From Department of Women's and Children's Health
Karolinska Institutet, Stockholm, Sweden

GENETIC IMPLICATIONS OF INDIVIDUAL INTERVENTION AND NEURONAL DYSFUNCTION IN NEURODEVELOPMENTAL DISORDERS

Danyang Li

李丹阳

Karolinska Institutet

Stockholm 2022

Genetic implications of individual intervention and neuronal dysfunction in neurodevelopmental disorders
THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Danyang Li

Atrium, Wargentinhuset entrance floor, Nobels väg 12B, Karolinska Insititutet, Solna
February 25th, 2022, 09:30 am

Principal Supervisor:
Docent Kristiina Tammimies
Karolinska Institutet
Department of Women's and Children's Health
Division of Neuropsychiatry

Co-supervisor(s):
Professor Sven Bölte
Karolinska Institutet
Department of Women's and Children's Health
Division of Neuropsychiatry

Dr. Hong Jiao
Karolinska Institutet
Department of Biosciences and Nutrition

Opponent:
Docent William Hennah
University of Helsinki
Neuroscience Center

Examination Board:
Docent Catharina Lavebratt
Karolinska Institutet
Department of Molecular Medicine and Surgery
Division of Translational Psychiatry

Docent Nick Tobin
Karolinska Institutet
Department of Oncology-Pathology

Docent Linda Halldner Henriksson
Umeå Universitet
Department of Clinical Sciences
Division of Child and Adolescent Psychiatry

To everyone who shows kindness in my life

ABSTRACT

Neurodevelopmental disorders (NDDs) are a group of conditions appearing in childhood, with developmental deficits that produce impairments of functioning. Autism spectrum disorder (ASD) is a common NDD with a high heritability affected by complex genetic factors, including both common and rare variants. Behavior interventions such as social skills group training (SSGT) have been widely used in school-aged autistic individuals to relieve social communication difficulties in a group setting. Studies have confirmed that intervention outcomes can be influenced by sex and age, but how the genetic risk contributes to the outcome variability remains elusive. Furthermore, although large population cohorts have been well studied and have found numerous genes associated with ASD and NDDs, the molecular and neuronal outcomes of risk variants and genes are unclear. Therefore, this thesis included four studies in which the effects of genetic factors on intervention outcomes and cellular level neuronal functions were investigated. Results from this thesis may provide a genetic perspective for further studies to explore potential individualized treatments for ASD and other NDDs.

Specifically, In STUDY 1-3, exome sequencing and microarray were performed on individuals from a randomized controlled trial of SSGT (KONTAKT®). Common and rare variants, including copy number variations (CNVs) and exome variants, were tested for association effects with SSGT and standard care intervention outcomes. Polygenic risk scores (PRSs) were calculated from common variants, and clinically significant rare CNVs and rare exome variants were prioritized. Molecular diagnoses were identified in 12.6% of the autistic participants. PRSs and carrier status of clinically significant rare variants were associated with intervention outcomes, although with varied effects on both SSGT and standard care. In addition, genetic scores representing variant loads in specific gene sets were obtained from rare and common variants in ASD-related pathways. Outcomes of interventions were differentially associated with genetic scores for ASD-related gene sets including synaptic transmission and transcription regulation from RNA polymerase II. After combining genetic information and behavior measures, a machine learning model was able to select important features and confirm that the intervention outcomes were predictable.

In STUDY 4, genetic variants affecting Calcium/Calmodulin Dependent Serine Protein Kinase (*CASK*) gene, a risk gene for NDDs, were examined using human induced pluripotent stem cell-derived neuronal models to identify the cellular effects of these mutation

consequences. CASK protein was reduced in maturing neurons from mutation carriers. Bulk RNA sequencing results revealed that the global expression of genes from presynaptic development and CASK network were downregulated in CASK-deficient neurons compared to controls. Neuronal cells influenced by *CASK* mutations showed a decrease of inhibitory presynapse size and changed excitatory-inhibitory (E/I) balance in developing neural circuitries.

In summary, this is the first study to investigate the association of genome-wide rare and common variants with ASD intervention outcomes. Differential variant effects were found for individuals receiving SSGT or standard care. Future studies should include genetic information at different levels to improve molecular genetic testing for diagnoses and intervention plans. Presynapses and E/I imbalance could be an option to be developed for the treatment of CASK-related disorders.

LIST OF SCIENTIFIC PAPERS

I. Association between Copy Number Variation and Response to Social Skills Training in Autism Spectrum Disorder.
K Tammimies, **D Li**, I Rabkina, S Stamouli, M Becker, V Nicolaou, S Berggren, C Coco, T Falkmer, U Jonsson, N Choque-Olsson, S Bölte. *Sci. Rep.* **9**, 9810 (2019).

II. The influence of common polygenic risk and gene sets on social skills group training response in autism spectrum disorder.
D Li, N Choque-Olsson, H Jiao, N Norgren, U Jonsson, S Bölte, K Tammimies. *npj Genomic Med.* **5**, 45 (2020).

III. Rare variants in the outcome of social skills group training for autism.
D Li, N Choque-Olsson, M Becker, A Arora, H Jiao, N Norgren, U Jonsson, S Bölte, K Tammimies. *Autism Res.* 1-13 (2021). https://doi.org/10.1002/aur.2666

IV. Presynaptic dysfunction in CASK-related neurodevelopmental disorders.
M Becker, F Mastropasqua, J Philipp Reising, S Maier, ML Ho, I Rabkina, **D Li**, J Neufeld, L Ballenberger, L Myers, V Moritz, M Kele, J Wincent, C Willfors, R Sitnikov, E Herlenius, BM Anderlid, A Falk, S Bölte, K Tammimies. *Transl Psychiatry.* **10**, 312 (2020).

CONTENTS

LIST OF ABBREVIATIONS

NDD	Neurodevelopmental disorders
ASD	Autism spectrum disorder
ID	Intellectual disability
ADHD	Attention-deficit/hyperactivity disorder
MICPCH	Mental retardation and microcephaly with pontine and cerebellar hypoplasia
DSM-5	Diagnostic and Statistical Manual of Mental Disorders Fifth Edition
CNV	Copy number variation
SNV	Single nucleotide variant
SNP	Single-nucleotide polymorphism
PRS	Polygenic risk score
Indel	Insertion or deletion
VCS	Variant of clinical significance
VUS	Variant of uncertain significance
LOF	Loss-of-function
GWAS	Genome-wide association study
QC	Quality control
GSASD	Genetic score for ASD-related gene set
GSSyT	Genetic score for synaptic transmission genes
GSPol	Genetic score for regulation of transcription from RNA polymerase II promoter genes
SRS	Social Responsiveness Scale
ADOS	Autism Diagnostic Observation Schedule
SSGT	Social skills group training
ML	Machine learning
SVM	Support vector machine
CASK	Calcium/calmodulin-dependent serine protein kinase
iPSC	Induced pluripotent stem cell
NES cell	Neuroepithelial-like stem cell
GSEA	Gene set enrichment analysis

1 INTRODUCTION

Neurodevelopmental disorders (NDDs) are a group of conditions appearing in childhood, with developmental deficits that produce a range of functioning impairments. Some of the disorders in this category are intellectual disability (ID), attention-deficit/hyperactivity disorder (ADHD), and autism spectrum disorder (ASD) [1]. Most NDDs last for a life time and are frequently comorbid with other psychiatric conditions [1]. The prevalence of NDDs is greater in males compared to females. Studies have indicated a high heritability with NDDs influenced by complex genetic and environmental factors [2].

Autism spectrum disorder (ASD) is characterized by persistent deficits in social communication and interaction, together with a range of restricted, repetitive patterns of behavior, interests or activities [1]. The number of cases has been continuously growing worldwide over recent decades reaching an average of 1% in the population [3]. The majority of autistic individuals have co-occurring disorders with varied symptoms from mild to severe [4]. ASD is highly heritable, and complex and diverse genetic factors can contribute to the disorder. Both common and rare variants have substantial contributions, and they converge in specific biological pathways that induce cellular dysfunction [5]. Nevertheless, few studies have revealed the role of genetic information on ASD interventions, which have limited benefits and varied outcomes in individuals with ASD [6,7]. Additionally, the molecular and neuronal outcomes of many risk variants and genes of ASD and NDDs are unclear, especially in humans.

Therefore, in this thesis, a randomized controlled trial of social skills group training (SSGT) [8], a common intervention applied to school-aged individuals with ASD, was used to examine the effect of different levels of genetic variants on intervention outcomes of SSGT and standard care. Common and rare variants including rare copy number variations (CNVs) and rare sequencing variants were tested and combined with clinical evaluations to gain insight into how different genetic moderators influence individual outcomes. Furthermore, using human induced pluripotent stem cell (iPSC)-derived neuronal models, this thesis also applied transcriptional, and functional analyses to elucidate the consequence of deleterious variants from the *CASK* gene, which has been strongly implicated in ASD and a range of NDDs.

2 LITERATURE REVIEW

2.1 NEURODEVELOPMENTAL DISORDERS

Neurodevelopmental disorders (NDDs) are introduced as a group of conditions with onset during the developmental period, often before grade school [1]. The disorders are characterized by developmental deficits that impair personal, social, academic, or occupational functioning [1]. Some of the disorders in this category include intellectual disability (ID), attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), and learning disabilities, and these disorders usually have a high co-occurrence with other psychiatric conditions. Most NDDs have numerous contributing factors affecting complex mechanisms, and the etiology in many individual cases is still unclear. Therefore, given the varied symptoms and causality, diagnoses and treatments in NDDs are difficult. A combination of treatments is often applied involving professional therapy, pharmaceuticals, and home- or school-based programs. In this thesis, ASD is selected for further discussion as a paradigmatic example of the NDDs group.

2.2 AUTISM SPECTRUM DISORDER (ASD)

ASD is characterized by persistent deficits in social communication and social interaction, and by a range of restricted, repetitive patterns of behavior, interests or activities [1]. The word "spectrum" represents a wide range of disorder manifestations which depend on the autistic severity, developmental level, and chronological age. According to the Diagnostic & Statistical Manual of Mental Disorders Fifth Edition (DSM-5), individuals with autistic disorder, pervasive developmental disorder not otherwise specified and Asperger's disorder are integrated into the single diagnosis of ASD [1]. The severity of the autistic condition may vary by individual, can fluctuate over time, and is specified by the presence or absence of accompanying language or intellectual impairment. With heterogeneous symptoms ranging from mild to severe, ASD is a lifelong condition requiring diverse support from families, schools, clinicians, and societies [9].

2.2.1 Epidemiology

The global prevalence of ASD is about 1% of the population, with similar estimates in children and adults [3,10]. In Nordic countries, over 1% of children in Finland and Sweden and around 1.5% of children in Denmark have ASD [11]. The highest prevalence was recently

reported in the United States with 1 in 44 children aged eight years estimated to be autistic, although the prevalence differed among racial groups, proximity to services, urban areas, and socioeconomic status [12]. Rates of ASD prevalence have also been continuously growing over recent decades, a trend especially pronounced in autistic individuals with average or above-average IQ [3,13]. Reasons for the increased diagnoses could be an expansion of the diagnostic criteria, improved awareness and services, and/or a true frequency increase of ASD [13].

2.2.2 Etiology

Various studies have confirmed that ASD is a highly heritable disorder. A meta-analysis combining multiple twin studies estimated that 64–91% of ASD risk is heritable [14]. A recent Swedish study also showed a high heritability rate of 0.88 to 0.97 of ASD diagnoses, and changes in the genetic and environmental contributions were weak over time [15]. Many genetic and environmental risk factors have been acknowledged to be associated with ASD. A number of studies have indicated both older maternal and older paternal age can increase ASD risk [16]. During pregnancy, maternal metabolic conditions, weight gain, and medication usage (e.g. valproic acid) have also been linked to the disorder [17,18]. Supplements, such as folic acid at preconception reduce the risk of both ASD and general developmental disabilities [19]. In addition to the environment, genetic factors are a crucial cause of ASD. Therefore, more studies are needed to understand the contributions of both common and rare variants, combined with environmental factors, and the complex interactions among these to explain the occurrence and heterogeneity of the disorder.

2.2.3 Diagnoses

The essential features of ASD presented in the early developmental period are persistent impairments in communication and interaction, and restricted, repetitive sensory-motor behaviors. Based on clinical criteria, ASD can be classified into syndromic and nonsyndromic. The syndromic diagnosis represents conditions in which ASD occurs together with additional phenotypes and/or dysmorphic features [20]. The etiology for these conditions is mostly clear and can be linked with chromosomal abnormalities, or mutations in a single gene, such as fragile X syndrome or Rett syndrome. Nonsyndromic ASD, on the other hand, is the classic form of autism with no presentation of additional symptoms. It also refers to "idiopathic" where the etiology is unknown for most cases [20]. As no reliable biomarker exists so far, ASD diagnoses are based on clinically significant impairments in social, occupational, educational, or other functioning and these disturbances are not explained by ID or global

4

developmental delay [1].

Abnormal individuals can be typically detected in the second year of life but maybe younger if the developmental delays are severe, or later than 24 months if symptoms are more subtle [1,21]. Cultural and socioeconomic factors may affect recognition or diagnosis age considering the cultural differences in norms for social interaction and nonverbal communication [1]. Additionally, the diagnosis of ASD is around four times higher in males compared to females [1,12] and girls with ASD tend to also have intellectual impairments or language delays [1]. At present, standardized ASD diagnostic instruments include the Screening Tool for Autism in Toddlers and Young Children (STAT) for young children, and the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R) for both children and adults [22–24]. Symptom assessment can be measured using various scales, such as the Childhood Autism Rating Scale (CARS), the Social Responsiveness Scale (SRS), and the Social Communication Questionnaire (SCQ) [25–27].

2.2.4 Coexisting conditions

Most autistic individuals have co-existing disorders, including neurological disorders (seizures, sleep disorders), genetic disorders (fragile X syndrome), psychiatric conditions (anxiety, depression), and other NDDs (ID, ADHD). Up to 70% of people with ASD have at least one concurring condition [4], which may exacerbate or mask ASD symptoms, and may require individualized intervention.

Intellectual disability is one of the most frequent comorbidities in autistic individuals. With the sharpening diagnostic criteria in ASD, dual diagnoses of ID and ASD have declined to 35% recently [12,28]. Given a considerable phenotypic and etiologic overlap, the line between autism and ID is blurred in the clinic. As social function evaluation is included in the conceptualization of both ID and ASD, determining whether observed social deficits are attributable to ID or ASD remains a challenge. At a genetic level, autism and ID share many of the same genes, some of which are more common in ASD and others are more common in ID [29]. Understanding the relationship of the two conditions may operationalize criteria for the diagnosis and eventually lead to improved treatments.

ADHD is another NDDs that often co-occurs with ASD. A recent meta-analysis showed an overall pooled prevalence estimate of 28% for ADHD in the autism population [30]. In addition to their high co-occurrence, both conditions can involve abnormal manifestations such as language delays, heightened sensory responses, and emotional regulation problems. Similar to ASD, ADHD usually appears in childhood and is more often diagnosed in boys [31].

Genetic studies of common and rare variants are starting to find at least partly shared risk for ASD and ADHD [32].

Additionally, around 10-20% of autistic children have a diagnosable genetic syndrome [33]. Some of these conditions are obvious while others are subtle and need specialized testing. It is therefore recommended to have an evaluation from a clinical geneticist following an ASD diagnosis [1]. For instance, mental retardation and microcephaly with pontine and cerebellar hypoplasia (MICPCH) is a comorbid genetic disorder characterized by intellectual disability and pontocerebellar hypoplasia. The affected gene is Calcium/Calmodulin Dependent Serine Protein Kinase (*CASK*) located on chromosome X and females tend to have a higher diagnosis rate than males [34]. Like most genetic disorders, there is no cure for MICPCH at present.

