Common genetic overlap in childhood psychiatric disorders

Jansen, Adriana Gerarda

2022

document version
Publisher's PDF, also known as Version of record

Common genetic overlap in childhood psychiatric disorders

Arija Jansen

Common genetic overlap in childhood psychiatric disorders

Arija Jansen

Paranimfen

Marco Benjamin Lutters, CEO
Rogier Visser, MSc

Design and lay out: Arija Jansen
Cover: ism Proefschrift maken
Artwork, including cover art: Mia Adriana Lutters
Printing: Proefschrift maken (www.proefschriftmaken.nl)
ISBN: 978-94-6423-677-4

VRIJE UNIVERSITEIT

COMMON GENETIC OVERLAP IN CHILDHOOD PSYCHIATRIC DISORDERS

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Bètawetenschappen
op woensdag 20 april 2022 om 9.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

Adriana Gerarda Jansen

geboren te Hendrik-Ido-Ambacht

promotoren:	prof.dr. D. Posthuma
	prof.dr. F.C. Verhulst
copromotoren:	dr. J.C. Polderman
	dr. G.C. Dieleman
promotiecommissie:	prof.dr. A. Ronald
	prof.dr. N.N.J. Lambregts-Rommelse
	prof.dr. C.J.H. Verweij
	dr. B. Dierckx
	prof.dr. M.P. van den Heuvel

#OHANA

Voor haar,

met de wilde krullen

en zonlicht,

wiens lach de hemelen verraadt

Table of contents

1

Introduction

Introduction

1.1 General introduction

Psychiatric disorders are debilitating disorders which cause a tremendous amount of suffering for patients and their families (Costello et al., 2005). Many of these psychiatric disorders arise during childhood as 50% of the symptoms of psychiatric disorders emerge by the age of 14 and 75% by the age of 24 (Kessler et al., 2007). As such, these kinds of disorders are rightfully high on the research agenda which has resulted in a better understanding of the origins of these disorders (Visscher et al., 2017). Psychiatric disorders and psychological traits such as scores on neuroticism scales or risk-taking behavior are influenced by genetic and environmental factors, better known as respectively nature and nurture, and a complex interplay between the two. Family, twin and adoption studies have shown that to various degrees all psychiatric disorders and psychological traits are heritable (Polderman et al., 2015) with a complex genetic and environmental signature, classifying them as complex traits (Plomin et al., 2016). The heritabilities of the psychiatric disorders and psychological traits relevant for this thesis are provided in Chapter 2. In this thesis, when I use the terms psychiatric disorders, I refer to traits that have been measured categorically, a psychiatric disorder is either present or absent as determined by clinical standards, for example autism spectrum disorder or schizophrenia. It is proposed that these disorders should not (only) be classified as dichotomous traits but would benefit from a continuous measurement which would do justice to the heterogeneity of the disorders. When I use the term psychological trait, I refer to a personal trait which is present in the whole population and which is observed as a score on a standardized assessment scale for example neuroticism (high to low continuum) or educational attainment (high to low continuum). However, a psychological trait can also be categorical, as can be seen in the traits smoking initiation (yes/no), risk taking behavior (are you a person who takes many risks yes/no).

In this thesis I focus on the genetic aspects of psychiatric disorders. Where family, twin and adoption studies have shown us to what extent certain psychiatric disorders and psychological traits are heritable, genome wide association studies (GWAS) have given us the first insights into which parts of the genome are associated with these psychiatric disorders and psychological traits. GWAS are often large case-control studies attempting to show which single nucleotide polymorphisms (SNPs)/ specific locations in the genome are associated with a disorder. The genetic architecture of psychiatric disorders and psychological traits turns out to be polygenic, meaning that multiple SNPs that each make a small contribution to a psychiatric disorder or psychological trait liability are involved. This was for instance shown in a hallmark GWAS on schizophrenia (SCZ)(schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) in which they identified 128 independent genetic associations spanning 108 loci. Soon after, GWAS of other psychiatric disorders were published with similar results (Demontis et al., 2019; Wray et al., 2018). Now, capitalizing on the successes

of GWAS in identifying associated SNPs, the next step is interpreting these findings. A big challenge in interpreting GWAS results is the correct annotation of SNPs to genes as most found SNPs are located intergenic and hence can be in close vicinity of multiple genes or they can be related to the function of a gene located further up or downstream in the genome. A second important issue is the fact that genes do not function by themselves but are embedded in functional pathways and a gene can be part of multiple pathways (Schadt, 2009). These challenges have given rise to tools as gene-set analysis (de Leeuw et al., 2016), functional annotation tools (Watanabe et al., 2017) and polygenic risk scores (PRS) (Euesden et al., 2015; Wray et al., 2020) designed to identify candidate genes and pathways associated with the disorders and traits. These methods are often utilized as follow up tools in large GWAS (P. R. Jansen et al., 2019; Nagel et al., 2018; Savage et al., 2018). GWAS, gene-set analysis and PRS are the main tools I have used in this thesis and are explained in paragraph 1.2 of this Chapter.

In addition to the difficulty of interpreting the associated SNPs as identified in GWAS there are the issues of genetic heterogeneity (McClellan & King, 2010) (one disorder but multiple combinations of risk increasing SNPs are possible), clinical heterogeneity (Hodgson et al., 2017) (one disorder but patients can show a different subset of symptoms), comorbidity (van Oudheusden et al., 2015) (a patient has multiple psychiatric diagnoses) and symptom overlap between psychiatric disorders (several disorders share the same symptoms) (*DSM 5*, 2013, p. 5). These issues result in the conclusion that there is not one explanation for the manifestation of a psychiatric disorder or psychological trait in all patients as all patients have a unique genetic signature (and environmental context), a unique symptom signature and unique comorbidities.

Genetic and clinical heterogeneity, symptom overlap, comorbidity and the fact that mental disorders and psychological traits are not only present or absent but can exist on a continuum have resulted in approaches aimed at understanding and mapping the dimensions and interrelatedness of psychiatric disorders resulting in the concept of a *p*-factor (*p*) (Caspi et al., 2014; Caspi & Moffitt, 2018; Selzam et al., 2018; Smith et al., 2020). *The p*-factor is based on a familiar concept in psychology, *g.* Where *g* conceptualizes general intelligence on a low to high scale, *p* conceptualizes psychopathology severity in a low to high scale (Caspi & Moffitt, 2018), more or less life impairment, worse developmental history (Caspi et al., 2014) and general liability to develop a mental disorder (Selzam et al., 2018). This *p*-factor has been observed at several levels, internalizing (Caspi et al., 2014; Neumann et al., 2016), externalizing (Caspi et al., 2014; Neumann et al., 2016), thought disorder (Caspi et al., 2014) and a general psychopathology level (Caspi et al., 2014; Neumann et al., 2016; Selzam et al., 2018). In addition, the *p*-factor has been reported to have an estimated SNP heritability of 38% (Neumann et al., 2016) to 43% (Smith et al., 2020). In depth information on the *p*-factor is provided in paragraph 1.2.4.

Although many advances have been made in the field of psychiatric genetics, they have not yet resulted in new cures or treatment methods for psychiatric disorders (Gandal et

al., 2016; Sullivan & Geschwind, 2019) and many research questions remain to be investigated.

In this thesis I focus on the role of shared genetic factors for different psychiatric disorders, a new research area at the beginning of this project in 2008. At the time, it was acknowledged that these types of disorders often co-occur but a scientific explanation as to why was not available. This results in the main aim of the research presented in this thesis, which is: investigating the shared genetic factors between different psychiatric disorders in children and adolescents. To answer this question, in Chapter 3, I have focused on the broad phenotype "diagnostic status" captured in a case-control study comparing undiagnosed controls to children diagnosed with one or more psychiatric disorders. The common genetic variation that is tested in this chapter was captured in a PRS for each individual and per individual several PRS for different disorders were computed. In chapter 4 I zoom in on neurodevelopmental disorders (i.e. a group comprising ASD and ADHD). This subset of disorders is present since birth, has a high comorbidity, and is highly heritable (A. G. Jansen et al., 2019). This is a large group in our sample (see Chapter 2 for an extensive sample description) which provided a unique opportunity to investigate the genetic overlap between ASD and ADHD. Finally, in Chapter 5, I focus on ASD by running gene-set analysis on publicly available data to investigate the contribution of common genetic variation to biological pathways functionally involved in ASD.

1.2 Theoretical background

In this thesis several well-established statistical genetics methods and concepts regarding general psychopathology have been applied. These are described below.

1.2.1 Genome-wide Association Studies (GWAS)

In the last 10 years GWAS have been the method of choice for many geneticists to find genetic variants associated with complex traits under study (Visscher et al., 2012, 2017). Complex traits are often polygenic with possibly hundreds of genetic variants involved (Watanabe et al., 2019), some already implicated by other studies, others still unknown. The hypothesis free statistical method of GWAS enables the researcher to scan the genome on associated genetic variants without any prior knowledge. For this, a large group of individuals comprising cases and controls is genotyped on a selection of single nucleotide polymorphisms (SNPs). SNPs are genetic differences that may contribute to behavioral differences between people. It is then tested if certain genetic variants are more prevalent in the cases versus the controls. The genotyped SNPs are intended to be representative for the whole genome due to naturally occurring correlation patterns called linkage disequilibrium (Pritchard & Przeworski, 2001) (LD) between groups of SNPs in the genome. The ungenotyped SNPs can be added due to this LD by the process of imputation. Ancestry must be taken into account to avoid issues with

population stratification. This is the difference in allele frequency between different ethnical groups which can lead to measuring systematic differences in ancestry rather than measuring genes possibly associated with diseases (Freedman et al., 2004). There are SNPs common in the population (these are present in >1% of the population) and rare ones (present in <0.1% of the population). GWAS focusses on the common SNPs which are not under genetic pressure, hence can stay in the population at a moderate to high frequency due to a small effect on the individual (Risch & Merikangas, 1996). As a GWAS tests all SNPs together, around 6,000,000 tests are executed resulting in the need for a stringent multiple-testing correction to reduce false positives. A SNP is considered significantly associated if it reaches a P-value < 5×10^{-8}. Due to this stringent significance threshold large sample sizes are needed to gain enough statistical power. Sample sizes as large as 40,675 cases and 64,643 controls (Pardiñas et al., 2018) are common in GWAS exploring psychiatric disorders and sample sizes of 1,232,091 (Liu et al., 2019) are not exceptional in GWAS that examine continuous trait measures. Even with these already substantial sample sizes, there is still a large missing heritability which is the difference between heritability as estimated from the significantly associated SNPs found in GWAS and the heritability estimated by twin and family studies (Zuk et al., 2012). To close this gap even larger samples must be acquired or different genotyping methods like whole genome sequencing must be used to create even better effect size estimates and more accurate P-values.
In addition, it is important to focus on reliable interpretation of the GWAS hits to uncover biological mechanisms underlying complex traits such as psychiatric disorders. Several methods have been developed for this purpose including polygenic risk scoring (Wray et al., 2014) (see below), gene-set analysis (de Leeuw et al., 2015, 2016) (see below) and online functional mapping and genetic annotation software have been created (Watanabe et al., 2017).

1.2.2 Gene-set analysis

To enhance the interpretation of GWAS hits, methods have been developed that test the GWAS findings simultaneously in a biological plausible context. The goal is to facilitate the generation of hypotheses concerning causal biological mechanisms (Schadt, 2009). Gene-set analysis (Mooney & Wilmot, 2015) is an often used tool for analyzing GWAS results and does justice to the fact that SNPs do not act in isolation, and the fact that even genes do not act by themselves but work together in pathways of functionally related genes that regulate biological mechanisms (Zhu et al., 2008). An underlying assumption is the fact that significantly associated SNPs accumulate in genes with similar cellular functions, or similar expression pattern across tissues or cell types. SNPs are assigned to genes and genes are assigned to gene-sets to assess the joint effect. This reduces the number of tests executed and multiple-testing correction is as such reduced, compared to GWAS (de Leeuw et al., 2015, 2016).

1.2.3 Polygenic Risk Score (PRS) analysis

A way to utilize GWAS findings is by means of PRS analysis (Wray et al., 2014). With this method, for each individual the number of risk alleles associated with a given trait per SNP is multiplied by their effect size as estimated by the GWAS. The sum of these weighted risks across all SNPs results in a single measure approximating an individual's genetic predisposition for a complex trait.

PRS is useful in research at group level into disorder etiology and genetic overlap but do not have enough finesse to be useful by themselves in a clinical setting for prediction for individual patients. The explained variance of the PRS is relatively low and limited by the SNP heritability of the disorder, leaving a lot of room for other risk factors to contribute to the development of a trait or disorder (Wray et al., 2020). This means PRS are not suitable to use as a single diagnostic tool but can be used 1) in an existing diagnostic test battery as a guide into more extensive and optimized disease screening and interpretation of disease screens, as for example, is currently beneficial to diabetes 1 patients of whom 15% are wrongly diagnosed as type 2 diabetes patients as not al biomarkers are exclusive for type 1 or type 2 diabetes and age lines between type 1 and type 2 diabetes are becoming increasingly more blurred (Oram et al., 2016; Wray et al., 2020). The use of a type 1 diabetes PRS as an additional tool in the diagnostic process can result in more frequent monitoring of insulin levels in individuals with a high diabetes type 1 PRS. This can avoid unseen escalation to the critical to the critical insulin deficiency type (Wray et al., 2020). 2) As an informant into therapeutic interventions to prevent or treat disease, hence as a guide into treatment respondent subtypes, an area of research in which the inflammatory bowel disease community of research is a frontrunner (Graham & Xavier, 2020; Wray et al., 2020). 3) For informative life planning and preventative health management strategies, for example, individuals in the highest PRS quartile of coronary artery disease will benefit from optimal lifestyle habits to reduce the overall disease risk by nearly half (Torkamani et al., 2018). However, in mental health care the clinical use of PRS has not yet taken flight. A lack of population wide screening programs concerning psychiatric disorders makes it currently unfit for screening in the healthy population as it cannot latch on to existing structures (Murray et al., 2021). When the lifetime risk of a disorder is 1%, as is the case for schizophrenia, even when the PRS reaches is maximum potential (where it currently is not due to limited GWAS power), and we only look at the top 1% of schizophrenia PRS, we see that the lifetime risk goes up 13-fold. This is relatively a large increase but in absolute measures it still means that most individuals in that category will not suffer schizophrenia (Murray et al., 2021). For disorders with a 15% lifetime risk, such as depression, the liability goes up 3-fold, rendering the same conclusion. Most individuals in the top 1% PRS will not suffer from depression (Murray et al., 2021). This is another reason PRS are at least currently unfit for general population screening. PRS can become a valid tool for the help-seeking population in which a PRS can tilt the scale, just like a family history of a disorder can, in the decision making process (Murray et al., 2021).

In addition to clinical use, PRS are useful in a research setting trying to uncover biological underpinnings of disorders (Sullivan & Geschwind, 2019) and can be used in case-control studies for evaluation and prediction purposes of the trait or disorder (Wray et al., 2020).

1.2.4 Structure of psychopathology

The current diagnostic process regarding psychiatric disorders is partly based on the DSM 5 (*DSM 5*, 2013,) which classifies behavior based on criteria applicable to psychiatric disorders. When an individual meets x amount of criteria and the psychiatric exam generates concurring results, a diagnosis can be given. Individuals can be given multiple diagnoses as they meet criteria for several due to overlap.

The DSM however is not based on the empirical insights in the dimensions, nature and interrelatedness of symptoms or problem behaviors and does not do justice to the continuous nature of these disorders (Kotov et al., 2017). In addition, there is a large clinical heterogeneity blurring the meaning to be given to the received diagnosis and psychiatric codiagnoses are more norm than exception (Smith et al., 2020) which increases the notion that psychiatric disorders and their etiologies might not be as specific as assumed. These issues can be addressed by looking at the dimensional structure of psychiatric disorders which finds an origin in the "internalizing" and "externalizing" factors as for instance already described by Achenbach in 1966 (Smith et al., 2020). This work is based on the covarying symptoms within the internalizing (internalizing problems are generally considered to belong to the subgroup of psychopathology that involves disturbances in emotion or mood such as sadness, guilt or worry) and externalizing (externalizing problems tend to refer to dysregulations in behavior) categories, which highly correlate in children ($r = 0.66$), adolescents ($r = 0.72$) and adults ($r = 0.51$) (Smith et al., 2020).

This paves the way for more in-depth research on a broader overarching general psychopathology factor (*p*-factor). Research has shown the *p*-factor to be stable over time in children, adolescents and adults (Liu et al., 2019) hinting it to be a heritable trait. The heritability has indeed been shown and has been estimated to be 38% (Neumann et al., 2016) to 43%(Smith et al., 2020). The interpretations of the *p*-factor are manifold and include dispositional negative emotionality, impulsive responsivity to emotion, low cognitive functioning, thought dysfunction and impairment (Smith et al., 2020). Of all traits, impairment is the most plausible, as the other four interpretations suffer in their capacity to explain all or at least several impairments as they each focus on a specific component of impairment. Each of the variables loads onto the *p*-factor, but it is unclear how each specific interpretation explains the variance of all the variables loading on the *p*-factor (Smith et al., 2020). The impairment hypothesis does not have this problem as it is clearer to see how impairment can explain all variables loading onto the *p*-factor (Smith et al., 2020).

This leads to the question whether the *p*-factor has clinical utility. Although with caution, a first advantage, when the p-factor is refined in the future, is that it could turn into a clear measuring tool, making it less subject to clinician biases. A second advantage could be

that it creates a frame work to shape patient treatment (Smith et al., 2020). For example, a patient presenting with only bulimia nervosa symptoms who is functioning well at school and does not have a history of mental health and a patient with the same symptoms but in addition does not function well in school, has depressive episodes and a history of mental health issues will benefit from different treatments (Smith et al., 2020). A reliable measure of the *p*-factor will provide additional insights into which treatment is best suited, the duration and hence costs of this treatment and the best possible outcome for each individual patient, being complete remission of symptoms or an improved quality of life and better handle on the daily struggles without full remission of symptoms (Smith et al., 2020).

References

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clinical Psychological Science: A Journal of the Association for Psychological Science*, *2*(2), 119–137. https://doi.org/10.1177/2167702613497473

Caspi, A., & Moffitt, T. E. (2018). All for One and One for All: Mental Disorders in One Dimension. *The American Journal of Psychiatry*, *175*(9), 831–844. https://doi.org/10.1176/appi.ajp.2018.17121383

Costello, E. J., Egger, H., & Angold, A. (2005). 10-Year Research Update Review: The Epidemiology of Child and Adolescent Psychiatric Disorders: I. Methods and Public Health Burden. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(10), 972–986. https://doi.org/10.1097/01.chi.0000172552.41596.6f

de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Computational Biology*, *11*(4), e1004219. https://doi.org/10.1371/journal.pcbi.1004219

de Leeuw, C. A., Neale, B. M., Heskes, T., & Posthuma, D. (2016). The statistical properties of gene-set analysis. *Nature Reviews. Genetics*, *17*(6), 353–364. https://doi.org/10.1038/nrg.2016.29

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J. I., Grasby, K. L., Grove, J., … Neale, B. M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics*, *51*(1), 63–75. https://doi.org/10.1038/s41588-018-0269-7

Diagnostic and Statistical Manual of mental disorders (Vol. 5). (2013). American Psychiatric Association.

Euesden, J., Lewis, C. M., & O'Reilly, P. F. (2015). PRSice: Polygenic Risk Score software. *Bioinformatics (Oxford, England)*, *31*(9), 1466–1468. https://doi.org/10.1093/bioinformatics/btu848

Freedman, M. L., Reich, D., Penney, K. L., McDonald, G. J., Mignault, A. A., Patterson, N., Gabriel, S. B., Topol, E. J., Smoller, J. W., Pato, C. N., Pato, M. T., Petryshen, T. L., Kolonel, L. N., Lander, E. S., Sklar, P., Henderson, B., Hirschhorn, J. N., & Altshuler, D. (2004). Assessing the impact of population stratification on genetic association studies. *Nature Genetics*, *36*(4), 388–393. https://doi.org/10.1038/ng1333

Gandal, M. J., Leppa, V., Won, H., Parikshak, N. N., & Geschwind, D. H. (2016). The road to precision psychiatry: Translating genetics into disease mechanisms. *Nature Neuroscience*, *19*(11), 1397–1407. https://doi.org/10.1038/nn.4409

Graham, D. B., & Xavier, R. J. (2020). Pathway paradigms revealed from the genetics of inflammatory bowel disease. *Nature*, *578*(7796), 527–539. https://doi.org/10.1038/s41586-020-2025-2

Hodgson, K., McGuffin, P., & Lewis, C. M. (2017). Advancing psychiatric genetics through dissecting heterogeneity. *Human Molecular Genetics*, *26*(R2), R160–R165. https://doi.org/10.1093/hmg/ddx241

Jansen, A. G., Dieleman, G. C., Jansen, P. R., Verhulst, F. C., Posthuma, D., & Polderman, T. J. C. (2019). Psychiatric Polygenic Risk Scores as Predictor for Attention Deficit/Hyperactivity Disorder and Autism Spectrum Disorder in a Clinical Child and Adolescent Sample. *Behavior Genetics*. https://doi.org/10.1007/s10519-019-09965-8

Jansen, P. R., Watanabe, K., Stringer, S., Skene, N., Bryois, J., Hammerschlag, A. R., Leeuw, C. A. de, Benjamins, J. S., Muñoz-Manchado, A. B., Nagel, M., Savage, J. E., Tiemeier, H., White, T., Tung, J. Y., Hinds, D. A., Vacic, V., Wang, X., Sullivan, P. F., Sluis, S. van der, … Posthuma, D. (2019). Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nature Genetics*, *51*(3), 394. https://doi.org/10.1038/s41588-018-0333-3

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Ustün, T. B. (2007). Age of onset of mental disorders: A review of recent literature. *Current Opinion in Psychiatry*, *20*(4), 359–364. https://doi.org/10.1097/YCO.0b013e32816ebc8c

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., Zhan, X., Choquet, H., Docherty, A. R., Faul, J. D., Foerster, J. R.,

Fritsche, L. G., Gabrielsen, M. E., Gordon, S. D., Haessler, J., … Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, *51*(2), 237. https://doi.org/10.1038/s41588-018-0307-5

McClellan, J., & King, M.-C. (2010). Genetic heterogeneity in human disease. *Cell*, *141*(2), 210–217. https://doi.org/10.1016/j.cell.2010.03.032

Mooney, M. A., & Wilmot, B. (2015). Gene set analysis: A step-by-step guide. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, *168*(7), 517–527. https://doi.org/10.1002/ajmg.b.32328

Murray, G. K., Lin, T., Austin, J., McGrath, J. J., Hickie, I. B., & Wray, N. R. (2021). Could Polygenic Risk Scores Be Useful in Psychiatry?: A Review. *JAMA Psychiatry*, *78*(2), 210–219. https://doi.org/10.1001/jamapsychiatry.2020.3042

Nagel, M., Jansen, P. R., Stringer, S., Watanabe, K., de Leeuw, C. A., Bryois, J., Savage, J. E., Hammerschlag, A. R., Skene, N. G., Muñoz-Manchado, A. B., 23andMe Research Team, White, T., Tiemeier, H., Linnarsson, S., Hjerling-Leffler, J., Polderman, T. J. C., Sullivan, P. F., van der Sluis, S., & Posthuma, D. (2018). Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genetics*, *50*(7), 920–927. https://doi.org/10.1038/s41588-018-0151-7

Neumann, A., Pappa, I., Lahey, B. B., Verhulst, F. C., Medina-Gomez, C., Jaddoe, V. W., Bakermans-Kranenburg, M. J., Moffitt, T. E., van IJzendoorn, M. H., & Tiemeier, H. (2016). Single Nucleotide Polymorphism Heritability of a General Psychopathology Factor in Children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *55*(12), 1038-1045.e4. https://doi.org/10.1016/j.jaac.2016.09.498

Oram, R. A., Patel, K., Hill, A., Shields, B., McDonald, T. J., Jones, A., Hattersley, A. T., & Weedon, M. N. (2016). A Type 1 Diabetes Genetic Risk Score Can Aid Discrimination Between Type 1 and Type 2 Diabetes in Young Adults. *Diabetes Care*, *39*(3), 337–344. https://doi.org/10.2337/dc15-1111

Pardiñas, A. F., Holmans, P., Pocklington, A. J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S. E., Bishop, S., Cameron, D., Hamshere, M. L., Han, J., Hubbard, L., Lynham, A., Mantripragada, K., Rees, E., MacCabe, J. H., McCarroll, S. A., Baune, B. T., Breen, G., … CRESTAR Consortium. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature Genetics*, *50*(3), 381–389. https://doi.org/10.1038/s41588-018-0059-2

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 Replicated Findings from Behavioral Genetics. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *11*(1), 3–23. https://doi.org/10.1177/1745691615617439

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), 702–709. https://doi.org/10.1038/ng.3285

Pritchard, J. K., & Przeworski, M. (2001). Linkage disequilibrium in humans: Models and data. *American Journal of Human Genetics*, *69*(1), 1–14. https://doi.org/10.1086/321275

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science (New York, N.Y.)*, *273*(5281), 1516–1517. https://doi.org/10.1126/science.273.5281.1516

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J. R. I., Grasby, K. L., Hammerschlag, A. R., Kaminski, J. A., Karlsson, R., Krapohl, E., Lam, M., Nygaard, M., Reynolds, C. A., Trampush, J. W., … Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, *50*(7), 912–919. https://doi.org/10.1038/s41588-018-0152-6

Schadt, E. E. (2009). Molecular networks as sensors and drivers of common human diseases. *Nature*, *461*(7261), 218–223. https://doi.org/10.1038/nature08454

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427. https://doi.org/10.1038/nature13595

Selzam, S., Coleman, J. R. I., Caspi, A., Moffitt, T. E., & Plomin, R. (2018). A polygenic p factor for major psychiatric disorders. *Translational Psychiatry*, *8*(1), 205. https://doi.org/10.1038/s41398-018-0217-4

Smith, G. T., Atkinson, E. A., Davis, H. A., Riley, E. N., & Oltmanns, J. R. (2020). The General Factor of Psychopathology. *Annual Review of Clinical Psychology*, *16*, 75–98. https://doi.org/10.1146/annurev-clinpsy-071119-115848

Sullivan, P. F., & Geschwind, D. H. (2019). Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell*, *177*(1), 162–183. https://doi.org/10.1016/j.cell.2019.01.015

Torkamani, A., Wineinger, N. E., & Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews. Genetics*. https://doi.org/10.1038/s41576-018-0018-x

van Oudheusden, L. J. B., Meynen, G., & van Balkom, A. J. L. M. (2015). [Psychiatric comorbidity: Theoretical and clinical dilemmas]. *Tijdschrift Voor Psychiatrie*, *57*(9), 664–671.

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *American Journal of Human Genetics*, *90*(1), 7–24. https://doi.org/10.1016/j.ajhg.2011.11.029

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*, *101*(1), 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T. J. C., van der Sluis, S., Andreassen, O. A., Neale, B. M., & Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics*. https://doi.org/10.1038/s41588-019-0481-0

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, *8*(1), 1826. https://doi.org/10.1038/s41467-017-01261-5

Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A. E., Dudbridge, F., & Middeldorp, C. M. (2014). Research review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *55*(10), 1068–1087. https://doi.org/10.1111/jcpp.12295

Wray, N. R., Lin, T., Austin, J., McGrath, J. J., Hickie, I. B., Murray, G. K., & Visscher, P. M. (2020). From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2020.3049

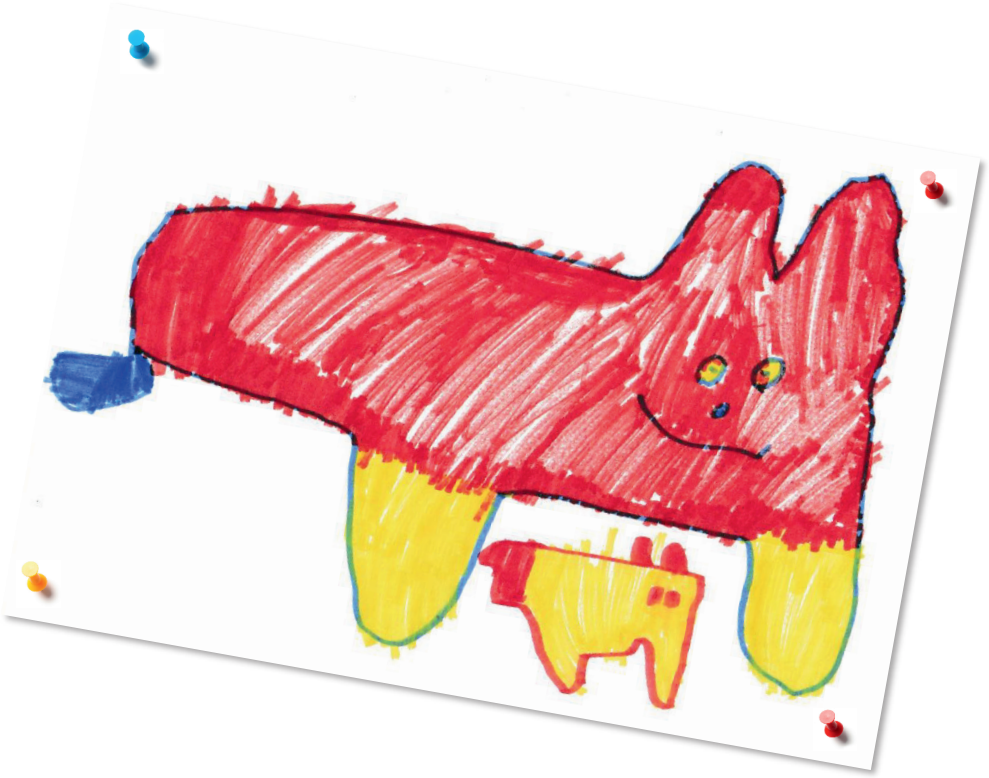
Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Adams, M. J., Agerbo, E., Air, T. M., Andlauer, T. M. F., Bacanu, S.-A., Bækvad-Hansen, M., Beekman, A. F. T., Bigdeli, T. B., Binder, E. B., Blackwood, D. R. H., Bryois, J., Buttenschøn, H. N., Bybjerg-Grauholm, J., … Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, *50*(5), 668–681. https://doi.org/10.1038/s41588-018-0090-3

Zhu, J., Zhang, B., & Schadt, E. E. (2008). A systems biology approach to drug discovery. *Advances in Genetics*, *60*, 603–635. https://doi.org/10.1016/S0065-2660(07)00421-X

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(4), 1193–1198. https://doi.org/10.1073/pnas.1119675109

2

Sample description and GWAS data used in this thesis



Sample description and GWAS data used in this thesis

2.1 Clinical sample "inside-out"

2.1.1 Sample description

In this thesis we used, for the first time, data of a large Dutch child and adolescent psychiatric sample, called 'Inside-Out'. Data were collected at the department of Child and Adolescent Psychiatry of the Sophia Children's Hospital, Erasmus Medical Center in Rotterdam from January 2001 until January 2012 resulting in a psychiatric outpatient sample. For this study, ethical approval of the Erasmus Medical center was obtained. The diagnostic classification was performed by a clinician according to the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV). Before the first visit, parents and children filled out the Child Behavior Check List (CBCL) from the Achenbach System of Empirically Based Assessment (ASEBA) (Achenbach & Rescorla, 2001). The on-site procedure consisted of a semi-structured interview with the child based on the Semi-structured Clinical Interview for Children and Adolescents (McConaughy & Achenbach, 2001), the Diagnostic Interview Schedule for Children IV-P (Shaffer et al., 2000), an interview with parents, and the Autism Diagnostic Observation Schedule-Generic (Lord et al., 1989) in case of a suspected autism spectrum disorder (ASD). This procedure was part of the standard clinical practice.

DNA was extracted from saliva and genotyping was performed on the Illumina PsychChip array(https://www.illumina.com/products/by-type/micproarray-kits/infinium-psycharray.html) and Global Screening Array (GSA; ; https://www.illumina.com/products/by-type/microarray-kits/infinium-global-screening.html) at two different points in time. 32% was genotyped on the GSA chip and 68% on the Illumina PsychChip array, see Table 2.1 for details.

The study presented in Chapter 3 used data of the 1402 children diagnosed with one or more of the following DSM-IV classification categories, and subcategories of mentioned classifications: autism spectrum disorders (ASD), attention-deficit/hyperactivity disorders (ADHD), behavioral disorders, tic disorders, obsessive compulsive disorders (OCD), anxiety disorders, mood disorders, eating disorders, other disorders of infancy, childhood or adolescence, motor, learning and communication disorders, somatoform disorders, trauma- and stress-related disorders and comorbid intellectual disability. Of the 1402 participants, 61% was male. The age range was 1-21 years, (mean 9.59, SD 3.69).

The study presented in Chapter 4 focused on a selection of participants from the Inside-Out sample who were diagnosed with a neurodevelopmental disorder that was either ASD, ADHD or both (N = 688, age range 2.5–18.5, mean: 8.96, SD: 3.07, 76% male) resulting in three samples to be tested: (1) a single diagnosis of ASD (N = 295, age range: 2.5–18.3 years, mean: 9.02, SD: 3.55, 73% male), (2) a single diagnosis of ADHD (N = 280, age range: 3.3–18.5 years, mean: 9.06, SD: 2.66, 75% male), and (3) combining the two diagnoses, thus subjects with either ASD, ADHD or both (N = 688, age range 2.5–18.5, mean: 8.96, SD: 3.07, 76% male). All participants in this study were genotyped on the Illumina PsychChip array.

Table 2.1 Diagnostic classifications and the array used for genotyping.

