KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Human Genetics

KU LEUVEN

DOCTORAL SCHOOL
BIOMEDICAL SCIENCES

MOLECULAR DIAGNOSTICS IN AUTISM SPECTRUM DISORDER

NEXT GENERATION SEQUENCING AND FACIAL PHENOTYPING IN MEDICAL GENETICS

Yoeri SLEYP

Jury

Supervisor:	Prof. dr. Hilde Peeters
Co-supervisor:	Prof. dr. Peter Claes
Chair examining committee:	Prof. dr. Karel Allegaert
Chair public defence:	Prof. dr. Eric Legius
Jury members:	Prof. dr. Benedikt Hallgrímsson
	Prof. dr. Bert Callewaert
	Prof. dr. Els Ortibus
	Prof. dr. Jeroen Breckpot

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor in Biomedical Sciences

October 2022

Author: Yoeri Sleyp
Cover design: Yoeri Sleyp

Printed by Procopia

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

3D: three-dimensional

ADHD: attention-deficit/hyperactivity disorder

ASD: autism spectrum disorder

ASID: Autism Spectrum/Intellectual Disability

AUC: area under the curve

BMP: bone morphogenetic protein

CCDG: Centers for Common Disease Genomics

CNCC: cranial neural crest cells

CNV: copy number variant

CMA: chromosomal microarray

DSM: Diagnostic and Statistical Manual of Mental Disorders

FGF: fibroblast growth factor

GWAS: genome-wide association study

ID: intellectual disability

Indel: insertion or deletion

IQ: intelligence quotient

gnomAD: Genome Aggregation Database

kb: kilobase

LCR: low copy repeat

LD: linkage disequilibrium

LGD: likely gene-disruptive

MAF: minor allele frequency

Mb: megabase

MIS30: missense variants with a combined annotation dependent depletion score above 30

MPA: minor physical anomaly

MRI: magnetic resonance imaging

NDD: neurodevelopmental disorder

NGS: next generation sequencing

PDD: pervasive developmental disorder

PDD-NOS: pervasive developmental disorder not otherwise specified

PRS: polygenic risk score

ROC: receiver operating characteristic

SNP: single nucleotide polymorphism

SNV: single nucleotide variant

SV: structural variant

WES: whole exome sequencing

WGS: whole genome sequencing

STATEMENT

This thesis is aimed at improving molecular diagnostics for individuals with autism who come to consultation with a request for a genetic workup. Therefore, the findings in this thesis should not be extrapolated to the broader autism community. The work described in this thesis is not aimed at diagnosing, curing, or eradicating autism, nor is it intended to contribute to such type of research.

CHAPTER 1

GENERAL INTRODUCTION

1 General introduction

1.1 Autism diagnosis

1.1.1 The complicated history of autism diagnosis

In 1908, Eugen Bleuler coined the term "autism" to describe a detachment from reality in individuals with schizophrenia [1]. The idea for using autism as a diagnostic concept was however started by Leo Kanner in 1943 when he published the paper "Autistic Disturbances of Affective Contact" [2]. He described children who displayed social isolation, abnormal language development, sensory sensitivities and repetitive behaviors [2]. These observations were groundbreaking because the children's clinical presentation did not fit within known psychiatric disorders at that time but rather constituted a novel clinical entity, which later became known as "early infantile autism" [3]. In 1944, Hans Asperger reported on "autistic psychopathy" in children who particularly had social difficulties and restrictive interests, but who did not have the language problems described in Leo Kanner's work [4]. Hans Asperger therefore also described a novel autistic-like disorder, which was later used to define "Asperger's Syndrome" [5].

Although the work of Leo Kanner and Hans Asperger laid the groundwork for the development of autism as a diagnostic concept, it took until 1980 before the first autistic-like disorders were included in the Diagnostic and Statistical Manual of Mental Disorders (DSM), the widely used manual to diagnose and classify mental disorders. The DSM-III established the concept of autism under a new class of conditions called pervasive developmental disorders (PDDs), which contained infantile autism, childhood-onset PDD and atypical PDD [6]. The diagnostic criteria of these disorders focused primarily on problems with social skills, language, and behavior. To allow for more flexible criteria that would be useful across ages and developmental levels, DSM-III-R revised the concept of autism in 1987 and changed the PDD subclasses to autistic disorder and PDD not otherwise specified (PDD-NOS) [6]. The criteria were also divided into three domains of impairment: 1) impairments in reciprocal social interaction, 2) impairments in communication, and 3) restricted interests, resistance to change and repetitive movements [6]. The next change occurred in 1994 when the DSM-IV introduced three novel PDDs: Asperger's syndrome, Rett's disorder and child disintegrative disorder [6]. The concept of autism now consisted of five PDDs each of which was linked to a specific set of diagnostic criteria. However, the large heterogeneity in the number and severity of autism symptoms both within and across these PDDs made it difficult to distinguish them from each other [6], suggesting that these disorders represent a continuum of the same disorder rather than separate disorders. An important shift in the conceptualization of autism happened in 2013 when the DSM-V introduced the umbrella term autism spectrum disorder (ASD) encompassing autistic disorder, PDD-NOS, Asperger's syndrome

and child disintegrative disorder [6]. This change ended the many attempts to categorize the clinical heterogeneity of autism into subgroups. The DSM-V diagnostic criteria are discussed below.

1.1.2 The diagnosis of ASD

ASD is a neurodevelopmental disorder (NDD) characterized by social and communication deficits, as well as restricted interests and repetitive behaviors [7]. The diagnostic criteria for ASD specified in the DSM-V are currently the most widely used worldwide. The core criteria are divided in two symptom domains: A) persistent deficits in social communication and social interaction, and B) restricted, repetitive patterns of behavior, interests, or activities (**table 1.1**). To meet the diagnostic criteria for ASD according to the DSM-V, a person must have all three symptoms in the domain of social communication and interaction (A1 through A3 in **table 1.1**) plus at least two of the four symptoms in the domain of behavior (B1 through B4 in **table 1.1**) [7]. Additional criteria state that the symptoms must be present in the early developmental period, cause clinically significant impairment in social, occupational, or other important areas of current functioning, and should not be better explained by intellectual disability (ID) or global developmental delay [7]. The current estimated prevalence of ASD in Europe and the United States ranges between 0.38% and 1.85% of the population [8]. These estimates however vary widely between and within countries [8].

1.1.3 The clinical heterogeneity of ASD

An important feature of ASD is its clinical heterogeneity. The "autism spectrum" refers to the strong individual variations in the type and severity of ASD symptoms. The diagnosis of ASD can manifest itself very differently depending where an individual is on the spectrum (see examples **table 1.1**). Consequently, the DSM-V introduced three levels of ASD symptom severity that indicate how much support a person with ASD needs in daily life in the areas of social interaction/communication and behavior [7]. Co-morbidities are also commonly observed in individuals with ASD and often contribute to the severity of ASD symptoms. ASD is frequently co-diagnosed with ID [8], which complicates the assessment of ASD symptoms. Co-occurring mental health diagnoses are also often reported such as attention-deficit/hyperactivity disorder (ADHD), anxiety disorders, sleep-wake disorders, and depressive disorders [8–10]. In addition, ASD often co-occurs with gastrointestinal disorders and with neurological disorders such as epilepsy and migraine/headaches [8,11].

Table 1.1: DSM-V criteria for ASD. Table adapted from Hyman et al. [12].		
Domains	**Symptoms**	**Examples**
A. Persistent deficits in social communication and social interaction across multiple contexts, as manifested by the following, currently or by history; must have all 3 symptoms in this domain	1. Deficits in social-emotional reciprocity	Abnormal social approach and failure of normal back-and-forth conversation; reduced sharing of interests, emotions, or affect; failure to initiate or respond to social interactions
	2. Deficits in nonverbal communicative behaviors used for social interaction	Poorly integrated verbal and nonverbal communication; abnormalities in eye contact and body language or deficits in understanding and use of gestures; total lack of facial expressions and nonverbal communication
	3. Deficits in developing, maintaining, and understanding relationships	Difficulties adjusting behavior to suit various social contexts; difficulties in sharing imaginative play or in making friends; absence of interest in peers
B. Restricted, repetitive patterns of behavior, interests, or activities, as manifested by at least 2 of the following, currently or by history; must have 2 of the 4 symptoms	1. Stereotyped or repetitive motor movements, use of objects, or speech	Simple motor stereotypies, lining up toys or flipping objects, echolalia, idiosyncratic phrases
	2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns or verbal nonverbal behavior	Extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat food every day
	3. Highly restricted, fixated interests that are abnormal in intensity or focus	Strong attachment to or preoccupation with unusual objects, excessively circumscribed or perseverative interest
	4. Hyper- or hyporeactivity to sensory input or unusual interests in sensory aspects of the environment	Apparent indifference to pain/temperature, adverse response to specific sounds or textures, excessive smelling or touching of objects, visual fascination with lights or movement

1.2 The genetic architecture of ASD

1.2.1 Epidemiology and heritability of ASD

Epidemiological studies have demonstrated that ASD has a strong genetic basis. Twin studies reported higher concordance rates in monozygotic twins as compared to dizygotic twins [13]. A meta-analysis of twin studies estimated the heritability of ASD between 64% and 91%, indicating that genetic factors explain a large proportion of the phenotypic variance of ASD [14]. In contrast, shared environmental factors are expected to account for little of the phenotypic variance of ASD [14]. Family-based population studies have reported increased rates of ASD in children of families with a history of ASD [15–18]. The relative recurrence risk of ASD in a family-based population study was 153.0 for monozygotic twins, 8.2 for dizygotic twins, 10.3 for full siblings, 3.3 for maternal half siblings, 2.9 for paternal half siblings, and 2.0 for cousins [18]. This study shows that, in families with a history of ASD, the risk for ASD in a child depends on the amount of genetic information that is shared between the child and an affected family member [19]. Stronger genetic relatedness between the child and the affected family member increases the risk of ASD in the child, indicating that the etiology of ASD has a strong genetic component.

1.2.2 Genetic risk factors in ASD

With the strong technological advancements in genetics in the past decades, a lot of progress has been made in understanding the genetic variation that underlies ASD risk. Like other complex disorders, genetic variants with different minor allele frequencies (MAF) and phenotypic effect sizes have been identified as ASD risk factors (**figure 1.1**). Common genetic variants (MAF > 5%) with small effect sizes collectively explain a large proportion of the variance in ASD liability but each contribute little to individual liability [20]. Genome-wide association studies (GWAS) of ASD have not been very successful so far in identifying common genetic variants associated with ASD. The largest GWAS to date included 18381 individuals with ASD and 27969 controls but found only five significant loci associated with ASD [21]. Hence, the genomic regions harboring common risk variants of ASD are not well characterized. Low-frequency variants (MAF 1-5%) with intermediate effect sizes also contribute to the variance in ASD liability but it is not known to what extent. Rare variants (MAF < 1%) with large effect sizes collectively explain only a modest proportion of the variance in ASD liability but each contribute substantially to individual liability [20]. Hundreds of genes and genomic regions have been linked to ASD through rare variant association studies [22]. Therefore, our knowledge on genes involved in ASD risk originates mainly from research on rare variants [23].

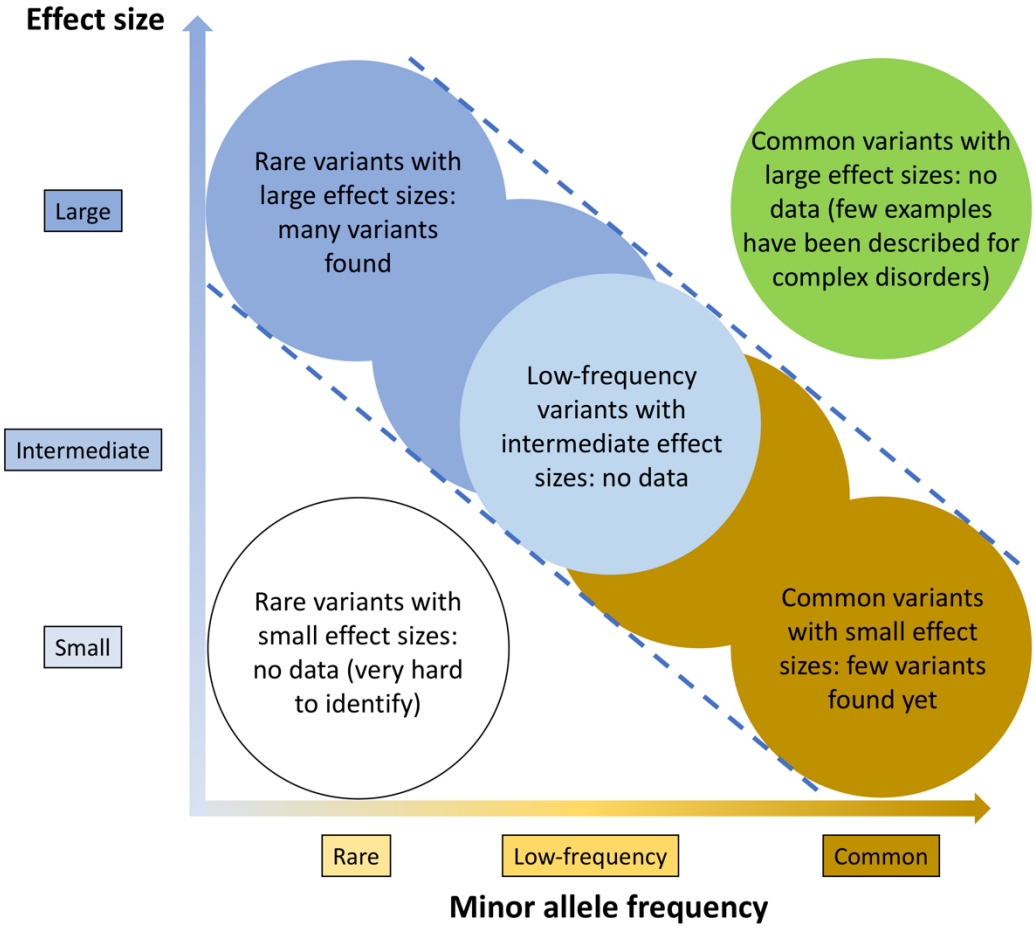


Figure 1.1: Effect sizes and allele frequencies of genetic risk variants associated with ASD. Rare variants with large effect sizes, low-frequency variants with intermediate effect sizes and common variants with small effect sizes are expected to shape the genetic architecture of ASD. Most genetic risk variants that have been associated with ASD so far are rare variants with large effect sizes. Concept adapted from Manolio et al. [24].

1.2.3 Rare variant association studies in ASD

The greatest progress in understanding the genetic architecture of ASD has been made through rare variant association studies. These studies typically study the burden of rare genetic variants between individuals with ASD and controls, while taking into account the mode of inheritance of the variants and the type of genetic variation. They typically distinguish between variants that occurred *de novo* and variants that were inherited from a parent. Different types of rare genetic variants have been identified as ASD risk factors including single nucleotide variants (SNVs), small insertions or deletions (indels) and structural variants (SVs). SVs are defined as DNA rearrangements of 50 bases or larger in size. The most studied type of SVs in genetic research on ASD are copy number variants (CNVs), which include both deletions and duplications. CNVs can contribute to ASD risk through various mechanisms such as gene dosage, gene disruption or position effects [25]. Studies have reported a significant increase in rare CNVs, especially *de novo*

events affecting multiple genes, in individuals with ASD as compared to unaffected siblings and controls [26–32]. SNVs and indels are defined as small genetic variants that affect respectively one base or up to 49 bases of DNA. Most SNVs and indels that have been currently linked to ASD are in the protein coding regions of the genome (i.e., the exome). These variants mainly contribute to ASD risk by affecting protein function through loss-of-function, gain-of-function, and dominant-negative effects. Studies have found a significant increase in rare SNVs and indels, particularly likely gene-disruptive (LGD) and *de novo* variants, in individuals with ASD as compared to unaffected siblings and controls [33–42]. LGD variants, also referred to as protein-truncating variants, are predicted to shorten the coding sequence of genes, often resulting in loss of protein function. They typically include nonsense, splice-site, and frameshift variants.

Altogether, rare variant association studies show that individuals with ASD harbor more rare *de novo* variants than expected by chance. Although such *de novo* variants explain little of the variance in ASD liability, they are major contributors to individual liability for ASD. Some studies have also reported higher rates of rare *de novo* variants in individuals with ASD from simplex families (i.e. families with only one affected individual) as compared to multiplex families (i.e. families with multiple affected individuals) [26,32]. However, this probably depends on how the recruitment of individuals with ASD was done since other studies did not observe differences in the rates of *de novo* variants between individuals of simplex families and multiplex families [28,29]. Nevertheless, many studies have established that rare *de novo* variants play an important role in the etiology of autism in simplex families since they may be a clear explanation for the sporadic presentation of ASD in these families [26,30–38].

1.2.4 Genetic causes of ASD

The search for genetic causes of ASD is complicated since various types of genetic variants with different allele frequencies, effect sizes and modes of inheritance across hundreds of genomic regions can contribute to individual liability for ASD. In addition, these genomic regions, when mutated, do not only confer risk for ASD but also for ID and other NDDs [43]. The genetic risk variants identified so far are also not sufficient to cause ASD on their own. Instead, a combination of genetic risk variants, along with non-genetic factors, is needed to surpass a risk threshold that results in ASD (**figure 1.2**) [19,20,23,44,45]. This makes it difficult to understand the precise genetic cause of ASD in most individuals.

In clinical genetics, an etiological distinction is usually made between Mendelian causes and multifactorial causes of ASD. Mendelian causes are driven by highly penetrant rare variants with large effect sizes, usually occurring *de novo*, that confer high risk for ASD (**figure 1.2A**) [45]. Other

genetic variants in the genome (i.e. genetic background) will ultimately determine whether an individual develops ASD by contributing to the genetic burden or by modulating the effects of rare variants (**figure 1.2A**) [45]. The genetic background will therefore influence the penetrance of the rare variants and contribute to the variable expressivity and severity in individuals with Mendelian causes of ASD [45]. In contrast, multifactorial causes of ASD are driven by a combination of many genetic variants with variable allele frequencies and effect sizes, as well as by environmental factors (**figure 1.2B-D**) [45]. For example, a lowly penetrant rare risk variant (**figure 1.2B**), a high burden of common risk variants (**figure 1.2C**) or environmental factors (**figure 1.2D**) together with a substantial amount of other genetic variants and non-genetic factors can result in multifactorial causes of ASD [45]. Most individuals with ASD have a multifactorial cause, while a minority has a Mendelian cause. There is currently no clear answer to which percentage of individuals with ASD have a Mendelian cause because genetic testing is usually done in clinical populations rather than unselected populations.

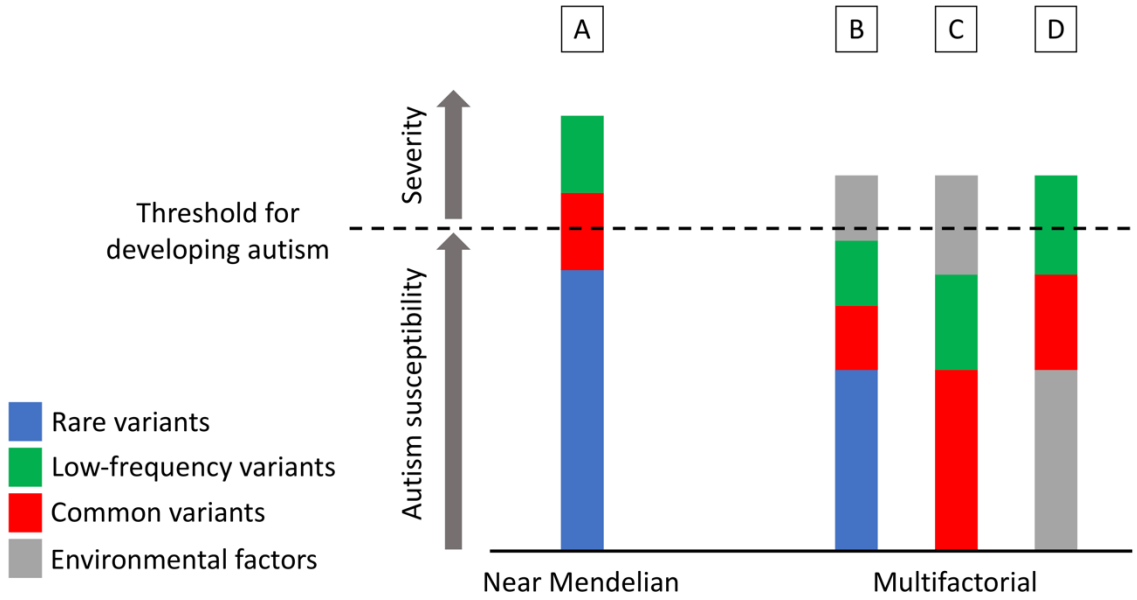


Figure 1.2: Schematic overview of the genetic causes of ASD. A combination of genetic risk variants, along with non-genetic factors, is needed to surpass a risk threshold that results in ASD. Causes can be divided in two categories: Mendelian and multifactorial. Mendelian causes of ASD are driven by highly penetrant rare variants with large effect sizes that combine with other genetic variants to cause ASD (panel A). Multifactorial causes of ASD are driven by a combination of many genetic variants with variable allele frequencies and effect sizes, as well as by environmental factors. Multifactorial causes can occur in the presence of a lowly penetrant rare risk variant (panel B), high burden of common risk variants (panel C) or environmental factors (panel D) and a substantial amount of other genetic variants and non-genetic factors. Concept adapted from Walsh et al. [45].

1.2.5 Biological mechanisms underlying ASD risk

Advances in genetics of ASD have led to a better understanding of the biological mechanisms underlying ASD risk. ASD risk genes have been shown to be predominantly expressed during prenatal brain development [39,46–53]. Furthermore, ASD risk genes have been found to be prenatally expressed in most brain regions linked to ASD such as the cortex, cerebellum, amygdala, hippocampus, and striatum [50,53]. This implicates that the cause of ASD usually originates from the prenatal period and can involve different brain regions [54]. Aberrations of various neurodevelopmental processes across different developmental stages have been linked to ASD including abnormal cell proliferation, neurogenesis, cell migration, cell fate, neurite outgrowth, synaptogenesis, synapse function and neural network formation/functioning (**figure 1.3)** [39,47–49,51–53]. Many ASD risk genes affect multiple of these processes and developmental stages (**figure 1.3**) [53]. The pathogenesis of ASD can start as early as the first trimester and continue into the postnatal period (**figure 1.3**) [53,54]. Thus, in most individuals, ASD likely results from disruptions of several neurodevelopmental processes during multiple developmental stages rather than from disruption of one process during one developmental stage [53,54].

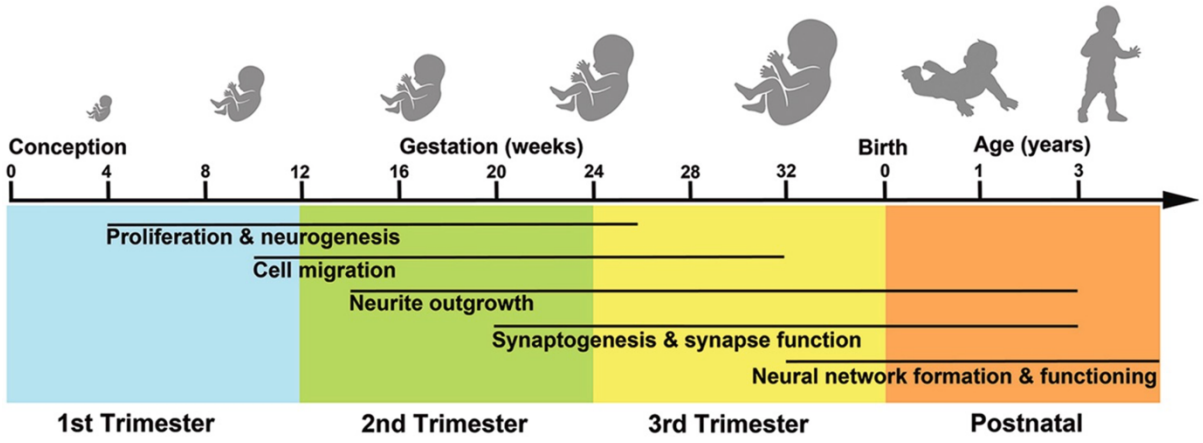


Figure 1.3: Schematic overview of the neurodevelopmental processes and developmental stages that are implicated in the pathogenesis of ASD. Figure adapted from Courchesne et al. [53].

1.3 Molecular diagnostics for ASD

1.3.1 Genetic testing for ASD

Current molecular diagnostics for ASD is aimed at finding highly penetrant rare variants that drive Mendelian causes. The reason is that the effect sizes of these rare variants are large enough to be clinically actionable. In contrast, variants contributing to multifactorial causes are currently not clinically actionable. Below we provide an overview of the most used tests in molecular diagnostics for the detection of rare variants and the diagnostic yield of these tests for individuals with ASD:

- G-banded karyotyping allows to detect large chromosomal abnormalities up to 3 megabases (Mb) but abnormalities smaller than 10 Mb are often missed [55]. Rare large chromosomal abnormalities are currently recognized as important risk factors for ASD. Typical examples are sex chromosome aneuploidies (47,XXY; 47,XYY; 48,XXYY) [56]. Studies have reported diagnostic yields in individuals with ASD from 2.23% up to 9.7% for G-banded karyotyping but these numbers are likely overestimations due to ascertainment biases [57–61].

- Fragile X testing allows to detect CGG repeat expansions in *FMR1*, which causes fragile X syndrome [62]. Approximately 50% of males and 20% females with Fragile X syndrome have ASD [62]. Studies have reported diagnostic yields in individuals with ASD from 0.04% up to 2.2% for Fragile X testing but these numbers are likely overestimations due to ascertainment biases [57–61]. Fragile X testing is recommended in individuals with ASD if the clinical features and the family history are suggestive for Fragile X syndrome [63].

- Chromosomal microarrays (CMA) allow for the detection of CNVs across the entire genome. The overall resolution of the array strongly depends on the design of the array but usually varies between tens to hundreds of kilobases (kb) [55,64]. In addition, the resolution can be higher in certain regions of the genome, allowing for the detection of smaller CNVs in those regions [55]. With the widespread application of CMA, rare recurring CNVs in specific loci have been identified at greater frequency in individuals with ASD as compared to controls. Typical examples are 1q21.1 deletions and duplications, 2p16.3 deletions, 7q11.23 duplications, 15q11.2–13.1 duplications, 16p11.2 deletions and duplications, and 22q11.2 deletions and duplications [30,31,64]. The diagnostic yield of CMA in individuals with ASD across different studies ranges from 1.5% to 20.5% (median 8.1%) [65]. This strong variation in yield is likely the result of different ascertainment methods and array designs across studies. CMA platforms were recommended as first-tier genetic tests for ASD in the past decade [55].

- With the advent of next generation sequencing (NGS), it became possible to detect multiple types of genetic variation (SNVs, indels and SVs) across the entire genome in an unprecedentedly rapid manner. As a result, genetic variants in many individual genes were identified as Mendelian causes of ASD. Examples include SNVs and indels in *ADNP* [66], *CHD8* [67], *KMT2E* [68], *MYT1L* [69] and *POGZ* [70]. Consequently, NGS was quickly introduced in molecular diagnostics for ASD. If the clinical presentation is suggestive for a specific Mendelian cause, single-gene sequencing (e.g., *MECP2*, *PTEN*) can be performed. However, with the large number of genes associated with ASD, other approaches are currently

preferred over single gene testing. Targeted sequencing of a panel of genes allows to detect genetic variants in exons of genes of interest. However, there is no consensus about which genes should be included in an ASD panel. A recent comparison of 18 commercial gene panels for ASD revealed that the number of genes in these panels ranged from 27 to 2641 and the diagnostic yields ranged from 0.22% to 10.02% [71]. Moreover, new genes are regularly linked to ASD, making the panels rapidly outdated and requiring a new panel design. Another NGS approach that circumvents this problem and that is increasingly being used for genetic testing in individuals with ASD is whole exome sequencing (WES). WES allows to detect genetic variants in the entire exome (i.e., all exons in the genome), which comprises approximately 2% of the genome. Although CNV analysis is not typically done for WES, genetic laboratories are increasingly doing CNV calling from WES data. Studies have reported diagnostic yields for WES in individuals with ASD between 8% to 26% (median 15%) [65]. Since the application of WES results in a higher diagnostic yield as compared to CMA, WES is now recommended as a first-tier genetic test for individuals with ASD [72,73]. The most superior NGS approach is whole genome sequencing (WGS) since it detects genetic variants across the entire genome and improves the calling of SVs [74]. For example, WGS can pick up small CNVs affecting ASD risk genes that are typically missed by CMA or WES [75,76]. In addition, WGS allows to detect ASD-associated risk variants in noncoding regions. Ruzzo et al. identified recurrent deletions in the promoters of *DLG2* and *NR3C2* as risk factors for ASD using WGS [77]. Studies have also reported significant enrichment of rare *de novo* variants in noncoding regions in individuals with ASD using WGS [75,78–80], but the clinical significance of these variants remains poorly understood. WGS is currently not implemented in most genetic laboratories due to high costs and difficult interpretation of noncoding variants [74]. Therefore, limited information is currently available about the diagnostic yield of WGS in large cohorts of individuals with ASD. One study reported a diagnostic yield of 11.2% for WGS in 2620 individuals with ASD [51].

1.3.2 Clinical indications for a genetic workup for ASD

Although Mendelian causes are only found in a minority of individuals with ASD, current guidelines recommend genetic testing for all individuals with ASD [55,72]. There are however several clinical parameters that increase the diagnostic yield of genetic testing in individuals with ASD (i.e., clinical parameters that are associated with Mendelian causes of ASD). Studies have reported higher diagnostic yields in individuals with ASD who have ID, dysmorphic features and congenital anomalies [60,65,81,82]. Therefore, a clinical geneticist will first thoroughly evaluate the individual with ASD before they order a genetic test. Information is gathered about the developmental history, current behavioral problems, and other medical problems in the

individual with ASD [83]. This can be done by reviewing the medical records or by interacting with the affected individual, the parents, or the caregivers. In addition, a multi-generation pedigree is created in collaboration with the family to gain insight in neurodevelopmental, behavior and/or psychiatric problems in other family members [83]. Information about the family history can help to suspect familial or sporadic ASD. During the evaluation of the individual with ASD, the clinical geneticist will also perform a clinical examination [83]. This includes the measurement of an individual's height, weight and head circumference to identify growth problems, weight problems and/or the presence of microcephaly/macrocephaly [83]. Furthermore, a dysmorphology assessment is performed during the clinical examination to detect minor physical anomalies (MPAs) in the individual with ASD [83]. These anomalies are morphological deviations with little to no health consequences that are rare in the general population (typically <2,5%) and are often called dysmorphic features in the field of clinical genetics [84]. We refer to the Elements of Morphology for the full catalogue of dysmorphic features that is used in clinical genetics [85]. The presence of MPAs is strongly determined by genetic factors as shown by a study where a higher correlation in the type and amounts of MPAs was found within monozygotic twins as compared to dizygotic twins [86]. Furthermore, the presence of multiple MPAs is used to define dysmorphism and is of great value to clinical geneticists because it is often an indication of disturbed development due to an underlying Mendelian cause. Despite that several clinical parameters are suggestive for an underlying Mendelian cause in individuals with ASD, consensus statements about clinical indications for a genetic workup or a referral to a genetics center for ASD are currently missing.

1.3.3 Value of dysmorphology assessment in molecular diagnostics of ASD

Studies have consistently reported higher rates of MPAs in individuals with ASD as compared to controls [87]. In 2000, Miles and Hillman investigated whether an elevated number of MPAs in individuals with ASD can be used to predict an underlying Mendelian cause [88]. They found that individuals with six or more MPAs were 10 times more likely to have a known genetic syndrome [88]. In 2005, Miles et al. defined different subtypes of ASD with different outcomes and recurrence risks based on clinical examination including a dysmorphology assessment [89]. The term "complex autism" was used to describe individuals with ASD who have microcephaly or significant dysmorphism (more than five MPAs), while the term "essential autism" was used for individuals without microcephaly and without dysmorphism (less than three MPAs) [89]. Individuals with ASD without microcephaly who had between three and five MPAs were referred to as having "equivocal autism", which is a grey area between the complex and essential autism subtypes [89]. Individuals in the complex group were found to have lower cognitive levels, worse adaptive behavior, more seizures, more abnormal electroencephalograms, more abnormal brain

magnetic resonance imaging (MRI) findings, more gastrointestinal problems, higher number of medication intake and lower quality of life than individuals in the essential group [89,90]. In contrast, individuals in the essential group were found to have higher sibling recurrence, more relatives with autism and higher male to female ratio as compared to the complex group [89]. Studies have reported higher diagnostic yields in individuals with complex autism as compared to individuals with essential autism [82,91,92], demonstrating the added value of a dysmorphology assessment in molecular diagnostics of ASD.

A dysmorphology assessment is also needed to distinguish between syndromic and non-syndromic ASD. "Syndromic ASD" refers to individuals with a clinically defined pattern of somatic abnormalities (e.g. dysmorphism, congenital anomalies) and neurobehavioral problems including ASD [93]. Most individuals with syndromic ASD have a known Mendelian cause [93]. In contrast, "non-syndromic ASD" refers to individuals where ASD is the primary diagnosis [94]. The etiology for ASD in individuals with non-syndromic ASD usually remains unknown and these individuals are therefore often described as having "idiopathic ASD" [94]. The classification of individuals with ASD into syndromic and non-syndromic groups allows to identify individuals that are more likely to benefit from genetic testing, which again illustrates the importance of a dysmorphology assessment in molecular diagnostics of ASD.

1.3.4 Benefits and drawbacks of genetic testing in ASD

The identification of genetic variants that contribute to the underlying etiology of ASD can have several benefits for the individual with ASD and their families. First, genetic testing allows the clinician to provide information about recurrence risks, which may help the parents and/or individual with ASD in reproductive decision making [65]. The recurrence risk for a family however strongly depends on the genetic variant identified. For example, a pathogenic *de novo* variant usually poses little risk for ASD in the next child, while a pathogenic inherited variant often poses relatively high risk and may explain neurodevelopmental problems in other family members. Second, a genetic finding can provide information about the prognosis, allowing patient tailored support and resources in some instances [65]. In case a Mendelian cause is found, it may be possible to prevent, identify and/or treat medical problems [65]. Third, establishing a genetic cause can have psychosocial benefits for the parents [65]. Genetic testing is often preceded by years of medical evaluations in the child, which cause a lot of uncertainty and anxiety for the parents [65]. Finding a genetic cause may bring an end to this uncertainty and has been shown to reduce guilt in the parents of children with ASD [95]. Furthermore, when a Mendelian cause is found, parents can connect with support groups that share experiences and information about this specific cause [65].

There are also several downsides to genetic testing. First, the genetic complexity of ASD makes genetic counseling very challenging. Many variants contribute to ASD risk in an individual, making the clinical significance of genetic variants difficult to assess. Therefore, genetic counseling for ASD is usually hampered by limited and incomplete genetic information, which often results in overinterpretation of the clinical significance. Second, a molecular diagnosis can only be made in a minority of individuals with ASD (i.e., those with Mendelian causes). However, there is a tendency to expect too much from genetic testing for ASD with the recent advances in high-throughput genotyping methods. The lack of consensus statements about the indications for a genetic workup or a referral to a genetics center means that the value of genetic testing for ASD is often overinterpreted by clinicians and miscommunicated to individuals with ASD and their families. Third, the return of genetic findings may have a negative psychological impact on the individual with ASD and their families such as increase anxiety, feeling of guilt, worry and stress [96,97]. For example, a Mendelian cause of ASD that was inherited from a parent may cause a feeling of guilt in the parent [95]. Fourth, genetic findings may negatively impact family relationships because it may reveal unwanted information about family members [96]. Fifth, the presence of a genetic cause in an individual with ASD may lead to stigmatization and discrimination by others [96,97]. For example, genetic information may be misused by future partners, family members, employers, insurance companies and schools.

The benefits and drawbacks of genetic testing for ASD are typically discussed during pre-test genetic counseling, which enables the patient and parents to make an informed decision. For patients with suspected multifactorial ASD, it allows to temper expectations regarding genetic testing. For patients with suspected Mendelian ASD, it allows to prepare the patient and their families for possible consequences of genetic findings. However, pre-test genetic counseling is time-consuming. To avoid that too much time is spent on counseling of patients with ASD who do not benefit from genetic testing (i.e., patients with multifactorial ASD), guidelines on the indications for a genetic workup or even a referral to a genetics center for ASD are needed.

1.4 The face predicts the brain

1.4.1 Shared development of brain and face

Evidence from both animal models and human disorders suggest that the brain and face develop in close coordination. First, the development of the brain and face is tightly linked due to their common origin [98]. During early development, cranial neural crest cells (CNCC) arise as the neural tube (i.e. the primordium of the central nervous system) forms [99]. CNCC then migrate from the neural tube to the cephalic region of the embryo where they will contribute to the formation of the facial skeleton [100]. The cell fate of CNCC is strongly influenced by their location

of origin along the anterior-posterior axis of the neural tube [98]. Second, physical interactions exist between the brain and face during development since they develop close to each other [98]. The brain is believed to act as a structural platform for the face to grow on [98]. Therefore, aberrant early brain growth rates may influence the positioning of the facial tissues [98]. Third, the developing brain and face express various molecular signals that direct the development of one another [98]. The location, timing, and duration of these molecular signals are crucial for normal brain and face development. Finally, both brain and face are affected in several human disorders [98]. Holoprosencephaly is a structural brain disorder that results from an inadequate division of the developing forebrain into two hemispheres and usually affects both forebrain and facial development. In addition, many NDDs co-occur with distinct facial features such as Down syndrome and Williams syndrome.

1.4.2 Facial shape analysis in individuals with ASD

In 1964, Demyer coined the phrase "the face predicts the brain" to describe the correlation between the severity of brain and facial anomalies in patients with holoprosencephaly [101]. Building on this observation, this phrase was later used to suggest that facial anomalies may indicate aberrant brain development in individuals with ASD [102–104]. Although it is well recognized that ASD is not associated with a distinct facial appearance, several studies have identified facial features that distinguish individuals with ASD from controls (**table 1.2**) [102–113]. However, the findings of these studies are often contradictory and cannot be replicated by others (**table 1.2**) [105]. For example, inconsistent findings have been reported from studies measuring intercanthal distance, interpupillary distance, philtrum height and nasal width (**table 1.2**) [105]. This inconsistency in findings is not surprising given that most studies use different ascertainment methods and suffer from small sample sizes that only capture a small part of the clinical and genetic heterogeneity of ASD (**table 1.2**). Furthermore, the use of different methods to assess facial morphology may also have contributed to the lack of consistent findings across studies (**table 1.2**) [105]. Some studies performed manual anthropometry using the Waldrop Physical Anomaly Scale, which involves making qualitative observations and taking quantitative measurements using a ruler and measuring tape [107,108,110,112,113]. Other studies used MRI or three-dimensional (3D) stereophotogrammetry to retrieve objective facial measurements [102–104,106,109,111].

In clinical genetics, an assessment of facial features by experienced dysmorphologists is currently considered the gold standard. However, these assessments are time-consuming, subject to examiner bias, and require a lot of training and experience [114]. In the past decade, objective facial phenotyping methods using 3D surface images have become available [115]. Several of the

studies summarized in **table 1.2** have employed objective facial phenotyping methods to perform an analysis of facial morphology in individuals with ASD. Aldridge et al. analyzed a cohort of prepubertal boys with essential ASD using 3D images and identified a small subgroup of boys with facial morphology that differed from the other boys with ASD and from the controls [103]. The facial morphology of this subgroup was characterized by decreased height of the facial midline and increased mouth widths [103]. Hammond et al. calculated facial asymmetry in boys with ASD of multiplex families using 3D images [102]. Increased facial asymmetry was found in the ASD group compared to the control group, especially greater right-dominant depth-wise asymmetry in the supraorbital and periorbital regions [102]. Boutrus et al. calculated horizontal-, vertical- and depth-wise asymmetry in individuals with ASD using 3D facial images [104]. Greater depth-wise asymmetry was observed in individuals with ASD as compared to unaffected siblings and controls, notably right-dominant depth-wise asymmetry around the eyes, nostrils, and sides of the upper lip [104]. Tan et al. modelled facial masculinity in prepuberal children with ASD using 3D facial images and found an increase in facial masculinity in both males and females with ASD as compared to controls [106].

The current studies summarized in **table 1.2** have focused primarily on identifying specific facial features that distinguish individuals with ASD from controls. Given that such an exploratory approach has resulted in inconsistent findings, knowledge about the underlying etiology of ASD could help to explain the differences in facial morphology between individuals with ASD and controls [105]. However, none of studies described in **table 1.2** have linked facial features to underlying genetic causes. Evidence from clinical genetics suggests that Mendelian causes for ASD often co-occur with dysmorphic facial features. For example, individuals with pathogenic *ADNP* variants often have a prominent forehead, high anterior hairline, wide and depressed nasal bridge, and short nose with full, upturned nasal tip [66]; individuals with pathogenic *CHD8* variants often have a pronounced supraorbital brow ridges, wide-set eyes with down-slanted palpebral fissures, broad nose with full nasal tip, and pointed chin [67]; and individuals with pathogenic *KMT2E* variants often have a high forehead, deep-set eyes, periorbital fullness, prominent cheeks, and prominent nasolabial folds [68]. These findings illustrate the importance of studying facial features in relation to underlying genetic causes in individuals with ASD.

Table 1.2: Summary of studies that compared facial features of individuals with ASD to controls				
Study	**ASD cohort size (M; F)**	**Age range; mean age**	**Assessment method**	**Main findings**
Walker (1977) [107]	74 ASD (60M; 14F)	4y–25y; unknown	Manual anthropometry	Higher frequency of hypertelorism in ASD group.
Rodier (1997) [108]	61 ASD (50M; 11F)	6y-15y; 11y	Manual anthropometry	Reduction in interpupillary distance in ASD group. No difference in inter canthal distance, nasolabial distance, or frequency of epicanthic folds between ASD group and control group.
Hardan (2006) [109]	40 ASD (unknown)	8y-45y; 19y	MRI	Higher frequency of hypotelorism in ASD group with low IQ. No difference in interpupillary distance or frequency of hypertelorism and hypotelorism between entire ASD group and control group.
Hammond (2008) [102]	72 ASD (72M)	2y-18y; 9y	Stereophotogrammetry	Increased nasal width and facial asymmetry in ASD group.
Tripi (2008) [110]	24 ASD (22M; 2F)	unknown; 7y	Manual anthropometry	No difference in nasal width and philtrum length between ASD group and control group. No difference in frequency of prominent forehead, anteverted nostrils, micrognathia, epicanthic folds, hypertelorism, hypotelorism, ptosis or thin upper lip between ASD group and control group.
Aldridge (2011) [103]	64 ASD (64M)	8y-12y; 10y	Stereophotogrammetry	Increased width of the mouth, orbits, and upper face in ASD group. Higher frequency of flattened nasal bridge and hypotelorism in ASD group. Reduced height of the philtrum and maxillary region in ASD group.
Cheung (2011) [111]	36 ASD (30M; 6F)	7y-16y; 11y	MRI	Higher frequency of hypertelorism in ASD group.
Ozgen (2011) [112]	224 ASD (186M; 38F)	3y-18y; 9y	Manual anthropometry	Increased facial asymmetry and asymmetry of eyes and mouth in ASD group. Increased frequency of hypertelorism, coarse face, prominent forehead, deeply set eyes, prominent premaxilla, macrostomia, microstomia, short philtrum, upper lip cupid bow, full cheeks, periorbital fullness, large nose, prominent nasal bridge, upturned nose, full lower lip, thin upper lip, prominent philtrum, prominent lower jaw, and grooved chin in ASD group.
Manouilenko (2014) [113]	50 ASD (26M; 24F)	20y-47y; 30y	Manual anthropometry	No difference in frequency of hypertelorism, hypotelorism and epicanthic folds between ASD group vs control group.
Tan (2017) [106]	74 ASD (54M; 20F)	4y-12y; 8y	Stereophotogrammetry	Increased facial masculinity in ASD group (larger alar-base width, nose height, upper lip height and outer-canthal distance).
Boutrus (2019) [104]	72 ASD (58M; 14F)	unknown; 8y	Stereophotogrammetry	Increased facial asymmetry in ASD group.
Abbreviations: ASD, autism spectrum disorder; F, female; IQ, intelligence quotient; M, male; MRI, magnetic resonance imaging; y, years.				

CHAPTER 2

OBJECTIVES

2 OBJECTIVES

The very rapid evolution in genomics and high-throughput genotyping methods has greatly improved our understanding of the genetic architecture of ASD. Hundreds of genes and loci have been shown to be associated with ASD through rare variant association studies. However, the clinical significance of rare variants in part of these genes remains unknown since many were linked to ASD based on statistical analyses without further clinical validation. In addition, it is expected that many genes involved in the etiology of ASD remain to be discovered. At the same time, this genomic revolution has resulted in increased access to genetic testing for different medical conditions including ASD. Genome-wide sequencing approaches are increasingly used to make a causal molecular diagnosis in individuals with ASD. However, these approaches are currently only useful for those with Mendelian causes and not for those with multifactorial causes. Like for other NDDs, the clinician needs to decide on the usefulness of a genetic workup for each individual with ASD based on the medical files, family history and the clinical examination. This is not evident as Mendelian causes of ASD range from syndromic conditions associated with ID and facial dysmorphism to conditions associated with borderline intellectual functioning or normal cognition and subtle or no dysmorphic facial features. Guidelines on the indications for a genetic workup or even a referral to a genetics center for ASD are currently missing.