2.2.5 Interventions

Although it is not possible yet to completely relieve problems for autistic individuals, interventions for ASD should start as soon as possible after diagnosis to reduce impairments, and to help individuals learn new skills and consolidate strength. In addition to some medications being used to treat syndromes concurring with ASD [35–38], different behavioral interventions have been used to alleviate ASD symptoms and prevent the onset of comorbidities. For instance, cognitive-behavioral therapy in children and adolescents with ASD has shown positive effects [39]. Applied Behavior Analysis is another well-known behavioral intervention for young children, in which Early Intensive Behavioral Intervention is a widely disseminated comprehensive program [40].

Social skills interventions have become the most attracted program among researchers and clinicians to promote social skills in children, adolescents and young adults with ASD [7]. For school-aged children and adolescents with ASD, social skills group training (SSGT) is one of the most commonly applied behavioral interventions. It aims to alleviate social communication difficulties in individuals within average to high intellectual ability in a group setting. This program applies multiple elements such as cognitive-behavioral therapy, computer-based cognitive training, behavioral activation, psycho-education, observational learning, and parent involvement through various treatment formats [8,41]. Recently, a randomized controlled trial of SSGT in Sweden (KONTAKT®) included individuals (7-17 years) with ASD diagnosis and IQ > 70. Each participant had one or more common neurodevelopmental or psychiatric diagnosis, including ADHD, depression or anxiety [8]. Participants randomly received standard care, including any commonly provided

6

interventions or standard care plus SSGT. However, compared to standard care only, SSGT had a moderate effect, with varied responses influenced by age and sex [8].

2.3 ASD GENETICS

2.3.1 Genetic architecture of ASD

The architecture of genetic risk is complex and diverse in ASD. Common variants with additive polygenic contributions and rare variants following monogenic inheritance coexist, creating a broad spectrum of genetic backgrounds among individuals. Variants found in ASD-related genes are mostly rare variants with normally < 1% frequency in the population. Both *de novo* and inherited rare variants appear in at least 10-30% of autistic individuals [42]. Common variants, despite modest individual effect, have a large contribution to heritability when considered cumulatively [43].

2.3.2 Genetic pleiotropy with other conditions

Substantial heritability has been reported in ASD and psychiatric disorders from twin and family studies, and phenotype overlaps and comorbidities are common for many twin pairs [5]. Similar to ASD, most psychiatric disorders are highly polygenic. Some of the genetic mutations are specific to one disorder, but most of them are pleiotropic with the effects conferring broad phenotypic risks across diagnostic categories and between disorders and behavioral traits. These large-effect pleiotropic variants may disrupt key neurodevelopmental processes rather than causing one specific clinical phenotype [44]. Thus, some copy number variations (CNVs) and *de novo* rare variants associated with ASD can also confer risk for ID, ADHD, and schizophrenia [45–47]. Genetic correlation from genome-wide association studies (GWASs) measures the average effect of pleiotropy for common alleles and has shown substantial connections in psychiatric and behavioral phenotypes [48]. A recent study of eight psychiatric disorders revealed a complex genetic structure, of which ASD is significantly correlated with schizophrenia, ADHD, and major depression (Figure 2.3.1) [49]. Additionally, significant genetic sharing is found between ASD, social skills and early life cognitive measures (college attainment, years of education, and intelligence) suggesting a link between cognitive performance and the genetic risk for the disorder and related phenotypes [50,51].
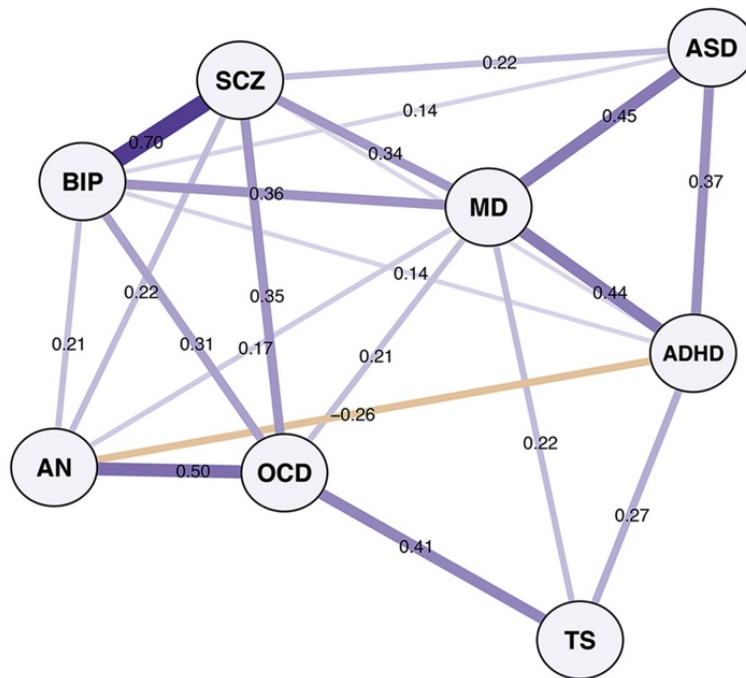
Figure 2.3.1. SNP-based genetic correlations between eight disorders. ADHD: attention deficit hyperactivity disorder; ASD: autism spectrum disorder; AN: anorexia nervosa; BIP: bipolar disorder; MD: major depression; OCD: obsessive-compulsive disorder; SCZ: schizophrenia; TS: Tourette syndrome. The figure was obtained from Smoller *et al.* [49]

2.3.3 Common variants and polygenic risk score

From the early 20th century, different models asserted an additive genetic effect of large numbers of genetic loci contributing to both complex and binary traits [52,53]. With the completion of the Human Genome Project, HapMap, and collaborative biological projects, new technologies, genomic data and the advent of biobanks have motivated the emergence of GWAS, which focuses on associations between whole genome single-nucleotide polymorphisms (SNPs) and different traits in humans and other organisms. After the first GWAS finished in 2002 [54], thousands of significant SNP associations have been detected in human traits [55]. However, due to limited samples in clinical research, most common variants with weaker effects cannot be detected by traditional GWAS.

Polygenic risk score (PRS) is the most commonly applied approach of genetic risk estimation. PRS aggregates the numbers and the effects of risk alleles for each individual, where the numbers are acquired from individual genotyping, and the effects are derived from GWAS summary statistics [56]. PRS can combine multiple loci below genome-wide significance to predict disease risk even with a limited sample size. It has been widely used to stratify individuals with distinct risks for disease prevention, early detection, and response to treatment in psychiatric disorders, coronary artery disease, and cancers [57].

Despite only a few significant SNPs indicated from ASD GWAS, the polygenic variation recapitulated by PRS for ASD showed that a cumulative variation effect was correlated with ASD and autistic traits [58,59]. In autistic individuals, an increased polygenic

burden for autism is associated with methylome variation [60], and can be over-transmitted to ASD children in simplex families [61]. In the general population, ASD PRS is also associated with cortical thickness and infant neurodevelopment such as gross motor function and receptive language development [62,63]. In schizophrenia cases, the correlation of ASD PRS has been linked to left amygdala function and emotion recognition [64]. A recent study demonstrated that PRS for ASD was negatively associated with the reduction of depressive symptoms after cognitive behavior therapy, providing promising implications for PRS application in psychiatric behavioral treatments [65].

2.3.4 Rare variants

For decades, rare variants have been discovered in both Mendelian disorders caused by one single gene as well as complex diseases influenced by multiple genes and environmental factors. Each variant tends to have a large effect even if it only counts for a small part of heritability in individuals. Depending on variation size, rare variants range from single nucleotide variations (SNVs), several base pairs insertions or deletions (indels), to CNVs of kilobases or larger in length. Furthermore, if variants are obtained from parent-child trios, both *de novo* and inherited mutations can be distinguished [66]. The effect of variant function can be divided into loss-of-function and gain-of-function, with more detailed classification based on Muller's morphs (amorph, hypomorph, hypermorph, antimorph and neomorph) [67].

From 2004, array hybridization methods have been used to detect CNVs [68,69]. A few years later, next generation sequencing was widely available to cost-effectively identify rare alleles in genes underlying Mendelian disorders. Currently, results from chromosome microarray and DNA sequencing help clinicians make accurate diagnoses of disorders with atypical manifestations, to screen genetic disorders before conception, birth or postpartum, and to test tumor DNA for early cancer detection [70,71]. In addition, rare variants detected from genetic testing can identify numerous targets for therapies and drugs providing new evidence in the areas of therapygenetics and pharmacogenetics [72,73].

2.3.4.1 Rare CNVs in ASD

CNVs are considered one of the most common causes of ASD, with some regions such as 16p11.2, 7q11.23, 2p16.3 and 15q11-q13 highlighted as ASD associated [74,75]. Compared to the neurotypical population, studies have shown that rare CNVs are enriched in autistic individuals, especially when considering variants previously implicated in ASD or ID [76]. In simplex families, *de novo* CNVs are more common in probands than in siblings, and tend to

be larger covering more genes [75,77]. Although larger CNVs are more frequently linked to ASD, there is no evidence of any single driver gene in the majority of large CNVs [78–80]. Instead, transgenic model organisms suggest that CNV related developmental phenotypes are attributed to dosage effects of multiple genes, indicating a potential complex effect of CNVs on gene regulatory networks [81].

2.3.4.2   Rare SNVs and indels in ASD

Initially, genes contributing to ASD risk were identified from syndromes associated with ASD such as fragile X syndrome and Rett syndrome [82,83]. After entering the era of next-generation sequencing, the first exome sequencing of ASD, launched in 2011, suggested associations with several *de novo* mutations [84]. Landmark studies in 2012 indicated additional genes carrying *de novo* mutations with significant ASD liability, and individuals with ASD tend to carry an excess of rare, loss-of-function coding variants [85–87]. Using genome sequencing, Werling *et al.* and Ruzzo *et al.* recently indicated rare *de novo* noncoding variants probably have a modest effect compared to rare *de novo* coding variants [88,89]. However, noncoding risk variants show enrichment at transcription factor binding sites especially in the distal promoter [90], and involvement in synaptic transmission and neuronal development [91]. In addition to *de novo* variants, studies have implicated novel ASD risk genes affected by rare inherited variants, providing a substantial extension of *de novo* variant findings [89,92]. Currently, genome and exome sequencing from thousands of individuals has found hundreds of associated genes with *de novo* and inherited pathogenic variants in both coding and non-coding regions, suggesting a high genetic heterogeneity and complex inheritance in ASD.

2.3.5 Relationship between common and rare variants in ASD

Both common and rare variants have proved to contribute to ASD. A recent study confirmed that autistic individuals with *de novo* variants had higher PRS for ASD compared to controls, but no PRS difference was shown in ASD group with or without *de novo* mutations [61]. Similarly, children with severe NDDs including ID, ASD, and developmental delay had comparable PRSs no matter with pathogenic variants or not [93]. On the contrary, Klei and colleagues implicated burden of common risk variation harbored by ASD subjects with pathogenic variants was lower in non-carrier ASD but higher in control subjects [94]. Although these results fit imperfectly, all these studies agree that PRS and rare variants related to ASD and other NDDs could have an additive contribution to the risk of the conditions in individuals. Additional investigations of common and rare variant relations are needed to

provide a broader and more comprehensive insight into the genetic architectures underlying NDDs.

2.3.6 Associated genes in ASD

Hundreds of genes have been implicated in association with ASD risk. The consequence of different types of mutations within these associated ASD genes could depend on how they directly act on the gene function. For example, some ASD susceptibility genes encode DNA or RNA binding proteins which can directly regulate the gene expression [81]. Variants that disrupt the protein can alter the regulation of other target genes and further propagate by downstream regulatory proteins. For non-coding region variants, additional efforts are needed to pinpoint genes affected by the risk variants to understand the pathogenicity [95]. Considering many cell type variations and long developmental trajectories in the brain, variants affecting gene regulation or transcription are specific to regions, cells, and time stages at cellular and tissue levels [96].

One example of a susceptible gene to ASD and NDD risk is the *CASK* gene located on chromosome Xp11.4. Genetic variants including loss-of-function variants, and missense variants in *CASK* have been reported in cases with MICPCH, ASD, developmental delay and ID [29,34,97,98]. CASK is highly expressed in the developing human brain and is necessary for pre- and postsynaptic function, such as regulation of synaptic vesicle exocytosis and neuronal cell adhesion at the presynapse, and regulation of ionotropic receptor trafficking at the postsynapse [99,100]. Excitatory–inhibitory imbalance in neurons was also indicated in CASK-negative mice [101]. Despite that, more studies are required to explore the effect of *CASK* variants on neuronal development, especially in humans.

2.3.7 Convergent pathways in ASD

Multiple genes associated with ASD are connected within biological interaction networks. Therefore, mutations from a single gene can influence the function of related genes magnifying their effects throughout regulatory networks. A subset of ASD genes that encode DNA binding or RNA binding proteins can regulate gene expression, such as fragile X mental retardation protein (FMRP) and CELF4 [102–104]. ASD genes with altered expression also overlap with developmentally regulated genes associated with cell cycle and proliferation, and cortical patterning and neuronal differentiation [105,106]. Analysis of protein-protein interaction networks for ASD-associated genes highlighted common pathways including synaptic transmission, transcriptional and translational regulation, chromatin remodeling,

cell adhesion, and ion transport [74,78,80,89]. Additionally, the regulation of cell-proliferation pathways especially mitogen activated protein kinase (MAPK), mammalian target of rapamycin (mTOR), and Wnt signaling, is related to ASD risk genes [76,107,108]. Gene-set analysis combining common and rare variations related to ASD has presented a converged enrichment within specific gene networks such as synaptic function, gene regulation pathways and FMRP targets [78,104].

2.3.8 Functional genetics in ASD

Studies have observed cellular types, brain regions and developmental stages in which the ASD susceptibility genes are enriched. The co-expression network reflected the enrichment of ASD risk variants in cortical glutamatergic neurons from the prefrontal cortex during fetal development [109]. Likewise, a large-scale exome sequencing study reported that significant genes are highly expressed in the early development of human cortex and enriched in maturing and mature neurons of both excitatory and inhibitory lineages [29]. Genes located in the regions of ASD CNVs were also shown to have a high co-expression in the developing cortex [110]. Using single-nucleus RNA sequencing of cortical tissues from individuals with autism, a recent study has revealed that cells affected by autism are mostly upper-layer excitatory neurons and microglia cells [111]. By integrating chromatin interaction and gene expression, GWAS results also support the enrichment of enhancer marks and ASD target genes during corticogenesis in the developing brain [58]. All these results emphasize the prenatal development of the cortical cerebrum is important in ASD pathophysiology.

2.3.9 Clinical and therapeutic implications

2.3.9.1 Genetic testing

With the progress of the human genome understanding, the application of genetic testing has grown enormously from rare disorders to a broader scope of complex disease and personal use [112,113]. Findings of the ASD research of molecular genetics have been widely applied in clinical studies [114,115]. Currently, Fragile X syndrome and chromosome microarray are recommended as the standard genetic testing for autistic individuals identifying around 15-20% carrying pathogenic CNVs of the ASD population [116,117]. A recent meta-analysis and consensus statement also highlighted exome sequencing outperformed chromosome microarray for unexplained NDDs including ASD, and should be placed at the beginning for evaluation [118]. Studies have also indicated individuals with more complex and severe morphological phenotypes have higher molecular diagnostic yield in comparison with milder

category [79,117,119]. Given an additive contribution of common SNPs on the ASD genetic liability, a polygenic score aggregating the risk of common variants could be considered a potential biomarker. However, due to low predictive ability at present, its application to clinical use is minimal [58]. Thus, current results of genetic testing are binary depending on whether a rare clinically significant variants can be found in the proband's genome. In clinical practice, currently, only a small part of the ASD population received clinical genetic testing showing a dissonance with recommendations [120]. Still, the application of genetic testing in most cases is merely for molecular diagnoses without additional usage including prevention or treatment.