Genotype chip	Phenotypes	Chapter
Illumina PsychChip array	ASD, ADHD, eating disorders, OCD, tic disorder	4
Global Screening Array (GSA)	Anxiety disorders, disruptive behaviors, eating disorders, disorder of infancy, childhood or adolescence NOS, mood disorders, motor, learning and communication disorders, somatoform disorders, trauma/stress	3, 4

If children received more than one diagnosis with any one of those diagnoses being autism spectrum disorder (ASD), attention-deficit/hyperactivity disorders (ADHD), eating disorder, obsessive compulsive disorders (OCD) or tic disorder, the child was genotyped on the Illumina PsychChip array.

2.1.2 Diagnoses

The diagnostic classifications in the sample have been collapsed into the follow overarching categories: ASD, ADHD, anxiety disorders, disruptive behavior, eating disorders, OCD, disorder of infancy, childhood or adolescence, mood disorders, motor, learning, or communication developmental disorder, somatic symptoms and trauma/stress. Figure 1 shows the diagnostic composition as the amount of cases in the full sample per classification category and the amount of cases per classification category in diagnosed with intellectual disability, as well as the percentage of the full sample with an intellectual disability diagnosis, intellectual disability status. Comorbidities are allowed in the sample and do occur. To provide more insights into the comorbidities a table with comorbidities per classification category has been added below (Table 2.2). In addition, for the three largest classification categories, i.e. ASD, ADHD, and anxiety disorders, Figure 2 has been added.

Figure 1. Diagnostic composition Inside-Out and intellectual disability status details

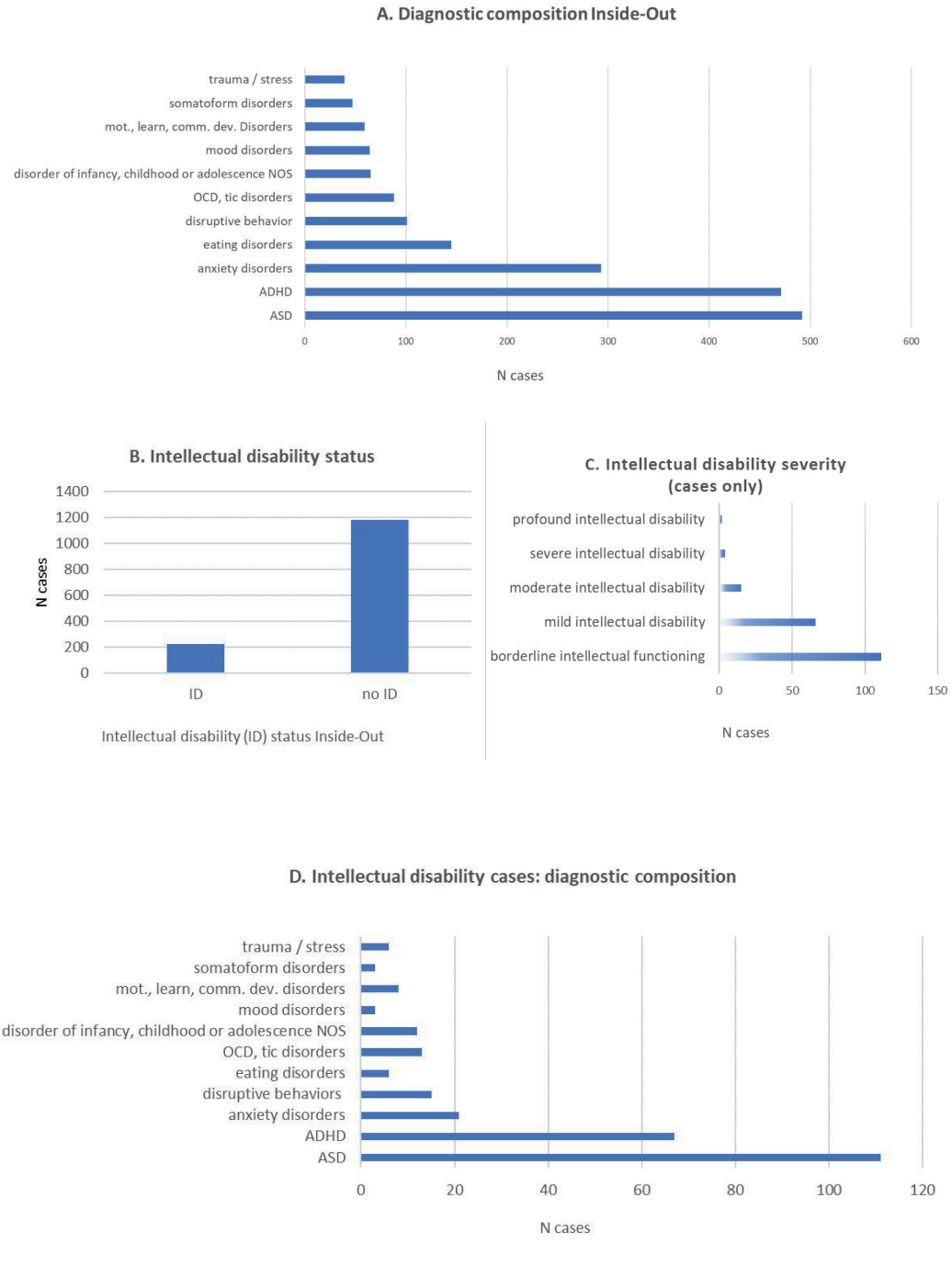
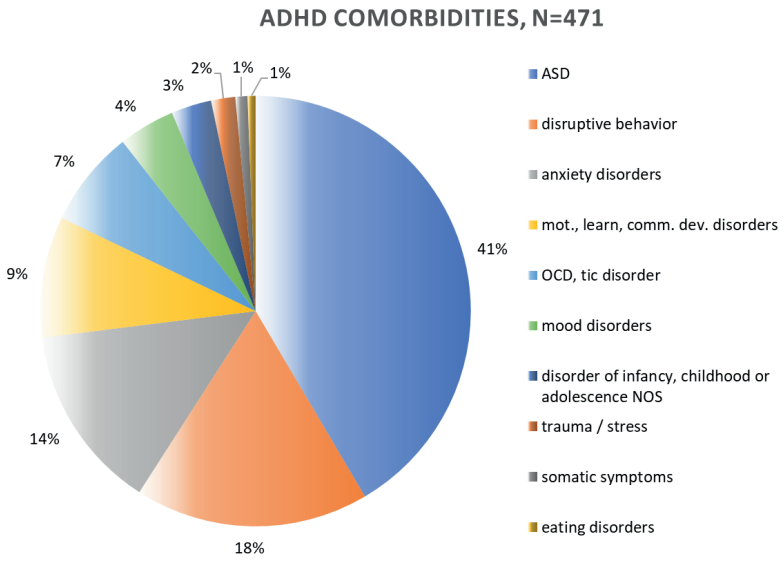


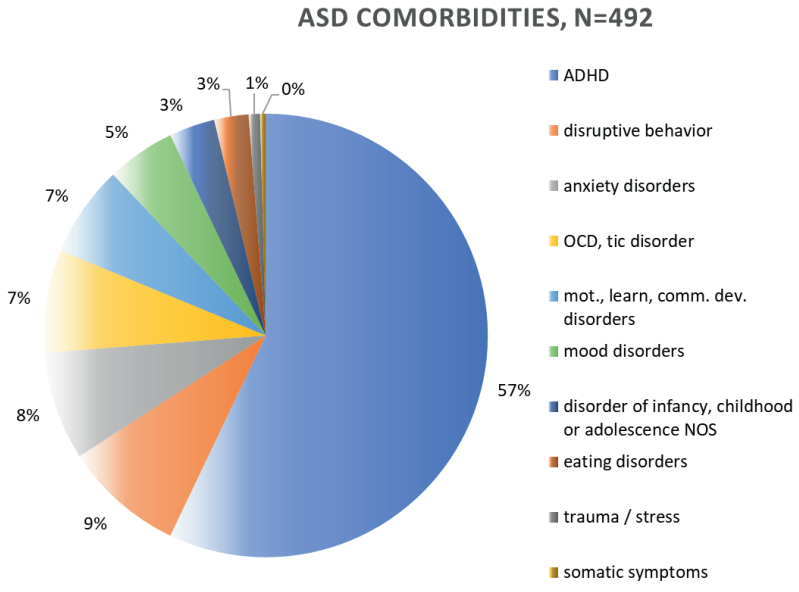
Table 2.2 Overview of the comorbidities per diagnostic category in Inside-out

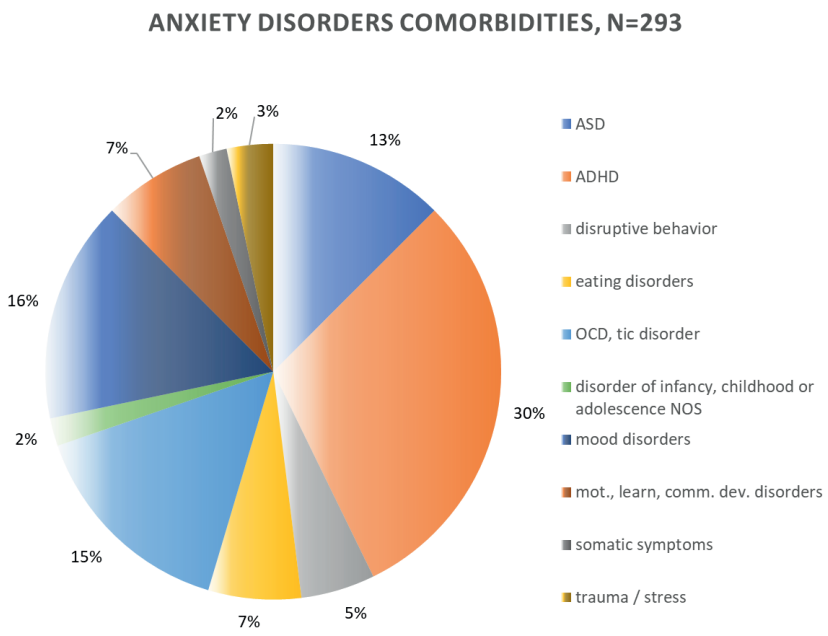
	ASD	ADHD	AD	DB	ED	OCD	TD	DI-NOS	MD	MLC-DD	SS	TS
ASD (N=492)												
ADHD (N=471)	137											
anxiety disorders (N=293)	19	46										
disruptive behavior (N=101)	21	58	8									
eating disorder (N=145)	6	2	10	0								
OCD (N=43)	5	5	21	2	1							
tic disorder (N=50)	13	19	14	3	0	4						
disorder of infancy, childhood or adolescence NOS (N=65)	8	10	3	2	0	0	0					
mood disorders (N=64)	12	14	24	7	5	4	3	0				
mot., learn, comm. dev. disorder (N=59)	16	30	11	1	0	0	1	2	6			
somatic symptoms (N=47)	1	3	3	0	1	0	0	0	3	3		
trauma / stress (N=39)	2	6	5	1	4	1	0	0	1	1	1	

ASD: autism spectrum disorders, ASDHD: attention deficit / hyperactivity disorder, AD: anxiety disorder, DB: disruptive behavior, ED: eating disorder, OCD: obsessive compulsive disorder, TD: tic disorder, DI-NOS: disorder of infancy, childhood or adolescence NOS, MD: mood disorder, MLC -DD: motor, learning, communication. developmental disorder, SS: somatic symptoms, TS: trauma / stress

Figure 2 Visualization of the comorbidities in the three largest diagnostic categories (ADHD, ASD, anxiety disorders) in Inside-Out







2.2 Traits of GWAS data studied in this thesis

We selected outcomes of several existing GWAS to calculate PRS in our sample, aiming to quantify the predictive capacity of common genetic variation of a variety of psychiatric disorders and psychological traits as captured by their PRS. To limit the burden of multiple testing correction, we tested 14 polygenic risk scores (PRS) instead of all psychiatric and psychological traits of which GWAS were available. The choice of psychiatric and psychological traits was based on A) presence in our sample: we aimed to include all the disorders present in Inside-out, or B) reported moderate to high genetic correlations with the disorders present in our sample, and C) a GWAS being powerful enough to generate reliable estimated betas (i.e., > 7,000 cases, and >14,000 controls in the GWAS). This criterium resulted in the exclusion of GWAS data of obsessive compulsive disorder (International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS), 2018), anorexia nervosa (Duncan et al., 2017), and Tourette Syndrome (Scharf et al., 2013), at the time those samples were to small. We included ASD, ADHD, schizophrenia, major depression disorder, bipolar disorder, alcohol dependence, anxiety disorders, neuroticism, insomnia, anti-social behavior, risk taking behavior, smoking initiation, intelligence and educational attainment. The genetic correlations between those 14 traits can be found in Chapter 3.

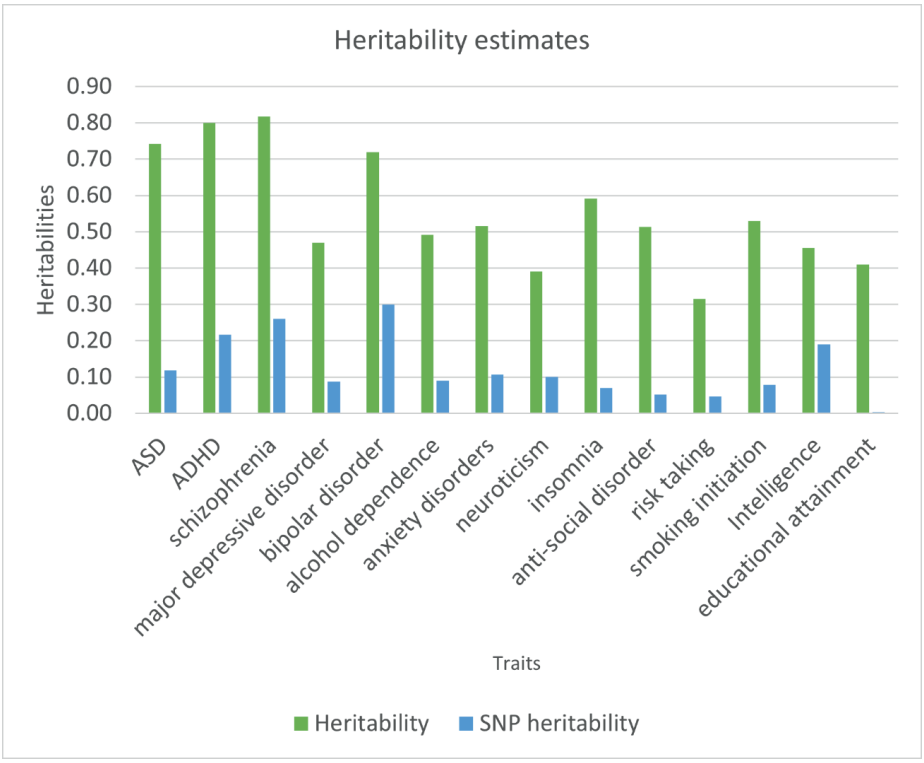
2.2.1 Schematic overview of the heritability of the traits and disorders based on GWAS

Many human traits, including psychiatric disorders, are heritable (Polderman et al., 2015). All psychiatric disorders and psychological traits discussed in this thesis have a moderate to high heritability. Figure 3 shows twin-based heritability estimates (green), which range from 0.3 for risk taking behavior to 0.8 for schizophrenia (see also Table 2.3). The SNP based heritability estimates (blue) are all much lower, ranging from 0.003 for educational attainment to 0.3 for bipolar disorder, showing the magnitude of the missing heritability. The missing heritability is defined as the difference between the estimated heritability based on twin and family research and the estimated SNP heritability based on GWAS (Eichler et al., 2010).

Table 2.3 Twin and SNP based heritability estimates, as reported in the literature

Trait	Twin heritability	SNP heritability
ASD	0.74 (Tick et al., 2016)	0.12 (Grove et al., 2019)
ADHD	0.70 - 0.80 (Brikell et al., 2015)	0.21(Demontis et al., '19)
Schizophrenia	0.80 (Sullivan et al., 2003)	0.26(Pettersson et al., '19)
Major depressive disorder	0.40 (Wray et al., 2018)	0.09 (Wray et al., 2018)
Bipolar disorder	0.70 (Stahl et al., 2019)	0.30 (Stahl et al., 2019)
Alcohol dependence	0.49 (Walters et al., 2018)	0.09 (Walters et al., 2018)
Anxiety disorders	0.30 - 0.50 (Otowa et al., 2016)	0.11 (Otowa et al., 2016)
Neuroticism	0.39 (Vukasović et al., 2015)	0.10 (Nagel et al., 2018)
Insomnia	0.38 - 0.59 (P.R. Jansen et al., '19)	0.07 (P.R. Jansen et al., '19)
Anti-social behavior	0.50 (Tielbeek et al., 2017)	0.05 (Tielbeek et al., 2017)
Risk taking behavior	0.30 (Linnér et al., 2019)	0.05 (Linnér et al., 2019)
Smoking initiation	0.50 (Hicks et al., 2011)	0.08 (Liu et al., 2019)
intelligence	0.45 (Polderman et al., 2015)	0.19 (Savage et al., 2018)
Educational attainment	0.40 (Branigan et al., 2013)	0.003 (Lee et al., 2018)

Figure 3 Twin and SNP based heritability estimates based on the estimates above



2.2.2 Overview of the GWAS used per chapter

Genetic correlations and polygenic risk scores (PRS) as applied in this thesis were based on summary statistics of the largest most up-to-date GWAS data at the time. The target sample Inside-Out was not included in any of the GWAS used.

2.2.2.1 Discovery GWAS in Chapter 3

Table 2.4 Schematic overview of GWAS studies used in Chapter 3

Trait	Authors and sample sizes
Dichotomous traits:	
ASD:	Grove et al., 2019 (Cases N: 18,381 Controls N: 27,969).
ADHD:	Demontis et al., 2019 (Cases N: 20,183 Controls N: 35,191)
Schizophrenia:	Pardiñas et al., 2018 (Cases N: 40,675 Controls N: 64,643)
Major depressive disorder:	Wray et al., 2018 (*Cases N: 135,458 Controls N: 344,901)
Bipolar disorder:	Stahl et al., 2019 (Cases N: 20,352 Controls N: 31,358)
Alcohol dependence:	Walters et al., 2018 (Cases N: 14,904 Controls N: 37,944)
Anxiety disorders:	Otowa et al., 2016 (Cases N: 7,016 Controls N: 14,745)
Continuous traits:	
Neuroticism:	Nagel et al., 2018 (N: 390,278)
Insomnia:	P. R. Jansen et al., 2019 (N: 386,533)
Anti-social behavior:	Tielbeek et al., 2017 (N: 32,000)
Risk taking:	Linnér et al., 2019 (939,908)
Smoking initiation:	Liu et al., 2019 (N: 1,232,091)
Intelligence:	Savage et al., 2018 (N: 269,867)
Educational attainment:	Lee et al., 2018 (N: 1,131,881)

2.2.2.2 Discovery GWAS in Chapter 4

Table 2.5 Schematic overview of studies used in Chapter 4, all dichotomous traits.

Trait	Authors and sample sizes
ASD:	Grove et al., 2019 (cases N: 18,382 and controls N: 27,969)
ADHD:	Demontis et al., 2019 (cases N: 20,183 and controls N: 35,191)
Schizophrenia:	Pardiñas et al., 2018 (cases N: 40,675 and controls N: 64,643)

2.2.2.3 Discovery GWAS in Chapter 5

For Chapter 5 only one set of GWAS summary statistics was used. The PGC-ASD workgroup generated summary statistics from a meta-analysis of 5,305 ASD diagnosed cases of ASD and 5,305 pseudocontrols of European descent. These summary statistics have been used in several publications (A. Jansen et al., 2017; Robinson et al., 2016) but have not led to a published GWAS at the time. The summary statistics were obtained from the PGC website (https://www.med.unc.edu/pgc/download-results/) March 2015.

2.2.3 A closer look at the traits presented in the GWAS

2.2.3.1 Autism Spectrum Disorder (ASD)

ASD is a neurodevelopmental disorder characterized by persistent deficits in social communication and social interaction across multiple contexts and restricted, repetitive patterns of behavior, interests, or activities (*DSM 5*, 2013)**.** These symptoms must be present in the early developmental period and cause clinically significant impairment in social, occupational, or other important areas of current functioning. It can be specified with or without intellectual disability and/or language impairment or other neurodevelopmental, mental or behavioral disorders (*DSM 5*, 2013). It has a reported prevalence around 1% (Tomlinson et al., 2014). In addition, the gradual rise in ASD prevalence can also be explained by increased awareness, different assessment strategies and broader diagnostic criteria possibly including milder neurodevelopmental disorder cases which border on typical individuals (Graf et al., 2017). ASD is a genetic complex trait with a heritability of 74% (Tick et al., 2016) and a SNP heritability of 11% (Grove et al., 2019). This large difference between family based heritability and SNP heritability might have its origin in the genetic architecture of ASD which might include more rare variants than common variants compared to other complex psychiatric disorders. GWAS have implicated several biological pathways relating to neural function and brain development (Grove et al., 2019**)**. Large-scale exome sequencing studies supported and refined these findings by implicating genes which are expressed and enriched in inhibitory and excitatory neuronal cell-lineages and stating that most of the implicated genes regulate other genes or affect synapses (Satterstrom et al., 2020).

2.2.3.2 Attention Deficit /Hyperactivity Disorder (ADHD)

ADHD is a neurodevelopmental disorder which is diagnosed if six or more symptoms of inattention and/or more than six symptoms of hyperactivity/impulsivity have persisted for over six months with the behavior being inconsistent with the developmental stage of the individual. The behavior needs to have a negative impact on social or occupational activities,

be present in more than two settings and several symptoms must be present before the age of 7 according to DSM-IV (American Psychiatric Association, 1994) and the age of 12 according to DSM 5 (*DSM 5*, 2013). ADHD has a high population prevalence of 2.5% in adults and 5% in children (Demontis et al., 2019) hence carrying a high social burden. It is a genetically complex trait with a heritability ranging between 70 and 80% (Brikell et al., 2015) for both adults and children and a reported SNP heritability of 21.6% (Demontis et al., 2019). The largest GWAS to date reports that the significantly associated loci are located in or near genes that implicate neurodevelopmental processes (Demontis et al., 2019).

2.2.3.3 Schizophrenia

Schizophrenia is a mental disorder that presents itself with psychotic symptoms of hallucinations, delusions and disorganized speech and negative symptoms as diminished expressiveness, lowered motivation and impaired executive functioning, speed of mental processing and impaired memory. To receive a schizophrenia diagnosis, for at least a one month period, a person must present at least hallucinations, delusions or disorganized speech in addition to disorganized behavior or negative symptoms (*DSM 5*, 2013). It is a psychotic disorder with a population prevalence of 1% (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). It has a heritability of 80% (Sullivan et al., 2003) and a SNP heritability of 26% (Pettersson et al., 2019). GWAS show a role for mutation intolerant genes involved in glutamatergic neurotransmission (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), genes encoding voltage-gated calcium channels (Pardiñas et al., 2018; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), targets of FMRP (Pardiñas et al., 2018), abnormal behavior (Pardiñas et al., 2018), 5-HT2c receptor complex (Pardiñas et al., 2018), abnormal nervous system electrophysiology (Pardiñas et al., 2018) and abnormal long-term potentiation (Pardiñas et al., 2018).

2.2.3.4 Major depressive disorder

Major depressive disorder is a debilitating mood disorder in which a person experiences depressive symptoms for at least two weeks. Five or more of the following symptoms must be present: recurrent thoughts of death, diminished ability to think or concentrate, feelings of guilt or worthlessness, fatigue or loss of energy, engaging in purposeless movements, weight loss or decrease or increase of appetite, loss of interest and pleasure in activities and depressed mood most of the day almost every day. At least one of the symptoms needs to be (1) loss of interest or pleasure or (2) depressed mood (*DSM 5*, 2013). It has a point prevalence of 4.7% (Ferrari et al., 2013), a heritability of 40% (Wray et al., 2018) and a SNP heritability of 9% (Wray et al., 2018). The GWAS meta-analysis of which we use the summary statistics is a thoroughly reviewed sample combining 7 cohorts in which they describe the phenotype as

major depression. GWAS findings report that (1) gene expression patterns best match the prefrontal cortex and the anterior cingulate cortex, two regions that show differences in major depressive disorder case-control studies and (2) developmental gene regulatory processes are implicated (Wray et al., 2018).

2.2.3.5 Bipolar disorder

Bipolar disorder is a psychotic disorder consisting of alternating manic and depressive episodes with or without psychotic episodes. To be classified as mania, this state must last for one week and be present most days, most of the day (*DSM 5*, 2013). More than three of the following symptoms must be present: grandiosity, increased goal activity, decreased need for sleep, easily distracted, racing thoughts, increased talkativeness, engaging in activities that hold potential painful consequences. To be classified as a depressive episode, a person must experience depressive symptoms for at least two weeks. Five or more of the following symptoms must be present: recurrent thoughts of death, diminished ability to think or concentrate, feelings of guilt or worthlessness, fatigue or loss of energy, engaging in purposeless movements, weight loss or decrease or increase of appetite, loss of interest and pleasure in activities and depressed mood most of the day almost every day (*DSM 5*, 2013). It has a lifetime prevalence of 1-2% (Stahl et al., 2019) and a high heritability of 70% (Stahl et al., 2019) with a SNP heritability of 30% (Stahl et al., 2019). The most recent GWAS implicates a role for neurotransmitter function and brain calcium channels, however further studies with larger sample sizes need to be conducted to gain more conclusive insights.

2.2.3.6 Alcohol dependence

Alcohol use disorder as reported in the DSM 5 is diagnosed based on the presence of several of 11 criteria. A mild manifestation of the disorder has a presence of two to three symptoms, a moderate manifestation has a presence of four to five symptoms and a severe manifestation has six or more symptoms.
The 11 criteria are: (1) alcohol is often consumed in larger quantities over a longer period than intended, (2) there is a desire and there are unsuccessful efforts to limit the alcohol use, (3) a large amount of time is spent obtaining alcohol, using alcohol and recovering from the alcohol use, (4) cravings, (5) the alcohol use results in failure to fulfill obligations at work or at home, (6) continued alcohol use despite of recurrent problems due to this, (7) important activities are given up due to the alcohol use, (8) recurrent alcohol use in situations where it is dangerous, (9) continued alcohol use despite of know physical or psychological problems due to this, (10) tolerance, (11) withdrawal (*DSM 5*, 2013). Alcohol dependence has an estimated heritability of 49% and a SNP heritability of 9% (Walters et al., 2018) and a prevalence of 5.8% (*Section 5 PE Tables – Results from the 2018 National Survey on Drug Use and Health: Detailed*

Tables, Sections 1 - 3, SAMHSA, CBHSQ, n.d.). The top associations reported in the most recent GWAS point towards biological pathways affecting alcohol metabolism (Walters et al., 2018).

2.2.3.7 General anxiety disorder

Individuals suffering from general anxiety disorder show excessive worrying in times when no actual threat is present. Criteria assessed are the duration of worrying (> six months), possibility of controlling the worrying (does the worrying easily shift from one topic to another?) and the presence of at least three of the following symptoms in adults and one in children: restlessness, irritability, fatigue, muscle aches, impaired concentration, difficulty sleeping (*DSM 5*, 2013). Anxiety is one of the most commonly occurring mental disorders with a prevalence of 16.4% (Cía et al., 2018) and a heritability of 30 – 50% (Otowa et al., 2016) with a reported SNP heritability of 10.6% (Otowa et al., 2016). The summary statistics used for this phenotype come from a GWAS aiming to identify genetic variants contributing to genetic susceptibility shared across interview-generated DSM-based anxiety disorders (Otowa et al., 2016). To date, GWAS have not yet presented robust biological pathways due to a lack of sample size. However, some interesting loci have been reported among which *CAMKMT* which encodes for a calmodulin-lysine N-methyltransferase. This locus is known for two gene-deletion syndromes, the 2p21 deletion syndrome and the hypotonia-cystinuria syndrome. In addition calcium dependent signaling has also been reported in SCZ and bipolar disorder GWAS (Otowa et al., 2016).

2.2.3.8 Neuroticism

In psychology a well-known model regarding personality traits is the “big five personality theory” by Goldberg (Goldberg, 1993). The five traits included in this theory are used to describe certain aspects of an individual’s personality. Neuroticism is one of the five traits included in this model, next to openness to experience, conscientiousness, extraversion and agreeableness. Neuroticism is viewed as a continuous trait and can be measured by questionnaires such as the Eysenck Personality Questionnaire Revised short form (EPQ-RS) (Eysenck, B.G. et al., 1985). Individuals scoring high on the neuroticism scale are more likely than average scoring individual to experience feelings of fear, anxiety, worry, guilt loneliness and depression. As it is a continuous personality trait a population prevalence cannot be given. It is a heritable trait with an estimated heritability of 15% (Power & Pluess, 2015)and a SNP heritability of 10% (Nagel et al., 2018). A recent GWAS (Nagel et al., 2018) reporting on a meta-analysis of three samples, each measuring neuroticism with a questionnaire of 8 or 12 items, reported the involvement of specific cell types as dopaminergic neuroblasts, medium spiny neurons and serotonergic neurons.

2.2.3.9 Insomnia

Insomnia is a psychiatric disorder described in the DSM5 as a sleep-wake disorder which is characterized by recurrent poor quality or quantity of sleep which manifests itself as difficulty falling asleep, staying asleep, waking up early and being unable to get back to sleep. This results in impairment in daily functioning due to symptoms like (extreme) fatigue, difficulty focusing and feeling irritable. Diagnostic criteria are unhappiness with the quality of sleep and daytime symptoms of impairment of functioning due to the lack of good quality of sleep (*DSM 5*, 2013). With 33% of the population reporting insomnia complaints it is the second most prevalent mental disorder (P. R. Jansen et al., 2019). Insomnia has a heritability estimate of 38% for females and 59% for males (Lind et al., 2015) and a SNP heritability of 7% (P. R. Jansen et al., 2019). The most recent GWAS, based on a meta-analysis with participants providing a representative insomnia measure via questionnaires, implicated axonal parts of neurons, cortical and subcortical tissues and specific cell types including striatal, hypothalamic and claustrum neurons (P. R. Jansen et al., 2019).

2.2.3.10 Anti-social behavior

Anti-social behavior is a broad construct that that encompasses a large range of behaviors including violent felonies, aggression, theft, hostility and deceitfulness (Tielbeek et al., 2017). Anti-social behavior has an estimated heritability of 50% and a SNP heritability of 5.2% (Tielbeek et al., 2017). I use summary statistics of a GWAS reporting on a meta-analysis utilizing eight samples with several measures for anti-social behavior such as DSM-IV conduct disorder, child behavioral checklists conduct problems items as filled out by the mother, and rule breaking behavior as reported by the teacher (Tielbeek et al., 2017). This GWAS does not report on biological pathways but a repeatedly hypothesized biological pathway leading to ASB is the short variant of the serotonin transporter gene polymorphism (5-HTTLPR). A meta-analysis of the literature confirmed this possibility but many methodological issues are reported, hence the study of biological mechanisms in ASB needs larger sample sizes and more solid research designs (Tielbeek et al., 2016).

2.2.3.11 Risk taking

The GWAS we used for this study measures risk taking with the self-report question: would you describe yourself as someone who takes risks? As the resulting variable is on a continuous scale a population prevalence cannot be given. A heritability of 30% (Linnér et al., 2019) and a SNP heritability of 5% (Linnér et al., 2019) has been reported and glutamatergic and GABAergic neurotransmission have been implicated (Linnér et al., 2019).

2.2.3.12 Smoking initiation

Smoking initiation is used as a dichotomous trait providing information on if a person has ever smoked regularly (Liu et al., 2019). It has an estimated heritability of 50% (Hicks et al., 2011) and a SNP heritability of 8% (Liu et al., 2019). A recent large GWAS reported evidence for pathways involved in nicotinic, dopaminergic and glutamatergic neurotransmission in cortical (inferior temporal pathways, dorsolateral and medial prefrontal cortex) and subcortical (hippocampus, caudatus, striatum) regions in the brain (Liu et al., 2019).

2.2.3.13 Intelligence

Intelligence is a psychological trait which can be measured with cognitive and/or IQ tests for example the WAIS (Hartman, 2009), WISC (Na & Burns, 2016) or the SON (Geerlings et al., 2014). It is a trait with a heritability estimate increasing during development from 46 to 65% (Polderman et al., 2015) and a SNP heritability of 19% (Savage et al., 2018). The latest GWAS of intelligence comprised 14 independent cohorts, measuring intelligence through various neurocognitive tests primarily quantifying fluid domains of cognitive functioning (Savage et al., 2018). Genes associated with intelligence are highly brain expressed specifically in hippocampal pyramidal neurons and medium spiny neurons (Savage et al., 2018).

2.2.3.14 Educational attainment

Educational attainment is a trait highly correlated to physical and psychological health and to many other psychological traits (Conti et al., 2010; Lee et al., 2018; Lynch & von Hippel, 2016). It is often measured as years of schooling or highest level of education. The estimated heritability is 40% (Branigan et al., 2013) and a SNP heritability of 0.3% (Lee et al., 2018) was reported. A GWAS investigating the number of years of schooling an individual has completed, reported that the significantly associated SNPs are disproportionally found in genomic regions regulating gene expression in the fetal brain (Okbay et al., 2016). A following GWAS replicated this finding but expanded the list of prioritized genes with postnatally expressed genes (Lee et al., 2018).

2.2.4 A closer look at the phenotypes in Inside-out not represented by GWAS data

Due to a lack of well powered GWAS data for some disorders present in Inside-out have not been described above. These disorders are described here by means of DSM-IV or 5 criteria. The clinical classifications of the participants in the Inside-Out sample were based on DSM-IV criteria, but for the description of clusters of disorders in this thesis they have been grouped

according to DSM 5 criteria, doing just to the most recent scientific insights. These disorders are present in the sample used in Chapter 3.

2.2.4.1 Eating disorders

Eating disorder is a category disorders which is composed of anorexia nervosa, bulimia nervosa, binge eating disorder, other specified feeding and eating disorder, pica, rumination disorder, avoidant/restrictive food intake disorder, unspecified feeding or eating disorder, other: muscle dysmorphia and orthorexia nervosa. In Inside-Out, the main eating disorders are anorexia nervosa and bulimia nervosa. The description will be limited to these two disorders.