The aim of this work is to contribute to improved molecular diagnostics in ASD. The objectives are:

1) To contribute to knowledge on genotype-phenotype correlations in ASD
2) To bridge the gap between novel insights about the genetics of ASD and molecular diagnostics in clinical practice
3) To study objective facial phenotyping as a tool to better recognize Mendelian causes of ASD

CHAPTER 3

NGS IN MOLECULAR DIAGNOSTICS OF ASD

Explanatory note

In this chapter, we study how advances in the genetics of ASD and in high-throughput genotyping methods contribute to better molecular diagnostics in ASD. In part 3.1, we participated in the international "Autism Spectrum/Intellectual Disability (ASID)" consortium [116], which uses large-scale NGS approaches for the identification of novel NDD/ASD risk genes. The consortium performed targeted sequencing of 270 NDD candidate genes in over ten thousand patients with NDDs including 1894 ASD patients from Leuven. Our participation contributed to the rare variant association studies by the consortium and allowed for the evaluation of the clinical significance of rare variants in candidate genes in individual patients. In part 3.2, we participated in a collaborative large-scale genome sequencing effort known as the "Centers for Common Disease Genomics (CCDG)" initiative. In this initiative, WES/WGS is performed in over twenty-five thousand patients with ASD including family-based WGS for 242 ASD patients from Leuven. At the same time, we also performed in-house family-based WGS in 63 ASD patients from Leuven. Analyses of WGS data allowed us to study the value of WGS in molecular diagnostics of ASD. A major advantage of performing in-house WGS was our scientific autonomy, which allowed us to search for patients with similar phenotypic and genotypic profiles through matchmaking platforms. In part 3.3, we demonstrate how this autonomy led to the identification of *KLHL20* as a novel gene for a rare NDD.

3.1 Large-scale targeted sequencing to identify novel NDD risk genes

- Yoeri Sleyp[1,#], Valentine Delanote[2,#], Nele Cosemans[1], Laura Vandenhove[1], Hanne Teule[2], Hilde Peeters[1,2]

  [1]Department of Human Genetics, KU Leuven, Leuven, Belgium. [2]Center for Human Genetics, University Hospitals Leuven, Leuven, Belgium. [#]These authors contributed equally.
- Status: this work was part of a broader research project with the ASID consortium (co-author on one of the consortium papers).
- Adapted from:
  - Stessman HAF, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M, et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nat Genet. 2017;49(4):515–26.
  - Wang T, Hoekzema K, Vecchio D, Wu H, Sulovari A, Coe BP, et al. Large-scale targeted sequencing identifies risk genes for neurodevelopmental disorders. Nat Commun. 2020;11(1):4932.
- Personal contribution: performed Sanger sequencing experiments for segregation analyses of variants, interpreted the clinical significance of variants, and drafted this section.

3.2 Family-based WGS for the genetic workup of ASD

- Yoeri Sleyp[1], Hilde Peeters[1,2]

  [1]Department of Human Genetics, KU Leuven, Leuven, Belgium. [2]Center for Human Genetics, University Hospitals Leuven, Leuven, Belgium.
- Status: most of this work was done as part of an ongoing research project within the CCDG initiative.
- Personal contribution: prepared and shipped DNA samples, analyzed WGS data, performed Sanger sequencing experiments for segregation analyses of variants, interpreted the clinical significance of variants, and drafted this section.

3.3 *KLHL20* is a novel gene implicated in ID, epilepsy and ASD

- Yoeri Sleyp[1], Irene Valenzuela[2], Andrea Accogli[3,4], Katleen Ballon[5], Bruria Ben-Zeev[6,7], Samuel F Berkovic[8], Martin Broly[9,10], Patrick Callaerts[1], Raymond C Caylor[11], Perrine Charles[12], Nicolas Chatron[13], Lior Cohen[14,15], Antonietta Coppola[16], Dawn Cordeiro[17], Claudia Cuccurullo[16], Ivon Cuscó[2], Janette diMonda[18], Ramon Duran-Romaña[19,20], Nina Ekhilevitch[21], Paula Fernández-Alvarez[2], Christopher T Gordon[22], Bertrand Isidor[9], Boris Keren[23], Gaetan Lesca[13], Jarymke Maljaars[24], Saadet Mercimek-Andrews[17,25], Michelle M Morrow[26], Alison M Muir[27], University of Washington Center for Mendelian Genomics, Frederic Rousseau[19,20], Vincenzo Salpietro[4,28], Ingrid E Scheffer[8,29,30], Rhonda E Schnur[26], Joost Schymkowitz[19,20],

Erika Souche[31], Jean Steyaert[32], Elliot S Stolerman[11], Jaime Vengoechea[33], Dorothée Ville[34], Camerun Washington[11], Karin Weiss[21,35], Rinat Zaid[21], Lynette G Sadleir[36], Heather C Mefford[27,37,38], Hilde Peeters[1,31,38]

[1]Department of Human Genetics, KU Leuven, Leuven, Belgium. [2]Department of Clinical and Molecular Genetics, University Hospital Vall d´Hebron and Medicine Genetics Group, Vall d'Hebron Research Institute, Barcelona, Spain. [3]Medical Genetics Unit, IRCCS Istituto Giannina Gaslini, Genoa, Italy. [4]Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Genoa, Italy. [5]Centre for Developmental Disabilities, University Hospitals Leuven, Leuven, Belgium. [6]Pediatric Neurology Institute, The Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Ramat Gan, Israel. [7]Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel. [8]Epilepsy Research Centre, Department of Medicine, University of Melbourne, Austin Health, Heidelberg, VIC, Australia. [9]Centre Hospitalier Universitaire (CHU) de Nantes, Service de Génétique Médicale, Nantes, France. [10]Medigenome, Swiss Institute of Genomic Medicine, Geneva, Switzerland. [11]Greenwood Genetic Center, Greenwood, SC, USA. [12]Salpêtrière Hospital Genetic Department and Reference Center for Rare Intellectual Disabilities, APHP, Paris, France. [13]Department of Medical Genetics, Hospices Civils de Lyon and NeuroMyogene Institute, CNRS UMR 5310 - INSERM U1217, Université Claude Bernard Lyon 1, Lyon, France. [14]Genetic Institute, Barzilai University Medical Center, Ashkelon, Israel. [15]Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer Sheba, Israel. [16]Department of Neuroscience and Reproductive and Odontostomatological Sciences, Federico II University, Naples, Italy. [17]Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada. [18]Department of Human Genetics, Emory Healthcare, Atlanta, GA, USA. [19]Switch Laboratory, VIB-KU Leuven Center for Brain & Disease Research, Leuven, Belgium. [20]Department of Cellular and Molecular Medicine, KU Leuven, Leuven, Belgium. [21]The Genetics Institute, Rambam Health Care Campus, Haifa, Israel. [22]Laboratory of embryology and genetics of human malformations, Institut National de la Santé et de la Recherche Médicale (INSERM), Institut Imagine, Université de Paris, Paris, France. [23]AP-HP.Sorbonne Université, Hôpital Pitié-Salpêtrière, Département de Génétique, Paris, France. [24]Parenting and Special Education Research Unit, KU Leuven, Leuven, Belgium. [25]Department of Medical Genetics, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, Canada. [26]GeneDx, Inc., Gaithersburg, MD, USA. [27]Division of Genetic Medicine, Department of Pediatrics, University of Washington, Seattle, WA, USA. [28]Pediatric Neurology and Muscular Diseases Unit, IRCCS Istituto Giannina Gaslini, Genoa, Italy. [29]Department of Paediatrics, University of Melbourne, Royal Children's Hospital, VIC, Australia. [30]Florey and Murdoch Children's Research Institutes, Melbourne, Australia. [31]Center for Human Genetics, University Hospitals Leuven, Leuven, Belgium. [32]Center for Developmental Psychiatry, KU Leuven, Leuven, Belgium. [33]Department of Human Genetics, Emory University, Atlanta, GA, USA. [34]Department of Neuropediatrics, Lyon University Hospital, Lyon, France. [35]The Ruth and Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel. [36]Pediatric Neurology Department, University of Otago, Wellington, New Zealand. [37]Center for Pediatric Neurological Disease Research, St. Jude Children's Research Hospital, Memphis, TN, USA. [38]These authors contributed equally: Heather C. Mefford, Hilde Peeters.

- Status: accepted in Genetics in Medicine, August 2022 (first author).
- Personal contribution: identified the *KLHL20* variant in one patient which started this project, initiated the collaboration between researchers/clinicians of different genetic centers to bring together phenotypic and molecular data, initiated the collaboration with the Switch lab of KU Leuven to perform *in silico* analysis of the variants, coordinated the project, and drafted this section.

3 NGS in molecular diagnostics of ASD

3.1 Large-scale targeted sequencing to identify novel NDD risk genes

3.1.1 Abstract

With the decreasing costs of NGS, large-scale sequencing studies and subsequent rare variant association analyses are increasingly used to find novel NDD risk genes. However, such approaches require large datasets because they rest on finding multiple rare variants in the same gene or locus in several unrelated patients. The ASID consortium consists of multiple groups of researchers and clinicians around the world that aim to accelerate the identification of novel NDD risk genes by sharing DNA samples, technology, genetic findings, and clinical data. As part of a large-scale sequencing project, the ASID consortium performed targeted sequencing of 270 candidate genes in over ten thousand patients with NDDs. We contributed to this project by providing 1894 DNA samples of ASD patients. Large-scale targeted sequencing, in combination with rare variant association analyses, provided evidence for the association of 148 genes to NDDs. Clinical validation of 67 variants in these genes in ASD patients from Leuven revealed 20 clinically significant variants and contributed to knowledge on genotype-phenotype correlations for *NCKAP1*, *SPEN* and *TCF12*. This study demonstrates that large-scale sequencing with subsequent rare variant association analysis is a strong approach to identify novel NDD risk genes, but clinical validation is essential to extrapolate findings to clinical practice.

3.1.2 Introduction

In the past, one of the cornerstones of disease gene identification was phenotypic similarity between individual patients [116]. However, this approach is limited to genetic disorders characterized by typical craniofacial features or rare combinations of congenital anomalies or symptoms. ASD is a complex NDD in which the underlying causes vary greatly between different patients and are almost never identifiable based on phenotypic features. Although ID, dysmorphic features and congenital anomalies are associated with Mendelian causes of ASD [60,65,81,82], they cannot be used to delineate genetically homogeneous groups for the purpose of novel gene identification in ASD because they are usually not specific to any cause and are very common in other developmental disorders. With the advent of high-throughput genotyping methods, researchers adopted a genotype-first approach in which genotyping precedes extensive phenotyping [116]. In such an approach, patients with variants in the same gene or locus are extensively phenotyped to find meaningful genotype-phenotype correlations [116]. However, large datasets are necessary to find several unrelated patients with rare variants in the same gene or locus.

We participated in a large-scale sequencing project of an international consortium termed the ASID network by contributing 1894 DNA samples of ASD patients for targeted sequencing. The consortium sequenced 270 candidate genes in over ten thousand patients with NDDs and subsequently performed mutation burden and *de novo* enrichment analyses to find statistical associations between individual genes and NDDs [117,118]. In addition, clinical follow-up was done by the consortium for a subset of genes to gain knowledge on genotype-phenotype correlations [117,118]. For the Leuven cohort, we performed clinical validation of 67 variants in 36 NDD candidate genes to determine how the findings from large-scale sequencing studies can be translated to the clinic.

3.1.3 Materials and methods

Targeted sequencing by the ASID consortium

DNA sample collection was organized and performed by the international consortium termed the ASID network. The sequencing of candidate genes was part of two consortium studies to which we contributed 1894 DNA samples of ASD patients from Leuven [117,118]. In a first study, the coding and splicing regions of 208 NDD candidate genes in over 11730 patients with NDDs were sequenced using single-molecule molecular inversion probes [118]. Candidate genes were selected from previously published sequencing studies in patients with NDDs. The selection was mainly based on the number of published *de novo* recurrences, overlap with a CNV morbidity map [119], biological function and pathway analysis [47], or absence of *de novo* variants in exomes of unaffected siblings [37,40,118]. In a follow-up study, the coding and splicing regions of 62 novel NDD candidate genes were sequenced in 16294 patients with NDDs using single-molecule molecular inversion probes [117]. These candidate genes were selected from published studies on NDDs based on the number of *de novo* variants, evidence of ASD sex bias [120], association with high-functioning ASD [118], or overlap with a CNV candidate [117]. In addition, 63 candidate genes that were included in the first consortium study [118] were sequenced in 6211 newly recruited patients with NDDs during the follow-up study [117]. These candidate genes were reselected based on the number of *de novo* variants in published studies and the number of ultra-rare variants identified in sequencing data from previous analyses [117,118]. By combining the new targeted sequencing data with already published data of ASID samples, the final dataset of the follow-up study contained sequencing data for 125 NDD candidate genes in over 16000 patients with NDDs [117]. In total, 270 NDD candidate genes were sequenced across both consortium studies. Detailed information on ASID sample collection, single-molecule molecular inversion sequencing and candidate gene selection is available in the consortium papers [117,118].

Mutation burden and *de novo* enrichment analyses

To identify NDD risk genes, mutation burden and *de novo* enrichment analyses were performed in the candidate genes. The analyses focused on variants in the candidate genes that were private (i.e. found in only one family in the study) [118] or ultra-rare (i.e. MAF < 0.01%) [117]. In addition, the analyses were limited to high-impact variants including LGD variants (here defined as nonsense, stop-loss, start-loss, frameshift, or splice-site variants) and missense variants with a combined annotation dependent depletion score above 30 (MIS30) [117,118]. For the mutation burden analysis, the number of LGD and/or MIS30 variants in NDD candidate genes was compared between patients with NDDs and non-psychiatric controls from the Exome Aggregation Consortium database [117,118]. For the *de novo* enrichment analysis, the number of *de novo* LGD, missense and/or MIS30 variants in NDD candidate genes was calculated and compared to the expected number of *de novo* variants estimated by statistical models [117,118]. The *de novo* enrichment analysis was mainly done on (published) trio exome sequencing data since inheritance information was not available for most variants identified through single-molecule molecular inversion probes sequencing [117,118]. Detailed information on the mutation burden and *de novo* enrichment analyses can be found in the consortium papers [117,118].

Leuven samples

In Leuven, 1894 DNA samples of ASD patients had been collected for targeted sequencing by the ASID consortium. A multi-disciplinary diagnosis of autism was established by the Expert Center for Autism (University Hospitals of Leuven, Leuven, Belgium) based on DSM-IV-TR criteria [121]. All ASD patients were examined by a clinical geneticist and none of the patients had a clinically recognizable genetic syndrome (e.g., Fragile X syndrome, Tuberous Sclerosis). Although there were no specific inclusion criteria besides the presence of ASD, the Leuven cohort was enriched with ASD patients without ID. A genetic workup by CMA analysis was previously done in most patients but did not reveal an underlying Mendelian cause. Sequencing of the coding and splicing regions of 270 NDD candidate genes (208 candidate genes by Stessman et al. and 62 genes by Wang et al.) was performed for the Leuven samples [117,118].

Clinical validation of variants in NDD candidate genes in the Leuven cohort

To gain insight in the clinical significance of the findings from the two large-scale sequencing studies of the ASID consortium, a clinical validation was performed for a subset of the genetic variants that were detected in the Leuven cohort. This validation was only done for variants in genes that reached statistical significance in the mutation burden and *de novo* enrichment analyses by the consortium or in genes that were previously associated with Mendelian ASD. For

the clinical validation, we recalled the patient and their family members to the clinic, and we performed segregation analysis of the variants. Patients and their families were invited for a clinical consultation where data from the medical file such as ASD symptoms, cognitive ability, language ability, motor ability, behavioral and psychiatric disorders were reconsidered. The patients also underwent a clinical examination including an evaluation of MPAs. In addition, the history of ASD, psychiatric disorders and developmental problems in other family members was questioned. Segregation analysis of the genetic variants was performed for all available family members using Sanger sequencing. We followed the standard protocol and used custom-made primers for Sanger sequencing. The clinical significance of the variants was ultimately assessed by evaluating 1) segregation data, 2) clinical presentation of the patient and family members in relation to segregation data, 3) clinical presentation of the patient in relation to cases from literature, 4) variant population frequency data, 5) variant pathogenicity scoring by other laboratories, 6) variant location in relation to other pathogenic variants in the gene, 7) functional studies of the variants, and/or 8) mechanism of disease of the gene if known. A variant was deemed clinically significant when the variant was thought to be causal for ASD in the individual patient.

3.1.4 Results

Summary of rare variant association analyses (see also consortium papers [117,118])

Mutation burden and *de novo* enrichment analyses were performed for 270 NDD candidate genes. In 64 genes, a significant burden of LGD and/or MIS30 variants was found in patients with NDDs as compared to controls. In 134 genes, an excess of *de novo* variants was observed for patients with NDDs. Accounting for overlap in significant findings between the mutation burden analysis and the *de novo* enrichment analysis, the consortium provided statistical evidence for the association of 148 genes with NDDs. To show that some of these NDD risk genes were associated with a distinct neurodevelopmental phenotype, clinical validation was done for patients with genetic variants in a small subset of these genes. Clinical data of the patients and segregation analysis of the variants allowed to identify or extend genotype–phenotype correlations for eleven NDD risk genes (*NAA15*, *KMT5B*, *ASH1L*, *CTCF*, *HNRNPU*, *KCNQ3*, *ZBTB18*, *TCF12*, *SPEN*, *LEO1*, *NCKAP1*). Clinical follow-up for *NAA15*, *KMT5B*, *ASH1L*, *SPEN*, *LEO1* and *NCKAP1* revealed novel associations of these genes with neurodevelopmental phenotypes. For example, LGD variants in *NCKAP1* were found to cause a novel NDD that mainly consisted of ASD or autistic features, language delay, motor delay, and ID or learning difficulties. Another example involved LGD variants in *SPEN*, which were found to cause developmental delay or ID, ASD, muscular hypotonia, tall stature, poor motor coordination and ocular abnormalities. Phenotypic assessment of patients with variants in *CTCF*, *HNRNPU*, *KCNQ3*, *ZBTB18* and *TCF12* helped to refine the

phenotypic spectrum of the NDDs that are associated with these genes. For example, *TCF12* variants were found to cause ASD and ID, in addition to craniosynostosis, which was originally associated with this gene. We contributed to the genotype-phenotype analyses for *NCKAP1*, *SPEN* and *TCF12* by providing segregation data and phenotypic data. These results show that large-scale sequencing with subsequent rare variant association analysis is a strong approach to identify novel NDD risk genes, and in combination with clinical validation can identify or expand genotype-phenotype correlations.

Clinical validation of variants in NDD candidate genes in the Leuven cohort

Targeted sequencing of 270 NDD candidate genes in 1894 ASD patients from Leuven revealed 212 variants in 73 candidate genes. The variants comprised 18 nonsense variants, 41 frameshift variants, 21 splice-site variants, 2 non-frameshift deletions and 130 missense variants. Out of the 73 candidate genes, 58 genes showed a significant mutational burden and/or *de novo* excess in the rare variant association analyses of the consortium. The genes that were the most frequently mutated in patients from Leuven were *RELN* and *NRXN1*, containing 18 and 14 variants respectively. We determined the clinical significance for 67 variants in 36 genes that reached statistical significance in the mutation burden analysis and/or *de novo* enrichment analysis (**supplementary table 3.1**). Segregation analysis was performed for 62/67 variants and resulted in 17 *de novo* LGD variants, 1 *de novo* missense variant, 13 inherited LGD variants and 31 inherited missense variants. Although all 67 variants were considered high-impact variants by the consortium and were in NDD candidate genes that reached significance in the consortium analyses, only 20 variants were retained as clinically significant (**supplementary table 3.1**). Variants that were rated as clinically significant were usually *de novo* variants and LGD variants (**figure 3.1, supplementary table 3.1**). In contrast, variants that could not be retained as clinically significant were usually inherited variants and missense variants (**figure 3.1, supplementary table 3.1**).

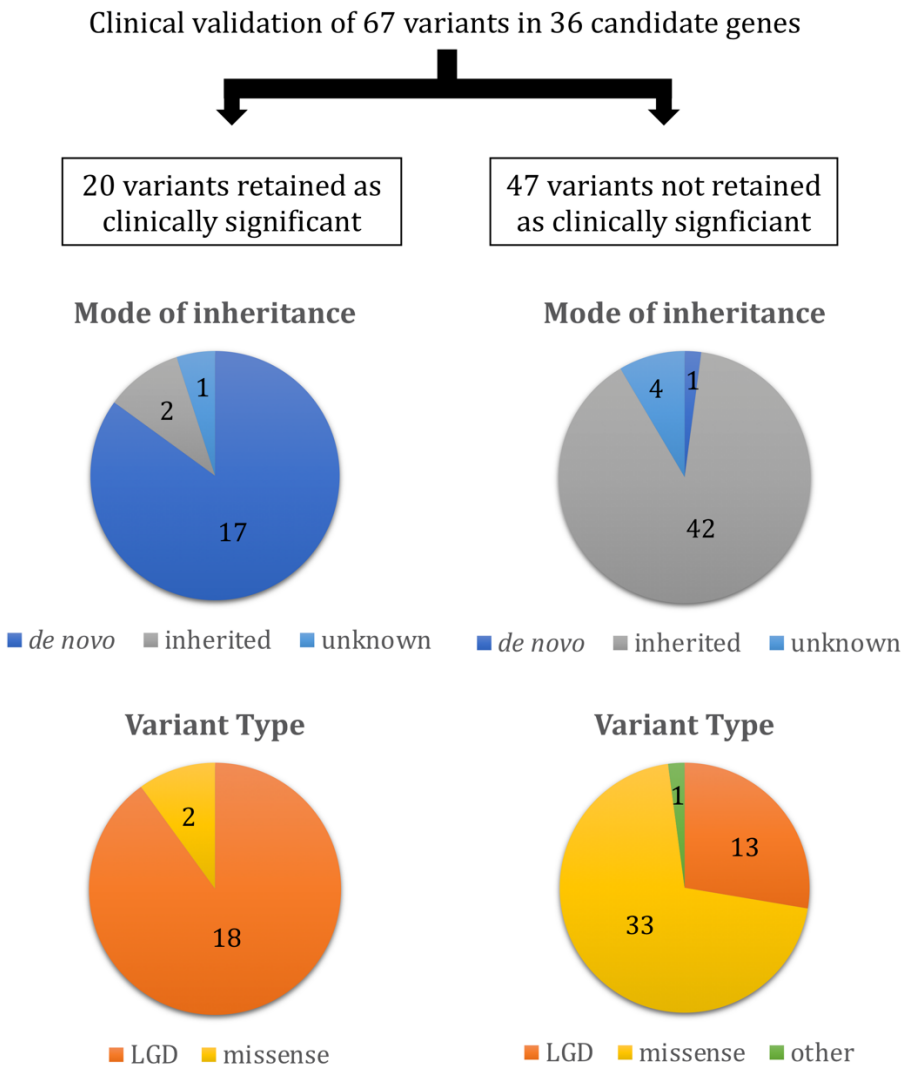


Figure 3.1: Overview of the results from the clinical validation of variants in NDD candidate genes in the Leuven cohort.

Assessment of the cognitive phenotype of ASD patients with clinically significant variants revealed that many presented with ID (12/20; 60%). However, several patients also displayed low to average cognitive abilities including patient P30 with a *de novo* frameshift deletion in *KMT2E*, P32 with a *de novo* frameshift deletion in *KMT5B*, P38 with a maternally inherited frameshift deletion in *NCKAP1*, P40 with a *de novo* splice-site variant in *NFIA*, P44 with a *de novo* nonsense variant in *PHIP*, P58 with *de novo* frameshift deletion in *SETD2*, P62 with a *de novo* frameshift deletion in *SPEN*, and P64 with a *de novo* frameshift deletion in *TCF12* (**supplementary table 3.1**). Furthermore, we identified mild neurodevelopmental problems in parents of two patients with an inherited clinically significant variant. For example, patient P12 had ASD, mild ID and hyperactive behavior and had a paternally inherited splice-site variant in *CHD8* (**supplementary table 3.1**). No formal testing for ASD was ever done in the father, but the

grandparents provided a clear history of unusual and very shy behavior in the father as compared to his siblings. The father attended regular school as a child and has no cognitive impairment. Segregation analysis in the grandparents revealed that the *CHD8* variant occurred *de novo* in the father. These findings allowed us to retain the variant in P12 as clinically significant. The other example involved a maternally inherited frameshift deletion in *NCKAP1* in patient P38 who had ASD and borderline intellectual functioning (**supplementary table 3.1**). Her mother had anxiety from childhood and had several episodes of psychosis in adulthood. Segregation data of maternal grandparents of patient P38 could not be obtained. Nevertheless, these examples illustrate that extensive clinical information of patients and parents, and/or a multigenerational segregation analysis is essential for the correct interpretation of inherited variants. Altogether, these results show that clinically significant variants are usually found in ASD patients with ID but can also be identified in patients or parents with milder neurodevelopmental phenotypes.

We also performed a clinical validation of two variants in NDD candidate genes that were previously associated with Mendelian ASD but that did not reach statistical significance in the mutation burden and *de novo* enrichment analyses. This included a missense variant in *SMAD4* in patient P61 and a nonsense variant in *ARID2* in patient P5 (**supplementary table 3.1**). Clinical assessment revealed that both ASD patients presented with cognitive impairment and facial dysmorphism. In addition, segregation analysis showed that the variants occurred *de novo*. This allowed us to retain both variants as clinically significant. These results show that large-scale sequencing studies are still underpowered to detect all meaningful associations between candidate genes and NDDs.

3.1.5 Discussion

Large-scale targeted sequencing of 270 NDD candidate genes and subsequent rare variant association analyses provided statistical evidence for the association of 148 candidate genes with NDDs [117,118]. Follow-up clinical validation of 67 rare variants in these genes in the Leuven cohort revealed that 20 could be retained as clinically significant. These findings illustrate a discrepancy between promising insights from rare variant association analyses and the clinical value for individual patients. A major cause for this discrepancy is that variants that could not be retained as clinically significant were often not driving the significance signal in the rare variant association analyses (e.g., clinical validation of missense variants in genes that reached statistical significance due to a burden of LGD variants). However, the type of variant that was driving the significance was largely unknown to us at the time of clinical validation and therefore could not be taken into account. Another cause for the discrepancy is that rare variants driving the significance signal in the rare variant association analyses are simply not always clinically

significant. Nevertheless, this validation contributed to knowledge on novel genotype-phenotype correlations for *NCKAP1*, *SPEN* and *TCF12*. This study demonstrates that large-scale sequencing with subsequent rare variant association analysis is a strong approach to identify novel NDD risk genes, but clinical validation is essential to extrapolate findings to clinical practice.

We observed that most ASD patients with clinically significant variants had ID (12/20; 60%). Mendelian causes have indeed been frequently found in ASD patients with ID in the past [43]. However, we observed that some patients with clinically significant variants in NDD risk genes had low-to-normal cognitive abilities. Such mild cognitive phenotypes had previously not extensively been described for some genes including *KMT5B* [122], *SETD2* [123] and *KMT2E* [68]. In recent years, an increasing number of Mendelian causes have been found in ASD patients without ID such as in patients with variants in *CHD8* [67] and *NFIA* [124]. Furthermore, we observed mild neurodevelopmental phenotypes in two parents of ASD patients who carried clinically significant inherited variants, which is in line with Guo et al. who previously found that parents carrying rare LGD variants in ASD/NDD risk genes often display milder autism-related neurodevelopmental phenotypes [125]. These findings demonstrate the variable expressivity associated with many Mendelian causes of ASD, especially in terms of cognitive functioning.

In contrast to *de novo* variants which in general were easy to interpret because of sporadic ASD and associated clinical features, establishing the clinical significance of inherited variants was usually complicated because carrier parents were apparently unaffected, as they did not have cognitive impairment, ASD, or other neuropsychiatric problems. However, we were able to classify two inherited variants as clinically significant by collecting extensive clinical information of patients and parents and/or by performing a multigenerational segregation analysis. This was the case for a paternally inherited variant *CHD8* variant in P12 (NM_001170629: c.2730+1G>A) and a maternally inherited *NCKAP1* variant in P38 (NM_013436: c.2274delG). Such an approach was not possible for most inherited variants, so we cannot exclude that some of the inherited variants that were not retained as clinically significant in the current study may in fact be associated with a Mendelian cause of ASD. Nevertheless, our results demonstrate that extensive phenotyping of patient and parents and/or a multigenerational segregation analysis is a valuable approach to conclude on the significance of inherited variants.

LGD variants were more often retained as clinically significant than missense variants in our study. In general, they were easier to interpret because loss-of-function was often a known mechanism of disease for the candidate genes. In contrast, predicting the impact of a specific missense variant on gene function remains very difficult and often relies on computational tools.

In the absence of an accurate impact prediction, segregation data remains key to assessing the clinical significance of missense variants.

We also confirmed the clinical significance of two variants in genes that were previously associated with Mendelian ASD but that did not reach statistical significance in the rare variant association analyses (*ARID2* and *SMAD4*). This suggests that there were not enough patients with clinically significant rare variants in these genes in the current cohort since the association of these genes with Mendelian ASD indicates that they are ASD/NDD risk genes. Future studies will have to further increase sample sizes to find all meaningful associations between candidate genes and NDDs. International collaborations between genetic centers will be essential to accomplish this since sufficient patients with clinically significant variants for each gene need to be identified.

In conclusion, large-scale targeted sequencing, in combination with rare variant association analyses, is a strong approach to identify novel NDD risk genes. Clinical validation is however essential to determine how these findings can be extrapolated to the clinic. Our study highlights the challenges of molecular diagnostics in ASD as many variants in risk genes could not be retained as clinically significant. An interesting insight in this study was that some patients with clinically significant variants had low-to-normal cognitive abilities, even in genes for which this had not been reported before. We also demonstrated the importance of extensive phenotyping of patients and parents and/or multigenerational segregation analysis to correctly interpret the clinical significance of inherited variants.

3.1.6 Supplementary information

Supplementary table 3.1: Overview of clinical validation of variants in ASD patients from Leuven. Sixty-seven variants are in genes that reached statistical significance in the rare variant association analyses by the consortium. Two variants are in genes that are associated with Mendelian ASD but that did not reach statistical significance in the rare variant association analyses by the consortium (*ARID2* and *SMAD4*).

Patient	Variant	Gene	Variant type	Mode of inheritance	Clinically significant
P1	NM_001282532:c.1402_1403del p.(E468Tfs*1)	*ADNP*	frameshift deletion	*de novo*	+
P2	NM_001282532:c.1287dupT p.(A430Cfs*9)	*ADNP*	frameshift insertion	*de novo*	+
P3	NM_017519:c.G3266A p.(R1089Q)	*ARID1B*	missense	maternal	-
P4	NM_017519:c.G3838A p.(E1280K)	*ARID1B*	missense	maternal	-
P5	NM_001347839:c.C2377T p.(Q793X)	*ARID2*	nonsense	*de novo*	+
P6	NM_018489:c.5446dupA p.(I1816Nfs*12)	*ASH1L*	frameshift insertion	*de novo*	+

P7	NM_018489:c.C8459T p.(S2820L)	*ASH1L*	missense	maternal	-
P8	NM_001042572:c.28_30del p.(E11del)	*CHD2*	non-frameshift deletion	paternal	-
P9	NM_001271:c.G5033A p.(R1678Q)	*CHD2*	missense	paternal	-
P10	NM_001271:c.G5033A p.(R1678Q)	*CHD2*	missense	paternal	-
P11	NM_001170629:c.G2099T p.(R700L)	*CHD8*	missense	paternal	-
P12	NM_001170629:c.2730+1G>A	*CHD8*	splice-site	paternal	+
P13	NM_001170629:c.G5129T p.(G1710V)	*CHD8*	missense	maternal	-
P14	NM_177560:c.C508T p.(R170X)	*CSNK2A1*	nonsense	*de novo*	+
P15	NM_006565:c.G848A p.(R283H)	*CTCF*	missense	*de novo*	+
P16	NM_001103148:c.C733T p.(R245X)	*GIGYF2*	nonsense	maternal	-
P17	NM_001103148:c.C887T p.(S296L)	*GIGYF2*	missense	unknown	-
P18	NM_000834:c.G1570A p.(D524N)	*GRIN2B*	missense	unknown	+
P19	NM_000834:c.2862_2863del p.(C954fs*0)	*GRIN2B*	nonsense	*de novo*	+
P20	NM_000834:c.1780+1G>A	*GRIN2B*	splice-site	*de novo*	+
P21	NM_004501:c.334dupG p.(A112Gfs*32)	*HNRNPU*	frameshift insertion	*de novo*	+
P22	NM_178229:c.C3640T p.(Q1214X)	*IQGAP3*	nonsense	maternal	-
P23	NM_178229:c.G3496A p.(E1166K)	*IQGAP3*	missense	maternal	-
P24	NM_178229:c.3831delC p.(M1278Wfs*59)	*IQGAP3*	frameshift deletion	paternal	-
P25	NM_178229:c.C1474T p.(R492X)	*IQGAP3*	nonsense	maternal	-
P26	NM_178229:c.2361delT p.(F787Lfs*7)	*IQGAP3*	frameshift deletion	paternal	-
P27	NM_031303:c.G589A p.(A197T)	*KATNAL2*	missense	maternal	-
P28	NM_001204824:c.G347A p.(R116H)	*KCNQ3*	missense	paternal	-
P29	NM_001197104:c.2312dupC p.(S774Vfs*11)	*KMT2A*	frameshift insertion	*de novo*	+
P30	NM_018682:c.3524_3527del p.(T1176Rfs*15)	*KMT2E*	frameshift deletion	*de novo*	+
P31	NM_018682:c.G3376A p.(E1126K)	*KMT2E*	missense	maternal	-
P32	NM_001300909:c.87delA p.(Q29Hfs*11)	*KMT5B*	frameshift deletion	*de novo*	+
P33	NM_001300908:c.G1850A p.(R617H)	*KMT5B*	missense	unknown	-
P34	NM_006059:c.G2972A p.(G991D)	*LAMC3*	missense	paternal	-

P35	NM_006059:c.C2242T p.(Q748X)	*LAMC3*	nonsense	maternal	-
P36	NM_015335:c.A6610G p.(N2204D)	*MED13L*	missense	maternal	-
P37	NM_015335:c.2570-1G>A	*MED13L*	splice-site	paternal	-
P38	NM_013436:c.2274delG p.(I759Lfs*17)	*NCKAP1*	frameshift deletion	maternal	+
P39	NM_013436:c.G1603A p.(E535K)	*NCKAP1*	missense	maternal	-
P40	NM_001145512:c.1081+1G>C	*NFIA*	splice-site	*de novo*	+
P41	NM_001042424:c.G2125A p.(E709K)	*NSD2*	missense	maternal	-
P42	NM_001252119:c.C2608T p.(Q870X)	*PASK*	nonsense	maternal	-
P43	NM_001033561:c.2236dupG p.(D746Gfs*76)	*PHF12*	frameshift insertion	*de novo*	-
P44	NM_017934:c.A4276T p.(K1426X)	*PHIP*	nonsense	*de novo*	+
P45	NM_017934:c.C5227T p.(R1743X)	*PHIP*	nonsense	unknown	-
P46	NM_145796:c.G1960A p.(G654S)	*POGZ*	missense	maternal	-
P47	NM_005045:c.G2689A p.(D897N)	*RELN*	missense	unknown	-
P48	NM_005045:c.G3862A p.(A1288T)	*RELN*	missense	maternal	-
P49	NM_005045:c.G6632A p.(R2211H)	*RELN*	missense	maternal	-
P50	NM_005045:c.C1113A p.(N371K)	*RELN*	missense	paternal	-
P51	NM_005045:c.C194G p.(P65R)	*RELN*	missense	paternal	-
P52	NM_005045:c.C1913T p.(P638L)	*RELN*	missense	paternal	-
P53	NM_005045:c.G1328T p.(G443V)	*RELN*	missense	maternal	-
P54	NM_001040143:c.C2566T p.(R856X)	*SCN2A*	nonsense	*de novo*	+
P55	NM_001040143:c.G100A p.(A34T)	*SCN2A*	missense	paternal	-
P56	NM_001130110:c.C645A p.(C215X)	*SETBP1*	nonsense	maternal	-
P57	NM_001349370:c.G5419T p.(A1807S)	*SETD2*	missense	maternal	-
P58	NM_001349370:c.1407delA p.(K469Nfs*1)	*SETD2*	frameshift deletion	*de novo*	+
P59	NM_001080517:c.G2645A p.(R882Q)	*SETD5*	missense	paternal	-
P60	NM_001145357:c.G2896T p.(V966L)	*SIN3A*	missense	paternal	-
P61	NM_001347839:c.C2377T p.(Q793X)	*SMAD4*	missense	*de novo*	+
P62	NM_015001:c.6959delA p.(V2321Wfs*31)	*SPEN*	frameshift deletion	*de novo*	+

P63	NM_001322151:c.1189-2A>G	*TCF12*	splice-site	maternal	-
P64	NM_001306219:c.182delC p.(H62Tfs*12)	*TCF12*	frameshift deletion	*de novo*	+
P65	NM_015088:c.C4310T p.(A1437V)	*TNRC6B*	missense	paternal	-
P66	NM_015088:c.C4468T p.(R1490C)	*TNRC6B*	missense	paternal	-
P67	NM_001284216:c.T3074C p.(V1025A)	*TRIP12*	missense	paternal	-
P68	NM_001284214:c.C62T p.(A21V)	*TRIP12*	missense	maternal	-
P69	NM_001199298:c.C1996T p.(Q666X)	*UIMC1*	nonsense	paternal	-

3.2 Family-based WGS for the genetic workup of ASD

3.2.1 Abstract

In the past decade, the genetic workup of ASD patients mainly consisted of CMA for the detection of CNVs, sometimes followed by sequencing of a panel of ASD/NDD risk genes for the detection of SNVs and indels. However, WGS is expected to eventually become the preferred test in the genetic workup of ASD because it can detect SNVs, indels, CNVs and other types of genetic variation across the genome. Here, we perform family-based WGS in 305 ASD patients in whom there was no Mendelian cause found after genetic testing with CMA (n=305) and/or gene panel sequencing (n=222). We retained 34 variants in known NDD risk genes as clinically significant, of which 2 CNVs and 32 SNVs or indels. These variants were previously missed because no gene panel sequencing was performed (n=15), because they were in genes that were only recently associated with NDDs and thus not detected through gene panel sequencing (n=12), because they were not retained as clinically significant before (n=3), because of poor coverage of CMA and gene panels (n=2) or because they were in genes not included in the gene panels (n=2). Furthermore, we found three clinically significant variants outside known NDD risk genes, one of which allowed for the delineation of a novel NDD caused by *de novo* missense variants in *KLHL20*. We conclude that, once a genetic workup is justified for an ASD patient, family-based WGS is a valuable first-tier test. However, it will only reach its full potential once we better understand the clinical significance of noncoding variants and SVs.

3.2.2 Introduction

With the advent of high-throughput genotyping methods, there has been considerable progress in understanding the genetic architecture of ASD. Rare genetic variants in hundreds of ASD risk genes and loci have been identified as important risk factors for ASD [22]. As a result, the demand for genetic testing for ASD has significantly increased. Patients and their families can benefit from genetic testing since it can help with family planning, inform on the prognosis, improve medical management and promote psychosocial well-being of the parents [65]. However, genetic testing is only useful for ASD patients with Mendelian causes. Such Mendelian causes have predominantly been found in ASD patients with ID, dysmorphic features and/or congenital anomalies [60,65,81,82].

CMA analysis for the detection of causal CNVs was recommended as a first-tier genetic test for ASD in the past decade [55]. Additional targeted sequencing of a panel of genes adds the possibility of detecting SNVs and indels in exons of genes of interest. However, there is currently no consensus on which genes should be included in these gene panels and panels are quickly outdated with the continued progress in Mendelian gene discovery [71]. As a result, WES is now

recommended as a first-tier genetic test for ASD since it can identify SNVs, indels and CNVs affecting exonic regions of any gene [72,73]. WGS is expected to eventually become the preferred test in the genetic workup of ASD because it can detect SNVs, indels, CNVs and other types of genetic variation across the genome. However, WGS is currently not implemented in most genetic laboratories due to high costs over WES and the limited clinical value of genetic variants in noncoding regions [74].

We explore the Mendelian causes that can be found by means of family-based WGS in 305 ASD patients in whom there was no Mendelian cause found after testing with CMA (n=305) and/or gene panels (n=222). The objectives of this study were 1) to evaluate the value of WGS in molecular diagnostics of ASD, 2) to investigate the cognitive phenotype and family characteristics of patients in whom a Mendelian cause can be found, and 3) to determine whether Mendelian causes can be found outside known NDD risk genes.

3.2.3 Materials and methods

Patient recruitment

The total cohort comprised 305 ASD patients from 239 families: 170 ASD patients from 118 families were recruited as part of a prospective study from the Leuven Autism Research consortium in 2008-2009, and 135 ASD patients from 121 families were recruited through the genetics outpatient department where they were seen for a clinical genetic workup and genetic counseling for ASD. All patients had been examined by a clinical geneticist and none of the patients had a clinically recognizable genetic syndrome (e.g., Fragile X syndrome, Tuberous Sclerosis). ASD patients were only included if a multi-disciplinary diagnosis of autism was established based on DSM-IV-TR criteria or DSM-V criteria [7,121]. This multi-disciplinary diagnosis was conducted in the Center for Developmental Disorders or in the Expert Center for Autism (University Hospitals of Leuven, Leuven, Belgium). CMA analysis was previously performed for all patients using the Illumina Omni2.5-v8 single nucleotide polymorphism (SNP) array (n=175 patients) [126] or using the Oxford Gene Technology oligonucleotide array (n=130 patients). In addition, targeted sequencing was previously done for 222 patients using a panel containing 270 NDD candidate genes [117,118] (n=218 patients) or using a clinical exome panel containing more than 6000 disease-associated genes (n=4 patients). Patients for whom a highly penetrant rare risk variant (i.e., Mendelian cause) had previously been found using CMA or gene panel testing were not included in the study. Patients for whom a lowly penetrant rare risk CNV had previously been found using CMA testing were included in the study (n=35 patients) (**supplementary table 3.2**).

Genome sequencing

Family-based WGS was performed for all 239 families. Sequencing with an average depth of 30X was done on an Illumina platform using an Illumina PCR-free library preparation kit according to the manufacturer's instructions. Sequencing was conducted at the New York Genome Center in the United States as part of the CCDG initiative (n=179 families) or at the Genomics Core Leuven in Belgium (n=60 families). Alignment and variant calling were performed according to the pipelines of the centers where the sequencing data were generated. In short, raw reads were aligned to the GRCh38 reference genome using Burrows–Wheeler Aligner-MEM [127], SNVs and indels were detected using the Genome Analysis Toolkit HaplotypeCaller [128] and CNVs were called using Canvas [129] or SeqCBS [130].

Variant prioritization of SNVs and indels

Variant prioritization of SNVs and indels was done using a custom-developed pipeline consisting of variant annotation, filtering, and interpretation. First, the variants were annotated with ANNOVAR [131]. Second, variants were filtered based on: (1) quality metrics, (2) mode of inheritance, (3) variant population frequency and (4) variant genomic location. For variant quality filtering, heterozygous variants with a b-allele frequency above 0.85 or below 0.15 and multi-allelic variants were removed in the variant files of the patients. The following inheritance patterns were analyzed: *de novo* dominant, inherited dominant and X-linked recessive (in males). Variant population frequencies were obtained from the Genome Aggregation Database (gnomAD) [132] and only ultra-rare variants (MAF < 0.01%) were kept. In addition, only variants in exonic regions or splice-sites were retained for further analysis. Finally, variants were interpreted according to international guidelines [133] including assessment of population data, computational data, functional data, segregation data, phenotypic data and/or pathogenicity classifications by others.

Variant prioritization of CNVs

Variant prioritization of CNVs was also done using a custom-developed pipeline consisting of variant annotation, filtering, and interpretation. First, CNVs were annotated with AnnotSV [134]. Second, CNVs were filtered based on: (1) overlap of the CNV with uniquely mappable regions, (2) mode of inheritance, (3) variant population frequency and (4) variant genomic location. For variant quality filtering, only CNVs that overlap more than 50% with uniquely mappable regions were retained for further analysis [135]. The following inheritance patterns were analyzed: *de novo* dominant, inherited dominant and X-linked recessive (in males). Variant population frequencies were obtained from four different databases (Database of Genomic Variants, 1000 genomes, gnomAD, Ira Hall's Genome Database) [132,136–138] if more than 70% length of the

CNV of interest overlapped with a CNV in the database. CNVs with MAF < 2% were retained for further analysis. Moreover, only CNVs affecting exonic regions were kept. Finally, variants were interpreted according to international guidelines [139] including assessment of the CNV size, genomic content of the CNV and overlap with CNVs from external databases.