2.3.9.2 *Interactions between gene and environment in ASD*

Both genes and environmental factors are associated with the onset of ASD, so it is crucial to explore their interactions and understand their roles for the disorder. Some environmental factors, such as vitamin D deficiency could increase the mutation rates and impair DNA repairing in ASD-associated genes [121]. Air pollutants, phthalates and maternal use of prenatal vitamins also interact with genetic variants to increase the risk of ASD [122–124]. In addition, the expression of genes which are essential for key molecular pathways in neurodevelopment could be regulated by environmental exposures. For example, folate can affect DNA methylation, and its deficiency can cause impairment of neurodevelopment [125]. Valproic acid, on the other hand, inhibits the histone deacetylase influencing the epigenetic function in the brain [126]. As such, evidence of the interplay between genetic and environmental factors could fundamentally help to evaluate genetic predisposition in the context of specific environmental factors in clinical practice. Another area is the exploration of PRS and environmental interaction. However, currently, there is no study focusing on NDDs and no significant evidence supporting their interaction in brain disorders including depression PRS and childhood maltreatment, as well as dementia PRS and lifestyle factors [127,128].

2.3.9.3 *Intervention prediction using genetic information*

Both genetic and clinical heterogeneity are high in ASD. Identifying the role of genetic information could be a promising way to improve the intervention effect for individuals. A few case studies [129–131] indicated a potential to use highly penetrant rare genetic variants as a target for the ASD intervention. In common variants, PRS has provided appealing results for behavioral and pharmacological treatments [56,65,132–134]. For instance, cognitive behavior therapy has shown an association with ASD PRS in major depressive disorder [65]. Additionally,

gene-set analysis has been used to unravel molecular processes and cellular functions for interventions, and it has been applied to interventions in psychiatry [135–139]. For example, the efficacy of short-term antipsychotic medication can be implied by genetic variations in glutamatergic or NMDA neurotransmission, and the calcium signaling pathway has been shown to associate with the response to selective serotonin reuptake inhibitors in obsessive-compulsive disorder [137,139]. Machine learning (ML) is a method to classify and make predictions in rich data using learning algorithms. Currently, it has been successfully applied to predict risk genes, pathogenic variants, and disorder diagnosis using genetic information in different traits, including ASD [140–142]. Until recently, ML methods have predicted treatment outcomes in psychiatric disorders such as depression and obsessive-compulsive disorder [143,144]. However, how to combine the various genetic information and clinical parameters and apply them to intervention choice is still a challenge in clinical settings.

2.4 UNSOLVED ISSUES IN ASD

2.4.1 Detected variants explain a small part of ASD genetic risk

The genetic risk of ASD is derived from a large number of common variants of additive contribution with a smaller proportion of rare de novo and inherited variations. Studies of ASD to date estimated the SNP heritability from 12% to 40% [43,58], accounting for a moderate part of the heritability from twin and family studies. PRS representing the SNP cumulative effect also explained low variance on ASD pathogenicity [58]. Although over hundreds of genes enriched by rare variants were discovered, most of them have been driven by *de novo* variants especially on coding regions [78–80,89]. Recently using genome sequencing data, Wilfert *et al.* found majority of rare inherited variants resided outside of genes enriched for *de novo* mutations contributing to at least 4.5% of autism risk in the population [92]. Moreover, the phenotypic variance explained by different types of variations is changeable in subgroups. Autistic individuals with ID have more risk for rare de novo variants in mutation-intolerant genes but less risk explained by the additive effect of common variants compared to those without ID [58,145]. In ASD subtypes, the heritability explained by common variants is higher in Asperger's syndrome than in childhood autism and unspecified pervasive developmental disorders [58]. Even by combining different types of known variants, the explained genetic risk is still limited. One possible explanation might be phenotypic heterogeneity in different study populations. However, Chaste *et al.* grouped individuals based on ASD sub-phenotyping and found that reducing phenotypic heterogeneity had at most a modest impact on genetic homogeneity [146]. With increased population power and more types of variations identified

14

including structural and non-coding variants, the role of genetic factors and their interactions is needed to evaluate further to explain the risk of genetic variation.

2.4.2 Unclear molecular and neuronal outcomes for many of the identified risk variants and genes in human neurons

Large-scale omics analyses have confirmed a large set of sequencing and structural variants with considerable effect conferring to ASD risk. However, biological understanding of these variants within genes or regions is still challenging in human neurons. Specifically, one of the hurdles is the difficulty to distinguish the relevant contribution of a single gene effect under varied multigenic backgrounds, especially its impact on social functioning [147]. For example, the expression of neurodevelopmental phenotypes in ASD individuals with 16p12.1 deletion was correlated with the number of other gene-disruptive variants suggesting an oligogenic or polygenic mechanism for variants contributing to ASD [148]. With numerous genes identified as ASD targets, large locus heterogeneity confuses whether identifying the putative function of an individual gene is helpful to understand ASD pathology in a broader insight. More complex, the widespread pleiotropy effects in ASD genes are more likely to have changeable expressivity and penetrance, which can contribute to phenotypic variability. Furthermore, as the human brain is a complicated and mostly inaccessible organ, model systems derived from human such as human embryonic stem cells and recent human induced pluripotent stem cells (iPSCs) can dissect basic molecular mechanisms related to human diseases [149]. Although it is relatively straightforward to determine the consequence of an ASD mutation in the model system, understanding how the system at the molecular-, cellular- or circuit-level links to specific features of social functioning is challenging. Also, how each variant interacts with distinct genetic backgrounds and environmental factors for phenotype presentation is needed to be answered.

2.4.3 Diverse intervention responses and limited genetic information used for drug development and individual treatment

Many autistic children are treated with medical interventions, but little to moderate levels of evidence support benefits for most of them [6,7,150]. Studies stated the different genetic, cognitive, and environmental aspects can be varied in the ASD phenotype which influences the effect of an intervention [151]. Additionally, other issues, including small sample sizes, scarcity of severely dysfunctional participants, and inconsistent outcome measures, affect the treatment efficacy in different clinical trials [151]. Understanding the biological basis is a crucial

step to facilitate the well-characterized distinction of ASD subgroups and develop more precise therapies. Although many studies have looked into the role of genetic factors, the effect measured is mainly limited to the cause of ASD. Currently, syndromic ASD and some nonsyndromic ASD cases with highly penetrant rare protein-truncating mutations are favorable to receive early interventions aiming at the genetic etiology to restore the normal protein levels. However, translating genetic findings into potential treatment to other nonsyndromic ASD cases will be a lot challenging, as a consequence of the clinical severity, the mutation frequency in the population, the reliability of the mutation displaying measurable developmental impairment, and the ability to influence the biological targeting in a clinically meaningful way [147]. There is no biomarker at present to classify patients based on genetic, biochemical, and circuit information. As numerous genes have been found to contribute to ASD risk, the genetic heterogeneity of ASD hinders the one-size-fits-all treatment. Even worse, one single candidate gene could also involve multiple targets for potential therapies [152]. Therefore, treating individuals with ASD based on every genetic mutation is formidable. One possible way is to identify the convergence upon biological pathways and spatiotemporal patterns across implicated genes, which can be aimed by a few of interventions [147,153,154] (Fig 2.4.1). Also, large population studies could continue to increase sample sizes to identify rare recurrent variants with large effects that are worthy to explore the intervention further [147].
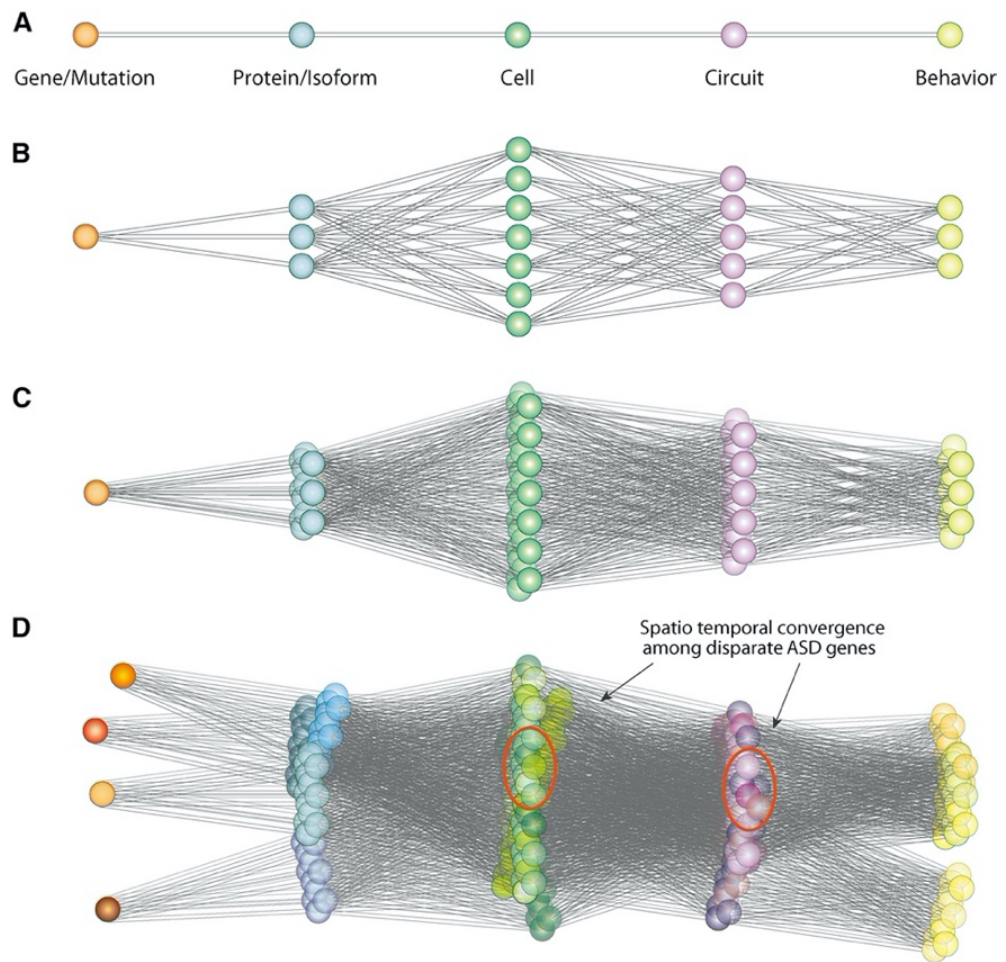
Figure 2.4.1. Spatiotemporal Convergence among High-Confidence ASD Risk Genes. The figure was obtained from Sestan *et al*. [154]

3 RESEARCH AIMS

This thesis is composed of two parts containing four studies:

3.1 STUDY 1-3

In the first part, three studies were included to understand the contribution of common and rare variants on intervention outcomes of social skills group training (SSGT) and standard care in the population of ASD (Fig 3.1.1). Some specific aims of this part are:

1) To identify molecular diagnoses in autistic individuals with normative intellectual abilities.

2) To estimate the role of common and rare variants on outcomes of SS

3) To examine the contribution of ASD-related biological pathways o

4) To predict the intervention responses with genetic and clinical info
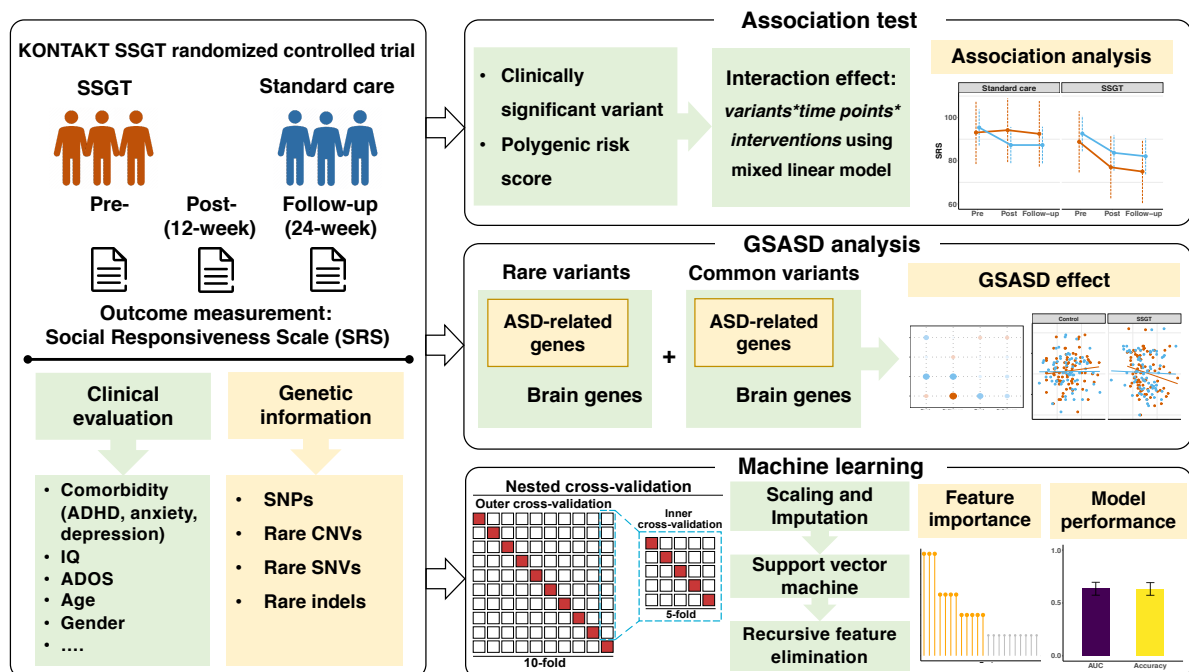

Figure 3.1.1. Overview and performed analysis in STUDY 1-3. SSGT: social skills group training; ADHD: attention deficit hyperactivity disorder; ADOS: Autism Diagnostic Observation Schedule; SNP: single nucleotide polymorphism; CNV: copy number variation; GSASD: genetic score for ASD-related gene sets; SNV: single nucleotide variant; indel: insertion or deletion.

3.2 STUDY 4

In the second part, the fourth study was to explore the effect of *CASK* mutations using induced pluripotent stem cell-derived neurons. Specifically,

1) To unravel the consequence of *CASK* mutations on gene transcription and interacting proteins.

2) To delineate neuronal effects of *CASK* pathogenic variants using iPSC-derived models.

4 MATERIALS AND METHODS

4.1 STUDY 1-3

4.1.1 Participants

All participants were recruited from a randomized controlled trial of social skills group training (SSGT) in Sweden (KONTAKT®, clinical trial identifier: NCT01854346) [8]. Briefly, all included participants were diagnosed with ASD following the criteria of International Classification of Diseases 10th Revision [155], had IQ > 70, and at least one common neurodevelopmental and psychiatric diagnosis including attention deficit hyperactivity disorder (ADHD), anxiety disorder, and depression. Both Children (7-12 years) and adolescents (13-17 years) were randomly assigned to either standard care only with any usual clinical intervention (n = 146), or standard care plus SSGT (n = 150). The primary outcome was recorded at baseline, post-intervention after 12-week and follow-up after 12-week post-intervention, using parent-reported Social Responsiveness Scale (SRS) total score with a higher score indicating lower social ability in certain aspects. Studies in this thesis included 207 participants (SSGT group: n = 105, standard care group: n = 102) who had an agreement of genetic part, saliva samples, and SRS measurement at pre-intervention, and either post-intervention or follow-up. As a part of the analysis, a binary outcome with a > 10% decrease of pre- to post-intervention in the total SRS score was classified as "response to interventions".

4.1.2 Genotyping and exome sequencing

Saliva samples were collected using a recommended procedure of the Oragene•DNA OG-500 tubes (DNA Genotek, Inc., Ottawa, Ontario, Canada) at home. DNA was extracted using Chemagen kit (PerkinElmer Chemagen, Baesweiler, Germany) with Chemagic- STAR®-robot (Hamilton Robotics, Reno, NV, USA). Genotyping was done on the Affymetrix CytoScan™ HD microarray platform (Santa Clara, CA, USA) containing 2.7 million probes for copy number variations (CNVs) and 743304 probes for common single nucleotide polymorphisms (SNPs). Genotyping data were transformed from Affymetrix .CEL format to .tped format using "Affy2sv" package v1.0.14 in R.