Anorexia nervosa: the main three criteria are 1) restriction of energy intake in relation to the required intake leading to a less than minimally expected bodyweight for the appropriate sex, age, health and developmental trajectory, 2) disturbances in the way in which one's body weight is experienced and 3) fear of getting fat and gaining weight even at very low weight. There are two subtypes, the restrictive type which accomplishes weight loss through fasting, dieting and/or excessive exercise and the binge-eating/purging subtype which remains low weight through self induced vomiting, or misuse of diuretics, laxatives or enemas. The severity level is determined by BMI with the mild levels (BMI > 17kg/m^2), moderate levels (BMI 16 – 16.99 kg/m^2), severe levels (BMI 15 – 15.99 kg/m^2) and extreme levels (BMI <15 kg/m^2)(*DSM 5*, 2013).

Bulimia nervosa: the main criteria are 1) recurrent episodes of binge eating characterized by a sense of lack of control over eating during the episode as well as eating in a certain period of time an amount of food much larger than most individuals would during such a time 2) recurrent inappropriate compensatory behaviors to prevent the gain of weight such as self induced vomiting, misuse of diuretics, laxatives or other medications, 3) these two behaviors both occur at least three times a week, 4) self evaluation is influenced by body shape and weight and 5) the disturbance does not only occur during the anorexia nervosa episode. There are levels of severity from mild (average of 1-3 episodes per week), moderate (average of 4-7 episodes per week), severe (average of 8-13 episodes per week) or extreme (average of>14 episodes per week) (*DSM 5*, 2013).

2.2.4.2 Tic disorders

Tic disorders is a category which includes Tourette's disorder. For a tic disorder one must have a motor tic (such as blinking or shrugging the shoulders) or a vocal tic (such as clearing the throat, humming, or yelling out a word or phrase). The tics must have occurred of and on for at least a year. For Tourette's disorder two or more motor and 1 or more vocal tics must have

been present at some time during the illness, though not necessarily concurrently (*DSM 5*, 2013). The symptoms must have occurred before the age of 18 and not be due to physiological effects of substance use or other medical conditions.

2.2.4.3 Disorder of infancy, childhood or adolescence Not otherwise specified (NOS)

A child receives this diagnosis when it is clear that there ought to be a diagnosis of a childhood disorder but when it is not clear due to incomplete information which diagnosis this should be. Over time, when more tests can provide additive information the diagnosis can become more specific (American Psychiatric Association, 1994). This is a DSM-IV diagnostic category. Under the DSM 5, Disorder of Infancy, Childhood, or Adolescence NOS (F32.9) is to be diagnosed as Unspecified depressive disorder, Other specified anxiety disorder, Unspecified anxiety disorder or Consider other diagnosis(es) Disorder of Infancy, Childhood, or Adolescence NOS (F41.9). For practical means we have kept this category as is in our data.

2.2.4.4 Motor, learning or communication developmental disorders

Developmental coordination disorder: in the DSM-IV this disorder was categorized as a learning disorder, in the DSM5 however, it is categorized as a motor disorder in the category neurodevelopmental disorders. The DSM5 describes the child with this disorder as having motor coordination below the level expected for the chronologic age. These children may have had delays in walking and crawling and other early motor milestones and have been described as clumsy. The difficulties the child experiences interfere with daily living activities or academic achievements. The difficulties are not attributable to a medical condition and if they occur in combination with intellectual disability the delay is in excess of those expected for the child's intellectual abilities (*DSM 5*, 2013).

Specific learning disorder, as learning disorders are also known, are neurodevelopmental disorders being during school age and affecting the ability in reading (dyslexia), writing and/or math. The specific diagnostic criteria are difficulties in the school age years which persisted after 6 months of targeted intervention, the individual performs significantly lower on achievement tests than age appropriately expected, the difficulties become apparent in the first years of schooling and another diagnosis such as sensory impairment or low intellectual disability are not better suited explanations (*DSM 5*, 2013).

Communication developmental disorders are characterized by persistent difficulties in the social use of verbal and nonverbal communication, this results in limited effective communication, social participation academic achievement occupational performance or a combination of these and the onset of symptoms is in the early developmental phase. In addition, another (medical) condition is not a more plausible alternative (*DSM 5*, 2013).

2.2.4.5 Somatoform disorders

The DSM-IV somatoform disorder has changed in the DSM5 into somatic symptom and related disorder. The diagnostic criteria for the DSM5 disorder are:

1) One or more somatic symptoms that are distressing or result in significant disruption of daily life.

2) Excessive thoughts, feelings, or behaviors related to the somatic symptoms or associated health concerns as manifested by at least one of the following: a. Disproportionate and persistent thoughts about the seriousness of one's symptoms. b. Persistently high level of anxiety about health or symptoms and c. Excessive time and energy devoted to these symptoms or health concerns

3) although any one somatic symptom may not be continuously present, the state of being symptomatic is persistent (typically more than 6 months).

The main difference between the DSM-IV and DSM5 is that under the DSM-IV the somatic symptoms needed to be medically unexplained, a criterium no longer included in the DSM5 (American Psychiatric Association, 1994; *DSM 5*, 2013).

2.2.4.6 Trauma and stressor related disorders

This group of disorders occurs after a traumatic or stressful event. The three main disorders are post traumatic stress disorder, acute stress disorder and adjustment disorder.

Diagnostic criteria for post-traumatic stress disorder are experience of a traumatic event involving death (actual or threatened), serious injury, or sexual violence, intrusion symptoms that begin after the traumatic event, avoidance of triggering stimuli following the event, negatively affected cognition that begins or worsens after the event, altered reactivity or arousal beginning or worsening after the event. The duration of symptoms last > 1 month following the traumatic event. The affected individual has been experiencing significant distress or impaired social and/or occupational functioning since the traumatic event and the symptoms are not explained by substance misuse or another medical condition (*DSM 5*, 2013).

Diagnostic criteria for acute stress disorder are exposure to death (actual or threatened), injury, or sexual abuse. The duration of symptoms lasts from 3 days to 1 month following the traumatic event. The affected individual has been experiencing significant distress or impaired social and/or occupational functioning since the traumatic event and symptoms are not explained by substance misuse or another medical condition.

The diagnostic criteria for adjustment disorder are emotions or behaviors in response to a stressor that occur within thee months of onset, a clinically significant responses that include

≥ 1 of the following: A level of distress that is disproportionate to the expected response to the stressor, impaired functioning in social, occupational, and/or other important areas, the symptoms are not explained by another mental disorder and are not explained by a normal response to grief. Additionally, the symptoms last ≤ 6 months following resolution of the stressor (*DSM 5*, 2013).

2.2.4.7 Obsessive compulsive and related disorders

This category comprises obsessive compulsive disorder (OCD), trichotillomania (excessive hair pulling), hoarding disorder (difficulty discarding or parting with possessions), skin picking disorder (constant and recurring skin picking) and body dysmorphic disorder (preoccupation with an imagined or minor flaw in one's appearance causing distress.). OCD is characterized by the presence of obsessions and compulsion or both over distress associated with certain unwanted thoughts followed by various behaviors to resist or neutralize the thoughts. These behaviors are time consuming and create significant limitations in other domains of an individuals life, they are not due to physiological effects of medicine or another somatic cause and the disorder cannot be explained better by another psychiatric disorder (*DSM 5*, 2013).

2.2.4.8 Mood disorders

The GWAS used is based on major depressive disorder. In our clinical sample the diagnostic category depression includes more mood disorders. This category in the DSM 5 contains disruptive mood dysregulation disorder, persistent depressive disorder and premenstrual dysphoric disorder. A main symptom these disorders have in common is the presence of depressed affect and diminished pleasure and interest in most or all daily activities influencing the capacity to function on a daily basis (*DSM 5*, 2013)**.**

2.2.4.9 Anxiety disorders

The GWAS used for generating the PRS is based on general anxiety. In our sample the diagnostic category anxiety disorders comprises all disorders included in the category anxiety disorders making our sample phenotypically more heterogeneous compared to the GWAS sample. Included disorders are selective mutism, separation anxiety disorder, specific phobia, generalized anxiety disorder, social phobia and panic disorder with and without agoraphobia. These disorders have in common that the diagnosed individual suffers from excessive fear and anxiety and related behavioral disturbances (*DSM 5*, 2013, p. 5).

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. University of Vermont, Research Center for Children, Youth, & Families.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of mental disorders, 4th* (4th ed.). APA.

Branigan, A. R., McCallum, K. J., & Freese, J. (2013). Variation in the Heritability of Educational Attainment: An International Meta-Analysis. *Social Forces*, *92*(1), 109–140. https://doi.org/10.1093/sf/sot076

Brikell, I., Kuja-Halkola, R., & Larsson, H. (2015). Heritability of attention-deficit hyperactivity disorder in adults. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, *168*(6), 406–413. https://doi.org/10.1002/ajmg.b.32335

Cía, A. H., Stagnaro, J. C., Aguilar Gaxiola, S., Vommaro, H., Loera, G., Medina-Mora, M. E., Sustas, S., Benjet, C., & Kessler, R. C. (2018). Lifetime prevalence and age-of-onset of mental disorders in adults from the Argentinean Study of Mental Health Epidemiology. *Social Psychiatry and Psychiatric Epidemiology*, *53*(4), 341–350. https://doi.org/10.1007/s00127-018-1492-3

Conti, G., Heckman, J., & Urzua, S. (2010). THE EDUCATION-HEALTH GRADIENT. *The American Economic Review*, *100*(2), 234–238. https://doi.org/10.1257/aer.100.2.234

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J. I., Grasby, K. L., Grove, J., … Neale, B. M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics*, *51*(1), 63–75. https://doi.org/10.1038/s41588-018-0269-7

Diagnostic and Statistical Manual of mental disorders (Vol. 5). (2013). American Psychiatric Association.

Duncan, L., Yilmaz, Z., Walters, R., Goldstein, J., Anttila, V., Bulik-Sullivan, B., Ripke, S., Thornton, L., Hinney, A., Daly, M., Sullivan, P., Zeggini, E., Breen, G., & Bulik, C. (2017). Genome-Wide Association Study Reveals First Locus for Anorexia Nervosa and Metabolic Correlations. *The American Journal of Psychiatry*, *174*(9), 850–858. https://doi.org/10.1176/appi.ajp.2017.16121402

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., & Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews. Genetics*, *11*(6), 446–450. https://doi.org/10.1038/nrg2809

Eysenck, B.G., Eysenck, H.J., & Barrett, P.A. (1985). *Eysenck, B. G., Eysenck, H. J. & Barrett, P. A revised version of the psychoticism scale. Pers. Individ. Dif. 6, 21–29 (1985). 3*. *6*, 21–29.

Ferrari, A. J., Somerville, A. J., Baxter, A. J., Norman, R., Patten, S. B., Vos, T., & Whiteford, H. A. (2013). Global variation in the prevalence and incidence of major depressive disorder: A systematic review of the epidemiological literature. *Psychological Medicine*, *43*(3), 471–481. https://doi.org/10.1017/S0033291712001511

Geerlings, H., Laros, J. A., Tellegen, P. J., & Glas, C. A. W. (2014). Testing the difficulty theory of the SON-R 5(1/2)-17, a non-verbal test of intelligence. *The British Journal of Mathematical and Statistical Psychology*, *67*(2), 248–265. https://doi.org/10.1111/bmsp.12017

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *The American Psychologist*, *48*(1), 26–34. https://doi.org/10.1037//0003-066x.48.1.26

Graf, W. D., Miller, G., Epstein, L. G., & Rapin, I. (2017). The autism "epidemic": Ethical, legal, and social issues in a developmental spectrum disorder. *Neurology*, *88*(14), 1371–1380. https://doi.org/10.1212/WNL.0000000000003791

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Awashti, S., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Christensen, J. H., … 23andMe Research Team. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, *51*(3), 431–444. https://doi.org/10.1038/s41588-019-0344-8

Hartman, D. E. (2009). Wechsler Adult Intelligence Scale IV (WAIS IV): Return of the gold standard. *Applied Neuropsychology*, *16*(1), 85–87. https://doi.org/10.1080/09084280802644466

Hicks, B. M., Schalet, B. D., Malone, S. M., Iacono, W. G., & McGue, M. (2011). Psychometric and genetic architecture of substance use disorder and behavioral disinhibition measures for gene association studies. *Behavior Genetics*, *41*(4), 459–475. https://doi.org/10.1007/s10519-010-9417-2

International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS). (2018). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Molecular Psychiatry*, *23*(5), 1181–1188. https://doi.org/10.1038/mp.2017.154

Jansen, A., Dieleman, G. C., Smit, A. B., Verhage, M., Verhulst, F. C., Polderman, T. J. C., & Posthuma, D. (2017). Gene-set analysis shows association between FMRP targets and autism spectrum disorder. *European Journal of Human Genetics: EJHG*. https://doi.org/10.1038/ejhg.2017.55

Jansen, P. R., Watanabe, K., Stringer, S., Skene, N., Bryois, J., Hammerschlag, A. R., Leeuw, C. A. de, Benjamins, J. S., Muñoz-Manchado, A. B., Nagel, M., Savage, J. E., Tiemeier, H., White, T., Tung, J. Y., Hinds, D. A., Vacic, V., Wang, X., Sullivan, P. F., Sluis, S. van der, … Posthuma, D. (2019). Genome-wide analysis of insomnia in 1,331,010 individuals identifies

new risk loci and functional pathways. *Nature Genetics*, *51*(3), 394. https://doi.org/10.1038/s41588-018-0333-3

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., Fontana, M. A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P. N., Walters, R. K., Willoughby, E. A., … Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*(8), 1112–1121. https://doi.org/10.1038/s41588-018-0147-3

Lind, M. J., Aggen, S. H., Kirkpatrick, R. M., Kendler, K. S., & Amstadter, A. B. (2015). A Longitudinal Twin Study of Insomnia Symptoms in Adults. *Sleep*, *38*(9), 1423–1430. https://doi.org/10.5665/sleep.4982

Linnér, R. K., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., Lebreton, M., Tino, S. P., Abdellaoui, A., Hammerschlag, A. R., Nivard, M. G., Okbay, A., Rietveld, C. A., Timshel, P. N., Trzaskowski, M., Vlaming, R. de, Zünd, C. L., Bao, Y., Buzdugan, L., … Beauchamp, J. P. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics*, *51*(2), 245. https://doi.org/10.1038/s41588-018-0309-3

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., Zhan, X., Choquet, H., Docherty, A. R., Faul, J. D., Foerster, J. R., Fritsche, L. G., Gabrielsen, M. E., Gordon, S. D., Haessler, J., … Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, *51*(2), 237. https://doi.org/10.1038/s41588-018-0307-5

Lord, C., Rutter, M., heemsbergen, J., Jordan, H., Mawhood, l, & Schopler, e. (1989). *Autism diagnostic observation schedule: A standardized observation of communicative and social behavior*. *19(2)*, 185–212.

Lynch, J. L., & von Hippel, P. T. (2016). An education gradient in health, a health gradient in education, or a confounded gradient in both? *Social Science & Medicine (1982)*, *154*, 18–27. https://doi.org/10.1016/j.socscimed.2016.02.029

McConaughy, S. H., & Achenbach, T. M. (2001). *Manual for the Semi-structured Clinical Interview for Children and Adolescents*.

Na, S. D., & Burns, T. G. (2016). Wechsler Intelligence Scale for Children-V: Test Review. *Applied Neuropsychology. Child*, *5*(2), 156–160. https://doi.org/10.1080/21622965.2015.1015337

Nagel, M., Jansen, P. R., Stringer, S., Watanabe, K., de Leeuw, C. A., Bryois, J., Savage, J. E., Hammerschlag, A. R., Skene, N. G., Muñoz-Manchado, A. B., 23andMe Research Team, White, T., Tiemeier, H., Linnarsson, S., Hjerling-Leffler, J., Polderman, T. J. C., Sullivan, P. F., van der Sluis, S., & Posthuma, D. (2018). Meta-analysis of genome-wide association studies

for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genetics*, *50*(7), 920–927. https://doi.org/10.1038/s41588-018-0151-7

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., Turley, P., Chen, G.-B., Emilsson, V., Meddens, S. F. W., Oskarsson, S., Pickrell, J. K., Thom, K., Timshel, P., de Vlaming, R., Abdellaoui, A., Ahluwalia, T. S., Bacelis, J., Baumbach, C., … Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*(7604), 539–542. https://doi.org/10.1038/nature17671

Otowa, T., Hek, K., Lee, M., Byrne, E. M., Mirza, S. S., Nivard, M. G., Bigdeli, T., Aggen, S. H., Adkins, D., Wolen, A., Fanous, A., Keller, M. C., Castelao, E., Kutalik, Z., Van der Auwera, S., Homuth, G., Nauck, M., Teumer, A., Milaneschi, Y., … Hettema, J. M. (2016). Meta-analysis of genome-wide association studies of anxiety disorders. *Molecular Psychiatry*, *21*(10), 1391–1399. https://doi.org/10.1038/mp.2015.197

Pardiñas, A. F., Holmans, P., Pocklington, A. J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S. E., Bishop, S., Cameron, D., Hamshere, M. L., Han, J., Hubbard, L., Lynham, A., Mantripragada, K., Rees, E., MacCabe, J. H., McCarroll, S. A., Baune, B. T., Breen, G., … CRESTAR Consortium. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature Genetics*, *50*(3), 381–389. https://doi.org/10.1038/s41588-018-0059-2

Pettersson, E., Lichtenstein, P., Larsson, H., Song, J., Attention Deficit/Hyperactivity Disorder Working Group of the iPSYCH-Broad-PGC Consortium, Autism Spectrum Disorder Working Group of the iPSYCH-Broad-PGC Consortium, Bipolar Disorder Working Group of the PGC, Eating Disorder Working Group of the PGC, Major Depressive Disorder Working Group of the PGC, Obsessive Compulsive Disorders and Tourette Syndrome Working Group of the PGC, Schizophrenia CLOZUK, Substance Use Disorder Working Group of the PGC, Agrawal, A., Børglum, A. D., Bulik, C. M., Daly, M. J., Davis, L. K., Demontis, D., Edenberg, H. J., Grove, J., Gelernter, J., Neale, B. M., Pardiñas, A. F., Stahl, E., Walters, J. T. R., Walters, R., … Polderman, T. J. C. (2019). Genetic influences on eight psychiatric disorders based on family data of 4 408 646 full and half-siblings, and genetic data of 333 748 cases and controls. *Psychological Medicine*, *49*(7), 1166–1173. https://doi.org/10.1017/S0033291718002039

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), 702–709. https://doi.org/10.1038/ng.3285

Power, R. A., & Pluess, M. (2015). Heritability estimates of the Big Five personality traits based on common genetic variants. *Translational Psychiatry*, *5*, e604. https://doi.org/10.1038/tp.2015.96

Robinson, E. B., St. Pourcain, B., Anttila, V., Kosmicki, J. A., Bulik-Sullivan, B., Grove, J., Maller, J., Samocha, K. E., Sanders, S. J., Ripke, S., Martin, J., Hollegaard, M. V., Werge, T., Hougaard, D. M., Neale, B. M., Evans, D. M., Skuse, D., Mortensen, P. B., Børglum, A. D., …

Daly, M. J. (2016). Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nature Genetics*, *48*(5), 552–555. https://doi.org/10.1038/ng.3529

Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M. S., Artomov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., … Walters, R. K. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*. https://doi.org/10.1016/j.cell.2019.12.036

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J. R. I., Grasby, K. L., Hammerschlag, A. R., Kaminski, J. A., Karlsson, R., Krapohl, E., Lam, M., Nygaard, M., Reynolds, C. A., Trampush, J. W., … Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, *50*(7), 912–919. https://doi.org/10.1038/s41588-018-0152-6

Scharf, J. M., Yu, D., Mathews, C. A., Neale, B. M., Stewart, S. E., Fagerness, J. A., Evans, P., Gamazon, E., Edlund, C. K., Service, S., Tikhomirov, A., Osiecki, L., Illmann, C., Pluzhnikov, A., Konkashbaev, A., Davis, L. K., Han, B., Crane, J., Moorjani, P., … Pauls, D. L. (2013). Genome-wide association study of Tourette Syndrome. *Molecular Psychiatry*, *18*(6), 721–728. https://doi.org/10.1038/mp.2012.69

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427. https://doi.org/10.1038/nature13595

Section 5 PE Tables – Results from the 2018 National Survey on Drug Use and Health: Detailed Tables, Sections 1—3, SAMHSA, CBHSQ. (n.d.). Retrieved March 12, 2020, from https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHDetailedTabs2018R2/NSDUHDetTabsSect5pe2018.htm#tab5-4b

Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*(1), 28–38. https://doi.org/10.1097/00004583-200001000-00014

Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J. R. I., Gaspar, H. A., de Leeuw, C. A., Steinberg, S., Pavlides, J. M. W., Trzaskowski, M., Byrne, E. M., Pers, T. H., Holmans, P. A., Richards, A. L., Abbott, L., … Bipolar Disorder Working Group of the Psychiatric Genomics Consortium. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics*, *51*(5), 793–803. https://doi.org/10.1038/s41588-019-0397-8

Sullivan, P. F., Kendler, K. S., & Neale, M. C. (2003). Schizophrenia as a complex trait: Evidence from a meta-analysis of twin studies. *Archives of General Psychiatry*, *60*(12), 1187–1192. https://doi.org/10.1001/archpsyc.60.12.1187

Tick, B., Bolton, P., Happé, F., Rutter, M., & Rijsdijk, F. (2016). Heritability of autism spectrum disorders: A meta-analysis of twin studies. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *57*(5), 585–595. https://doi.org/10.1111/jcpp.12499

Tielbeek, J. J., Johansson, A., Polderman, T. J. C., Rautiainen, M.-R., Jansen, P., Taylor, M., Tong, X., Lu, Q., Burt, A. S., Tiemeier, H., Viding, E., Plomin, R., Martin, N. G., Heath, A. C., Madden, P. A. F., Montgomery, G., Beaver, K. M., Waldman, I., Gelernter, J., … Broad Antisocial Behavior Consortium collaborators. (2017). Genome-Wide Association Studies of a Broad Spectrum of Antisocial Behavior. *JAMA Psychiatry*, *74*(12), 1242–1250. https://doi.org/10.1001/jamapsychiatry.2017.3069

Tielbeek, J. J., Karlsson Linnér, R., Beers, K., Posthuma, D., Popma, A., & Polderman, T. J. C. (2016). Meta-analysis of the serotonin transporter promoter variant (5-HTTLPR) in relation to adverse environment and antisocial behavior. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, *171*(5), 748–760. https://doi.org/10.1002/ajmg.b.32442

Tomlinson, M., Yasamy, M. T., Emerson, E., Officer, A., Richler, D., & Saxena, S. (2014). Setting global research priorities for developmental disabilities, including intellectual disabilities and autism. *Journal of Intellectual Disability Research: JIDR*, *58*(12), 1121–1130. https://doi.org/10.1111/jir.12106

Vukasović, T., & Bratko, D. (2015). Heritability of personality: A meta-analysis of behavior genetic studies. *Psychological Bulletin*, *141*(4), 769–785. https://doi.org/10.1037/bul0000017

Walters, R. K., Polimanti, R., Johnson, E. C., McClintick, J. N., Adams, M. J., Adkins, A. E., Aliev, F., Bacanu, S.-A., Batzler, A., Bertelsen, S., Biernacka, J. M., Bigdeli, T. B., Chen, L.-S., Clarke, T.-K., Chou, Y.-L., Degenhardt, F., Docherty, A. R., Edwards, A. C., Fontanillas, P., … Agrawal, A. (2018). Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature Neuroscience*, *21*(12), 1656–1669. https://doi.org/10.1038/s41593-018-0275-1

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Adams, M. J., Agerbo, E., Air, T. M., Andlauer, T. M. F., Bacanu, S.-A., Bækvad-Hansen, M., Beekman, A. F. T., Bigdeli, T. B., Binder, E. B., Blackwood, D. R. H., Bryois, J., Buttenschøn, H. N., Bybjerg-Grauholm, J., … Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, *50*(5), 668–681. https://doi.org/10.1038/s41588-018-0090-3

3

The predictive capacity of psychiatric and psychological polygenic risk scores for distinguishing cases in a child and adolescent psychiatric sample from controls

Jansen AG, Jansen PR, Savage JE, Kraft J, Skarabis N, Polderman TJC, Dieleman GC

J Child Psychol Psychiatry. 2021 Sep; 62(9): 1079–1089

Abstract

Psychiatric traits are heritable, highly comorbid and genetically correlated, suggesting that genetic effects that are shared across disorders are at play. The aim of the present study is to quantify the predictive capacity of common genetic variation of a variety of traits, as captured by their PRS, to predict case-control status in a child and adolescent psychiatric sample including controls to reveal which traits contribute to the shared genetic risk across disorders.
Method: polygenic risk scores (PRS) of 14 traits were used as predictor phenotypes to predict case-control status in a clinical sample. Clinical cases (N= 1,402), age 1–21, diagnostic categories: autism spectrum disorders (N= 492), attention-deficit/ hyperactivity disorders (N = 471), anxiety (N= 293), disruptive behaviors (N= 101), eating disorders (N= 97), OCD (N= 43), tic disorder (N= 50), disorder of infancy, childhood or adolescence NOS (N= 65), depression (N= 64), motor, learning and communication disorders (N= 59), anorexia nervosa (N= 48), somatoform disorders (N= 47), trauma/stress (N= 39) and controls (N= 1,448, age 17–84) of European ancestry. First, these 14 PRS were tested in univariate regression analyses. The traits that significantly predicted case-control status were included in a multivariable regression model to investigate the gain in explained variance when leveraging the genetic effects of multiple traits simultaneously.
Results: in the univariate analyses, we observed significant associations between clinical status and the PRS of educational attainment (EA), smoking initiation (SI), intelligence, neuroticism, alcohol dependence, ADHD, major depression and anti-social behavior. EA (p-value: 3.53E-20, explained variance: 3.99%, OR: 0.66), and SI (p-value: 4.77E-10, explained variance: 1.91%, OR: 1.33) were the most predictive traits. In the multivariable analysis with these eight significant traits, EA and SI, remained significant predictors. The explained variance of the PRS in the model with these eight traits combined was 5.9%.
Conclusion: our study provides more insights into the genetic signal that is shared between childhood and adolescent psychiatric disorders. As such, our findings might guide future studies on psychiatric comorbidity and offer insights into shared etiology between psychiatric disorders. The increase in explained variance when leveraging the genetic signal of different predictor traits supports a multivariable approach to optimize precision accuracy for general psychopathology.

Introduction

In this study, we aim to quantify the predictive capacity of common genetic variation of a variety of traits to reveal which traits contribute to the shared genetic risk across disorders as it is well known that psychiatric disorders are highly comorbid. High comorbidity rates have for instance been shown between anxiety disorders (anxiety), major depressive disorder (MDD), attention-deficit/ hyperactivity disorder (ADHD), autism spectrum disorder (ASD), schizophrenia, alcohol dependence and eating disorders (Katzman, Bilkey, Chokka, Fallu, & Klassen, 2017; Klimkiewicz, Klimkiewicz, Jakubczyk, Kieres-Salomo_nski, & Wojnar, 2015; Ulfvebrand, Birgeg_ard, Norring, H€ogdahl, & von Hausswolff-Juhlin, 2015). Next to the comorbidity there is also extensive symptom overlap (American Psychiatric Association, 2013). This overlap has been described for MDD and anxiety (Tiller, 2013), ADHD, ASD, tic disorders and obsessive compulsive disorder (OCD; Huisman-van Dijk, van de Schoot, Rijkeboer, Mathews, & Cath, 2016). Interestingly, the occurrence of psychiatric disorders is also correlated with psychological traits in the general population such as lower educational attainment (EA, Lee et al., 2018), lower intelligence (Savage et al., 2018), higher substance use among which earlier smoking initiation (Liu et al., 2019), higher neuroticism scores (Nagel et al., 2018), and insomnia (Jansen et al., 2019).

EA comprises cognitive abilities (intelligence), noncognitive abilities (patience, self-control, temperament, motivation, self-discipline, time preference), health endowments, and family background (Conti, Heckman, & Urzua, 2010). There is a phenotypic link between EA and health (Lynch & Hippel, 2016) as shown by previous research involving EA and adult success on the labor market and adult health including psychopathology, with a focus on depression, which demonstrates an important role for both cognitive abilities in early life and noncognitive abilities (Conti et al., 2010). Intelligence by itself also plays a major role in health and wellbeing with higher intelligence being associated with lower risk of mental health problems (Savage et al., 2018). Focusing on substance use behaviors, the literature shows that smoking behavior is related to a host of psychiatric disorders among which schizophrenia, ADHD, eating disorders, mood disorders, anxiety and substance use disorders (Boksa, 2017), and in a US population patients with a psychiatric diagnosis have a 3.23 times greater odds of smoking compared to individuals with no diagnosis (Smith, Mazure, & McKee, 2014). High scores on neuroticism questionnaires are associated with psychiatric disorders (Hettema, Neale, Myers, Prescott, & Kendler, 2006; Nagel et al., 2018), and insomnia is one of the most common comorbidities of psychiatric disorders (Jansen et al., 2019).

Next to phenotypic overlap, extensive genetic overlap between psychiatric and psychological traits has been observed. These traits are at least moderately heritable (Polderman et al., 2015) with an underlying genetic architecture of rare and common genetic variation (Claussnitzer et al., 2020). A common genetic overlap has been shown extensively in the brain disorder (Bulik-Sullivan et al., 2015; CrossDisorder Group of the Psychiatric Genomics Consortium, 2019), the psychiatric disorder (Demontis et al., 2019; Grove et al., 2019; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Walters

et al., 2018; Wray et al., 2018) and psychological trait literature (Jansen et al., 2019; Linner et al., 2019; Nagel et al., 2018; Tielbeek et al., 2017; Walters et al., 2018). Health-related traits as EA (Lee et al., 2018; Satterstrom et al., 2020), intelligence (Savage et al., 2018), smoking initiation (Liu et al., 2019), insomnia (Jansen et al., 2019), risk-taking behavior (RTB; Linner et al., 2019) and anti-social behavior (Tielbeek et al., 2017) show genetic correlations with psychiatric disorders and with each other. These studies show that psychiatric disorders, psychological traits and closely related phenotypes show genotypic overlap that might be due to pleiotropy (Watanabe et al., 2019; a locus affecting more than one trait) and polygenicity (Watanabe et al., 2019; multiple loci affecting one trait).

Building on the existing phenotypic and genetic overlap as summarized above, research on a theorized underlying general psychopathology factor, the 'p factor', tries to identify an underlying higher order dimension for psychopathology in general, and specific domains below this overarching p factor, such as internalizing, externalizing or psychotic experience domains (Caspi & Moffitt, 2018). This hierarchical clustering is based on the hypothesis that each mental disorder has a broadly shared and a unique genetic component. The shared genetic component is thought to capture the genetic part of the broad range of symptoms that are common across disorders, while the unique genetic component is thought to capture disorder specific symptoms (Caspi et al., 2014; Murray, Eisner, & Ribeaud, 2016). In addition, it is suggested that the p factor can combine all psychiatric disorders on a low to high psychopathology severity scale. The hypothesis is that a person's score on this scale is informative of family history, developmental history, brain functioning and adult life impairment with higher p factor scores representing worse outcomes. (Caspi & Moffitt, 2018).

The findings regarding the p factor, genetic overlap, pleiotropy and polygenicity in psychopathology provide support for studies exploring shared genetic variation of nonspecific, shared psychiatric problems as present in clinical psychiatric samples. The shared heritability between traits and disorders (Brainstorm Consortium et al., 2018; Bulik-Sullivan et al., 2015) can be examined by means of polygenic risk scores (PRS; Chatterjee, Shi, & Garcıa-Closas, 2016; Wray et al., 2014). A PRS is an individual's weighted sum of risk alleles for a trait based on previously determined effects of those alleles for that trait (Euesden, Lewis, & O'Reilly, 2015). At group level, the PRS has the potential to distinguish cases from controls. For example, the ADHD PRS has been shown to distinguish cases from controls in an ADHD and in an Autism Spectrum Disorder (ASD) /ADHD combined sample (Jansen et al., 2019), and the schizophrenia PRS differentiated patients who developed schizophrenia from patients who did not in a first episode psychosis sample (Vassos et al., 2017). Despite this capacity to distinguish cases from controls at a group level, the explained variance of the PRS is limited, often below 5% (Jansen et al., 2019). The predictive capacity of the PRS can be improved by making predictions based on multiple traits and disorders that share genetic influences (Brainstorm Consortium et al., 2018; Bulik-Sullivan et al., 2015), by using multivariate approaches (Abdellaoui et al., 2018), or creating a multi-trait predictor (Krapohl et al., 2018; Maier et al., 2018). These methods seem promising as, for example, a multi-polygenic score (Krapohl et al., 2018) explained 4.8% of the variance in general cognitive ability and 10.9% in

educational achievement in an adolescent sample, capturing 1.1% more variance than the best single-score predictors.

The aim of the present study is to quantify the predictive capacity of common genetic variation of a variety of traits, as captured by their PRS, to predict case-control status in a child and adolescent psychiatric sample with a variety of psychiatric disorders to reveal which traits contribute to the shared genetic risk across disorders. Disorders present in the sample, and closely related traits were used as predictive traits. Both child or adolescent and adult mental traits have been included as we expect genetic overlap, since the majority of the adult psychiatric disorders usually have their onset during childhood or adolescence (Kessler et al., 2007) and the genetic make-up of an individual is fixed during life. To add, to our knowledge no genetic studies into addiction in child and adolescent samples have been presented. Firstly, we examined which individual PRS of these phenotypes significantly predicted receiving a diagnosis. Secondly, we evaluated whether this prediction could be improved by combining the joint genetic signals of the significantly associated phenotypes. The findings of this study will contribute to the identification of a shared genetic signal across disorders.