3.2.4 Results

Molecular genetic findings in known NDD risk genes

In the total cohort of 305 ASD patients, 34 variants in known NDD risk genes were retained as clinically significant after family-based WGS (**supplementary table 3.3**). These variants were rarely located in the same genes or genomic loci (**figure 3.2, panel A**). Thirty-two variants were SNVs or indels and two variants were CNVs (**figure 3.2, panel B**). Clinically significant variants included 10 nonsense variants, 8 frameshift variants, 2 splice-site variants, 12 missense variants, 1 deletion and 1 duplication. Thirty variants occurred *de novo*, three variants were maternally inherited, and one variant was paternally inherited (**figure 3.2, panel C**).

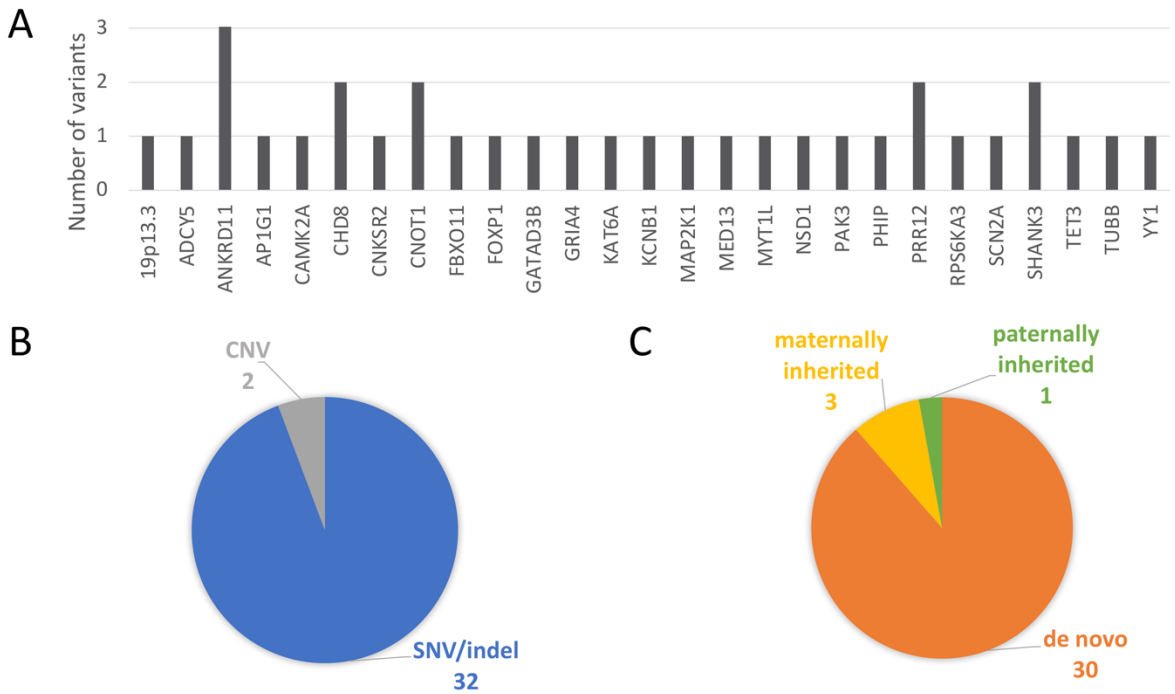


Figure 3.2: Molecular genetic findings in known NDD risk genes following WGS. (A) Number of clinically significant variants detected for each gene or locus. (B) Variant type of clinically significant variants. (C) Mode of inheritance of clinically significant variants.

In the total cohort of 305 ASD patients in whom there was no Mendelian cause found after CMA analysis for the detection of CNVs, only in two patients a clinically significant CNV was detected by means of family-based WGS. This included a 485kb deletion at chromosome 19p13.3 that was missed by CMA in patient P1 and a 40kb intragenic duplication of *NSD1* that was not retained as causal before in patient P22 (**supplementary table 3.3**). In the group of 222 ASD patients in whom there was no Mendelian cause found after gene panel sequencing for the detection of SNVs and indels, 17 clinically significant SNVs or indels were found after family-based WGS. Twelve of these variants were not picked up through gene panel testing because they were in genes that were associated with NDDs in recent years (*AP1G1*, *CAMK2A*, *CNOT1*, *FBXO11*, *GRIA4*, *MED13*, *MYT1L*, *PRR12*, *TET3*, *YY1*). The other five variants were in well-known NDD risk genes (*CHD8*, *CNKSR2*, *KCNB1*, *MAP2K1*, *SHANK3*). Those five variants were missed because 1) the genes were not included in the gene panel (*KCNB1*, *MAP2K1*), 2) the variants were not retained as causal before (*CHD8*, *CNKSR2*), or 3) the variant was missed by gene panel sequencing due to poor coverage (*SHANK3*).

In the cohort of 35 patients in whom a lowly penetrant rare risk CNV was previously found during CMA analyses, we found a clinically significant variant in four patients who carried a lowly penetrant proximal 15q11.2 deletion (BP1-BP2). This included a *de novo* missense in *CAMK2A* in patient P7, a maternally inherited missense variant in *GRIA4* in patient P16, a maternally inherited intragenic duplication in *NSD1* in patient P22, and a *de novo* frameshift deletion in *PRR12* in patient P25 (**supplementary table 3.3**). These results show that the presence of a lowly penetrant rare risk CNV (such as a proximal 15q11.2 deletion) does not exclude the presence of a highly penetrant rare risk variant.

Assessment of the cognitive phenotype and family characteristics of ASD patients with clinically significant variants revealed that many presented with ID (22/33; 67%) and were part of simplex families (31/33; 94%). However, we also detected clinically significant variants in several patients with low-to-average cognitive abilities including patient P3 with a paternally inherited nonsense variant in *ANKRD11*, patient P8 with a *de novo* missense variant in *CHD8*, patient P10 with a *de novo* frameshift deletion in *CNKSR2*, patient P11 with a *de novo* frameshift deletion in *CNOT1*, patient P12 with a *de novo* nonsense variant in *CNOT1*, patient P13 with a *de novo* missense variant in *FBXO11*, patient P16 with a maternally inherited missense variant in *GRIA4*, patient P20 with a *de novo* missense variant in *MED13*, patient P22 with a maternally inherited intragenic duplication in *NSD1*, patient P31 with a *de novo* missense variant in *TET3*, and patient P32 with a *de novo* missense variant in *TUBB* (**supplementary table 3.3**). Moreover, a clinically significant variant was found in two multiplex families. One example included a *de novo* nonsense

variant in *AP1G1* in patient P6 who has mild ID and ASD, and who has a brother with ASD and normal cognition (**supplementary table 3.3**). The other example involved a paternally inherited nonsense variant in *ANKRD11* in patient P3 who has ASD, borderline intellectual functioning and ADHD, and who has a brother with ASD and an average cognitive ability (**supplementary table 3.3**). Other risk factors must contribute to ASD in these families since the clinically significant variants did not segregate with the ASD phenotype. The findings in the family of patient P3 also revealed milder neurodevelopmental problems in the father in comparison to patient P3. In essence, the father of patient P3 was able to function quite well but was the only one with learning difficulties in a family of five children. Segregation analysis for the *ANKRD11* nonsense variant in the paternal grandparents of patient P3 had not been performed. Mild learning difficulties were also seen in the mother of patient P16 with a maternally inherited missense variant in *GRIA4* and in the mother of patient P22 with a maternally inherited intragenic duplication affecting *NSD1* (**supplementary table 3.3**). Altogether, these results show that clinically significant variants are mainly found in ASD patients with ID and a sporadic presentation but can also be identified in patients or parents with milder neurodevelopmental phenotypes and even in multiplex families.

Molecular genetic findings outside known NDD risk genes

We also found three clinically significant variants outside known NDD risk genes. We detected a *de novo* missense variant in *KLHL20* in patient P34 whose clinical presentation is mainly characterized by ID, epilepsy, ASD and ADHD (**supplementary table 3.4**). Through matchmaking platforms, other patients with the same *KLHL20* variant and similar phenotypic features were identified. This allowed us to retain the *KLHL20* variant as clinically significant and to implicate *KLHL20* in a novel NDD. Furthermore, we found a maternally inherited missense variant in *SCN11A* in patient P35 who has ASD and familial episodic pain syndrome (**supplementary table 3.4**). This variant was retained as clinically significant because it segregated with the rare episodic pain disorder in the family. Missense variants in *SCN11A* are known to cause sensory and autonomic neuropathy and episodic pain syndrome [140]. However, there was no evidence for this variant to explain the occurrence of ASD in the proband (patient P35). Moreover, we found one incidental finding that involved a paternally inherited frameshift deletion in *BRCA2* in patients P36-P39 (**supplementary table 3.4**). Pathogenic variants in *BRCA2* increase the risk for several cancers including breast cancer, ovarian cancer and prostate cancer [141]. The *BRCA2* variant was detected because it had already been reported 20 times as pathogenic by other laboratories in the ClinVar database.

3.2.5 Discussion

Here, we report on the molecular genetic findings following family-based WGS in 305 ASD patients in whom there was no Mendelian cause found after testing with CMA (n=305) or gene panels (n=222). WGS analyses revealed 34 clinically significant variants in known NDD risk genes, of which 2 CNVs and 32 SNVs or indels. Most of these variants were previously missed because no gene panel sequencing was performed (n=15). WGS offers the advantage that both SNVs, indels and CNVs can be detected in one single test. Moreover, some variants were not picked up by gene panels because they were in genes that were only recently associated with NDDs (n=12). These variants would have been detected by WGS but could not have been retained as clinically significant at the time gene panel sequencing was done. However, WGS data can be reanalyzed at later timepoints to interpret the clinical significance of variants in recently discovered NDD risk genes. Such a reanalysis has been shown to improve diagnostic yield of WES/WGS by resolving previously unsolved cases [142], and will likely be an important part of future molecular diagnostics. Only in a few cases the variants were missed because they were not retained as clinically significant before (n=3). Furthermore, some variants in well-known NDD risk were not picked up because they were not included in the gene panels (n=2) or because of poor coverage of CMA and gene panels (n=2). WGS allows for a uniform coverage across the genome, so variants in all NDD risk genes can be detected and will less likely be missed because of poor coverage. In addition, WGS can detect clinically significant variants outside known NDD risk genes. This allows for the discovery of novel ASD/NDD risk genes. For example, we found a *de novo* missense variant in *KLHL20* in one patient, which allowed us to search for additional patients with *KLHL20* variants and to find the cause of a novel NDD. Furthermore, WGS also allows for the detection of multiple rare variants affecting different genes or loci in one single patient. For example, we found that four patients in whom a proximal 15q11.2 deletion was previously found through CMA analysis had an additional SNV or indel in a well-known NDD risk gene. Altogether, we conclude that WGS has several advantages in molecular diagnostics of ASD compared to CMA and/or gene panels.

Examination of the cognitive phenotype and family characteristics revealed that most ASD patients with clinically significant variants in NDD risk genes had ID and a sporadic presentation. This is not unexpected because patients with Mendelian causes for ASD have indeed been more frequently found in patients with ID and in simplex families (i.e. families with one affected individual) [37,43,78]. We observed mild cognitive and neurodevelopmental phenotypes in some patients and in carrier parents, which is in line with the variable expressivity of rare variants implicated in Mendelian causes of ASD [125]. In addition, we found a clinically significant variant in two multiplex families (i.e., families with multiple affected individuals). However, the *de novo* nonsense variant in *AP1G1* in patient P6 and the paternally inherited nonsense variant in

ANKRD11 in patient P3 did not segregate with ASD in the family. Yuen et al. also reported that rare variants in NDD risk genes in multiplex families usually do not segregate with ASD [76], which suggests that other factors contribute to ASD risk in these families. However, these variants could potentially explain the more severe neurodevelopmental phenotypes in affected children who are carrier of the clinically significant variant as compared to their affected siblings who are not carrier. Similarly, two studies previously reported that patients with multiple hits in ASD/NDD risk genes show more severe neurodevelopmental phenotypes [125,143].

We limited ourselves to the interpretation of rare variants affecting exonic regions. The main reason being that the interpretation of rare variants outside exonic regions (i.e., in noncoding regions) is nearly impossible without any functional data. As the exome represents less than 2% of the genome, rare genetic variation in most of WGS data in the current cohort of ASD patients thus remain unexplored. Over the past years, RNA sequencing has been increasingly applied in molecular diagnostics as a complementary functional assay to WES and WGS [144]. RNA sequencing is a high-throughput transcriptome-wide assay that enables the quantification of transcript levels and isoforms, and thereby allows for the identification of aberrant gene expression, aberrant splicing of genes, and allelic-specific expression of variants [144]. The use of RNA sequencing has been shown to greatly improve the diagnostic yield of WES/WGS for Mendelian disorders by facilitating the interpretation of variants in both exonic and noncoding regions [144–146]. Combining RNA sequencing data with existing WGS data of ASD patients will allow for improved interpretation of rare variants in both exonic and noncoding regions and will likely reveal additional clinically significant variants in the cohort.

The WGS analyses did also not include the detection and interpretation of SVs besides CNVs. There is currently no SV calling algorithm that can accurately detect the complete range of SVs from short-read WGS data [147]. This is the result of the large variability in subtype and size of SVs, and the proximity of SVs to repetitive regions [147]. In addition, SV calling algorithms apply different approaches, which each have their own strengths and limitations [147]. One strategy for more accurate detection of SVs in short-read sequencing data involves combining multiple SV calling algorithms [147]. Such a strategy was not implemented at the genomic centers where the WGS data were sequenced, but its application on the WGS data of the ASD patients could potentially reveal clinically significant SVs. However, multiplatform strategies (i.e. combining strengths of multiple genomic platforms) will likely be necessary to detect the complete range of SVs [147]. Integrating short-read sequencing data of ASD patients with long-read sequencing data or optical genome mapping may be needed to detect all SVs and to find all clinically significant SVs.

Taken together, family-based WGS revealed several clinically significant variants in ASD patients for whom no Mendelian cause was found after CMA and/or gene panel testing. We found that WGS offers several advantages as compared to CMA and gene panels such as a re-analysis of data, more uniform coverage, the detection of variants outside known NDD risk genes and detection of multiple types of genetic variants in a single patient with one analysis. We conclude that once a genetic workup is justified in a patient with ASD, family-based WGS is a valuable first-tier test. However, it will only reach its full potential once we better understand the clinical significance of noncoding variants and SVs.

3.2.6 Supplementary information

Supplementary table 3.2: Lowly penetrant rare risk CNVs previously found using CMA analyses. 34 patients in the cohort carried one risk CNV and one patient carried two risk CNVs.

Frequency in the cohort	Variant	Mode of inheritance	Affected gene(s)
7	2p16.3 deletion	*de novo* (4); maternal (3)	*NRXN1*
2	7q11.23 duplication	maternal (1); paternal (1)	*CASTOR2;RCC1L;GTF2IRD2B;SPDYE14;SPDYE13;SPDYE15;TRIM73;POM121C;SPDYE5;HIP1;CCL26;CCL24;RHBDD2;POR;TMEM120A;STYXL1;MDH2;SRRM3;HSPB1;YWHAG;SSC4D;ZP3;DTX2;UPK3B*
1	7q36.2 deletion	maternal (1)	*DPP6*
1	9q22.1-2 deletion	*de novo* (1)	*DAPK1;CTSL;SPATA31E1;SPATA31C1;CDK20;SPATA31C2;SPIN1;NXNL2;S1PR3;SHC3;CKS2;SECISBP2;SEMA4D;GADD45G*
2	14q31.1 deletion	maternal (2)	*NRXN3*
16	15q11.2 deletion (BP1-BP2)	paternal (9); maternal (7)	*NIPA1;NIPA2;CYFIP1;TUBGCP5*
1	15q13.2-13.3 deletion	*de novo* (1)	*GOLGA8Q;GOLGA8H;ARHGAP11B;FAN1;MTMR10;TRPM1;KLF13;OTUD7A;CHRNA7*
2	16p11.2 duplication (BP2-BP3)	maternal (2)	*NPIPB7;EIF3C;NPIPB9;ATXN2L;TUFM;SH2B1;ATP2A1;RABEP2;CD19;NFATC2IP;SPNS1;LAT*
1	16p11.2 deletion (BP4-BP5)	*de novo* (1)	*SPN;QPRT;C16orf54;ZG16;KIF22;MAZ;PRRT2;PAGR1;MVP;CDIPT;SEZ6L2;ASPHD1;KCTD13;TMEM219;TAOK2;HIRIP3;INO80E;DOC2A;C16orf92;TLCD3B;ALDOA;PPP4C;TBX6;YPEL3;GDPD3;MAPK3;CORO1A*
2	16p11.2 duplication (BP4-BP5)	*de novo* (1); paternal (1)	*SPN;QPRT;C16orf54;ZG16;KIF22;MAZ;PRRT2;PAGR1;MVP;CDIPT;SEZ6L2;ASPHD1;KCTD13;TMEM219;TAOK2;HIRIP3;INO80E;DOC2A;C16orf92;TLCD3B;ALDOA;PPP4C;TBX6;YPEL3;GDPD3;MAPK3;CORO1A*
1	22q11.2 duplication	*de novo* (1)	*FAM246C;DGCR2;ESS2;TSSK2;GSC2;SLC25A1;CLTCL1;HIRA;MRPL40;C22orf39;UFD1;UFD1;CDC45;CLDN5;SEPTIN5;SEPT5;GP1BB;TBX1;GNB1L;RTL10;TXNRD2;COMT;ARVCF;TANGO2;DGCR8;TRMT2A;RANBP1;ZDHHC8;CCDC188;RTN4R;DGCR6L;ZNF74;SCARF2;KLHL22;MED15;PI4KA;SERPIND1;SNAP29;CRKL;AIFM3;LZTR1;THAP7;P2RX6;SLC7A4;LRRC74B*

Supplementary table 3.3: Overview of molecular genetic findings in known NDD risk genes				
Patient	**Variant**	**Gene**	**Variant type**	**Mode of inheritance**
P1	chr19:1175899-1660700 (hg38)	/	deletion	*de novo*
P2	NM_001199642:c.T1121C p.(I374T)	*ADCY5*	missense	*de novo*
P3	NM_001256183:c.C2716T p.(R906X)	*ANKRD11*	nonsense	paternal
P4	NM_001256183:c.6968_6975del p.(A2323Gfs*206)	*ANKRD11*	frameshift	*de novo*
P5	NM_001256183:c.A1343C p.(Q448P)	*ANKRD11*	missense	*de novo*
P5	NM_001256183:c.A1343C p.(Q448P)	*ANKRD11*	nonsense	*de novo*
P6	NM_001030007:c.C610T p.(R204X)	*AP1G1*	nonsense	*de novo*
P7	NM_001369025:c.C775T p.(R259C)	*CAMK2A*	missense	*de novo*
P8	NM_001170629:c.A3962C p.(E1321A)	*CHD8*	missense	*de novo*
P9	NM_001170629:c.3563dupG p.(Q1189Tfs*5)	*CHD8*	frameshift	*de novo*
P10	NM_001168647:c.255delA p.(T87Pfs*6)	*CNKSR2*	frameshift	*de novo*
P11	NM_001265612:c.3805_3809del p.(D1269Kfs*7)	*CNOT1*	frameshift	*de novo*
P12	NM_001265612:c.C76T p.(R26X)	*CNOT1*	nonsense	*de novo*
P13	NM_001190274:c.G1231A p.(V411I)	*FBXO11*	missense	*de novo*
P14	NM_001244813:c.1021_1024del p.(D341Nfs*32)	*FOXP1*	frameshift	*de novo*
P15	NM_020699:c.C1432T p.(R478X)	*GATAD2B*	nonsense	*de novo*
P16	NM_000829:c.G2090C p.(R697P)	*GRIA4*	missense	maternal
P17	NM_006766:c.3116_3117del p.(S1039*)	*KAT6A*	nonsense	*de novo*
P18	NM_004975:c.G1237A p.(V413I)	*KCNB1*	missense	*de novo*
P19	NM_002755:c.C371T p.(P124L)	*MAP2K1*	missense	*de novo*
P20	NM_005121:c.G3839T p.(R1280L)	*MED13*	missense	*de novo*
P21	NM_001303052:c.G391T p.(E131X)	*MYT1L*	nonsense	*de novo*
P22	chr5:177167878-177208287	*NSD1*	intragenic duplication	maternal
P23	NM_001128166:c.431-2A>G	*PAK3*	splice-site	*de novo*
P24	NM_017934:c.100delG p.(V34Cfs*2)	*PHIP*	frameshift	*de novo*
P25	NM_020719:c.671delC p.(P226Lfs*89)	*PRR12*	frameshift	*de novo*
P26	NM_020719:c.C4926G p.(Y1642X)	*PRR12*	nonsense	*de novo*
P27	NM_004586:c.C986T p.(T329M)	*RPS6KA3*	missense	maternal
P28	NM_001040142:c.3521-2A>G	*SCN2A*	splice-site	*de novo*
P29	NM_033517:c.C1069T p.(R357X)	*SHANK3*	nonsense	*de novo*

P30	NM_033517:c.2649delC p.(S884Rfs*58)	*SHANK3*	frameshift	*de novo*
P31	NM_001366022:c.G2453A p.(R818Q)	*TET3*	missense	*de novo*
P32	NM_178014:c.229C>T p.(R77C)	*TUBB*	missense	*de novo*
P33	NM_003403:c.T1149G p.(Y383X)	*YY1*	nonsense	*de novo*

Supplementary table 3.4: Overview of molecular genetic findings outside known NDD risk genes

Patient(s)	Variant (hg38)	Gene	Variant type	Mode of inheritance
P34	NM_014458:c.G1069A p.(G357R)	*KLHL20*	missense	*de novo*
P35	NM_014139:c.C49T p.(R17C)	*SCN11A*	missense	maternal
P36, P37, P38, P39	NM_000059.4:c.5213_5216del p.(Thr1738IlefsTer2)	*BRCA2*	frameshift	paternal

3.3 *KLHL20* is a novel gene implicated in ID, epilepsy and ASD

3.3.1 Abstract

KLHL20 is part of a CUL3-RING E3 ubiquitin ligase involved in protein ubiquitination. KLHL20 functions as the substrate adaptor that recognizes substrates and mediates the transfer of ubiquitin to the substrates. Although KLHL20 regulates neurite outgrowth and synaptic development in animal models, a role in human neurodevelopment has not yet been described. We report on a NDD caused by *de novo* missense variants in *KLHL20*. Patients were ascertained by investigators through Matchmaker Exchange. We studied 14 patients with *de novo* missense variants in *KLHL20*, delineating a genetic syndrome with patients having mild to severe ID, febrile seizures or epilepsy, ASD, hyperactivity and subtle dysmorphic facial features. We observed a recurrent *de novo* missense variant in 11 patients (NM_014458.4:c.1069G>A p.(Gly357Arg)). The recurrent missense and the three other missense variants all clustered in the Kelch-type β-propeller domain of the KLHL20 protein, which shapes the substrate binding surface. Our findings implicate *KLHL20* in a NDD characterized by ID, febrile seizures or epilepsy, ASD and hyperactivity.

3.3.2 Introduction

Protein ubiquitination is a post-translational modification that is important for proper protein function and degradation. In the ubiquitination cascade, E3 ubiquitin ligases confer substrate specificity by promoting the transfer of ubiquitin to specific substrates. Increasing evidence shows that E3 ubiquitin ligases play a prominent role during neurodevelopment [148,149]. One of the E3 ubiquitin ligases that has been linked to neurodevelopment in animal models is the KLHL20-associated E3 ubiquitin ligase, a multi-subunit CUL3-RING E3 ubiquitin ligase that includes KLHL20, CUL3 and RBX1. In this complex, KLHL20 functions as the substrate adaptor that recognizes substrates and mediates the transfer of ubiquitin to these substrates [150]. By promoting ubiquitin-dependent degradation of the ARHGEF11 substrate, KLHL20 stimulates neurite outgrowth and arborization in hippocampal and cortical rat neurons [151]. Furthermore, the *Drosophila melanogaster* homologue of human KLHL20, known as Dbo, regulates synaptic development and function through negatively controlling Pak [152]. KLHL20 also mediates the ubiquitin-dependent degradation of DAPK1, a well-known regulator of neuronal cell death that has been implicated in epilepsy [153]. DAPK1 overexpression was previously reported in the brain of patients with temporal lobe epilepsy, and pharmacological inhibition of DAPK1 activity in epilepsy mouse models showed strong antiepileptic effects [154–156]. Although KLHL20 has been shown to play a role during neurodevelopment in animal models and is indirectly linked to epilepsy, its role in human neurodevelopment remains unknown. Our study describes a causal

link between genetic variants in *KLHL20* and a neurodevelopmental syndrome that is characterized by ID, febrile seizures or epilepsy, ASD and hyperactivity.

3.3.3 Materials and methods

Patients

We evaluated 11 patients with a recurrent *de novo* missense variant (P1-11) and three patients with unique *de novo* missense variants (P12-14) in *KLHL20* (GenBank: NM_014458.4). Investigators from different centers were connected through Matchmaker Exchange (GeneMatcher, MyGene2) [157].

Genetic and clinical data analysis

Exome sequencing or genome sequencing was performed for 12/14 patients according to the protocols and platforms of each center. P4 was diagnosed by means of targeted Sanger sequencing of the recurrent *KLHL20* missense variant based on clinical assessment and phenotypic resemblance with patient P3. The *KLHL20* variant in P13 was identified by a targeted research candidate gene panel. In general, the identification of *KLHL20* variants in the sequencing data was performed by filtering for: (1) *de novo* variants, (2) variants with an ultra-rare allele frequency in the population (not present in gnomAD) and/or (3) variants located in exonic regions or splice sites. The *KLHL20* variants were validated in patients P1, P2, P3, P4, P5, P6, P7, P9, P10, P13 and P14 by targeted Sanger sequencing. The following intelligence quotient (IQ) cutoffs were used to define ID: profound ID (IQ<20), severe ID (IQ: 20-34), moderate ID (IQ: 35-49), mild ID (IQ: 50-69).

Protein structure analysis of KLHL20 missense variants

The effect of the variants on the binding between the KLHL20 protein and the DAPK1 protein was computed using FoldX (FoldX consortium) [158]. First, the co-crystal structure of the KLHL20 Kelch domain and the DAPK1 peptide (PDB: 6GY5) was refined using the repair function in FoldX. Next, a structural model for all four *KLHL20* variants was generated using the BuildModel function. We then analyzed how each *KLHL20* variant affects the binding of the KLHL20 protein to the DAPK1 substrate by calculating the interaction energy between both molecules using the AnalyseComplex function. The difference in interaction energy between the altered KLHL20-DAPK1 and wild type KLHL20-DAPK1 (called ΔΔG, in kcal/mol) was then calculated. Variants could have a stabilizing effect (ΔΔG < -0.5 kcal/mol), neutral effect (-0.5 kcal/mol ≥ ΔΔG ≥ 0.5 kcal/mol), destabilizing effect (ΔΔG > 0.5 kcal/mol), or highly destabilizing effect (ΔΔG > 2 kcal/mol) on the interaction between the KLHL20 protein and the DAPK1 protein.

3.3.4 Results

The clinical features of 11 patients (P1-P11) with the recurrent *de novo KLHL20* missense variant c.1069G>A p.(Gly357Arg) are summarized in **table 3.1 and table 3.2**. In addition, the clinical features of three patients (P12-14) with unique *de novo KLHL20* missense variants are included: P12 with c.1214G>A p.(Ser405Asn), P13 with c.1262A>G p.(Gln421Arg) and P14 with c.1777G>T p.(Gly593Trp). Detailed information for all genetic examinations performed in the patients is provided in **supplement S1**. Clinical descriptions and relevant genetic findings are reported in **supplement S2**. Clinical images are shown in **figures 3.3 and 3.4**.

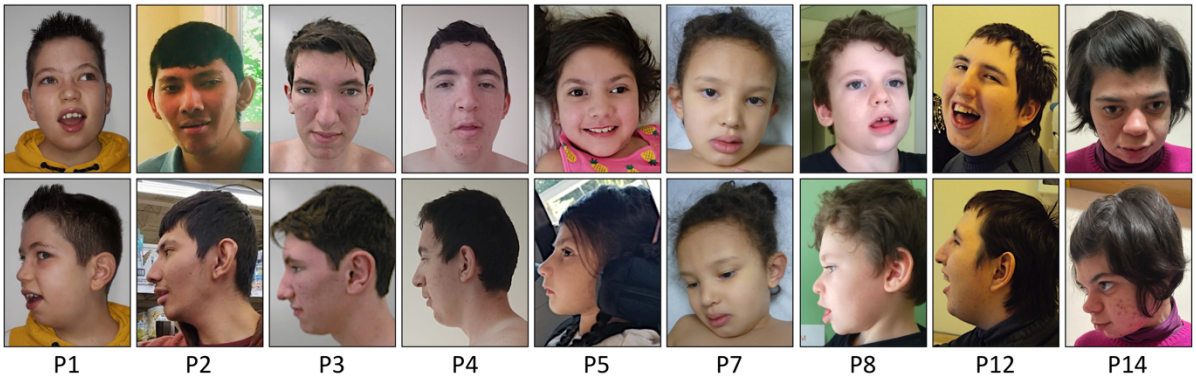


Figure 3.3: Facial photographs of patients. Facial features of patients P1-8 with the recurrent *de novo KLHL20* missense variant NM_014458.4:c.1069G>A p.(Gly357Arg), P12 with a unique *de novo KLHL20* missense variant NM_014458.4:c.1214G>A p.(Ser405Asn), and P14 with a unique *de novo KLHL20* missense variant NM_014458.4:c.1777G>T p.(Gly593Trp). Dysmorphic facial features including hypertelorism, micrognathia, bulbous nasal tip, full lower lip and low-set and large ears are present in some patients. Age of the patients (in years): P1, 12; P2, 22; P3, 25; P4, 18; P5, 8; P7, 6; P8, 8; P12, 22; P14, 23.

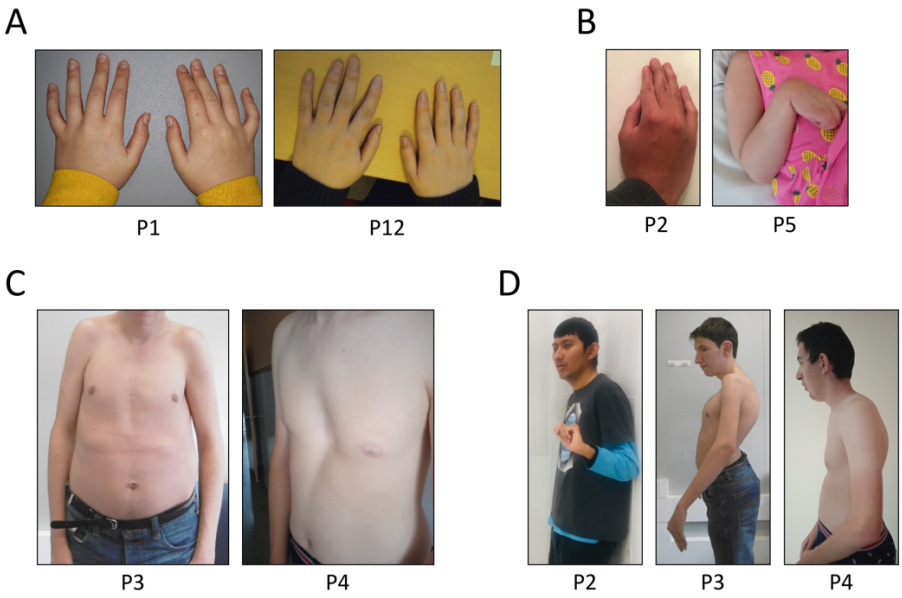


Figure 3.4: Photographs demonstrating skeletal abnormalities in patients. External skeletal features associated with the recurrent *de novo KLHL20* missense variant NM_014458.4:c.1069G>A p.(Gly357Arg) in patients P1-5 and the unique *de novo KLHL20* missense variant NM_014458.4:c.1214G>A p.(Ser405Asn) in patient P12: (A) tapered fingers, (B) dystonic posture, (C) narrow thorax and pectus excavatum, (D) kyphosis and lumbar lordosis.

Table 3.1. General and physical features of patients with *KLHL20* missense variants.								
	P1-P11	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
GENERAL								
Genetic Variant (NM_014458.4)	11/11 recurrent	c.1069G>A	c.1069G>A	c.1069G>A	c.1069G>A	c.1069G>A	c.1069G>A	c.1069G>A
Variant Type	11/11 missense	missense	missense	missense	missense	missense	missense	missense
Inheritance	10/10 *de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*
Sex	8M, 3F	M	M	M	M	F	M	F
Age		12y	22y	25y	18y	8y	3y	6y
Birth Weight	11/11 normal	3.995 kg	3.630 kg	3.555 kg	4.200 kg	3.799 kg	3.095 kg	2.815 kg
Birth Length	9/9 normal	52 cm	53 cm	51 cm	50 cm	NR	51 cm	48 cm
Birth OFC	9/9 normal	34 cm	37.5 cm	38 cm	36 cm	NR	36 cm	34 cm
Gestational age (weeks/days)	10/10 term delivery	38	40	38/5	38	41/5	40	38
BIOMETRY								
Stature	2/11 Tall	Short* (<2SD)	Tall (>2SD)	Normal	Tall (>2SD)	Normal	Normal	Normal
OFC	2/11 MIC	Normal	MIC (<2SD)	Normal	Normal	Normal	Normal	Normal
CRANIOFACIAL PHENOTYPE								
Hypertelorism	5/11	+	-	+	+	+	-	-
Micrognathia	3/11	+	+	-	-	+	-	-
Bulbous nasal tip	3/11	+	-	-	-	-	-	-
Full lower lip	5/11	+	-	+	+	-	-	-
Large ears	4/10	+	+	+	+	NR	-	-
Other Facial Anomalies		Inner Epicanthic Folds	Upslanting Palpebral Fissures	Inner Epicanthic Folds	-	-	Inner Epicanthic Folds	-
Skull Deformity	2/11	-	BC	-	-	-	BC	-
Strabismus	3/11	+	-	+	-	-	-	-
ORTHOPEDIC AND SKELETAL ANOMALIES								
Kyphosis	3/11	-	+	+	+	-	-	-
Scoliosis	1/11	-	-	+	-	-	-	-
Pectus Excavatum	2/10	-	+	-	+	-	-	-
Distal Joint Laxity	3/11	+	-	+	+	-	-	-
OTHER								
Vascular Anomalies	2/11	-	-	+	-	-	-	+
Cardiac Anomalies	1/11	-	-	-	-	-	-	-
GERD	2/11	+	-	-	-	+	-	-
Chronic Constipation	9/11	+	+	+	+	+	-	-
Drooling	5/11	+	+	-	+	-	+	-

BC, brachycephaly; F, female; GERD, gastroesophageal reflux disease; M, male; m, months; MAC, macrocephaly; MIC, microcephaly; NR, not reported; OFC, occipitofrontal circumference; SD, standard deviations; y, years. *Patient P1 carries a *SHOX* deletion, which can explain his short stature.

Table 3.1. Continued

	P8	P9	P10	P11	P12	P13	P14
GENERAL							
Genetic Variant (NM_014458.4)	c.1069G>A	c.1069G>A	c.1069G>A	c.1069G>A	c.1214G>A	c.1262A>G	c.1777G>T
Variant Type	missense	missense	missense	missense	missense	missense	missense
Inheritance	*de novo*	*de novo*	*de novo*	unknown	*de novo*	*de novo*	*de novo*
Sex	M	F	M	M	M	M	F
Age	8y	27y	8y	35y	22y	12y	23y
Birth Weight	3.730 kg	2.630 kg	2.900 kg	3062 kg	3.720 kg	3.950 kg	3.100 kg
Birth Length	55.5 cm	46 cm	NR	48 cm	50 cm	Normal	49 cm
Birth OFC	35 cm	32 cm	34.5 cm	NR	36.5 cm	Normal	35 cm
Gestational age (weeks/days)	37/5	38	NR	42	39	42	At Term
BIOMETRY							
Stature	Normal	Normal	Normal	Normal	Normal	Normal	Normal
OFC	MIC (<2SD)	Normal	Normal	Normal	MIC (<2SD)	Normal	MIC (<2SD)
CRANIOFACIAL PHENOTYPE							
Hypertelorism	-	-	-	+	-	-	+
Micrognathia	-	-	-	-	+	-	+
Bulbous nasal tip	+	-	+	-	-	-	+
Full lower lip	+	-	+	-	-	-	+
Large ears	-	-	-	-	-	-	+
Other Facial Anomalies	Midface Hypoplasia	-	Inner Epicanthic Folds	-	Thin Upper Lip, Narrow Eye Lids	Broad Nasal Tip	Small Forehead, Broad Nasal Tip
Skull Deformity	-	-	-	-	NR	NR	NR
Strabismus	-	+	-	-	-	+	NR
ORTHOPEDIC AND SKELETAL ANOMALIES							
Kyphosis	-	-	-	-	+	-	+
Scoliosis	-	-	-	-	-	-	+
Pectus Excavatum	-	-	NR	-	NR	+	NR
Distal Joint Laxity	-	-	-	-	+	-	-
OTHER							
Vascular Anomalies	-	-	-	-	-	-	NR
Cardiac Anomalies	-	+	-	-	-	-	-
GERD	-	-	-	-	-	+	-
Chronic Constipation	+	+	+	+	-	+	+
Drooling	+	-	-	-	+	+	+

BC, brachycephaly; F, female; GERD, gastroesophageal reflux disease; M, male; m, months; MAC, macrocephaly; MIC, microcephaly; NR, not reported; OFC, occipitofrontal circumference; SD, standard deviations; y, years.

Table 3.2. Neurological and behavioral features of patients with *KLHL20* missense variants.								
	P1-P11	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
Age		12y	22y	25y	18y	8y	3y	6y
NEUROLOGICAL PROBLEMS								
ID	11/11	Severe	Moderate	Mild	Moderate	Severe	DD	Severe
Age of Walking		18m	24m	16m	15m	12m (Lost Motor Skills)	22m	15m
Speech Problems	11/11	Few Single Words	Speech Delay	Speech Delay	Speech Delay	Nonverbal	Speech Delay	Nonverbal
Developmental regression	3/10	-	-	-	-	+	+	-
Seizures (age of onset and offset)	11/11	+ (36m to 5y)	+ (6m to 17y)	+ (6m to 2y)	+ (24m to 10y)	+ (36m - unknown)	+ (6m - ongoing)	+ (unknown to 6y)
Seizure Types (**initial type**) (triggers)		**TC** (fever)	Unknown	**TC** (fever)	**TC** (fever)	**GTC**	**M**, At, GTC, Ab	**TC** (fever)
Epilepsy Syndrome		FS	DEE	FS	FS → epilepsy type unknown	DEE	GE	FS → epilepsy type unknown
EEG		5y: N	4y, 8y: slow, MFD, slow, GSW; 9y: slow, MFD; 17y: slow; 20 y: slow, MFD, GSW	15y: MFD	7y: N	3y: N; 3.8y: slow, FD; 6y, 9y: slow, MFD	2.5y, 2.7y, 3y, 3.8y: GSW, MFD	Unknown
Truncal Hypotonia	5/11	+	-	+	+	+	+	-
Dystonia or Spasticity	4/11	-	Dystonic Hand Posture	-	-	Dystonia/ Status Dystonicus	-	-
Stereotypic Movements	8/10	+	+	+	+	Unknown	+	-
Abnormal brain MRI findings	4/9	-	+	+	-	+	+	-
BEHAVIORAL PROBLEMS								
ASD or autistic features	9/11	ASD	ASD	Autistic features	Autistic features	-	ASD	Autistic features
Hyperactivity	8/10	+	+	+	+	+	+	-
Aggression	7/11	+	+	-	+	-	+	+

Age: m, months; y, years. **Seizure types:** Ab, absence seizure; At, atonic seizure; FIAS, focal impaired awareness seizure; FBTC, focal to bilateral tonic-clonic seizure; GTC, generalized tonic-clonic seizure; M, myoclonic seizure; NCS, nonconvulsive status epilepticus; SE, status epilepticus; T, tonic seizure; TC, tonic-clonic seizure. **Epilepsy syndromes:** DEE, developmental and epileptic encephalopathy; FS, febrile seizures; GE, generalized epilepsy. **EEG:** GSW, generalized spike and wave; FD, focal discharges; MFD, multifocal discharges; N, normal; PSW, polyspike and wave. **Other:** ASD, autism spectrum disorder; DD, developmental delay; ID, intellectual disability; MRI, magnetic resonance imaging; N/A, not applicable.

Table 3.2. Continued							
	P8	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**
Age	8y	27y	8y	35y	22y	12y	23y
NEUROLOGICAL PROBLEMS							
ID	Severe-Profound	Moderate-Severe	Severe	Mild	Moderate	Severe	Severe-Profound
Age of Walking	36m	18m	48m	15m	24m	14m	7y (Lost Motor Skills)
Speech Problems	Nonverbal	Nonverbal	Nonverbal	Speech Delay	Speech Delay	Speech Delay	Nonverbal
Developmental regression	+	-	Unknown	-	-	-	+
Seizures (age of onset and offset)	+ (16m - ongoing)	+ (18m - ongoing)	+ (6m - unknown)	+ (6m - unknown)	None	+ (12m - ongoing)	+ (15days - ongoing)
Seizure Types (**initial type**) (triggers)	**FIAS,** TC, M, SE (fever)	**FIAS**, T, TC, SE	**GTC**	**TC**	N/A	**GTC** (+/- fever), At, T, M, Multifocal, NCS	**FBTC**, GTC, T with perioral myoclonia, SE (fever)
Epilepsy Syndrome	DEE	DEE	Unknown	Unknown	N/A	DEE	DEE
EEG	EEG since 4y: slow, MFD, GSW	2y, 5y: FD, GSW; 13y: slow	8y: FD	Unknown	N/A	2.1y: N; 2.7y, 3y: slow, MFD; 4.5y: slow, GSW, PSW; 6y: slow, MFD, GSW; 11y: slow, GSW, MFD; 13y: FD	Multiple from 6m: slow, MFD
Truncal Hypotonia	-	-	-	-	Unknown	+	Unknown
Dystonia or Spasticity	Unstable Gait at Young Age/ Later Spasticity of Limbs	Dystonic Posture/ Internal Rotation of Hips/ Dystonic Gait	-	-	Unknown	-	Unstable Gait/ Spasticity of Limbs
Stereotypic Movements	Hand Movements	+	+	-	+	-	+
Abnormal brain MRI findings	-	-	Unknown	Unknown	Unknown	-	-
BEHAVIORAL PROBLEMS							
ASD or autistic features	Autistic features	Autistic features	Autistic features	-	Autistic features	Autistic features	Autistic features
Hyperactivity	+	-	Unknown	+	+	-	Unknown
Aggression	+	-	+	-	+	-	Unknown

Age: m, months; y, years. **Seizure types:** Ab, absence seizure; At, atonic seizure; FIAS, focal impaired awareness seizure; FBTC, focal to bilateral tonic-clonic seizure; GTC, generalized tonic-clonic seizure; M, myoclonic seizure; NCS, nonconvulsive status epilepticus; SE, status epilepticus; T, tonic seizure; TC, tonic-clonic seizure. **Epilepsy syndromes:** DEE, developmental and epileptic encephalopathy; FS, febrile seizures; GE, generalized epilepsy. **EEG:** GSW, generalized spike and wave; FD, focal discharges; MFD, multifocal discharges; N, normal; PSW, polyspike and wave. **Other:** ASD, autism spectrum disorder; DD, developmental delay; ID, intellectual disability; MRI, magnetic resonance imaging; N/A, not applicable.

Clinical synopsis of patients with the recurrent *KLHL20* variant (c.1069G>A, p.(Gly357Arg))

All children were born at term without problems in the neonatal period and normal growth parameters. The clinical presentation of all 11 patients with the recurrent variant was characterized by early global developmental delay and seizures (**table 3.2**). There was limited epilepsy phenotype information available for most of these children; however, two had only febrile seizures, two had febrile seizures that progressed to an unclear epilepsy type, one had a generalized epilepsy, four had a developmental and epileptic encephalopathy, and for two, the type of epilepsy was unclear. Seizure onset ranged from 6 to 36 months (mean 16 months). Seizures stopped in 5 of the 8 patients with available data (range 2 to 17 years) and there was drug-resistance in 6 of the 9 patients with available data. Developmental delay was usually present in the first year of life with evolution to ID. The cognitive abilities of the patients ranged from mild to severe-profound ID. All patients had delayed speech and five were nonverbal. Developmental regression with the onset of seizures was reported for patients P5 and P8. For patient P6 a period of regression limited to language skills was reported. Detailed neurocognitive and behavioral data for patient P1 and P2 are available in **supplement S3**. For the other patients, available data on instruments and corresponding scores are presented in the individual patient descriptions in **supplement S2**.