Genomic DNA (50 ng) from each saliva sample was used for library preparation with the Twist Human Core Exome v1.3 Enrichment Kit (Twist Bioscience). The following modifications: xGen Dual Index UMI adapters (6-nucleotide unique molecular identifiers (UMI), 0.6 mM, Integrated DNA Technologies) were used for the ligation, and xGen Library Amp Primer (2 mM, Integrated DNA Technologies) was used for polymerase

chain reaction (PCR) amplification. Target enrichment was performed in a multiplex fashion with a library amount of 187.5 ng. The libraries were hybridized to Exome probes v1.3 (Twist Bioscience), xGen Universal Blockers - TS Mix (Integrated DNA Technologies), and COT Human DNA (Life Technologies). The post-capture PCR was performed with xGen Library Amp Primer (0.5 mM, Integrated DNA Technologies). Sequencing was done on NovaSeq 6000 using paired-end 2 × 150 readouts. Demultiplexing was done using Casava v2.20. Exome sequencing was performed at the Clinical Genomics Stockholm core facility, Karolinska Institutet, and Science for Life Laboratory, Stockholm, Sweden.

4.1.3 Variants calling, quality control, and prioritization

4.1.3.1 Common variants

Quality control (QC) of the genotyping data was performed based on per-individual QC within each genotyping batch and on per-marker QC [156,157]. Individuals with discordant sex, heterozygosity rate > 3 standard deviation (SD), individual genotype failure rate > 0.03, and relatedness were removed. Ancestry of the participants was estimated using principal component analysis (PCA) based on the HapMap Phase III (HapMap3) data using EIGENSOFT v7.2.1[158]. Participants were restricted to the European ancestry. The first four principal component values of each individual were included in the statistical test for ancestry adjustment. As there was no detected batch effect, the following criteria were used to exclude low-quality markers: minor allele frequency < 0.05, individual missingness < 0.1, marker missingness < 0.05, and Hardy-Weinberg equilibrium < 1e-06. After marker QC, the final genotype data consisted of 539106 SNPs.

4.1.3.2 Rare CNVs

Calls from two algorithms Chromosome Analysis Suite (ChAS) software v.3.1 (Affymetrix), and Partek Genomics Suite software, version 6.6 (Partek Inc., St.Louis, MO, USA) were incorporated for CNVs identification. Variants called by both algorithms and spanned at least 25 kb and 25 consecutive probes were included for the following analyses [159]. CNVs with more than 0.1% frequency in the general population, and overlapped with more than five variants (50% reciprocal length) in the Database of Genomic Variants (DGV) were removed [159,160]. After that, CNVs were excluded if they had no overlap with any coding exons based on RefSeq annotation. The final included rare genic CNVs were classified according to their pathogenicity [161], and by three size groups: 25–100 kb (small), 101– 500 kb (middle),

and >500 kb including chromosome aneuploidies (large) [74,162]. Real-time PCR validation was then performed for 15 identified CNVs.

4.1.3.3   Rare SNVs and indels

After sequencing, Mutation Identification Pipeline (MIP) was used to conduct QC, reads alignment, variant calling, and annotations by Clinical Genomics Stockholm core facility, Karolinska Institutet, and Science for Life Laboratory, Stockholm, Sweden [163]. The average sequencing coverage of coding exons within OMIM genes across all subjects was between 77.9 and 98.2% at 10X sequence depth, and sequence depth of at least 10X and 20X separately covered 89.9% and 88.5% of the whole exome. After calling, variants were annotated by Ensembl Variant Effect Predictor (VEP) v92 and were removed if they were low quality or common (minor allele frequency > 0.1%). Rare variants with good quality were selected as rare deleterious variants if they were annotated as loss-of-function or damaging missense. Selected variants that were from developmental disorder-related genes and followed gene inheritance patterns were categorized as pathogenic/likely pathogenic, uncertain clinical effect, or benign/likely benign based on ACMG guidelines [164]. All pathogenic/likely pathogenic variants were classified as variants of clinical significance (VCSs). All VCSs and variants of uncertain significance (VUSs) were validated by Sanger sequencing.

4.1.4   Common variants imputation

After QC, SNPs were separated into autosomes, and haplotypes were inferred based on 1000 Genomes phase III haplotype data using SHAPEIT v2 [165]. Imputation was conducted in 5Mb windows for each phased autosome using IMPUTE2 v2.3.2 referencing 1000 Genomes phase III data [166,167]. Then all imputed regions were combined and filtered using the following criteria: info score < 0.8, minor allele frequency < 0.05, Hardy-Weinberg equilibrium < 1×10^{-6}, marker missing rate < 0.05, and individual missing rate < 0.1 [157,168]. In total, there were 5126694 SNPs after imputation QC for the polygenic risk score calculation (PRS).

4.1.5   Polygenic risk score calculation

The largest GWAS studies of ASD, ADHD, and educational attainment (EA) during the analysis time were used as reference data to calculate PRSs [58,169,170]. Only summary results from European ancestry individuals of each GWAS study were included. SNPs were pruned using clumping with a cutoff of $r^2 >= 0.1$ within a 500 kb window. PRS was calculated using

PRSice v2.1.4 with five P-value thresholds (Pts) (< 0.01, < 0.05, < 0.1, < 0.5, < 1) [171], and standardized (mean = 0, SD = 1) for association.

4.1.6 Gene sets calculation

At first, MAGMA v1.0.6 was used for gene and gene-set analyses of common variants [172]. Reference gene sets were acquired from a previous study [104], including 32 gene sets within five categories: synaptic, glial, FMRP, glutamate, and mitochondrial. The changes in the parent-reported SRS between post-intervention or follow-up and pre-intervention were used as outcomes. Age, sex, intervention methods, four largest principal components, and CNV carrier status were added in the model as cofactors.

To combine the effect of both common and rare variants, a genetic score for ASD-related gene sets (GSASDs) was constructed to analyze gene set specific genetic load for common (GSASD$_c$) and rare variants (GSASD$_r$). Basically, synaptic transmission (SyT) genes (GSSyT) (GO:0007268), regulation of transcription from RNA polymerase II promoter (Pol II) genes (GSPol) (GO:0006357), curated developmental disorder-related gene list, and genes from Simons Foundation Autism Research Initiative (SFARI, category 1, 2, and syndromic) were used as ASD-related gene sets for analyses. For rare variants, the number of genes having rare deleterious exome variants or rare CNVs in the ASD-related gene sets and brain expressed gene set were used. The number of ASD-related genes captured by these rare variants in each ASD-related gene set was added and divided by rare variants affected brain-expressed gene numbers in each individual (Equation 4.1.1).

$$GSASD_r \;=\; \frac{ASD_{rc}+ASD_{re}}{Brain_{rc}+Brain_{re}} \qquad (4.1.1)$$

- Where ASD$_{rc}$ and ASD$_{re}$ is the number of genes from ASD-related gene sets affected by rare CNVs and rare deleterious exome variants. Brain$_{rc}$ and Brain$_{re}$ is the number of brain-expressed genes affected by rare CNVs and rare deleterious exome variants.

For common variants, set-based PRS (https://www.prsice.info/quick_start_prset/) was calculated based on the largest ASD GWAS study [58]. The ratio between each ASD-related set PRS and brain PRS was represented as GSASD for common variants (Equation 4.1.2).

$$GSASD_c \;=\; \frac{ASD_{PRS}}{Brain_{PRS}} \qquad (4.1.2)$$

4.1.7 Statistical analyses

4.1.7.1 Distribution of different types of variations in individuals

Student's t-test was used to test the difference of ASD PRS in individuals with or without clinically significant CNVs. The distribution of carriers with clinically significant CNVs and exome VCSs/VUSs was examined by Pearson's χ^2 test. The distribution of PRS for ASD between individuals with or without VCSs/VUSs was assessed using one-way ANOVA.

4.1.7.2 Association of intervention outcomes

Mixed linear models were applied to assess the effect of different types of variants on intervention outcomes measured by parent-reported SRS total score. In general, a three-way interaction: variants*time points*interventions together with lower-order interactions were added in the model adjusting for age and sex (Equation 4.1.3). The clinics and individual IDs were considered random factors in the model. For rare CNVs, carrier status of all rare genic CNVs, clinically significant CNVs, and large size CNVs were tested separately. PRSs for ASD, ADHD and EA were used separately for interactions with time points and interventions after adjusting CNV carrier status and other cofactors. For rare exome variants, the interaction effect of VCSs/VUSs carrier status with time points and interventions was examined adjusted for clinically significant CNVs and PRS for ADHD based on previous two model results. Other factors including sex, age and random factors were also included in the model. To measure the effect of different variation types on SSGT and standard care, a two-way interaction of variants*time points was used with the same mixed linear model of each type of variants in SSGT and standard care subgroups. Furthermore, linear regressions were performed as a secondary model for ASD PRS with VCSs/VUSs and clinically significant CNVs, respectively, to test if different types of variants independently affected the intervention outcomes. The changes of SRS total score at post-intervention and follow-up compared to pre-intervention were used as outcomes. Three-way interactions: VCSs/VUSs*PRS*interventions, as well as clinically significant CNVs*PRS*interventions were added separately, with sex and age adjusted in the models.

$$\text{SRS total score} \sim \text{variants*time points*interventions} + \text{age} + \text{sex} +$$
$$(1 \mid \text{clinic/individual}) \tag{4.1.3}$$

CNV: rare genic CNV, clinically significant CNV, large size CNV
PRS: PRS for ASD, ADHD, EA
Exome variant: VCS, VUS

Similarly, to investigate if GSASD was associated with the intervention outcomes, a mixed linear model was applied combining two three-way interactions of

GSASD$_r$*time points*interventions and GSASD$_c$*time points*interventions with all lower-order interactions, age, sex as fixed factors, and the same random factors as earlier. Furthermore, linear regression was also examined for the interaction of GSASD$_c$ and GSASD$_r$ using changes in SRS total score as the outcome with sex and age as cofactors.

4.1.8 Machine learning prediction

A binary outcome representing the changes in the parent-reported SRS total score was used in the machine learning prediction model. The responders were categorized as a decrease >10% of SRS score at post-intervention, while non-responders were the other participants that did not fulfill this requirement. Prediction features in the model included participant characteristics, comorbidity diagnoses, interventions, autism-related scales assessed at baseline, and genetic information of both rare and common variants (Table 4.1.1). A linear support vector machine (SVM) was applied using recursive feature elimination to select the most important features. Nested cross-validation with 10-fold outer cycle and 5-fold inner cycle was implemented for model training and validation. The area under the receiver operating characteristic curve (AUC) was calculated as the main measurement of model performance. Other metrics, such as sensitivity, specificity, and accuracy, were also evaluated during the cross-validation.

Table 4.1.1. All features used in the machine learning model

Feature name	Type
Sample characteristics	
IQ	Integer
Age	Integer
Sex	Binary (0 = female, 1 = male)
Comorbidities	
Attention deficit hyperactivity disorder (ADHD)	Binary (0 = no, 1 = yes)
Depression	Binary (0 = no, 1 = yes)
Anxiety	Binary (0 = no, 1 = yes)
Others	Binary (0 = no, 1 = yes)
Treatments	
Social skills group training (SSGT)	Binary (0 = no, 1 = yes)
Cognitive behavioral therapy (CBT)	Binary (0 = no, 1 = yes)
Pharmaceutical treatment	Binary (0 = no, 1 = yes)
Counsel treatment	Binary (0 = no, 1 = yes)
Scales	
Social Responsiveness Scale (SRS)	Integer
Autism Diagnostic Observation Schedule (ADOS)	Integer

Developmental Disabilities Modification of Children's Global Assessment Scale (DD-CGAS)	Integer
Adaptive Behavior Assessment System (ABAS)	Integer
Clinical Global Impression (CGI)	Integer
Genetics	
Rare copy number variation (CNV) size	Integer
Number of genes in rare CNVs	Integer
Carrier status of clinically significant rare CNVs	Binary (0 = non-carrier, 1 = carrier)
Carrier status of large size rare CNVs	Binary (0 = non-carrier, 1 = carrier)
Polygenic risk score (PRS) for ASD	Float
PRS for ADHD	Float
Carrier status of rare exome VCS/VUS	Category (0 = non-carrier, 1 = VUS, 2 = VCS)
Number of rare damaging SNVs/indels	Integer
Number of developmental disorder-related rare SNVs/indels	Integer
GSSyT$_r$	Float
GSSyT$_c$	Float
GSPol$_r$	Float

Abbreviations: VCS: variant of clinical significance; VUS: variant of uncertain significance; SNV: single nucleotide variation; GSSyT$_r$: genetic score of rare variants for synaptic transmission genes; GSSyT$_c$: genetic score of common variants for synaptic transmission genes; GSPol$_r$: genetic score of rare variants for regulation of transcription from RNA polymerase II promoter genes

4.2 STUDY 4

4.2.1 Identification of mutation carriers

Two mutation carriers were recruited from The Roots of Autism and ADHD study in Sweden (RATSS) and Clinical Genetics at the Karolinska University Hospital. Details of the ASD cases and his cotwin, as well as the girl with MICPCH has been described elsewhere [173–175]. Microarray for cases and exome sequencing for the twins and their parents were performed using extracted DNA from whole blood and saliva samples. CNV detection has been reported earlier [175,176]. Clinically relevant variants from exome sequencing were categorized based on ACMG guidelines [164].

4.2.2 Cell culture of iPS and neuroepithelial-like stem (NES) cells

Human iPSC cells were obtained from the two *CASK*-mutation carriers diagnosed with MICPCH and ASD, respectively [177]. NES cells were derived from iPS cells using dual-SMAD inhibition [178,179], and were seeded on plastic surface, precoated with 20 μg/ml polyornithine and 1 μg/ml laminin in DMEM/F12+ glutamax medium supplemented with

0.5× B-27, 1× N-2, and 10 U/ml penicillin/streptomycin. Cells were maintained in a 5% CO_2 atmosphere at 37 °C and changed every second day of differentiation. Neurons were differentiated on day 8, 16, or 28.

4.2.3 RNA sequencing

4.2.3.1 Single-cell RNA sequencing

Cells were dissociated and resuspended, and single cells were sorted by size into lysis buffer. Library preparation and sequencing of Smart-Seq2 were performed by the Eukaryotic Single Cell Genomics facility in the SciLifeLab, Stockholm. After sequencing, there were 384 sorted wells and 209.6M reads in total with an average 550000 reads of sequence depth per cell. Dropout expression values were imputed using the scImpute package from the count matrix of per gene and cell (GSE140572) [180]. Cells were excluded if they had less than 50000 read counts in less than 2000 genes. Gene expression differentiation and cell clustering were calculated using the Seurat package [181].

4.2.3.2 Bulk RNA sequencing

RNA samples were obtained per cell line seeded at different dates and different passages. All samples were delivered for library preparation and sequencing at SciLifeLab National Genomics Infrastructure Sweden. There were 50.4 million reads on average per sample and at least 96.9% reads aligned to protein-coding regions. The DESeq2 v1.24.0 package was used to calculate differential gene expression based on gene counts from the count matrix (GSE140572). ASD CASK case with male control and MICPCH CASK case with female control were compared separately to obtain differential expressions. Top ranking genes were selected with adjusted P value of 1e−5 (Benjamini-Hochberg), more than 20 reads and log2 fold change greater than 0.5. Gene set enrichment analysis (GSEA) was performed using the ranked gene expression list to obtain the enriched sets which were visualized in Cytoscape v3.7.0 with Enrichment Map v3.1.0 and AutoAnnotate [182]. Additionally, GSEA was done on genes from CASK protein-protein interactions (download from Pathway Commons PCViz: CASK), in-house curated developmental disorder-related gene list and SFARI gene sublist.

4.2.4 Immunofluorescence and protein quantification

Cells were cultured on glass coverslips and fixed in formaldehyde. Primary used antibodies were CASK-NBP2-41181, 1:500 (Novus bio), MAP2-M2320, 1:500 (Sigma Aldrich),

VGLUT1-135304, 1:250 (Synaptic systems), Homer-1-160011, 1:250 (Synaptic systems), VGAT- 131003, 1:500 (Synaptic systems), Synapsin-1/2-106006, 1:500 (Synaptic systems), Gephyrin-147021, 1:250 (Synaptic systems), Nestin-MAB5326-KC, 1:1000 (Merck-Millipore), and SOX2-AB5603, 1:1000 (Merck-Millipore). All images were taken using LSM 700 Zeiss Confocal Microscope (Zeiss Plan-Apochromat 63×/1.40na Oil DIC Objective M27), with 63× magnification at 1024- × 1024- pixel resolution. CASK puncta and cell nuclei were counted by ImageJ Particle Analyzer and quantification of particle number and size was calculated by R. Synaptic marker particle size and number were quantified with the ImageJ plugin Synapse Counter.