Methods

Participants

Clinical sample: 'Inside-out'. Data for this clinical sample ('Inside-out') were collected from January 2001 until January 2012 at the department of Child and Adolescent Psychiatry of the Sophia Children's Hospital, Erasmus Medical Center in Rotterdam. The diagnostic classification was performed by a clinician according to the Diagnostic and Statistical Manual of Mental Disorders, fourth edition. This procedure consisted of an interview with parents, a semi structured interview with the child based on the Semi-structured Clinical Interview for Children and Adolescents (McConaughy & Achenbach, 2001), the Diagnostic Interview Schedule for Children IV-P (Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000), and the Autism Diagnostic Observation Schedule-Generic (Lord et al., 1989) in case of a suspected Autism Spectrum Disorder. The above-mentioned procedure was part of standard clinical practice. Additionally, DNA was extracted from saliva and genotyping was performed on the Illumina Psych Chip array and Global Screening Array (see Data section). For this study, ethical approval of the Erasmus Medical center was obtained. The full sample (N = 1909) consisted of children that received a clinical diagnosis (N = 1594), and a group of children that did not receive a diagnosis (N = 315). The current study used data of the 1,402 children (192 cases were removed after genetic quality control) diagnosed with one or more DSM-IV disorders (ASD, ADHD, tic disorder, OCD, MDD, anxiety, anorexia nervosa (AN), eating disorder NOS, and subcategories of mentioned disorders). Intellectual disability was present in 16% of the sample.

Control sample

A Dutch population sample was used (NESCOG, N = 943, age range: 17.0–79.0), previously described by Polderman et al. (2013). Data were collected on various behavioral symptoms, cognitive functioning, personality, environmental factors, and life events, in addition to genetic information. To correct for undiagnosed ASD, ADHD or anxiety status we excluded participants scoring 3 SD above the mean on the Autism Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), the attention problems scale of the Young Adult Self Report (YASR; Achenbach, 1997), the Conners' Adult ADHD Rating Scale (CAARS; Conners, Erhardt, & Sparrow, 1999) or the Beck Anxiety Inventory (BAI). Genotyping was performed on the Illumina Psych Chip. This resulted in a sample of 939 participants (age range 17–79, 38% male). In addition, we used a German sample, the Berlin Psychosis Study (BePS; Skarabis & Ripke, 2017) of healthy adult individuals (N = 509, age range 18–84, 31% male). Participants whom reported having received a bipolar disorder (BiP) or other psychotic disorder, ADHD, OCD, MDD, anxiety, AN, or alcohol dependence diagnosis were excluded (N = 31). The total control sample consisted of 1,448 individuals (age range 17–84, 35% males).

To provide a sense of the nature of comorbidities and diagnoses we have the following Tables/Figures: Table 1 shows sample specifics such as sample size, age range and genotyping array for the cases and controls. Figure 1 shows the diagnostic composition as the amount of cases in the full sample per disorder and the amount of cases per disorder in that part of the sample diagnosed with intellectual disability, as well as the percentage of the full sample with an intellectual disability diagnosis, intellectual disability status and the sex distribution. Tables S1 and S2 provide an overview of comorbidities.

Genotyping

Genotyping of part of the clinical sample (ADHD, ASD, tic disorder and AN diagnosis) and of the NESCOG control sample was performed on the Illumina PsychChip array. The PsychChip SNP array contains HumanCore, Human Exome and custom content to capture genetic variants previously linked with psychiatric disorders (https://www.illumina.com/products/by-type/micproarray-kits/infinium-psycharray.html).

The remaining part of the clinical sample and the BePS controls were genotyped on the Illumina Infinium Global Screening Array (GSA; https://www.illumina.com/products/ by-type/microarray-kits/infinium-global-screening.html).

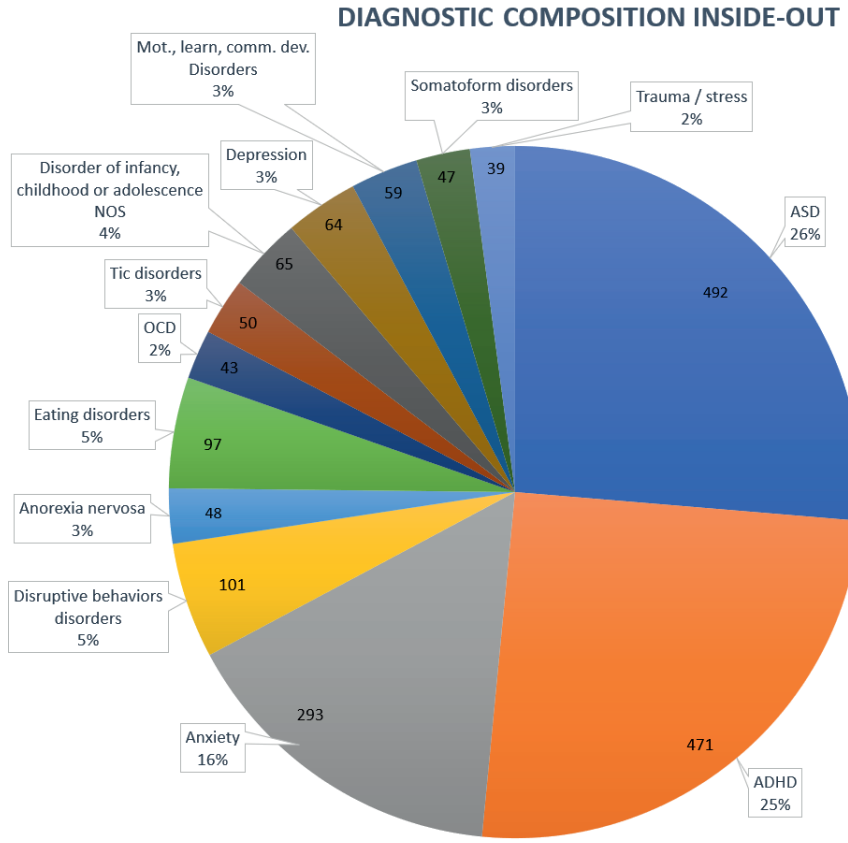
For SNP harmonization purpose between arrays, all samples were imputed in the Michigan imputation server. After imputation, the samples were combined. We used the Michigan imputation server pipeline which uses the Haplotype Reference Consortium (McCarthy et al., 2016; HRC) as a reference panel and poorly imputed variants were excluded based on their imputation score ($R^2 < 0.9$). In all samples, SNPs were filtered on MAF (<1%), SNP call rate (<95%) and Hardy–Weinberg disequilibrium ($p < .00001$). In the control samples, individual quality control filtering was based on missingness (>5%), relatedness (pairwise IBD > 0.185), ancestry (within the range of 1,000 Genomes CEU population on the first two principal

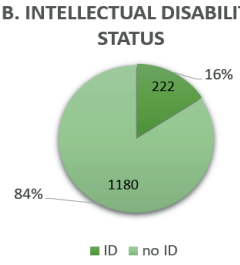
components (PCs)), outlying heterozygosity (excluded if > 3 x SD from the mean of the heterozygosity rate), gender mismatch and missing phenotypes. In the clinical sample, individuals were filtered based on genotype and sex mismatch, outlying heterozygosity and non-European ancestry (4 SD outside the range of the first two genetic principal components of the HapMap3 European founder population (CEU)), missingness (>5%) and relatedness (pairwise Identity-By-Descent (IBD) >0.185).

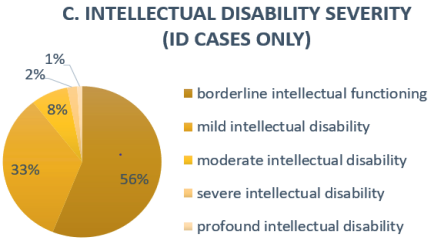
Table 1 Sample description

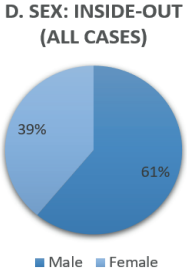
	Cases	Controls		
	Clinical sample	NESCOG	BePS	Total controls
Sample size	1,402	939	509	1,448
Age range (mean, SD), years	1–20 (9.54, 3.71)	17–79 (40.7, 17.3)	18–84 (30.2, 12.1)	17–84 (37.0)
Gender % male	61	38	31	35
Genotyping array	GSA (32%), Psych chip (68%)	Psych chip	GSA	GSA (35%), Psych chip (65%)

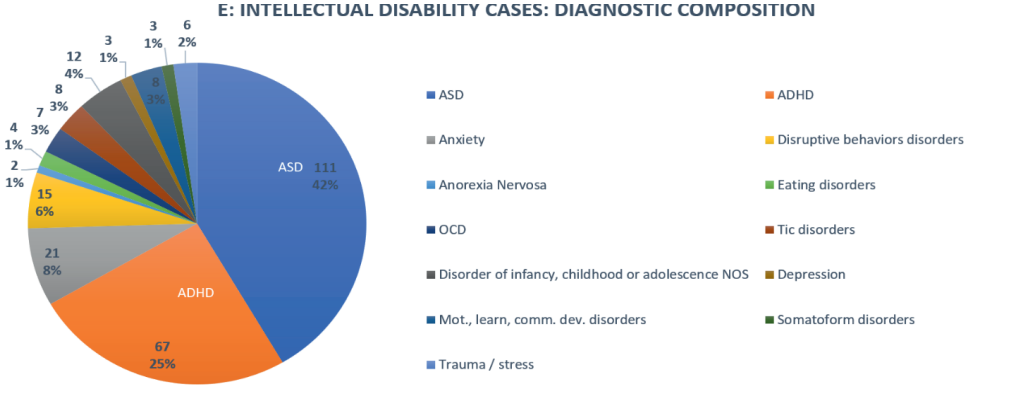
Figure 1, see next page: For readability shown on the following page, Diagnostic composition Inside-out. Abbreviations: ASD = autism spectrum disorder; ADHD = attention-deficit/hyperactivity disorder; OCD = obsessive compulsive disorder, disorder of infancy, childhood full diagnostic term: disorder of infancy, childhood or adolescence NOS, mot., learn., comm. dev. disorder comprises motor, learning and communication developmental disorders, ID: intellectual disability. A: Numbers per disorder are based on a total 1,402 cases. Comorbid disorders are included therefore totaling more than 1,402 diagnoses. B: Intellectual Disability (ID) status for all 1,402 cases. C: ID severity for all ID cases. ID severity known for 89% of all ID cases. D: sex distribution in Inside-out. E: the chart is based on all ID cases (N 222). Comorbid disorders are included, therefore totaling more than 222 diagnoses.


DIAGNOSTIC COMPOSITION INSIDE-OUT
Mot., learn, comm. dev. Disorders 3%
Somatoform disorders 3%
Trauma / stress 2%
Depression 3%
Disorder of infancy, childhood or adolescence NOS 4%
Tic disorders 3%
OCD 2%
Eating disorders 5%
Anorexia nervosa 3%
Disruptive behaviors disorders 5%
Anxiety 16%
ASD 26%
ADHD 25%
492
471
293
101
48
97
43
50
65
64
59
47
39


B. INTELLECTUAL DISABILITY STATUS
222
16%
1180
84%
ID
no ID


C. INTELLECTUAL DISABILITY SEVERITY (ID CASES ONLY)
1%
2%
8%
33%
56%
borderline intellectual functioning
mild intellectual disability
moderate intellectual disability
severe intellectual disability
profound intellectual disability


D. SEX: INSIDE-OUT (ALL CASES)
39%
61%
Male
Female


E: INTELLECTUAL DISABILITY CASES: DIAGNOSTIC COMPOSITION
ASD 111 42%
ADHD
67 25%
21 8%
15 6%
2 1%
4 1%
7 3%
8 3%
12 4%
3 1%
3 1%
6 2%
ASD
ADHD
Anxiety
Disruptive behaviors disorders
Anorexia Nervosa
Eating disorders
OCD
Tic disorders
Disorder of infancy, childhood or adolescence NOS
Depression
Mot., learn, comm. dev. disorders
Somatoform disorders
Trauma / stress

Polygenic scoring

The PRS is the sum of an individual's 'risk' alleles for a certain phenotype weighted by the allele effect sizes, which are typically derived from linear association coefficients from a genome wide association study (GWAS). For the PRS creation a SNP p-value inclusion threshold of < 1 is used (Choi, Mak, & O'Reilly, 2018; Maier et al., 2018). Using large publicly available summary statistics from GWA studies, PRS were constructed for EA (Lee et al., 2018), intelligence (Savage et al., 2018), smoking initiation (Liu et al., 2019), neuroticism (Nagel et al., 2018), insomnia (Jansen et al., 2019), RTB (Linner et al., 2019), anti-social behavior (Tielbeek et al., 2017), ADHD (Demontis et al., 2019), ASD (Grove et al., 2019), schizophrenia (Pardinas et al., 2018), MDD (Wray et al., 2018), anxiety (Otowa et al., 2016), alcohol dependence (Walters et al., 2018) and BiP (Stahl et al., 2018) see Appendix S1. Table S3 provides an overview and details of the selected GWA studies. The selected GWAS studies are large enough to use for this type of analyses as shown by their LD intercept which show no worrisome potential inflation (see Table S4). Inside-out, NESCOG and the BePS samples are independent samples not included in any of the GWAS.

The polygenic scoring was performed using PRSice2 (Euesden et al., 2015). Prior to polygenic scoring SNPs in high LD were clumped using PRSice2 (LD $R^2 < 0.1$, 250 kb pair window). For interpretational purposes the results were standardized to mean 0 and SD 1.

Statistical analysis

Genetic correlation with LDSC regression.

Using linkage disequilibrium score (LDSC) regression (Bulik-Sullivan et al., 2015), we calculated genetic correlations across all included traits based on the GWAS summary statistics we used for the PRS calculations of our predictor phenotypes. In addition, we computed the genetic correlations of the PRS in SPSS in our clinical sample to compare to the LDSC results.

Regression analyses.

First, we performed 14 univariate analyses to investigate which PRS were able to distinguish between cases and controls (outcome variable). Although all participants of the discovery and target sample were of European descent, the baseline model included eight PCs to account for potential population stratification The baseline model included, in addition to the eight PCs to account for population stratification, chip (GSA or Psych Chip, to correct for array effects), and sex as covariates. Age was not added as a covariate as all cases are children and all controls are adults. The PCs were calculated on all samples together and were based on the pruned data with Eigensoft (Price et al., 2006; version 3.0) software. After Bonferroni multiple testing correction for 14 tests we assessed the significance (p-value) of each predictor phenotype as well as its explained variance. The explained variance of the PRS is

based on Nagelkerke pseudo R^2 (i.e., the difference between the full model R^2 and the covariate only (baseline) model R^2). As 16% of Inside-out cases are co-diagnosed with intellectual disability (ID) we investigated if the results were driven by the ID subgroup by comparing the diagnostic distribution of the whole sample to the ID part of the sample. Additionally, we ran the univariate analysis for the intelligence and EA PRS on the part of Inside out without ID (N: 1,180, see Figure 1) to adjust the analysis if needed. The same covariates were included and the results were assessed the same way as the full sample results. Second, the significantly associated PRS from the univariate analyses were tested for their significance in a multivariable analysis.

Results

Genetic correlations between the predictor phenotypes

The genetic correlations as calculated from the summary statistics (Figure 2, Table S5) were overall in line with the available literature. As shown in the correlational matrix in Table S5, all included traits showed intermediate to high correlations with at least two other traits. Therefore all 14 PRS of predictive phenotypes were included in the subsequent analyses. The Pearson correlations between the prs in our clinical sample (Table S6) show some differences with the genetic correlations between the GWAS summary statistic of the phenotypes. These differences can be partially explained by a difference in sample size. The clinical sample is much smaller than the GWAS sample sizes resulting in a less precise estimate. In general, the significances are quite similar giving no reason for concern.

Regression analyses

The univariate logistic regression analyses showed eight significantly associated PRS (P Bonferroni corrected $_{(bf)} < 0.05$); EA, intelligence, smoking initiation, neuroticism, anti-social behavior, ADHD, MDD and alcohol dependence (Table 2a and Figure 3a). Presented p-values are Bonferroni corrected. Fit statistics are provided in Table S7.

When comparing the whole clinical sample to the ID cases in our sample, the ID cases sub-sample showed a larger proportion of ASD cases (42% vs 26%) and fewer anxiety cases (8% vs. 16%), and eating disorders (2% vs. 8%). Proportions of the other diagnostic groups were very similar between the full sample and the ID cases. The additional regression analyses in the sample without ID cases gave similar results for the EA and intelligence PRS (see Table S8).

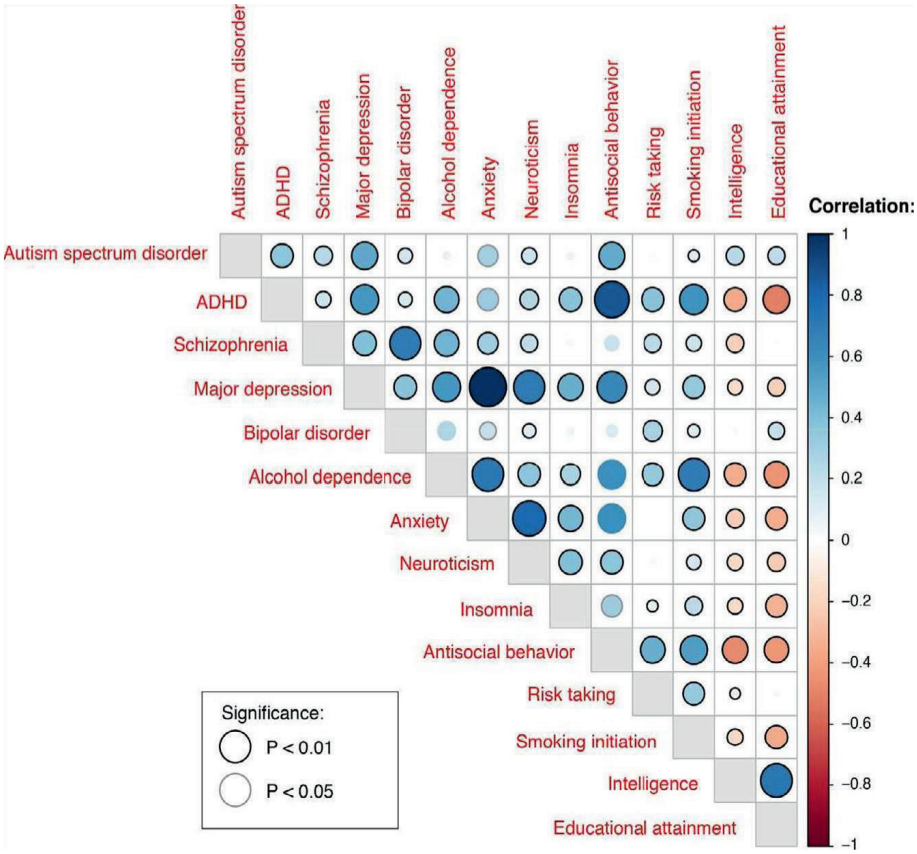
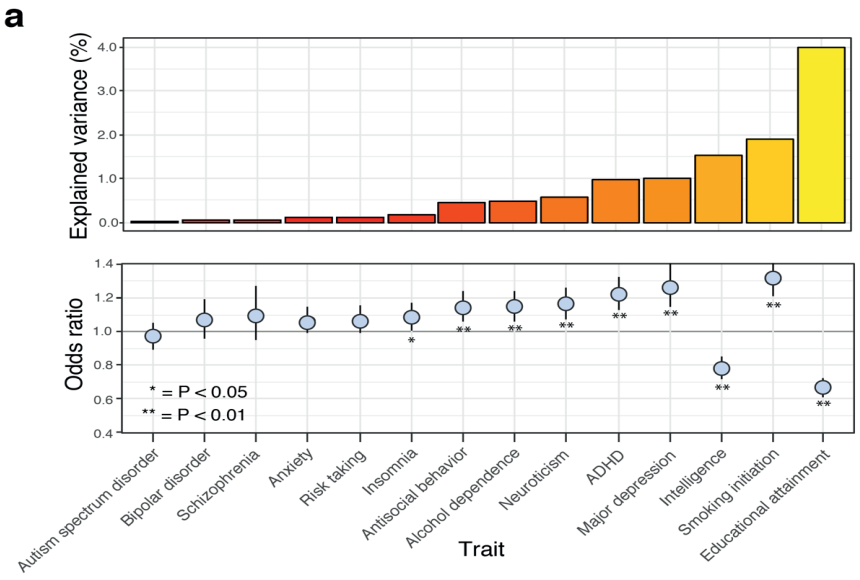


Figure 2 Overview of the genetic correlations based on the GWAS summary statistics

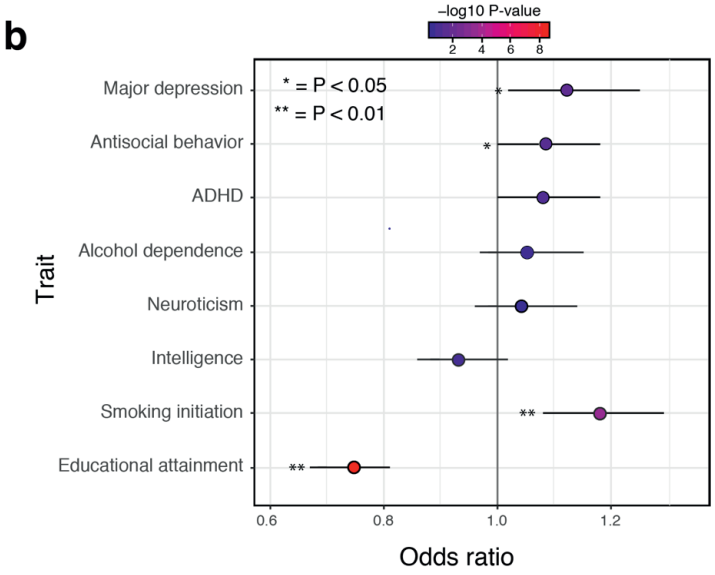
In a second instance, we retained only PRS with P_bf < 0.05 (EA, intelligence, smoking initiation, neuroticism, anti-social behavior, ADHD, MDD and alcohol dependence) for inclusion in a multivariable model. Of the eight PRS included in the full multivariable model (EA, intelligence, smoking initiation, neuroticism, anti-social behavior, ADHD, MDD and alcohol dependence) two remained statistically significant (EA, SI; see Table 2b and Figure 3b). The full model has an explained variance of 17.8%. Of this, the PRS account for 5.9% of the variance, which is an increase of 1.91% compared to the explained variance of the highest scoring univariate PRS (EA) of 3.99%. The remaining part of the explained variance can be attributed to the covariates. Details are shown in Table S9.

Another large diagnostic group in our sample is the anxiety group. Surprisingly, this PRS did not distinguish between cases and controls. As the anxiety PRS is based on a GWAS with a small sample (7,016 cases, 14,745 controls) it is likely to be underpowered.

Figure 3



The explained variance for all tested traits in the univariate analyses. The explained Variance is based on Nagelkerke R^2. The effect sizes are shown as Odds Ratios (ORs).



The ORs based on the multivariable analyses

Table 2 (a) Univariate logistic regression analysis. (b) Multivariable logistic regression analysis

(a) PRS	**% explained variance PRS**	**p**	**p_bf[a]**		**OR**	**95% CI for OR[c]**
EA	3.99	2.52E-21	3.53E-20		0.66	0.61–0.72
SI	1.91	3.41E-11	4.77E-10		1.33	1.22–1.44
IQ	1.53	3.22E-09	4.51E-08		0.78	0.72–0.85
MDD	1.02	1.00E-06	1.40E-05		1.27	1.15–1.40
ADHD	0.99	2.00E-06	2.80E-05		1.22	1.13–1.32
NEU	0.58	2.47E-04	3.46E-03		1.16	1.07–1.26
AD	0.50	6.27E-04	8.78E-03		1.15	1.06–1.24
ASB	0.47	9.84E-04	1.38E-02		1.14	1.06–1.24
INS	0.19	3.76E-02	5.27E-01		1.09	1.01–1.18
RTB	0.13	7.92E-02	1		1.07	0.99–1.16
ANX	0.11	1.16E-01	1		1.06	0.99–1.15
SCZ	0.07	1.98E-01	1		1.10	0.95–1.27
BiP	0.06	2.36E-01	1		1.07	0.96–1.19
ASD	0.03	4.30E-01	1		0.97	0.89–1.05
(b) PRS	**p**	**p_bf[b]**		**OR**		**95% CI for OR[c]**
EA	2.43E-09	1.94E-08		0.74		0.76–0.81
SI	2.43E-04	1.94E-03		1.18		1.08–1.29
MDD	1.96E-02	1.57E-01		1.13		1.02–1.25
ASB	3.92E-02	3.14E-01		1.09		1.00–1.18
ADHD	6.48E-02	5.19E-01		1.08		1.00–1.18
IQ	1.48E-01	1		0.93		0.86–1.02
AD	2.12E-01	1		1.05		0.97–1.15
NEU	3.18E-01	1		1.04		0.96–1.14

Clinical sample (N cases: 1,402, N controls: 1,448). Baseline model covariates: 8 PCs, sex and chip.

All included PRS have SNP p value threshold < 1.

AD = alcohol dependence; ADHD = attention-deficit/hyperactivity disorder; ANX = anxiety; ASB = anti-social behavior; ASD = autism spectrum disorder; BiP = bipolar disorder; EA = educational attainment; INS = insomnia; IQ = intelligence; MDD = major depressive disorder; NEU = neuroticism; RTB = risk-taking behavior; SCZ = schizophrenia; SI = smoking initiation

a After Bonferonni multiple testing (p-bf) correction for 14 tests.

b After Bonferonni multiple testing correction for 8 tests.

c Upper and lower limits are shown.

Discussion

Polygenic risk scores of EA and smoking initiation are the main predictors of case-control status in our clinical psychiatric child and adolescent sample. PRS of Intelligence, neuroticism, anti-social behavior, ADHD, MDD and alcohol dependence are the other predictor phenotypes that in univariate analyses significantly distinguished between cases and controls. The multivariable analysis, testing the joint genetic signal of multiple predictor PRS had a higher predictive capacity compared to single PRS analysis. The increase in explained variance highlights the usability of multiple PRS in joint models to optimize precision accuracy for general psychopathology.

Based on the first series of univariate analyses, we included eight significantly predicting traits in the multivariable analysis (EA, intelligence, smoking initiation, neuroticism, anti-social behavior, ADHD, MDD and alcohol dependence). Significant prediction came either from phenotypes that were based on a larger GWAS and hence had likely more statistical power (EA, intelligence, smoking initiation, neuroticism, MDD; Lee et al., 2018; Liu et al., 2019; Nagel et al., 2018; Savage et al., 2018; Wray et al., 2018) or were the more prevalent disorders in the sample (ADHD; Demontis & Walters, 2017), with a couple of exceptions (i.e., smaller GWAS and low prevalence in sample; anti-social behavior and alcohol dependence; Tielbeek et al., 2017; Walters et al., 2018). Four of these traits (EA, smoking initiation, antisocial behavior, MDD) remained significant predictors for case-control status in the multivariable analysis of which EA and smoking initiation survived multiple testing correction. Due to the high genetic correlations between respectively EA and intelligence (r .7), smoking initiation and alcohol dependence (r .8), neuroticism and MDD (r .7), and anti-social behavior and ADHD (r .9) the significance of one of both traits may be random, or based on only subtle differences between them. The significant positive association for the anti-social behavior PRS is a somewhat surprising finding as this GWAS is smaller and hence less powerful. We hope this result will be replicated in a future study. When comparing the results of the univariate and multivariable analysis we see that the explained variance increased from 3.99% (the highest explained variance result of the EA PRS as a single predictor) to 5.90%. The rise in explained variance shows that the addition of phenotypes is not only useful in studies with small, hence low powered GWAS summary statistics as shown before (Maier et al., 2018), but that it is also useful for more general defined phenotypes as diagnostic status. Yet, it is important to note that the general psychopathology construct might be less representative of some disorders, in particular ASD, given the reported PRS associations. We did not find a statistically significant association between six traits (insomnia, RTB, ASD, schizophrenia, anxiety, BIP) and case-control status in our sample. For insomnia, this is possibly due to the genetic correlation of insomnia being larger in mood disorders than in eating disorders and ASD (Jansen et al., 2019). Our sample composition might be too varied with not enough mood disorder cases included to generate a detectable enrichment of common genetic variation for INS. RTB was included in our model as it has a genetic correlation with ADHD and antisocial behavior and a phenotypic overlap between these three traits is present. However, the genetic correlation might be too weak to predict clinical status. In addition, our sample comprises not only groups

whom we expect to take more risk (ADHD) but also groups who are less likely to take risk (MDD/ anxiety) or with no relation to the risk phenotype (ASD, OCD/TIC disorders). As BiP and schizophrenia in general have a later age of onset compared to other psychiatric disorders (Abidi et al., 2017; Lijster et al., 2017; Patten, 2017) they are not present in our sample as this is a child and adolescent sample hence the common genetic variation of schizophrenia and BiP might not be enriched in Inside-out despite the link between ASD and schizophrenia (Zheng, Zheng, & Zou, 2018). The ASD PRS was expected to be associated as this is one of the larger diagnostic groups within the sample. However, this PRS has not been associated with the ASD subgroup in Inside-out in previous research, nor was the schizophrenia PRS (Jansen et al., 2019). Several explanations may explain this finding. First, ASD might have a unique genetic signature including an important contribution of rare variants (Satterstrom et al., 2020) that is not captured by the PRS, and second, the PRS might be inaccurate due to a smaller GWAS sample size. Lastly, given the high heterogeneity of ASD, the diagnostic composition within the ASD Inside-out sample may differ from the ASD cases included in the GWAS sample.

Another large diagnostic group in our sample is the anxiety group. Surprisingly, this PRS did not distinguish between cases and controls. As the anxiety PRS is based on a GWAS with a small sample (7,016 cases, 14,745 controls) it is likely to be underpowered.

Limitations

A weak point of the study is the referral bias present in samples generated in one institute. However, when comparing the diagnostic composition of Inside-out to a general psychiatric European sample of older children, we see a similar diagnostic composition (Gerritsen et. al. Milestone, oral communication). The main difference is the higher percentage of depression diagnoses in Milestone. As younger children are less likely to be diagnosed with depression and more likely to be diagnosed with anxiety, which later develops into depression, it seems likely these children will be included in different groups at younger ages, quite possibly in the ADHD, ASD and anxiety groups. With this in mind, in general, in both samples ADHD, ASD and anxiety show the highest prevalence, which suggests Inside-out is a good representation of the broader general child and adolescent psychiatric population. It would have been interesting to take educational level, cognitive performance and substance use into account. However, this information was not available for all samples used. In addition, this would remove part of the shared variance (Loe & Feldman, 2007), we did perform a sensitivity analysis excluding ID cases. A general concern regarding PRS studies in clinical samples is the limited clinical usability (Torkamani, Wineinger, & Topol, 2018). Our effect sizes are in line with the current literature (Jansen et al., 2019) meaning 5.9% still leaves room for many other contributing factors. Another point to keep in mind is the relatively small sample size of 1,402 cases and 1,448 controls adding up to 2,850 participants. However, with acceptable standard errors and 95% CI intervals for the OR in the regression analysis we feel this study adds value and can function as a pilot study leading into larger studies in this direction. Finally, It would

be informative to run additional sensitivity analyses to rule out that the results are being driven by the ASD/ADHD part of the sample. Due to power issues this is currently not an option. However, the neurodevelopmental part of the paper has been analyzed extensively (Jansen et al., 2019) and besides a significant association with the ADHD PRS, no significant associations with the schizophrenia and ASD PRS have been observed. Still, we cannot rule out that other predictors might be associated specifically in this sample due to the over representation of ADHD and ASD.

Strong points of the study are our carefully curated sample and the comparison between children as cases and adults as controls. As the controls are adults, the chance of them receiving an additional diagnosis of ADHD or ASD is small making them pure controls for these traits. In addition, the NESCOG control sample was corrected for undiagnosed ADHD, ASD and anxiety status. In the BePS sample, no psychiatric diagnoses were allowed.

Future Directions

Future directions in this area of research are replication in a larger independent child and adolescent sample, and preferably in an adult sample as well. Next to replication, research into causality is of great importance. We show association between diagnostic status and low EA and smoking initiation but are not able to address the issue of causality. Still, if low EA and smoking initiation are good predictors of psychiatric disorders, studies exploring early interventions targeting EA and smoking initiation to clarify their role in the development of psychiatric disorders can be useful. In addition, future studies will benefit from a longitudinal design to investigate how PRS correlate to later life outcomes. Based on this a risk profile of a group of individuals can be generated identifying individuals at risk whom might benefit from early interventions. Our PRS selection is a first step in identifying PRS suitable for this type of study.

To conclude, our findings suggest that a lot of the genetic variance influencing psychiatric disorders influence a myriad of mental health-related traits. Hence, a genetic vulnerability for low EA and SI are potential predictors for general psychopathology in children and adolescents which can be taken into account as some of the potential factors in the development of psychiatric symptoms. In addition, a genetic vulnerability for low EA and SI might contribute to specific comorbidity patterns as observed between psychiatric symptoms and to the broad range of psychiatric symptoms and as such might represent important contributors to the p factor. Our findings can guide future studies on psychiatric comorbidity, and the p factor, and studies addressing the causal directions between EA, SI and general psychopathology.

References

Abdellaoui, A., Nivard, M.G., Hottenga, J.J., Fedko, I., Verweij, K., Baselmans, B.M.L., ... & Cacioppo, J.T. (2018). Predicting loneliness with polygenic scores of social, psychological and psychiatric traits. Genes,Brain, andBehavior, 17, e12472.

Abidi, S., Mian, I., Garcia-Ortega, I., Lecomte, T., Raedler, T., Jackson, K. ... & Addington, D. (2017). Canadian guidelines for the pharmacological treatment of schizophrenia spectrum and other psychotic disorders in children and youth.