Neurological examination was abnormal in most patients with truncal hypotonia in five patients, dystonia in three patients, and spasticity of the lower limbs in one patient. Patient P5 had early loss of her motor skills with progression to status dystonicus, patient P8 had an unstable gait at a young age that evolved to spasticity of the lower limbs, and patient P9 had dystonic gait with dystonic internal rotation of the hips. Behavioral problems were present in all patients: ASD or autistic features (9/11 patients), hyperactivity (8/10 patients), and aggression (7/11 patients). Drooling (5/11 patients) and chronic constipation (9/11 patients) were also common. Mild craniofacial features were present with hypertelorism (5/11 patients), micrognathia (3/11 patients), bulbous nasal tip (3/11 patients), full lower lip (5/11 patients), large ears (4/10 patients), and brachycephaly (2/11 patients). Other clinical features were strabismus (3/11 patients), distal joint laxity (3/11 patients), pectus excavatum (2/10 patients), scoliosis (1/11 patients), kyphosis (3/11 patients), vascular anomalies (2/11 patients) and mild non-specific changes in brain MRI (4/9 patients). Biometric parameters were normal in most patients except for the presence of tall stature (2/11 patients) and microcephaly (2/11 patients).

The three patients with unique *de novo* KLHL20 variants had similar clinical presentations with moderate to severe ID (3/3 patients), speech delay (3/3 patients with 1/3 nonverbal), autistic

features (3/3 patients), and drug-resistant developmental and epileptic encephalopathy (2/3 patients).

Variant interpretation of the identified *KLHL20* missense variants

All four missense variants (p.(Gly357Arg), p.(Ser405Asn), p.(Gln421Arg), p.(Gly593Trp)) are absent in population variation databases, cluster in the Kelch-type β-propeller domain of the KLHL20 protein, affect well-conserved amino acids and are predicted to be damaging, with the exception of p.(Gln421Arg), by most computational algorithms in Varsome [159]. Causality of the p.(Gly357Arg) variant was evident from the recurrent *de novo* aspect in multiple patients with sporadic mild to severe ID and other neurodevelopmental problems. Similarly, the *de novo* occurrence and the large overlap in clinical presentation with patients with the recurrent *KLHL20* variant suggests causality for the p.(Ser405Asn), p.(Gln421Arg) and p.(Gly593Trp) missense variants.

Protein structure analysis of KLHL20 missense variants

All four missense variants are in the Kelch-type β-propeller domain of the KLHL20 protein, which consists of six Kelch repeats (labeled I to VI in **figure 3.5A and B**) [160]. Each Kelch repeat is composed of four anti-parallel β strands (labeled A to D in **figure 3.5C**) and connecting loops (labeled AB, BC, CD, DA in **figure 3.5C**) [160]. Within these Kelch repeats, the variants are either in the largely buried C-terminal part of DA loops (p.(Gly357Arg), p.(Ser405Asn) and p.(Gly593Trp)) or in the long BC loops (p.(Gln421Arg)), which are known to shape the substrate binding surface of the KLHL20 ubiquitin ligase (**figure 3.5B and C**). Since all variants are in the substrate binding surface, we determined the effect of the variants on the interaction with DAPK1, a substrate that was upregulated in the brain of patients with temporal lobe epilepsy in previous studies [154–156]. Both the p.(Gly357Arg) and p.(Gly593Trp) variant were predicted to be destabilizing (ΔΔG=2.49 kcal/mol and ΔΔG=15.77 kcal/mol respectively) to the interaction between the KLHL20 protein and the DAPK1 protein. The p.(Ser405Asn) and p.(Gln421Arg) variants were found to disrupt electrostatic interactions and hydrogen bonding with the DAPK1 protein respectively but were not predicted to be destabilizing (ΔΔG=0.09 kcal/mol and ΔΔG=0.10 kcal/mol respectively) to the binding of KLHL20 with DAPK1.

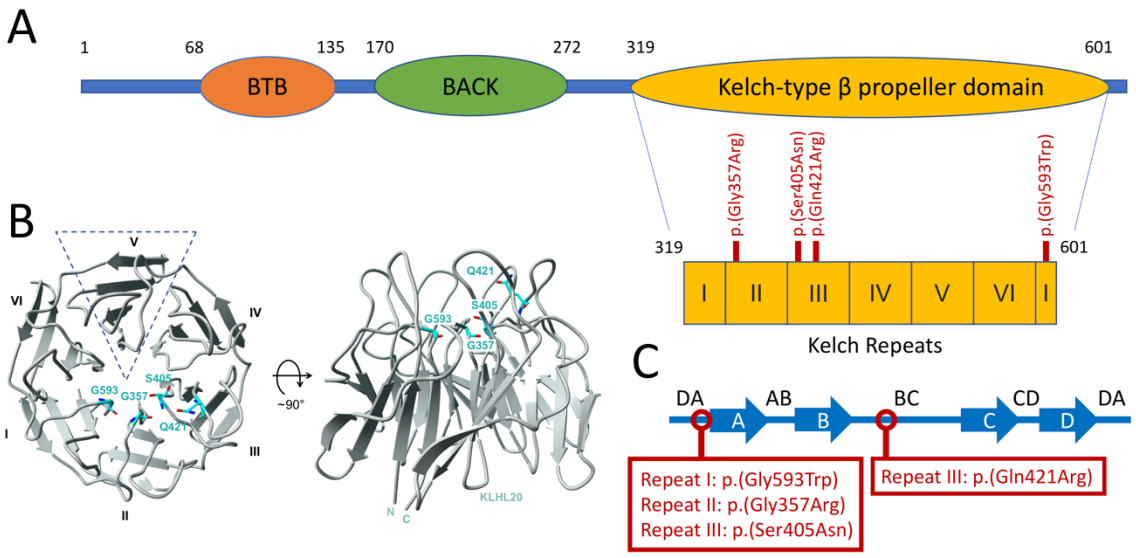


Figure 3.5: KLHL20 protein structure. (A) Representation of linear protein structure of KLHL20 with its protein domains. Three protein domains have been described for the KLHL20 protein: BTB, BACK and Kelch-type β propeller domain. The Kelch-type β propeller domain consists of six Kelch repeats (labeled I to VI) in which the variants are located. (B) Overview of the 3D structure of Kelch-type β propeller domain of the KLHL20 protein (PDB: 6GY5). Residues corresponding to the identified variants are highlighted. The structure of Kelch repeat V is shown by the dashed lines as an example. (C) Representation of the general composition of a Kelch repeat. Each repeat consists of four antiparallel β strands (arrows labeled A-D) and connecting loops (labeled AB, BC, CD and DA). The arrows correspond to the arrows of individual Kelch repeats in panel B. The position of the variants in the Kelch repeats is shown.

3.3.5 Discussion

We report 11 patients with a neurodevelopmental syndrome caused by a recurrent *de novo* missense variant in *KLHL20* (c.1069G>A p.(Gly357Arg)). The core features of this disorder were mild to severe ID (with half of the patients being nonverbal), febrile seizures or epilepsy, ASD and hyperactivity. Additional features in a subset of patients include spasticity or dystonia, strabismus, distal joint laxity, pectus excavatum, scoliosis, kyphosis and vascular anomalies. Recurring craniofacial features included hypertelorism, micrognathia, bulbous nasal tip, full lower lip, low-set and large ears, and brachycephaly. Furthermore, several features that are common in NDDs were frequently observed such as truncal hypotonia, drooling and constipation. Biometry at birth was normal but microcephaly was observed in two patients at later ages. It is unclear yet whether tall stature, which is present in two patients, may be a feature of this syndrome. Mild non-specific changes in brain MRI were found in four patients. Of interest, the clinical observations in patient P3 with scoliosis, kyphosis, distal joint laxity, cerebral artery tortuosity and aortic root dilatation may represent a phenotypic overlap with connective tissue disorders. In addition to the recurrent missense variant, we also report three patients with unique missense variants in *KLHL20* who have similar phenotypic features to the patients with the recurrent variant. The core features including mild to severe ID and autistic features were present

in all three patients. A developmental and epileptic encephalopathy was present in two out of three patients. The large phenotypic similarity with patients with the recurrent *KLHL20* variant, the *de novo* occurrence, the absence in population variation databases, and the clustering in the same protein domain also strongly support causality for these unique *KLHL20* missense variants.

KLHL20 shows no clear intolerance to loss-of-function variants (probability of loss-of-function intolerance=0.01; observed/expected ratio=0.3) but shows some intolerance to missense variants (z-score=3.61; observed/expected ratio=0.47) according to constraint metrics from the gnomAD browser [132]. The spatial clustering of the four identified missense variants in the Kelch-type β propeller domain of the KLHL20 protein suggests that the pathophysiological mechanism is likely a gain-of-function or dominant-negative effect [161].

The Kelch-type β propeller domain of the KLHL20 protein is crucial for the substrate recognition and interaction function of the KLHL20-associated E3 ubiquitin ligase [160]. We found that the variants are in the largely buried C-terminal part of DA loops or in the long BC loops of the Kelch repeats, which shape the substrate binding surface of the KLHL20 ubiquitin ligase [160]. We therefore hypothesize that *KLHL20* variants interfere with the recognition and interaction of substrates by the KLHL20-associated ligase in some specific manner. Consistent with this hypothesis, the amino acids that are altered in our patients were previously identified as interacting residues between the KLHL20 ubiquitin ligase and the DAPK1 substrate [160]. DAPK1 is a well-known regulator of neuronal cell death that has previously been linked to epilepsy [153–156]. FoldX predictions showed that the variants p.(Gly357Arg) and p.(Gly593Trp), but not p.(Ser405Asn) and p.(Gln421Arg), are detrimental to the binding of KLHL20 with DAPK1. Another interesting substrate of the KLHL20-associated ubiquitin ligase is ARHGEF11 due to its role in neural morphogenesis. It was previously found that the KLHL20-associated ubiquitin ligase stimulates neurite outgrowth and arborization by promoting ubiquitin-dependent degradation of ARHGEF11, an activator of the RhoA signaling pathway [151]. Furthermore, KLHL20 itself may be a substrate of the KLHL20-associated ubiquitin ligase. It was previously reported that protein abundance of dbo (*Drosophila melanogaster* homologue of human KLHL20) is regulated through auto-ubiquitination [162]. The *KLHL20* variants could therefore interfere with the auto-ubiquitination of KLHL20 and its subsequent degradation. Follow-up studies are necessary to elucidate the pathophysiological mechanism underlying this syndrome.

In conclusion, we define a NDD that is characterized by mild to severe ID, febrile seizures or epilepsy, ASD and hyperactivity caused by *de novo* missense variants in *KLHL20*. Associated clinical features are spasticity or dystonia, strabismus, distal joint laxity, pectus excavatum,

scoliosis, kyphosis, vascular anomalies, brachycephaly and mild dysmorphic facial features. Identification of additional affected individuals will further define the molecular and phenotypic spectrum of this disorder.

3.3.6 Supplementary information

Supplement S1: Overview genetic testing

P1: CMA analysis revealed a maternally inherited 2 Mb deletion at Xp22.33 ([hg19]: chrX:1491-2182681) including *SHOX*. *FMR1* screening was normal. Trio genome sequencing was then performed on a NovaSeq 6000 sequencing system (Illumina) with an average depth of coverage of 30X. Variants were called using the Genome Analysis Toolkit and analyzed using a custom-developed analysis tool. A rare *de novo* missense variant in *KLHL20* (NM_014458.4:c.1069G>A) was detected and confirmed by Sanger sequencing.

P2: CMA analysis did not reveal any pathogenic variants. *FMR1* screening was normal. Trio exome sequencing was then performed using the xGen Exome Research Panel v1.0 (Integrated DNA Technologies) on an Illumina platform with an average depth of coverage of 125x. Variants were analyzed using a custom-developed analysis tool. A rare *de novo* missense variant in *KLHL20* (NM_014458.4:c.1069G>A) was detected and confirmed by Sanger sequencing. Additionally, a rare maternally inherited missense variant was identified in *ATP1A2* (NM_000702.4:c.1148G>A) that was previously reported as pathogenic.

P3: Initial genetic screening revealed a normal male karyotype and normal *FMR1* gene. CMA analysis did not reveal any pathogenic variants. Trio exome sequencing was then performed with SeqCap EZ MedExome Kit (Roche) on a NextSeq 500 sequencing system (Illumina) with an average depth of coverage of 85X. A rare *de novo* missense variant was found in *KLHL20* (NM_014458.4:c.1069G>A). Sanger sequencing of the *KLHL20* variant in patient and parental samples confirmed that the variant occured *de novo*.

P4: Initial genetic screening revealed a normal male karyotype and normal *FMR1* gene. Targeted Sanger sequencing was then performed for the recurrent *KLHL20* variant (NM_014458.4:c.1069G>A) based on clinical assessment and phenotypic resemblance with patient P3. Sanger sequencing showed that the recurrent *KLHL20* variant was present. Sanger sequencing of the *KLHL20* variant in parental samples showed that the variant occured *de novo*.

P5: CMA analysis did not reveal any pathogenic variants. Trio exome sequencing was then performed using the xGen Exome Research Panel v1.0 Kit (Integrated DNA Technologies) on an

Illumina platform with an average depth of coverage of 100x. Variants were analyzed using a custom-developed analysis tool. A rare *de novo* missense variant in *KLHL20* was detected (NM_014458.4:c.1069G>A) and confirmed by Sanger sequencing. No pathogenic variants were found after mitochondrial genome sequencing.

P6: Initial genetic screening showed a normal *FMR1* gene. CMA analysis did not reveal any pathogenic variants. Trio exome sequencing was done using the xGen Exome Research Panel v2 Kit (Integrated DNA Technologies) on a NovaSeq 6500 sequencing system (Illumina). Variants were analyzed using the Genoox data analysis platform (Genoox). A rare *de novo* missense variant in *KLHL20* was detected (NM_014458.4:c.1069G>A) and confirmed by Sanger sequencing.

P7: CMA analysis did not reveal any pathogenic variants. Trio exome sequencing revealed a rare *de novo* missense variant in *KLHL20* (NM_014458.4:c.1069G>A). This variant was confirmed by Sanger sequencing.

P8: CMA analysis did not reveal any pathogenic variants. *FMR1* screening was normal. Trio exome sequencing was then performed using SeqCap EZ MedExome Kit (Roche) on a NextSeq 500 sequencing system (Illumina) with an average depth of coverage of 71X. Variants were called using Genome Analysis Toolkit and analyzed using SnpEff. A rare *de novo* missense variant in *KLHL20* (NM_014458.4:c.1069G>A) was detected.

P9: Conventional karyotype revealed a normal female result. Trio exome sequencing was performed and revealed a rare *de novo* variant in *KLHL20* (NM_014458.4:c.1069G>A). This variant was confirmed by Sanger sequencing.

P10: Initial genetic screening revealed a normal male karyotype and normal *FMR1* gene. CMA analysis did not reveal any pathogenic variants. Trio exome sequencing was then performed using the xGen capture kit (Integrated DNA Technologies) on a HiSeq 2500 platform (Illumina). Variants were analyzed using the Genoox data analysis platform (Genoox). A rare *de novo* missense variant in *KLHL20* was detected (NM_014458.4:c.1069G>A) and confirmed by Sanger sequencing.

P11: Initial genetic screening revealed a normal male karyotype and normal *FMR1* gene. CMA analysis did not reveal any pathogenic variants. Singleton exome sequencing was performed using the Agilent SureSelectXT Clinical Research Exome kit on a NovaSeqTM 6000 System. A rare

missense variant in *KLHL20* was detected (NM_014458.4:c.1069G>A). Additionally, a rare nonsense variant in *TNRC18* was detected (NM_001080495.2:c.1756C>T).

P12: Initial genetic screening revealed a normal male karyotype and normal *FMR1* gene. CMA analysis did not reveal any pathogenic variants. Trio exome sequencing was then done using SeqCap EZ MedExome Kit (Roche) and performed on a NextSeq 500 sequencing system (Illumina). Variants were called using Genome Analysis Toolkit and analyzed using SnpEff. A rare *de novo* missense variant in *KLHL20* (NM_014458.4:c.1214G>A) was detected.

P13: CMA analysis did not reveal any pathogenic variants. Analysis of singleton exome sequencing did not reveal any plausible variants in known epilepsy genes. Targeted capture and sequencing of a panel of candidate epilepsy genes including *KLHL20* was performed and identified a rare variant in *KLHL20* (NM_014458.4:c.1262A>G). Sanger sequence analysis of parents confirmed that the variant occurred *de novo*.

P14: Trio exome sequencing was performed using the SureSelectXT Clinical Research Exome kit (Agilent) and sequenced on a HiSeq platform (Illumina). A rare *de novo* missense variant in *KLHL20* (NM_014458.4:c.1777G>T) was detected and confirmed by Sanger sequencing.

Supplement S2: Individual patient descriptions

P1 is a 12-year-old boy of European ancestry who presented at the age of three years with psychomotor and language delay, hyperactive behavior, and epilepsy. He has short stature explained by a maternally inherited deletion of *SHOX* ([hg19]: chrX:1491-2182681). He has tapered fingers with distal joint hyperlaxity. He has hypertelorism and exotropia, inner epicanthic folds, micrognathia and open bite. He has excessive hypersalivation, chronic constipation, and idiopathic vomiting. His overall current level of functioning meets the DSM-5 criteria for severe ID, both in terms of intelligence (based on the Snijders-Oomen Nonverbal Intelligence Test) and adaptive functioning (based on the Vineland-Z) [163,164]. He has pronounced speech impairment and only uses a few simple words. At the behavioral level, he meets DSM-5 criteria for ADHD and ASD, the latter was confirmed with the Autism Diagnostic Observation Schedule and the ASD screening algorithm of the Developmental Behavior Checklist [165,166]. His current behavior is mainly characterized by an atypical social approach, limited social reciprocity, stereotyped/repetitive motor movements, and hypersensitivities (e.g., to auditory stimuli, food selectivity). Furthermore, relatively high levels of externalizing psychopathology, such as hyperactivity, screaming, and aggressive/destructive behavior (e.g. temper tantrums, hitting

others, destroying materials), are reported by parents and the teacher on the Developmental Behavior Checklist [167].

P2 is a 22-year-old male of mixed Japanese/Mexican origin. He has Lennox-Gastaut syndrome and mild spasticity and dystonia. He has tall stature (height 185 cm, height father 168 cm, height mother 164 cm). He also has microcephaly, prominent kyphosis, pectus excavatum, inverted nipples, brachycephaly, slightly upslanting palpebral fissures and micrognathia. He has chronic constipation. Brain MRI revealed mild asymmetry of the hippocampi and cerebellar hemispheres. Based on an evaluation at the age of 20 using the Wechsler Adult Intelligence Scale and the Adaptive Behavior Assessment System, his current level of functioning meets criteria for moderate ID [168,169]. Results of the Social Responsiveness Scale were consistent with his previous ASD diagnosis [170]. Based on the Achenbach System of Empirically Based Assessment questionnaires, he also showed behavior consistent with a diagnosis of ADHD [171]. During childhood, several additional problems were reported, with mainly eating and sleeping difficulties. His fine motor skills are below his chronological age expectancy level based on the Beery-Buktenica Developmental Test of Visual-Motor Integration [172]. Besides the *de novo* *KLHL20* variant, a maternally inherited variant in *ATP1A2* ([NM_000702.4]: c.1148G>A, p.R383H) was found, which was previously reported to cause familial hemiplegic migraine [173]. This was consistent with the occurrence of severe migraine with and without aura in the mother.

P3 is a 25-year-old Spanish male who presented at the age of 6 months with febrile seizures. He has psychomotor and language delay, kyphosis, scoliosis, and tapered fingers with distal joint laxity. He has hypertelorism, inner epicanthic folds and exotropia. During infancy, he had chronic constipation. Brain MRI was normal at the age of two, but vertebral arterial tortuosity and hypoplasia of the inferior vermis was observed at age 18. Further investigations revealed aortic root dilatation. He has anxiety and has been diagnosed with mild ID and hyperactivity. He attended regular school with an adapted program.

P4 is an 18-year-old Spanish male that that presented with at the age of 24 months with febrile seizures. He had mild developmental delay. On physical examination he has tall stature (height 190 cm, height father 180 cm, height mother 170 cm), pectus excavatum, kyphosis, and distal joint laxity. He has excessive hypersalivation. He has aggressive behavior, hyperactivity, and temper tantrums. No standardized neurobehavioral test data are available.

P5 is an 8-year-old girl of mixed Eastern European/Filipino origin. She had globally delayed milestones before seizure onset and had a period of psychomotor regression at the onset of the

seizures at around the age of three years. She has drug-resistant epilepsy, failure to thrive, global developmental delay and regression with hypotonia and dystonia. She could speak words but lost her speech and motor skills and now has severe cognitive dysfunction. She sits with support and cannot stand or walk alone. Currently at age eight, she lacks all fine motor skills. She has hypertelorism and micrognathia. She has gastroesophageal reflux disease and constipation. She was reported to be hyperactive at age 4 years during clinic visits. Brain MRI revealed white matter volume loss, mild dilatation of the right trigone and thinning of the posterior part of the corpus callosum.

P6 is a 3-year-old Sephardic/Ashkenazi Jewish boy who presented at the age of six months with myoclonic seizures. This was followed by atonic and generalized tonic-clonic seizures at the age of 18 months which were drug resistant by three years. He has global developmental delay with pronounced language delay and a period of mild regression of language. His development had been assessed at the age of 3 years (score <50 Bayley III mental scale and Adaptive Behavior Assessment System II). In the same assessment period, he had been diagnosed with ASD (Autism Diagnostic Observation Scale 2 module 1 and DSM-V). He has inner epicanthic folds and brachycephaly. He has hyperactivity, aggressive and destructive behavior with temper tantrums, hitting and destroying materials. Brain MRI showed a mild enlargement of the Sylvian fissures.

P7 is a 6-year-old girl with global developmental delay and absence of speech, behavioral disturbances and recurrent episodes of febrile tonic-clonic seizures. She has a Vein of Galen aneurysmal malformation.

P8 is an 8-year-old boy with early delayed psychomotor milestones, microcephaly, severe to profound ID and epilepsy. Developmental regression at seizure onset was reported with improvement after the start of anti-seizure medication. He had an unstable gait when he was young and developed spasticity of the limbs afterwards. He has midface hypoplasia. His behavior is characterized by autistic features, hyperactivity, and hand stereotypies.

P9 is a 27-year-old Anglo-Australian female with delayed psychomotor development noted at 16 months. She is nonverbal and has moderate-severe ID. Her developmental age score was 24 months at 5,5 years (Psycho-Educational-Profile-Revised). She has a developmental and epileptic encephalopathy with predominant tonic and tonic-clonic seizures. Her seizures began at 18 months and febrile seizures were not noted. Seizures were predominantly nocturnal and tended to cluster. She had pulmonary stenosis, small muscular ventricular septal defect and small anomalous right pulmonary vein draining into low superior vena cava. She has dystonic posture

with internal rotation of the hips and dystonic gait. She has strabismus. She demonstrates stereotypic movements with figure-eight head movements. Chronic constipation and sleep disturbance were important comorbidities.

P10 is an 8-year-old Arab boy that presented at the age of six months with seizures and developmental delay. He has been diagnosed with severe ID, is nonverbal and has chronic constipation. His current behavior is mainly characterized by autistic features, stereotyped/repetitive hand movements and aggressive/destructive behavior.

P11 is a 35-year-old male who was referred to the genetic clinic for a history of ID. He was able to graduate high school with a certificate and currently lives independently. He has obesity, irritable bowel syndrome, and pre-diabetes. He has widely spaced eyes.

P12 is a 22-year-old adult male patient with moderate ID, behavioral difficulties and hyperactivity. He has narrow eyelids, a thin upper lip, low set ears and microcephaly. He has tapered fingers. His behavior is characterized by agitation, stereotypies, and hyperactivity. He was born with a normal head circumference developing at centile 25-50 until the age of 11. After 11, his head circumference did not further increase, resulting in microcephaly in adulthood (occipitofrontal circumference 53 cm at age 20).

P13 is a 12-year-old boy with a developmental and epileptic encephalopathy. He was noted to have delayed visual maturation at 2 months of age and was diagnosed with congenital exotropia, hypospadias, and undescended testis. He had five simple febrile seizures between the ages of 12 and 26 months before he presented with an explosive onset of afebrile generalized tonic–clonic and myoclonic seizures. He subsequently developed tonic and atonic seizures at 2 years and 9 months, multifocal seizures at 2 years and 11 months, and episodes of non-convulsive status epilepticus at 7 years. His development was felt to be within normal limits until 2 years at which stage it plateaued and he now has severe ID (based on an evaluation of his adaptive functioning at the age of 12 using the Vineland Adaptive Behavior Scales - third edition) with mild autistic features [174]. He can now speak in simple sentences. He developed a mild tremor and ataxia at 3 years and is now ambulant. Initial EEG was normal but at 26 months they showed diffuse slowing and multifocal spikes. By 4.5 years there was generalized spike and polyspike wave in addition to multifocal discharges which became almost continuous in sleep at 6 years. EEG at 12 years of age shows occipital discharges. MRIs at 6 months, 1.5 and 5 years (3T) were normal. He has had multiple hospital admissions and is drug-resistant. He continues to have tonic seizures and episodes of non-convulsive status epilepticus. He has some autonomic abnormalities with

altered temperature regulation and extreme flushing particularly with exercise and before sleep. He has mild pectus excavatum, mild bilateral fifth finger clinodactyly and a broad nasal tip.

P14 is a 23-year-old female who presented with neonatal convulsions at day 15. Later she developed seizures with fever, generalized tonic-clonic seizures, and tonic seizures with sometimes perioral myoclonias. She also has head-drop attacks. She has severe to profound ID. She had a normal head circumference at birth however since the age of 6 months a head circumference -2.5 standard deviations has been documented until adulthood. She had gross motor delay with unstable gait at the age of seven years and loss of walking after regression at the age of 12. She developed spasticity. She has tetraparesis. Her behavior is characterized by autistic features and stereotypies that are triggered by sound.

Supplement S3: Neurocognitive and behavioral data for P1 and P2

Summary of psychiatric diagnoses and corresponding test results for P1 and P2			
		P1	**P2**
Intellectual disability	Documented diagnosis (age)	Severe intellectual disability (5y)	Moderate intellectual disability (10y)
	Intelligence test (test(age):result)[a]	• *SON-R 2,5-7 (10y): developmental IQ ≈ 30 (M = 100, SD = 15)*	• *WAIS-IV (22y): Full Scale IQ = 42 (standard score with M = 100, SD = 15)*
	Adaptive behavior scale (test(age):result)[a]	• *Vineland-Z (9y): developmental age = 1;4 years*	• *ABAS-3 (22y): General Adaptive Composite = 50 (standard score with M = 100, SD = 15)*
Autism spectrum disorder	Diagnosis (age)	Autism spectrum disorder (3y)	Pervasive developmental disorder not otherwise specified (10y)
	Observation scale / interview / questionnaire autism spectrum disorder (test(age):result)[a]	• *ADOS – module 1 (unknown): 16 (cut-off autism = 12)* • *DBC-P – autism screening algorithm (parent report, 10y): 41 (cut-off autism = 17)*	• *SRS-2 (16y): > cut-off autism*
Attention-deficit/hyperactivity disorder	Diagnosis (age)	Attention-deficit/hyperactivity disorder (11y)	Attention-deficit/hyperactivity disorder (10y)
	Questionnaire attention-deficit/hyperactivity disorder (test(age):result) [a]	• *No information available*	• *CBCL (parent report, 10y) : Attention Problems (clinical range)* • *TRF (teacher report, 10y): Attention Problems (clinical range)*
Additional behavior or developmental problems	Clinical description	• Eating problems (e.g., food selectivity, pica) • Verbal problem behavior (e.g., screaming) • Aggressive/destructive behavior (e.g., temper tantrums, hitting/biting others, destroying materials)	• Motor problems • Eating problems • Sleep problems
	Questionnaire (test(age):result) [a]	• *DBC-P – Total Behavior Problem Score (parent report, 10y): 124 (clinical cut-off = 46)* • *DBC-T – Total Behavior Problem Score (teacher report, 10y): 65 (clinical cut-off = 30)*	• *DTVMI – Visual-Motor Integration (22y): < 45 (standard score with M = 100, SD = 15)* • *CBCL (parent report, 22y): Withdrawn/Depressed, Somatic Complaints, Social Problems, Thought Problems, Attention Problems, Aggressive Behavior (clinical range)* • *TRF (teacher report, 10y): Thought Problems, Attention Problems*

[a]Only the most recent test results available are mentioned. **Abbreviations:** ABAS, Adaptive Behavior Assessment System; ADOS, Autism Diagnostic Observation Schedule; CBCL, Child Behavior Checklist; DBC, Developmental Behavior Checklist; DTVMI, Developmental Test of Visual Motor Integration; IQ, intelligence quotient; SD, standard deviations; SON-R, Snijders-Oomen Nonverbal Intelligence Test; SRS, Social Responsiveness Scale; TRF, teacher report form; WAIS, Wechsler Adult Intelligence Scale; y, years.

CHAPTER 4
FACIAL PHENOTYPING IN MOLECULAR DIAGNOSTICS OF ASD

Explanatory note

In this chapter, we study how facial phenotyping can contribute to better molecular diagnostics of ASD. The chapter starts with a basic research study aimed at better understanding the genetic link between the brain and the face. A close developmental relationship between the brain and face is evident from the co-occurrence of neurodevelopmental problems, structural brain anomalies and dysmorphic facial features in patients with NDDs. We investigated in part 4.1 to which extent this relationship extends to common human genetic variation by comparing GWAS data of human brain and face shape. In part 4.2, we illustrate with a case study the importance of a facial dysmorphology assessment and deep phenotyping in children with a NDD even in the absence of ID. The opinion of dysmorphologists is currently considered the gold standard to assess facial dysmorphism but this is often subject to examiner bias and is highly dependent on training and clinical experience. Therefore, we collected 3D facial images of ASD patients to perform an objective analysis of facial dysmorphism in part 4.3. We then investigated whether objective facial phenotyping could help to improve the recognition of ASD patients with Mendelian causes.

4.1 Shared heritability of human face and brain shape

- Sahin Naqvi[1,2,#], Yoeri Sleyp[3,#], Hanne Hoskens[3,4], Karlijne Indencleef[4,5], Jeffrey P Spence[6], Rose Bruffaerts [7,8,9], Ahmed Radwan[4,10], Ryan J Eller[11], Stephen Richmond[12], Mark D Shriver[13], John R Shaffer[14,15], Seth M Weinberg[14,15,16], Susan Walsh[11], James Thompson[17], Jonathan K Pritchard[6], Stefan Sunaert[4,10], Hilde Peeters[3], Joanna Wysocka[18,19,20], Peter Claes[21,22,23,24]

[1]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA. [2]Departments of Genetics and Biology, Stanford University School of Medicine, Stanford, CA, USA. [3]Department of Human Genetics, KU Leuven, Leuven, Belgium. [4]Medical Imaging Research Center, University Hospitals Leuven, Leuven, Belgium. [5]Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium. [6]Departments of Genetics and Biology, Stanford University School of Medicine, Stanford, CA, USA. [7]Department of Neurosciences, KU Leuven, Leuven, Belgium, Hasselt University, Hasselt, Belgium. [8]Neurology Department, University Hospitals Leuven, Leuven, Belgium, Hasselt University, Hasselt, Belgium. [9]Biomedical Research Institute Hasselt University Hasselt Belgium, Hasselt University, Hasselt, Belgium. [10]Department of Imaging and Pathology, Translational MRI, KU Leuven, Leuven, Belgium. [11]Department of Biology, Indiana University Purdue University Indianapolis, Indianapolis, IN, USA. [12]Applied Clinical Research and Public Health, School of Dentistry, Cardiff University, Cardiff, UK. [13]Department of Anthropology, Pennsylvania State University, State College, PA, USA. [14]Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, USA. [15]Department of Oral and Craniofacial Sciences, Center for Craniofacial and Dental Genetics, University of Pittsburgh, Pittsburgh, PA, USA. [16]Department of Anthropology, University of Pittsburgh, Pittsburgh, PA, USA. [17]Department of Psychology, George Mason University, Fairfax, VA, USA. [18]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA. [19]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA. [20]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA. [21]Department of Human Genetics, KU Leuven, Leuven, Belgium. [22]Medical Imaging Research Center, University Hospitals Leuven, Leuven, Belgium. [23]Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium. [24]Murdoch Children's Research Institute, Melbourne, Victoria, Australia. #Contributed equally.
- Status: published in Nature Genetics (shared first author).
- Adapted from: Naqvi S, Sleyp Y, Hoskens H, Indencleef K, Spence JP, Bruffaerts R, et al. Shared heritability of human face and brain shape. Nat Genet. 2021;53(6):830–9.
- Personal contribution: performed a literature study on the shared development of brain and face shape and analyzed/interpreted GWAS signals through gene set enrichment and literature analyses.

4.2 Deep phenotyping hints towards an underlying Mendelian disorder: a case report

- Yoeri Sleyp[1], Ann Swillen[1], Kris Van Den Bogaert[1], Guy Massa[2], Peter Aerssens[2], Hilde Peeters[1]
  [1]Department of Human Genetics, KU Leuven, Leuven, Belgium. [2]Department of Pediatrics, Jessa Hospital, Hasselt, Belgium.
- Status: published in Clinical Dysmorphology (first author).
- Personal contribution: put together molecular and phenotypic data of literature cases, interpreted genetic data of proband, and drafted this section.
- Adapted from: Sleyp Y, Swillen A, Van Den Bogaert K, Massa G, Aerssens P, Peeters H. A de-novo 15q24.2 deletion involving SIN3A is associated with emotional, behavioural, motor problems and hypersensitivity in a girl with above average intelligence and typical facial features. Clin Dysmorphol. 2020;29(4):210–3.

4.3 Towards 3D facial analysis for recognizing Mendelian causes of ASD

- Yoeri Sleyp[1], Harold S Matthews[1,2,3], Michiel Vanneste[1,4], Laura Vandenhove[1], Valentine Delanote[4], Hanne Hoskens[1,2], Karlijne Indencleef[1,2], Hanne Teule[4], Maarten H D Larmuseau[1,5,6], Jean Steyaert[7], Koenraad Devriendt[1,4], Peter Claes[1,2,3,8], Hilde Peeters[1,4]
  [1]Department of Human Genetics, KU Leuven, Leuven, Belgium. [2]Medical Imaging Research Center, UZ Leuven, Leuven, Belgium. [3]Facial Sciences Research Group, Murdoch Children's Research Institute, Parkville, Victoria, Australia. [4]Center for Human Genetics, University Hospitals Leuven, Leuven, Belgium. [5]Antwerp Cultural Heritage Sciences, ARCHES, UAntwerpen, Antwerpen, Belgium. [6]Histories vzw, Gent, Belgium. [7]Center for Developmental Psychiatry, KU Leuven, Leuven, Belgium. [8]Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium.
- Status: in preparation for submission to European Journal of Human Genetics (first author).
- Personal contribution: collected 3D facial images, analyzed facial imaging and genetic data, coordinated the project, and drafted this section.

4 Facial phenotyping in molecular diagnostics of ASD

4.1 Shared heritability of human face and brain shape

4.1.1 Abstract

Evidence from model organisms and clinical genetics suggests a developmental interplay between the brain and face, but the role of this link in common genetic variation remains unknown. In the current study, comparison of GWAS results from brain shape and face shape revealed 76 overlapping loci. Shared loci included transcription factors involved in craniofacial development, as well as members of signaling pathways implicated in brain–face cross-talk. We do not detect a significant overlap between shared brain–face GWAS signals and variants affecting behavioral–cognitive traits. These results suggest that early in embryogenesis, the face and brain mutually shape each other through both structural effects and paracrine signaling, but this interplay may not be related to brain development associated with cognitive function.

4.1.2 Introduction

The development of the brain and face is highly integrated due to shared developmental lineage, spatial proximity and signaling cross-talk between both structures [98]. Early in embryonic development, the rostral end of the ectodermally derived neural tube gives rise to the forebrain, which in turn gives rise to the cerebrum that encompasses the cerebral cortex [175]. Just before forebrain formation, a subset of neuroepithelial cells within the neural folds give rise to facial progenitor cells called CNCCs [176]. Following specification, CNCCs undergo an epithelial-to-mesenchymal transition and migrate ventrally [177], giving rise to most of the craniofacial skeleton and connective tissue [178]. Early brain growth rates can modulate both positioning and outgrowth of the facial prominences [179,180], as well as induce flexion and bone deposition of CNCC-derived basicranial bones [181,182] and neurocranial sutures [183,184], respectively. Finally, paracrine factors secreted by either the developing forebrain [185–188] or CNCCs [189–191] modulate facial or brain development, respectively.

These physical and molecular interactions have been detailed by studies in developing chick and mouse embryos, but are also supported by widespread co-occurrence of neurodevelopmental and craniofacial malformations in rare human syndromes [192]. This phenomenon was noticed by DeMyer et al. [101] in 1964, who coined the phrase “the face predicts the brain” to describe correlations between the severity of brain and face malformations in patients with holoprosencephaly. While in some cases this co-occurrence may be caused by pleiotropic gene functions, a number of human syndromes have been mapped to genes functioning in brain–face cross-talk through paracrine signaling [193–195]. Nonetheless, close developmental links

between face and brain are underappreciated; whether and how they extend to common human genetic variation influencing brain and face shape is unknown.

4.1.3 Materials and methods

GWAS data

Brain shape GWAS was done as part of this study. We adapted our previously published data-driven phenotyping approach [196] to brain shape, as measured by MRI scans of 19,644 individuals in the UK Biobank. We focused on the mid-cortical surface (midway between the white–gray matter interface and the pial surface with the cerebrospinal fluid, as extracted using FreeSurfer), which we refer to as brain shape. Using mid-cortical surfaces represented by a mesh of 3D vertices, the method segments brain shape in a global-to-local manner, yielding brain segments at different hierarchical levels of scale. Additional information on how the brain shape GWAS was conducted can be found in the original publication [197]. Face shape GWAS data was collected from a previous study [198]. We also collected publicly available genome-wide summary statistics for 26 auxiliary traits encompassing neuropsychiatric disorders [21,199–203], behavioral–cognitive traits [204–206], subcortical volume measures [207–209] and immune-related disorders (as negative controls) [210–213].

GWAS peak overlap and annotations

In determining overlap between lead SNPs from different GWASs, we used the following strategy: two lead SNPs tag the same genetic locus if they are within 10 kb of each other or if they are within 1 Mb of each other and with an $r2 > 0.2$. To quantify the overlap between the brain shape GWAS (this study) and the GWAS results of 430 other studies from the NHGRI-EBI GWAS Catalog, we defined linkage disequilibrium (LD) blocks of 0.2 around the brain shape loci using PLINK v1.9, and then calculated the odds ratio and P value for the overlap between these blocks and any given GWAS using bedtools v2.27.1 with the fisher function.

In determining brain–face shared loci, we first considered the genome-wide lead SNPs from the brain GWAS and looked for any SNP within 10 kb or within 1 Mb and $LD > 0.2$ of these lead SNPs with at least a genome-wide suggestive association ($P < 5 \times 10-7$) association with face shape. This resulted in a first set of loci with evidence of association in brain and face shape. Then we took the genome-wide lead SNPs reported in the face GWAS and clumped them if two lead SNPs were within 10 kb or within 10 Mb with an $r2 > 0.01$. For the resulting independent genome-wide facial lead SNPs, we selected any SNP within 10 kb or within 1 Mb and with $r2 > 0.2$ with at least suggestive ($P < 5 \times 10-7$) association with brain shape. This resulted in a second set of loci with

evidence of association in brain and face shape and, together with the first set of loci, they were clumped (within 10 kb or within 1 Mb and an r2 > 0.2) into a final set of brain–face shared loci.

We manually identified candidate genes in the vicinity of the brain–face shared loci. For each locus, we first considered all genes within 500 kb of the lead SNP. We primarily relied on evidence for involvement of these genes in a human craniofacial or neurodevelopmental syndrome, or for evidence of craniofacial or neurodevelopmental defects in knockouts of their orthologs in mice. We also considered associations with gene ontology terms related to craniofacial development, neurodevelopment or skeletal system development. In some cases (that is, SOX9, where enhancer–promoter interactions over 1 Mb have been described [214]), we extended the window to within 750 kb from the lead SNP.

Clinical gene-panel overlap

Gene panels were downloaded from the Genomics England PanelApp website. Only panels used for clinical interpretation in the 100000 Genomes Project were selected (provided by PanelApp [215]). The clinical gene panels were merged in disease (sub)categories according to the 100000 Genomes Project criteria (for example, the clinical gene panel "Intellectual Disability" belongs to the subcategory "Neurodevelopmental Disorders", which is part of the "Neurology and Neurodevelopment" disease category). Only genes with high confidence for gene–disease association were included in the clinical gene panels. We calculated the overlap between genes from clinical panels/subcategories/categories and different gene-sets allowing for a 200kb, 500kb or 1Mb window around the loci. Significance was tested by generating 10000 random panels for each clinical panel subcategory/category with equal size using a list of 19198 protein-coding genes. P-values were obtained by dividing "the number of times the overlap random panel and gene-set was larger than the overlap clinical gene-panel/subcategory/category and gene-set" and "number of random gene-panels created (10000)". Clinical panels/subcategories/categories were interpreted as strongly or weakly enriched if they showed significance ($P < .05$) across three or two different gene-sets respectively.

Quantifying sharing of signals between pairs of GWAS

To assess the extent to which genome-wide profiles of association were shared between a pair of GWAS, we computed a Spearman correlation between two vectors of LD-block organized association P values. First, genome-wide SNPs were selected to overlap with the HapMap3 SNPs [216], and SNPs within the major histocompatibility complex region were removed. Second, we organized SNPs within 1,725 blocks in the human genome that can be treated as approximately independent in individuals of European ancestry [217]. For every LD block, we computed the

mean SNP −log10(P value), and then computed a rank-based Spearman correlation using the averaged association value (n = 1,725) for each LD block. A standard error of the Spearman correlation was estimated using statistical resampling with 100 bootstrap cycles with replacement from the 1,725 LD blocks.

4.1.4 Results

Loci affecting both brain and face shape

To test for sharing of genetic effects between brain and face shape, we intersected 472 loci associated with brain shape with 203 loci associated with face shape in individuals of European ancestry. Thirty-seven of the loci for brain shape were linked ($r2 > 0.2$) to at least one of the face shape loci, significantly above random expectation ($P = 2.03 \times 10-22$, odds ratio = 10.6) and greater than the overlap with other traits that have similar numbers of genome-wide significant associations in the NHGRI-EBI GWAS Catalog [218] (**online supplementary data**). Identifying signals showing a genome-wide significant association with one of brain or face shape and a suggestive ($P < 5 \times 10-7$) association with the other resulted in 76 brain–face shared loci (**figure 4.1**).

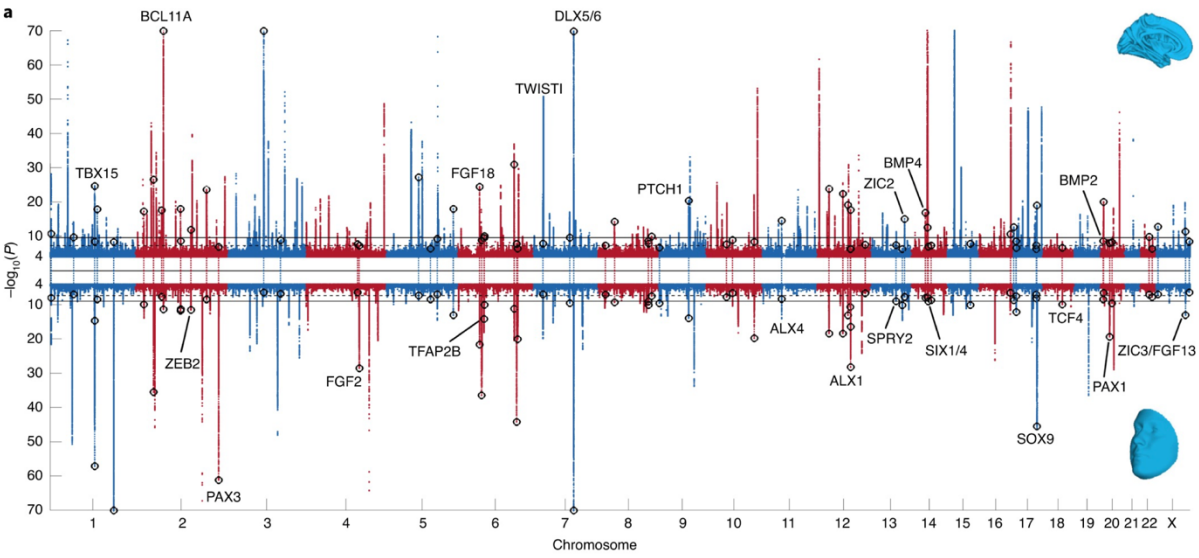


Figure 4.1: Loci affecting both brain and face shape. Miami plot of GWAS results for brain (top) and face (bottom) shape. All 76 loci reaching genome-wide significance ($P < 5 \times 10-8$) in one study and genome-wide suggestive significance ($P < 5 \times 10-7$) in the other are highlighted by unfilled circles.

