97–102

- **Männik K**, Parkel S, Palta P, Žilina 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 the general population in Estonia" **Eur J Med Genet 2011** Mar-Apr;54(2):136–43
- 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, Blakemore 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

- Van der Aa N, Rooms L, Vandeweyer G, van den Ende J, Reyniers E, Fichera M, Romano C, Delle Chiaie B, Mortier G, Menten B, Destrée A, Maystadt I, **Männik K**, Kurg A, Reimand T, McMullan D, Oley C, Brueton L, Bongers EM, van Bon BW, Pfund R, Jacquemont S, Ferrarini A, Martinet D, Schrander-Stumpel C, Stegmann AP, Frints SG, de Vries BB, Ceulemans B, Kooy RF "Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome" Eur J Med Genet **2009** Mar–Jun;52(2–3):94–100
- Puusepp H, Zilina O, Teek R, **Männik K**, Parkel S, Kruustük K, Kuuse K, Kurg A, Ounap K "5.9 Mb microdeletion in chromosome band 17q22-q23.2 associated with tracheo-esophageal fistula and conductive hearing loss" Eur J Med Genet **2009** Jan–Feb;52(1):71–4
- Kousoulidou L*, **Männik K***, Sismani C, Žilina 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
  * These authors contributed equally to this work
- Kousoulidou L, **Männik K**, Zilina O, Parkel S, Palta P, Remm M, Kurg A, and 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** May; 46(5):722–4
- Patsalis PC, Kousoulidou L, **Mannik K**, Sismani C, Zilina O, Parkel S, Puusepp H, Tonisson N, Palta P, Remm M, Kurg A. "Detection of small genomic imbalances using microarray-based multiplex amplifiable probe hybridisation" **Eur J Hum Genet 2007** Feb;15(2):162–72
- Puusepp H, **Männik K**, Zilina O, Parkel S, Kurg A, Õunap K "Vaimse arengu mahajäämuse geneetilised põhjused: X-liiteline vaimse arengu mahajäämus" **Eesti Arst 2007** 4:239–45 (in Estonian)
- Patsalis PC, Kousoulidou L, Sismani C, **Mannik K**, Kurg A "MAPH: from gels to microarrays" **Eur J Med Genet 2005** Jul–Sep;48(3):241–9

ELULOOKIRJELDUS

Ees- ja perekonnanimi: Katrin Männik
Sünniaeg ja -koht: 1. aprill 1975; Kuressaare, Eesti

Kontaktandmed

Aadress:	Center for Integrative Genomics, Génopode Building University of Lausanne, 1015 Lausanne, Switzerland
Telefon:	+ 41(0) 21 692 3991
E-post:	Katrin.Mannik@unil.ch

Haridus ja omandatud kraadid

Alates 2011	doktorant, Integreeritud Genoomika Keskus, Lausanne'i Ülikool, Šveits
Alates 2009	doktorant, geenitehnoloogia õppekava, loodus- ja tehnoloogia-teaduskond, Tartu Ülikool, Eesti
2001–2004	MSc, molekulaardiagnostika õppekava, bioloogia-geograafia-teaduskond, Tartu Ülikool, Eesti
1996–2001	BSc, biomeditsiini ja biotehnoloogia õppekava, bioloogia geograafiateaduskond, Tartu Ülikool, Eesti
1993–1996	ravi eriala, arstiteaduskond, Tartu Ülikool, Eesti

Teenistuskäik

Alates 2008	teadur, Molekulaar- ja Rakubioloogia Instituut, Tartu Ülikool, Eesti
2006–2008	erakorraline teadur, Molekulaar- ja Rakubioloogia Instituut, Tartu Ülikool, Eesti
2004–2006	erakorraline teadur, Eesti Biokeskus, Eesti
2001–2004	vanemlaborant, Eesti Biokeskus, Eesti

Täiendkoolitus

Juuni 2012	Šveitsi Bioinformaatika Instituut, "Analysis of differential gene expression", Lausanne, Šveits
Sept 2011	Lausanne'i Ülikool, Vaud'i Ülikooli Keskhaigla südame-vere-soonkonna ja metabolismi doktorikursus "Brain nutrients and hormone sensing in feeding and glycemic control", Lausanne, Šveits