Canadian Journal of Psychiatry, 62, 635–647.

Achenbach, T.M. (1997). Manual for the young adult self-report and young adult behavior checklist. Burlington, VT: Department of Psychiatry, University of Vermont.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (Vol. 5). Arlington: American Psychiatric Association.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. Journal of Autism and Developmental Disorders, 31, 5–17.

Boksa, P. (2017). Smoking, psychiatric illness and the brain. Journal of Psychiatry and Neuroscience, 42, 147.

Brainstorm Consortium, Anttila, V., Bulik-Sullivan, B., Finucane, H.K., Walters, R.K., Bras, J., ... & Murray, R. (2018). Analysis of shared heritability in common disorders of the brain. Science, 360, eaap8758.

Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., ... & Neale, B.M. (2015). An atlas of genetic correlations across human diseases and traits. Nature Genetics, 47, 1236–1241

Caspi, A., Houts, R.M., Belsky, D.W., Goldman-Mellor, S.J., Harrington, H., Israel, S., ... & Moffitt, T.E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? Clinical Psychological Science : A Journal of the Association for Psychological Science, 2, 119– 137.

Caspi, A., & Moffitt, T.E. (2018). All for one and one for all: Mental disorders in one dimension. American Journal of Psychiatry, 175, 831–844.

Chatterjee, N., Shi, J., & Garcıa-Closas, M. (2016). Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nature Reviews Genetics, 17, 392–406.

Choi, S.W., Mak, T.S.H., & O'Reilly, P.F. (2018). A guide to performing Polygenic Risk Score analyses. bioRxiv, https:// doi.org/10.1101/416545

Claussnitzer, M., Cho, J.H., Collins, R., Cox, N.J., Dermitzakis, E.T., Hurles, M.E., ... & McCarthy, M.I. (2020). A brief history of human disease genetics. Nature, 577, 179–189.

Conners, C., Erhardt, D., & Sparrow, E. (1999). Conners' adult ADHD rating scales. North Tonawanda, NY: Multi-Health Systems.

Conti, G., Heckman, J., & Urzua, S. (2010). The educationhealth gradient. The American Economic Review, 100, 234– 238.

Cross-Disorder Group of the Psychiatric Genomics Consortium. (2019). Electronic address: plee0@mgh.harvard.edu & Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell, 179,

1469–1482.e11 de Lijster, J.M., Dierckx, B., Utens, E.M.W.J., Verhulst, F.C., Zieldorff, C., Dieleman, G.C., Legerstee, J.S. (2017). The age of onset of anxiety disorders. Canadian Journal of Psychiatry, 62, 237–246.

Demontis, D., & Walters, R.K. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nature Genetics, 51, 63–75.

Demontis, D., & Walters, R.K. (2017). Discovery of the first genome-wide significant risk loci for ADHD. bioRxiv,

https://doi.org/10.1101/145581

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

Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., ... & Børglum, A.D. (2019). Identification of common genetic risk variants for autism spectrum disorder. Nature Genetics, 51, 431–444.

Hettema, J.M., Neale, M.C., Myers, J.M., Prescott, C.A., & Kendler, K.S. (2006). A population-based twin study of the relationship between neuroticism and internalizing disorders. American Journal of Psychiatry, 163, 857–864.

Huisman-van Dijk, H.M., van de Schoot, R., Rijkeboer, M.M., Mathews, C.A., & Cath, D.C. (2016). The relationship between tics, OC, ADHD and autism symptoms: A crossdisorder symptom analysis in Gilles de la Tourette syndrome patients and family-members. Psychiatry Research, 237, 138–146.

Jansen, A.G., Dieleman, G.C., Jansen, P.R., Verhulst, F.C., Posthuma, D., & Polderman, T.J.C. (2019). Psychiatric polygenic risk scores as predictor for attention deficit/ hyperactivity disorder and autism spectrum disorder in a clinical child and adolescent sample. Behavior Genetics, 50, 203–212.

Jansen, P.R., Watanabe, K., Stringer, S., Skene, N., Bryois, J., Hammerschlag, A.R., ... & Posthuma, D. (2019). Genomewide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nature Genetics, 51, 394.

Katzman, M.A., Bilkey, T.S., Chokka, P.R., Fallu, A., & Klassen, L.J. (2017). Adult ADHD and comorbid disorders: Clinical implications of a dimensional approach. BMC Psychiatry, 17, 302.

Kessler, R.C., Amminger, G.P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Bedirhan Ustun, T. (2007). Age of onset of mental disorders: A review of recent literature. Current Opinion in Psychiatry, 20, 359–364.

Klimkiewicz, A., Klimkiewicz, J., Jakubczyk, A., Kieres-Salomonski, I., & Wojnar, M. (2015). Comorbidity of alcohol dependence with other psychiatric disorders. Part I. Epidemiology of dual diagnosis. Psychiatria Polska, 49, 265– 275.

Krapohl, E., Patel, H., Newhouse, S., Curtis, C.J., von Stumm, S., Dale, P.S., ... & Plomin, R. (2018). Multi-polygenic score approach to trait prediction. Molecular Psychiatry, 23, 1368–1374.

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature Genetics, 50, 1112–1121.

Linner, R.K., Biroli, P., Kong, E., Meddens, S.F.W., Wedow, R., Fontana, M.A., ... & Beauchamp, J.P. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nature Genetics, 51, 245.

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D.M., Chen, F., ... & Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nature Genetics, 51, 237.

Loe, I.M., & Feldman, H.M. (2007). Academic and educational outcomes of children with ADHD. Journal of Pediatric Psychology, 32, 643–654.

Lord, C., Rutter, M., Heemsbergen, J., Jordan, H., Mawhood, I., & Schopler, E. (1989). Autism diagnostic observation schedule: A standardized observation of communicative and social behavior. Journal of Autism and Developmental Disorders, 19, 185–212.

Lynch, J.L., & von Hippel, P.T. (2016). An education gradient in health, a health gradient in education, or a confounded gradient in both? Social Science and Medicine, 1982(154), 18–27.

Maier, R.M., Zhu, Z., Lee, S.H., Trzaskowski, M., Ruderfer, D.M., Stahl, E.A., ... & Robinson, M.R. (2018). Improving genetic prediction by leveraging genetic correlations among human diseases and traits. Nature Communications, 9, 989. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., ... & Marchini, J. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nature Genetics, 48, 1279–1283.

McConaughy, S.H., & Achenbach, T.M. (2001). Manual for the semi-structured clinical interview for children and adolescents. Burlington, VT: University of Vermont.

Murray, A.L., Eisner, M., & Ribeaud, D. (2016). The development of the general factor of psychopathology 'p Factor' through childhood and adolescence. Journal of Abnormal Child Psychology, 44, 1573–1586.

Nagel, M., Jansen, P.R., Stringer, S., Watanabe, K., de Leeuw, C.A., Bryois, J., ... & Posthuma, D. (2018). Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. Nature Genetics, 50, 920–927.

Otowa, T., Hek, K., Lee, M., Byrne, E.M., Mirza, S.S., Nivard, M.G., ... & Hettema, J.M. (2016). Meta-analysis of genomewide association studies of anxiety disorders. Molecular Psychiatry, 21, 1391–1399.

Pardinas, A.F., Holmans, P., Pocklington, A.J., Escott-Price,~ V., Ripke, S., Carrera, N., ... & Walters, J.T.R. (2018). Common schizophrenia alleles are enriched in mutationintolerant genes and in regions under strong background selection. Nature Genetics, 50, 381–389. Patten, S.B. (2017). Age of onset of mental disorders. Canadian Journal of Psychiatry, 62, 235–236.

Polderman, T.J.C., Hoekstra, R.A., Vinkhuyzen, A.A.E., Sullivan, P.F., van der Sluis, S., & Posthuma, D. (2013). Attentional switching forms a genetic link between attention problems and autistic traits in adults. Psychological Medicine, 43, 1985–1996.

Polderman, T.J.C., Benyamin, B., de Leeuw, C.A., Sullivan, P.F., van Bochoven, A., Visscher, P.M., ... & Posthuma, P. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nature Genetics, 47, 702–709.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38, 904–909.

Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.Y., ... & Buxbaum, J.D. (2020). Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell, 180, 568–584.e23.

Savage, J.E., Jansen, P.R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C.A., ... & Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nature Genetics, 50, 912–919.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature, 511, 421– 427.

Shaffer, D., Fisher, P., Lucas, C.P., Dulcan, M.K., & SchwabStone, M.E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. Journal of the American Academy of Child and Adolescent Psychiatry, 39, 28–38.

Skarabis, N., & Ripke, S. (2017). Beps – Berlin psychosis study. European Neuropsychopharmacology, 27, S466.

Smith, P.H., Mazure, C.M., & McKee, S.A. (2014). Smoking and mental illness in the US population. TobaccoControl, 23, e147.

Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., ... & Sklar, P. (2018). Genome-wide association study identifies 30 Loci Associated with Bipolar Disorder. bioRxiv, https://doi.org/10.1101/173062

Tielbeek, J.J., Johansson, A., Polderman, T.J.C., Rautiainen, M.R., Jansen, P., Taylor, M., ... & Posthuma, D. (2017). Genome-wide association studies of a broad spectrum of antisocial behavior. JAMA Psychiatry, 74, 1242–1250.

Tiller, J.W.G. (2013). Depression and anxiety. Medical Journal of Australia, 199, S28–S31.

Torkamani, A., Wineinger, N.E., & Topol, E.J. (2018). The personal and clinical utility of polygenic risk scores. Nature Reviews Genetics, 19, 581–590.

Ulfvebrand, S., Birgegard, A., Norring, C., Hogdahl, L., & von€ Hausswolff-Juhlin, Y. (2015). Psychiatric comorbidity in women and men with eating disorders results from a large clinical database. Psychiatry Research, 230, 294–299.

Vassos, E., Di Forti, M., Coleman, J., Iyegbe, C., Prata, D., Euesden, J., ... & Breen, G. (2017). An examination of polygenic score risk prediction in individuals with firstepisode psychosis. Biological Psychiatry, 81, 470–477.

Walters, R.K., Polimanti, R., Johnson, E.C., McClintick, J.N., Adams, M.J., Adkins, A.E., ... & Agrawal, A. (2018). Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Nature Neuroscience, 21, 1656–1669.

Watanabe, K., Stringer, S., Frei, O., Umicevi c Mirkov, M., de Leeuw, C., Polderman, T.J.C., ... & Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. Nature Genetics, 51, 1339–1348.

Wray, N.R., Lee, S.H., Mehta, D., Vinkhuyzen, A.A.E., Dudbridge, F., & Middeldorp, C.M. (2014). Research review: Polygenic methods and their application to psychiatric traits. Journal of Child Psychology and Psychiatry, 55, 1068–1087.

Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., & Abdellaoui, A., ... & the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nature Genetics, 50, 668–681.

Zheng, Z., Zheng, P., & Zou, X. (2018). Association between schizophrenia and autism spectrum disorder: A systematic review and meta-analysis. Autism Research: Official Journal of the International Society for Autism Research, 11, 1110– 1119.

Chapter 3 Supplementary Material

Table of Contents

Appendix S1. Selection of predictive polygenic risk scores (PRS)

The choice of predictive polygenic risk scores (PRS) was based on well-established genetic correlations between all disorders in our sample ASD, ADHD, MDD and anxiety (Demontis & Walters, 2017; Grove et al., 2019; Nagel et al., 2018; Wray et al., 2018) and genetic correlations between obsessive-compulsive disorder (OCD) and ASD have been shown by significant PRS analysis (Guo et al., 2017), in a Tic disorder sample genetic correlations between Tic disorder and mood, anxiety or disruptive behaviors may be accounted for by ADHD or OCD (Hirschtritt et al., 2015) and eating disorders are genetically correlated to each other (Anorexia Nervosa (AN), Bulimia nervosa (BN) and Binge eating disorder) (Bulik et al., 2019), to OCD (Cederlöf et al., 2015), alcohol dependence (Munn-Chernoff et al., 2015) and MDD (Wade et al., 2000).

The carefully selected predictive phenotypes are chosen due to their presence in Inside-out or their known genetic correlation with the traits present in Inside-out. If well-powered GWAS summary statistics were unavailable for a disorder included in Inside-out, a genetically closely related trait or disorder was chosen. A GWAS with a sample size under 10,000 cases was classified as underpowered (Cichon & Ripke, 2016) and excluded from this study (AN (Duncan et al., 2017), Tic disorder (Scharf et al., 2013), OCD (International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS), 2018)), with the exception of anxiety. As anxiety is one of the three large diagnostic groups (anxiety, ASD and ADHD) in Inside-out, we chose to include the GWAS (Otowa et al., 2016) with 7,016 cases and a total sample size of 17,310. To add more nuance to the genetic component of the anxiety phenotype, neuroticism (Nagel et al., 2018) and MDD (Wray et al., 2018), being genetically correlated to anxiety and being measures for internalizing behavior, were included to represent the genetic signal possibly missed by the anxiety PRS. Late onset psychiatric disorders such as schizophrenia, bipolar disorder (BiP) and alcohol dependence were considered for inclusion since the observed problems and diagnoses in the cases might currently be classified as anxiety but eventually lead to a schizophrenia or BiP diagnosis. For these cases the relevant late onset disorder PRS are valuable to include. schizophrenia, intelligence and EA are genotypically correlated to ASD (Grove et al., 2019) hence, they are included. Alcohol dependence (Walters et al., 2018). Risk taking behavior (RTB) (Linnér et al., 2019) and smoking initiation(Liu et al., 2019) are included as they can be seen as a measure for externalizing behavior which is closely related to ADHD (Ahmad & Hinshaw, 2017). Anti-social behavior (Tielbeek et al., 2017) is included due to its genetic correlation to ADHD and neuroticism and insomnia is genetically correlated to several psychiatric traits (Jansen et al., 2019).

Ahmad, S. I., & Hinshaw, S. P. (2017). Attention-Deficit/Hyperactivity Disorder, Trait Impulsivity, and Externalizing Behavior in a Longitudinal Sample. *Journal of Abnormal Child Psychology*, *45*(6), 1077–1089. https://doi.org/10.1007/s10802-016-0226-9

Bulik, C. M., Blake, L., & Austin, J. (2019). Genetics of Eating Disorders: What the Clinician Needs to Know. *The Psychiatric Clinics of North America*, *42*(1), 59–73. https://doi.org/10.1016/j.psc.2018.10.007

Cederlöf, M., Thornton, L. M., Baker, J., Lichtenstein, P., Larsson, H., Rück, C., Bulik, C. M., & Mataix-Cols, D. (2015). Etiological overlap between obsessive-compulsive disorder and anorexia nervosa: A longitudinal cohort, multigenerational family and twin study. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *14*(3), 333–338. https://doi.org/10.1002/wps.20251

Cichon, S., & Ripke, S. (2016). Chapter 3—Insights From Genome-Wide Association Studies (GWAS). In T. Abel & T. Nickl-Jockschat (Eds.), *The Neurobiology of Schizophrenia* (pp. 39–50). Academic Press. https://doi.org/10.1016/B978-0-12-801829-3.00011-2

Demontis, D., & Walters, R. K. (2017). *Discovery Of The First Genome-Wide Significant Risk Loci For ADHD*. https://doi.org/10.1101/145581

Duncan, L., Yilmaz, Z., Walters, R., Goldstein, J., Anttila, V., Bulik-Sullivan, B., Ripke, S., Thornton, L., Hinney, A., Daly, M., Sullivan, P., Zeggini, E., Breen, G., & Bulik, C. (2017). Genome-Wide Association Study Reveals First Locus for Anorexia Nervosa and Metabolic Correlations. *The American Journal of Psychiatry*, *174*(9), 850–858. https://doi.org/10.1176/appi.ajp.2017.16121402

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Awashti, S., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Christensen, J. H., … 23andMe Research Team. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, *51*(3), 431–444. https://doi.org/10.1038/s41588-019-0344-8

Guo, W., Samuels, J. F., Wang, Y., Cao, H., Ritter, M., Nestadt, P. S., Krasnow, J., Greenberg, B. D., Fyer, A. J., McCracken, J. T., Geller, D. A., Murphy, D. L., Knowles, J. A., Grados, M. A., Riddle, M. A., Rasmussen, S. A., McLaughlin, N. C., Nurmi, E. L., Askland, K. D., … Shugart, Y. Y. (2017). Polygenic risk score and heritability estimates reveals a genetic relationship between ASD and OCD. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, *27*(7), 657–666. https://doi.org/10.1016/j.euroneuro.2017.03.011

Hirschtritt, M. E., Lee, P. C., Pauls, D. L., Dion, Y., Grados, M. A., Illmann, C., King, R. A., Sandor, P., McMahon, W. M., Lyon, G. J., Cath, D. C., Kurlan, R., Robertson, M. M.,

Osiecki, L., Scharf, J. M., Mathews, C. A., & Tourette Syndrome Association International Consortium for Genetics. (2015). Lifetime prevalence, age of risk, and genetic relationships of comorbid psychiatric disorders in Tourette syndrome. *JAMA Psychiatry*, *72*(4), 325–333. https://doi.org/10.1001/jamapsychiatry.2014.2650

International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS). (2018). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Molecular Psychiatry*, *23*(5), 1181–1188. https://doi.org/10.1038/mp.2017.154

Jansen, P. R., Watanabe, K., Stringer, S., Skene, N., Bryois, J., Hammerschlag, A. R., Leeuw, C. A. de, Benjamins, J. S., Muñoz-Manchado, A. B., Nagel, M., Savage, J. E., Tiemeier, H., White, T., Tung, J. Y., Hinds, D. A., Vacic, V., Wang, X., Sullivan, P. F., Sluis, S. van der, … Posthuma, D. (2019). Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nature Genetics*, *51*(3), 394. https://doi.org/10.1038/s41588-018-0333-3

Linnér, R. K., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., Lebreton, M., Tino, S. P., Abdellaoui, A., Hammerschlag, A. R., Nivard, M. G., Okbay, A., Rietveld, C. A., Timshel, P. N., Trzaskowski, M., Vlaming, R. de, Zünd, C. L., Bao, Y., Buzdugan, L., … Beauchamp, J. P. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics*, *51*(2), 245. https://doi.org/10.1038/s41588-018-0309-3

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., Zhan, X., Choquet, H., Docherty, A. R., Faul, J. D., Foerster, J. R., Fritsche, L. G., Gabrielsen, M. E., Gordon, S. D., Haessler, J., … Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, *51*(2), 237. https://doi.org/10.1038/s41588-018-0307-5

Munn-Chernoff, M. A., Grant, J. D., Agrawal, A., Sartor, C. E., Werner, K. B., Bucholz, K. K., Madden, P. A. F., Heath, A. C., & Duncan, A. E. (2015). Genetic overlap between alcohol use disorder and bulimic behaviors in European American and African American women. *Drug and Alcohol Dependence*, *153*, 335–340. https://doi.org/10.1016/j.drugalcdep.2015.05.043

Nagel, M., Jansen, P. R., Stringer, S., Watanabe, K., de Leeuw, C. A., Bryois, J., Savage, J. E., Hammerschlag, A. R., Skene, N. G., Muñoz-Manchado, A. B., 23andMe Research Team, White, T., Tiemeier, H., Linnarsson, S., Hjerling-Leffler, J., Polderman, T. J. C., Sullivan, P. F., van der Sluis, S., & Posthuma, D. (2018). Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genetics*, *50*(7), 920–927. https://doi.org/10.1038/s41588-018-0151-7

Otowa, T., Hek, K., Lee, M., Byrne, E. M., Mirza, S. S., Nivard, M. G., Bigdeli, T., Aggen, S. H., Adkins, D., Wolen, A., Fanous, A., Keller, M. C., Castelao, E., Kutalik, Z., Van der Auwera, S., Homuth, G., Nauck, M., Teumer, A., Milaneschi, Y., … Hettema, J. M. (2016). Meta-analysis of genome-wide association studies of anxiety disorders. *Molecular Psychiatry*, *21*(10), 1391–1399. https://doi.org/10.1038/mp.2015.197

Scharf, J. M., Yu, D., Mathews, C. A., Neale, B. M., Stewart, S. E., Fagerness, J. A., Evans, P., Gamazon, E., Edlund, C. K., Service, S., Tikhomirov, A., Osiecki, L., Illmann, C., Pluzhnikov, A., Konkashbaev, A., Davis, L. K., Han, B., Crane, J., Moorjani, P., … Pauls, D. L. (2013). Genome-wide association study of Tourette Syndrome. *Molecular Psychiatry*, *18*(6), 721–728. https://doi.org/10.1038/mp.2012.69

Tielbeek, J. J., Johansson, A., Polderman, T. J. C., Rautiainen, M.-R., Jansen, P., Taylor, M., Tong, X., Lu, Q., Burt, A. S., Tiemeier, H., Viding, E., Plomin, R., Martin, N. G., Heath, A. C., Madden, P. A. F., Montgomery, G., Beaver, K. M., Waldman, I., Gelernter, J., … Broad Antisocial Behavior Consortium collaborators. (2017). Genome-Wide Association Studies of a Broad Spectrum of Antisocial Behavior. *JAMA Psychiatry*, *74*(12), 1242–1250. https://doi.org/10.1001/jamapsychiatry.2017.3069

Wade, T. D., Bulik, C. M., Neale, M., & Kendler, K. S. (2000). Anorexia nervosa and major depression: Shared genetic and environmental risk factors. *The American Journal of Psychiatry*, *157*(3), 469–471. https://doi.org/10.1176/appi.ajp.157.3.469

Walters, R. K., Polimanti, R., Johnson, E. C., McClintick, J. N., Adams, M. J., Adkins, A. E., Aliev, F., Bacanu, S.-A., Batzler, A., Bertelsen, S., Biernacka, J. M., Bigdeli, T. B., Chen, L.-S., Clarke, T.-K., Chou, Y.-L., Degenhardt, F., Docherty, A. R., Edwards, A. C., Fontanillas, P., … Agrawal, A. (2018). Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature Neuroscience*, *21*(12), 1656–1669. https://doi.org/10.1038/s41593-018-0275-1

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Adams, M. J., Agerbo, E., Air, T. M., Andlauer, T. M. F., Bacanu, S.-A., Bækvad-Hansen, M., Beekman, A. F. T., Bigdeli, T. B., Binder, E. B., Blackwood, D. R. H., Bryois, J., Buttenschøn, H. N., Bybjerg-Grauholm, J., … Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, *50*(5), 668–681. https://doi.org/10.1038/s41588-018-0090-3

For readability of all Tables the following abbreviations have been used: educational attainment (EA), smoking initiation (SI), intelligence (IQ), major depressive disorder (MDD), attention deficit/hyperactivity disorder (ADHD), neuroticism (NEU), alcohol dependence (AD), anti-social behavior (ASB), insomnia (INS), risk taking behavior (RTB), anxiety, (ANX), schizophrenia (SCZ), bipolar disorder (BiP), autism spectrum disorder (ASD), obsessive-compulsive disorder (OCD) and motor, learning or communication developmental disorder (Mot., learn, comm. dev. disorder).

Table S1

Overview diagnostic composition including codiagnoses

diagnosis (not main diagnosis)	comorbid co-diagnoses	N
ADHD (N=471)	ASD	137
	Disruptive behavior	58
	Anxiety	46
	Mot., learn, comm. dev. disorder	30
	OCD	5
	Tic disorder	19
	Depression	14
	Disorder of infancy, childhood or adolescence NOS	10
	Trauma / stress	6
	Somatic symptoms	3
	Eating disorder	2
	Anorexia Nervosa	0
Total		**330**

diagnosis (not main diagnosis)	co-diagnoses	N
ASD O(N=492)	ADHD	137
	Disruptive behavior	21
	Anxiety	19
	OCD	5
	Tic disorder	13
	Mot., learn, comm. dev. disorder	16
	Depression	12
	Disorder of infancy, childhood or adolescence NOS	8
	Eating disorder	6
	Trauma / stress	2
	Somatic symptoms	1
	Anorexia Nervosa	0
Total		**240**

diagnosis (not main diagnosis)	co-diagnoses	N
ANX (N=293)	ASD	19
	ADHD	46
	Disruptive behavior	8
	Eating disorder	9
	OCD	21
	Tic disorder	14
	Disorder of infancy, childhood or adolescence NOS	3
	Depression	24
	Mot., learn, comm. dev. disorder	11
	Anorexia Nervosa	1
	Somatic symptoms	3
	Trauma / stress	5
Total		**164**

diagnosis (not main diagnosis)	co-diagnoses	N
Disruptive behavior (N=101)	ADHD	58
	ASD	21
	ANX	8
	Depression	7
	OCD	2
	Tic disorder	3
	Disorder of infancy, childhood or adolescence NOS	2
	Mot., learn, comm. dev. disorder	1
	Trauma / stress	1
	Eating disorder	0
	Anorexia Nervosa	0
	Somatic symptoms	0
Total		**103**

diagnosis (not main diagnosis)	co-diagnoses	N
Eating disorder (N=97)	Anorexia Nervosa	48
	ANX	9
	ASD	6
	Depression	3
	Trauma / stress	3
	ADHD	2
	OCD	1
	Tic disorder	0
	Somatic symptoms	1
	Disruptive behavior	0
	Disorder of infancy, childhood or adolescence NOS	0
	Mot., learn, comm. dev. disorder	0
Total		**73**

diagnosis (not main diagnosis)	co-diagnoses	N
OCD (N=43)	ANX	21
	ADHD	5
	ASD	5
	Depression	4
	Disruptive behavior	2
	Eating disorder	1
	Mot., learn, comm. dev. disorder	0
	Trauma / stress	1
	Disorder of infancy, childhood or adolescence NOS	0
	Anorexia Nervosa	0
	Tic disorder	4
	Somatic symptoms	0
Total		**43**

diagnosis (not main diagnosis)	co-diagnoses	N
Tic disorder (N=50)	ANX	34
	ADHD	19
	ASD	13
	Depression	3
	Disruptive behavior	3
	Eating disorder	0
	Mot., learn, comm. dev. disorder	1
	Trauma / stress	0
	Disorder of infancy, childhood or adolescence NOS	0
	Anorexia Nervosa	0
	OCD	4
	Somatic symptoms	0
Total		**77**

diagnosis (not main diagnosis)	co-diagnoses	N
Disorder of infancy, childhood or adolescence NOS	ADHD	10
(N=65)	ASD	8
	ANX	3
	Disruptive behavior	2
	Mot., learn, comm. dev. disorder	2
	Eating disorder	0
	OCD	0
	Tic disorder	0
	Depression	0
	Anorexia Nervosa	0
	Somatic symptoms	0
	Trauma / stress	0
Total		**25**

diagnosis (not main diagnosis)	co-diagnoses	N
Depression (N=64)	ANX	24
	ADHD	14
	ASD	12
	Disruptive behavior	7
	OCD	4
	Tic disorder	3
	Mot., learn, comm. dev. disorder	6
	Eating disorder	3
	Somatic symptoms	3
	Anorexia Nervosa	2
	Trauma / stress	1
	Disorder of infancy, childhood or adolescence NOS	0
Total		**79**

diagnosis (not main diagnosis)	co-diagnoses	N
Mot., learn, comm. dev. Disorder (N=59)	ADHD	30
	ASD	16
	ANX	11
	Depression	6
	Somatic symptoms	3
	Disorder of infancy, childhood or adolescence NOS	2
	Disruptive behavior	1
	OCD	0
	Tic disorder	1
	Trauma / stress	1
	Eating disorder	0
	Anorexia Nervosa	0
Total		**71**

diagnosis (not main diagnosis)	co-diagnoses	N
Anorexia Nervosa (N=48)	Depression	2
	ANX	1
	Trauma / stress	1
	ASD	0
	ADHD	0
	Disruptive behavior	0
	Eating disorder	0
	OCD	0
	Tic disorder	0
	Disorder of infancy, childhood or adolescence NOS	0
	Mot., learn, comm. dev. disorder	0
	Somatic symptoms	0
Total		**4**

diagnosis (not main diagnosis)	co-diagnoses	N
Somatic symptoms (N=47)	ADHD	3
	ANX	3
	Depression	3
	Mot., learn, comm. dev. disorder	3
	ASD	1
	Eating disorder	1
	Trauma / stress	1
	Disruptive behavior	0
	OCD	0
	Tic disorder	0
	Disorder of infancy, childhood or adolescence NOS	0
	Anorexia Nervosa	0
Total		**15**

diagnosis (not main diagnosis)	co-diagnoses	N
Trauma / stress (N=39)	ADHD	6
	ANX	5
	Eating disorder	3
	ASD	2
	Disruptive behavior	1
	OCD	1
	Tic disorder	0
	Depression	1
	Mot., learn, comm. dev. disorder	1
	Anorexia Nervosa	1
	Somatic symptoms	1
	Disorder of infancy, childhood or adolescence NOS	0
Total		**22**

Table S2:
Most common comorbidities in the sample

Diagnosis	N	Co-diagnose	N co-diagnoses
ASD	492	ADHD	137
		Disruptive behavior	21
		Anxiety	19
ADHD	471	ASD	137
		Disruptive behavior	58
		Anxiety	46
Anxiety	293	ASD	19
		ADHD	46
		Disruptive behavior	8
Disruptive behavior	101	ADHD	58
		ASD	21
		ANX	8
Eating disorder	97	Anorexia Nervosa	48
		ANX	9
		ASD	6
OCD	43	ANX	14
		ASD	5
		ADHD	5
Tic disorder	50	ANX	34
		ADHD	24
		ASD	18
Disorder of infancy, childhood or adolescence NOS	65	ADHD	10
		ASD	8
		ANX	3
Depression	64	ANX	24
		ADHD	14
		ASD	12
Mot., learn, comm. dev. disorder	59	ADHD	30
		ASD	16
		ANX	11
Anorexia Nervosa	48	Depression	2
		ANX	1
		Trauma / stress	1
Somatoform disorders	47	ADHD	3
		ANX	3
		Depression	3
Trauma / stress	39	ADHD	6
		ANX	5
		Eating disorder	3

Table S3
GWAS discovery sample information

Phenotypes included	Author	Year	Journal	Sample size	Study type	Included samples
Educational attainment	Lee et. al.	2018	Nature Genetics	N =766,345	continuous trait	SS GAC
Intelligence	Savage et. al.	2018	Nature genetics	N: 269,867	continuous trait	UKB, COGENT, RS, GENR, STR, S4S, HiQ/HRS, TEDS, DTR-MADT, DTR-LSADT, IMAGEN, BLTS-Children, BLTS-Adolescents, GfG, STSA-SATSA + GENDER, STSA-HARMONY
Neuroticism	Nagel et. al.	2018	Nature genetics	N: 390,278	continuous trait	UKB, GPC
Insomnia	Jansen et. al.	2019	Nature genetics	N: 386,533	continuous trait	UKB
Risk tolerance	Karlsson Linnér et. al.	2019	Nature genetics	N: 466,571	continuous trait	UKB 23 and me (UKB, N: 431,126)
Antisocial behavior	Tielbeek et. al.	2017	JAMA psychiatry	N: 32,000	continuous trait	Broad ABC, Eagle

Table S3 continued

Phenotypes included	Author	Year	Journal	Sample size	Study type	Included samples
smoking initiation	Liu et. al.	2018	Nature genetics	N=1,232,091	case / control	23andMe, ALSPAC, ARIC, BEAGESS, BLS, iPSYCH, PGC
Attention Deficit /Hyperactivity Disorder	Demontis et. al.	2018	Nature genetics	Cases N: 20,183 Controls N: 35,191	case / control	iPSYCH, PGC
Autism Spectrum Disorder	Grove et. al.	2019	Nature genetics	Cases N: 18,381 Controls N: 27,969	case / control	iPSYCH, PGC (comprises: ACE, AGP, AGRE, MONBOS, SSC)
Schizophrenia	Pardinas et. al.	2018	Nature genetics	Cases N: 40,675 Controls N: 64,643	case / control	PGC, CLOZUK
Major Depressive Disorder	Wray et. al.	2018	Nature genetics	Cases N: 135,458 Controls N: 344,901	case / control	PGC29, deCODE, GenSCot, GERA,iPSYCH
Anxiety	Otowa et. al	2016	Mol psychiatry	Cases N: 7,016 Controls N: 14,745 sumstats 17,310	case / control	MGS Controls (43), PsyCoLaus (44), RS (45), SHIP (46) QIMR (47), TRAILS (48), NESDA (49) / NTR (50;51)
Alcohol dependence	Walters et. al.	2018	Nature neuroscience	Cases N: 14,904 Controls N: 37,944 sumstats 46,568	case / control	CATS, CHDS, COGA-cc, FSCD, GESGA, GEDI-GSMS, CADD, PAGES, COGEND-Nico, COGEND SAGE, Spit for Science, NIAAA, CITA, BLTS, GEDI-VTSABD, MCTFR, CEDAR, STR
Bipolar disorder	Stahl et. al.	2018	bioarchives	Cases N: 20,352 Controls N: 31,358 sumstats 46,582	case / control	PGC

Table S4
Inflation statistics GWAS studies

Phenotype	LD intercept
ADHD	1.0287 (0.01)
ASD	0.9992 (0.0079)
SCZ	1.0699 (0.0113)
MDD	0.9946 (0.0079)
NEU	1.0239 (0.0107)
IQ	1.0754 (0.0118)
ASB	0.9983 (0.0076)
INS	1.0133 (0.0086)
AD	1.0119 (0.0061)
ANX	1.0009 (0.0066)
EA	1.0279 (0.0134)
RTB	1.0101 (0.0086)
SI	0.9007 (0.0086)
BiP	1.0185 (0.0077)