Genes near the 76 brain–face shared loci were strongly enriched for disease associations, including “skeletal disorders” and “hearing and ear disorders”, consistent with the contribution of CNCCs to craniofacial skeleton and ear structures. We next manually scanned the 76 brain–face shared loci for genes with known roles in craniofacial or brain development from human syndromes and/or knockout mouse models (**online supplementary data**). We observed that

many of the shared brain–face loci included genes encoding transcription factors involved in neural crest formation and/or craniofacial skeletal development. Some of those transcription factors (for example, DLX5/6, SOX9, ZEB2, ZIC2, ZIC3 and TCF4) have known functions in both neural crest and brain development, and this pleiotropy may account for the shared brain–face genetic signals. However, other shared brain–face signals are associated with transcription factors thought to function primarily during neural crest development rather than brain development, and whose mutation causes specific craniofacial defects; those transcription factors include ALX1 and ALX4 (associated with frontonasal dysplasias [219,220]), TWIST1 (associated with Saethre–Chotzen syndrome [221]), PAX3 (associated with Waardenburg syndrome [222]) and TFAP2B (associated with CHAR syndrome [223]).

Interactions between face and brain can be architectural, with the forebrain acting as a structural support for facial development, and facial skeletal structures flexing to accommodate early brain growth [224]. However, these interactions can also involve paracrine signaling, with fibroblast growth factor (FGF), Hedgehog and bone morphogenetic protein (BMP) pathways known to mediate the signaling from the developing brain to the face [185–187]. Interestingly, genes encoding members of all three pathways, FGF (*FGF2*, *FGF13*, *FGF18* and *SPRY2*), Hedgehog (*PTCH1*) and BMP (*BMP2* and *BMP4*) are among the shared brain–face loci. For example, mutations in *PTCH1*, encoding the receptor for the sonic hedgehog ligand, cause holoprosencephaly [225], a congenital, structural forebrain anomaly with associated craniofacial malformations. Conversely, CNCCs secrete anti-BMP signaling molecules that modulate forebrain development [189,190]; expression of these BMP antagonists is dependent on the SIX family of transcription factors, whose perturbation in CNCCs leads to both craniofacial malformations and secondary pre-otic brain defects [226]. *SIX1* and *SIX4* are also among the 76 brain–face shared loci (**figure 4.1**). Furthermore, genes linked to other signaling pathways, including Wnt (*DAAM1*, *DAAM2*, *TNKS*, *AHI1*, *FBXW11* and *MCC*) and transforming growth factor beta (*LEMD3* and *PPP2R3A*), are among the shared brain–face loci.

Genome-wide sharing of signals with neuropsychiatric disorders and behavioral–cognitive traits

We compared the degree to which brain shape and face shape shares GWAS signals with neuropsychiatric disorders, behavioral–cognitive traits and subcortical volumes. As approximate negative controls, we used four immune-related diseases shown to have minimal genetic correlation with schizophrenia and bipolar disorder [227]. Brain shape shares significant signal with most neuropsychiatric traits, as well as all behavioral–cognitive and subcortical volume traits analyzed (**figure 4.2**). In contrast, face shape does not show significant sharing with any of the neuropsychiatric disorders or behavioral–cognitive traits, and significant but weaker sharing

with the subcortical volume measures (**figure 4.2**). Sharing between brain shape and the immune diseases was generally lower than with neuropsychiatric disorders, behavioral–cognitive traits or subcortical volumes, but reached significance for type 1 diabetes and rheumatoid arthritis (**figure 4.2**). This overlap may be because these immune traits have genetic correlation with brain-related traits other than those tested previously (schizophrenia and bipolar disorder). Altogether, the substantial sharing of signals between brain and face shape appears to be mostly independent of neuropsychiatric disorder risk and behavioral–cognitive traits, perhaps because mutual influences of face and brain shape on each other involve phenotypic effects on brain shape distinct from those influencing neuropsychiatric disorder risk and behavioral–cognitive traits.

4.1.5 Discussion

Here, we found a striking convergence of common genetic variation affecting brain and face shape. These observations suggest a larger than previously appreciated role of the face in shaping development of the brain and its morphological variation between individuals. However, these shared genetic effects do not appear to substantially impact neuropsychiatric disorder risk or cognitive functions. Our results are therefore consistent with a model whereby CNCCs and their derived cranial structures substantially influence brain shape through both physical interactions and paracrine signaling early in embryogenesis, but later shaping of cortical morphology, through processes such as the folding of the cortical surface [228], has a greater impact on cognitive traits. Nevertheless, we cannot exclude the possibility that future GWASs of cognitive traits show more substantial overlap with brain–face shared genetic effects, perhaps due to alternative trait definitions or to greater statistical power.

A number of developmental mechanisms could mediate shared brain–face genetics. One potential contribution comes from the common neuroepithelial origins of the two structures, with genes influencing growth, patterning and cell fate decisions within the neural plate ultimately affecting cell allocation within distinct parts of the brain and face; examples of such neural plate genes within brain–face shared loci include *ZIC2* and *ZIC3* [229–231]. Another potential mechanism entails common genetic variation modulating expression of genes with independent roles in both brain and face development. *SOX9*, encoding a transcription factor with key functions in neural crest development and chondrogenesis, but which is also required for gliogenesis [232], is an attractive candidate for this mechanism. In addition, proximity-based mechanisms, which can be either structural or mediated by paracrine signaling, have also been linked to shared brain-face development before [224].

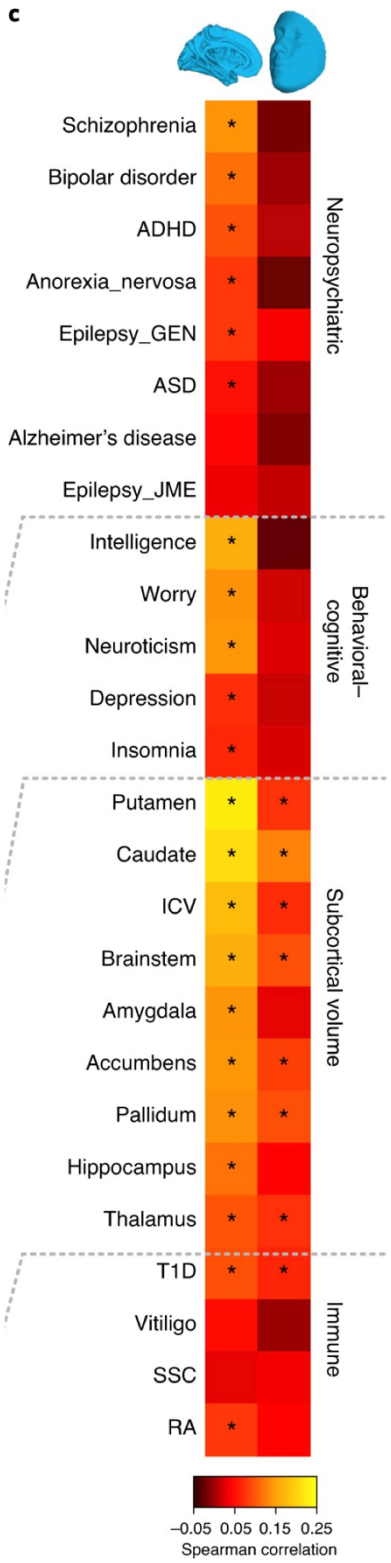


Figure 4.2: Genome-wide sharing of signals of brain shape and face shape with neuropsychiatric disorders, behavioral–cognitive traits, subcortical volumes and immune disorders. Genome-wide sharing of signals between any two given GWASs was assessed by Spearman correlation of LD block-average SNP −log10(P values) (material and methods). Spearman correlations between shape effects on the full brain (left) or face (right) with the indicated traits. *5% false discovery rate based on bootstrapped P value (Materials and methods). ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder; GEN, generalized epilepsy; JME, juvenile myoclonic epilepsy; ICV, intracranial volume; RA, rheumatoid arthritis; SSC, systemic sclerosis; T1D, type 1 diabetes.

While brain and face development must be tightly coordinated, the former is thought to have greater structural effects on craniofacial development, as the forebrain can serve as structural support for facial development as well as induce flexion of the basicranium and bone deposition at coronal sutures through growth-dependent tensile forces [182,183,224]. However, we find multiple brain–face shared loci near transcription factors with known roles in, and expression specific to, CNCCs and their derivatives. Furthermore, mutations in genes encoding these transcription factors result in malformations of the frontal facial skeleton, such as coronal synostosis (*TWIST1*) [221] or frontonasal dysplasias (*ALX1* and *ALX4*) [219,220]. One explanation for these results is that these transcription factors control regulatory programs ultimately modulating the ability of the craniofacial skeleton to respond to and accommodate brain growth, causing subtle changes in brain shape. It is also possible, however, that these transcription factors exert some phenotypic effects on brain shape by regulating expression of signaling ligands secreted from the face. For example, CNCCs secrete BMP antagonists that modulate forebrain development by blocking BMP and FGF production in the anterior neural ridge [189,190]. BMP antagonist production in CNCCs is regulated by the SIX family transcription factors [226], with *SIX1*/*SIX4* lying near a shared brain–face GWAS signal (**figure 4.1**). In the reverse direction, studies in chick embryos have shown that Fgf, Shh and BMP ligands are secreted by the forebrain and regulate the formation of the frontonasal ectodermal zone, a signaling center that in turn patterns the frontonasal prominence of the developing face [185–188,233]. Notably, our study implicates all three of these signaling pathways, nominating specific ligands and receptors whose modulation may be associated with the brain–face cross-talk. Furthermore, our study nominates other pathways, such as Wnt and transforming growth factor beta, for roles in paracrine brain–face signaling. Altogether, we uncovered common genetic variants yielding numerous candidate molecular players whose diverse mechanistic roles in mediating brain–face interactions during development can be examined in future studies.

4.1.6 Supplementary information

All supplementary data can be found in the original publication [197].

4.2 Deep phenotyping hints towards an underlying Mendelian disorder: a case report

4.2.1 Summary/Case study

The proband is the second child of unrelated parents. She was born at term by caesarean section with a birth weight of 2800 grams (-1.6 SD) and a length of 46 cm (-2.2 SD). The pregnancy was complicated by hypertension caused by IgA nephropathy for which the mother was treated with labetalol. Neonatal screening revealed elevated thyroid stimulating hormone levels indicating congenital hypothyroidism. Treatment with L-thyroxine was started at the age of 14 days. There was neonatal hypotonia, feeding difficulties and failure to thrive. At the age of 2 years there was short stature (-1.7 SD), peculiar facial features as described at an older age and behavior characterized by emotional instability, temper tantrums (often in response to frustration), social anxiety, poor attention span and hyperactivity. She had delayed motor milestones with standing with support at 12 months and crawling at 18-19 months. She was raised in a bilingual environment (Flemish/English). Her speech was slightly delayed with first words at 1 year 10 months and difficulties with articulation until the age of 5 years. However, she had good language abilities later on. At age 7 years 4 months, she was clinically assessed because of concerns about her academic progress (particularly within the domain of mathematics) as well as fine and gross motor difficulties, problems with attention and focus, and social skill delays. On examination she had peculiar facial features (**figure 4.3** at age 10), short stature with height 119.5 cm (-1.1 SD) and weight 23.0 kg (-0.4 SD). Currently at the age of 11 years, she follows regular school with some educational support for mathematics mainly because of problems with automatization. She still has fine motor problems: poor handwriting, slower pace, and problems with routine daily motor tasks.

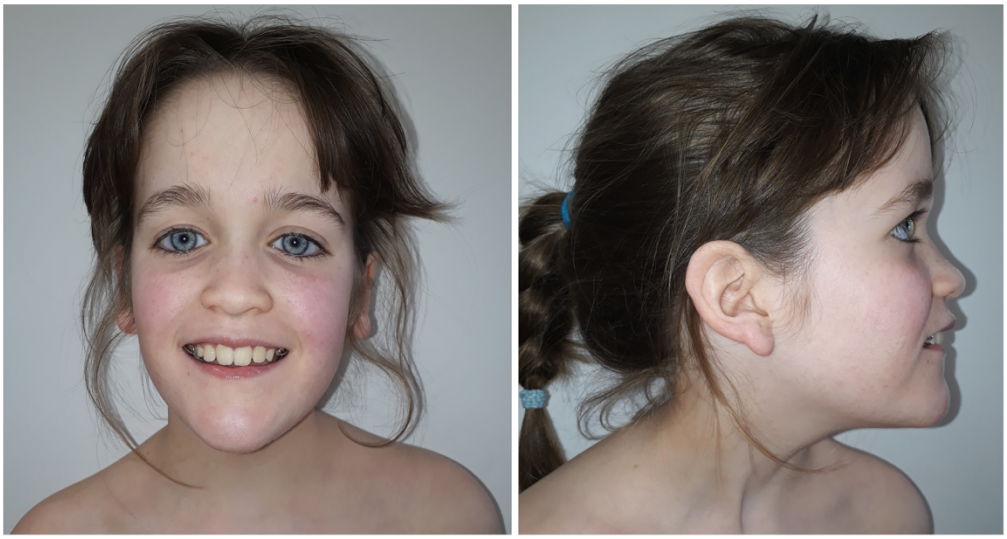


Figure 4.3: Facial photograph of the proband. The proband, aged 10 years, showing a high forehead, pointed chin, low-set ears, long and downslanting palpebral fissures.

4.2.2 Investigations

Several instruments have been used to get insight into the probands neurocognitive and socio-emotional developmental profile. **Table 4.1** gives an overview of her results across the cognitive domains and visual-motor integration. Her overall cognitive abilities (WISC-IV) were above average, however significantly lower scores were seen for processing speed. Academic assessment (WJ III ACH) revealed above average scores for reading and spelling but a low average score for math fluency. Visual Perceptual and Fine Motor Skill Assessment (Beery VMI-6) indicated that she had significant difficulties with integrating visual perceptual and finger hand movements. The diagnosis of developmental coordination disorder was made. In terms of social and emotional adjustment, there were many areas of concern: multi-rater (parents and teacher) results on the Behavior Assessment System for Children and the Adaptive Behavior Assessment System revealed clinically significant ratings in hyperactivity and inattention, atypicality and withdrawal [234,235]. She meets DSM-5 criteria for ADHD (predominantly inattentive presentation). On the social level, she has difficulties with social interactions with peers while social interactions with adults are adequate. Scores on the Social Responsiveness Scale-2 reveals a T-score of 69 (90% interval score 64-73) which is within the clinical range for mild-moderate problems with social responsiveness [236]. Because of some issues with sensory processing, the Sensory Profile questionnaire was used [237]. The latter revealed very high scores for auditory hypersensitivity (sensitivity to noise), visual hypersensitivity (very sensitive to light and to the sun), high tactile sensitivity (tags on clothes), and problems with modulation of tone and activity. Microarray analysis was part of the routine genetics workup at the age of 2 years towards the etiology of multiple medical and behavioral problems as described.

Genomic DNA was isolated from peripheral blood samples of the patient and her parents and was analysed on a 180k CytoSure Constitutional v3 array (Oxford Gene Technology). Data analysis was performed using bench lab CNV (Agilent Technologies) and was based on hg19. Array comparative genomic hybridization analysis revealed a *de novo* heterozygous deletion in the 15q24 locus. The deletion had a minimal size of 84 kb ([hg19]:chr15:75589027-75673311) and a maximal size of 164 kb ([hg19]:chr15:75535348-75699660). The deletion contains at least four protein-coding genes (*COMMD4*, *NEIL1*, *MAN2C1*, *SIN3A*) and is the smallest deletion within the 15q24 microdeletion region reported so far (**figure 4.4**).

Table 4.1: Overview of the results from neurocognitive and academic assessment.			
Instrument	**Scaled score/standard score (CI)**	**Percentile rank**	**Classification**
The Wechsler Intelligence Scale for Children IV Indices (WISC-IV) [238]			
General Ability Index	119 (113-124)	90	Above average
Verbal comprehension	120 (111-126)	91	Above average
Perceptual Reasoning	114 (104-121)	82	Above average
Working Memory	100 (90-108)	50	Average
Processing speed	83 (76-93)	13	Below average
Woodcock-Johnson III Tests of Achievement (WJ III ACH) [239]			
Broad reading	108 (104-112)	70	Above average
Broad math	97 (89-105)	42	Average
Broad written expression	99 (92-105)	46	Average
Beery-Buktenica Developmental Test of Visual-Motor Integration 6 (Beery VMI-6) [240]			
Total score	78	7	Below Average

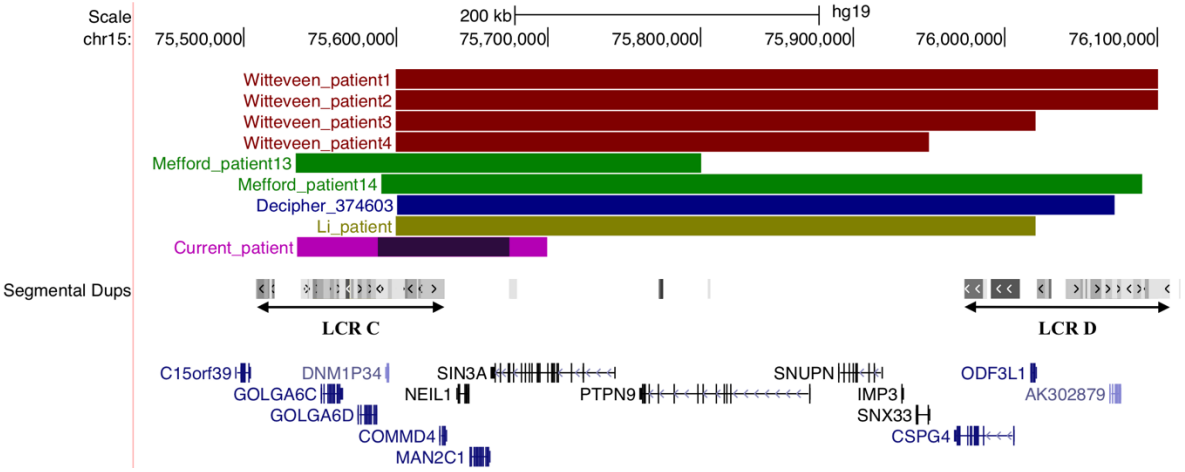


Figure 4.4: Overview of patients with 15q24 microdeletion encompassing *SIN3A*. The deletion in the current patient has a minimal size of 84kb ([hg19]:chr15:75589027-75673311) and a maximal size of 164 kb ([hg19]:chr15: 75535348-75699660). The deletion affects at least four protein-coding genes (*COMMD4*, *NEIL1*, *MAN2C1* and *SIN3A*). This figure was visualized using the UCSC genome browser (hg19).

4.2.3 Discussion

The 15q24 microdeletion syndrome is caused by recurrent deletions of varying size in the 15q24 locus, a complex region of 5.6 Mb that harbours 5 low copy repeats (LCRs) referred to as A, B, C, D and E. Witteveen et al. delineated the clinical phenotype that is associated with loss of function mutations in *SIN3A* and found overlapping features with patients with deletions in the region flanked by the segmental duplication blocks C and D [241]. Core features are ID and specific facial dysmorphism characterised by a long face with pointed chin, a high and broad forehead, downslanting palpebral fissures and a low nasal bridge. The deletion in the current patient involving a breakpoint within *SIN3A* is the smallest deletion within the 15q24 C-D region reported so far (**figure 4.4**). She has specific dysmorphic features (**figure 4.3**), however unlike other patients she has above average intelligence and medical problems including congenital hypothyroidism, hypersensitivity, automatization problems and problems with sensory processing that have not been specified in other patients so far. A comparison of the features in patients with 15q24.2 deletions within the LCR C-D region and mutations involving *SIN3A* is provided in **table 4.2** [241–244]. No information was obtained after request on DECIPHER patient 374603 (**figure 4.4**).

We illustrate with this case a rather favourable outcome of a child with several medical and developmental problems in infancy related to a 15q24 microdeletion affecting *SIN3A*. We further show that *SIN3A* haploinsufficiency is not always causing ID and may be accompanied by sensory processing difficulties. These findings may allow for a more precise counseling and surveillance of development in young children with deletions and mutations involving *SIN3A.*

Table 4.2: Phenotypic features of patients with 15q24.2 deletions within the LCR C-D region and a series of patients with point mutations involving *SIN3A*.

	current patient	**W1** [241]	**W2** [241]	**W3** [241]	**W4** [241]	**M13** [242]	**M14** [242]	**L1** [243]	**W5-13** [241] **and N1** [244]
Genetic variant	DEL	DEL	DEL	DEL	DEL	DEL	DEL	DEL	SNV
Position on chr15 (Megabase)	75.59-75.67	75.60-76.10	75.60-76.10	75.60-76.02	75.60-75.95	75.53-75.80	75.59-76.09	75.60-76.02	*SIN3A*
Size	84 kb	500 kb	500 kb	420 kb	350 kb	270 kb	500 kb	420 kb	/
Inheritance	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	UNK	4 *de novo*, 4 inherited, 2 UNK
Sex	F	F	F	F	M	UNK	F	F	6M, 4F
Birth weight	P3	P2.3	P25	P16	P50	NR	NR	P25	2/7 low
Neonatal hypotonia	+	NR	NR	NR	NR	NR	NR	NR	1/7
Congenital hypothyroidism	+	NR	NR	NR	NR	NR	NR	NR	NR
Short stature	-	+	-	-	-	-	+	NR	5/9
Microcephaly	-	+	+	-	-	+	-	NR	3/8
High/broad forehead	+	+	+	+	+	NR	+	NR	7/8
Nasal bridge	low	high	NR	low	NR	NR	flat	flat	2/8 (1 depressed, 1 flat)
Downslanting palpebral fissures	+	+	+	-	+	NR	+	NR	6/8
Pointed chin	+	+	+	+	+	NR	+	NR	5/8
Intellectual disability	-	mild	mild	mild	mild	mild	border	NR	1/10 severe, 7/10 mild 2/10 border
Psychiatric disorders	ADHD, DCD	ASD, CD	-	-	-	OCD	possible ASD	NR	5/7 (ASD, DCD, ADHD, OCD)
Hypersensitivity	+	NR	NR	NR	NR	NR	NR	NR	NR
Temper tantrums	+	NR	NR	NR	NR	NR	NR	NR	2/2
Epilepsy	-	-	+	-	-	NR	NR	NR	2/9
Brain anomalies	-	DC, DCC, WMA	DC, DCC, WMA	NR	NR	NR	DCC, VD	NR	4/4 (DC, DCC, WMA, VD)
Cardiac anomalies	-	-	-	-	-	-	-	TOF	0/10
Hypermobile joints	-	+	-	+	-	+	-	NR	3/7
Digital abnormalities	-	+	+	-	-	+	+	NR	4/7
Ectodermal abnormalities	-	-	thin hair	thin hair	-	NR	NR	NR	1/7 (thin hair, brittle nails, teeth anomalies)
Constipation	-	+	NR	NR	+	NR	NR	NR	1/1
Hearing impairment	-	+	-	-	-	NR	NR	NR	3/10
Visual anomalies	-	+	-	-	-	-	-	NR	2/7
Hypotonia	+	+	-	-	-	-	-	NR	3/7

+, present; −, absent; /, not applicable; ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorders; border, borderline intellectual functioning; CD, conversion disorder; DEL, deletion; DC, dysgenesis of cortex; DCC, dysgenesis of corpus callosum; DCD, developmental coordination disorder; f, female; m, male; NR, not reported; OCD, obsessive compulsive disorder; SNV, single nucleotide variant; TOF, tetralogy of Fallot; UNK, unknown; VD, ventricle dilatation; WMA, white matter abnormalities.

4.3 Towards 3D facial analysis for recognizing Mendelian causes of ASD

4.3.1 Abstract

Recognizing Mendelian causes of ASD remains a cornerstone in molecular diagnostics and counseling but is often complicated by the large variability of clinical features that may represent the milder end of the spectrum of Mendelian ASD. In this work, we explore the potential of objective facial phenotyping to improve the recognition of ASD patients with Mendelian causes. 3D facial images were collected from 152 ASD patients who came to consultation for a molecular diagnostic workup. Univariate computational dysmorphism and asymmetry scores indexing unusual 3D facial development were calculated based on the 3D facial shape of ASD patients and were investigated in relation to the presence of a Mendelian cause. Univariate dysmorphism ratings by experts were used to determine the possible added value of 3D facial shape analysis in addition to the clinical opinion in recognizing Mendelian ASD. Computational dysmorphism scores and asymmetry scores were significantly increased in ASD patients with Mendelian causes. Computational scores improved the recognition of Mendelian ASD beyond using individual expert ratings alone. We conclude that 3D facial analysis is a promising method for the objective evaluation of facial shape that may assist clinicians in recognizing Mendelian causes of ASD.

4.3.2 Introduction

ASD is a NDD typically characterized by deficits in social interaction and communication, as well as restricted and repetitive patterns of behavior, interests, or activities. With the advent of NGS, many Mendelian causes for ASD have been identified where the main contributor to ASD risk is a single genetic variant. As with other NDDs, Mendelian causes for ASD are often associated with ID and complex morphological phenotypes [60,65,81,82]. However, Mendelian ASD also comprises conditions with borderline intellectual functioning or normal cognition and subtle or no dysmorphic facial features. However, this large variability in the number and severity of clinical features in patients with Mendelian causes makes it difficult to distinguish them from ASD patients with multifactorial causes using current methods. In clinical practice, recognizing Mendelian ASD is of critical importance, as NGS is the standard of care in the molecular diagnostic workup of suspected Mendelian ASD but not of multifactorial ASD. Therefore, there is a need for better phenotyping strategies to improve the stratification of ASD patients according to underlying genetic causes.

In NDDs, craniofacial dysmorphism is an important feature suggesting a Mendelian cause. Craniofacial dysmorphism is defined as the presence of multiple minor anomalies, which are morphological deviations with little to no functional or medical consequences that are rare

(typically <2.5%) in the general population [84]. Several studies have previously reported an increased prevalence of minor facial anomalies in ASD patients compared to controls [105]. Facial asymmetry has consistently been found to be increased in ASD patients relative to controls [102,104,245]. However, inconsistent results have been reported across studies for specific anomalies that are more likely to occur in ASD [105], likely due to differences in inclusion criteria, measurement methods, and the genetic heterogeneity of ASD. Currently, clinical assessment by a dysmorphologist is the gold standard to assess minor facial anomalies and craniofacial dysmorphism. However, these assessments are subjective and largely depend on training and experience of the dysmorphologist. Furthermore, facial asymmetry is not routinely assessed in a standard dysmorphological workup and its value in recognizing Mendelian ASD is unknown.

In this work, we explore the utility of objective facial phenotyping to improve the recognition of ASD patients with Mendelian causes. We investigated the 3D facial shape of 152 ASD patients in relation to the presence of ID on the one hand and to the presence of a Mendelian cause on the other hand. We used the assessment of facial gestalts by experienced dysmorphologists to determine the possible added value of 3D facial shape analysis in addition to the clinical opinion in recognizing Mendelian ASD.

4.3.3 Materials and methods

Participants

Recruitment of participants is described in the **supplementary methods**. Demographic factors (i.e., sex, age, self-reported ancestry) were registered for all participants in this study. Only participants with self-reported European ancestry were included. Welch's t-test (two-tailed) was used to calculate whether there were differences between mean ages of the patient and control groups, while Fisher's exact test (two-tailed) was used to measure differences in sex (**supplementary table 4.1**). To minimize influences of environmental factors on facial morphology, we excluded participants who were obese (body mass index > 30) or underweight (body mass index < 18) and those who had orthognathic surgery in the past.

Genotyping and genetic analyses

Genetic testing was performed for all patients in this study by means of WGS except for patients for whom a Mendelian cause was identified either by CMA analysis or by targeted sequencing of NDD risk genes. "Mendelian causes" were defined as causes where the main contributor to ASD risk is a single highly penetrant rare variant such as a pathogenic SNV, indel or CNV. "Susceptibility variants" were defined as rare CNVs that are known to be associated with ASD, but which are characterized by low penetrance. ASD patients in whom no Mendelian cause or

susceptibility variant was found are further referred to as patients with an "unknown cause". The flow of genetic testing of patient samples is shown in **supplementary figure 4.1**. Identified Mendelian causes and susceptibility variants per patient are provided in **supplementary table 4.2**.

Spatially dense facial quasi-landmarking

After 3D facial image acquisition (described in **supplementary methods**), in-house software was used to manually edit the 3D facial images to remove hair, ears, and dissociated surface polygons. Next, all images were manually annotated with five positioning landmarks to establish rough facial orientation. Non-rigid dense surface registration of the images was then performed in Matlab 2017b using the MeshMonk toolbox [246]. Essentially, this samples each facial surface at 7160 points, by fitting a predefined template face onto each image. The facial surface of each participant is then represented by spatially-dense configuration of 3D points that correspond anatomically across all participants (e.g., if point number x of the template is located on the tip of the nose this point is placed on the tip of the nose for each participant), which allows the images of multiple participants to be compared in multivariate statistical analysis.

Facial signatures

A facial signature, as first described by Hammond et al., consists of z-scores calculated for each point on the face and is an objective description of how a participant differs from an age-, sex- and ethnicity-matched reference population [115]. We used the publicly available 3D growth curves of individuals of European ancestry to define the expected face and standard deviations per point on the face in each direction (x: lateral-medial; y: inferior-superior; z: anterior-posterior; and locally perpendicular: inward-outward), for each age and sex [247]. Each participant was aligned to their age and sex appropriate expected face using a robust Procrustes transformation and their displacement from the expected face in each direction at each point was normalized according to the corresponding standard deviation. Inward and outward displacements perpendicular to the surface were visualized in color maps by blue and red colors respectively.

Facial metrics

1. Expert dysmorphism scores

To determine the presence of facial dysmorphism according to the clinical expert opinion, 10 experienced dysmorphologists from six genetic centers in Belgium, hereafter referred to as "experts", evaluated the facial gestalts of the ASD patients. All experts were aware that they were scoring the facial gestalt of ASD patients only. Each evaluation was done based on three two-

dimensional facial images (frontal, left, right view of the face) and one video of the 3D image (left-right lateral rotation). Facial gestalts were scored as non-dysmorphic (score 0), uncertain (score 1) or dysmorphic (score 2). An average expert dysmorphism score was calculated for each patient by taking the average score across all 10 experts and was considered to reflect clinically assessed dysmorphism in each patient. Light's Kappa was used to estimate the inter-rater agreement and was calculated in RStudio version 1.1.383 by taking the average of Cohen's Kappa across all rater pairs. We interpreted the kappa coefficient as recommended by Landis and Koch: poor agreement if < 0.00, slight agreement if 0.00–0.20, fair agreement if 0.21–0.40, moderate agreement if 0.41–0.60, substantial agreement if 0.61–0.80 and almost perfect agreement if 0.81–1.00 [248].

2. Computational dysmorphism scores

Computational dysmorphism scores were calculated using the facial signatures by taking the sum of the absolute value of z-scores at 7160 points across the face in x, y, and z directions. Computational dysmorphism scores represent the overall magnitude of deviation from the reference population and should therefore reflect the magnitude of facial dysmorphism in the patient.

3. Computational asymmetry scores

Following image processing with MeshMonk each face was represented by 7160 quasi-landmarks. The facial template used was constructed to include bilaterally paired and midline quasi-landmarks [249]. To remove effects of overall size of each configuration on asymmetry, all faces were isotropically scaled to unit size. Asymmetry was then calculated according to the Klingenberg protocol [250]. First, a reflected copy of the entire configuration of quasi landmarks was created by reversing the sign of the x coordinate of each landmark, reflecting the configuration about the Y-Z coordinate plane. The indexes (position in the list of landmark coordinates) of each bilateral pair of landmarks in the reflected copy were interchanged left to right. This effectively relabels each point on the left of midline as the corresponding point on the right and vice-versa. This reflected and relabeled copy was then aligned to the original configuration via Procrustes superimposition. The residual distances between the two configurations represent asymmetry; each point on the left of the midline on the original face is compared to the corresponding point on the right of the original (left of the reflected relabeled copy) and vice-versa. The root mean square of these distances was used as the computational asymmetry score for each participant.

Statistical analysis

Between-group comparisons

A family structure preserving permutation analysis was performed to test for significant differences in distributions of average expert dysmorphism scores, computational dysmorphism scores and computational asymmetry scores between ASD patients, unaffected siblings and unselected controls, or between ASD subgroups. The family structure in our data needed to be preserved to satisfy the assumption of exchangeability in permutation testing. For each permutation, each family in the dataset was randomly assigned a binary label that indicated whether the group labels of all members of the family would be flipped. The z-score approximation of the Wilcoxon rank-sum statistic was used as the test statistic. This was re-calculated for each permuted dataset to estimate the empirical null distribution. The one-tailed p-value of the test was calculated as the proportion of the null distribution above the actual value of the test statistic. The Benjamini-Hochberg procedure was used to control the false discovery rate at 0.05 and to adjust p-values for multiple testing [251]. Area under the curve (AUC) of the receiver operating characteristic (ROC) curve was used as a measure of the effect size of the difference between two groups [252]. Spearman's Rho was used to measure the correlations among average expert dysmorphism scores, computational dysmorphism scores and computational asymmetry scores.

Logistic regression models

We fitted logistic binomial regression models to estimate the probabilities that an ASD patient has a Mendelian cause (i.e., ability to discriminate patients with Mendelian causes from patients with unknown causes) as a function of expert dysmorphism scores, computational dysmorphism scores and/or computational asymmetry scores. Regression models were trained and tested using stratified k-fold cross validation (k=3), ensuring the same proportion of Mendelian causes in test and training sets. The cross-validated AUC for each model was calculated by computing the standard ROC curve and AUC for each fold and by taking the average AUC across all three folds. The average AUC value indexed the discriminative capacity of each regression model. AUC values were interpreted as the following: no discrimination 0.40-0.60, some discrimination 0.61-0.70, fair discrimination 0.71-0.80, good discrimination 0.81-0.90, excellent discrimination 0.91-1.

4.3.4 Results

High average expert dysmorphism scores in ASD patients are associated with Mendelian causes, susceptibility variants and ID

Facial dysmorphism in the group of 152 patients with ASD was assessed by 10 experts. In a total of 1520 assessments, faces were considered 771 times as non-dysmorphic (51%), 308 times as uncertain (20%) and 441 times as dysmorphic (29%) (**figure 4.5A**). The inter-rater coefficient of agreement between experts, also known as kappa coefficient, was 0.26 for all ASD patients and 0.25 for ASD patients with Mendelian causes. This corresponds to a fair agreement between the different experts. The proportion of ASD patients rated as non-dysmorphic, uncertain, or dysmorphic varied strongly among the experts and ranged between 21% and 76% for non-dysmorphic, between 6% and 36% for uncertain and between 9% and 51% for dysmorphic (**figure 4.5B**). These results show that there are strong differences in the individual expert opinion of facial dysmorphism. The average score across experts, referred to as the average expert dysmorphism score, was considered to reflect clinically assessed dysmorphism in each patient. Comparison of average expert dysmorphism scores of patients with Mendelian causes, susceptibility variants and unknown causes revealed that patients with Mendelian causes had significantly higher average expert dysmorphism scores than patients with unknown causes ($z=4.598$, $p<.001$, AUC=.744) (**figure 4.5C**). Although less pronounced, also patients with susceptibility variants had significantly higher average expert dysmorphism scores than patients with unknown causes ($z=2.294$, $p=.027$, AUC=.690) (**figure 4.5C**). No significant difference was observed between the average expert dysmorphism scores of patients with Mendelian causes and patients with susceptibility variants ($z=0.697$, $p=.309$, AUC=.563) (**figure 4.5C**). A significant increase in the average expert dysmorphism scores was observed for patients with ID as compared to patients without ID ($z=5.400$, $p<.001$, AUC=.779) (**figure 4.5D**). These results indicate that high average expert dysmorphism scores, as a reflection of clinically assessed dysmorphism, were associated with Mendelian causes, susceptibility variants and ID in the ASD cohort.

High computational dysmorphism and asymmetry scores in ASD patients are associated with Mendelian causes

To allow for more objective facial phenotyping, we created computational scores based on 3D images that measure facial dysmorphism and asymmetry. We first tested the computational dysmorphism scores on patients with Noonan syndrome and 22q11.2 deletion and the computational asymmetry scores on patients with oculo-auriculo-vertebral spectrum (**supplementary methods, supplementary figure 4.2A, 4.2B**).

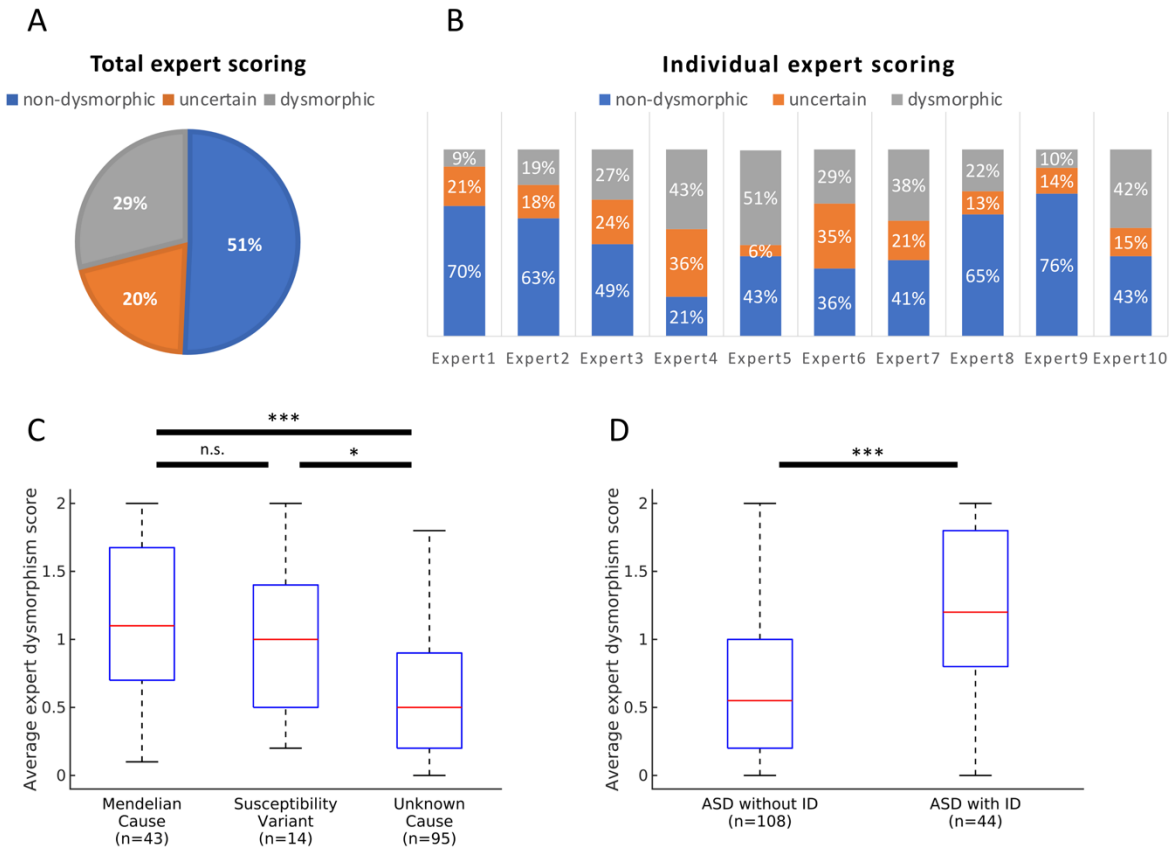


Figure 4.5: Expert scoring of facial dysmorphism in ASD patients. (A-B) Proportions of patients rated non-dysmorphic, uncertain or dysmorphic across all experts (A) and for each individual expert (B). (C-D) Differences in average expert dysmorphism scores between ASD patients of different genetic subgroups (C) or between ASD patients with and without ID (D). *p<.05; **p<.01; ***p<.001; n.s., not significant.

We then compared the distribution of computational dysmorphism and asymmetry scores for ASD patients, unaffected siblings, and unselected controls (**figure 4.6A, 4.6B)**. Both computational dysmorphism and asymmetry scores were significantly elevated in ASD patients as compared to unselected controls (z=4.930, p<.001, AUC=.631; z=4.641, p<.001, AUC=.623 respectively) (**figure 4.6A, 4.6B**) and to unaffected siblings (z=4.139, p<.001 AUC=.655; z=2.041, p=.028, AUC=.576 respectively) (**figure 4.6A, 4.6B**). No significant difference was found between computational dysmorphism or asymmetry scores of unaffected siblings and unselected controls (z=0.428, p=.375, AUC=.514; z=1.700, p=.051, AUC=.554 respectively) (**figure 4.6A, 4.6B**). Subsequent comparison of computational dysmorphism and asymmetry scores of ASD patients with Mendelian causes, susceptibility variants and unknown causes revealed that patients with Mendelian causes had significantly higher computational dysmorphism and asymmetry scores than patients with unknown causes (z=2.542, p=.014, AUC=.636; z=3.802, p<.001, AUC=.703 respectively) (**figure 4.6C, 4.6D**). In addition, patients with susceptibility variants had significantly elevated computational asymmetry scores, but not computational dysmorphism

scores, as compared to patients with unknown causes (z=2.124, p=.028, AUC=.677; z=1.753, p=.058, AUC=.646 respectively) (**figure 4.6C, 4.6D**). There was no significant difference in computational dysmorphism or asymmetry scores between patients with Mendelian causes and susceptibility variants (z=0.213, p=.431, AUC=.520; z=0.362, p=.375, AUC=.533 respectively) (**figure 4.6C, 4.6D**). Patients with ID had significantly elevated computational dysmorphism scores, but not computational asymmetry scores, as compared to patients without ID (z=2.228, p=.029, AUC=.616; z=0.908, p=.241, AUC=.547 respectively) (**figure 4.6E, 4.6F**). In essence, both high computational dysmorphism and asymmetry scores in ASD patients were associated with Mendelian causes, while high computational dysmorphism and asymmetry scores were also associated with ID and susceptibility variants respectively.

<u>The potential of expert and computational scores to discriminate between Mendelian and unknown causes in the ASD cohort</u>

Since high average expert dysmorphism scores, computational dysmorphism scores and computational asymmetry scores were associated with Mendelian causes of ASD, we explored the potential of these scores to discriminate between patients with Mendelian causes and unknown causes in the current cohort. Logistic regression models showed that the computational dysmorphism score and the computational asymmetry score had some discriminative capacity to distinguish between Mendelian causes and unknown causes (average AUC between 0.61-0.70), while the average expert dysmorphism score had fair discriminative capacity (average AUC between 0.71-0.80) (**table 4.3**). However, most individual expert dysmorphism scores only had some discriminative capacity (average AUC between 0.61-0.70) (**table 4.3**). These results show that individual expert dysmorphism scores often do not outperform the computational dysmorphism score or computational asymmetry score in predicting Mendelian ASD in our study cohort. Adding the computational dysmorphism and/or asymmetry score to the individual expert dysmorphism scores improved the discriminative capacity of the individual experts (**table 4.3**). In addition, there was an improvement in the discriminative capacity of the average expert dysmorphism score when adding the asymmetry score (**table 4.3**). These results indicate that computational scores can contribute to the prediction of Mendelian ASD and have added value for this task over expert opinion alone.

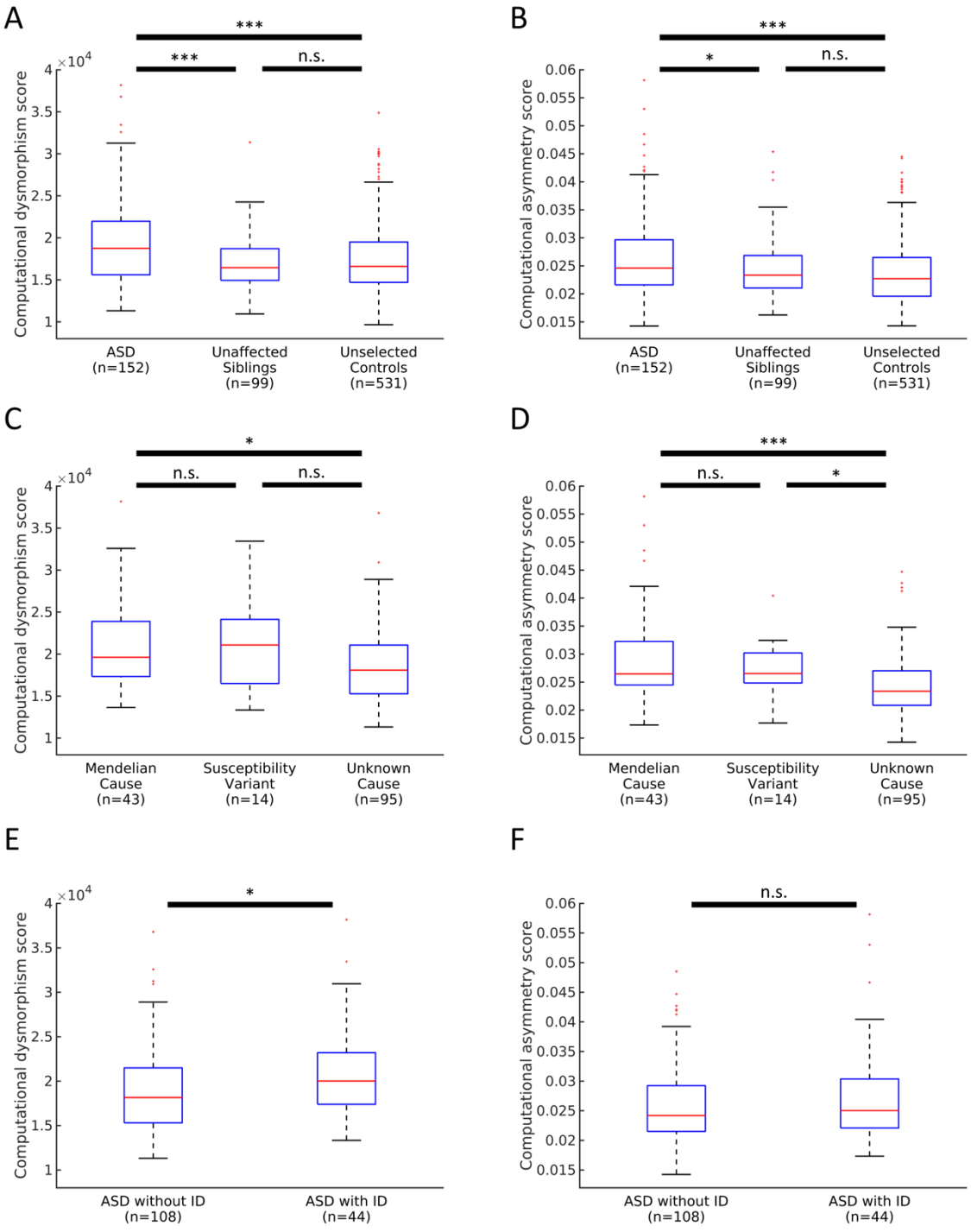


Figure 4.6: Computational scoring of facial dysmorphism and asymmetry in ASD patients. (A-F) Differences in computational dysmorphism or asymmetry scores between ASD patients, unaffected siblings of ASD patients and unselected controls (A-B), between ASD patients of different genetic subgroups (C-D), or between ASD patients with and without ID (E-F). *p<.05; **p<.01; ***p<.001; n.s., not significant.