Juuli 2011	ICVS, Minho Ülikool täienduskursus "Mental Retardation: from clinic to gene and back", Braga, Portugal
2010	Arstiteaduskond, Tartu Ülikool "Biomeditsiini eetika"; Tartu, Eesti
Jan 2010	Wellcome Trust'i konverentsikeskus, Clinical Genetics Society, Specialist Advisory Committee of the Royal College Physicians "Fundamentals of Clinical Genetics"; Hinxton, Suurbritannia
Sep 2008	EU FP6 Marie Curie – Genome Architecture in Relation to Disease 2nd Workshop "Genome bioinformatic techniques"; Braga, Portugal
Sept 2007	EU FP6 Marie Curie – Genome Architecture in Relation to Disease 1st Workshop "Array techniques to identify copy number variations"; Helsinki, Soome
Sep 2003	European School of Genetic Medicine "5th Course in Molecular Cytogenetics and DNA Arrays"; Bertinoro, Itaalia

Tunnustused ja stipendiumid

2011	Jérôme Lejeune Foundation; stipendium täiendkoolitusel osalemiseks
2011	Sciex-NMSch teadusstipendium töötamiseks Prof. Alexandre Reymond'i töögrupis, Centre Intégratif de Génomique, Université de Lausanne
2009	Euroopa Inimesegeneetika Ühing; rahvuslik stipendium noorteadlasele
2008	Eesti Inimesegeneetika Ühingu 10. Aastakonverentsi teaduskomitee; parima posterettekande auhind
2007	13th International Workshop on Fragile X and X-linked Mental Retardation teaduskomitee; reisistipendium noorteadlasele
2004	Eesti Teaduste Akadeemia, magistritöö autasustatud I preemiaga üliõpilastööde konkursil

Avaldatud teadusartiklid

- Preiksaitiene E, **Männik K**, Dirse V, Utkus A, Ciuladaite Z, Kasnauskiene J, Kurg A, Kucinskas V „A novel de novo 1.8 Mb microdeletion of 17q21.33 associated with intellectual disability and dysmorphic features", **Eur J Med Genet**, **2012** Jul 26
- Žilina O, Reimand T, Zjabolovskaja P, **Männik K**, Männamaa M, Traat A, Puusepp-Benazzouz H, Kurg A, Õunap K "Maternally and paternally inherited deletion of 7q31 involving the FOXP2 gene in two families", **Am J Med Genet, 2011** Nov 21;
- Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, Martinet D, Shen Y, Valsesia A, Beckmann ND, Thorleifsson G, Belfiore M, Bouquillon S, Campion D, de Leeuw N, de Vries BB, Esko T, Fernandez BA, Fernández-Aranda F, Fernández-Real JM, Gratacòs M, Guilmatre A,