LD intercept calculated in LDSR

Table S5
Genetic correlations (SE) based on LDSC regression analysis

	ASD	ADHD	SCZ	MDD	BiP	AD	ANX
ASD							
ADHD	0.36** (0.05)						
SCZ	0.24** (0.04)	0.17** (0.03)					
MDD	0.49** (0.04)	0.56** (0.04)	0.39** (0.03)				
BiP	0.15** (0.04)	0.14** (0.04)	0.69** (0.02)	0.37** (0.03)			
AD	0.06 (0.08)	0.44** (0.08)	0.44** (0.05)	0.56** (0.07)	0.26 (0.06)		
ANX	0.29* (0.14)	0.32* (0.13)	0.30** (0.08)	0.99** (0.18)	0.20* (0.09)	0.71** (0.18)	
NEU	0.17** (0.03)	0.25** (0.03)	0.21** (0.02)	0.70** (0.02)	0.13** (0.02)	0.36** (0.06)	0.79** (0.15)
INS	0.05 (0.04)	0.37** (0.03)	0.04 (0.02)	0.46** (0.03)	0.05 (0.03)	0.28** (0.06)	0.43** (0.11)
ASB	0.48** (0.18)	0.86** (0.24)	0.18 (0.10)	0.63** (0.20)	0.12 (0.11)	0.60 (0.32)	0.60 (0.38)
RTB	0.02 (0.05)	0.37** (0.03)	0.22** (0.02)	0.16** (0.03)	0.28** (0.02)	0.33** (0.06)	0.0042 (0.07)
SI	0.09** (0.03)	0.58** (0.02)	0.17** (0.02)	0.34** (0.03)	0.11** (0.02)	0.69** (0.06)	0.35** (0.09)
IQ	0.23** (0.03)	-0.36** (0.03)	-0.22** (0.02)	-0.16** (0.03)	-0.03 (0.02)	-0.34** (0.06)	-0.23** (0.08)
EA	0.21** (0.02)	-0.51** (0.02)	0.01 (0.02)	-0.22** (0.02)	0.19** (0.02)	-0.44** (0.05)	-0.34** (0.08)

	NEU	INS	ASB	RTB	SI	IQ	EA
NEU							
INS	0.39** (0.02)						
ASB	0.36** (0.11)	0.31* (0.14)					
RTB	-0.04 (0.02)	0.09** (0.03)	0.46** (0.15)				
SI	0.15** (0.02)	0.22** (0.02)	0.53** (0.17)	0.33**(0.02)			
IQ	-0.18** (0.02)	-0.17** (0.02)	-0.47** (0.14)	-0.08** (0.02)	-0.18** (0.02)		
EA	-0.23** (0.02)	-0.33** (0.02)	-0.43** (0.14)	0.03 (0.02)	-0.36** (0.02)	0.72** (0.01)	

Table S6
Correlational matrix of the PRS within Inside out

	ASD	ADHD	SCZ	MDD	BiP	AD	ANX	NEU	INS	ASB	RTB	SI	IQ	EA
ASD														
ADHD	0.40**													
SCZ	0.47**	0.30**												
MDD	0.42**	0.38**	0.61**											
BiP	0.39**	0.28**	0.77**	0.52**										
AD	0.06*	0.15**	0.15**	0.18**	0.12**									
ANX	0.08**	0.06*	0.11**	0.13**	0.09**	0.08**								
NEU	0.09**	0.11**	0.12**	0.19**	0.11**	0.02	0.10**							
INS	0.08**	0.10**	0.21**	0.21**	0.15**	0.06*	0.05*	0.21**						
ASB	0.09**	0.11**	0.21**	0.19**	0.19**	0.12**	0.04	0.03	0.08**					
RTB	0.09**	0.10**	0.19**	0.14**	0.17**	0.05*	0.02	-0.02	0.10**	0.11**				
SI	0.19**	0.24**	0.36**	0.29**	0.31**	0.18**	0.07*	0.12**	0.17**	0.15**	0.22**			
IQ	0.09**	-0.11**	-0.07**	-0.08**	-0.04	-0.13**	-0.81**	-0.12**	-0.05	-0.06*	-0.02	-0.13**		
EA	-0.05	-0.26**	-0.23**	-0.22**	-0.23**	-0.11**	-0.07**	-0.17**	-0.18**	-0.11**	-0.020	-0.30**	0.41**	

** sig. 0.01 level
* sig. 0.05 level

Table S7
Univariate logistic regression analysis

Clinical sample (N cases: 1402, N controls: 1448)

PRS	% explained variance PRS*	BETA	SE	Wald	P Bonferonni corrected**	OR	95% CI for OR***	P	Included in multi-variate analysis	Nagelkerke R^2 full model	Nagelkerke R^2 baseline model	Nagelkerke R^2 PRS
EA	3.99	-0.41	0.044	89.89	3.53E-20	0.66	0.61 - 0.72	2.52E-21	yes	0.159	0.119	0.040
SI	1.91	-0.28	0.043	43.93	4.77E-10	0.75	0.69 - 0.82	3.41E-11	yes	0.138	0.119	0.019
IQ	1.53	-0.25	0.041	35.04	4.51E-08	0.78	0.72 - 0.85	3.22E-09	yes	0.134	0.119	0.015
MDD	1.02	0.24	0.049	23.68	1.40E-05	1.27	1.15 - 1.40	1.00E-06	yes	0.129	0.119	0.010
ADHD	0.99	0.20	0.041	22.98	2.80E-05	1.22	1.13 - 1.32	2.00E-06	yes	0.129	0.119	0.010
NEU	0.58	0.15	0.041	13.44	3.46E-03	1.16	1.07 - 1.26	2.47E-04	yes	0.125	0.119	0.006
AD	0.50	0.14	0.040	11.69	8.78E-03	1.15	1.06 - 1.24	6.27E-04	yes	0.124	0.119	0.005
ASB	0.47	0.13	0.040	10.86	1.38E-02	1.14	1.06 - 1.24	9.84E-04	yes	0.124	0.119	0.005
INS	0.19	0.08	0.040	4.32	5.27E-01	1.09	1.01 - 1.18	3.76E-02	no	0.121	0.119	0.002
RTB	0.13	0.07	0.040	3.08	1	1.07	0.99 - 1.16	7.92E-02	no	0.120	0.119	0.001
ANX	0.11	0.06	0.040	2.47	1	1.06	0.99 - 1.15	1.16E-01	no	0.120	0.119	0.001
SCZ	0.07	0.09	0.073	1.66	1	1.10	0.95 - 1.27	1.98E-01	no	0.120	0.119	0.001
BiP	0.06	0.07	0.056	1.41	1	1.07	0.96 - 1.19	2.36E-01	no	0.120	0.119	0.001
ASD	0.03	-0.03	0.044	0.62	1	0.97	0.89 - 1.05	4.30E-01	no	0.119	0.119	0.000

* Based on Nagelkerke R^2 of the PRS

**Bonferonni correction for 14 tests applied.

Nagelkerke R^2 baseline model: 0.119

*** upper and lower limits are shown

Table S8
Univariate logistic regression analysis with ID cases excluded from the sample

Clinical sample: intellectual disability cases excluded (N cases: 1180, N controls: 1448) (excluded 222)

PRS	% explained variance* PRS	BETA	SE	Wald	P	OR	95% CI for OR **	Include in multi-variate analysis	P Bonferonni corrected 14 tests	Nagelkerke R^2 full model	Nagelkerke R^2 baseline model	Nagelkerke R^2 PRS
EA	3.65	-0.40	0.046	75.74	3.24E-18	0.67	0.62 - 0.74	yes	4.53E-17	0.155	0.118	0.037
IQ	1.24	-0.22	0.043	26.11	3.22E-07	0.80	0.74 - 0.87	yes	4.50E-06	0.131	0.118	0.012

* Based on Nagelkerke R^2 of the PRS
** upper and lower limits are shown

Table S9
Multivariable logistic regression analysis

Model	**Nagelkerke R^2**	**Explained variance in %**				
baseline model	0.119	11.9				
Full model	0.178	17.8				
only the PRSs	0.059	5.9				
INCLUDED						
PRS	**BETA**	**SE**	**Wald**	**P**	**OR**	**95% CI for OR ***
EA	-0.299	0.050	35.598	2.43E-09	0.741	0.67 - 0.81
SI	-0.166	0.045	13.469	0.000243	0.847	0.78 - 0.93
MDD	0.123	0.053	5.449	0.019585	1.131	1.02 - 1.25
ASB	0.086	0.042	4.252	0.039212	1.089	1.00 - 1.18
ADHD	0.081	0.044	3.409	0.06484	1.085	1.00 - 1.18
IQ	-0.068	0.047	2.090	0.148281	0.934	0.86 - 1.02
AD	0.053	0.042	1.559	0.21182	1.054	0.97 - 1.15
NEU	0.043	0.043	0.998	0.317823	1.044	0.96 - 1.14

baseline model covariates: 8 PCs, sex and chip
full model: baseline plus **univariate significant PRS**
All included PRS have SNP P-val threshold <1
*upper and lower limits are shown

4

Psychiatric Polygenic Risk Scores as Predictor for Attention Deficit/Hyperactivity Disorder and Autism Spectrum Disorder in a Clinical Child and Adolescent Sample

Jansen AG, Dieleman GC, Jansen PR, Verhulst FC, Posthuma D, Polderman TJC
Psychiatric Behav Genet. 2020 Jul;50(4):203-212

Abstract

Neurodevelopmental disorders such as attention deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) are highly heritable and influenced by many single nucleotide polymorphisms (SNPs). SNPs can be used to calculate individual polygenic risk scores (PRS) for a disorder. We aim to explore the association between the PRS for ADHD, ASD and for Schizophrenia (SCZ), and ADHD and ASD diagnoses in a clinical child and adolescent population. Based on the most recent genome wide association studies of ADHD, ASD and SCZ, PRS of each disorder were calculated for individuals of a clinical child and adolescent target sample (N = 688) and for adult controls (N = 943). We tested with logistic regression analyses for an association with (1) a single diagnosis of ADHD (N = 280), (2) a single diagnosis of ASD (N = 295), and (3) combining the two diagnoses, thus subjects with either ASD, ADHD or both (N = 688). Our results showed a significant association of the ADHD PRS with ADHD status (OR 1.6, $P = 1.39 \times 10^{-07}$) and with the combined ADHD/ASD status (OR 1.36, $P = 1.211 \times 10^{-05}$), but not with ASD status (OR 1.14, $P = 1$). No associations for the ASD and SCZ PRS were observed. In sum, the PRS of ADHD is significantly associated with the combined ADHD/ASD status. Yet, this association is primarily driven by ADHD status, suggesting disorder specific genetic effects of the ADHD PRS.

Introduction

Psychiatric disorders are heritable complex traits with varying heritability estimates. At the top end of the heritability range, reported heritabilities vary from 74% for ASD (Tick et al. 2016) to 80% for ADHD (Brikell et al. 2015), and 81% for SCZ (Sullivan et al. 2003). These traits likely have a similar genetic architecture with a role for common and rare variants, including de novo mutations and copy number variants (CNV) playing an important role (Gratten et al. 2014). Common genetic variation can be captured in a polygenic signal that contains a multitude of single nucleotide polymorphisms (SNPs) from many genes (Gratten et al. 2014; Sullivan et al. 2012). Genome-wide association studies (GWAS) are a highly successful method to identify the common variants that influence these disorders (Visscher et al. 2017). GWAS reveal increasingly more significantly associated loci. These represent the most associated part of the genetic signal. The most recent GWAS for ADHD, ASD, and SCZ identified 12, 5, and 145 independent associated loci, respectively (Demontis and Walters 2017; Grove et al. 2017; Pardiñas et al. 2018).

However, given the polygenicity of disorders like ADHD and ASD, also non-significantly associated SNPs are likely to contribute to the disorder (Wray et al. 2014). Hence, it is also of interest to investigate the non-genomewide significant component of the genetic signal.

One method to include the non-genome-wide significant component of the common genetic variation is the polygenic risk scores (PRS) approach. PRS are the sum of risk alleles weighted by their estimated effect size as determined in an independent GWAS sample, and can serve as such as an estimation of an individual's polygenic risk (Torkamani et al. 2018; Weiner et al. 2017; Wray et al. 2014). PRS estimated from an independent sample can be used for prediction between groups (e.g., cases and controls), or for stratifying groups of people according to high or low genetic risk as defined by their PRS. For example, in a sample of children from the general population, the SCZ PRS has shown positive associations with behavioral and emotional problems in children as young as 3 years old (Jansen et al. 2017). Similarly, the ADHD PRS has been associated with attention problems in children from the general population (Groen-Blokhuis et al. 2014), and with attentional and hyperactive-impulsive traits in another general population sample (age ~ 7 year, 7 months) (Martin et al. 2014).

As previous research indicates, the common genetic burden of different psychiatric disorders partially overlaps (Mitchell 2011). To add, both ADHD and ASD, as well as SCZ, are regarded neurodevelopmental disorders (NDD) (Mullin et al. 2013; Rapoport et al. 2012) and genetic studies have shown positive genetic correlations of 0.36 for ASD/ADHD (Grove et al. 2017), 0.211 for ASD/SCZ (Grove et al. 2017), and 0.122 for ADHD/SCZ (Demontis and Walters 2017).

In addition, it was shown that the prevalence of SCZ is significantly higher in an ASD sample compared to controls (OR 3.55, 95% CI 2.08–6.05, $P < 0.001$), and the prevalence of ASD in an SCZ samples ranges between 3.4 and 52% compared to 1% in the general population (Zheng et al. 2018). To add, ASD and SCZ share clinical features among which

social cognition (Cheung et al. 2010; *DSM 5* 2013), while ASD and ADHD share inattention (Craig et al. 2015; *DSM 5* 2013).

The current study adds to this literature by investigating associations of the ADHD, ASD, and SCZ PRS in a sample of children and adolescents referred to an outpatient university clinic. The children in this sample were assessed with standardized procedures generating clinical (DSM-IV) diagnoses as well as continuous rating scale scores on behavioral/emotional problems. We aim to investigate whether PRS for ADHD (Demontis and Walters 2017), ASD (Grove et al. 2017) and SCZ (Pardiñas et al. 2018) can distinguish ADHD and ASD cases from controls in this sample. Findings from genetic studies suggest a partly shared genetic diathesis underlying neurodevelopmental disorders (including SCZ, ASD and ADHD) (Bulik-Sullivan et al. 2015). We therefore hypothesized that the ADHD, ASD and SCZ PRS would be associated with the ADHD/ASD (either ASD, ADHD or both) diagnostic status. In addition, we expected both the ADHD and ASD PRS to be associated with ADHD and ASD respectively. In addition, we expected the SCZ PRS to be associated with ASD status given the genetic overlap previously reported (Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium 2017), although conflicting results with low (Cross-Disorder Group of the Psychiatric Genomics Consortium et al. 2013) or no (Vorstman et al. 2013) genetic association between ASD and SCZ have been reported as well. As a sensitivity analysis, we aim to perform a follow up correlation analysis and subsequently a linear regression analysis with the Child Behavioral Checklist (CBCL) subscales to validate the robustness of our findings and gain additional information on the link between associated genetic signals and specific behavioral or emotional problems, given a particular clinical diagnosis.

Methods

Sample

Psychiatric outpatient sample: "Inside-Out"

A new psychiatric outpatient sample called "Inside-Out" is analyzed. Data were collected from January 2001 until January 2012 at the department of Child and Adolescent Psychiatry at the Sophia Children's Hospital at Erasmus Medical Center in Rotterdam, resulting in a psychiatric outpatient sample. Before the first visit, parents and children received the CBCL from the Achenbach System of Empirically Based Assessment (ASEBA) (Achenbach and Rescorla 2001). In addition, DNA was extracted from saliva and genotyping was performed on the Illumina PsychChip array (see data). The procedure was approved by the ethical committee of the Erasmus Medical Center. The total Inside-Out sample comprises 1941 children diagnosed with one or more DSMIV disorders (ASD, ADHD, tic disorder, obsessive compulsive disorder (OCD), depression, anxiety, anorexia nervosa (AN), eating disorder NOS, RETT syndrome and subcategories of mentioned disorders) and children with a delayed diagnostic status or children who did not receive a DSM diagnosis (27.9%).

The diagnostic procedure consisted of an interview with parents, a semi-structured interview with the child based on the Semi-structured Clinical Interview for Children and Adolescents (McConaughy and Achenbach 2001), the Diagnostic Interview Schedule for Children IV-P (Shaffer et al. 2000) and the Autism Diagnostic Observation Schedule-Generic (Lord et al. 1989) in case of a suspected autism spectrum disorder. Diagnostic classification was done by a clinician according to the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV). The above-mentioned procedure was part of standard clinical practice for the current study, genetic and clinical information was used of the children who received an ADHD diagnosis, no ASD co-diagnosis allowed (N = 280, age range: 3.3–18.5 years, mean: 9.06, SD: 2.66) or an ASD diagnosis, no ADHD co-diagnosis allowed (RETT excluded) (N = 295, age range: 2.5–18.3 years, mean: 9.02, SD: 3.55). In addition we used a sample of combined ADHD and ASD diagnoses where comorbidity of ADHD and ASD was allowed, adding another 113 children to this combined sample (N = 688, age range 2.5–18.5, mean: 8.96, SD: 3.07). The target sample was diagnosed with the DSM-IV and includes many cases with Asperger and pervasive developmental disorder-not otherwise specified (PDD-NOS) diagnoses (82% of total ASD sample). ADHD and ASD co-diagnosed children (N = 113) were not included in the ADHD and ASD sample. For sample specifics see Tables 1 and 2.

Population-based control sample

As a control sample, we used a Dutch population sample (NESCOG, N = 943, age range: 17.0–79.0) previously described by Polderman et al. (2013). NESCOG comprises a general population and a family-based sample of which closely related individuals were excluded. Data were collected on cognitive tasks, behavioral conditions (such as ADHD and ASD symptoms), life events, personality and environmental factors, as well as genetic information. Moreover, to correct for undiagnosed ADHD status, participants scoring 3 SD above the mean on the Conners' Adult ADHD Rating Scale (CAARS) (Conners et al.), or the Attention Problems scale of the Young Adult Self Report (YASR) (Achenbach 1997) were excluded. Participants scoring three SD above mean on the Autism Quotient (AQ) (Baron-Cohen et al. 2001) were also excluded, resulting in a final control sample of 943 participants (age range 17–79, 38% male), see Tables 1 and 2.

Data

Genotyping of the cases and controls was performed on the same Illumina PsychChip array. The PsychChip SNP array contains HumanCore, Human Exome and custom content to accurately capture genetic variants previously linked with psychiatric disorders (https ://www.illum ina.com/produ cts/ by-type/microa rray- kits/infini um-psycharray. htm l).

Genetic variants in the clinical sample were filtered based on minor allele frequency (MAF < 1%), Hardy–Weinberg disequilibrium ($P < 1 \times 10^{-6}$) and SNP call rate (< 95%). Individuals were subsequently filtered based on relatedness (pairwise Identity-By-Descent (IBD) > 0.185), genotype and phenotypic sex mismatch, outlying heterozygosity and non European ancestry (4 SD outside the range of the first two genetic principal components of the HapMap3 European founder population (CEU)) resulting in a clinical sample of 812 patients of which 688 are diagnosed with ADHD, ASD or both. The remaining part of the children in this sample (N = 124) are diagnosed with either Rett syndrome, anorexia nervosa or other eating disorders, tourette disorder, or other disorders. Another subset of the sample is currently being genotyped and includes children diagnosed with anxiety disorder, affective disorder or other disorders. In the control sample, SNP filtering was based on MAF (< 1%) Hardy–Weinberg disequilibrium ($P < 0.00001$) and SNP call rate (< 95%). Individual QC was based on missingness (> 5%), ancestry (within the range of 1000G CEU population on first PCs), relatedness (pairwise IBD > 0.185), gender mismatch, outlying heterozygosity and missing phenotypes.

Sex differences in the samples

The case and control samples differed in sex distribution (cases are 75% and the controls 25% males). Therefore, we compared allele frequencies between males and females in an independent sample, GoNL (see www.nlgenome.nl for more information), by means of correlation. The Pearson correlation coefficient between the male and female allele frequencies is 0.99, removing concerns of different allele frequencies in the two samples due to sex differences.

Table 1 Sample overview

Sample	N	Reference
Discovery samples used for computation of PRS		
Discovery sample ADHD	Cases: 20,183 Controls: 35,191	Demontis et al. (2019)
Discovery sample ASD	Cases: 18,381 Controls: 27,969	Grove et al. (2019)
Discovery sample SCZ	Cases: 40,675 Controls: 64,643	Pardiñas et al. (2018)

Sample	N	Additional information
Target samples used for case control studies: Inside-out (logistic regression)		
ADHD/ASD sample	688	ADHD/ASD comorbidity allowed, therefore including 113 extra children
ADHD sample	280	Subset of ADHD/ASD sample based on diagnostic status. ADHD/ASD comorbidity NOT allowed
ASD sample	295	Subset of ADHD/ASD sample based on diagnostic status. ADHD/ASD comorbidity NOT allowed
Control sample	943	NESCOG general population sample corrected for high scores on AQ and CAARS
Target sample used for sensitivity analysis: inside-out (correlations PRS-syndrome and CBCL scales)		
ADHD/ASD sample	530	Subset of ADHD/ASD sample based on the presence of the CBCL for age 6-18 and hence, diagnostic age. ADHD/ASD comorbidity allowed

ASD autism spectrum disorder, ADHD attention deficit/hyperactivity disorder, SCZ schizophrenia, CBCL child behavioral checklist, AQ autism quotient, CAARS Conners' Adult ADHD Rating Scale

Table 2 Sample description

	Sample logistic regression				**Sample Correlation analysis**
	ADHD /ASD	**ADHD**	**ASD**	**Control**	**ADHD /ASD ***
N	688	280	295	943	530
Age range (mean, SD) in years	2.5- 18.5 (8.96, 3.07)	3.3 -18,5 (9.06, 2.66)	2.5- 18.3 (9.02, 3.55)	17.0- 79.0 (44.47, 13.94)	6.05 – 18.52 (9.7, 2.60)
Gender % male	76	75	73	38	75

ADHD/ASD = ADHD (280) + ASD (295) plus children codiagnosed with ADHD and ASD (113)

*Sample size differs from the sample size for the logistic regression due to CBCL 6-18 (age range) availability

Polygenic risk scoring

The PRS is constructed as the sum of risk alleles weighted by their effect size. Per disorder several PRS were calculated with different *P* value inclusion thresholds (*P*-values: < 0.01, < 0.05, < 0.1, < 0.2, < 0.3, < 0.4, < 0.5, < 1). Starting from a low *P*-value threshold moving up to *P*-value 1, an optimal *P*-value threshold with the highest explained variance was identified, including the most truly associated positives. After this threshold more false positives will be included dampening the true signal (Wray et al. 2014). Prior to our calculation of the PRS, the SNPs were pruned (LD R^2 < 0.1, 250 kb pair window) to remove variants in LD. Polygenic scoring was performed with the software package PRSice (Euesden et al. 2015). The PRS for ASD, ADHD and SCZ were constructed using the most recent summary statistics from GWAS with the largest publicly available sample size, ADHD (Demontis and Walters 2017) (20,183 cases and 35,191 controls), ASD (Grove et al. 2017) (18,382 cases and 27,969 controls), and SCZ (Pardiñas et al. 2018) (40,675 cases and 64,643 controls). Of note, the Inside-Out and the control sample are independent samples, not included in these GWAS. After polygenic scoring the results were standardized to mean 0 and SD 1 for interpretational purposes. For the number of SNPs included in the scores see Supplementary Table S1.

Behavioral measurements

Child emotional and behavioral problems were assessed using the Dutch version of the Child Behavior Checklist/6–18 (CBCL) (Achenbach and Rescorla 2001) filled out by the parent before the first visit to the hospital. The CBCL contains 113 problem items that can be scored on eight syndrome scales (Anxious/Depressed N_{item} = 13, Withdrawn/ Depressed N_{item} = 8, Somatic Complaints N_{item} = 11, Social Problems N_{item} = 11, Thought Problems N_{item} = 15, Attention Problems N_{item} = 10, Rule Breaking Behavior N_{item} = 17 and Aggressive Behavior N_{item} = 18). Parents score each problem on a three-point scale (0: not true, 1: somewhat or sometimes true, 2: very or often true). This follow up analysis included children with a CBCL 6–18 report, completed by the parent less then a year before receiving the diagnosis. If a CBCL from within a year before diagnosis was not present the person was excluded from this part of the analysis. In all analyses, sum scores on the CBCL syndrome scales were used.

Statistical analysis

Case control analysis on the association between PRS and disease status

We performed logistic regression analyses to investigate if the ADHD, ASD or SCZ PRS can distinguish between cases and controls in a sample (1) with a diagnosis of ADHD, ASD not permitted as co-diagnosis (ADHD, N = 280), (2) with a diagnosis of ASD, ADHD not permitted as co-diagnosis (ASD, N = 295), and (3) combining the first two samples, thus subjects with either ASD, ADHD or both (ADHD/ ASD, N = 688). For each PRS, eight different SNP inclusion thresholds were tested. All *P*-values were corrected for multiple testing by means of Bonferroni correction (72 tests: three samples (ADHD, ASD, ADHD/ASD), three PRS (ADHD, ASD, SCZ), eight PRS thresholds (0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 1) per disorder). To account for population stratification we included eight principal components (PCs). The PCs were calculated based on the pruned data with Eigensoft (Price et al. 2006) (version 3.0) software. Additionally, sex was added as a covariate. Age was not added as a covariate as all cases are children and all controls are adults.

Sensitivity analysis: correlation and association between CBCL syndrome scales and PRS

We aim to provide additional evidence for the significant association of the PRS and the disorders as measured by the CBCL score severity. Given statistical power, we tested the association with symptom severity only in the combined ADHD/ASD sample by calculating the correlation between the significantly associated PRS (i.e., ADHD) and the syndrome scales of the CBCL. Age was added as a covariate in addition to the previously used eight PCs and sex. All analyses were performed in IBM SPSS statistics 21.

Results

Case control analysis on the association between PRS and disorder status

The ADHD PRS showed significant associations before multiple testing correction with disorder status in all three samples (Table 3). As shown in Fig. 1, all ADHD PRS *P*-value thresholds remained significant after Bonferroni multiple testing correction in both the combined ADHD/ASD and ADHD sample, but not the ASD sample. The most stringent *P*-value threshold of 0.01 generated a positive association in the ADHD/ASD sample OR 1.28 ($P = 1.3 \times 10^{-3}$), and ADHD sample OR 1.4 ($P = 3.6 \times 10^{-4}$). The most optimal *P*-value threshold as defined by explained variance, OR and P-value was 0.3 for the ADHD/ASD sample ($R^2 = 0.02$, OR 1.36, $P = 1.21 \times 10^{-05}$), and 0.4 for the ADHD sample ($R^2 = 0.045$, OR 1.62, $P = 5.75 \times 10^{-08}$).

The most lenient *P*-value threshold of $P < 1$ had a significant association in the ADHD/ASD sample, OR 1.35 ($P = 1.9 \times 10^{-5}$), and also in the ADHD sample OR 1.62 ($P = 4.73 \times 10^{-8}$). In the ASD sample none of the results remained significant after Bonferroni correction.

The ASD and SCZ PRS were not significantly associated with the ADHD, ASD, or combined ADHD/ASD status. The SCZ PRS including all SNPs (*P*-value threshold $P < 1$) showed a trend towards association in the ADHD/ASD sample (OR 1.13, $P = 5.72 \times 10^{-2}$) (Supplementary Tables S2 and S3 and Figs. S1 and S2).

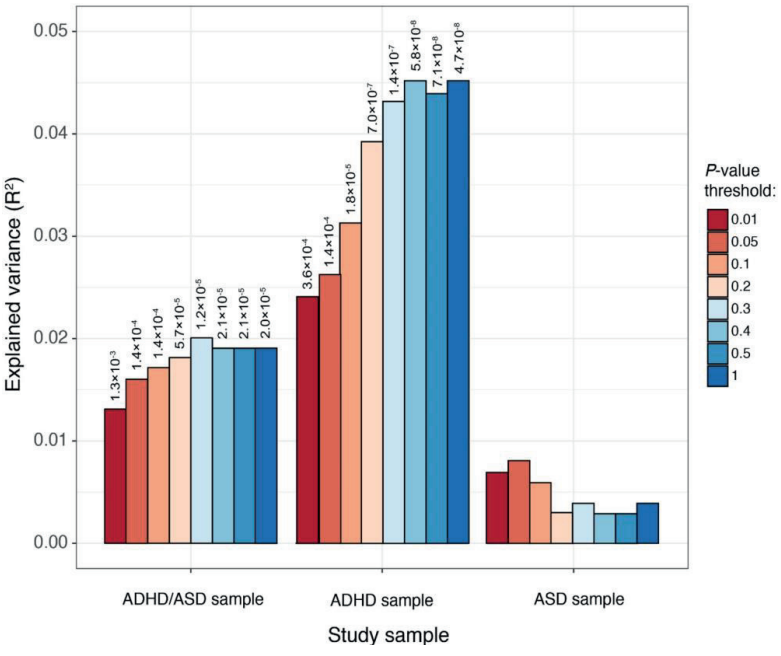


Fig. 1
Variance explained (Nagelkerke R^2) by the ADHD PRS. All SNP inclusion *P*-value thresholds are shown from low to high. *P*-values are shown on top of each bar and are Bonferroni corrected

Table 3 Results of the logistic regression analyses for the ADHD PRS

ADHD PRS uncorrected threshold	B	Wald p	Bonferroni corr. Wald p	OR	Nagelkerke R^2 PRS
ADHD/ASD sample (N = 688)					
0.01	0.243	**1.80×10^{-05}**	**1.30×10^{-03}**	1.275	0.013
0.05	0.274	**2.00×10^{-06}**	**1.44×10^{-04}**	1.316	0.016
0.1	0.278	**2.00×10^{-06}**	**1.44×10^{-04}**	1.321	0.017
0.2	0.287	**7.91×10^{-07}**	**5.70×10^{-05}**	1.333	0.018
0.3	0.304	**1.68×10^{-07}**	**1.21×10^{-05}**	1.355	0.020
0.4	0.297	**2.96×10^{-07}**	**2.13×10^{-05}**	1.346	0.019
0.5	0.297	**2.88×10^{-07}**	**2.07×10^{-05}**	1.346	0.019
1	0.297	**2.71×10^{-07}**	**1.95×10^{-05}**	1.346	0.019
ADHD sample (N = 280)					
0.01	0.337	**5.00×10^{-06}**	**3.60×10^{-04}**	1.401	0.024
0.05	0.356	**2.00×10^{-06}**	**1.44×10^{-04}**	1.428	0.026
0.1	0.401	**2.52×10^{-07}**	**1.82×10^{-05}**	1.493	0.031
0.2	0.454	**9.68×10^{-09}**	**6.97×10^{-07}**	1.574	0.039
0.3	0.472	**1.93×10^{-09}**	**1.39×10^{-07}**	1.603	0.043
0.4	0.482	**7.98×10^{-10}**	**5.75×10^{-08}**	1.620	0.045
0.5	0.479	**9.87×10^{-10}**	**7.11×10^{-08}**	1.614	0.044
1	0.485	**6.57×10^{-10}**	**4.73×10^{-08}**	1.625	0.045
ASD sample (N = 295)					
0.01	0.176	**1.45×10^{-02}**	1	1.192	0.007
0.05	0.201	**7.33×10^{-03}**	5.28×10^{-01}	1.222	0.008
0.1	0.169	**2.35×10^{-02}**	1	1.184	0.006
0.2	0.132	7.82×10^{-02}	1	1.141	0.003
0.3	0.135	6.83×10^{-02}	1	1.144	0.004
0.4	0.119	1.05×10^{-01}	1	1.127	0.003
0.5	0.129	7.83×10^{-02}	1	1.138	0.003
1	0.130	7.68×10^{-02}	1	1.139	0.004

All models have eight PCs and sex as covariate (baseline model). Bonferroni P-value corrected for 72 tests. Sig. P-values are shown in bold. Results of the logistic regression analyses for the ASD and SCZ PRS are presented in the Supplementary Tables 3 and 4

Sensitivity analysis: association between CBCL syndrome scales and the ADHD PRS

Based on the correlational structure in the ADHD/ASD sample (Supplemental Material Table 4) between the CBCL syndrome scale scores and the ADHD PRS *P*-value thresholds, we concluded the correlation was too low (all correlations ≤ 0.1) to warrant the linear regression analysis. Mean scores and standard deviations for the CBCL syndrome scale scores for the ADHD/ASD sample are provided in Supplementary Table 5.

Discussion

This study investigated the associations of PRS for ADHD, ASD and SCZ, with ADHD and ASD status in a clinical child and adolescent population. As hypothesized, we found a significant association between the ADHD PRS and the combined ADHD/ASD status, and the separate ADHD status. The PRS SNP-inclusion thresholding resulted in a rise of explained variance with increasing *P*-value thresholds, showing that in addition to the GWAS significant hits, the non-significant SNPs in the ADHD GWAS also contribute to the associations with diagnostic status. Given the comorbidity between ADHD and ASD, and previously reported genetic correlations, we expected that the ADHD PRS would also be associated with ASD status, however, this association was not observed in our data. In contrast, the current results suggest a disorder specific effect of ADHD associated SNPs instead of a shared common genetic mechanism with ASD. The ADHD PRS is based on the most recent GWAS results, and explained variance up to 4.5% in our sample, which is in line with the results from the initial GWAS (Demontis and Walters, 2017) who reported an explained variance of 5.5%, making it a promising PRS for further use in research on ADHD.

Contrary to our expectation, the ASD and SCZ PRS were not associated with any of the diagnostic groups. The null results for the ASD PRS are unexpected as the initial GWAS (Grove et al. 2017) reported an explained variance of 2.45% in an independent sample, and their summary statistics were used for the analysis. Given that the discovery sample size of ASD was only slightly smaller than the ADHD sample, and the SCZ sample was even larger, we do not expect that sample size alone explains these findings. Moreover, apart from sample size, power analyses usually take several parameters into account, including the heritability and population prevalence of traits, the amount of SNPs included in the GWAS, the effective number of chromosome segments, and the proportion of cases in discovery and target sample (Lee et al. 2017). In our study, the discovery and target samples were for most of the parameters similar across disorders, except for prevalence rates (ASD and SCZ have a population prevalence of 1%, and ADHD has a population prevalence of 5%).