Cells were dissociated and lysed, and protein from cells was loaded for Simple-Western WES (ProteinSimple) quantification with antibodies for CASK (1:500 Novus Biologicals NBP2-41181), beta-actin (1:100 Abcam ab8227), and GAPDH (1:5000 Sigma G9545). Protein-specific peaks were identified by Compass Software For Simple Western v4.0.0 and used for protein quantification. CASK protein was then normalized for the house-keeping protein. Statistical significance between cell lines was tested by ANOVA and post hoc Tukey HSD in R.

5 RESULTS

5.1 STUDY 1-3

To identify the contribution of common and rare variants on intervention outcomes of social skills group training (SSGT) and standard care in ASD, participants from the Swedish randomized controlled trial of SSGT (KONTAKT®) who had saliva samples and recorded outcome measures were included in STUDY 1-3. After removing low quality samples, 207 individuals were included in genetic analyses.

5.1.1 Molecular diagnoses of clinically significant variants

There were 71 (34.8%) individuals who carried at least one rare genic CNV ≥ 25 kb, of which 17 (8.2%) individuals carried at least one clinically significant CNVs. After excluding samples with low quality of exome sequencing, in the remaining 204 individuals, nine variants of clinically significance (VCSs) from nine carriers (4.4%) and 24 variants of uncertain significance (VUSs) from 23 carriers (11.3%) were identified. Except one individual carried a VUS plus a VCS, and another individual had two VUSs, the remaining 29 carriers had either one VCS or VUS.

5.1.2 Correlation with polygenic risk scores and clinically significant variants

Combining the information of PRS and rare variants, no difference was observed in ASD PRS between neither exome VCS/VUS carriers and non-carriers (P = 0.37, Fig 5.1.1a), nor carriers of clinically significant CNVs and non-carriers (P = 0.72, Fig 5.1.1b). Two individuals carried both VUSs and clinically significant rare CNVs, but no overlap between carriers of clinically significant rare CNVs and VCSs (P = 0.65, Fig 5.1.1c). The distribution of ASD PRS and carriers with clinically significant rare CNVs and VCSs/VUSs were presented in Fig 5.1.1d.
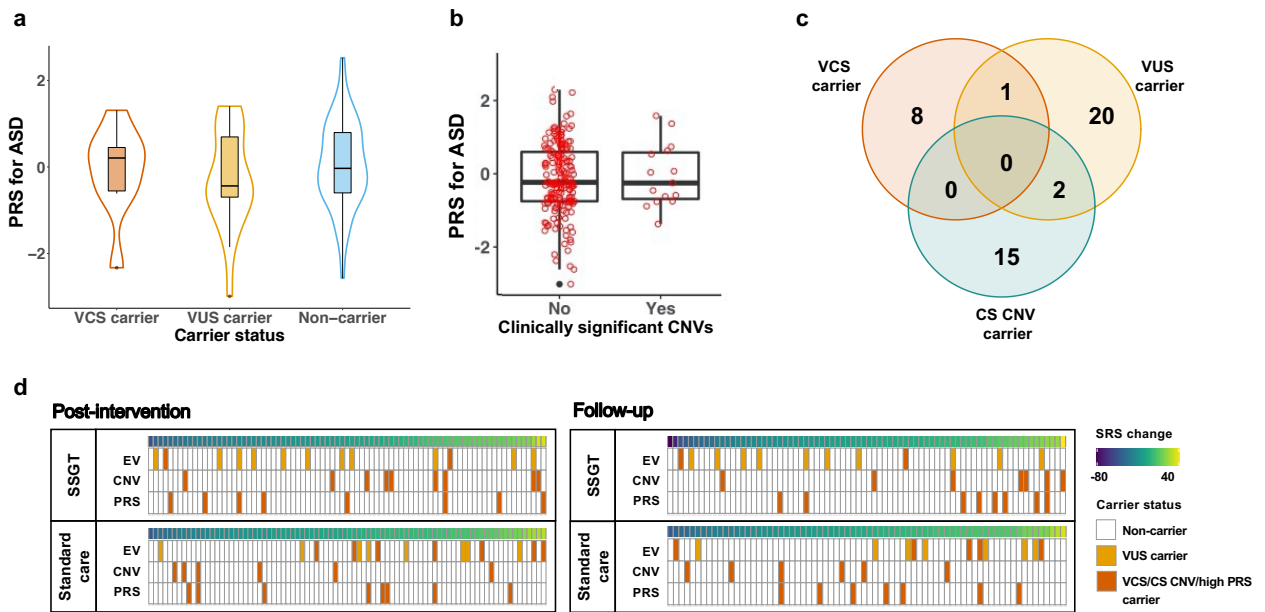
Figure 5.1.1. Distribution of the prioritized genetic variants within the study cohort. **a**. The distribution of ASD PRS in VCS/VUS carriers and non-carriers **b**. The distribution of PRS for ASD in individuals with and without clinically significant CNVs. **c**. The number and overlap of carriers with VCSs/VUSs and clinically significant (CS) CNVs. **d**. Distribution of CS CNVs, exome VCSs/VUSs, and 10% highest PRS for ASD carriers. EV: exome variant, high PRS carrier: an individual with the 10% highest PRS for ASD in the study cohort.

5.1.3 Association between intervention outcomes and carrier status

To begin with, the association of intervention outcomes with rare CNVs such as clinically significant rare CNVs were tested. Carriers of clinically significant CNVs had less improvement compared to non-carriers in the whole trial or SSGT subgroup at post-intervention (whole trial: $\beta = 17.34$, P = 0.047) and follow-up (whole trial: $\beta = 23.64$, P = 0.008; subgroup: $\beta = 16.59$, P = 0.010) (Fig 5.1.2a). For the association of PRS, inferior outcomes were shown in individuals with higher ASD PRS and ADHD PRS in the SSGT group compared to the standard care at follow-up (ASD: $\beta = 6.47$, P = 0.019, ADHD: $\beta = 6.67$, P = 0.016, Fig 5.1.2b). Results of the two-way interaction indicated overall better outcomes of higher ADHD PRS after interventions (post-intervention: $\beta = -4.747$, P = 0.0129; follow-up: $\beta = -5.309$, P= 0.0083). Subgroup association in the SSGT and standard care demonstrated higher ADHD PRS was associated with better outcomes after standard care intervention (post-intervention: $\beta = -4.729$, P = 0.00647; follow-up: $\beta = -5.277$, P = 0.00394). Combining VCS and VUS, individuals carrying the VCSs/VUSs tended to have less improvement at post-intervention in the whole cohort ($\beta = 9.22$, P = 0.057, Fig 5.1.2c). Similarly, VCS carriers significantly improved less at post-intervention compared to non-carriers ($\beta = 14.84$, P = 0.029). No association was found with changes of SRS total score in the three-way interaction (Fig 5.1.2c). Further subgroup analysis confirmed inferior outcomes were associated with VCS carriers as well as the combined VCS and VUS carriers

compared to non-carriers in the standard care group at post-intervention (VCS carriers: $\beta = 14.86$, P = 0.017; VCS and VUS carriers: $\beta = 9.35$, P = 0.036). Secondary linear models observed no interaction of ASD PRS with carriers of clinically significant rare CNVs or VCSs/VUSs, demonstr[...]d rare clinically significant variants on S[...]
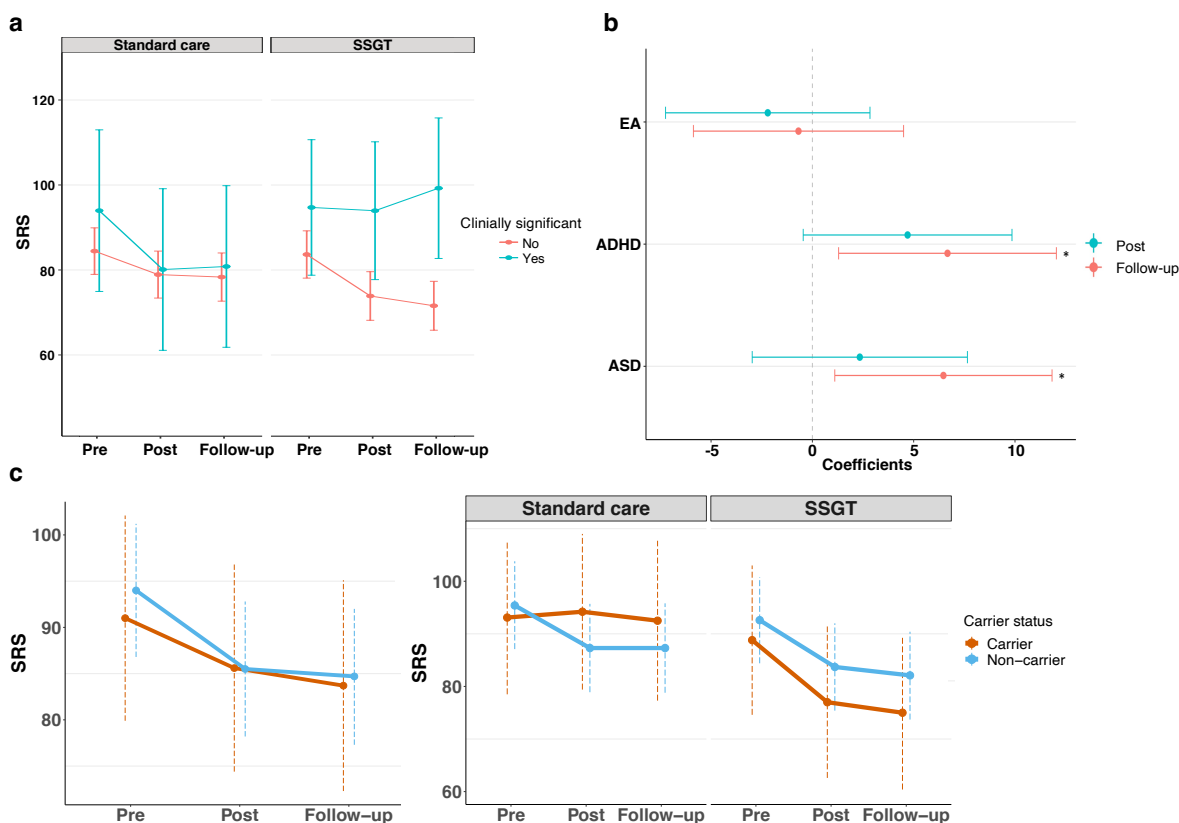

Figure 5.1.2. Association between different genetic variations and intervention outcomes. **a.** Least-square m[...] score in the standard care and SSGT groups stratified by [...]Vs. **b**. Association of PRS for ASD, ADHD, and ed[...]w-up represented by correlation coefficients and 95% co[...] VCS+VUS carrier status and intervention outcomes sh[...]

intervention outcomes. Five gene sets were [...] post-intervention (intracellular signal transdu[...], excitability, and GPCR signaling) and follow-up (RNA and protein synthesis, folding and breakdown), all belonging to the synaptic group.

Next, to test the association of both common and rare variants enriched gene sets and biological pathways with intervention outcomes, a pilot scheme was generated to calculate gene set scores from ASD-related gene sets (GSASDs) on rare (GSASD$_r$) and common variants (GSASD$_c$), including synaptic transmission genes (GSSyT), regulation of

transcription from RNA polymerase II promoter genes (GSPol), developmental disorder-related genes, and genes from Simons Foundation Autism Research Initiative (SFARI). Using a mixed linear model, SSGT had significantly better effect among individuals with higher $GSSyT_c$ and $GSPol_r$ compared to standard care (post-intervention: $GSSyT_c$: $\beta = -5.52$, $P = 0.033$; $GSPol_r$: $\beta = -6.38$, $P = 0.018$; follow-up: $GSPol_r$: $\beta = -6.79$, $P = 0.014$, Fig 5.1.3). However, a smaller effect of SSGT was observed for individuals with higher $GSSyT_r$ compared to standard care (follow-up: $\beta = 8.30$, $P = 0.0044$, Fig 5.1.3). The significance remained in the results of $GSSyT_r$ and $GSPol_r$ after multiple testing.
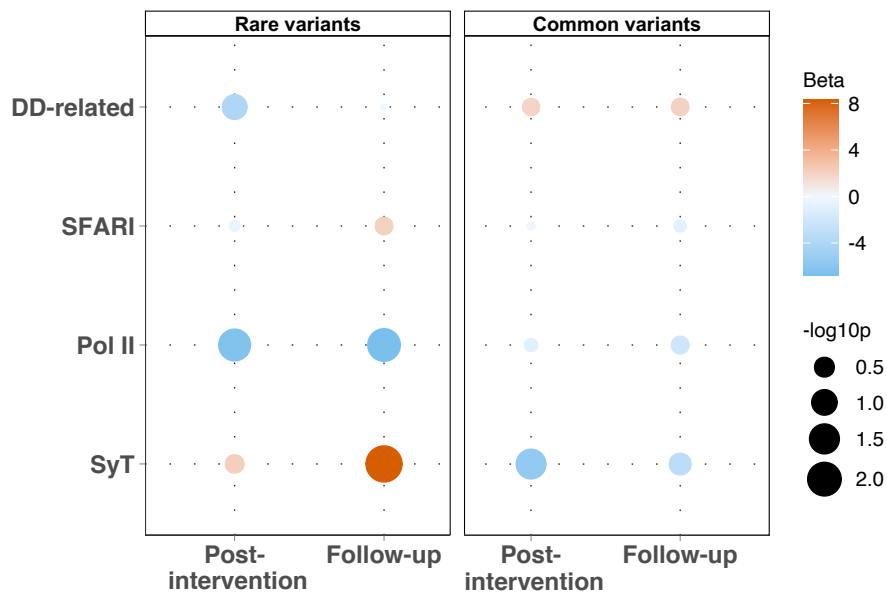

Figure 5.1.3. The effect of rare and common variant genetic scores on intervention outcomes at post-intervention and follow-up. Included gene sets were synaptic transmission (SyT), regulation of transcription from RNA polymerase II promoter (Pol II), Simons Foundation Autism Research Initiative (SFARI), and developmental disorder-related (DD-related) gene list.

5.1.5 Intervention response prediction using machine learning

Lastly, to explore whether the individual intervention response could be predicted, a machine learning model was developed by combining clinical and genetic information from the KONTAKT®. After evaluation, linear support vector machine (SVM) was applied on prioritized twelve features. The average area under the receiver operating characteristic curve (AUC) and accuracy for the model was 0.632 (SD 0.0599) and 0.634 (SD 0.0610). The sensitivity and specificity for linear SVM to predict individuals' responses was 60% (SD 0.128) and 66.4% (SD 0.131), respectively.

5.2 STUDY 4

5.2.1 Mutant transcripts and reduction of CASK expression in maturing neurons from mutation carriers

This study included two families with two different *CASK* genetic mutations, representing ASD and MICPCH disorders. A splice-site mutation and a de novo tandem duplication in *CASK* were identified and referred to as ASD_{CASK_SS} and $MICPCH_{CASK\_dup4/5}$, respectively. The fibroblasts from individuals were transformed into iPSCs and were further differentiated to self-renewing neuroepithelial-like stem (NES) cells. The expression of *CASK* variants was investigated during the differentiation of NES cells to neurons. ASD_{CASK_SS} cells showed reduced expression of mRNA (P = 0.02) and protein (P = 0.01) at day 28 compared to controls. $MICPCH_{CASK\_dup4/5}$ cells showed increased mRNA instead of protein levels until day 16 (day 0 P < 0.001, day 8 P < 0.001, day 16 P = 0.001) and leveled with controls on day 28 (P = 1). Using the N-terminal binding antibody, CASK puncta had similar number per nuclei in all cell lines (ASD_{CASK_SS} vs. controls: P = 0.99, $MICPCH_{CASK\_dup4/5}$ vs. controls: P = 0.83) but was smaller in ASD_{CASK_SS} neurons compared to controls (P = 0.049) and $MICPCH_{CASK\_dup4/5}$ (P = 0.011).