Table 4.3: The capacity of expert and computational scores to discriminate between Mendelian causes and unknown causes in the ASD cohort.

	Average AUC (minimum AUC – maximum AUC)			
	Self	**With computational dysmorphism score**	**With computational asymmetry score**	**With computational dysmorphism and asymmetry score**
Average expert dysmorphism score	0.744 (0.713-0.798)	0.732 (0.699-0.797)	0.791 (0.788-0.796)	0.784 (0.768-0.794)
Computational dysmorphism score	0.642 (0.570-0.714)	NA	0.695 (0.671-0.723)	NA
Computational asymmetry score	0.7 (0.674-0.735)	0.695 (0.671-0.723)	NA	NA
Expert 1 dysmorphism score	0.626 (0.569-0.667)	0.685 (0.649-0.743)	0.741 (0.719-0.768)	0.737 (0.721-0.746)
Expert 2 dysmorphism score	0.689 (0.625-0.734)	0.722 (0.667-0.783)	0.765 (0.740-0.781)	0.75 (0.720-0.768)
Expert 3 dysmorphism score	0.669 (0.603-0.718)	0.701 (0.611-0.763)	0.759 (0.725-0.781)	0.755 (0.716-0.779)
Expert 4 dysmorphism score	0.643 (0.592-0.742)	0.67 (0.585-0.783)	0.716 (0.663-0.748)	0.714 (0.683-0.748)
Expert 5 dysmorphism score	0.689 (0.667-0.708)	0.723 (0.697-0.759)	0.773 (0.743-0.815)	0.768 (0.741-0.783)
Expert 6 dysmorphism score	0.716 (0.685-0.769)	0.75 (0.705-0.801)	0.805 (0.732-0.893)	0.797 (0.719-0.893)
Expert 7 dysmorphism score	0.586 (0.520-0.644)	0.637 (0.585-0.739)	0.697 (0.634-0.753)	0.7 (0.674-0.714)
Expert 8 dysmorphism score	0.617 (0.505-0.748)	0.657 (0.555-0.759)	0.708 (0.662-0.741)	0.706 (0.643-0.754)
Expert 9 dysmorphism score	0.647 (0.583-0.757)	0.702 (0.645-0.806)	0.777 (0.757-0.799)	0.768 (0.746-0.799)
Expert 10 dysmorphism score	0.653 (0.609-0.714)	0.684 (0.611-0.763)	0.738 (0.730-0.751)	0.738 (0.725-0.754)

Computational dysmorphism scores versus average expert dysmorphism scores in ASD patients with Mendelian causes

To examine how an objective assessment of facial dysmorphism compares to a clinical assessment, we investigated the extent to which computational dysmorphism scores correlated with average expert dysmorphism scores for the ASD patients in our cohort. A significant positive correlation was found between computational and average expert dysmorphism scores (Spearman's ρ=0.45, p<.001) for the whole patient group (**supplementary figure 4.3A**). Next, we determined the correlation between these scores in the subgroup of patients with Mendelian causes. The Spearman correlation between the computational and average expert dysmorphism scores for patients with Mendelian ASD was 0.45 (p=.003) (**figure 4.7A**). We then evaluated discrepancies between these scores for each patient with a Mendelian cause. Individual patients are numbered and represented by their computational and average expert dysmorphism scores in **figure 4.7B** and by their facial signatures in **figure 4.7C**. Quadrants (Q1-Q4) in **figure 4.7** are defined by the median of signature scores of all ASD patients and by an average expert dysmorphism score of "1", indicating uncertainty among experts about the presence of facial dysmorphism. In Q1, six patients (patients 1-6) had high computational dysmorphism scores, while average expert dysmorphism scores were low. Visual inspection of the facial signatures of these patients showed that the deviations from the reference population, as indicated by red and blue colors (outward and inward displacement of points compared to the reference population respectively), were shared in some facial regions across the patients (**figure 4.7C**). Shared regions were the zygomatic arch and malar eminence (inward displacement in patient 1 and outward displacement in patient 2) and buccal area (outward displacement in patient 2 and inward displacement in patient 3). Concordant with this, clinical pictures showed a long narrow face for patient 1 and 3 and a round face for patient 2. No minor facial anomalies were present in any of these patients (patients 1-3). In Q4, five patients (patients 39-43) had low computational dysmorphism scores, while average expert dysmorphism scores were high. Visual inspection of the facial signatures of these five patients revealed deviations from the reference population around the eyes in patients 39 and 40. However, no other shared facial regions with obvious deviations from the reference population were apparent. Clinical pictures showed minor facial anomalies in all five patients. Patient 39 had a small mouth, a short philtrum and hypertelorism; patient 40 had hypertelorism and a unilateral epicanthic fold; patient 41 had anteverted nares; patient 42 had deep set eyes and large ears; and patient 43 had a long and prominent philtrum, a thin lower lip, anteverted nares, hypertelorism and unilateral epicanthic fold.

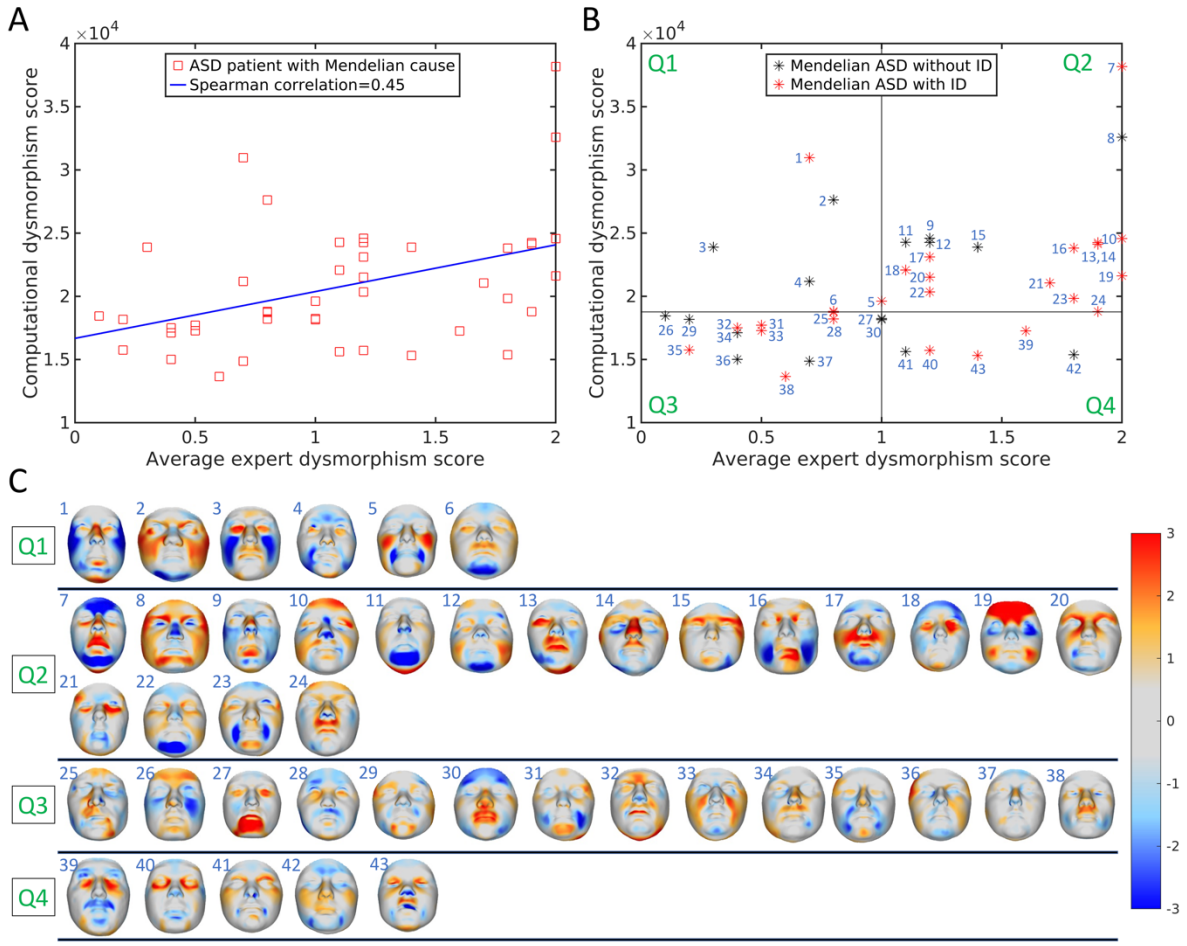


Figure 4.7: Comparison of computational and average expert dysmorphism scores in ASD patients with Mendelian causes. (A) Correlation between computational and average expert dysmorphism scores in ASD patients with Mendelian causes. (B) Comparison of computational and average expert dysmorphism scores in individual ASD patients with Mendelian causes. Individual patients are numbered and divided in quadrants (Q1-Q4). These quadrants are defined by the median of signature scores of all ASD patients and by an expert score of "1", indicating uncertainty among experts about the presence of facial dysmorphism. (C) Facial signatures of ASD patients with Mendelian causes. Signatures are represented as z-scores and visualized as color heat maps. Inward and outward displacements perpendicular to the surface were visualized by blue and red colors respectively. Numbers and quadrants refer to panel B.

4.3.5 Discussion

Here, we studied the utility of 3D facial phenotyping to improve the recognition of ASD patients with Mendelian causes. Evidence from clinical genetics suggests that Mendelian causes for ASD often co-occur with dysmorphic facial features. We confirm this finding using both expert and computational dysmorphism scores. Our findings also suggest, for the first time, that ASD patients with Mendelian causes show elevated facial asymmetry relative to those with unknown causes.

Currently, an assessment of facial dysmorphism by experts is the gold standard in clinical genetics. We found only fair agreement between the dysmorphology assessments of 10 experts, indicating substantial individual differences in expert assessments. Similarly, Lumaka et al. previously reported a fair agreement between experts when scoring dysmorphism in African children with ID [253]. Strong differences across experts were found in the proportion of ASD patients rated as non-dysmorphic, uncertain and dysmorphic, indicating that experts may use different operational definitions of craniofacial dysmorphism and/or different thresholds for considering a face to be dysmorphic. These findings point to the subjectivity of individual expert dysmorphology assessments and demonstrate the need for a more objective measure of facial dysmorphism.

Objective facial phenotyping methods using 3D surface images have shown promise in the past for analyzing individual facial dysmorphism [254–259]. To our knowledge, this is the first study to measure facial dysmorphism using 3D facial images in a large cohort of patients with ASD. Our findings suggest that ASD patients with Mendelian causes show elevated facial dysmorphism. This is not surprising given that minor facial anomalies have been described for many of the Mendelian disorders that were found in the patients in this study. Examples include minor facial anomalies associated with *ARID2* [260], *ADNP* [66], *FOXP1* [261], *PHIP* [262] and *GATAD2B* [263]. Furthermore, we also found that facial dysmorphism in the ASD patients was associated with ID. An obvious explanation is that ASD patients with ID have more complex morphological phenotypes than ASD patients without ID.

Facial asymmetry has been reported to be increased in ASD patients relative to controls [102,104,245]. We detected a significant increase in facial asymmetry in the ASD patients as compared to controls. Previous studies suggested that inherited genetic factors might underlie facial asymmetry in ASD patients since parents of ASD patients also showed increased facial asymmetry [102,264]. Our findings suggest that patients with Mendelian causes show elevated facial asymmetry. However, there is currently no literature evidence that facial asymmetry is associated with Mendelian ASD since facial asymmetry is not routinely assessed during a molecular genetic workup. Nevertheless, our findings contribute to the idea that facial asymmetry in ASD patients is caused by genetic factors.

The computational dysmorphism score and average expert dysmorphism score (as a proxy for clinically assessed dysmorphism) are moderately correlated in general but can give discrepant assessments of individual patients. The discrepancies could be explained by the different relative contribution of minor facial anomalies to the computational versus the expert dysmorphism

score. Minor facial anomalies are often used as an operational definition of clinically assessed dysmorphism and are therefore considered by experts to be of great importance in the assessment of dysmorphism. These minor anomalies however affect only small parts of the face and do not necessarily result in higher computational dysmorphism scores since they do not affect many vertices in the facial signatures (**figure 4.7B, 4.7C**, patients 39-43). In contrast, deviations in global shape affect many vertices in a facial signature and will therefore result in high computational dysmorphism scores. In the absence of minor facial anomalies however, these deviations are less meaningful to an expert and will therefore likely result in low expert dysmorphism scores (**figure 4.7B, 4.7C**, patients 1-3). However, further studies are needed to investigate which aspects of these discrepancies are interesting from a molecular diagnostic perspective and how 3D facial analysis may be optimized accordingly. One approach could be to give more weight to the vertices that are typically affected in signatures of patients with Mendelian conditions.

We also investigated to what extent the two computational scores can complement the expert opinion in discriminating between patients with Mendelian causes and patients with unknown causes in our study cohort. We found that adding the computational dysmorphism score to expert scores improved the discriminative capacity of the individual expert scores but not the average expert score. These results show the added value of objective analysis of dysmorphism in the form of improving individual expert assessments, allowing for a better recognition of patients with Mendelian causes. At the same time, it shows that computational dysmorphism scores require further optimizations. Adding the computational asymmetry scores to expert scores improved the discriminative capacity of the individual expert scores and average expert score. This suggest that an assessment of facial dysmorphism in ASD patients should not be limited to the standard Elements of Morphology [84] but should include an assessment of facial asymmetry as well. It is not known to what extend asymmetry contributes to the clinical impression or assessment of dysmorphism. So far, there is also no standardized way to clinically assess facial asymmetry, while methods to measure facial asymmetry from 3D data are well established. Also in the current study, the expert opinion on facial asymmetry was not investigated. However, this may to some extend be complementary to the assessment of facial dysmorphism in predicting Mendelian conditions.

This study contains the largest collection of 3D images of patients with ASD published to date. Moreover, it provides the first attempt to objectively study facial dysmorphism and facial asymmetry in ASD patients in relation to underlying genetic causes. However, we also recognise the limitations of the current study. The cohort of ASD patients was ascertained through an

outpatient genetics department and thus includes patients for whom genetic counseling and a diagnostic workup was requested. This resulted in a relatively high rate of Mendelian causes in the current ASD cohort. Thus, with this study, we can only draw conclusions about the utility of objective facial phenotyping for ASD patients with a stronger than usual suspicion of a Mendelian cause. In addition, the average expert dysmorphism scores were used as a proxy for clinically assessed facial dysmorphism in the current study. However, this evaluation is not as accurate as the assessment of facial dysmorphism by means of a clinical examination. Furthermore, the expert dysmorphism score was used to evaluate the possible contribution of objective facial analysis to recognize Mendelian ASD. An important note is that in clinical practice many sources of information are used besides the presence of facial dysmorphism to suspect a Mendelian condition. Therefore, our study does not yet allow to draw conclusions on the actual point-of-care use of 3D facial analysis.

Although Mendelian causes of ASD are rare, it is extremely important they are diagnosed in individual patients because they allow for accurate counseling including discussion of risk prediction and reproductive choices. Therefore, recognizing Mendelian causes remains a cornerstone in molecular diagnostics and counseling for ASD. Among other clinical features, craniofacial dysmorphism is an important factor suggesting a Mendelian cause of ASD. Based on our results, we consider 3D facial analysis as a promising method for the objective evaluation of dysmorphism in ASD patients. Ultimately, the clinical expert would benefit from such an analysis by better recognizing dysmorphism. However, further optimizations are needed since the current method may insufficiently differentiate between deviations that are strongly influenced by environmental factors or other sources of phenotypic diversity versus deviations that are strongly determined by genetic variants underlying Mendelian ASD. Future research should focus on whether 3D facial analysis can uncover aspects that are difficult to recognize by the clinical expert but may be important indications of Mendelian ASD. One such aspect may be facial asymmetry. Our results are promising for the development of applications for clinicians who are less experienced in dysmorphology both in clinical diagnostics and in training.

4.3.6 Supplementary information

Supplementary methods

3D facial image acquisition

3D facial surface images were acquired using 3D digital stereophotogrammetry systems: the 3dMDtrio camera system (3dMD, Atlanta, GA, USA) or the Vectra H1 system (Canfield Scientific, Fairfield, NJ, USA). 3D facial images were taken using standard facial image acquisition protocols. Participants who did not maintain a neutral facial expression, participants with poor quality images and participants with facial hair were excluded from the study.

Patient recruitment

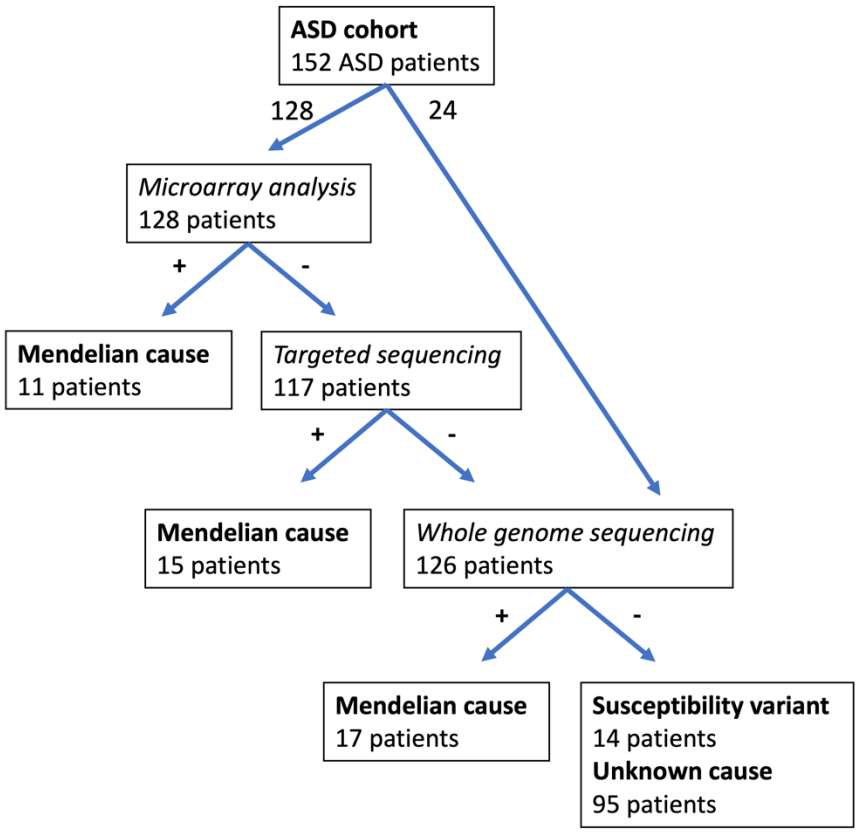
After 3D image data curation described in "3D facial image acquisition", the study sample consisted of 152 ASD patients (124 families), 99 unaffected siblings (69 families), 531 unselected controls (281 families), 49 patients with Noonan syndrome, 67 patients with 22q11.2 deletion syndrome, and 15 patients with oculo-auriculo-vertebral spectrum. ASD patients and their unaffected siblings were recruited through the outpatient department in the Center for Human Genetics (University Hospitals of Leuven, Leuven, Belgium). The patients presented for a molecular diagnostic workup for ASD. All ASD patients and unaffected siblings had been examined by a clinical geneticist and none of the patients had a clinically recognizable genetic syndrome (e.g., Fragile X syndrome, Tuberous Sclerosis). Patients were included if a multi-disciplinary diagnosis of autism was established based on DSM-IV-TR or DSM-V criteria. The latter was done in the Center for Developmental Disorders or in the Expert Center for Autism (University Hospitals of Leuven, Leuven, Belgium). Undiagnosed ASD was ruled out in the unaffected siblings either by behavioral history provided by the parents or by assessment using the Dutch version of the Social Responsiveness Scale. The cohort of unselected controls was recruited through various media channels at the Center for Human Genetics (University Hospitals of Leuven, Leuven, Belgium) and at Technopolis (Flemish Center for Science Communication, Mechelen, Belgium). Unselected controls were unrelated to the ASD patients. The patients with Noonan syndrome, 22q11.2 deletion syndrome and oculo-auriculo-vertebral spectrum were derived from established databases: (1) the FaceBase repository (www.facebase.org, FB00000861); (2) the database of the Western Australian Health Department; and/or (3) Peter Hammond's legacy 3D dysmorphology dataset hosted at KU Leuven, Belgium.

Testing of computational scores

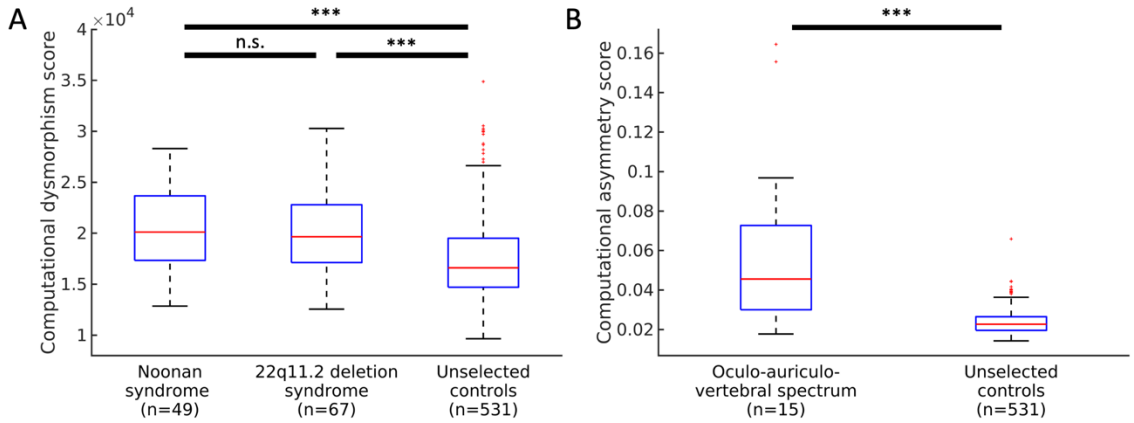
To test the computational scores, we compared the distribution of the computational dysmorphism scores between patients with Noonan syndrome or 22q11.2 deletion syndrome (as positive controls) and unselected controls, and by comparing the distribution of the

computational asymmetry scores between patients with oculo-auriculo-vertebral spectrum (as positive controls) and unselected controls. We found that the computational dysmorphism scores were significantly elevated in patients with Noonan syndrome (z=4.957, p<.001, AUC=.714) and 22q11.2 deletion syndrome (z=5.593, p<.001, AUC=.710) as compared to unselected controls (**supplementary figure 4.2A**). In contrast, no significant difference was found between computational dysmorphism scores of patients with Noonan syndrome and 22q11.2 deletion syndrome (z=0.134, p=.453, AUC=.508) (**supplementary figure 4.2A**). In addition, we found that computational asymmetry scores were significantly elevated in patients with oculo-auriculo-vertebral spectrum as compared to unselected controls (z=4.937, p<.001, AUC=.874) (**supplementary figure 4.2B**).

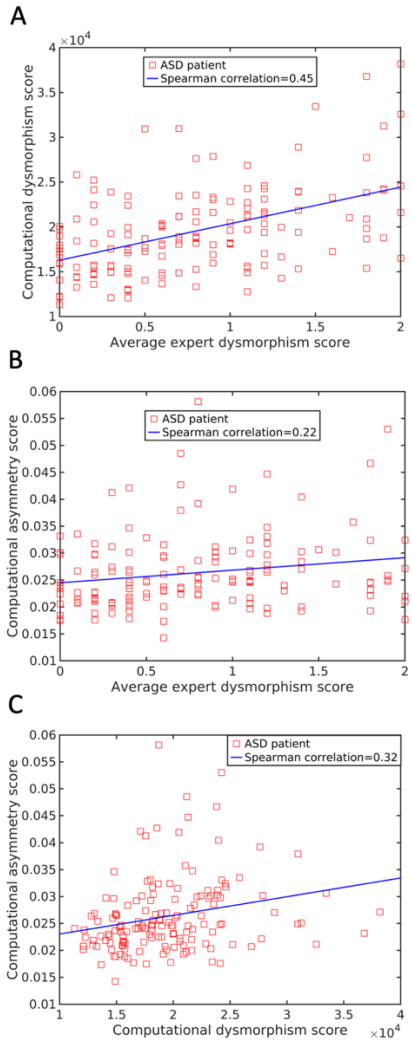
Supplementary figures



Supplementary figure 4.1: Overview genetic testing in ASD patients



Supplementary figure 4.2: Testing of computational scores. (A) Comparison of computational dysmorphism scores between patients with Noonan syndrome, patients with 22q11.2 deletion syndrome, and unselected controls. (B) Comparison of computational asymmetry scores between patients with oculo-auriculo-vertebral spectrum and unselected controls. *p<.05; **p<.01; ***p<.001; n.s., not significant.



Supplementary figure 4.3: Correlations between the expert and computational scores for all ASD patients.

Supplementary tables

Supplementary table 4.1: cohort characteristics				
	Sample size n	**Mean age ± SD (years)**	**Males n (%)**	**ID n (%)**
ASD	152	16.7 ± 5.9	107 (70%)	44 (29%)
ASD with ID	44	15.9 ± 7.1	24 (55%)	44 (100%)
ASD without ID	108	17.0 ± 5.4	83 (77%)	0 (0%)
ASD with Mendelian cause	43	14.8 ± 6.9	27 (63%)	26 (60%)
ASD with susceptibility variant	14	16.8 ± 7.8	7 (50%)	4 (29%)
ASD with unknown cause	95	17.5 ± 5.0	73 (77%)	14 (15%)
US	99	16.9 ± 7.6	35 (35%)	NA
UC	531	9.6 ± 3.6	243 (46%)	NA
NS	49	11 ± 6.4	30 (61%)	NA
22q11.2 DS	67	9.3 ± 4.4	40 (60%)	NA
OAVS	15	13.8 ± 8.3	8 (53%)	NA

- Abbreviations: 22q11.2 DS, 22q11.2 deletion syndrome; ASD, autism spectrum disorder; ID, intellectual disability; NA, not applicable; NS, Noonan syndrome; OAVS, oculo-auriculo-vertebral spectrum; SD, standard deviation; UC, unselected controls; US, unaffected siblings.

- Age: ASD vs US (p=.874); ASD vs UC (p<.001); NS vs UC (p=.122); 22q11.2 DS vs UC (p=.677); OAVS vs UC (p=.070); US vs UC (p<.001); NS vs 22q11.2 DS (p=.114); ASD with ID vs without ID (p=.338); ASD with Mendelian cause vs susceptibility variant (p=.412); ASD with Mendelian cause vs unknown cause (p=.024); ASD with susceptibility variant vs unknown cause (p=.730).

- Sex: ASD vs US (p<.001); ASD vs UC (p<.001); NS vs UC (p=.051); 22q11.2 DS vs UC (p=.037); OAVS vs UC (p=.607); US vs UC (p=.061); NS vs 22q11.2 DS (p=1); ASD with ID vs without ID (p=.010); ASD with Mendelian cause vs susceptibility variant (p= .532); ASD with Mendelian cause vs unknown cause (p=.102); ASD with susceptibility variant vs unknown cause (p=.050).

Supplementary table 4.2: Overview genetic findings in ASD patients					
ID	**Genetic finding**	**Genetic variant (Genomic position hg38)**	**Mode of inheritance**	**Variant type**	**Affected protein-coding gene(s)**
1	Mendelian cause	chr12:13564375-13564376 delCA	*de novo*	Nonsense	*GRIN2B*
2	Mendelian cause	chr7:107739450-116755368 duplication	*de novo*	Duplication	*CBLL1;SLC26A3;DLD;LAMB1;LAMB4;NRCAM;PNPLA8;THAP5;DNAJB9;IMMP2L;LRRN3;DOCK4;ZNF277;IFRD1;LSMEM1;TMEM168;BMT2;GPR85;SMIM30;PPP1R3A;FOXP2;MDFIC;TFEC;TES;CAV2;CAV1;MET*
3	Mendelian cause	chr2:47832601 C>T	*de novo*	Missense	*FBXO11*
4	Mendelian cause	chr5:177167878-177208287	maternal	Intragenic duplication	*NSD1*
5	Mendelian cause	chr19:1175899-1660700 deletion	*de novo*	Deletion	*STK11;CBARP;ATP5F1D;MIDN;CIRBP;FAM174C;EFNA2;PWWP3A;NDUFS7;GAMT;DAZAP1;RPS15;APC2;C19orf25;PCSK4;REEP6;ADAMTSL5;PLK5;MEX3D;MBD3;UQCR11;TCF3*

6	Mendelian cause	chr22:50694813 C>T	*de novo*	Nonsense	*SHANK3*
7	Mendelian cause	chr12:45850500 C>T	*de novo*	Nonsense	*ARID2*
8	Mendelian cause	chr18:51078306 A>G	*de novo*	Missense	*SMAD4*
9	Mendelian cause	chr1:61359275 G>C	*de novo*	Splice-site	*NFIA*
10	Mendelian cause	chr20:50893312 delCT	*de novo*	Frameshift	*ADNP*
11	Mendelian cause	chr14:21402057 T>G	*de novo*	Missense	*CHD8*
12	Mendelian cause	chr17:60062-1757071 deletion	*de novo*	Deletion	*SCGB1C2;DOC2B;RPH3AL;C17orf97;RFLNB;VPS53;TLCD3A;GEMIN4;GLOD4;MRM3;NXN;TIMM22;ABR;BHLHA9;TRARG1;YWHAE;CRK;MYO1C;INPP5K;PITPNA;SLC43A2;SCARF1;RILP;PRPF8;TLCD2;WDR81;SERPINF2*
13	Mendelian cause	chr3:70977852 delTGTC	*de novo*	Frameshift	*FOXP1*
14	Mendelian cause	chr16:89283826 G>A	unknown	Nonsense	*ANKRD11*
15	Mendelian cause	chr1:10027033-12567954 deletion	*de novo*	Deletion	*UBE4B;KIF1B;PGD;CENPS;CORT;DFFA;PEX14;CASZ1;C1orf127;TARDBP;MASP2;SRM;EXOSC10;MTOR;ANGPTL7;UBIAD1;DISP3;FBXO2;FBXO44;FBXO6;MAD2L2;DRAXIN;AGTRAP;C1orf167;MTHFR;CLCN6;NPPA;NPPB;KIAA2013;PLOD1;MFN2;MIIP;TNFRSF8;TNFRSF1B;VPS13D:DHRS3*
16	Mendelian cause	chr15:66436825 C>T	*de novo*	Missense	*MAP2K1*
17	Mendelian cause	chr2:165344558 C>T	*de novo*	Nonsense	*SCN2A*
18	Mendelian cause	chr6:79077729 delC	*de novo*	Frameshift	*PHIP*
19	Mendelian cause	chr1:153812120 G>A	*de novo*	Nonsense	*GATAD2B*
20	Mendelian cause	chr16:67612017 G>A	*de novo*	Missense	*CTCF*
21	Mendelian cause	chr14:100277504 T>G	*de novo*	Nonsense	*YY1*
22	Mendelian cause	chr1:173757077 G>A	*de novo*	Missense	*KLHL20*
23	Mendelian cause	chr7:129283790-132982039 deletion	*de novo*	Deletion	*AHCYL2;STRIP2;SMKR1;NRF1;UBE2H;ZC3HC1;KLHDC10;TMEM209;SSMEM1;CPA2;CPA4;CPA5;CPA1;CEP41;MEST;COPG*

					2;TSGA13;KLF14;MKLN1;PODXL;PLXNA4;CHCHD3
24	Mendelian cause	chr20:487484 G>A	*de novo*	Nonsense	*CSNK2A1*
25	Mendelian cause	chr5:88820588-88837452 deletion	*de novo*	Deletion	*MEF2C*
26	Mendelian cause	chrX:21432638 delA	*de novo*	Frameshift	*CNKSR2*
27	Mendelian cause	chr11:68189990 delT	*de novo*	Frameshift	*KMT5B*
28	Mendelian cause	chr14:21403167 insC	*de novo*	Frameshift	*CHD8*
29	Mendelian cause	chrX:154337943-154450235	maternal	Duplication	*FLNA;EMD;RPL10;DNASE1L1;TAFAZZIN;ATP6AP1;GDI1;FAM50A*
30	Mendelian cause	chr15:92851736-92910792 deletion	*de novo*	Deletion	*CHD2*
31	Mendelian cause	chr14:21408311 C>T	paternal	Splice-site	*CHD8*
32	Mendelian cause	chr2:1943096 C>A	*de novo*	Nonsense	*MYT1L*
33	Mendelian cause	chr16:9849836 G>C	*de novo*	Missense	*GRIN2A*
34	Mendelian cause	chr2:182953211 delC	maternal	Frameshift	*NCKAP1*
35	Mendelian cause	chrX:154337943-154450235	maternal	Duplication	*FLNA;EMD;RPL10;DNASE1L1;TAFAZZIN;ATP6AP1;GDI1;FAM50A*
36	Mendelian cause	chr7:105108997 delAGAC	*de novo*	Frameshift	*KMT2E*
37	Mendelian cause	chr3:196008500-197620100 duplication	*de novo*	Duplication	*TFRC;ZDHHC19;SLC51A;PCYT1A;DYNLT2B;TM4SF19;UBXN7;RNF168;SMCO1;WDR53;FBXO45;NRROS;CEP19;PIGX;PAK2;SENP5;NCBP2;PIGZ;MELTF;DLG1;BDH1*
38	Mendelian cause	chr15:23223829-28402494 duplication	*de novo*	Duplication	*GOLGA8S;GOLGA6L2;MKRN3;MAGEL2;NDN;NPAP1;SNRPN;SNURF;UBE3A;ATP10A;GABRB3;GABRA5;GABRG3;OCA2;HERC2;GOLGA8F*
39	Mendelian cause	chr15:22460684-32223648 duplication	*de novo*	Duplication	*GOLGA6L22;NIPA1;NIPA2;CYFIP1;TUBGCP5;GOLGA6L1;GOLGA8S;GOLGA6L2;MKRN3;MAGEL2;NDN;NPAP1;SNRPN;SNURF;UBE3A;ATP10A;GABRB3;GABRA5;GABRG3;OCA2;HERC2;GOLGA8F;GOLGA8G;GOLGA8M;GOLGA6L7;APBA2;FAM189A1;NSMCE3;TJP1;GOLGA8J;GOLGA8T;CHRFAM7A;GOLGA8R;GOLGA8Q;GOLGA8H;ARHGAP11B;FAN1;MTMR10;TRPM1;KLF13;OTUD7A;CHRNA7*

40	Mendelian cause	chrX:251088-50322053 deletion; chrX:50327117-156005042 duplication	*de novo*	Isodicentric Chromosome X	/
41	Mendelian cause	chr1:151407264 delC	unknown	Frameshift	*POGZ*
42	Mendelian cause	chr11:105933765 G>C	maternal	Missense	*GRIA4*
43	Mendelian cause	chr9:127669957 A>G	*de novo*	Missense	*STXBP1*
/	Susceptibility variant	chr16:29578137-30187604 duplication	paternal	Duplication	*SPN;QPRT;C16orf54;ZG16;KIF22;MAZ;PRRT2;PAGR1;MVP;CDIPT;SEZ6L2;ASPHD1;KCTD13;TMEM219;TAOK2;HIRIP3;INO80E;DOC2A;C16orf92;TLCD3B;ALDOA;PPP4C;TBX6;YPEL3;GDPD3;MAPK3;CORO1A*
/	Susceptibility variant	chr15:22602234-23123157 deletion	paternal	Deletion	*NIPA1;NIPA2;CYFIP1;TUBGCP5*
/	Susceptibility variant	chr22:18948382-21110341 duplication	*de novo*	Duplication	*FAM246C;DGCR2;ESS2;TSSK2;GSC2;SLC25A1;CLTCL1;HIRA;MRPL40;C22orf39;UFD1;UFD1;CDC45;CLDN5;SEPTIN5;SEPT5;GP1BB;TBX1;GNB1L;RTL10;TXNRD2;COMT;ARVCF;TANGO2;DGCR8;TRMT2A;RANBP1;ZDHHC8;CCDC188;RTN4R;DGCR6L;ZNF74;SCARF2;KLHL22;MED15;PI4KA;SERPIND1;SNAP29;CRKL;AIFM3;LZTR1;THAP7;P2RX6;SLC7A4;LRRC74B*
/	Susceptibility variant	chr9:87415877-90413818 deletion	*de novo*	Deletion	*DAPK1;CTSL;SPATA31E1;SPATA31C1;CDK20;SPATA31C2;SPIN1;NXNL2;S1PR3;SHC3;CKS2;SECISBP2;SEMA4D;GADD45G*
/	Susceptibility variant	chr15:22571875-23224922 deletion	paternal	Deletion	*NIPA1;NIPA2;CYFIP1;TUBGCP5;GOLGA6L1*
/	Susceptibility variant	chr2:50226168-50435824 deletion	*de novo*	Deletion	*NRXN1*
/	Susceptibility variant	chr15:22572476-23123147 deletion	maternal	Deletion	*NIPA1;NIPA2;CYFIP1;TUBGCP5*
/	Susceptibility variant	chr7:154280017-154523882 deletion	maternal	Deletion	*DPP6*
/	Susceptibility variant	chr15:30528147-32224181 deletion	*de novo*	Deletion	*GOLGA8Q;GOLGA8H;ARHGAP11B;FAN1;MTMR10;TRPM1;KLF13;OTUD7A;CHRNA7*
/	Susceptibility variant	chr15:22598734-23122386 deletion; chr7:74990016-76034395 duplication	paternal; paternal	Deletion; Duplication	*NIPA1;NIPA2;CYFIP1;TUBGCP5 + CASTOR2;RCC1L;GTF2IRD2B;SPDYE14;SPDYE13;SPDYE15;TRIM73;POM121C;SPDYE5;HIP1;CCL26;CCL24;RHBDD2;POR;TMEM120A;STYXL1*
/	Susceptibility variant	chr15:22571875-23224922 deletion	paternal	Deletion	*NIPA1;NIPA2;CYFIP1;TUBGCP5;GOLGA6L1*

/	Susceptibility variant	chr14:79000439-79639594 deletion	maternal	Deletion	*NRXN3*
/	Susceptibility variant	chr14:79000439-79639594 deletion	maternal	Deletion	*NRXN3*
/	Susceptibility variant	chr15:22707170-23119409 deletion	paternal	Deletion	*NIPA1;NIPA2;CYFIP1;TUBGCP5*

CHAPTER 5

GENERAL DISCUSSION

5 General discussion

5.1 Contributing to knowledge on genotype-phenotype correlations in ASD

Since the advent of NGS technologies, there has been a strong increase in Mendelian gene discovery for NDDs. One strategy to identify novel Mendelian causes involves large-scale international sequencing projects and subsequent rare variant association studies [265]. By providing 1894 DNA samples of ASD patients, we contributed to an international large-scale sequencing effort by the ASID consortium where targeted sequencing of 270 NDD candidate genes was performed in over ten thousand patients with NDDs. Matching genotypes of patients from Leuven with other patients in the consortium allowed to gain knowledge on genotype–phenotype correlations for several NDD risk genes including *NCKAP1*, *SPEN* and *TCF12*. A more traditional strategy to detect novel Mendelian causes are family-based studies [265]. They are often combined with data-sharing approaches to find sufficient patients with the same Mendelian cause [265]. In the current thesis, family-based WGS was performed in 305 ASD patients through either an international large-scale sequencing initiative known as CCDG or through in-house sequencing. Rare variants in potential candidate genes that were identified through analyses of in-house sequencing data were put in matchmaking platforms to find additional patients with variants in these genes. This allowed us to build a cohort of patients with *de novo* variants in *KLHL20* who share a similar NDD that is mainly characterized by ID, epilepsy and ASD. The *KLHL20* variant in the patient from Leuven would not have been found through clinical diagnostic laboratories since they typically focus on analyzing variants in known disease-related genes. This demonstrates the importance of further assessing genetic data in a research context. Extensive collaborations between clinical laboratories and research laboratories are needed to extract all useful information from genetic data, which will ultimately lead to improved molecular diagnostics. Altogether, these findings show that large-scale sequencing, family-based approaches, and data sharing are key strategies in the identification of novel Mendelian causes of ASD. It is expected that thousands of Mendelian disorders remain to be discovered [266]. With the increasing use of WES and WGS in the genetic workup of ASD/NDD, translational research will be needed to ensure that variants in yet unknown ASD/NDD risk genes are identified and that Mendelian gene discovery for NDDs continues to advance.

5.2 Bridging the gap between novel insights about the genetics of ASD and molecular diagnostics in clinical practice

The results of this PhD thesis provided further insight in the genetics of ASD and how this information can be used in clinical practice. Most clinically significant variants identified through analyses of sequencing data occurred *de novo* and were easy to interpret in a context of a sporadic

presentation of ASD and other clinical features in the patients. However, an important observation in this work is that clinically significant variants can also be inherited from parents with mild neurodevelopmental phenotypes. These findings are consistent with the idea that highly penetrant rare variants associated with Mendelian causes of ASD can result in a wide spectrum of neurodevelopmental problems but are often not sufficient to result in ASD in the absence of other risk variants [125]. In line with this, we observed both during the analyses of targeted sequencing data and WGS data that some patients with Mendelian causes present with milder cognitive phenotypes than previously described in literature. Similarly, we described a patient with a microdeletion involving *SIN3A* who had, in contrast to literature cases, an above average cognitive ability. In addition, patients with the recurrent *KLHL20* variant showed varying degrees of ID ranging from mild ID to profound ID. The most obvious explanation for this variability in cognitive functioning associated with Mendelian causes of ASD is that intelligence is a highly polygenic trait [206] and that other variants in the genetic background likely modulate the phenotypic effects of rare variants associated with Mendelian causes of ASD. In addition, unknown environmental factors may also contribute to the variable expressivity of Mendelian causes of ASD. It is important that clinical laboratories take the possibility of clinically significant inherited variants into account during genetic data analysis and interpretation so that these variants are not overlooked. We found that collecting extensive clinical information of patients and parents, and/or performing a multigenerational segregation analysis is a valuable approach for the interpretation of inherited variants. Clinical geneticists should also become more aware that mild neurodevelopmental phenotypes can be associated with Mendelian causes of ASD. Nevertheless, we found that the presence of ID in ASD patients often indicates an underlying Mendelian cause, which is common knowledge in molecular diagnostics for neuropsychiatric disorders.

In the past, Mendelian causes for ASD have predominantly been found in simplex families (i.e. families with only one affected individual) [37,78]. Most clinically significant variants in our patients were also detected in simplex families, while they were rarely found in multiplex families. This is not surprising given that most clinically significant variants occurred *de novo* and thus fitted with the sporadic presentation in simplex families. In multiplex families where a clinically significant variant was found, we observed that the variants did not segregate with ASD, indicating that other genetic variants and/or unknown environmental factors play an important role in the etiology of ASD in multiplex families. We found that clinically significant variants in multiplex families could potentially explain the more severe neurodevelopmental phenotype in affected children who are carrier as compared to their affected siblings who are not carrier. This

is consistent with the idea that multiple hits in ASD/NDD risk genes lead to more severe neurodevelopmental phenotypes [125,143].