Regarding the null result for the ASD PRS one explanation might be a difference in the diagnostic sample composition of the ASD GWAS discovery sample compared to the target ASD sample. The target sample was diagnosed with the DSM-IV and includes many cases with asperger, and PDD-NOS diagnoses (82% of total ASD sample), which might differ from the

discovery sample. Moreover, about one-third of the discovery sample were trio data (i.e. case pseudo control design), of which it has been suggested that the un-transmitted chromosomes contain increased polygenic burden, and as such the genetic signal based on these data might be decreased (Peyrot et al., 2016). Additionally, the genetic architecture of ADHD might differ from ASD, e.g., rare genetic variants might comprise a more important part of the genetic contribution to ASD (Geschwind and State, 2015) compared to ADHD. With growing sample sizes, genetic discoveries will increase and become more reliable, potentially allowing the identification of rare variants.

The choice of including the SCZ PRS was based partly on the higher prevalence rate of SCZ in ASD individuals compared to the general population, a recent systematic review reports a significantly higher SCZ prevalence in ASD individuals compared to the general population (OR 3.55, 95% confidence interval (CI) 2.08–6.05, P < 0.001) (Zheng et al., 2018). If the actual SCZ prevalence rate in an ASD population resides at the lower end of the of the 95% CI the enrichment of common SCZ SNPs might not be detectable in our relatively small sample. Additionally, the genetic correlation of 0.211 between ASD and SCZ (Grove et al., 2017) and 0.122 between SCZ and ADHD (Demontis and Walters, 2017) might be too small to detect the genetic overlap between the two disorders in our data. Finally, it is possible that ASD has a different genetic underpinning with more rare variants than SCZ although some overlap has been reported in rare genetic variation between ASD and SCZ (Sanders et al., 2015). Recent whole-genome sequence research on height fully recovered the heritability of this trait, meaning that next to the previously established common variants, all rare variants have been discovered (Wainschtein et al., 2019). Whole-genome sequence research into ASD, SCZ and ADHD might shed light on this issue revealing the genetic architecture of these traits.

The sensitivity analyses exploring the associations between scores on the syndrome scales of the CBCL and the ADHD PRS showed low correlations between these two measures, as such we decided not to pursue the follow-up analysis further. A reason for the low correlations can be the amount variance explained by the ADHD PRS. The explained variance of 4.5% might not be enough to give meaningful results in follow-up analysis using the CBCL in a smaller sample like “Inside out”. In addition, a diagnosis is not based solely on the CBCL results but includes careful evaluation by an experienced psychologist/psychiatrist based on a patient interview, a parent interview and if possible an evaluation by a third party like a school teacher of the child.

Strengths and limitations

A strength of our study is the adult control sample as, in contrast to a child sample, the chance that adult individuals will receive a future ADHD or ASD diagnosis is limited compared to young individuals i.e., these disorders are usually diagnosed during childhood (Nylander et al., 2013), while DNA sequences are fixed during life.

One concern might be the difference in sex distribution between the samples, with the clinical sample consisting of 75% males and the control sample having an opposite skew in sex distribution, as this could potentially affect the observed associations between the PRS and diagnoses. However, we compared the allele frequencies between males and females in an independent sample (GoNL (Genome of the Netherlands Consortium 2014)) and found no differences. Yet, due to the skewed sex distribution we could not examine sex-PRS interactions, or sex specific associations, which would both be interesting to investigate given the higher prevalence of males in both ADHD and ASD.

We also need to take into account that the ADHD/ASD group comprises the ADHD and ASD groups and that this is no official diagnostic disorder classification. The results should be replicated in a comparable independent sample first before firm conclusions can be drawn.

Overall, despite the fact that symptoms overlap between the neurodevelopmental disorders, our study does not directly imply that the umbrella of NDD is present at the common genetic level as captured in the PRS. As the ASD and SCZ PRS do not distinguish cases from controls in any of our diagnostic samples it is possible that ADHD, ASD and SCZ have a different common genetic signature. Moreover, the results should be replicated in one or more independent samples.

A final remark can be made on the cross sectional nature of the sample. Unlike longitudinal studies, measures are available for one point in time for most of the subjects. This presents the possibility that children might receive additional diagnoses later on in life resulting in a change in diagnostic status from ADHD or ASD to the ADHD/ASD codiagnosed group, or to other comorbidities.

Conclusions

In conclusion, the PRS of ADHD is significantly associated with the combined ADHD/ASD and ADHD status. Yet, this association is primarily driven by ADHD status, suggesting disorder specific genetic effects of the ADHD PRS. Nevertheless, it is of interest to explore the genetic predictive value of other psychiatric disorders besides neurodevelopmental disorders. Improving genetic prediction in neurodevelopmental disorders by using a multi-trait predictor instead of single-trait predictors is also an interesting option (Maier et al., 2018). Lastly, it is of interest to delve deeper into the association between the ADHD PRS and the specific emotional and behavioral problems in larger samples as those data may provide additional information on specific problems or the severity of problems within a diagnostic status.

References

Achenbach TM (1997) Manual for the young adult self-report and young adult behavior checklist. Department of Psychiatry, University of Vermont, Burlington

Achenbach TM, Rescorla LA (2001) Manual for the ASEBA schoolage forms & profiles. University of Vermont, Research Center for Children, Youth, & Families, Burlington

American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, vol 5. American Psychiatric Association, Washington, DC

Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium (2017) Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol Autism 8:21. https: //doi.org/10.1186/s13229-017-0137- 9

Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E (2001) The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. J Autism Dev Disord 31(1):5–17

Brikell I, Kuja-Halkola R, Larsson H (2015) Heritability of attention-deficit hyperactivity disorder in adults. Am J Med Genet B 168(6):406–413. https ://doi.org/10.1002/ajmg.b.32335

Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R et al (2015) An atlas of genetic correlations across human diseases and traits. Nat Genet. https ://doi.org/10.1038/ng.3406

Cheung C, Yu K, Fung G, Leung M, Wong C, Li Q et al (2010) Autistic disorders and schizophrenia: related or remote? An anatomical likelihood estimation. PLoS ONE 5(8):e12233. https ://doi. org/10.1371/journ al.pone.00122 33

Conners CK, Erhardt D, Sparrow E (z.d.). Conner's adults ADHD rating scales. North Tonawanda, New York

Craig F, Lamanna AL, Margari F, Matera E, Simone M, Margari L (2015) Overlap between autism spectrum disorders and attention deficit hyperactivity disorder: searching for distinctive/ common clinical features. Autism Res 8(3):328–337. https ://doi. org/10.1002/aur.1449

Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM et al (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 45(9):984–994. https ://doi. org/10.1038/ng.2711

Demontis D, Walters RK (2017) Discovery of the first genomewide significant risk loci for ADHD. Nat Genet. https ://doi. org/10.1101/14558 1

Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, Baldursson G, Belliveau R, Bybjerg-Grauholm J, Bækvad-Hansen M et al (2019) Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet 51:63–75

Euesden J, Lewis CM, O'Reilly PF (2015) PRSice: polygenic risk score software. Bioinformatics 31(9):1466–1468. https ://doi. org/10.1093/bioin forma tics/btu84 8

Genome of the Netherlands Consortium (2014) Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet 46(8):818–825. https :// doi.org/10.1038/ng.3021

Geschwind DH, State MW (2015) Gene hunting in autism spectrum disorder: on the path to precision medicine. Lancet Neurol 14(11):1109–1120. https: //doi.org/10.1016/S1474-4422(15)00044 -7

Gratten J, Wray NR, Keller MC, Visscher PM (2014) Large-scale genomics unveils the genetic architecture of psychiatric disorders. Nat Neurosci 17(6):782–790. https ://doi.org/10.1038/nn.3708

Groen-Blokhuis MM, Middeldorp CM, Kan K-J, Abdellaoui A, van Beijsterveldt CEM, Ehli EA et al (2014) Attention-deficit/hyperactivity disorder polygenic risk scores predict attention problems in a population-based sample of children. J Am Acad Child Adolesc Psychiatry 53(10):1123–1129.e6. https ://doi.org/10.1016/j.

jaac.2014.06.014

Grove J et al (2017) Common risk variants identified in autism spectrum disorder. bioRxiv. https ://doi.org/10.1101/22477 4

Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, Pallesen J et al (2019) Identification of common genetic risk variants for autism spectrum disorder. Nat Genet 51(3):431–444. https ://doi. org/10.1038/s4158 8-019-0344-8

Jansen PR, Polderman TJC, Bolhuis K, van der Ende J, Jaddoe VWV, Verhulst FC et al (2017) Polygenic scores for schizophrenia and educational attainment are associated with behavioural problems in early childhood in the general population. J Child Psychol Psychiatry. https ://doi.org/10.1111/jcpp.12759

Lee SH, Weerasinghe WMSP, Wray NR, Goddard ME, van der Werf JHJ (2017) Using information of relatives in genomic prediction to apply effective stratified medicine. Sci Rep 7:42091. https: //doi. org/10.1038/srep4 2091

Lord C, Rutter M, Heemsbergen J, Jordan H, Mawhood L, Schopler E (1989) Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. J Autism Dev Disord 19(2):185–212

Maier RM, Zhu Z, Lee SH, Trzaskowski M, Ruderfer DM, Stahl EA et al (2018) Improving genetic prediction by leveraging genetic correlations among human diseases and traits. Nat Commun 9(1):989. https ://doi.org/10.1038/s4146 7-017-02769 -6

Martin J, Hamshere ML, Stergiakouli E, O'Donovan MC, Thapar A (2014) Genetic risk for attention-deficit/hyperactivity disorder contributes to neurodevelopmental traits in the general population. Biol Psychiatry 76(8):664–671. https ://doi.org/10.1016/j. biops ych.2014.02.013

McConaughy SH, Achenbach TM (2001) Manual for the semi-structured clinical interview for children and adolescents. Department of Psychiatry, University of Vermont, Burlington

Mitchell KJ (2011) The genetics of neurodevelopmental disease. Curr Opin Neurobiol 21:197–203

Mullin AP, Gokhale A, Moreno-De-Luca A, Sanyal S, Waddington JL, Faundez V (2013) Neurodevelopmental disorders: mechanisms and boundary definitions from genomes, interactomes and proteomes. Transl Psychiatry 3(12):e329. https ://doi.org/10.1038/ tp.2013.108

Nylander L, Holmqvist M, Gustafson L, Gillberg C (2013) Attentiondeficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) in adult psychiatry. A 20-year register study. Nord J Psychiatry 67(5):344–350. https ://doi.org/10.3109/08039

488.2012.74882 4

Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N et al (2018) Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet 50(3):381–389. https ://doi.org/10.1038/ s4158 8-018-0059-2

Peyrot WJ, Boomsma DI, Penninx BWJH, Wray NR (2016) Disease and polygenic architecture: avoid trio design and appropriately account for unscreened control subjects for common disease.

Am J Hum Genet 98(2):382–391. https ://doi.org/10.1016/j. ajhg.2015.12.017

Polderman TJC, Hoekstra RA, Vinkhuyzen AAE, Sullivan PF, van der Sluis S, Posthuma D (2013) Attentional switching forms a genetic link between attention problems and autistic traits in adults. Psychol Med 43(9):1985–1996. https: //doi.org/10.1017/S003329171 20028 63

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38(8):904–909. https ://doi.org/10.1038/ng184 7

Rapoport JL, Giedd JN, Gogtay N (2012) Neurodevelopmental model of schizophrenia: update 2012. Mol Psychiatry 17(12):1228–1238.

https ://doi.org/10.1038/mp.2012.23

Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE et al (2015) Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron 87(6):1215–1233. https ://doi.org/10.1016/j.neuro n.2015.09.016

Shaffer D, Fisher P, Lucas CP, Dulcan MK, Schwab-Stone ME (2000) NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. J Am Acad Child Adolesc Psychiatry 39(1):28–38. https ://doi. org/10.1097/00004 583-20000 1000-00014

Sullivan PF, Kendler KS, Neale MC (2003) Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry 60(12):1187–1192. https: //doi.org/10.1001/archp syc.60.12.1187

Sullivan PF, Daly MJ, O'Donovan M (2012) Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat Rev Genet 13(8):537–551. https ://doi.org/10.1038/nrg32 40

Tick B, Bolton P, Happé F, Rutter M, Rijsdijk F (2016) Heritability of autism spectrum disorders: a meta-analysis of twin studies. J Child Psychol Psychiatry 57(5):585–595. https: //doi.org/10.1111/ jcpp.12499

Torkamani A, Wineinger NE, Topol EJ (2018) The personal and clinical utility of polygenic risk scores. Nat Rev Genet. https ://doi. org/10.1038/s4157 6-018-0018-x

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J (2017) 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet 101(1):5–22. https ://doi. org/10.1016/j.ajhg.2017.06.005

Vorstman JAS, Anney RJL, Derks EM, Gallagher L, Gill M, de Jonge MV et al (2013) No evidence that common genetic risk variation is shared between schizophrenia and autism. Am J Med Genet B 162B(1):55–60. https ://doi.org/10.1002/ajmg.b.32121

Wainschtein P, Jain DP, Yengo L, Zheng Z et al (2019) Recovery of trait heritability from whole genome sequence data. bioRxiv. https ://doi.org/10.1101/58802 0

Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J et al (2017) Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nat Genet 49(7):978–985. https: //doi. org/10.1038/ng.3863

Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM (2014) Research review: polygenic methods and their application to psychiatric traits. J Child Psychol Psychiatry 55(10):1068–1087. https ://doi.org/10.1111/jcpp.12295

Zheng Z, Zheng P, Zou X (2018) Association between schizophrenia and autism spectrum disorder: a systematic review and metaanalysis. Autism Res 11(8):1110–1119. https ://doi.org/10.1002/ aur.1977

Chapter 4 Supplementary information

Table of Contents

Figure S1

Variance explained (Nagelkerke R^2) by the ASD PRS under the best model, baseline with eight PCs, sex and the relevant PRS. All SNP inclusion *P*-value thresholds are shown. Given *P*-values are Bonferroni corrected, only sig. *P*-values are provided.

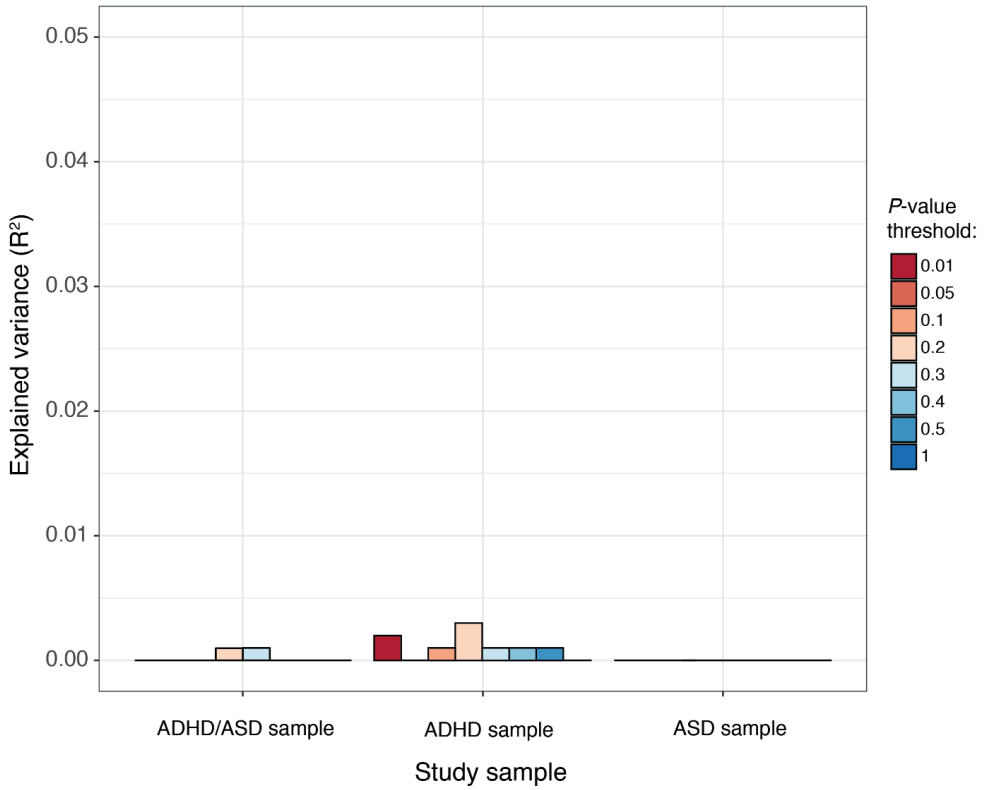


Figure S2

Variance explained (Nagelkerke R^2) by the SCZ PRS under the best model, baseline with eight PCs, sex and the relevant PRS. All SNP inclusion *P*-value thresholds are shown. Given *P*-values are Bonferroni corrected, only sig. *P*-values are provided.

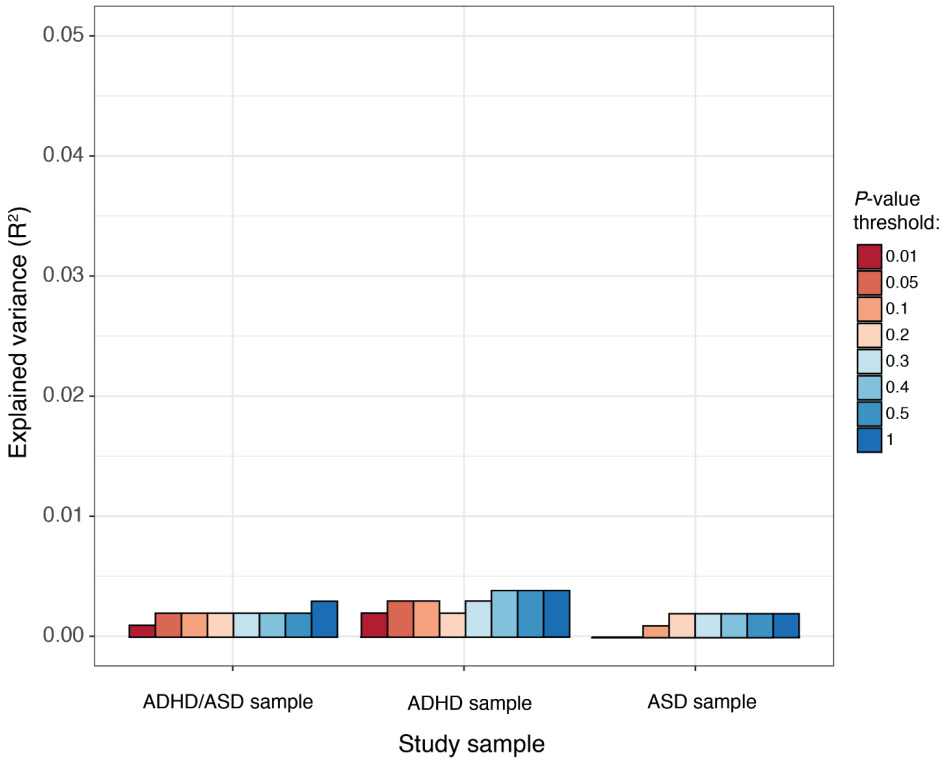


Table S1

Amount of SNPs included per P value threshold and per disorder

ADHD	
PRS P value threshold	**amount SNPS**
0.01	1409
0.05	5112
0.1	8930
0.2	15820
0.3	21923
0.4	27644
0.5	32980
1	55487
ASD	
PRS P value threshold	**amount SNPS**
0.01	1233
0.05	4935
0.1	8754
0.2	15785
0.3	22277
0.4	28227
0.5	33684
1	57277
SCZ	
PRS P value threshold	**amount SNPS**
0.01	4464
0.05	10214
0.1	14755
0.2	21545
0.3	27056
0.4	31997
0.5	36460
1	54524

Table S2 Results logistic regression ASD PRS and case control status

ADHD/ASD sample					
ASD PRS P value threshold	**B**	**Wald p uncorrected**	**Bonferroni corr. Wald p**	**OR**	**Nagelkerke R^2 PRS**
0.01	0.004	9.38E-01	1	1.004	0.000
0.05	0.016	7.74E-01	1	1.016	0.000
0.1	0.045	4.17E-01	1	1.046	0.000
0.2	0.065	2.42E-01	1	1.067	0.001
0.3	0.048	3.98E-01	1	1.049	0.001
0.4	0.034	5.43E-01	1	1.035	0.000
0.5	0.033	5.65E-01	1	1.033	0.000
1	0.028	6.26E-01	1	1.028	0.000
ADHD sample					
ASD PRS P value threshold	**B**	**Wald p uncorrected**	**Bonferroni corr. Wald p**	**OR**	**Nagelkerke R^2 PRS**
0.01	-0.092	2.09E-01	1	0.912	0.002
0.05	0.036	6.21E-01	1	1.037	0.000
0.1	0.078	2.86E-01	1	1.081	0.001
0.2	0.110	1.32E-01	1	1.116	0.003
0.3	0.065	3.78E-01	1	1.067	0.001
0.4	0.052	4.82E-01	1	1.053	0.001
0.5	0.062	4.04E-01	1	1.064	0.001
1	0.047	5.29E-01	1	1.048	0.000
ASD sample					
ASD PRS P value threshold	**B**	**Wald p uncorrected**	**Bonferroni corr. Wald p**	**OR**	**Nagelkerke R^2 PRS**
0.01	0.087	2.16E-01	1	1.09	0.000
0.05	0.000	9.99E-01	1	1.00	0.000
0.1	0.024	7.36E-01	1	1.02	0.000
0.2	0.034	6.36E-01	1	1.03	0.000
0.3	0.037	6.01E-01	1	1.04	0.000
0.4	0.012	8.69E-01	1	1.01	0.000
0.5	0.001	9.85E-01	1	1.00	0.000
1	0.010	8.88E-01	1	1.01	0.000

Note: Covariates included were eight PCs and sex. Multiple testing correction was applied for 72 tests

Table S3 Results logistic regression SCZ PRS and case control status

ADHD/ASD sample					
SCZ PRS P value threshold	**B**	**Wald p uncorrected**	**Bonferroni corr. Wald p**	**OR**	**Nagelkerke R^2 PRS**
0.01	0.077	1.87E-01	1	1.080	0.001
0.05	0.090	1.35E-01	1	1.095	0.002
0.1	0.100	1.05E-01	1	1.105	0.002
0.2	0.115	6.95E-02	1	1.121	0.002
0.3	0.109	8.77E-02	1	1.115	0.002
0.4	0.112	8.06E-02	1	1.118	0.002
0.5	0.115	7.26E-02	1	1.122	0.002
1	0.122	5.72E-02	1	1.130	0.003
ADHD sample					
SCZ PRS P value threshold	**B**	**Wald p uncorrected**	**Bonferroni corr. Wald p**	**OR**	**Nagelkerke R^2 PRS**
0.01	0.106	1.71E-01	1	1.112	0.002
0.05	0.129	1.10E-01	1	1.138	0.003
0.1	0.125	1.29E-01	1	1.133	0.003
0.2	0.120	1.53E-01	1	1.128	0.002
0.3	0.138	1.05E-01	1	1.148	0.003
0.4	0.152	7.43E-02	1	1.164	0.004
0.5	0.153	7.21E-02	1	1.166	0.004
1	0.154	7.06E-02	1	1.167	0.004
ASD sample					
SCZ PRS P value threshold	**B**	**Wald p uncorrected**	**Bonferroni corr. Wald p**	**OR**	**Nagelkerke R^2 PRS**
0.01	0.026	7.25E-01	1	1.027	0.000
0.05	0.044	5.70E-01	1	1.045	0.000
0.1	0.085	2.84E-01	1	1.089	0.001
0.2	0.111	1.75E-01	1	1.117	0.002
0.3	0.097	2.41E-01	1	1.102	0.002
0.4	0.097	2.42E-01	1	1.102	0.002
0.5	0.101	2.24E-01	1	1.106	0.002
1	0.115	1.64E-01	1	1.122	0.002

Note: Covariates included were eight PCs and sex. Multiple testing correction was applied for 72 tests

Table S4
Correlation between ADHD PRS in the ADHD/ASD sample.
Several thresholds and CBCL syndrome scales.

ADHD PRS treshold	Anxious Depressed	Withdrawn Depressed	Somatic complaints	Social problems
0.01	-0.07 (0.881)	-0.069 (0.11)	-0.016 (0.708)	-0.068 (0.119)
0.05	0.032 (0.461)	-0.029 (0.508)	0.031 (0.478)	-0.042 (0.335)
0.1	0.028 (0.519)	-0.075 (0.082)	0.014 (0.749)	-0.034 (0.439)
0.2	0.053 (0.222)	-0.5 (0.248)	0.025 (0.573)	0.011 (0.798)
0.3	0.053 (0.229)	-0.037 (0.389)	0.036 (0.409)	0.015 (0.722)
0.4	0.061 (0.165)	-0.035 (0.416)	0.03 (0.487)	0.01 (0.825)
0.5	0.063 (0.149)	-0.032 (0.466)	0.036 (0.409)	0.019 (0.661)
1	0.075 (0.084)	-0.023 (0.589)	0.047 (0.285)	0.03 (0.488)

ADHD PRS treshold	Thought problems	Attention problems	Rule breaking behavior	aggressive behavior
0.01	-0.019 (0.682)	0.04 (0.362)	-0.017 (0.7)	-0.041 (0.345)
0.05	0.037 (0.412)	0.052 (0.236)	0.001 (0.988)	0.023 (0.602)
0.1	0.058 (0.202)	0.064 (0.144)	0.042 (0.336)	0.069 (0.116)
0.2	0.054 (0.233)	0.085 (0.052)	0.056 (0.193)	0.093 (0.033*)
0.3	0.051 (0.258)	0.094 (0.03*)	0.062 (0.15)	0.091 (0.037*)
0.4	0.056 (0.219)	0.091 (0.037*)	0.074 (0.086)	0.107 (0.014*)
0.5	0.056 (0.221)	0.087 (0.045*)	0.068 (0.119)	0.109 (0.012*)
1	0.059 (0.192)	0.088 (0.044*)	0.079 (0.07*)	0.115 (0.008*)

in brackets: P values: no multiple testing correction
** correlated significantly at the .05 level*
Highlighted: sig. correlated syndrome scales and optimal PRS threshold

Table S5

Mean scores and standard deviations of the CBCL syndrome scales for the ADHD/ASD sample. Sample sizes differ slightly per scale depending on information provided by the parents

CBCL syndrome scale	Mean	SD	N
Anxious Depressed	6.65	4.74	519
Withdrawn Depressed	4.71	3.38	527
Somatic Complaints	2.78	2.92	522
Social Problems	7.29	3.89	523
Thought Problems	6.34	4.30	480
Attention Problems	10.56	3.58	522
Rule Breaking Behavior	3.92	3.14	526
Aggressive Behavior	12.21	7.19	519

5

Gene-set analysis shows association between FMRP targets and autism spectrum disorder

Jansen A, Dieleman GC, Smit AB, Verhage M, Verhulst FC, Polderman TJC, Posthuma D.
Eur J Hum Genet. 2017 Jun;25(7):863-868

Abstract

Autism spectrum disorder (ASD) is a heterogeneous group of disorders characterized by problems with social interaction, communication, and repetitive and restricted behavior. Despite its high heritability and the substantial progress made in elucidating genetic associations, the corresponding biological mechanisms are largely unknown. Our objective is to investigate the contribution of common genetic variation to biological pathways functionally involved in ASD. We conducted gene-set analyses to identify ASD-associated functional biological pathways using the statistical tools MAGMA and INRICH. Gene-set selection was based on previously reported associations with psychiatric disorders and resulted in testing of specific synaptic and glial sets, a glutamate pathway gene-set, mitochondrial gene-sets and gene-sets consisting of fragile X mental retardation protein (FMRP) targets. In total 32 gene-sets were tested. We used Psychiatric Genomics Consortium genome-wide association studies summary statistics of ASD. The study is based on the largest ASD sample to date (N=5305). We found one significantly associated gene-set consisting of FMRP-targeting transcripts (MAGMA: p corr.=0.014, INRICH: p corr.=0.031; all competitive P-values). The results indicate the involvement of FMRP-targeted transcripts in ASD in common genetic variation. This novel finding is in line with the literature as FMRP has been linked to fragile X syndrome, ASD and cognitive development in whole-exome sequencing and copy number variant studies. This gene-set has also been linked to Schizophrenia suggesting that FMRP-targeted transcripts might be involved in a general mechanism with shared genetic etiology between psychiatric disorders.

Introduction

The heterogeneous manifestation of autism spectrum disorders (ASD) consists of several characteristic features including markedly abnormal social interaction, impaired communication abilities, and repetitive and restricted patterns of behavior and interests. Symptoms vary in severity per case and can occur with or without intellectual or language impairment (American Psychiatric Association, 2013). Heritability is high, with an estimate of 60% (Polderman et.al., 2015) which has led to extensive research into the genetic variants underlying ASD.

There have been several genome-wide studies including genome wide association (GWA) and linkage studies, which reported promising associations (Hussman et. al., 2011; Ma et. al., 2009; Wang et. al., 2009). However, these associations explained only a small fraction of the genetic risk to ASD, showed little replicable results with P-values between 10^{-4} and 10^{-8}, and illustrated that most of the GWA studies up to date lack power to reliably determine a role for common variants in ASD (Ma et. al., 2009; Wang et. al., 2009; Anney et. al., 2010; Anney et. al., 2012; Connolly et. al., 2013; Ronald et. al., 2010; Sullivan et. al., 2012).

The examination of copy number variants (CNVs) in ASD patients has been more successful by revealing enrichment of genes important for several, mostly synaptic, functions (Chow et. al., 2012; Gai et. al., 2012; Pinto et. al., 2010; Sanders et. al., 2015). However, causal CNVs occur in 5–10% of ASD cases and although they have a large effect on the liability to ASD, they are very rare, generally not specific to ASD, and cover large genomic areas including multiple genes (Sanders et. al., 2012).

The most exciting genetic discoveries for ASD have been reported based on whole-exome sequencing (WES) data, in which several rare de novo variants in a diversity of genes have been linked to ASD (Sanders et. al., 2012; De Rubeis et. al., 2014; Iossifov et. al., 2012; Neale et. al., 2012; O'Roak et. al., 2012). These WES findings implicated a role for genes involved in chromatin remodeling (De Rubeis et. al., 2014; Iossifov et. al., 2012; O'Roak et. al., 2012; Iossifov et. al., 2014), synaptic formation (De Rubeis et. al., 2014), transcriptional regulation (De Rubeis et. al., 2014), and FMRP-associated genes (De Rubeis et. al., 2014; Iossifov et. al., 2012; Iossifov 2014) (see for reviews Sullivan et. al., 2012; Ronemus et. al., 2014).

The emerging picture is that ASD, like other psychiatric traits, is highly polygenic, and likely influenced by a mix of rare and common variants, of which functional implications still have to be determined (Sullivan et. al., 2012; De Rubeis et. al., 2015). The identification of pathways with genes dysfunctional in ASD may increase by investigating the combined effect of multiple variants, using gene-set analysis. This, because it evaluates the joint effect of multiple genetic variants grouped according to biological or cellular function, thereby decreasing the multiple testing problem and increasing effect size (Mooney et. al., 2015; Ruano et. al., 2010; Sullivan et. al. 2015; Torkamani et. al., 2008).

In the current study, our objective is to investigate the contribution of common genetic variation to biological pathways functionally involved in ASD. To this end, we investigated whether the joint effect of common genetic variants grouped into a priori selected gene-sets is associated with ASD. Because of our hypothesis driven top down approach, we performed

direct testing on all single nucleotide polymorphisms (SNPs) in a particular gene-set. This, instead of a functional enrichment analysis of top SNPs that does not require a priori hypotheses and uses top SNPs to define possible associated pathways based on functional enrichment.

We selected five categories of gene-sets, and limited ourselves to expert-curated gene-sets, resulting in testing 32 gene-sets. We included 19 curated synaptic (Ruano et. al., 2010) (category one) and three glial gene-sets (oligodendrocytes, astrocytes, and oligodendrocytes and astrocytes combined) (Duncan et. al., 2014) (category two), which have cell type-specific functions. Synaptic genes have been implicated to be among the top genes harboring variants associated with ASD (Sanders et. al., 2015; De Rubeis et. al., 2014) and other psychiatric disorders (Lips et. al., 2012). The glial sets are an exploratory approach to provide general insights and starting points for more specific hypothesis formation although previous research has pointed in the direction of a role for astrocytes (Zeidán-Chuliá et. al., 2014) and oligodendrocytes (Zeidán-Chuliá et. al., 2015) in ASD, and for oligodendrocytes in schizophrenia (SCZ) (Chavarria-Siles et. al., 2016). The third category we selected consists of genes which gene-transcripts are targeted by fragile X mental retardation protein (FMRP). FMRP is an RNA binding protein expressed in the brain coded by the FMRI gene located at Xq27.3 that has been linked to fragile X syndrome (FXS) (Pinto et. al., 2014) and evidence for an association with ASD is accumulating due to WES (De Rubeis et. al., 2014; Iossifov et. al. 2012; Iossifov et. al., 2014) and CNV studies (Darnell et. al., 2011), yet it has not been confirmed using common variants from genome-wide association studies (GWAS). Darnell et al. (2011) and Ascano et al. (2012) have provided insights into the biological underpinnings of FXS and ASD using human tissue and mouse models on FMRP and the RNA this binds to. Their research resulted in three gene-sets with FMRP target transcripts (Darnell gene-set, Ascano gene-set, Darnell and Ascano overlap gene-set). We aimed to test whether these sets of FMRP-targeted genes are associated with ASD in a large human sample not enriched with the FMR1 variant. Our fourth gene-set category is a glutamate pathway. With glutamate being the most important excitatory agent in the brain, glutamate and its receptors have been suggested to have a role in psychiatric diseases including ASD (Duncan et. al., 2014; Rojas et. al., 2014). The fifth category (six gene-sets) is based on mitochondrial genes (Duncan et. al., 2014). Mitochondria provide energy for the cell and with the brain being the organ using most of the energy, even a small reduction in energy production can result in impaired brain processes in the synapse (Duncan et. al., 2014). Mitochondrial dysfunction has cautiously been associated with psychiatric diseases, including ASD, based on abnormal mitochondrial biomarker values and high prevalence of mitochondrial diseases in ASD patients compared to a healthy subpopulation (Manji et. al., 2012; Rossignol et. al., 2012).