5.2.2 Reduced CASK levels alter presynaptic development and E/I balance

From bulk RNA sequencing, pooled analysis revealed upregulated pathways involved extracellular matrix components, morphogenesis and development of different tissues, and WNT and BMP signaling pathways (Fig 5.2.1a). The downregulated pathways were related to presynaptic synaptic vesicle components (Fig 5.2.1b). CASK protein-protein interactions (PPIs) were significantly enriched in down-regulated genes (P = 0.003, FDR = 0.003) instead of upregulated genes (P = 0.17, FDR = 0.25, Fig. 5.2.1c, d). Furthermore, by evaluating the synapse morphology through immunofluorescence co-staining, the particle size of postsynaptic excitatory marker Homer-1 and inhibitory marker Gephyrin was comparable between cell lines, but the size was significantly smaller for the presynaptic marker Synapsin-1/2 and the inhibitory presynaptic marker VGAT in ASD_{CASK_SS} compared to the male control (Synapsin-1/2: P = 0.021, VGAT: P = 0.018) and $MICPCH_{CASK\_dup4/5}$ (Synapsin-1/2: P = 0.014). However, excitatory presynaptic marker VGlut showed a similar size in all cell lines.

A   **Upregulated pathways**

B   **Downregulated pathways**

C

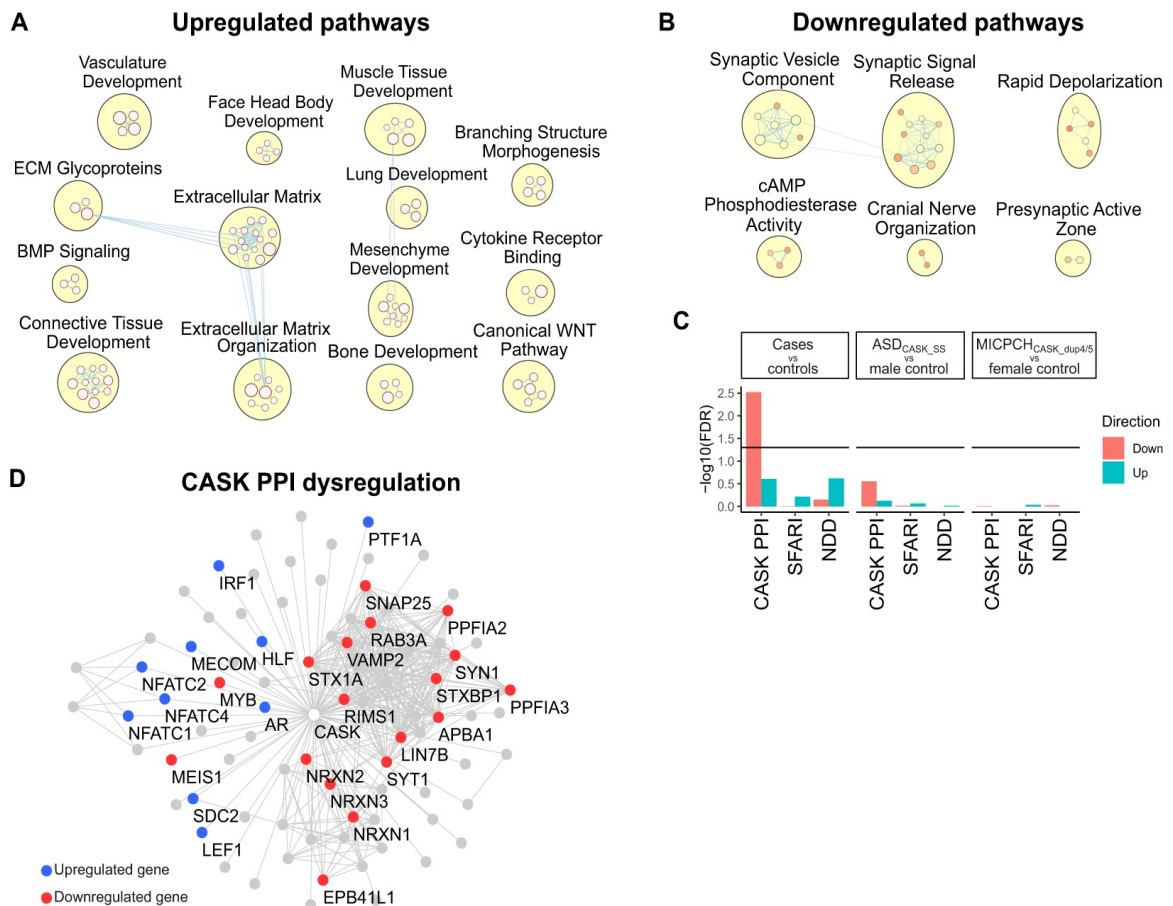**D**   **CASK PPI dysregulation**

Fig. 5.2.1. Consistent dysregulation of presynaptic and CASK-interacting genes in bulk RNA sequencing after 28 days of differentiation. **a**. Upregulated and **b.** downregulated gene sets emerging from GSEA in bulk RNA-sequencing data from mutation carriers compared with controls. **c**. GSEA for CASK-interacting proteins (CASK PPI), ASD risk genes (SFARI), and NDD genes (NDD) in bulk RNA-sequencing data from mutation carriers compared with controls. **d**. Network of CASK-interacting proteins with upregulation marked in blue and downregulation in red.

6 DISCUSSION

6.1 MOLECULAR DIAGNOSIS IN ASD INDIVIDUALS

In this thesis, autistic individuals with normal intellectual ability were confirmed to carry clinically significant variants. Of them, 4.4% of individuals carried clinically significant variants detected by exome sequencing and 8.2% had clinically significant rare CNVs. Both of the rates are within the range of earlier studies [79,119,183]. As no individuals with intellectual disability were included, the detected variants rate was slightly lower as indicated previously [79,119,183]. After combining the results, the overall proportion of identifiable genetic etiology was 12.6% in the cohort, suggesting intellectually abled autistic individuals should be capable of receiving genetic testing. Some interesting clinically significant variants implicated in ASD and social functioning were detected. For example, some large CNVs affected known risk CNV loci such as 15q11.2–q13.1, 9p24.3–p23 and smaller CNVs affected ASD susceptible genes including *GATAD2B, CHD8* and *KDM6A*. Exome sequencing variants from *CUL3* can contribute to ASD, social deficits, and anxiety-like behaviors [86,184]. As no carrier carried both kinds of clinically significant variants, a complementary role was also shown for chromosome microarray and exome sequencing in increasing the molecular diagnostic yield for autistic individuals.

6.2 RELATIONS BETWEEN RARE AND COMMON VARIANTS AND THEIR EFFECT ON INTERVENTIONS

An independent distribution of common and rare variants was indicated in the study population, in which the PRS for ASD showed no difference in individuals with or without pathogenic variants. Similar findings were also found in other studies using much larger cohorts in ASD and severe NDDs [61,93]. Moreover, association analyses also revealed no interaction effect for different types of variants including PRS for ASD with clinically significant rare CNVs and exome variants. Therefore, these results implied an additive or differential role of different variation types for the intervention outcomes and should be validated further.

Additionally, an inconsistent effect of different ASD-related pathways was found on interventions, such as synaptic transmission and regulation of transcription from RNA polymerase II promoter. These varied results may suggest complex pathway functions and connections for the interventions. However, due to the small sample size and limited evidence supporting the biological mechanism for specific intervention from other clinical trials [185],

further investigations are needed to explore the function of genes and pathways for different ASD treatments.

6.3 MUTATED GENES INFLUENCING CELLULAR FUNCTIONING RELATED TO NDDS

Applying iPSC-derived neurons from two *CASK*-mutation carriers, reduction of *CASK* was demonstrated to affect the transcription of presynaptic function and CASK protein interactors. Especially, genes interacting with *CASK* including the *NRXN* gene family, and *STXBP1* were downregulated, providing a phenotypic link to other NDD genes. Similarly, CASK interacted proteins were also dysregulated in heterozygous Cask-knockout mice [186]. Besides the significant disruption in the presynapse, CASK-interacting partners also showed dysregulation in other neuronal compartments including dendrites, and postsynapse. In addition, many upregulated developmental pathways were also identified, such as WNT signaling, mesenchyme, face and head development, which could be investigated to characterize their links to CASK-related phenotypes [187]. Therefore, *CASK* deficiency could result in NDD-related networks abnormality, especially on the presynapse development and the putative influence of additional neuronal functions.

Furthermore, using the presynaptic marker Synapsin-1/2 and the inhibitory marker VGAT, transcriptional changes induced by *CASK* mutations contributed to a smaller presynaptic size of inhibitory neurons. This reduction could indicate abnormal synaptogenesis and synaptic function, and lead to E/I imbalance. Decreased miniature inhibitory postsynaptic currents and increased miniature excitatory postsynaptic currents have been demonstrated in adult CASK-deficient neurons using electrophysiological measures [101,188]. Also, a postsynaptic deficiency of CASK could decrease the expression of glutamate receptor GluN2B and cause E/I imbalance [188]. Compared with the reduction of inhibitory presynapse, excitatory synapses measured by VGlut were unchanged in this study, suggesting the reduction of inhibitory presynapse is the primary mechanism to cause E/I imbalance in *CASK* mutation neurons.

6.4 COMBINE GENETIC AND CLINICAL EVALUATION TO IDENTIFY PHENOTYPE VARIETY AND PREDICT INTERVENTIONS

Using the KONTAKT® trial data, a machine learning model was built integrating measurements from clinical and genetic aspects to predict the intervention responses. Among various features, information including complexity and severity of the disorder, interventions,

and genetic moderators especially rare variants related characteristics, were selected that contributed most to the outcomes. Although the performance was not perfect, the tested model implied a potential to use related features for intervention outcomes prediction, which could further help clinicians and individuals with ASD to evaluate whether adding SSGT as an add-on of standard care could have a better response. The most important features of the prediction model could also be studied for the guidance of intervention decisions in the clinical practice, and the performance could be referenced by future studies for the improvement of outcome prediction. Due to the limited sample size, it was insufficient to test models for the subgroup of each intervention in the study. Some possible ways for the improvement of prediction performance could be increasing the sample size in the cohort, adding more evaluations of imaging, behavioral measures, and extracting different levels of genetic data to increase the liability and explain more variance of outcome changes.

6.5 LIMITATIONS

6.5.1 STUDY 1-3

Although participants of these three studies were from the largest clinical trial of SSGT to date, more samples are needed to identify the effect of single locus and associated genes as well as to stratify the variant effect in subgroups. The study population all had normative IQ and at least one neurodevelopmental and psychiatric comorbidities, so it limited the results translation to autistic individuals with different manifestations. Nevertheless, this thesis could provide inspiration for studies to explore more in-depth the heterogeneity of ASD interventions from a genetic perspective. In addition, both inherited and *de novo* variants were confirmed to be essential for ASD liability and were linked to cognitive phenotypes including IQ [81]. However, due to a lack of parental genetic information, the inheritance pattern of variants could not be differentiated. As varied interventions have been used to help individuals with ASD, integrating samples from other clinical trials is impractical to validate the genetic effect obtained from this thesis. Except for the KONTAKT® trial, genetic data are unavailable from other finished and ongoing SSGT clinical cohorts currently [189,190]. With increased genetic testing in the clinic, combining clinical evaluation and biometric information, genetic data could be applied to primary diagnoses and treatment to assist the clinical decisions and long-term management proactively [113]

6.5.2 STUDY 4

Based on the in vitro stem cell model, results from the study provided limited information to precisely map results to related brain regions and developmental periods to identify the variant effect at the individual level. As only the reduction effect of CASK protein was studied, the phenotype of CASK missense mutations at the cellular level remains to be understood further. Longitudinal studies are also required to examine if the CASK mutation effects are persistent in delaying neuronal maturation and influencing inhibitory presynaptic development.

6.6 ETHICAL CONSIDERATION

In STUDY 1-3, the clinical trial and sample collection were approved by the ethical review board in Stockholm (Dnr 2012/385-31/4). Since the participants in these studies were children and adolescents above 7 years, informed consent was obtained from both parents or legal guardians and children (8-12 years) /adolescents (13-17 years) after they received detailed written and verbal information about the treatment content, structure, and time plan in an intake interview. As participants were autistic, interviews, clinical judgment, medical records, and discussion with legal guardians were conducted to determine if the participants could provide informed consent. During the intervention, all participants had rights such as the ability to decline study procedures, withdraw participation, and the right to obtain information related to them. Responsible researchers ensured that information about participants would only be accessed by persons who were employed and who worked within the research project. Biological samples were voluntarily collected after receiving consent from participants. Saliva samples were collected from participants using a non-invasive method. All the samples were stored in KI biobank. Information from individual genetic data has been explained to participants by clinical geneticists. After finishing all the analysis, results are in processed of being returned to families who want to receive feedback. Only people who were responsible for the data analysis were allowed to access the database. As clinical and genetic data from the trial included personal information, according to general data protection regulation in Europe, all the data were stored and analyzed in a safe-protected server, and this requirement will be followed if the data is transferred to other places in the future.

STUDY 4 was approved by the regional and national ethical boards in Sweden (Dnr 2012/208-31/3, Dnr 2016/1452-31, Dnr 2015/1097-31/1). Written informed consent was acquired from both affected individuals and their parents. This study followed the

Declaration of Helsinki for medical research involving human subjects, including research on identifiable human material and data. Considering the individual privacy, information about participants was only accessible by persons within the research project. All the cell lines were stored in the Tammimies laboratory with standard procedure. Expression summary data of RNA and single cell sequencing was submitted after all analyses and openly published online.

7 CONCLUSIONS

7.1 STUDY 1-3

Using exome sequencing and chromosome microarray, molecular diagnoses can be identified in autistic individuals with normative intellectual abilities. Genetic information including clinically significant rare variants and common variant polygenic scores are significantly associated with intervention outcomes with varied effects for both SSGT and standard care. Gene set analyses combining rare and common variants reflected some potential pathways including synaptic transmission and transcription regulation from RNA polymerase II promoter, which could be explored to stratify biologically meaningful subgroups for the determination of intervention outcomes. Furthermore, the machine learning prediction model confirmed that the individual outcomes are predictable by using clinical and genetic features such as rare variants. In the future, genetic information from the variant-, gene- and circuit-level could be considered in the ASD clinical trials to improve the use of molecular genetic testing beyond diagnoses and help shape the individualized intervention plan.

7.2 STUDY 4

Reduced CASK protein levels can contribute to aberrant presynaptic development and decreased inhibitory presynapse size, which may lead to the E/I imbalance in developing neural circuitries. Targeting on presynapses and E/I imbalance in future clinical and pharmacological studies could be developed for treatments of CASK-related disorders.

8 POINTS OF PERSPECTIVE

8.1 MORE DETAILED PHENOTYPE CLASSIFICATION ADDING MOLECULAR OR GENETIC INFORMATION IN NDDS

Despite clear diagnostic criteria in ASD, clinical manifestations are broadly different among individuals. Genetic variability is one of the causes for phenotype heterogeneity. For example, Heyne *et al*. revealed 33 genes specifically associated with epilepsy comorbidity when examining the NDD population with or without epilepsy. Similarly, among 102 risk genes identified from the largest exome sequencing study of ASD, almost half of them showed higher frequencies in individuals ascertained to have a severe neurodevelopmental delay, whereas the other genes tended to be more frequent in individuals ascertained with ASD [29,191]. A genotype-first approach has been put forward as a superior classifier to define molecular phenotypes compared to nosological definitions with frequent overlap [192]. It may improve diagnoses of ASD subtypes and the development of therapies [192]. Although this objective still needs more development, shifting from consensus diagnosis boundaries to more biological and psychosocial defined disorders, will hopefully minimize the issue of heterogeneity and make a more personalized treatment based on etiology and pathophysiology. Before that day comes, large population-based studies are required to obtain a more thorough view of genotype-phenotype relationships using well-curated samples and longitudinal design. On the other hand, by using patient-specific cell lines or organoids, phenotypes can be characterized and compared among individuals with similar or different genetic profiles to design treatment for both personalized and larger case groups.