Hoyer J, Jarvelin MR, Kooy RF, Kurg A, Le Caignec C, **Männik K**, Platt OS, Sanlaville D, Van Haelst MM, Villatoro Gomez S, Walha F, Wu BL, Yu Y, Aboura A, Addor MC, Alembik Y, Antonarakis SE, Arveiler B, Barth M, Bednarek N, Béna F, Bergmann S, Beri M, Bernardini L, Blaumeiser B, Bonneau D, Bottani A, Boute O, Brunner HG, Cailley D, Callier P, Chiesa J, Chrast J, Coin L, Coutton C, Cuisset JM, Cuvellier JC, David A, de Freminville B, Delobel B, Delrue MA, Demeer B, Descamps D, Didelot G, Dieterich K, Disciglio V, Doco-Fenzy M, Drunat S, Duban-Bedu B, Dubourg C, El-Sayed Moustafa JS, Elliott P, Faas BH, Faivre L, Faudet A, Fellmann F, Ferrarini A, Fisher R, Flori E, Forer L, Gaillard D, Gerard M, Gieger C, Gimelli S, Gimelli G, Grabe HJ, Guichet A, Guillin O, Hartikainen AL, Heron D, Hippolyte L, Holder M, Homuth G, Isidor B, Jaillard S, Jaros Z, Jiménez-Murcia S, Helas GJ, Jonveaux P, Kaksonen S, Keren B, Kloss-Brandstätter A, Knoers NV, Koolen DA, Kroisel PM, Kronenberg F, Labalme A, Landais E, Lapi E, Layet V, Legallic S, Leheup B, Leube B, Lewis S, Lucas J, MacDermot KD, Magnusson P, Marshall C, Mathieu-Dramard M, McCarthy MI, Meitinger T, Mencarelli MA, Merla G, Moerman A, Mooser V, Morice-Picard F, Mucciolo M, Nauck M, Ndiaye NC, Nordgren A, Pasquier L, Petit F, Pfundt R, Plessis G, Rajcan-Separovic E, Ramelli GP, Rauch A, Ravazzolo R, Reis A, Renieri A, Richart C, Ried JS, Rieubland C, Roberts W, Roetzer KM, Rooryck C, Rossi M, Saemundsen E, Satre V, Schurmann C, Sigurdsson E, Stavropoulos DJ, Stefansson H, Tengström C, Thorsteinsdóttir U, Tinahones FJ, Touraine R, Vallée L, van Binsbergen E, Van der Aa N, Vincent-Delorme C, Visvikis-Siest S, Vollenweider P, Völzke H, Vulto-van Silfhout AT, Waeber G, Wallgren-Pettersson C, Witwicki RM, Zwolinksi S, Andrieux J, Estivill X, Gusella JF, Gustafsson O, Metspalu A, Scherer SW, Stefansson K, Blakemore AI, Beckmann JS, Froguel P "Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus" **Nature 2011** Aug 31;478(7367):97–102

- **Männik K**, Parkel S, Palta P, Žilina 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 the general population in Estonia" **Eur J Med Genet 2011** Mar-Apr;54(2):136–43
- 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, Blakemore 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

- Van der Aa N, Rooms L, Vandeweyer G, van den Ende J, Reyniers E, Fichera M, Romano C, Delle Chiaie B, Mortier G, Menten B, Destrée A, Maystadt I, **Männik K**, Kurg A, Reimand T, McMullan D, Oley C, Brueton L, Bongers EM, van Bon BW, Pfund R, Jacquemont S, Ferrarini A, Martinet D, Schrander-Stumpel C, Stegmann AP, Frints SG, de Vries BB, Ceulemans B, Kooy RF "Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome" Eur J Med Genet **2009** Mar–Jun;52(2–3):94–100
- Puusepp H, Zilina O, Teek R, **Männik K**, Parkel S, Kruustük K, Kuuse K, Kurg A, Ounap K "5.9 Mb microdeletion in chromosome band 17q22-q23.2 associated with tracheo-esophageal fistula and conductive hearing loss" Eur J Med Genet **2009** Jan–Feb;52(1):71–4
- Kousoulidou L*, **Männik K***, Sismani C, Žilina 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
  * These authors contributed equally to this work
- Kousoulidou L, **Männik K**, Zilina O, Parkel S, Palta P, Remm M, Kurg A, and 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** May; 46(5):722–4
- Patsalis PC, Kousoulidou L, **Mannik K**, Sismani C, Zilina O, Parkel S, Puusepp H, Tonisson N, Palta P, Remm M, Kurg A. "Detection of small genomic imbalances using microarray-based multiplex amplifiable probe hybridisation" **Eur J Hum Genet 2007** Feb;15(2):162–72
- Puusepp H, **Männik K**, Zilina O, Parkel S, Kurg A, Õunap K "Vaimse arengu mahajäämuse geneetilised põhjused: X-liiteline vaimse arengu mahajäämus" **Eesti Arst 2007** 4:239–45 (in Estonian)
- Patsalis PC, Kousoulidou L, Sismani C, **Mannik K**, Kurg A "MAPH: from gels to microarrays" **Eur J Med Genet 2005** Jul–Sep;48(3):241–9

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 crassland 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 erotonin. 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, 2008, 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 intracellular 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.