Evidence from clinical genetics suggests that Mendelian causes for ASD/NDD often co-occur with dysmorphic facial features. Our study on facial dysmorphism in ASD patients also revealed that facial dysmorphism, either assessed by experienced dysmorphologists or through computational methods, was significantly increased in patients with Mendelian causes of ASD as compared to patients with unknown causes. Furthermore, we illustrated with a case report that a small *SIN3A* microdeletion resulted in facial features that are typically associated with haploinsufficiency of *SIN3A* (i.e. long face, high forehead, pointed chin, low-set ears, long and downslanting palpebral fissures) [241]. However, the patient in this case report did have an above average cognitive ability while most patients described in literature have ID [241]. These results demonstrate that facial dysmorphism is an important indicator for underlying Mendelian causes in ASD patients, even in patients with mild neurodevelopmental problems.

Altogether, the findings in this PhD thesis show that ID, sporadic ASD and dysmorphic facial features are strong indicators for Mendelian causes in ASD patients. Further studies are necessary to determine whether other patient or family characteristics can be used as predictors for underlying Mendelian causes ASD. Neuroimaging studies could potentially reveal additional predictors as patients with Mendelian causes of ASD often have structural brain abnormalities [267]. Better prediction will make sure that patients with Mendelian causes are referred for further genetic testing, which provides opportunities to improve care for these patients.

5.3 Studying objective facial phenotyping as a tool to better recognize Mendelian causes of ASD

Assessments of facial gestalt by clinicians is part of the comprehensive evaluation of the ASD patient prior to genetic testing since it may provide a first clue to an underlying Mendelian cause. However, these assessments tend to be subjective and largely depend on training and experience of the clinicians [268]. In line with this, our study on the assessment of facial dysmorphism in ASD patients demonstrated that there are strong individual differences in dysmorphology assessments between experienced dysmorphologists. We found that these differences are likely the result of the use of different operational definitions of facial dysmorphism and/or the use of different thresholds for considering a face to be dysmorphic. These results showed the necessity to perform more objective facial phenotyping.

We performed objective facial phenotyping in ASD patients since ultimately it might improve the assessment of clinicians and therefore allow for a better recognition of patients with Mendelian causes of ASD. For this purpose, we collected 3D images of 152 ASD patients for which thorough genetic analyses were performed. These 3D images were subjected to morphometric analyses to calculate computational dysmorphism and asymmetry scores in individual patients. The results suggested that the computational dysmorphism scores at least partially detect the same facial anomalies as experienced clinicians. However, additional studies are necessary to see how the computational dysmorphism scores can be improved since they did not detect some minor facial anomalies and were strongly influenced by deviations in global facial shape. One strategy would be to give more weight to facial regions that are most often affected in patients with Mendelian disorders. A second strategy would be to incorporate a dysmorphic feature recognition tool, such as Cliniface [269], in the computational dysmorphism score. A third strategy would be to have the computational dysmorphism score learn from the average expert dysmorphism score. A fourth strategy would be to incorporate familial information in the computational dysmorphism scores so that familial facial traits can be taken into account in the objective analysis of facial dysmorphism. Nevertheless, we found that the current computational dysmorphism scores were already able to improve the assessments of individual clinicians in discriminating between ASD patients with Mendelian causes and unknown causes. Furthermore, we observed that computational asymmetry scores were also able to improve assessments of individual clinicians. Although facial asymmetry is not routinely assessed in a standard dysmorphological work-up, our results show that it could potentially help in recognizing Mendelian causes of ASD. Further studies are however necessary to confirm that facial asymmetry can be used as a predictor for Mendelian causes of ASD. Taken together, the results in this PhD thesis show that objective facial phenotyping is a valuable tool to complement the assessment of facial gestalt by clinicians for the recognition of ASD patients with Mendelian causes.

Predicting a specific Mendelian cause in an ASD patient is very difficult because the phenotypes of many NDDs overlap significantly. A recent evolution is that objective facial phenotyping tools are increasingly being used to predict which Mendelian causes are present in patients with Mendelian disorders [270–274]. Such tools typically measure facial similarity with known genetic syndromes. For example, the DeepGestalt facial phenotyping tool, which powers Face2Gene, quantifies similarities to hundreds of syndromes based on two-dimensional images [273]. Another example is the GestaltMatcher facial phenotyping tool, which uses two-dimensional images of syndromic patients to build a clinical face phenotyping space, in which distances between patients define syndromic similarity [274]. Current tools are mainly focused on the analysis of two-dimensional images and thus cannot be used on the 3D images of the ASD patients.

Nevertheless, the study of the dysmorphology of ASD patients in the context of known syndromes should certainly be explored. It bears further investigation whether similarity of the facial gestalt of ASD patients to known syndromes can help to better distinguish ASD patients with Mendelian causes from multifactorial causes. Combining the latter with our work on computational dysmorphism and asymmetry scores could potentially further improve the recognition of ASD patients with Mendelian causes. Furthermore, these phenotyping tools could potentially aid clinical genetic laboratories by facilitating analysis and interpretation of genetic data [273,274]. For example, they could help with the interpretation of the *SIN3A* microdeletion described in the case report of this thesis since the patient showed typical facial features associated with haploinsufficiency of *SIN3A* [241]. They could help with the interpretation of the genetic data of the ASD patients with Mendelian causes because some showed a facial gestalt that is typically associated with their underlying Mendelian cause. These tools could also help research laboratories with the identification of novel Mendelian disorders by matching patients from unrelated families with a yet unknown Mendelian disorder based on facial gestalt [274]. For example, they could reveal an unknown Mendelian cause in an ASD patient in our cohort by matching the facial gestalt of the ASD patient with the facial gestalts of other unrelated patients with the same unknown Mendelian cause. It is important to mention that not all Mendelian causes of ASD are associated with a typical facial gestalt or even dysmorphic features. For example, most patients with *KLHL20* variants did not present with a consistent facial phenotype. Nevertheless, objective facial phenotyping is expected to become a cornerstone in clinical genetics.

5.4 Molecular diagnostics for ASD: lessons learned

From this work we have also drawn some lessons that should be considered for future molecular diagnostics for patients with ASD. The first lesson is that "autism-specific" genes probably do not exist [43]. Recent studies have attempted to find such genes [52,118,275], but no gene has currently been identified that, when mutated, only confers risk to ASD, and not to other NDDs [43]. Therefore, if genetic testing for ASD is considered, all genes associated with NDDs should be tested. In line with this, we found that the clinically significant variants identified in our studies were in genes associated with a wide variety of neurodevelopmental problems. In addition, we found that ASD patients with Mendelian causes rarely had clinically significant variants in the same gene. We therefore argue against the use of any "ASD-specific" genetic test in molecular diagnostics of ASD. WES and WGS should currently be considered the gold standard for genetic testing in patients with ASD since they allow to detect different types of variants affecting any exonic region in the genome. The second lesson is that most individuals with ASD and their families do not benefit from genetic testing. We believe that genetic testing should currently only be considered in patients with ASD who are suspected to have a Mendelian cause since accurate

genetic counseling can only be given to those patients. We propose the classical clinical genetic approach to decide if a genetic workup is indicated or a referral to a genetics center is needed. The family history and clinical examination are cornerstones of this approach. It may reveal important indications of an underlying Mendelian cause such as (subtle) facial dysmorphism, an outlier phenotype as compared to the parents, or a severe or rather rare phenotype associated with ASD (e.g., ID, epilepsy, developmental regression, early psychosis, …). The third lesson is that reanalysis of genetic data will be an important part of future molecular diagnostics for ASD. The WGS analyses revealed that many clinically significant variants were in genes that had been associated with NDDs in the past years. With the increased use of WES and WGS in molecular diagnostics, it will become possible to re-analyze data at later time-points, allowing for the interpretation of variants in genes that were not associated with NDDs during the initial genetic analyses.

5.5 The role of common variants in future molecular diagnostics

The research in this thesis is mainly focused on better understanding the role of rare variants in the etiology ASD. This is because rare variants have a very large effect on individual ASD risk and are therefore useful in genetic counseling for ASD. However, rare variants only explain a small part of the heritability of ASD. It is expected that most of the heritability of ASD resides in common genetic variation [20]. Common variants are currently not used in molecular diagnostics in patients with ASD because they each only have a very small effect on individual ASD risk. In addition, few common variants have been robustly associated with ASD through GWAS [21]. This is probably due to a lack of statistical power as a result of small sample sizes and poor study designs [23]. Nevertheless, it is important that future studies keep searching for genomic loci harboring common risk variants since this may reveal novel genes or pathways associated with ASD, which will ultimately improve our understanding of the etiology and pathophysiology.

Common risk variants associated with ASD may also be combined into a polygenic risk score (PRS), which are increasingly being used in risk stratification for complex disorders. PRS is a weighted sum of the number of risk variants that an individual carries for a specific condition and provides an estimate of the genetic risk of an individual to that condition [276,277]. Although these scores are not informative for ASD due to lack of common variants identified through GWAS, it is possible that they will be used in ASD risk counseling in the future. There are several ways how PRSs can be applied. First, PRSs could be used to identify children at high genetic risk for ASD (i.e., risk stratification). This subgroup of high-risk individuals may then be enrolled in screening programs, potentially allowing for earlier intervention and better prognosis [278]. However, whether large-scale genetic testing and calculation of ASD risk scores in young children

should be undertaken is up for debate. Second, PRS might also be used to inform on the penetrance and clinical outcome associated with rare variants with large effect sizes [276,278]. For example, high PRS for breast cancer in women with *BRCA1* variants is associated with an earlier age of breast cancer diagnosis [279]. Similarly, high PRS for ASD could potentially better predict the clinical outcome associated with Mendelian causes of ASD. Third, PRSs may contribute to better choice of treatment [276,278]. For example, they might help to decide between pharmacological and non-pharmacological interventions in individuals with psychiatric disorders [276]. However, it remains to be seen how PRS for ASD might influence treatment choice since treatment of ASD is currently mainly focused on non-pharmacological interventions.

Although potentially many more individuals at high risk for ASD will be found by using PRS compared to rare variant analyses, genetic counseling for high PRS will probably be limited. PRSs are not able to accurately predict a diagnosis of ASD or NDDs since they do not capture all genetic variation (e.g. rare variants, SVs), do not consider nongenetic risk factors, usually assume an additive genetic architecture, and do not take epistasis into account [277]. Highly penetrant rare variants give rise to neurodevelopmental problems in most individuals. PRSs are also not informative for family planning since it is impossible to predict which of the common risk variants will be present in the offspring [277]. In contrast, for a rare variant associated with a Mendelian cause for patients with ASD, a more precise probability about the recurrence risk of the variant in the offspring can be given. Nevertheless, PRSs may find their way into molecular diagnostics for ASD since they could be used for risk stratification, refining penetrance or clinical outcome of rare risk variants, and improving treatment options. The first hurdle is however to detect common risk variants associated with ASD. Until this hurdle is not overcome, molecular diagnostics for patients with ASD will continue to focus on the identification of rare risk variants. It is important to mention that environmental factors may also contribute to the etiology of ASD, but little is currently known about these factors.

5.6 Future directions

The advent of NGS has enabled the rapid identification of rare genetic variants across the genome. The challenge for molecular diagnostics has now become the interpretation of these variants and to which extent they contribute to ASD risk. High-throughput in vitro and in vivo functional assays and multi-omics approaches will allow to investigate the functional impact of rare variants at unprecedently rapid manner and at different levels (e.g., RNA, protein, epigenome). At the same time, large-scale phenotyping studies will be needed to get a better insight into the phenotypic effects of rare variants. Finally, databases that integrate different types of individual data will

show how these data are correlated to each other, which may be needed to overcome this next challenge.

BIBLIOGRAPHY

1. Sharma SR, Gonda X, Tarazi FI. Autism Spectrum Disorder: Classification, diagnosis and therapy. Pharmacol Ther. 2018;190:91–104.
2. Kanner L. Autistic Disturbances Of Affective Contact. Nerv Child. 1943;2:217–50.
3. Kanner L. Problems of nosology and psychodynamics of early infantile autism. Am J Orthopsychiatry. 1949;19(3):416–26.
4. Asperger H. Die "Autistischen Psychopathen" im Kindesalter [Autistic Psychopathy in Childhood]. Arch für Psychiatr und Nervenkrankheiten. 1944;117:76–136.
5. Wing L. Asperger's syndrome: A clinical account. Psychol Med. 1981;11(1):115–29.
6. Rosen NE, Lord C, Volkmar FR. The Diagnosis of Autism: From Kanner to DSM-III to DSM-5 and Beyond. J Autism Dev Disord. 2021;51(12):4253–70.
7. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (5th ed.). Washington, DC; 2013.
8. Bougeard C, Picarel-Blanchot F, Schmid R, Campbell R, Buitelaar J. Prevalence of Autism Spectrum Disorder and Co-morbidities in Children and Adolescents: A Systematic Literature Review. Front Psychiatry. 2021;12:744709.
9. Lai MC, Kassee C, Besney R, Bonato S, Hull L, Mandy W, et al. Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. The Lancet Psychiatry. 2019;6(10):819–29.
10. Hollocks MJ, Lerh JW, Magiati I, Meiser-Stedman R, Brugha TS. Anxiety and depression in adults with autism spectrum disorder: A systematic review and meta-analysis. Psychol Med. 2019;49(4):559–72.
11. Pan PY, Bölte S, Kaur P, Jamil S, Jonsson U. Neurological disorders in autism: A systematic review and meta-analysis. Autism. 2021;25(3):812–30.
12. Hyman SL, Levy SE, Myers SM. Identification, Evaluation, and Management of Children With Autism Spectrum Disorder. Pediatrics. 2020;145(1):e20193447.
13. Ronald A, Hoekstra RA. Autism spectrum disorders and autistic traits: A decade of new twin studies. Am J Med Genet Part B Neuropsychiatr Genet. 2011;156(3):255–74.
14. Tick B, Bolton P, Happé F, Rutter M, Rijsdijk F. Heritability of autism spectrum disorders: A meta-analysis of twin studies. J Child Psychol Psychiatry Allied Discip. 2016;57(5):585–95.
15. Risch N, Hoffmann T, Anderson M, Croen L, Grether J, Windham G. Familial recurrence of autism spectrum disorder: evaluating genetic and environmental contributions. Am J Psychiatry. 2014;171(11):1206–13.
16. Hansen SN, Schendel DE, Francis RW, Windham GC, Bresnahan M, Levine SZ, et al. Recurrence Risk of Autism in Siblings and Cousins: A Multinational, Population-Based Study. J Am Acad Child Adolesc Psychiatry. 2019;58(9):866–75.
17. Constantino JN, Zhang Y, Frazier T, Abbacchi AM, Law P. Sibling recurrence and the genetic epidemiology of autism. Am J Psychiatry. 2010;167(11):1349–56.

18. Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A. The familial risk of autism. JAMA - J Am Med Assoc. 2014;311(17):1770–7.
19. Bourgeron T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. Nat Rev Neurosci. 2015;16(9):551–63.
20. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. Nat Genet. 2014;46(8):881–5.
21. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51(3):431–44.
22. Choi L, An JY. Genetic architecture of autism spectrum disorder: Lessons from large-scale genomic studies. Neurosci Biobehav Rev. 2021;128:244–57.
23. De La Torre-Ubieta L, Won H, Stein JL, Geschwind DH. Advancing the understanding of autism disease mechanisms through genetics. Nat Med. 2016;22(4):345–61.
24. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747–53.
25. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet. 2009;10:451–81.
26. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-martin C, Walsh T, et al. Strong Association of De Novo Copy Number Mutations with Autism. Science. 2007;316(5823):445–9.
27. Levy D, Ronemus M, Yamrom B, Lee Y ha, Leotta A, Kendall J, et al. Rare De Novo and Transmitted Copy-Number Variation in Autistic Spectrum Disorders. Neuron. 2011;70(5):886–97.
28. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010;466(7304):368–72.
29. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. Am J Hum Genet. 2014;94(5):677–94.
30. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, 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):863–85.
31. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron. 2015;87(6):1215–33.
32. Leppa VM, Kravitz SN, Martin CL, Andrieux J, Le Caignec C, Martin-Coignard D, et al. Rare Inherited and De Novo CNVs Reveal Complex Contributions to ASD Risk in Multiplex Families. Am J Hum Genet. 2016;99(3):540–54.
33. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet. 2011;43(6):585–9.
34. O ' Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012;485(7397):246–50.
35. O'Roak BJ, Stessman HA, Boyle EA, Witherspoon KT, Martin B, Lee C, et al. Recurrent de novo mutations implicate novel genes underlying simplex autism risk. Nat Commun. 2014;5:5595.

36. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De Novo Gene Disruptions in Children on the Autistic Spectrum. Neuron. 2012;74(2):285–99.
37. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014;515(7526):216–21.
38. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012;485(7397):237–41.
39. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014;515(7526):209–15.
40. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al. Excess of rare, inherited truncating mutations in autism. Nat Genet. 2015;47(6):582–8.
41. Wang T, Guo H, Xiong B, Stessman HAF, Wu H, Coe BP, et al. De novo genic mutations among a Chinese autism spectrum disorder cohort. Nat Commun. 2016;7:13316.
42. Wilfert AB, Turner TN, Murali SC, Hsieh PH, Sulovari A, Wang T, et al. Recent ultra-rare inherited variants implicate new autism candidate risk genes. Nat Genet. 2021;53(8):1125–34.
43. Myers SM, Challman TD, Bernier R, Bourgeron T, Chung WK, Constantino JN, et al. Insufficient Evidence for "Autism-Specific" Genes. Am J Hum Genet. 2020;106(5):587–95.
44. De Rubeis S, Buxbaum JD. Genetics and genomics of autism spectrum disorder: embracing complexity. Hum Mol Genet. 2015;24(R1):R24–31.
45. Walsh R, Tadros R, Bezzina CR. When genetic burden reaches threshold. Eur Heart J. 2020;41(39):3849–55.
46. Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell. 2013;155(5):997–1007.
47. Hormozdiari F, Penn O, Borenstein E, Eichler EE. The discovery of integrated gene networks for autism and related disorders. Genome Res. 2015;25(1):142–54.
48. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell. 2013;155(5):1008–21.
49. Chang J, Gilman SR, Chiang AH, Sanders SJ, Vitkup D. Genotype to phenotype relationships in autism spectrum disorders. Nat Neurosci. 2015;18(2):191–8.
50. Krishnan A, Zhang R, Yao V, Theesfeld CL, Wong AK, Tadych A, et al. Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. Nat Neurosci. 2016;19(11):1454–62.
51. Yuen RKC, Merico D, Bookman M, Howe JL, Thiruvahindrapuram B, Patel R V, et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nat Neurosci. 2017;20(4):602–11.
52. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell. 2020;180(3):568-584.e23.

53. Courchesne E, Pramparo T, Gazestani VH, Lombardo M V., Pierce K, Lewis NE. The ASD Living Biology: from cell proliferation to clinical phenotype. Mol Psychiatry. 2019;24(1):88–107.
54. Courchesne E, Gazestani VH, Lewis NE. Prenatal Origins of ASD: The When, What, and How of ASD Development. Trends Neurosci. 2020;43(5):326–42.
55. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. Am J Hum Genet. 2010;86(5):749–64.
56. Tartaglia NR, Wilson R, Miller JS, Rafalko J, Cordeiro L, Davis S, et al. Autism Spectrum Disorder in Males with Sex Chromosome Aneuploidy: XXY/Klinefelter Syndrome, XYY, and XXYY. J Dev Behav Pediatr. 2017;38(3):197–207.
57. Wassink TH, Piven J, Patil SR. Chromosomal abnormalities in a clinic sample of individuals with autistic disorder. Psychiatr Genet. 2001;11(2):57–63.
58. Shen Y, Dies KA, Holm IA, Bridgemohan C, Sobeih MM, Caronna EB, et al. Clinical genetic testing for patients with autism spectrum disorders. Pediatrics. 2010;125(4):e727-35.
59. Reddy KS. Cytogenetic abnormalities and fragile-x syndrome in Autism Spectrum Disorder. BMC Med Genet. 2005;6:3.
60. Roesser J. Diagnostic yield of genetic testing in children diagnosed with autism spectrum disorders at a regional referral center. Clin Pediatr (Phila). 2011;50(9):834–43.
61. Çöp E, Yurtbaşi P, Öner Ö, Münir KM. Genetic testing in children with autism spectrum disorders. Anadolu Psikiyatr Derg. 2015;16(6):426–32.
62. Fyke W, Velinov M. Fmr1 and autism, an intriguing connection revisited. Genes (Basel). 2021;12(8):1218.
63. Borch LA, Parboosingh J, Thomas MA, Veale P. Re-evaluating the first-tier status of fragile X testing in neurodevelopmental disorders. Genet Med. 2020;22(6):1036–9.
64. Woodbury-Smith M, Scherer SW. Progress in the genetics of autism spectrum disorder. Dev Med Child Neurol. 2018;60(5):445–51.
65. Savatt JM, Myers SM. Genetic Testing in Neurodevelopmental Disorders. Front Pediatr. 2021;9:526779.
66. Helsmoortel C, Vulto-Van Silfhout AT, Coe BP, Vandeweyer G, Rooms L, Van Den Ende J, et al. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. Nat Genet. 2014;46(4):380–4.
67. Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O, et al. Disruptive CHD8 mutations define a subtype of autism early in development. Cell. 2014;158(2):263–76.
68. O'Donnell-Luria AH, Pais LS, Faundes V, Wood JC, Sveden A, Luria V, et al. Heterozygous Variants in KMT2E Cause a Spectrum of Neurodevelopmental Disorders and Epilepsy. Am J Hum Genet. 2019;104(6):1210–22.
69. Blanchet P, Bebin M, Bruet S, Cooper GM, Thompson ML, Duban-Bedu B, et al. MYT1L mutations cause intellectual disability and variable obesity by dysregulating gene expression and development of the neuroendocrine hypothalamus. PLoS Genet. 2017;13(8):e1006957.

70. Stessman HAF, Willemsen MH, Fenckova M, Penn O, Hoischen A, Xiong B, et al. Disruption of POGZ Is Associated with Intellectual Disability and Autism Spectrum Disorders. Am J Hum Genet. 2016;98(3):541–52.
71. Ní Ghrálaigh F, McCarthy E, Murphy DN, Gallagher L, Lopez LM. Brief Report: Evaluating the Diagnostic Yield of Commercial Gene Panels in Autism. J Autism Dev Disord. 2022;
72. Srivastava S, Love-Nichols JA, Dies KA, Ledbetter DH, Martin CL, Chung WK, et al. Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. Genet Med. 2019;21(11):2413–21.
73. Arteche-López A, Rodríguez MJG, Calvin MTS, Quesada-Espinosa JF, Rosales JML, Milla CP, et al. Towards a change in the diagnostic algorithm of autism spectrum disorders: evidence supporting whole exome sequencing as a first-tier test. Genes (Basel). 2021;12(4):560.
74. Sun Y, Ruivenkamp CAL, Hoffer MJV, Vrijenhoek T, Kriek M, van Asperen CJ, et al. Next-Generation Diagnostics: Gene Panel, Exome, or Whole Genome? Hum Mutat. 2015;36(6):648–55.
75. Turner TN, Hormozdiari F, Duyzend MH, McClymont SA, Hook PW, Iossifov I, et al. Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. Am J Hum Genet. 2016;98(1):58–74.
76. Yuen RKC, Thiruvahindrapuram B, Merico D, Walker S, Tammimies K, Hoang N, et al. Whole-genome sequencing of quartet families with autism spectrum disorder. Nat Med. 2015;21(2):185–91.
77. Ruzzo EK, Pérez-Cano L, Jung JY, Wang L kai, Kashef-Haghighi D, Hartl C, et al. Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. Cell. 2019;178(4):850-866.e26.
78. Turner TN, Coe BP, Dickel DE, Hoekzema K, Nelson BJ, Zody MC, et al. Genomic Patterns of De Novo Mutation in Simplex Autism. Cell. 2017;171(3):710-722.e12.
79. Yuen RKC, Merico D, Cao H, Pellecchia G, Alipanahi B, Thiruvahindrapuram B, et al. Genome-wide characteristics of de novo mutations in autism. NPJ Genomic Med. 2016;1:160271–1602710.
80. An JY, Lin K, Zhu L, Werling DM, Dong S, Brand H, et al. Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. Science. 2018;362(6420):eaat6576.
81. Eriksson MA, Liedén A, Westerlund J, Bremer A, Wincent J, Sahlin E, et al. Rare copy number variants are common in young children with autism spectrum disorder. Acta Paediatr Int J Paediatr. 2015;104(6):610–8.
82. Tammimies K, Marshall CR, Walker S, Kaur G, Thiruvahindrapuram B, Lionel AC, et al. Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder. JAMA - J Am Med Assoc. 2015;314(9):595–903.
83. Miles JH. Autism spectrum disorders-A genetics review. Genet Med. 2011;13(4):278–94.
84. Hennekam RC, Biesecker LG, Allanson JE, Hall JG, Opitz JM, Temple IK, et al. Elements of morphology: General terms for congenital anomalies. Am J Med Genet Part A. 2013;161(11):2726–33.
85. Allanson JE, Biesecker LG, Carey JC, Hennekam RCM. Elements of morphology: Introduction. Am J Med Genet Part A. 2009;149(1):2–5.
86. Myers L, Anderlid BM, Nordgren A, Willfors C, Kuja-Halkola R, Tammimies K, et al. Minor physical anomalies in neurodevelopmental disorders: A twin study. Child Adolesc Psychiatry Ment Health. 2017;11:57.

87. Ozgen H, Hop JW, Hox JJ, Beemer FA, Van Engeland H. Minor physical anomalies in autism: A meta-analysis. Mol Psychiatry. 2010;15(3):300–7.
88. Miles JH, Hillman RE. Value of a clinical morphology examination in autism. Am J Med Genet. 2000;91(4):245–53.
89. Miles JH, Takahashi TN, Bagby S, Sahota PK, Vaslow DF, Wang CH, et al. Essential versus complex autism: Definition of fundamental prognostic subtypes. Am J Med Genet. 2005;135(2):171–80.
90. Flor J, Bellando J, Lopez M, Shui A. Developmental functioning and medical Co-morbidity profile of children with complex and essential autism. Autism Res. 2017;10(8):1344–52.
91. Annunziata S, Bulgheroni S, D'Arrigo S, Esposito S, Taddei M, Saletti V, et al. CGH Findings in Children with Complex and Essential Autistic Spectrum Disorder. J Autism Dev Disord. 2021;
92. Lovrečić L, Rajar P, Volk M, Bertok S, Gnidovec Stražišar B, Osredkar D, et al. Diagnostic efficacy and new variants in isolated and complex autism spectrum disorder using molecular karyotyping. J Appl Genet. 2018;59(2):179–85.
93. Fernandez BA, Scherer SW. Syndromic autism spectrum disorders: moving from a clinically defined to a molecularly defined approach. Dialogues Clin Neurosci. 2017;19(4):353–71.
94. Ziats CA, Patterson WG, Friez M. Syndromic Autism Revisited: Review of the Literature and Lessons Learned. Pediatr Neurol. 2021;114:21–5.
95. Hens K, Peeters H, Dierickx K. Genetic testing and counseling in the case of an autism diagnosis: A caregivers perspective. Eur J Med Genet. 2016;59(9):452–8.
96. Hens K, Peeters H, Dierickx K. The ethics of complexity. Genetics and autism, a literature review. Am J Med Genet Part B Neuropsychiatr Genet. 2016;171(3):305–16.
97. Scherer SW, Dawson G. Risk factors for autism: Translating genomic discoveries into diagnostics. Hum Genet. 2011;130(1):123–48.
98. Marcucio R, Hallgrimsson B, Young NM. Facial Morphogenesis. Physical and Molecular Interactions Between the Brain and the Face. Curr Top Dev Biol. 2015;115:299–320.
99. Rothstein M, Bhattacharya D, Simoes-Costa M. The molecular basis of neural crest axial identity. Dev Biol. 2018;444(Suppl 1):S170–80.
100. Le Douarin NM, Dupin E. The Pluripotency of Neural Crest Cells and Their Role in Brain Development. Curr Top Dev Biol. 1st ed. 2016;116:659–78.
101. DEMYER W, ZEMAN W, PALMER CG. THE FACE PREDICTS THE BRAIN: DIAGNOSTIC SIGNIFICANCE OF MEDIAN FACIAL ANOMALIES FOR HOLOPROSENCEPHALY (ARHINENCEPHALY). Pediatrics. 1964;34:256–63.
102. Hammond P, Forster-Gibson C, Chudley A, Allanson J, Hutton T, Farrell S, et al. Face–brain asymmetry in autism spectrum disorders. Mol Psychiatry. 2008;13(6):614–23.
103. Aldridge K, George ID, Cole KK, Austin JR, Takahashi TN, Duan Y, et al. Facial phenotypes in subgroups of prepubertal boys with autism spectrum disorders are correlated with clinical phenotypes. Mol Autism. 2011;2(1):15.
104. Boutrus M, Gilani SZ, Alvares GA, Maybery MT, Tan DW, Mian A, et al. Increased facial asymmetry in autism spectrum conditions is associated with symptom presentation. Autism Res. 2019;12(12):1774–83.

105. Boutrus M, Maybery MT, Alvares GA, Tan DW, Varcin KJ, Whitehouse AJO. Investigating facial phenotype in autism spectrum conditions: The importance of a hypothesis driven approach. Autism Res. 2017;10(12):1910–8.
106. Tan DW, Gilani SZ, Maybery MT, Mian A, Hunt A, Walters M, et al. Hypermasculinised facial morphology in boys and girls with Autism Spectrum Disorder and its association with symptomatology. Sci Rep. 2017;7(1):9348.
107. Walker H. Incidence of minor physical anomaly in autism. J Autism Child Schizophr. 1977;7(2):165–76.
108. Rodier PM, Bryson SE, Welch JP. Minor malformations and physical measurements in autism: Data from Nova Scotia. Teratology. 1997;55(5):319–25.
109. Hardan AY, Keshavan MS, Sreedhar S, Vemulapalli M, Minshew NJ. An MRI study of minor physical anomalies in autism. J Autism Dev Disord. 2006;36(5):607–11.
110. Tripi G, Roux S, Canziani T, Brilhault FB, Barthélémy C, Canziani F. Minor physical anomalies in children with autism spectrum disorder. Early Hum Dev. 2008;84(4):217–23.
111. Cheung C, McAlonan GM, Fung YY, Fung G, Yu KK, Tai KS, et al. MRI study of minor physical anomaly in childhood autism implicates aberrant neurodevelopment in infancy. PLoS One. 2011;6(6):e20246.
112. Ozgen H, Hellemann GS, Stellato RK, Lahuis B, Van Daalen E, Staal WG, et al. Morphological features in children with autism spectrum disorders: A matched case-control study. J Autism Dev Disord. 2011;41(1):23–31.
113. Manouilenko I, Eriksson JM, Humble MB, Bejerot S. Minor Physical Anomalies in Adults with Autism Spectrum Disorder and Healthy Controls. Autism Res Treat. 2014;2014:743482.
114. Hammond P. The use of 3D face shape modelling in dysmorphology. Arch Dis Child. 2007;92(12):1120–6.
115. Hammond P, Suttie M. Large-scale objective phenotyping of 3D facial morphology. Hum Mutat. 2012;33(5):817–25.
116. Stessman HA, Bernier R, Eichler EE. A Genotype-First Approach to Defining the Subtypes of a Complex Disease. Cell. 2014;156(5):872–7.
117. Wang T, Hoekzema K, Vecchio D, Wu H, Sulovari A, Coe BP, et al. Large-scale targeted sequencing identifies risk genes for neurodevelopmental disorders. Nat Commun. 2020;11(1):4932.
118. Stessman HAF, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M, et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nat Genet. 2017;49(4):515–26.
119. Coe BP, Witherspoon K, Rosenfeld JA, Van Bon BWM, Vulto-Van Silfhout AT, Bosco P, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet. 2014;46(10):1063–71.
120. Turner TN, Wilfert AB, Bakken TE, Bernier RA, Pepper MR, Zhang Z, et al. Sex-Based Analysis of De Novo Variants in Neurodevelopmental Disorders. Am J Hum Genet. 2019;105(6):1274–85.
121. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (4th ed., text rev.). Washington, DC; 2000.

122. Eliyahu A, Barel O, Greenbaum L, Zaks Hoffer G, Goldberg Y, Raas-Rothschild A, et al. Refining the Phenotypic Spectrum of KMT5B-Associated Developmental Delay. Front Pediatr. 2022;10:844845.
123. Marzin P, Rondeau S, Aldinger KA, Alessandri JL, Isidor B, Heron D, et al. SETD2 related overgrowth syndrome: Presentation of four new patients and review of the literature. Am J Med Genet Part C Semin Med Genet. 2019;181(4):509–18.
124. Revah-Politi A, Ganapathi M, Bier L, Cho MT, Goldstein DB, Hemati P, et al. Loss-of-function variants in NFIA provide further support that NFIA is a critical gene in 1p32-p31 deletion syndrome: A four patient series. Am J Med Genet Part A. 2017;173(12):3158–64.
125. Guo H, Wang T, Wu H, Long M, Coe BP, Li H, et al. Inherited and multiple de novo mutations in autism/developmental delay risk genes suggest a multifactorial model. Mol Autism. 2018;9:64.
126. Cosemans N, Claes P, Brison N, Vermeesch JR, Peeters H. Noise-robust assessment of SNP array based CNV calls through local noise estimation of log R ratios. Stat Appl Genet Mol Biol. 2018;17(2).
127. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
128. McKenna AH, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.
129. Roller E, Ivakhno S, Lee S, Royce T, Tanner S. Genome analysis Canvas : versatile and scalable detection of copy number variants. Bioinformatics. 2016;32(15):2375–7.
130. Shen JJ, Zhang NR. Change-Point Model on Nonhomogeneous Poisson Processes With Application in Copy Number Profiling by Next-Generation DNA Sequencing. Ann Appl Stat. 2012;6(2):476–96.
131. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
132. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434–43.
133. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405–24.
134. Geoffroy V, Herenger Y, Kress A, Stoetzel C, Piton A, Dollfus H, et al. AnnotSV : an integrated tool for structural variations annotation. Bioinformatics. 2018;34(20):3572–4.
135. Karimzadeh M, Ernst C, Kundaje A, Hoffman M. Umap and Bismap : quantifying genome and methylome mappability. Nucleic Acids Res. 2018;46(20):e120.
136. MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The Database of Genomic Variants: A curated collection of structural variation in the human genome. Nucleic Acids Res. 2014;42(Database issue):D986-992.
137. The 1000 Genomes Project Consortium, Auton A, Brooks L, Durbin R, Garrison E, Kang H, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68–74.
138. Abel HJ, Larson DE, Regier AA, Chiang C, Das I, Kanchi KL, et al. Mapping and characterization of structural variation in 17,795 human genomes. Nature. 2020;583(7814):83–9.

139. Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet Med. 2011;13(7):680–5.
140. Zhang P, Xiao F, Li X, Liang Y, Yi H, Hou M, et al. Familial episodic pain syndrome: a case report and literature review. Ann Transl Med. 2022;10(4):238.
141. Li S, Silvestri V, Leslie G, Rebbeck TR, Neuhausen SL, Hopper JL, et al. Cancer Risks Associated With BRCA1 and BRCA2 Pathogenic Variants. J Clin Oncol. 2022;40(14):1529–41.
142. Tan NB, Stapleton R, Stark Z, Delatycki MB, Yeung A, Hunter MF, et al. Evaluating systematic reanalysis of clinical genomic data in rare disease from single center experience and literature review. Mol Genet Genomic Med. 2020;8(11):e1508.
143. Guo H, Duyzend MH, Coe BP, Baker C, Hoekzema K, Gerdts J, et al. Genome sequencing identifies multiple deleterious variants in autism patients with more severe phenotypes. Genet Med. 2019;21(7):1611–20.
144. Smirnov D, Schlieben LD, Peymani F, Berutti R, Prokisch H. Guidelines for clinical interpretation of variant pathogenicity using RNA phenotypes. Hum Mutat. 2022;43(8):1056–70.
145. Stenton SL, Prokisch H. The Clinical Application of RNA Sequencing in Genetic Diagnosis of Mendelian Disorders. Clin Lab Med. 2020;40(2):121–33.
146. Lee H, Huang AY, Wang L kai, Yoon AJ, Renteria G, Eskin A, et al. Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. Genet Med. 2020;22(3):490–9.
147. Ho SS, Urban AE, Mills RE. Structural variation in the sequencing era. Nat Rev Genet. 2020;21(3):171–89.
148. Rape M. Ubiquitylation at the crossroads of development and disease. Nat Rev Mol Cell Biol. 2018;19(1):59–70.
149. Kawabe H, Stegmüller J. The role of E3 ubiquitin ligases in synapse function in the healthy and diseased brain. Mol Cell Neurosci. 2021;112:103602.
150. Chen HY, Liu CC, Chen RH. Cul3-KLHL20 ubiquitin ligase: physiological functions, stress responses, and disease implications. Cell Div. 2016;11(1):5.
151. Lin MY, Lin YM, Kao TC, Chuang HH, Chen RH. PDZ-RhoGEF ubiquitination by Cullin3-KLHL20 controls neurotrophin-induced neurite outgrowth. J Cell Biol. 2011;193(6):985–994.
152. Wang M, Chen PY, Wang CH, Lai TT, Tsai PI, Cheng YJ, et al. Dbo/Henji Modulates Synaptic dPAK to Gate Glutamate Receptor Abundance and Postsynaptic Response. PLoS Genet. 2016;12(10):e1006362.
153. Fujita Y, Yamashita T. Role of DAPK in neuronal cell death. Apoptosis. 2014;19(2):339–45.
154. Henshall DC, Schindler CK, So NK, Lan JQ, Meller R, Simon RP. Death-Associated Protein Kinase Expression in Human Temporal Lobe Epilepsy. Ann Neurol. 2004;55(4):485–94.
155. Williams S, Hossain M, Mishra S, Gonzalez-Martinez J, Najm I, Ghosh C. Expression and Functional Relevance of Death-Associated Protein Kinase in Human Drug-Resistant Epileptic Brain: Focusing on the Neurovascular Interface. Mol Neurobiol. 2019;56(7):4904–15.
156. Gan CL, Zou Y, Xia Y, Zhang T, Chen D, Lan G, et al. Inhibition of death-associated protein kinase 1 protects against epileptic seizures in mice. Int J Biol Sci. 2021;17(9):2356–66.

157. Sobreira NLM, Arachchi H, Buske OJ, Chong JX, Hutton B, Foreman J, et al. Matchmaker Exchange. Curr Protoc Hum Genet. 2017;95:9.31.1-9.31.15.
158. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: An online force field. Nucleic Acids Res. 2005;33(Web Server issue):W382-388.
159. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, et al. VarSome: the human genomic variant search engine. Bioinformatics. 2019;35(11):1978–80.
160. Chen Z, Picaud S, Filippakopoulos P, D'Angiolella V, Bullock AN. Structural Basis for Recruitment of DAPK1 to the KLHL20 E3 Ligase. Structure. 2019;27(9):1395-1404.e4.
161. Lelieveld SH, Wiel L, Venselaar H, Pfundt R, Vriend G, Veltman JA, et al. Spatial Clustering of de Novo Missense Mutations Identifies Candidate Neurodevelopmental Disorder-Associated Genes. Am J Hum Genet. 2017;101(3):478–84.
162. Zhou Z, Xu C, Chen P, Liu C, Pang S, Yao X, et al. Stability of HIB-Cul3 E3 ligase adaptor HIB Is Regulated by Self-degradation and Availability of Its Substrates. Sci Rep. 2015;5:12709.
163. Tellegen PJ, Laros JA. Snijders-Oomen Niet-verbale intelligentietest III. Nederlands-Duitse normen 2010. [Snijders-Oomen Non-verbal Intelligence Test III. Dutch-German standardization data 2010]. Göttingen, DE: Hogrefe; 2013.
164. de Bildt AA, Kraijer DW. Vineland-Z: Sociale redzaamheidsschaal voor kinderen en jeugdigen met een verstandelijke beperking [Vineland-Z: Adaptive behavior scale for children and youth with intellectual disability]. Leiden, NL: PITS; 2003.
165. Gotham K, Risi S, Pickles A, Lord C. The Autism Diagnostic Observation Schedule: Revised algorithms for improved diagnostic validity. J Autism Dev Disabil. 2007;4(37):613–27.
166. Brereton A V., Tonge BJ, Mackinnon AJ, Einfeld SL. Screening Young People for Autism with the Developmental Behavior Checklist. J Am Acad Child Adolesc Psychiatry. 2002;41(11):1369–75.
167. Einfeld SL, Tonge BJ. Manual for the Developmental Behaviour Checklist, 2nd edn, primary carer version and teacher version. Melbourne, Australia: University of New South Wales and Monash University; 2002.
168. Wechsler D. Wechsler Adult Intelligence Scale (4th ed.). San Antonio, TX: Pearson; 2008.
169. Harrison PL, Oakland T. Adaptive Behavior Assessment System, Third Edition [Manual]. Torrance, CA: Western Psychological Services; 2015.
170. Constantino JN, Gruber CP. Social Responsiveness Scale: Manual. Los Angeles, CA: Western Psychological Services; 2005.
171. Achenbach TM, Rescorla LA. Manual for the ASEBA school-age forms & profiles: an integrated system of multi-informant assessment. Burlington, VT: University of Vermont, Research Center for Children, Youth & Families; 2001.
172. Beery K. The Beery-Buktenica Developmental Test of Visual-Motor Integration: VMI with supplemental developmental tests of visual perception and motor coordination: Administration, scoring and teaching manual. Parsippany, NJ: Modern Curriculum Press; 1997.
173. Hiekkala ME, Vuola P, Artto V, Häppölä P, Häppölä E, Vepsäläinen S, et al. The contribution of CACNA1A, ATP1A2 and SCN1A mutations in hemiplegic migraine: A clinical and genetic study in Finnish migraine families. Cephalalgia. 2018;38(12):1849–63.

174. Sparrow SS, Cicchetti D V, Saulnier CA. Vineland Adaptive Behavior Scales - Third Edition. Bloomington, NM: Pearson; 2016.
175. Stiles J, Jernigan TL. The basics of brain development. Neuropsychol Rev. 2010;20(4):327–48.
176. Simões-Costa M, Bronner ME. Establishing neural crest identity: a gene regulatory recipe. Development. 2015;142(2):242–57.
177. Theveneau E, Mayor R. Neural crest delamination and migration: from epithelium-to-mesenchyme transition to collective cell migration. Dev Biol. 2012;366(1):34–54.
178. Santagati F, Rijli FM. Cranial neural crest and the building of the vertebrate head. Nat Rev Neurosci. 2003;4(10):806–18.
179. Diewert VM. A morphometric analysis of craniofacial growth and changes in spatial relations during secondary palatal development in human embryos and fetuses. Am J Anat. 1983;167(4):495–522.
180. Boughner JC, Wat S, Diewert VM, Young NM, Browder LW, Hallgrímsson B. Short-faced mice and developmental interactions between the brain and the face. J Anat. 2008;213(6):646–62.
181. Lieberman DE, Hallgrímsson B, Liu W, Parsons TE, Jamniczky HA. Spatial packing, cranial base angulation, and craniofacial shape variation in the mammalian skull: testing a new model using mice. J Anat. 2008;212(6):720–35.
182. Hallgrímsson B, Lieberman DE, Liu W, Ford-Hutchinson AF, Jirik FR. Epigenetic interactions and the structure of phenotypic variation in the cranium. Evol Dev. 2007;9(1):76–91.
183. Herring SW, Teng S. Strain in the braincase and its sutures during function. Am J Phys Anthropol. 2000;112(4):575–93.
184. Carver EA, Oram KF, Gridley T. Craniosynostosis in Twist heterozygous mice: a model for Saethre-Chotzen syndrome. Anat Rec. 2002;268(2):90–2.
185. Foppiano S, Hu D, Marcucio RS. Signaling by bone morphogenetic proteins directs formation of an ectodermal signaling center that regulates craniofacial development. Dev Biol. 2007;312(1):103–14.
186. Marcucio RS, Cordero DR, Hu D, Helms JA. Molecular interactions coordinating the development of the forebrain and face. Dev Biol. 2005;284(1):48–61.
187. Chong HJ, Young NM, Hu D, Jeong J, McMahon AP, Hallgrimsson B, et al. Signaling by SHH rescues facial defects following blockade in the brain. Dev Dyn. 2012;241(2):247–56.
188. Kaucka M, Petersen J, Tesarova M, Szarowska B, Kastriti ME, Xie M, et al. Signals from the brain and olfactory epithelium control shaping of the mammalian nasal capsule cartilage. Elife. 2018;7:e34465.
189. Le Douarin NM, Brito JM, Creuzet S. Role of the neural crest in face and brain development. Brain Res Rev. 2007;55(2):237–47.
190. Le Douarin NM, Couly G, Creuzet SE. The neural crest is a powerful regulator of pre-otic brain development. Dev Biol. 2012;366(1):74–82.
191. Aguiar DP, Sghari S, Creuzet S. The facial neural crest controls fore- and midbrain patterning by regulating Foxg1 expression through Smad1 activity. Dev. 2014;141(12):2494–505.