8.2 MULTI-OMICS APPLICATION FOR LARGE POPULATION AND INDIVIDUAL SAMPLES

Currently, studies using large-scale omics have been continuously unraveling the molecular basis of psychiatric disorders, and it becomes possible to integrate different omics data not only from individuals but also large population-based family cohorts. Combining longitudinal information and environmental exposures, collaborative data including genomics, transcriptomics, proteomics, and epigenomics can be collected to explore the developmental trajectories and interactions at variant, gene, and pathway levels. These efforts may help explain the missing heritability and improve our understanding of the genetic architecture of neurodevelopmental traits. Approaches from in vivo, in vitro, and in silico

environments can combine genetics and functional genomics from the same samples in tissues or cells for pathobiology and potential novel treatment targets.

8.3 COMBINATION OF GENETICS, IMAGING, BEHAVIORAL MEASURES FOR TREATMENT EVALUATION

In clinical trials, the efficacy for ASD treatment is mainly based on the clinically observed behavioral changes. However, the behaviorally-defined response is inadequate between different outcome measures and unpredictable for individuals [147]. Information of genomics, neuroimaging, and pathophysiology linked to associated genes are hopeful to be developed as potential biomarkers to help diagnose and predict the treatment response [193]. Integrating genetic testing and early symptom identification, including EEG, eye-tracking, and MRI, individuals could be possible receive a diagnosis at an earlier age [194]. In addition, circuit-based outcome measures may be more reliable for treatment efficacy compared to the global behavior functioning [153]. A combination of relevant biomarkers represents a unique profile for each patient, which may form an important part of the personalized trial design [153]. Furthermore, both behavioral intervention and pharmacological treatments can be combined in future trials to normalize neuronal and synaptic abnormalities as well as cognitive function. With a well-characterized group of subjects, a thorough understanding of disease mechanisms, and identified reliable biomarkers, clinical trials will have a higher possibility to be successful. All these outcomes will help to realize the notion of individualized intervention for ASD and other NDDs.

9 ACKNOWLEDGEMENTS

10 REFERENCES

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5), Fifth Edition*. (American Psychiatric Publishing, 2013).

2. Parenti, I., Rabaneda, L. G., Schoen, H. & Novarino, G. Neurodevelopmental Disorders: From Genetics to Functional Pathways. *Trends Neurosci.* **43**, 608–621 (2020).

3. Elsabbagh, M. *et al.* Global Prevalence of Autism and Other Pervasive Developmental Disorders. *Autism Res.* **5**, 160–179 (2012).

4. Simonoff, E. *et al.* Psychiatric disorders in children with autism spectrum disorders: Prevalence, comorbidity, and associated factors in a population-derived sample. *J. Am. Acad. Child Adolesc. Psychiatry* **47**, 921–929 (2008).

5. Sullivan, P. F. & Geschwind, D. H. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell* vol. 177 162–183 (2019).

6. McPheeters, M. L. *et al.* A Systematic Review of Medical Treatments for Children With Autism Spectrum Disorders. *Pediatrics* **127**, e1312–e1321 (2011).

7. Zheng, S., Kim, H., Salzman, E., Ankenman, K. & Bent, S. Improving Social Knowledge and Skills among Adolescents with Autism: Systematic Review and Meta-Analysis of UCLA PEERS® for Adolescents. *J. Autism Dev. Disord.* **51**, 4488–4503 (2021).

8. Choque Olsson, N. *et al.* Social Skills Training for Children and Adolescents With Autism Spectrum Disorder: A Randomized Controlled Trial. *J. Am. Acad. Child Adolesc. Psychiatry* **56**, 585–592 (2017).

9. Lord, C. *et al.* The Lancet Commission on the future of care and clinical research in autism. *Lancet* **6736**, (2021).

10. Brugha, T. S. *et al.* Epidemiology of autism spectrum disorders in adults in the community in England. *Arch. Gen. Psychiatry* **68**, 459–466 (2011).

11. Atladottir, H. O. *et al.* The increasing prevalence of reported diagnoses of childhood psychiatric disorders: a descriptive multinational comparison. *Eur. Child Adolesc. Psychiatry* **24**, 173–183 (2014).

12. Maenner, M. J. *et al.* Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2018. *MMWR Surveill. Summ.* **70**, 1–16 (2021).

13. Idring, S. *et al.* Changes in Prevalence of Autism Spectrum Disorders in 2001–2011: Findings from the Stockholm Youth Cohort. *J. Autism Dev. Disord.* **45**, 1766–1773 (2015).

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.* **57**, 585–595 (2016).

15. Taylor, M. J. *et al.* Etiology of Autism Spectrum Disorders and Autistic Traits Over Time. *JAMA Psychiatry* **77**, 936–943 (2020).

16. Sandin, S. *et al.* Autism risk associated with parental age and with increasing difference in age between the parents. *Mol. Psychiatry* **21**, 693–700 (2016).

17. Lyall, K., Ashwood, P., Van De Water, J. & Hertz-Picciotto, I. Maternal immune-mediated conditions, autism spectrum disorders, and developmental delay. *J. Autism Dev. Disord.* **44**, 1546–1555 (2014).

18. Christensen, J. *et al.* Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. *JAMA - J. Am. Med. Assoc.* **309**, 1696–1703 (2013).

19. Schmidt, R. J. *et al.* Maternal periconceptional folic acid intake and risk of autism spectrum disorders and developmental delay in the CHARGE (CHildhood Autism Risks from Genetics and Environment)

case-control study. *Am. J. Clin. Nutr.* **96**, 80–89 (2012).

20. Sztainberg, Y. & Zoghbi, H. Y. Lessons learned from studying syndromic autism spectrum disorders. *Nat. Neurosci.* **19**, 1408–1417 (2016).

21. Lord, C. *et al.* Autism from 2 to 9 years of age. *Arch. Gen. Psychiatry* **63**, 694–701 (2006).

22. Lord, C., Rutter, M., Dilavore, P. C. & Risi, S. *Autism Diagnostic Observation Schedule (ADOS)*. (Western Psychological Publishing, 1999).

23. Stone, W. L., Coonrod, E. E. & Ousley, O. Y. Brief Report: Screening Tool for Autism in Two-Year-Olds (STAT): Development and Preliminary Data. *J. Autism Dev. Disord.* **30**, 607–612 (2000).

24. Rutter, M., LeCouteur, A. & Lord, C. *Autism Diagnostic Interview-Revised (ADI-R)*. (Western Psychological Services, 2003).

25. Schopler, E., Reichler, R. J. & Renner, B. R. *The childhood autism rating scale (CARS)*. (Western Psychological Services, 1988).

26. Constantino, J. N. & Gruber, C. P. *Social Responsiveness Scale (SRS)*. (Western Psychological Services, 2005).

27. Rutter, M., Bailey, A. & Lord, C. *The social communication questionnaire: manual*. (Western Psychological Services, 2003).

28. Miller, J. S. *et al.* Autism Spectrum Disorder Reclassified: A Second Look at the 1980s Utah/UCLA Autism Epidemiologic Study. *J. Autism Dev. Disord.* **43**, 200–210 (2013).

29. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584.e23 (2020).

30. Lai, M. C. *et al.* Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. *The Lancet Psychiatry* **6**, 819–829 (2019).

31. Sayal, K., Prasad, V., Daley, D., Ford, T. & Coghill, D. ADHD in children and young people: prevalence, care pathways, and service provision. *The Lancet Psychiatry* **5**, 175–186 (2018).

32. Mattheisen, M. *et al.* Identification of shared and differentiating genetic risk for autism spectrum disorder , attention deficit hyperactivity disorder and case subgroups. *medRxiv* (2021) doi:https://doi.org/10.1101/2021.05.20.21257484.

33. Rylaarsdam, L. & Guemez-Gamboa, A. Genetic Causes and Modifiers of Autism Spectrum Disorder. *Front. Cell. Neurosci.* **13**, 385 (2019).

34. Hayashi, S. *et al.* Comprehensive investigation of CASK mutations and other genetic etiologies in 41 patients with intellectual disability and microcephaly with pontine and cerebellar hypoplasia (MICPCH). *PLoS One* **12**, e0181791 (2017).

35. Mcquade, R. D. & William, H. Aripiprazole in the Treatment of Irritability in Children and Adolescents With Autistic Disorder. *Pediatrics* **124**, 1533–1540 (2009).

36. Canitano, R. & Scandurra, V. Risperidone in the treatment of behavioral disorders associated with autism in children and adolescents. *Neuropsychiatric Disease and Treatment* vol. 4 723–730 (2008).

37. Sturman, N., Deckx, L. & van Driel, M. L. Methylphenidate for children and adolescents with autism spectrum disorder. *Cochrane Database Syst. Rev.* **2017**, CD011144 (2017).

38. Harfterkamp, M. *et al.* Atomoxetine in Autism Spectrum Disorder: No Effects on Social Functioning; Some Beneficial Effects on Stereotyped Behaviors, Inappropriate Speech, and Fear of Change. *J. Child Adolesc. Psychopharmacol.* **24**, 481–485 (2014).

39. Kreslins, A., Robertson, A. E. & Melville, C. The effectiveness of psychosocial interventions for anxiety in children and adolescents with autism spectrum disorder: a systematic review and meta-analysis. *Child Adolesc Psychiatry Ment Heal.* **9**, 22 (2015).

40. Reichow, B., Hume, K., Barton, E. E. & Boyd, B. A. Early intensive behavioral intervention (EIBI)

for young children with autism spectrum disorders (ASD). *Cochrane Database of Systematic Reviews* vol. 2018 CD009260 (2018).

41. Choque Olsson, N., Tammimies, K. & Bölte, S. Manualized social skills group training for children and adolescents with higher functioning autism spectrum disorder: protocol of a naturalistic multicenter, randomized controlled trial. *Transl. Dev. Psychiatry* **3**, 29825 (2015).

42. Vorstman, J. A. S. *et al.* Autism genetics: Opportunities and challenges for clinical translation. *Nat. Rev. Genet.* **18**, 362–376 (2017).

43. Gaugler, T. *et al.* Most genetic risk for autism resides with common variation. *Nat. Genet.* **46**, 881–885 (2014).

44. Geschwind, D. H. & Levitt, P. Autism spectrum disorders: developmental disconnection syndromes. *Curr. Opin. Neurobiol.* **17**, 103–111 (2007).

45. Singh, T., Neale, B. M. & Daly, M. J. Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia on behalf of the Schizophrenia Exome Meta-Analysis (SCHEMA) Consortium*. *medRxiv* 2020.09.18.20192815 (2020) doi:10.1101/2020.09.18.20192815.

46. Satterstrom, F. K. *et al.* Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. *Nat. Neurosci.* **22**, 1961–1965 (2019).

47. Kirov, G. *et al.* The Penetrance of Copy Number Variations for Schizophrenia and Developmental Delay. *Biol. Psychiatry* **75**, 378–385 (2014).

48. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

49. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482.e11 (2019).

50. Gustavsen, A. M. Longitudinal relationship between social skills and academic achievement in a gender perspective. *Cogent Educ.* **4**, 1411035 (2017).

51. Brainstorm Consortium, T. B. *et al.* Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757 (2018).

52. Fisher, R. A. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans. R. Soc. Edinburgh* **52**, 399–433 (1919).

53. Falconer, D. S. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann. Hum. Genet.* **29**, 51–76 (1965).

54. Ozaki, K. *et al.* Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32**, 650–654 (2002).

55. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).

56. Evans, D. M., Visscher, P. M. & Wray, N. R. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum. Mol. Genet.* **18**, 3525–3531 (2009).

57. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19**, 581–590 (2018).

58. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).

59. Bralten, J. *et al.* Autism spectrum disorders and autistic traits share genetics and biology. *Mol. Psychiatry* **23**, 1205–1212 (2018).

60. Hannon, E. *et al.* Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *Genome Med.* **10**, 1–13 (2018).

61. Weiner, D. J. *et al.* Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat. Genet.* **49**, 978–985 (2017).

62. Takahashi, N. *et al.* Association of Genetic Risks With Autism Spectrum Disorder and Early Neurodevelopmental Delays Among Children Without Intellectual Disability. *JAMA Netw. Open* **3**, e1921644 (2020).

63. Khundrakpam, B. *et al.* Neural correlates of polygenic risk score for autism spectrum disorders in general population. *Brain Commun.* **2**, fcaa092 (2020).

64. Qin, Y. *et al.* Polygenic risk for autism spectrum disorder affects left amygdala activity and negative emotion in schizophrenia. *Transl. Psychiatry 2020 101* **10**, 322 (2020).

65. Andersson, E. *et al.* Genetics of response to cognitive behavior therapy in adults with major depression: a preliminary report. *Mol. Psychiatry* **24**, 484–490 (2019).

66. Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **12**, 745–755 (2011).

67. Muller, H. J. Further studies on the nature and causes of gene mutations. in *Proceedings of the Sixth International Congress of Genetics* 213–255 (1932).

68. Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).

69. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science (80-.).* **305**, 525–528 (2004).

70. Brambati, C. *et al.* Droplet digital polymerase chain reaction for DNMT3A and IDH1/2 mutations to improve early detection of acute myeloid leukemia relapse after allogeneic hematopoietic stem cell transplantation. *Haematologica* vol. 101 e157–e161 (2016).

71. Adams, D. R. & Eng, C. M. Next-Generation Sequencing to Diagnose Suspected Genetic Disorders. *N. Engl. J. Med.* **379**, 1353–1362 (2018).

72. Drew, L. Pharmacogenetics: The right drug for you. *Nature* **537**, S60–S62 (2016).

73. Lester, K. J. & Eley, T. C. Therapygenetics: Using genetic markers to predict response to psychological treatment for mood and anxiety disorders. *Biol. Mood Anxiety Disord.* **3**, 4 (2013).

74. Pinto, D. *et al.* Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders. *Am. J. Hum. Genet.* **94**, 677–694 (2014).

75. Sanders, S. J. *et al.* Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. *Neuron* **70**, 863–885 (2011).

76. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).

77. Levy, D. *et al.* Rare De Novo and Transmitted Copy-Number Variation in Autistic Spectrum Disorders. *Neuron* **70**, 886–897 (2011).

78. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).

79. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).

80. Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215–1233 (2015).

81. Iakoucheva, L. M., Muotri, A. R. & Sebat, J. Getting to the Cores of Autism. *Cell* **178**, 1287–1298 (2019).

82. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23**, 185–188 (1999).

83. Brown, W. T. *et al.* Association of fragile X syndrome with autism. *Lancet* **319**, 100 (1982).

84. O'Roak, B. J. *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat. Genet.* **43**, 585–589 (2011).

85. Neale, B. M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).

86. O ' Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246–250 (2012).

87. Sanders, S. J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).

88. Werling, D. M. *et al.* An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. *Nat. Genet.* **50**, 727–736 (2018).

89. Ruzzo, E. K. *et al.* Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell* **178**, 850-866.e26 (2019).

90. An, J. Y. *et al.* Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science (80-.).* **362**, eaat6576 (2018).

91. Zhou, J. *et al.* Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. *Nat. Genet.* **51**, 973–980 (2019).

92. Wilfert, A. B. *et al.* Recent ultra-rare inherited variants implicate new autism candidate risk genes. *Nat. Genet.* **53**, 1125–1134 (2021).

93. Niemi, M. E. K. *et al.* Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature* **562**, 268–271 (2018).

94. Klei, L. *et al.* How rare and common risk variation jointly affect liability for autism spectrum disorder. *Mol. Autism* **12**, 66 (2021).

95. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **17**, (2021).

96. PsychEncode Consortium. Revealing the brain's molecular architecture. *Science (80-.).* **362**, 1262–1263 (2018).

97. Moog, U. *et al.* Phenotypic spectrum associated with CASK loss-of-function mutations. *J. Med. Genet.* **48**, 741–751 (2011).

98. Moog, U. *et al.* Phenotypic and molecular insights into CASK-related disorders in males. *Orphanet J. Rare Dis.* **10**, 44 (2015).

99. Lin, E. I., Jeyifous, O. & Green, W. N. CASK Regulates SAP97 Conformation and Its Interactions with AMPA and NMDA Receptors. *J. Neurosci.* **33**, 12067–12076 (2013).

100. LaConte, L. E. W. *et al.* CASK stabilizes neurexin and links it to liprin-α in a neuronal activity-dependent manner. *Cell. Mol. Life Sci.* **73**, 3599–3621 (2016).

101. Atasoy, D. *et al.* Deletion of CASK in mice is lethal and impairs synaptic function. *Proc. Natl. Acad. Sci.* **104**, 2525–2530 (2007).

102. Wagnon, J. L. *et al.* CELF4 Regulates Translation and Local Abundance of a Vast Set of mRNAs, Including Genes Associated with Regulation of Synaptic Function. *PLoS Genet.* **8**, e1003067 (2012).

103. Darnell, J. C. *et al.* FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism. *Cell* **146**, 247–261 (2011).

104. Jansen, A. *et al.* Gene-set analysis shows association between FMRP targets and autism spectrum disorder. *Eur. J. Hum. Genet.* **25**, 863–868 (2017).

105. Gandal, M. J. *et al.* Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science (80-.).* **359**, 693–697 (2018).

106. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).

107. Durak, O. *et al.* Chd8 mediates cortical neurogenesis via transcriptional regulation of cell cycle and Wnt signaling. *Nat. Neurosci.* **19**, 1477–1488 (2016).

108. Wang, H. & Doering, L. C. Reversing autism by targeting downstream mTOR signaling. *Front. Cell. Neurosci.* **7**, 28 (2013).

109. Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).

110. Lin, G. N. *et al.* Spatiotemporal 16p11.2 Protein Network Implicates Cortical Late Mid-Fetal Brain Development and KCTD13-Cul3-RhoA Pathway in Psychiatric Diseases. *Neuron* **85**, 742–754 (2015).

111. Velmeshev, D. *et al.* Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689 (2019).

112. Sequeiros, J. *et al.* The wide variation of definitions of genetic testing in international recommendations, guidelines and reports. *J. Community Genet.* **3**, 113–124 (2012).

113. Katsanis, S. H. & Katsanis, N. Molecular genetic testing and the future of clinical genomics. *Nat. Rev. Genet.* **14**, 415–426 (2013).

114. Finucane, B. M., Ledbetter, D. H. & Vorstman, J. A. Diagnostic genetic testing for neurodevelopmental psychiatric disorders: closing the gap between recommendation and clinical implementation. *Curr. Opin. Genet. Dev.* **68**, 1–8 (2021).

115. Havdahl, A. *et al.* Genetic contributions to autism spectrum disorder. *Psychol. Med.* 1–14 (2021) doi:10.1017/S0033291721000192.

116. Miller, D. T. *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.* **86**, 749–764 (2010).

117. Tammimies, K. *et al.* Molecular Diagnostic Yield of Chromosomal Microarray Analysis and Whole-Exome Sequencing in Children With Autism Spectrum Disorder. *JAMA* **314**, 895 (2015).

118. Srivastava, S. *et al.* Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet. Med.* **21**, 2413–2421 (2019).

119. Guo, H. *et al.* Genome sequencing identifies multiple deleterious variants in autism patients with more severe phenotypes. *Genet. Med.* **21**, 1611–1620 (2019).

120. Moreno-De-Luca, D. *et al.* Clinical genetic testing in autism spectrum disorder in a large community-based population sample. *JAMA Psychiatry* vol. 77 979–981 (2020).

121. Kinney, D. K., Barch, D. H., Chayka, B., Napoleon, S. & Munir, K. M. Environmental risk factors for autism: Do they help cause de novo genetic mutations that contribute to the disorder? *Med. Hypotheses* **74**, 102–106 (2010).

122. Volk, H. E. *et al.* Interaction of air pollution with the MET receptor tyrosine kinase gene. *Epidemiology* **25**, 44–47 (2014).

123. Schmidt, R. J. *et al.* Prenatal vitamins, one-carbon metabolism gene variants, and risk for Autism. *Epidemiology* **22**, 476–485 (2011).

124. Ponsonby, A. L. *et al.* Prenatal phthalate exposure, oxidative stress-related genetic vulnerability and early life neurodevelopment: A birth cohort study. *Neurotoxicology* **80**, 20–28 (2020).

125. Irwin, R. E. *et al.* The interplay between DNA methylation, folate and neurocognitive development. *Epigenomics* **8**, 863–879 (2016).

126. Phiel, C. J. *et al.* Histone Deacetylase Is a Direct Target of Valproic Acid, a Potent Anticonvulsant, Mood Stabilizer, and Teratogen*. *J. Biol. Chem.* **276**, 36734–36741 (2001).

127. Peyrot, W. J. *et al.* Does Childhood Trauma Moderate Polygenic Risk for Depression? A Meta-analysis of 5765 Subjects From the Psychiatric Genomics Consortium. *Biol. Psychiatry* **84**, 138–147 (2018).

128. Llewellyn, D. J. *et al.* Association of Lifestyle and Genetic Risk With Incidence of Dementia. *JAMA* **322**, 430–437 (2019).

129. Cucinotta, F. *et al.* FARP-1 deletion is associated with lack of response to autism treatment by early start denver model in a multiplex family. *Mol. Genet. Genomic Med.* **8**, e1373 (2020).

130. Serret, S. *et al.* Lithium as a rescue therapy for regression and catatonia features in two SHANK3 patients with autism spectrum disorder: Case reports. *BMC Psychiatry* **15**, 107 (2015).

131. Pini, G. *et al.* Repeated Insulin-Like Growth Factor 1 Treatment in a Patient with Rett Syndrome: A Single Case Study. *Front. Pediatr.* **2**, 52 (2014).

132. Amare, A. T. *et al.* Association of polygenic score for schizophrenia and HLA antigen and inflammation genes with response to lithium in bipolar affective disorder: A genome-wide association study. *JAMA Psychiatry* **75**, 65–74 (2018).

133. Zhang, J. P. *et al.* Schizophrenia polygenic risk score as a predictor of antipsychotic efficacy in first-episode psychosis. *Am. J. Psychiatry* **176**, 21–28 (2019).

134. Alemany-Navarro, M. *et al.* Do polygenic risk and stressful life events predict pharmacological treatment response in obsessive compulsive disorder? A gene–environment interaction approach. *Transl. Psychiatry* **9**, 70 (2019).

135. de Leeuw, C. A., Neale, B. M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nat. Rev. Genet.* **17**, 353–364 (2016).

136. Wigmore, E. M. *et al.* Genome-wide association study of antidepressant treatment resistance in a population-based cohort using health service prescription data and meta-analysis with GENDEP. *Pharmacogenomics J.* 1 (2019) doi:10.1038/s41397-019-0067-3.

137. Wang, Q. *et al.* Effect of Damaging Rare Mutations in Synapse-Related Gene Sets on Response to Short-term Antipsychotic Medication in Chinese Patients With Schizophrenia: A Randomized Clinical Trial. *JAMA Psychiatry* **75**, 1261–1269 (2018).

138. De Jong, S., Vidler, L. R., Mokrab, Y., Collier, D. A. & Breen, G. Gene-set analysis based on the pharmacological profiles of drugs to identify repurposing opportunities in schizophrenia. *J. Psychopharmacol.* **30**, 826–830 (2016).

139. Umehara, H. *et al.* Calcium Signaling Pathway Is Associated with the Long-Term Clinical Response to Selective Serotonin Reuptake Inhibitors (SSRI) and SSRI with Antipsychotics in Patients with Obsessive-Compulsive Disorder. *PLoS One* **11**, e0157232 (2016).

140. Krishnan, A. *et al.* Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat. Neurosci.* **19**, 1454–1462 (2016).

141. Asif, M. *et al.* Identification of biological mechanisms underlying a multidimensional ASD phenotype using machine learning. *Transl. Psychiatry* **10**, 43 (2020).

142. Chen, S. *et al.* An interactome perturbation framework prioritizes damaging missense mutations for developmental disorders. *Nat. Genet.* **50**, 1032–1040 (2018).

143. Lenhard, F. *et al.* Prediction of outcome in internet-delivered cognitive behaviour therapy for paediatric obsessive-compulsive disorder: A machine learning approach. *Int. J. Methods Psychiatr. Res.* **27**, e1576 (2018).

144. Kautzky, A. *et al.* The combined effect of genetic polymorphisms and clinical parameters on treatment outcome in treatment-resistant depression. *Eur. Neuropsychopharmacol.* **25**, 441–453 (2015).

145. Kosmicki, J. A. *et al.* Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat. Publ. Gr.* **49**, (2017).

146. Chaste, P. *et al.* A genome-wide association study of autism using the Simons simplex collection: Does reducing phenotypic heterogeneity in autism increase genetic homogeneity? *Biol. Psychiatry* **77**, 775–784 (2015).

147. Searles Quick, V. B., Wang, B. & State, M. W. Leveraging large genomic datasets to illuminate the pathobiology of autism spectrum disorders. *Neuropsychopharmacology* **46**, 55–69 (2020).

148. Pizzo, L. *et al.* Rare variants in the genetic background modulate cognitive and developmental phenotypes in individuals carrying disease-associated variants. *Genet. Med.* **21**, 816–825 (2019).

149. Nehme, R. & Barrett, L. E. Using human pluripotent stem cell models to study autism in the era of big data. *Mol. Autism* **11**, 21 (2020).

150. Kasari, C., Shire, S., Factor, R. & McCracken, C. Psychosocial Treatments for Individuals with Autism Spectrum Disorder Across the Lifespan: New Developments and Underlying Mechanisms. *Curr. Psychiatry Rep.* **16**, 512 (2014).

151. Siegel, M. & Beaulieu, A. A. Psychotropic medications in children with autism spectrum disorders: A systematic review and synthesis for evidence-based practice. *J. Autism Dev. Disord.* **42**, 1592–1605 (2012).

152. Mellios, N. *et al.* β2-Adrenergic receptor agonist ameliorates phenotypes and corrects microRNA-mediated IGF1 deficits in a mouse model of Rett syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9947–9952 (2014).

153. Sahin, M. & Sur, M. Genes, circuits, and precision therapies for autism and related neurodevelopmental disorders. *Science (80-.).* **350**, aab3897 (2015).

154. Sestan, N. & State, M. W. Lost in Translation: Traversing the Complex Path from Genomics to Therapeutics in Autism Spectrum Disorder. *Neuron* **100**, 406–423 (2018).

155. World Health Organization. *International Classification of Diseases and Related Health Problems 10th Revision*. (World Health Organization, 1992).

156. Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nat. Protoc.* **5**, 1564–1573 (2010).

157. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

158. Belmont, J. W. *et al.* The international HapMap project. *Nature* **426**, 789–796 (2003).

159. Uddin, M. *et al.* A high-resolution copy-number variation resource for clinical and population genetics. *Genet. Med.* **17**, 747–752 (2015).

160. MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L. & Scherer, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* **42**, D986–D992 (2014).

161. Kearney, H. M., Thorland, E. C., Brown, K. K., Quintero-Rivera, F. & South, S. T. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet. Med.* **13**, 680–685 (2011).

162. Martin, J. *et al.* Copy number variation and neuropsychiatric problems in females and males in the general population. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **180**, 341–350 (2019).

163. Stranneheim, H. *et al.* Rapid pulsed whole genome sequencing for comprehensive acute diagnostics of inborn errors of metabolism. *BMC Genomics* **15**, 1090 (2014).

164. Richards, S. *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.* **17**, 405–423 (2015).

165.   Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).

166.   Howie, B. N., Donnelly, P. & Marchini, J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet.* **5**, e1000529 (2009).

167.   Van Leeuwen, E. M. *et al.* Population-specific genotype imputations using minimac or IMPUTE2. *Nat. Protoc.* **10**, 1285–1296 (2015).

168.   Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).

169.   Martin, J. *et al.* A Genetic Investigation of Sex Bias in the Prevalence of Attention-Deficit/Hyperactivity Disorder. *Biol. Psychiatry* **83**, 1044–1053 (2018).

170.   Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).

171.   Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466–1468 (2015).

172.   de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11**, e1004219 (2015).

173.   Bölte, S. *et al.* The Roots of Autism and ADHD Twin Study in Sweden (RATSS). *Twin Res. Hum. Genet.* **17**, 164–176 (2014).

174.   Myers, L. *et al.* Minor physical anomalies in neurodevelopmental disorders: A twin study. *Child Adolesc. Psychiatry Ment. Health* **11**, 57 (2017).

175.   Wincent, J. *et al.* Copy number variations in children with brain malformations and refractory epilepsy. *Am. J. Med. Genet. Part A* **167**, 512–523 (2015).

176.   Stamouli, S. *et al.* Copy Number Variation Analysis of 100 Twin Pairs Enriched for Neurodevelopmental Disorders. *Twin Res. Hum. Genet.* **21**, 1–11 (2018).

177.   Uhlin, E. *et al.* Derivation of human iPS cell lines from monozygotic twins in defined and xeno free conditions. *Stem Cell Res.* **18**, 22–25 (2017).

178.   Falk, A. *et al.* Capture of Neuroepithelial-Like Stem Cells from Pluripotent Stem Cells Provides a Versatile System for In Vitro Production of Human Neurons. *PLoS One* **7**, e29597 (2012).

179.   Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27**, 275–280 (2009).

180.   Li, W. V. & Li, J. J. An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nat. Commun.* **9**, 997 (2018).

181.   Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol. 2018 365* **36**, 411–420 (2018).

182.   Reimand, J. *et al.* Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* **14**, 482–517 (2019).

183.   Tammimies, K. *et al.* Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder. *JAMA* **314**, 895–903 (2015).

184.   Dong, Z. *et al.* CUL3 Deficiency Causes Social Deficits and Anxiety-like Behaviors by Impairing Excitation-Inhibition Balance through the Promotion of Cap-Dependent Translation. *Neuron* **105**, 475-490.e6 (2020).

185.   Sprengers, J. J. *et al.* Bumetanide for Core Symptoms of Autism Spectrum Disorder (BAMBI): A Single Center, Double-Blinded, Participant-Randomized, Placebo-Controlled, Phase-2 Superiority Trial. *J. Am. Acad. Child Adolesc. Psychiatry* **0**, (2020).

186. Patel, P. A. *et al.* Haploinsufficiency of X-linked intellectual disability gene CASK induces post-transcriptional changes in synaptic and cellular metabolic pathways. *Exp. Neurol.* **329**, 113319 (2020).

187. McRae, J. F. *et al.* Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433–438 (2017).

188. Mori, T. *et al.* Deficiency of calcium/calmodulin-dependent serine protein kinase disrupts the excitatory-inhibitory balance of synapses by down-regulating GluN2B. *Mol. Psychiatry* **24**, 1079–1092 (2019).

189. Freitag, C. M. *et al.* Group-based cognitive behavioural psychotherapy for children and adolescents with ASD: the randomized, multicentre, controlled SOSTA - net trial. *J. Child Psychol. Psychiatry* **57**, 596–605 (2016).

190. Gates, J. A., Kang, E. & Lerner, M. D. Efficacy of group social skills interventions for youth with autism spectrum disorder: A systematic review and meta-analysis. *Clin. Psychol. Rev.* **52**, 164–181 (2017).

191. Wilfert, A. B., Sulovari, A., Turner, T. N., Coe, B. P. & Eichler, E. E. Recurrent de novo mutations in neurodevelopmental disorders: properties and clinical implications. *Genome Med.* **9**, 101 (2017).

192. Stessman, H. A., Bernier, R. & Eichler, E. E. A Genotype-First Approach to Defining the Subtypes of a Complex Disease. *Cell* **156**, 872–877 (2014).

193. Masi, A., DeMayo, M. M., Glozier, N. & Guastella, A. J. An Overview of Autism Spectrum Disorder, Heterogeneity and Treatment Options. *Neurosci. Bull.* **33**, 183–193 (2017).

194. Webb, S. J. *et al.* Biomarker Acquisition and Quality Control for Multi-Site Studies: The Autism Biomarkers Consortium for Clinical Trials. *Front. Integr. Neurosci.* **13**, 71 (2020).