192. Plummer JT, Gordon AJ, Levitt P. The Genetic Intersection of Neurodevelopmental Disorders and Shared Medical Comorbidities - Relations that Translate from Bench to Bedside. Front psychiatry. 2016;7:142.
193. Muenke M, Cohen MMJ. Genetic approaches to understanding brain development: holoprosencephaly as a model. Ment Retard Dev Disabil Res Rev. 2000;6(1):15–21.
194. Muenke M, Beachy PA. Genetics of ventral forebrain development and holoprosencephaly. Curr Opin Genet Dev. 2000;10(3):262–9.
195. Balk K, Biesecker LG. The clinical atlas of Greig cephalopolysyndactyly syndrome. Am J Med Genet A. 2008;146A(5):548–57.
196. Claes P, Roosenboom J, White JD, Swigut T, Sero D, Li J, et al. Genome-wide mapping of global-to-local genetic effects on human facial shape. Nat Genet. 2018;
197. Naqvi S, Sleyp Y, Hoskens H, Indencleef K, Spence JP, Bruffaerts R, et al. Shared heritability of human face and brain shape. Nat Genet. 2021;53(6):830–9.
198. White JD, Indencleef K, Naqvi S, Eller RJ, Hoskens H, Roosenboom J, et al. Insights into the genetic architecture of the human face. Nat Genet. 2021;53(1):45–53.
199. Lam M, Chen CY, Li Z, Martin AR, Bryois J, Ma X, et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. Nat Genet. 2019;51(12):1670–8.
200. Watson HJ, Yilmaz Z, Thornton LM, Hübel C, Coleman JRI, Gaspar HA, et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. Nat Genet. 2019;51(8):1207–14.
201. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat Genet. 2019;51(5):793–803.
202. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019;51(1):63–75.
203. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet. 2019;51(3):404–13.
204. Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat Genet. 2019;51(3):394–403.
205. Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J, et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. Nat Genet. 2018;50(7):920–7.
206. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet. 2018;50(7):912–9.
207. Adams HHH, Hibar DP, Chouraki V, Stein JL, Nyquist PA, Rentería ME, et al. Novel genetic loci underlying human intracranial volume identified through genome-wide association. Nat Neurosci. 2016;19(12):1569–82.

208. Hibar DP, Adams HHH, Jahanshad N, Chauhan G, Stein JL, Hofer E, et al. Novel genetic loci associated with hippocampal volume. Nat Commun. 2017;8:13624.
209. Satizabal CL, Adams HHH, Hibar DP, White CC, Knol MJ, Stein JL, et al. Genetic architecture of subcortical brain structures in 38,851 individuals. Nat Genet. 2019;51(11):1624–36.
210. Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, Kim CE, et al. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. PLoS Genet. 2011;7(9):e1002293.
211. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet. 2010;42(6):508–14.
212. López-Isac E, Acosta-Herrera M, Kerick M, Assassi S, Satpathy AT, Granja J, et al. GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. Nat Commun. 2019;10(1):4955.
213. Jin Y, Andersen G, Yorgov D, Ferrara TM, Ben S, Brownson KM, et al. Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. Nat Genet. 2016;48(11):1418–24.
214. Long HK, Osterwalder M, Welsh IC, Hansen K, Davies JOJ, Liu YE, et al. Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder. Cell Stem Cell. 2020;27(5):765-783.e14.
215. Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat Genet. 2019;51(11):1560–5.
216. Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467(7311):52–8.
217. Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics. 2016;32(2):283–5.
218. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017;45(D1):D896–901.
219. Uz E, Alanay Y, Aktas D, Vargel I, Gucer S, Tuncbilek G, et al. Disruption of ALX1 causes extreme microphthalmia and severe facial clefting: expanding the spectrum of autosomal-recessive ALX-related frontonasal dysplasia. Am J Hum Genet. 2010;86(5):789–96.
220. Kayserili H, Uz E, Niessen C, Vargel I, Alanay Y, Tuncbilek G, et al. ALX4 dysfunction disrupts craniofacial and epidermal development. Hum Mol Genet. 2009;18(22):4357–66.
221. Howard TD, Paznekas WA, Green ED, Chiang LC, Ma N, Ortiz de Luna RI, et al. Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome. Nat Genet. 1997;15(1):36–41.
222. Tassabehji M, Read AP, Newton VE, Patton M, Gruss P, Harris R, et al. Mutations in the PAX3 gene causing Waardenburg syndrome type 1 and type 2. Nat Genet. 1993;3(1):26–30.
223. Zhao F, Weismann CG, Satoda M, Pierpont ME, Sweeney E, Thompson EM, et al. Novel TFAP2B mutations that cause Char syndrome provide a genotype-phenotype correlation. Am J Hum Genet. 2001;69(4):695–703.

224. Marcucio RS, Young NM, Hu D, Hallgrimsson B. Mechanisms that underlie co-variation of the brain and face. Genesis. 2011;49(4):177–89.
225. Ming JE, Kaupas ME, Roessler E, Brunner HG, Golabi M, Tekin M, et al. Mutations in PATCHED-1, the receptor for SONIC HEDGEHOG, are associated with holoprosencephaly. Hum Genet. 2002;110(4):297–301.
226. Garcez RC, Le Douarin NM, Creuzet SE. Combinatorial activity of Six1-2-4 genes in cephalic neural crest cells controls craniofacial and brain development. Cell Mol Life Sci. 2014;71(11):2149–64.
227. Pouget JG, Han B, Wu Y, Mignot E, Ollila HM, Barker J, et al. Cross-disorder analysis of schizophrenia and 19 immune-mediated diseases identifies shared genetic risk. Hum Mol Genet. 2019;28(20):3498–513.
228. Gregory MD, Kippenhan JS, Dickinson D, Carrasco J, Mattay VS, Weinberger DR, et al. Regional Variations in Brain Gyrification Are Associated with General Cognitive Ability in Humans. Curr Biol. 2016;26(10):1301–5.
229. Nagai T, Aruga J, Minowa O, Sugimoto T, Ohno Y, Noda T, et al. Zic2 regulates the kinetics of neurulation. Proc Natl Acad Sci U S A. 2000;97(4):1618–23.
230. Elms P, Siggers P, Napper D, Greenfield A, Arkell R. Zic2 is required for neural crest formation and hindbrain patterning during mouse development. Dev Biol. 2003;264(2):391–406.
231. Inoue T, Ota M, Mikoshiba K, Aruga J. Zic2 and Zic3 synergistically control neurulation and segmentation of paraxial mesoderm in mouse embryo. Dev Biol. 2007;306(2):669–84.
232. Stolt CC, Lommes P, Sock E, Chaboissier MC, Schedl A, Wegner M. The Sox9 transcription factor determines glial fate choice in the developing spinal cord. Genes Dev. 2003;17(13):1677–89.
233. Hu D, Marcucio RS. Neural crest cells pattern the surface cephalic ectoderm during FEZ formation. Dev Dyn. 2012;241(4):732–40.
234. Reynolds CR, Kamphaus RW. Behavior assessment system for children, 2nd edition (BASC-2). 2nd ed. Circle Pines, MN: American Guidance Service; 2004.
235. Harrison PL, Oakland T. Adaptive Behavior Assessment – Second Edition Manual (ABAS-II). 2nd ed. San Antonio, TX: Harcourt Assessment; 2003.
236. Constantino JN, Gruber CP, Torrance CA, Marrus N, Glowinski AL, Jacob T, et al. Social Responsiveness Scale-Second Edition (SRS-2). 2nd ed. Los Angeles, CA: Western Psychological Services; 2012.
237. Dunn W, Rietman A. Sensory Profile-NL. Amsterdam: Harcourt Test Publishers; 2006.
238. Wechsler D. Wechsler Intelligence Scale for Children - Fourth Edition (WISC-IV). San Antonio, TX: The Psychological Corporation; 2003.
239. Woodcock RW, McGrew KS, Mather N. Woodcock-Johnson III tests of achievement. Rolling Meadows, IL: Riverside Publishing; 2007.
240. Beery KE, Buktenica NA, Beery NA. BEERY VMI-6: Beery-Buktenica Developmental Test of Visual-Motor Integration–Sixth Edition. Pearson; 2010.

241. Witteveen JS, Willemsen MH, Dombroski TCD, Van Bakel NHM, Nillesen WM, Van Hulten JA, et al. Haploinsufficiency of MeCP2-interacting transcriptional co-repressor SIN3A causes mild intellectual disability by affecting the development of cortical integrity. Nat Genet. 2016;48(8):877–87.
242. Mefford HC, Rosenfeld JA, Shur N, Slavotinek AM, Cox VA, Hennekam RC, et al. Further clinical and molecular delineation of the 15q24 microdeletion syndrome. J Med Genet. 2012;49(2):110–8.
243. Li D, Bodamer OA, Fan Y shan. A 420 Kb Deletion within the Minimum Critical Region of the 15q24 Microdeletion Syndrome in a Female Infant. North Am J Med Sci. 2014;7(4):171–5.
244. Narumi-Kishimoto Y, Araki N, Migita O, Kawai T, Okamura K, Nakabayashi K, et al. Novel SIN3A mutation identified in a Japanese patient with Witteveen-Kolk syndrome. Eur J Med Genet. 2018;62(9):103547.
245. Ozgen H, Hellemann GS, De Jonge M V., Beemer FA, Van Engeland H. Predictive value of morphological features in patients with autism versus normal controls. J Autism Dev Disord. 2013;43(1):147–55.
246. White JD, Ortega-Castrillón A, Matthews H, Zaidi AA, Ekrami O, Snyders J, et al. MeshMonk: Open-source large-scale intensive 3D phenotyping. Sci Rep. 2019;9(1):6085.
247. Matthews HS, Palmer RL, Baynam GS, Quarrell OW, Klein OD, Spritz RA, et al. Large-scale open-source three-dimensional growth curves for clinical facial assessment and objective description of facial dysmorphism. Sci Rep. 2021;11(1):12175.
248. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977;33(1):159–74.
249. Ekrami O, Claes P, White JD, Zaidi AA, Shriver MD, Van Dongen S. Measuring asymmetry from high-density 3D surface scans: An application to human faces. PLoS One. 2018;13(12):e0207895.
250. Klingenberg CP, Barluenga M, Meyer A. Shape analysis of symmetric structures: Quantifying variation among individuals and asymmetry. Evolution (N Y). 2002;56(10):1909–20.
251. Benjamini Y, Hochberg Y, Benjamini, Yoav HY. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995;57(1):289–300.
252. Ruscio J, Mullen T. Confidence Intervals for the Probability of Superiority Effect Size Measure and the Area Under a Receiver Operating Characteristic Curve. Multivariate Behav Res. 2012 Mar 30;47(2):201–23.
253. Lumaka A, Cosemans N, Lulebo Mampasi A, Mubungu G, Mvuama N, Lubala T, et al. Facial dysmorphism is influenced by ethnic background of the patient and of the evaluator. Clin Genet. 2017;92(2):166–71.
254. Hammond P, Suttie M, Hennekam RC, Allanson J, Shore EM, Kaplan FS. The Face Signature of Fibrodysplasia Ossificans Progressiva. Am J Med Genet Part A. 2012;158A(6):1368–80.
255. Suttie M, Foroud T, Wetherill L, Jacobson JL, Molteno CD, Meintjes EM, et al. Facial dysmorphism across the fetal alcohol spectrum. Pediatrics. 2013;
256. Chinthapalli K, Bartolini E, Novy J, Suttie M, Marini C, Falchi M, et al. Atypical face shape and genomic structural variants in epilepsy. Brain. 2012;

257. Hopman SMJ, Merks JHM, Suttie M, Hennekam RCM, Hammond P. 3D morphometry aids facial analysis of individuals with a childhood cancer. Am J Med Genet Part A. 2016;170(11):2905–15.
258. Roessler HI, Shields K, Grange DK, Knoers NVAM, van Haaften G, Hammond P, et al. Three-dimensional facial morphology in Cantú syndrome. Am J Med Genet Part A. 2020;182(5):1041–52.
259. Postema FAM, Matthews H, Hopman SMJ, Merks JHM, Suttie M, Hoskens H, et al. 3D analysis of facial morphology in Dutch children with cancer. Comput Methods Programs Biomed [Internet]. 2021;205:106093.
260. Shang L, Cho MT, Retterer K, Folk L, Humberson J, Rohena L, et al. Mutations in ARID2 are associated with intellectual disabilities. Neurogenetics. 2015;16(4):307–14.
261. Lozano R, Gbekie C, Siper PM, Srivastava S, Saland JM, Sethuram S, et al. FOXP1 syndrome: a review of the literature and practice parameters for medical assessment and monitoring. J Neurodev Disord. 2021;13(1):18.
262. Jansen S, Hoischen A, Coe BP, Carvill GL, Van Esch H, Bosch DGM, et al. A genotype-first approach identifies an intellectual disability-overweight syndrome caused by PHIP haploinsufficiency. Eur J Hum Genet. 2018;26(1):54–63.
263. Willemsen MH, Nijhof B, Fenckova M, Nillesen WM, Bongers EMHF, Castells-Nobau A, et al. GATAD2B loss-of-function mutations cause a recognisable syndrome with intellectual disability and are associated with learning deficits and synaptic undergrowth in Drosophila. J Med Genet. 2013;50(8):507–14.
264. Tan DW, Gilani SZ, Boutrus M, Alvares GA, Whitehouse AJO, Mian A, et al. Facial asymmetry in parents of children on the autism spectrum. Autism Res. 2021;14(11):2260–9.
265. Seaby EG, Rehm HL, O'Donnell-Luria A. Strategies to Uplift Novel Mendelian Gene Discovery for Improved Clinical Outcomes. Front Genet. 2021;12:674295.
266. Bamshad MJ, Nickerson DA, Chong JX. Mendelian Gene Discovery: Fast and Furious with No End in Sight. Am J Hum Genet. 2019;105(3):448–55.
267. Frewer V, Gilchrist CP, Collins SE, Williams K, Seal ML, Leventer RJ, et al. A systematic review of brain MRI findings in monogenic disorders strongly associated with autism spectrum disorder. J Child Psychol Psychiatry Allied Discip. 2021;62(11):1339–52.
268. Myers L, Anderlid BM, Nordgren A, Lundin K, Kuja-Halkola R, Tammimies K, et al. Clinical versus automated assessments of morphological variants in twins with and without neurodevelopmental disorders. Am J Med Genet Part A. 2020;182(5):1177–89.
269. Palmer RL, Helmholz P, Baynam G. CLINIFACE: PHENOTYPIC VISUALISATION AND ANALYSIS USING NON-RIGID REGISTRATION of 3D FACIAL IMAGES. Int Arch Photogramm Remote Sens Spat Inf Sci - ISPRS Arch. 2020;43(B2):301–8.
270. Ferry Q, Steinberg J, Webber C, FitzPatrick DR, Ponting CP, Zisserman A, et al. Diagnostically relevant facial gestalt information from ordinary photos. Elife. 2014;3:e02020.
271. van der Donk R, Jansen S, Schuurs-Hoeijmakers JHM, Koolen DA, Goltstein LCMJ, Hoischen A, et al. Next-generation phenotyping using computer vision algorithms in rare genomic neurodevelopmental disorders. Genet Med. 2019;21(8):1719–25.

272. Dudding-Byth T, Baxter A, Holliday EG, Hackett A, O'Donnell S, White SM, et al. Computer face-matching technology using two-dimensional photographs accurately matches the facial gestalt of unrelated individuals with the same syndromic form of intellectual disability. BMC Biotechnol. 2017;17(1):90.
273. Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. Nat Med. 2019;25(1):60–4.
274. Hsieh TC, Bar-Haim A, Moosa S, Ehmke N, Gripp KW, Pantel JT, et al. GestaltMatcher facilitates rare disease matching using facial phenotype descriptors. Nat Genet. 2022;54(3):349–57.
275. Coe BP, Stessman HAF, Sulovari A, Geisheker MR, Bakken TE, Lake AM, et al. Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. Nat Genet. 2019;51(1):106–16.
276. Lewis CM, Vassos E. Polygenic risk scores: From research tools to clinical instruments. Genome Med. 2020;12(1):44.
277. Wray NR, Lin T, Austin J, McGrath JJ, Hickie IB, Murray GK, et al. From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. JAMA Psychiatry. 2021;78(1):101–9.
278. Murray GK, Lin T, Austin J, McGrath JJ, Hickie IB, Wray NR. Could Polygenic Risk Scores Be Useful in Psychiatry?: A Review. JAMA Psychiatry. 2021;78(2):210–9.
279. Kuchenbaecker KB, McGuffog L, Barrowdale D, Lee A, Soucy P, Dennis J, et al. Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. J Natl Cancer Inst. 2017;109(7):djw302.

SCIENTIFIC SUMMARY

Despite considerable progress in the genetics of autism spectrum disorder (ASD) in the past decade, molecular diagnostics for patients with ASD remains a great challenge due to the complex genetic architecture of ASD. In clinical genetics, an etiological distinction is usually made between ASD patients with Mendelian causes (i.e., driven by a highly penetrant rare variant) and patients with multifactorial causes (i.e., driven by many genetic variants and environmental factors). Although Mendelian causes for ASD are rare, they are extremely important to diagnose in individual patients, as accurate genetic counseling is only possible for these causes. Therefore, the identification of highly penetrant rare variants contributing to Mendelian causes and the recognition of clinical features associated with Mendelian causes have become cornerstones in molecular diagnostics for patients with ASD. The work in this thesis is focused on improving molecular diagnostics for patients with ASD by contributing to knowledge on genotype-phenotype correlations in ASD, by bridging the gap between novel insights about the genetics of ASD and molecular diagnostics in clinical practice and by studying objective facial phenotyping as a tool to better recognize Mendelian causes of ASD.

With the decreasing costs of next generation sequencing (NGS), large-scale sequencing and subsequent rare variant association analyses are increasingly done to find novel risk genes for ASD and other neurodevelopmental disorders (NDDs). We participated in an international large-scale sequencing effort where targeted sequencing of 270 NDD candidate genes in over ten thousand patients with NDDs, including 1894 ASD patients from Leuven, was performed. Our participation contributed to the rare variant association analyses, which provided evidence for the association of 148 genes to NDDs. Furthermore, the evaluation of the clinical significance of rare variants in candidate genes in individual patients from Leuven contributed to knowledge on genotype-phenotype correlations for *NCKAP1*, *SPEN* and *TCF12*. We also participated in a large-scale sequencing initiative where whole exome sequencing or whole genome sequencing (WGS) was performed for twenty-five thousand patients with ASD including family-based WGS for 242 ASD patients from Leuven. At the same time, NGS is also increasingly being used in the genetic workup in patients with ASD. To that end, in-house family-based WGS in 63 ASD patients from Leuven was done. Analyses of WGS data revealed several clinically significant variants in known NDD risk genes that were previously missed by chromosomal microarrays or gene panels. Furthermore, the WGS analyses revealed clinically significant variants outside known NDD risk genes, one of which allowed for the delineation of a novel NDD. This novel genetic disorder is caused by *de novo* missense variants in *KLHL20* and is mainly characterized by mild to severe intellectual disability (ID), febrile seizures or epilepsy, ASD and hyperactivity. The genetic findings also provided useful insight in the rare variants and the clinical features that are

associated with Mendelian causes of ASD. The results in this thesis showed that these rare variants mainly occur *de novo* and are typically found in ASD patients with ID and sporadic ASD. However, we also illustrated that these variants can be inherited from parents with mild neurodevelopmental phenotypes, frequently occur in patients with low-to-average cognitive abilities and in rare cases can even be found in families with multiple individuals with ASD.

A close developmental relationship between the brain and face is evident from the co-occurrence of neurodevelopmental problems, structural brain anomalies and dysmorphic facial features in patients with NDDs. During a side project in this thesis, we investigated to which extent this relationship extends to common human genetic variation by comparing genome-wide association study data of human brain and face shape. This comparison revealed 76 overlapping loci including transcription factors involved in craniofacial development, as well as members of signaling pathways implicated in brain–face cross-talk. Evidence from clinical genetics suggests that the presence of dysmorphic facial features is a strong indicator for an underlying Mendelian cause in patients with ASD. We illustrate with a case study the importance of a facial dysmorphology assessment and deep phenotyping in children with a NDD, even in the absence of ID. The opinion of dysmorphologists is currently considered the gold standard to assess facial dysmorphism, but this opinion is often subject to examiner bias and is highly dependent on training and clinical experience. Therefore, we collected three-dimensional (3D) facial images of 152 patients with ASD to perform an objective analysis of facial dysmorphism. Univariate computational dysmorphism and asymmetry scores indexing unusual 3D facial development were calculated based on the 3D facial shape of ASD patients and were investigated in relation to the presence of a Mendelian cause. We found that both the computational dysmorphism scores and asymmetry scores were significantly increased in patients with Mendelian causes as compared to patient with unknown causes. Furthermore, we showed that these computational scores improved the assessment of facial dysmorphism by individual experts and thereby allowed for better recognition of Mendelian causes of ASD.

WETENSCHAPPELIJKE SAMENVATTING

Ondanks aanzienlijke vooruitgang in de genetica van autismespectrumstoornissen (ASS) in het afgelopen decennium, blijft moleculaire diagnostiek voor patiënten met ASS een grote uitdaging vanwege de complexe genetische architectuur van ASS. In de klinische genetica wordt meestal een etiologisch onderscheid gemaakt tussen ASS patiënten met Mendeliaanse oorzaken (d.w.z. gedreven door een zeer penetrante zeldzame variant) en patiënten met multifactoriële oorzaken (d.w.z. gedreven door vele genetische varianten en omgevingsfactoren). Hoewel Mendeliaanse oorzaken voor ASS zeldzaam zijn, zijn ze uiterst belangrijk om bij individuele patiënten te diagnosticeren, omdat alleen voor deze oorzaken een goede genetische counseling mogelijk is. De identificatie van zeer penetrante zeldzame varianten die bijdragen aan Mendeliaanse oorzaken en de herkenning van klinische kenmerken geassocieerd met Mendeliaanse oorzaken zijn daarom hoekstenen geworden van de moleculaire diagnostiek voor patiënten met ASS. Het werk in dit proefschrift is gericht op het verbeteren van de moleculaire diagnostiek voor patiënten met ASS door bij te dragen aan de kennis over genotype-fenotype correlaties in ASS, door het overbruggen van de kloof tussen nieuwe inzichten over de genetica van ASS en de moleculaire diagnostiek in de klinische praktijk en door het bestuderen van objectieve gelaatsfenotypering als een hulpmiddel om Mendeliaanse oorzaken van ASS beter te herkennen.

Dankzij de dalende kosten van next generation sequencing (NGS) worden grootschalige sequencingprojecten en zeldzame variant associatie analyses steeds meer uitgevoerd om nieuwe risicogenen voor ASS en andere ontwikkelingsstoornissen (OSs) te vinden. Wij namen deel aan een grootschalig internationaal project waarbij gerichte sequenering van 270 kandidaatgenen voor OSs in meer dan tienduizend patiënten met OSs, waaronder 1894 ASS patiënten uit Leuven, werd uitgevoerd. Onze deelname droeg bij tot de zeldzame variant associatie analyses, die bewijs leverden voor de associatie van 148 genen met OSs. Bovendien droeg het bestuderen van de klinische betekenis van deze zeldzame varianten in individuele patiënten uit Leuven bij tot kennis over genotype-fenotype correlaties voor *NCKAP1*, *SPEN* en *TCF12*. We hebben daarnaast ook deelgenomen aan een grootschalig project waarbij whole exome sequencing of whole genome sequencing (WGS) werd uitgevoerd voor vijfentwintigduizend patiënten met ASS, inclusief familie-gebaseerde WGS voor 242 Leuvense ASS patiënten. Tegelijkertijd wordt NGS ook steeds meer gebruikt in de genetische diagnostiek voor patiënten met ASS. Om die reden werd in-house familie-gebaseerde WGS uitgevoerd bij 63 ASS patiënten uit Leuven. Analyses van de WGS data brachten verschillende klinisch significante varianten in gekende risicogenen voor OSs aan het licht die voorheen gemist werden door chromosomale microarrays of genpanels. Bovendien brachten de WGS analyses klinisch significante varianten aan het licht buiten de gekende risicogenen voor OSs, waarvan er één toeliet om een nieuwe OS te beschreven. Deze nieuwe

genetische aandoening wordt veroorzaakt door *de novo* missense varianten in *KLHL20* en wordt voornamelijk gekenmerkt door milde tot ernstige verstandelijke beperking, koortsstuipen of epilepsie, ASS en hyperactiviteit. De genetische bevindingen verschaften ook inzicht in de zeldzame varianten en de klinische kenmerken die geassocieerd zijn met Mendeliaanse oorzaken van ASS. De resultaten in dit proefschrift toonden aan dat deze varianten voornamelijk *de novo* ontstaan en typisch worden gevonden bij ASS patiënten met verstandelijke beperking en een sporadische vorm van ASS. We illustreerden echter ook dat de varianten overgeërfd kunnen worden van ouders met mildere ontwikkelingsproblemen, frequent voorkomen bij patiënten met lage tot gemiddelde cognitieve mogelijkheden en in zeldzame gevallen zelfs gevonden kunnen worden in families met meerdere personen met ASS.

Een nauwe ontwikkelingsrelatie tussen de hersenen en het gelaat blijkt uit het samen voorkomen van ontwikkelingsproblemen, structurele hersenafwijkingen en dysmorfe gelaatskenmerken bij patiënten met OSs. In dit proefschrift hebben we onderzocht in welke mate deze relatie betrekking heeft tot veelvoorkomende menselijke genetische variatie door genoombrede associatestudie data van de menselijke hersenvorm en gelaatsvorm te vergelijken. Deze vergelijking onthulde 76 overlappende loci waaronder transcriptiefactoren die betrokken zijn bij craniofaciale ontwikkeling, evenals leden van signalisatie pathways die betrokken zijn bij hersen-gelaat crosstalk. Bewijs uit de klinische genetica suggereert dat de aanwezigheid van dysmorfe gelaatskenmerken een sterke indicator is voor een onderliggende Mendeliaanse oorzaak bij patiënten met ASS. Wij illustreerden met een casestudy het belang van een evaluatie van het gelaat en een diepgaande fenotypering bij kinderen met een OS, zelfs als deze geen verstandelijke beperking hebben. De opinie van dysmorfologen wordt momenteel beschouwd als de gouden standaard om dysmorfie in het gelaat te beoordelen, maar deze opinie verschilt sterk van persoon tot persoon en is sterk afhankelijk van training en ervaring. Daarom verzamelden wij driedimensionale (3D) gelaatsbeelden van 152 patiënten met ASS om een objectieve analyse van dysmorfie in het gelaat uit te voeren. Computationele dysmorfie en asymmetrie scores, die een abnormale gelaatsontwikkeling weerspiegelen, werden berekend op basis van de 3D gezichtsvorm van ASS patiënten en werden onderzocht in relatie tot de aanwezigheid van een Mendeliaanse oorzaak. Wij vonden dat zowel de computationele dysmorfie scores als de asymmetrie scores significant verhoogd waren bij patiënten met Mendeliaanse oorzaken in vergelijking met patiënten met nog onbekende oorzaken. Bovendien toonden we aan dat deze computationele scores de beoordeling van dysmorfie door individuele experts verbeterden en daardoor een betere herkenning van Mendeliaanse oorzaken van ASS mogelijk maakten.

ACKNOWLEDGEMENTS, PERSONAL CONTRIBUTION AND CONFLICT OF INTEREST STATEMENT

Scientific acknowledgements

First, we would like to thank all the patients and their families for participating in the projects of this thesis. We also acknowledge the clinicians of the Center for Developmental Disorders, the Expert Center for Autism and the Center for Human Genetics at the University Hospitals of Leuven for consultation, recruitment and further follow-up of the patients. Furthermore, we wish to thank the diagnostic laboratories of the Leuven Center for Human Genetics, the Genomics Core Leuven, the ASID consortium and the New York Genome Center for performing DNA extraction and sequencing of DNA samples. Finally, we thank all contributing authors for reviewing the manuscripts in this thesis.

Hilde Peeters is a senior clinical investigator of The Research Foundation – Flanders. The work in this thesis was approved by the Medical Ethical Committee of the University Hospitals of Leuven and all participants gave written informed consent.

Personal contribution

Chapter 3

- Section 1: performed Sanger sequencing experiments for segregation analyses of variants, interpreted the clinical significance of variants, and drafted this section.
- Section 2: prepared and shipped DNA samples, analyzed WGS data, performed Sanger sequencing experiments for segregation analyses of variants, interpreted the clinical significance of variants, and drafted this section.
- Section 3: identified the *KLHL20* variant in one patient which started this project, initiated the collaboration between researchers/clinicians of different genetic centers to bring together phenotypic and molecular data, initiated the collaboration with the Switch lab of KU Leuven to perform *in silico* analysis of the variants, coordinated the project, and drafted this section.

Chapter 4

- Section 1: performed a literature study on the shared development of brain and face shape and analyzed/interpreted GWAS signals through gene set enrichment and literature analyses.
- Section 2: put together molecular and phenotypic data of literature cases, interpreted genetic data of proband, and drafted this section.

- Section 3: collected 3D facial images, analyzed facial imaging and genetic data, coordinated the project, and drafted this section.

Conflict of interest statement

The authors declare no conflict of interest.

CURRICULUM VITAE AND LIST OF PUBLICATIONS

Personalia

Name: Yoeri Sleyp

Date of birth: 27/12/1995

Nationality: Belgian

Email: yoerisleyp@gmail.com

Education

Oct 2018 – Oct 2022	KU Leuven University Doctoral Training in Biomedical Sciences Thesis: molecular diagnostics in autism spectrum disorder
Sep 2016 – Jun 2018	KU Leuven University Master of Biomedical Sciences (magna cum laude) Thesis: the introduction of whole genome sequencing for rare human disorders
Sep 2013 – Aug 2016	KU Leuven University Bachelor of Biomedical Sciences

Other research activities

2019-2020	Supervision of the master thesis entitled "Clinical validation of variants in autism genes" of a master student from the faculty of Medicine

Attended meetings and conferences

2022	European Society of Human Genetics Conference, Vienna, Austria *Oral presentation*
2022	Joint Meeting Belgian Society for Human Genetics and Dutch Society for Human Genetics, Bruges, Belgium *Oral presentation*
2019	Marguerite-Marie Delacroix Autism Research Workshop, Brussels, Belgium *Oral presentation*
2019	Facial genetics symposium, Pittsburgh (PA), USA
2019	Genomics Core Workshop, Leuven, Belgium *Oral presentation*

Additional certificates and trainings

2021-2022	KUL course: Strategic management in the pharma sector
2019	VSC workshop: Matlab programming introduction
2018	VSC workshop: Linux tools
2018	VSC workshop: Linux scripting
2018	VSC workshop: Linux introduction

Publications in peer reviewed journals

- Sleyp Y, Valenzuela I, Accogli A, Ballon K, Ben-Zeev B, Berkovic SF, et al. *De novo* missense variants in the E3 ubiquitin ligase adaptor *KLHL20* cause a developmental disorder with intellectual disability, epilepsy and autism spectrum disorder. Genet Med. 2022. In press.

- Naqvi S, Sleyp Y, Hoskens H, Indencleef K, Spence JP, Bruffaerts R, et al. Shared heritability of human face and brain shape. Nat Genet. 2021;53(6):830–9.

- Sleyp Y, Swillen A, Van Den Bogaert K, Massa G, Aerssens P, Peeters H. A *de-novo* 15q24.2 deletion involving *SIN3A* is associated with emotional, behavioural, motor problems and hypersensitivity in a girl with above average intelligence and typical facial features. Clin Dysmorphol. 2020;29(4):210–3.

- Wang T, Hoekzema K, Vecchio D, Wu H, Sulovari A, Coe BP, Gillentine MA, Wilfert AB, Perez-Jurado LA, Kvarnung M, Sleyp Y, Earl RK, Rosenfeld JA, Geisheker MR, Han L, Du B, Barnett C, Thompson E, Shaw M, Carroll R, Friend K, Catford R, Palmer EE, Zou X, Ou J, Li H, Guo H, Gerdts J, Avola E, Calabrese G, Elia M, Greco D, Lindstrand A, Nordgren A, Anderlid B-M, Vandeweyer G, Van Dijck A, Van der Aa N, McKenna B, Hancarova M, Bendova S, Havlovicova M, Malerba G, Dalla Bernardina B, Muglia P, van Haeringen A, Hoffer MJV, Franke B, Cappuccio G, Delatycki M, Lockhart PJ, Manning MA, Liu P, Scheffer IE, Brunetti-Pierri N, Rommelse N, Amaral DG, Santen GWE, Trabetti E, Sedlacek Z, Michaelson JJ, Pierce K, Courchesne E, Kooy RF, Nordenskjold M, Romano C, Peeters H, Bernier RA, Gecz J, Xia K, Eichler EE. Large-scale targeted sequencing identifies risk genes for neurodevelopmental disorders. Nat Commun. 2020;11(1):4932.

Oral presentations

- Sleyp Y, Valenzuela I, Accogli A, Ballon K, Ben-Zeev B, Berkovic SF, Broly M, Callaerts P, Caylor RC, Charles P, Chatron N, Cohen L, Coppola A, Cordeiro D, Cuccurullo C, Cuscó I, diMonda J, Duran-Romaña R, Ekhilevitch N, Fernández-Alvarez P, Gordon CT, Isidor B, Keren B, Lesca G, Maljaars J, Mercimek-Andrews S, Morrow MM, Muir AM, Rousseau F, Salpietro V, Scheffer IE, Schnur RE, Schymkowitz J, Souche E, Steyaert J, Stolerman ES, Vengoechea J, Ville D, Washington C, Weiss K, Zaid R, Sadleir LG, Mefford HC, Peeters H. *De novo* missense variants in the E3 ubiquitin ligase adaptor *KLHL20* cause a developmental disorder with intellectual disability, epilepsy and autism spectrum disorder. Oral presentation at the European Society of Human Genetics Conference, Vienna, Austria, 11 Jun 2022.

- Sleyp Y, Matthews H, Vanneste M, Hoskens H, Claes P, Peeters H. The potential of 3D facial analysis to recognize monogenic causes in the autism spectrum. Oral presentation at the joint meeting of the Belgian Society for Human Genetics and Dutch Society for Human Genetics, Bruges, Belgium, 22 Apr 2022.

- Sleyp Y, Steyaert J, Ortibus E, Claes P, Peeters H. Recent advances and challenges in genetic diagnostics for autism spectrum disorder. Oral presentation at the Marguerite-Marie Delacroix Autism Research Workshop, Brussels, Belgium, 22 Nov 2019.

- Sleyp Y, Souche E, Van Esch H. The introduction of whole genome sequencing for rare human disorders. Oral presentation at the Annual Genomics Core Workshop, Leuven, Belgium, 25 Jan 2019.

- Sleyp Y, Hens G, Rayyan M, Peeters H, Devriendt K. Familial cleft palate and Klippel-Feil syndrome. Oral presentation at the 30th European Meeting on Dysmorphology, Strasbourg, France, 11-13 Sep 2019.

- Sleyp Y, Souche E, Devriendt K, Boulanger S, Van Esch H. Symphalangism is part of the clinical spectrum of Ellis-van Creveld syndrome. Oral presentation at the 18th Manchester Dysmorphology Conference, Manchester, United Kingdom, 5-8 Nov 2018.

DANKWOORD

Graag sluit ik deze thesis af met een dankwoord aan al diegene die mijn pad gekruist zijn tijdens mijn doctoraat. De voorbije jaren heb ik namelijk ontzettend veel bijgeleerd zowel op professioneel als persoonlijk vlak. Dit zou niet mogelijk zijn geweest zonder de bijdrage van velen en daarom zou ik graag iedereen willen bedanken voor de fijne tijd de afgelopen vier jaar. Jullie hebben er mede voor gezorgd dat ik mijn doctoraat tot een goed einde gebracht heb.

Eerst en vooral wil ik mijn promotor Prof. Hilde Peeters bedanken. Het was eerder per toeval dat ik de kans kreeg om een doctoraat te starten in jouw labo. Ik ben enorm blij dat ik deze kans met beide handen gegrepen heb. Ik denk dat er weinig promotors zijn die zo veel doen voor hun doctoraatsstudenten als jij. Je zorgde voor een omgeving waar ik me optimaal kon ontwikkelen. Je was ook steeds beschikbaar bij vragen en gaf raad wanneer nodig. Je gaf veel vertrouwen en was steeds optimistisch waardoor je het beste uit me kon halen. Daarnaast was je ook de eerste om me te verdedigen bij anderen en om ervoor te zorgen dat er niet te veel werk op mijn schouders terecht kwam. Het was ook steeds plezant met jou op (inter)nationale congressen, meetings en workshops. De resultaten van deze thesis zijn ook jouw verdienste. Je stelde alle nodige data ter beschikking, trok belangrijke projecten naar ons toe en gaf constructieve feedback wanneer nodig. Tijdens deze thesis heb ik mogen ervaren wat voor een geweldige klinisch geneticus en onderzoeker jij bent. Een promotor uit de duizend!

Daarnaast wil ik ook graag een dikke merci zeggen aan mijn co-promotor Prof. Peter Claes om me mee te begeleiden tijdens mijn doctoraat. Je was steeds bereid om mee te denken over de projecten van mijn thesis en de daarbij behorende analyses. Je stelde computers en servers ter beschikking om deze analyses mogelijk te maken en stond altijd paraat om vragen te beantwoorden. Jouw out of the box denken en inzichten waren erg waardevol de voorbije vier jaren. Bedankt ook voor jouw feedback en enthousiasme telkens wanneer ik mijn resultaten kwam voorstellen. Je creëerde vanaf het begin een informele sfeer waar ik me steeds op mijn gemak voelde. Je bent een fantastische onderzoeker en ik kijk al uit naar jouw toekomstige onderzoeksprojecten.

I would also like to thank all the members of the jury, Prof. Benedikt Hallgrímsson, Prof. Bert Callewaert, Prof. Els Ortibus, and Prof. Jeroen Breckpot, for their constructive feedback during the revision of the thesis. Your suggestions and remarks have allowed me to deliver the best possible version of this thesis. I would especially like to thank Prof. Els Ortibus and Prof. Jeroen Breckpot who, as internal jury members, annually monitored the progress of my PhD. In addition, I would

like to thank Prof. Karel Allegaert and Prof. Eric Legius for acting as chair of my reading committee and my public defence respectively.

Ook de (oud-)onderzoekers van het Hilde Peeters lab verdienen een dankwoordje. Eerst wil ik mijn voorgangers Nele en Laura bedanken. Het onderzoek dat jullie verricht hebben en de data die jullie verzameld hebben gaven me meteen een voorsprong bij de start van mijn doctoraat. Daarnaast wil ik ook mijn master thesis student Valentine bedanken. De data en resultaten van jouw thesis hebben me erg geholpen bij mijn onderzoek. Ik was erg onder de indruk van jouw zelfstandigheid en hoe snel je alles kon oppikken. Het was een zegen om jou als master thesis student te hebben. Als laatste wil ik graag Michiel bedanken. De afgelopen jaren heb je me vaak geholpen bij de interpretatie van mijn onderzoeksresultaten. Je bent ook altijd de eerste om te vragen of je iets moet doen of er iets moet nagelezen worden wat ik enorm apprecieer. Door het congres in Wenen weet ik dat je niet alleen een goede onderzoeker bent maar ook een heel fijn persoon bent met een goed gevoel voor humor.

Tijdens mijn doctoraat heb ik samengewerkt met verschillende labo's en werkte ik als gevolg regelmatig vanop verschillende plaatsen. Mijn vaste plaats was de studentenbureau op het CME. Hoewel ik de enige onderzoeker was van mijn labo op deze bureau voelde ik me hier steeds welkom. Bedankt aan Greet, Lisanne, Margot, Heleen, Eline, Lieselot, Matt, Nicolas, Jente en Elise voor de fijne gesprekken tijdens het werk, om vragen over het doctoraat te beantwoorden, om me te helpen bij de labo experimenten, voor al die keren dat we samen broodje/slaatje gingen halen, op jacht gingen naar traktaties, ... Daarnaast wil ook Sarah, Koen, Berardo, Stefania, Chiara, Jef, Charlotte, Tine, Tanja, Dhanya, Yan, Olga, Thomas, Armelle, Laura en Adrian bedanken voor de fijne werksfeer op de bureau of in het labo. Regelmatig kon je mij ook terugvinden op een bureau in het Medical Imaging Center voor de analyse van 3D gelaatsbeelden. I want to thank Harry for his help in writing Matlab scripts and for troubleshooting errors in Matlab, but also for the fun conversations in Pittsburgh or when we went to lunch together. Daarnaast kon ik bij Hanne, Karlijne, Seppe, Meng en Soha steeds terecht voor een gezellige babbel. Dominique zorgde er dan weer voor dat de computers en servers bleven draaien en hielp me wanneer ik mijn paswoord weer vergeten was. Naar het einde van mijn doctoraat toe kon je me dan weer vaak terugvinden in het fruitvliegenlabo. Graag zou ik Prof. Patrick Callaerts bedanken voor mij te introduceren in de wereld van de fruitvliegjes. Ik werd met open armen ontvangen door Mattias, Lenz, Cissy, Sofie, Veerle en Tom wanneer ik bij het fruitvliegenlabo arriveerde. Bedankt voor alle leuke en grappige momenten die ik bij jullie beleefd heb.

Graag bedank ik ook nog een aantal personen die verbonden zijn aan het CME. Hartelijk dank aan alle klinisch genetici waarmee ik de voorbij jaren heb mogen samenwerken. Met bijzonder aan Prof. Hilde Van Esch om een goed woordje over mij te doen waardoor ik dit doctoraat heb kunnen starten en ook aan Prof. Koen Devriendt voor zijn waardevolle suggesties bij mijn onderzoeksresultaten. Verder wil ik ook nog Prof. Joris Vermeesch bedanken om mij te laten deelnemen aan de groepsmeetings van zijn labo, mij te betrekken bij de WGS meetings en zich in te zetten voor de implementatie van WGS in de diagnostiek. Daarnaast wil ik ook Prof. Ann Swillen bedanken voor de hulp bij het schrijven van mijn eerste artikel. Ik heb tijdens mijn doctoraat ook kunnen rekenen op de mensen van de diagnostiek en de counselors van het CME. Kris, Valérie, Lies, Marijke en Hanne hielpen me bij de interpretatie en validatie van mijn onderzoeksresultaten of bij het rekruteren van patiënten. Bedankt ook aan het secretariaat van het CME om mij telkens op te bellen wanneer er weer iemand stond voor een 3D foto of mij te helpen wanneer ik weer een pakket moest versturen. Bruno van de DNA-extractiefaciliteit in het CME zorgde er dan weer voor dat er DNA stalen voor mij klaar stonden. Graag bedank ik ook nog Erika van de Genomics Core om ervoor te zorgen dat ik met betrouwbare sequencing data aan de slag kon. Tijdens het congres in Wenen heb ik ook een heleboel assistenten van het CME (beter) leren kennen. Bedankt aan Michiel, Laurens, Marie, Mio en Nathalie voor de fijne periode zowel tijdens het congres en de avonden na het congres.

Verder bedank ik ook graag alle leden van TIARA en LAuRes waarbij ik terecht kon voor een babbeltje tijdens de verschillende meetings die ze organiseerden. Dankzij jullie heb ik ook een andere kant van het autismeonderzoek leren kennen.

Vrienden zorgden voor de nodige ontspanning de afgelopen vier jaar. Hierbij bedank ik de homies Sander, Boris, Arnaud, Thomas, Stijn, Erland, Jan, Bram, Joren, Ewout en Bart voor de barbecues, weekendjes, fietstochtjes, beach volleybalmatchen, spelletjes, … Verder wil ik mijn oud-medestudenten biomedische wetenschappen Karen, Mattias en Sander bedanken voor de morele steun en de gezellige etentjes. Ook bedank ik graag alle wielrenners van WTC de Fonzies voor al die keren dat ik mee mocht rijden en in het wiel mocht zitten de afgelopen jaren. Hopelijk kan ik in de toekomst wat meer op kop rijden. Ook bedankt aan de mannen van BCO united voor de toffe sfeer en teamspirit tijdens de zaalvoetbal.

Met Vi heb ik het afgelopen anderhalf jaar ook onvergetelijke momenten meegemaakt. Wat begon als pianolessen tijdens mijn doctoraat evolueerde al snel naar samenwonen op een kleine studio. Hier hebben we samen een fijne tijd beleefd. Ook onze bubble tea momentjes, wandelingetjes, uitstapjes, … waren enorm leuk en hielpen me om mijn gedachten te verzetten wanneer nodig. Je

speelde ook publiek toen ik mijn presentaties aan het oefenen was. Dank je voor al jouw steun en liefde het afgelopen anderhalf jaar. Ik ben benieuwd wat de toekomst nog voor ons in petto heeft!

Ook mijn familie wil ik graag bedanken voor hun onvoorwaardelijke steun tijdens dit doctoraat. Met bijzonder mijn ouders die me steeds alle kansen en vrijheid hebben gegeven om te doen wat ik wou, wat uiteindelijk heeft geresulteerd in een doctoraat in de biomedische wetenschappen. Jullie hulp achter de schermen, samen met die van oma en peter, was van onschatbare waarde tijdens mijn doctoraat en alle jaren daarvoor. Ook bedankt aan moemoe om vanaf het begin van mijn opleiding biomedische wetenschappen in mij te geloven en bij iedereen te gaan stoefen dat ik een “dokter” ging worden. Het is dan toch gelukt (een “doctor” weliswaar).

Yoeri