In sum, our main goal is to directly test predefined sets of genes for their association with ASD. These gene-sets were selected because of previous associations with psychiatric diseases and test involvement of (1) synaptic processes, (2) glia cells, (3) FMRP, (4) glutamate and (5) mitochondrial involvement. The underlying hypothesis is that the polygenic nature of ASD shows convergence of genetic effects in biologically meaningful sets of genes.

Materials and methods

Sample

We used the publicly available GWAS summary statistics (PGC.ASD.euro. all.25Mar2015.txt.gz) downloaded from http://www.med.unc.edu/pgc/ results-and-downloads on 21 May 2015. More information on the sample can be found on the mentioned Psychiatric Genomics Consortium (PGC) website. Briefly, in their original study PGC used five cohorts: the Geschwind Autism Center of Excellence (ACE), the Autism Genome Project (AGP), the Autism Genetic Resource Exchange (AGRE), the NIMH Repository, the Montreal/Boston Collection (MONBOS), and the Simons Simplex Collection (SSC); see Table 1). The total number of ASD probands in this sample is 5305, and of pseudocontrols this is 5303. In a pseudocontrol setting, instead of a regular control group the nontransmitted parental allele is used as the control. All participants were of European descend. For the current gene-set analyses summary statistics (ie, P-values per SNP) of this PGC study were used.

Generation of gene-sets

We used 32 publicly available expert-curated gene-sets that we assigned to five distinct categories. The 19 synaptic gene-sets were published in previous studies (Sanders et. al., 2015; Ruano et. al., 2010; Lips et. al. 2012) in which they were defined based on assignment of subcellular function as determined by synaptic protein purification experiments and data mining for synaptic genes and gene ontology. Synaptic genes were subdivided into 19 functional groups (N genes 1047). The glial (146 genes), oligodendrocyte (52 genes), astrocyte (42 genes) and mitochondrial (six gene-sets, N genes 132) gene-sets were created and described by Duncan et al., (2014) who conducted a database search in the gene ontology database and REACTOME. They supplemented the identified genes with genes found via an in-depth literature study. In addition, we included three gene-sets consisting of FMRP targeting genes (N genes 1809) as defined by Darnell et al. (2011) and Ascano et al. (2012) with sequencing methods. All gene-sets as used in the present study are shown in Supplementary Tables S1 and S2.

MAGMA and INRICH gene-set analyses

Gene-set analyses can consist of self-contained testing and competitive testing. In a self-contained test the alternative hypothesis states that a gene-set is associated with the trait against the null hypothesis of no association, whereas in competitive testing the alternative hypothesis is that the gene-set is significantly stronger associated with the trait than genes not included in the gene-set.

We performed our analysis using two methods, MAGMA and INRICH, both providing competitive test results (de Leeuw et. al., 2015).

MAGMA (v1.01, http://ctglab.nl/software/magma) is a tool to perform gene and gene-set analysis which is distinguishable from other methods like INRICH, ALIGATOR, MAGENTA by having more statistical power, being less affected by linkage disequilibrium (LD; a SNP is in LD with another SNP when their specific alleles occur more often together than expected by chance, implicating that the independent association assumption is violated and you can predict one of the specific alleles with high certainty dependent on the other known allele.) and multi-marker associations due to its multiple regression approach and being computationally less demanding as it does not use a permutation based approach de Leeuw et. al., 2015; de Leeuw et. al., 2016). A significant hit in MAGMA indicates that multiple genes in the gene-set are associated. Although the SNPs included in these genes can have relatively high P values, only together they are responsible for a positive signal in a gene-set. The 1000 genomes European panel (reference file) and NCBI 37.3 (gene location) (downloaded from http://ctglab.nl/software/magma) were used for SNP annotation to genes.

INRICH (Lee et. al., 2012) is a permutation based GWA analysis tool that tests whether functionally related genes compiled in gene-sets show a stronger association with a phenotype than expected by chance. A significant hit in INRICH can occur with only a few highly associated SNPs in a gene-set. INRICH can be downloaded from http://atgu.mgh.harvard.edu/inrich/ downloads.html. Utilizing Plink (Purcell et. al., 2007) we computed several intervals for our analyses using different parameters for the LD clumping procedure. We applied the default INRICH values of 5000 first pass permutations, and 1000 s pass permutations.

Clumping is a method to reduce the amount of double signal in a data set due to LD. We assigned SNPs that are significant between a certain threshold to the same clump if they have an r^2 of 0.5 and are not yet assigned to another clump. For this clumping parameter we used several SNP P-value significance thresholds: 0.0001 and 0.01 (both fixed P-value thresholds) and 0.00896288 (1% cut-off P-value, computed in R studio v3.0.2, Boston, MA, USA) and 0.0008106764 (0.1% cut-off P-value, computed in R studio v3.0.2). These different thresholds influence which and how many SNPs are assigned to a clump. A stricter P-value cut-off results in a clump with less SNPs. Again, we used NCBI 37.1 for gene location.

Table 1 Overview of the cohorts that were included in the initial analysis by the PGC

Cohort	*Type of cohort*	*N cases/controls*	*Country*	*Ancestry*	*Diagnostic instrument*
UCLA Autism Center of Excellence (ACE)	Parent-Parent-Proband Trios	391/391	USA	Caucasian	ADI-R and/or ADOS
Autism Genome Project (AGP)	Parent-Parent-Proband Trios	2272/2272	USA, Europe	Caucasian	ADI-R and ADOS
Autism Genetic Resource Exchange (AGRE)	Multiplex families	974/974	USA	Caucasian	ADI-R and ADOS
NIMH Repository, the Montreal/Boston Collection (MONBOS)	Simplex and multiplex families	1396/1396	USA/Canada	Caucasian	ADI-R (NIMH). Autism Screening Questionnaire, ADI-R ADOS (MONBOS)
Simons Simplex Collection (SSC)	Parent-Parent-Proband Trios	2231/2231	USA	Caucasian	ADI-R and ADOS

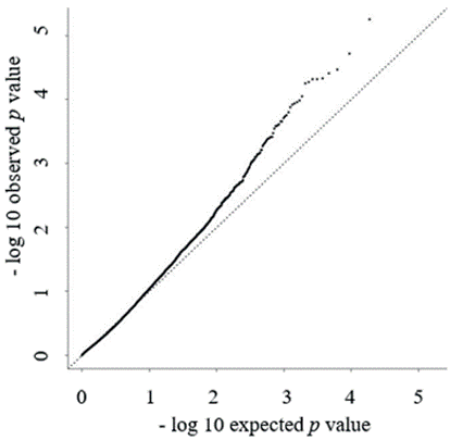
Abbreviations: ADI-R, Autism Diagnostic Interview-Revised; ADOS, Autism Diagnostic Observation Schedule; PGC, Psychiatric Genomics Consortium.

The summary statistics, containing the P-values per SNP of this PGC analysis, were used for the gene-set analysis.

Statistical testing

We applied multiple testing following our hypothesis driven approach: per hypothesis we multiplied the P-value by the number of gene-sets that was tested for that particular hypothesis. Although this correction is not as strict as the Bonferroni correction, it provides sufficient correction as we constructed independent hypotheses generating independent results per hypothesis. This allows for multiple testing correction per hypothesis instead of a more stringent multiple testing correction over all hypotheses. A consensus has not yet been reached so different studies parameterize and evaluate results differently (Sullivan et. al., 2015).

1a. All genes that have SNPs assigned



1b. All genes in the FMRP gene-set

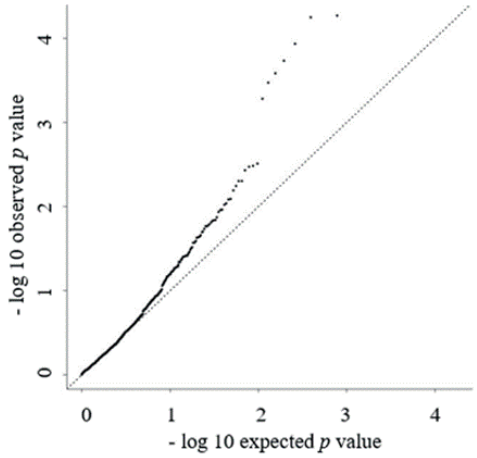
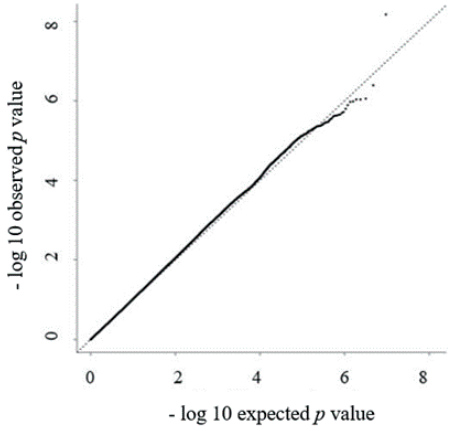


Figure 1

Visualization of the polygenic pattern at gene level. (a) QQ plot of P-values of all genes that have SNPs from the PGC data set assigned to them regardless of inclusion in a gene-set. This plot compares the observed P-values to the expected P-values at gene level. The non-linear pattern (the deviation from the diagonal) visualizes the polygenic signal in the genes. (b) QQ plot of the genes in the FMRP (Darnell) gene-set. The earlier lift off and larger deviation from the diagonal compared to (a) illustrates the signal in a is driven in part by genes in the FMRP gene-set by Darnell.

2a. All SNPs in the PGC dataset



2b. All SNPs in the FMRP gene-set

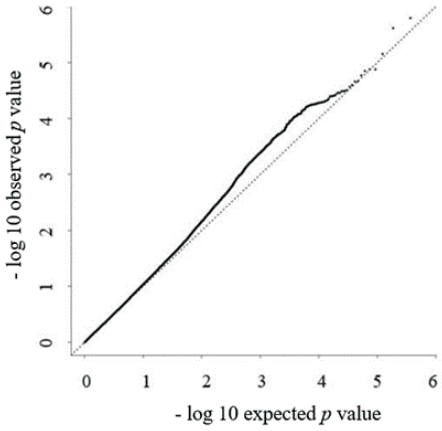


Figure 2 Visualization of the polygenic pattern at SNP level. (a) QQ plot from all SNPs in the PGC data set. This plot compares the observed P-values to the expected P-values at SNP level. The non-linear pattern (the deviation from the diagonal) visualizes the polygenic signal in the SNPs. (b) QQ plot from the SNPs in the FMRP (Darnell) gene-set. The earlier lift off and larger deviation from the diagonal compared to (a) illustrates the signal in a is driven partly by SNPs in the FMRP gene-set by Darnell

Results

MAGMA

Competitive gene-set analyses resulted in a statistically significant, multiple testing corrected association with the FMRP target gene-set by Darnell et al. (2011) (P=0.014). None of the other gene-sets showed a statistically significant association, after multiple testing correction (for all results see Supplementary Table S3. Also, see Figure 1 (gene level) and Figure 2 (SNP level). These figures show the polygenic signal (Figures 1a and 2a) and the role of the FMRP gene-set in this signal (Figures 1b and 2b).

Gene-based tests in MAGMA for the genes included in the significant gene-set resulted in multiple significantly associated genes, indicating the results were not driven by a few highly associated genes and illustrating the value of testing groups of genes together (Supplementary Table S4).

INRICH

We performed this method with several parameter settings, as described in the method section. Overall, different parameter settings regarding clumping thresholds did not meaningfully change the results. However, the FMRP targets showed again a significant association with ASD after multiple testing correction, P=0.031, for the clumping SNP P-value significance threshold: 0.0001. For all results see Supplementary Table S5.

Discussion

Previous research efforts have clearly shown that ASD is a highly polygenic disorder with reported heritability around 60% (Polderman et. al., 2015). Some genetic variants have consistently been identified but functional implications of current genetic findings are as yet modest (Sanders et. al., 2015; De Rubeis et. al., 2015; Sullivan et. al. 2015). In the present study, we tested whether the genetic variants of small effect on ASD tend to cluster in selected functional sets of genes. We tested expert-curated functional gene-sets that have been constructed previously in the context of their putative role in psychiatric disorders. Our multiple testing corrections were not as strict as a Bonferroni correction. If we had applied this correction no gene-set would be significantly associated. Still, we believe a sufficient correction was applied as we constructed, based on previous findings, independent hypotheses that as such generated independent results per hypothesis. Also, we applied competitive tests instead of the – far less stringent – self-contained tests, as usually used in this type of analyses. Moreover, the gene-sets are not independent, and thus a Bonferroni based on all gene-sets for all tested hypotheses is likely overly conservative.

The FRMP gene-set by Darnell was found to be significantly associated with the risk for ASD confirming previously reported associations in WES and CNV studies (Iossifov et. al., 2012; Pinto et. al., 2014; Darnell et. al., 2011; Ascano et. al., 2012). This gene-set of FMRP-targeting

proteins was constructed by Darnell et al. (2011) who identified FMRP interactions with mRNA in the mouse brain by means of high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation. Their study showed a connection between loss of function FMRP and ASD-associated symptoms in FXS and ASD patients. FMRP is important for translation of hundreds of neuronal mRNA's and its loss results in morphological and physiological neuronal defects resulting in FXS-like symptoms like cognitive impairment, seizures, anxiety and hyperactivity (De Rubeis et. al., 2011; Doll et. al., 2014; Fernández et. al., 2013). The link between ASD and FXS seems intuitive as FXS is the leading form of monogenetic inherited intellectual disability with many cognitive and behavioral symptoms which are also manifest in ASD (De Rubeis et. al., 2011). In addition, FXS shows comorbidity with ASD, about 30% of FXS patients are diagnosed with ASD, whereas 1 to 2% of ASD patients show FXS comorbidity (De Rubeis et. al., 2011). As an FXS diagnosis was an exclusion criterion our results are not likely due to inclusion of patients with this monogenetic disorder. However, in the FXS there are pre-mutations, less CGG repeats than FXS patients but more than healthy individuals, causing diseases like fragile X-associated tremor ataxia syndrome, and increasing the incidence of ASD and ADHD (Lozano et. al., 2014). Additional genetic phenotyping, including FMRP count, would be needed to ensure that samples are not enriched for FMRP pre-mutations.

A study on ASD rare variants (Purcell et. al., 2014) also reported associations with the Darnell FMRP gene-set, yet they found no evidence for overlap at the individual gene level. An association between this gene-set and SCZ has also been reported (Purcell et. al., 2014; Fromer et. al., 2014; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Szatkiewicz et. al. 2014). Two SCZ CNV studies (Fromer et. al., 2014; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Szatkiewicz et. al. 2014) showed genetic overlap between ASD at gene-set and individual gene level. As a whole, these results might point in the direction of a common biological basis between ASD and SCZ making it of interest to look further into this possible overlap.

A general concern in psychiatric disorders is phenotypic heterogeneity and in ASD heterogeneity in intellectual disability (ID) (Cervantes et. al., 2015) , is one of these concerns. As ID was not an exclusion criterion in our study, we cannot ensure with 100% certainty that our results are not partly driven by ID. Attempts have been made (Chaste et. al., 2015) to stratify samples into low (IQ<60) and high IQ (IQ>60) but a downside is that subsequent decreasing sample sizes reduce statistical power. Unfortunately, we only had access to GWAS summary statistics and not to IQ scores and raw genotypes of participants we could not perform such analyses in our current study.

A final point to address regarding our FMRP hypothesis is that, out of three FMRP gene-sets, only the Darnell gene-set remained significant after multiple testing correction. The Ascano and Ascano autism overlap gene-sets did not generate significant results. Possible explanations for these findings are the different ways the gene-sets were constructed. Darnell et. al. (2011) identified FMRP interactions within the mouse brain by means of high-throughput RNA sequencing and follow-up analysis, whereas Ascano et al.(2012) examined FMR1 family protein binding sites to identify and rank FMRP targets in human embryonic

kidney cells. These methods may have resulted in two different subsets of FMRP targets showing little overlap and expressing different biological properties. As a final remark, both gene-sets are large compared to all other tested sets. As the Ascano gene-set is larger than the Darnell gene-set, and given that effects of larger gene-sets are generally more easily to detect (de Leeuw et. al., 2016), it is unlikely that gene-set size explains the association of the Darnell set. In addition, both MAGMA and INRICH have a correct type I error rate which is independent of gene-set size (de Leeuw et. al., 2016).

Our results do not support a role of the glutamate pathway, mitochondrial, synaptic or glial pathway in ASD, suggesting it is unlikely that there are large effects of these pathways on the risk of ASD. However, gene-set definitions are dynamic, and with increased precision in pathway annotation, these results may change (Sullivan et. al., 2015).

Taken together, the current results provide evidence for a role of FMRP-targeted transcripts in ASD. As FMRP is associated with several psychiatric conditions a more thorough exploration of genes in this gene-set and their association with different psychiatric disorders might provide useful information on an underlying shared genetic etiology between several disorders.

To conclude, we performed a gene-set analysis aiming to find common variation clustered in functional pathways associated with ASD. Our significant hit in common genetic variants is an FMRP targeting gene-set that has been associated with ASD in rare variation and other psychiatric illnesses. These findings can point in the direction of a more general mechanism underlying psychiatric disorders making cross disorder research an important future component of the scientific repertoire.

References

Anney R, Klei L, Pinto D et al: A genome-wide scan for common alleles affecting risk for autism. Hum Mol Genet 2010; 19: 4072–4082.

Anney R, Klei L, Pinto D et al: Individual common variants exert weak effects on the risk for autism spectrum disorderspi. Hum Mol Genet 2012; 21: 4781–4792.

American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders. Washington, DC, USA: American Psychiatric Association, 2013, Vol 5.

Ascano M, Mukherjee N, Bandaru P et al: FMR1 targets distinct mRNA sequence elements to regulate protein expression. Nature 2012; 492: 382–386.

Cervantes PE, Matson JL: Comorbid symptomology in adults with autism spectrum disorder and intellectual disability. J Autism Dev Disord 2015; 45: 3961–3970.

Chaste P, Klei L, Sanders SJ 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 2015; 77: 775–784.

Chavarria-Siles I, White T, de Leeuw C et al: Myelination-related genes are associated with decreased white matter integrity in schizophrenia. Eur J Hum Genet 2016; 24: 381–386.

Chow ML, Pramparo T, Winn ME et al: Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. PLoS Genet 2012; 8: e1002592.

Connolly JJ, Glessner JT, Hakonarson H: A genome-wide association study of autism incorporating autism diagnostic interview-revised, autism diagnostic observation schedule, and social responsiveness scale. Child Dev 2013; 84: 17–33.

Darnell JC, Van Driesche SJ, Zhang C et al: FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell 2011; 146: 247–261.

de Leeuw CA, Mooij JM, Heskes T, Posthuma D: MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 2015; 11: e1004219.

de Leeuw CA, Neale BM, Heskes T, Posthuma D: The statistical properties of gene-set analysis. Nat Rev Genet 2016; 17: 353–364.

De Rubeis S, Bagni C: Regulation of molecular pathways in the fragile X syndrome: insights into autism spectrum disorders. J Neurodev Disord 2011; 3: 257–269.

De Rubeis S, He X, Goldberg AP et al: Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 2014; 515: 209–215.

De Rubeis S, Buxbaum JD: Recent advances in the genetics of autism spectrum disorder. Curr Neurol Neurosci Rep 2015; 15: 36.

Doll CA, Broadie K: Impaired activity-dependent neural circuit assembly and refinement in autism spectrum disorder genetic models. Front Cell Neurosc 2014; 8: 30.

Duncan LE, Holmans PA, Lee PH et al: Pathway analyses implicate glial cells in schizophrenia. PLoS ONE 2014; 9: e89441.

Fernández E, Rajan N, Bagni C: The FMRP regulon: from targets to disease convergence. Front Neurosci 2013; 7: 191.

Fromer M, Pocklington AJ, Kavanagh DH et al: De novo mutations in schizophrenia implicate synaptic networks. Nature 2014; 506: 179–184.

Gai X, Xie HM, Perin JC et al: Rare structural variation of synapse and neurotransmission genes in autism. Mol Psychiatry 2012; 17: 402–411.

Hussman JP, Chung R-H, Griswold AJ et al: A noise-reduction GWAS analysis implicates altered regulation of neurite outgrowth and guidance in autism. Mol Autism 2011; 2: 1.

Iossifov I, Ronemus M, Levy D et al: De novo gene disruptions in children on the autistic spectrum. Neuron 2012; 74: 285–299.

Iossifov I, O'Roak BJ, Sanders SJ et al: The contribution of de novo coding mutations to autism spectrum disorder. Nature 2014; 515: 216–221.

Lee PH, O'Dushlaine C, Thomas B, Purcell SM: INRICH: interval-based enrichment analysis for genome-wide association studies. Bioinformatics 2012; 28: 1797–1799.

Lips ES, Cornelisse LN, Toonen RF et al: Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. Mol Psychiatry 2012; 17: 996–1006.

Lozano R, Rosero CA, Hagerman RJ: Fragile X spectrum disorders. Intractable Rare Dis Res 2014; 3: 134–146.

Ma D, Salyakina D, Jaworski JM et al: A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. Ann Hum Genet 2009; 73: 263–273.

Manji H, Kato T, Di Prospero NA et al: Impaired mitochondrial function in psychiatric disorders. Nat Rev Neurosci 2012; 13: 293–307.

Mooney MA, Wilmot B: Gene set analysis: a step-by-step guide. Am J Med Genet Part B Neuropsychiatr Genet 2015; 168: 517–527.

Neale BM, Kou Y, Liu L et al: Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 2012; 485: 242–245.

O'Roak BJ, Vives L, Girirajan S et al: Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 2012; 485: 246–250.

Pinto D, Pagnamenta AT, Klei L et al: Functional impact of global rare copy number variation in autism spectrum disorders. Nature 2010; 466: 368–372.

Pinto D, Delaby E, Merico D et al: Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. Am J Hum Genet 2014; 94: 677–694.

Polderman TJC, Benyamin B, de Leeuw CA et al: Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat Genet 2015; 47: 702–709.

Purcell S, Neale B, Todd-Brown K et al: PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559–575.

Purcell SM, Moran JL, Fromer M et al: A polygenic burden of rare disruptive mutations in schizophrenia. Nature 2014; 506: 185–190.

Rojas DC: The role of glutamate and its receptors in autism and the use of glutamate receptor antagonists in treatment. J Neural Transm 2014; 121: 891–905.

Ronald A, Butcher LM, Docherty S et al: A genome-wide association study of social and non-social autistic-like traits in the general population using pooled DNA, 500 K SNP microarrays and both community and diagnosed autism replication samples. Behav Genet 2010; 40: 31–45.

Ronemus M, Iossifov I, Levy D, Wigler M: The role of de novo mutations in the genetics of autism spectrum disorders. Nat Rev Genet 2014; 15: 133–141.

Rossignol DA, Frye RE: Mitochondrial dysfunction in autism spectrum disorders: a systematic review and meta-analysis. Mol Psychiatry 2012; 17: 290–314.

Ruano D, Abecasis GR, Glaser B et al: Functional gene group analysis reveals a role of synaptic heterotrimeric G proteins in cognitive ability. Am J Hum Genet 2010; 86: 113–125.

Sanders SJ, Murtha MT, Gupta AR et al: De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 2012; 485: 237–241.

Sanders SJ, He X, Willsey AJ et al: Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron 2015; 87: 1215–1233.

Schizophrenia Working Group of the Psychiatric Genomics Consortium: Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014; 511: 421–427.

Sullivan PF, Daly MJ, O'Donovan M: Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat Rev Genet 2012; 13: 537–551.

Sullivan PF, Posthuma D: Biological pathways and networks implicated in psychiatric disorders. Curr Opin Behav Sci 2015; 58–68.

Szatkiewicz JP, O'Dushlaine C, Chen G et al: Copy number variation in schizophrenia in Sweden. Mol Psychiatry 2014; 19: 762–773.

Torkamani A, Topol EJ, Schork NJ: Pathway analysis of seven common diseases assessed by genome-wide association. Genomics 2008; 92: 265–272.

Verkerk AJ, Pieretti M, Sutcliffe JS et al: Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 1991; 65: 905–914.

Wang K, Zhang H, Ma D et al: Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 2009; 459: 528–533.

Zeidán-Chuliá F, Salmina AB, Malinovskaya NA, Noda M, Verkhratsky A, Moreira JCF: The glial perspective of autism spectrum disorders. Neurosci Biobehav Rev 2014; 38: 160–172.

Zeidán-Chuliá F, de Oliveira BHN, Casanova MF et al: Up-regulation of oligodendrocyte lineage markers in the cerebellum of autistic patients: evidence from network analysis of gene expression. Mol Neurobiol 2015; 53: 4019–4025.

Chapter 5 Supplementary Material

Table of Contents

Supplemental Table 1: Expert curated functional gene-sets	
GENE_SET	**N genes**
Synaptic: Cell adhesion and trans-synaptic signaling	81
Synaptic: Cell metabolism	57
Synaptic: Endocytosis	26
Synaptic: Excitability	59
Synaptic: Exocytosis	87
Synaptic: GPCR signaling	41
Synaptic: G-protein relay	27
Synaptic: intracellular signal transduction	150
Synaptic: Intracellular trafficking	80
Synaptic: Ion balance/transport	43
Synaptic: Ligand-gated ion channel signaling	36
Synaptic: Neurotransmitter metabolism	29
Synaptic: Peptide/neurotrophin signals	28
Synaptic: Protein cluster	47
Synaptic: RPSFB	71
Synaptic: Structural plasticity	98
Synaptic: Tyrosine kinase signaling	7
Synaptic: Unknown	61
Synaptic: Sanders FDR 0.01 large synaptic	19
FMRP targets (Ascano)	936
FMRP targets (Ascano Autism overlap)	93
FMRP targets (Darnell)	780
Glia astrocytes Duncan	42
Glia oligodendrocytes Duncan	52
Glia Duncan	146
Glutamate Duncan	156
Mitochondria: Duncan	74
Mitochondria: Crista Duncan	6
Mitochondria: Distribution Duncan	7
Mitochondria: Fission Duncan	12
Mitochondria: Fission_plus Duncan	24
Mitochondria: Fusion Duncan	9

Supplemental table 3: MAGMA results of the gene-sets analyses				
GENE_SET	**N genes**	**P_{comp}**	**P_{corr}**	**P_{self}**
Synaptic: Cell adhesion and trans-synaptic signaling	81	0.18192	0.9796	0.10887
Synaptic: Cellmetabolism	57	0.2548	0.9967	0.76364
Synaptic: Endocytosis	26	0.98686	1	0.99593
Synaptic: Excitability	58	0.41285	1	0.53495
Synaptic: Exocytosis	87	0.64415	1	0.7095
Synaptic: GPCRsignaling	41	0.4402	1	0.6031
Synaptic: G-proteinRelay	27	0.88081	1	0.85643
Synaptic: intracellularSignalTra...	150	0.020727	0.3307	0.028119
Synaptic: IntracellularTrafficking	79	0.65268	1	0.71324
Synaptic: IonBalance/transport	43	0.1892	0.9819	0.1258
Synaptic: Ligand-gated ion channel signaling	36	0.12412	0.9252	0.16304
Synaptic: Neurotransmitter metabolism	29	0.037016	0.5148	0.005359
Synaptic: Peptide/neurotrophin signals	28	0.22818	0.9921	0.057422
Synaptic: ProteinCluster	47	0.4644	1	0.32378
Synaptic: RPSFB	70	0.11831	0.9157	0.1212
Synaptic: StructuralPlasticity	96	0.016129	0.2649	0.1263
Synaptic: TyrosineKinaseSignaling	7	0.38377	0.9999	0.64281
Synaptic: Unknown	61	0.58128	1	0.74641
Synaptic: Sanders FDR 0.01 large synaptic	18	0.13436	0.9398	0.4698
GENE_SET	**N genes**	**P_{comp}**	**P_{corr}**	**P_{self}**
FMRP targets (Ascano)	936	0.65741	0.95070	0.39329
FMRP targets (Ascano Autism overlap)	93	0.51565	0.87120	0.56110
FMRP targets (Darnell)	780	0.00483	0.01380	0.08314
GENE_SET	**N genes**	**P_{comp}**	**P_{corr}**	**P_{self}**
Glia astrocytes Duncan	41	0.39517	0.64710	0.55910
Glia oligodendrocytes Duncan	52	0.62103	0.85580	0.71794
Glia Duncan	141	0.57262	0.82130	0.92142
GENE_SET	**N genes**	**P_{comp}**	**P_{corr}**	**P_{self}**
Glutamate Duncan	156	0.62348	0.62348	0.86420
GENE_SET	**N genes**	**P_{comp}**	**P_{corr}**	**P_{self}**
Mitochondria Duncan	70	0.77749	0.99750	0.31608
Mitochondria: Crista Duncan	6	0.85101	0.99940	0.85093
Mitochondria: Distribution Duncan	7	0.93720	1	0.91592
Mitochondria: Fission	11	0.99575	1	0.22899
Mitochondria: Fission plus	23	0.98723	1	0.33337
Mitochondria: Fusion	9	0.98366	1	0.17705

P_{SELF}= *self-contained P-value*, P_{COMP} = *competitive P-value*

P_{CORR} = *multiple testing corrected competitive P-value (comp P*N func. gene-sets per hypothesis)*

6

Summary and discussion

Summary and discussion

6.1 Summary

Many psychiatric disorders are already present, or nascent, in childhood and adolescence and pose a great burden on these children and their families (Costello et al., 2005). This thesis focuses on the common genetic variation that is associated with heritable neurodevelopmental and psychiatric disorders (Demontis & Walters, 2017; Grove & et.al., 2017; Polderman et al., 2015) occurring during childhood and adolescence such as ASD and ADHD. The underlying assumption is that complex disorders are influenced by genetic and environmental variation and by their interplay, called gene environment interaction and gene environment correlation. The genetic component comprises rare and common variations (Visscher et al., 2017). The focus of this thesis is on the common genetic variation (occurring >1% in the general population). Recent GWAS discoveries show that each associated common genetic variant has a very small effect size (Wray et al., 2020). We assume the effects to be cumulative, located in genetic and molecular networks which become impaired if enough local SNP mutations occur (Schadt, 2009). Due to the small effect size and assumed accumulation in specific networks it makes sense to not only analyze SNPs by themselves but additionally test the larger unit to which SNPs and genes (might) belong by means of gene-set analysis (de Leeuw et al., 2015; Duncan et al., 2014; Goudriaan et al., 2013). This approach is capable of identifying if common genetic variants grouped in a priori selected gene-sets are significantly associated with disorders, compared to genes not in the gene-set. A predictive tool based on GWAS discoveries is the PRS analysis. PRS involves calculating a weighted sum of SNPs consolidating the effect in a single measure (Wray et al., 2014, 2020) to be used in further analyses. Both gene-set analysis and PRS analysis are important tools used in this thesis.

Using the clinical child and adolescent sample ('Inside-Out' described in Ch. 2) and control samples (NESCOG and BePS, described in Ch. 3, 4 and 5 when included), this thesis aimed to provide insights on the overlap of associated SNPs between psychiatric disorders. The first main aim of this thesis is to quantify the predictive capacity of common genetic variation of a variety of traits, as captured by their PRS. I aim to predict case-control status in a child and adolescent psychiatric sample with a variety of psychiatric disorders to reveal which traits are associated with the genetic risk contributing to general psychiatric symptoms present in several psychiatric disorders (Chapter 3). The second aim was to investigate if the common genetic variation related to ASD and ADHD, as captured by their PRS, is capable of predicting case-control status in an ASD and/or ADHD sample (Chapter 4). The third aim was to test if SNPs significantly associated with ASD occur more often in gene-sets linked to SCZ, which might be genetically linked to ASD. This might be a result of a large accumulation of SNPs with a small effect size assigned to included genes, or be due to a select group of SNPs with a large effect size (Chapter 5). In Chapter 3 we quantified the predictive capacity of common genetic variation of a variety of traits as captured by their PRS. I aimed to predict diagnostic status (being diagnosed with a psychiatric disorder yes/no) in a child and adolescent psychiatric sample to reveal which PRS

contribute to the genetic risk underlying psychiatric symptoms shared between disorders. PRS of 14 traits were used to predict diagnostic status. Clinical cases, some with multiple diagnoses per participant, were compared to controls. An individual with multiple diagnoses can be included as they experience disorder specific symptoms and symptoms attributed to both disorders. The occurrence of several psychiatric disorders in one individual strengthens the assumption that there is a set of underlying common genetic variation being partly responsible for the shared symptoms. These 14 PRS were first individually tested. The traits that significantly predicted diagnostic status (being diagnosed with a psychiatric disorder yes/no) were included in a multivariable model to investigate the gain in explained variance when leveraging the genetic effects of multiple traits simultaneously. In the univariate analyses significant associations were observed between diagnostic status and the PRS of educational attainment, smoking initiation, intelligence, neuroticism, alcohol dependence, ADHD, major depression and anti-social behavior. However, these PRS are correlated to a certain extend making them not independent of each other. The PRS of educational attainment and smoking initiation showed the highest explained variance in case-control status. In the multivariable model with these eight significantly associated trait PRS and covariates, educational attainment and smoking initiation remained significant predictors. These results provide more insights into the genetic signal that is shared between childhood and adolescent psychiatric disorders. They suggest that a myriad of mental health-related traits are genetically associated with psychiatric disorders. General psychopathology in children and adolescents is potentially associated with a genetic vulnerability for low EA and SI. Hence a genetic vulnerability for low EA and SI might contribute to specific comorbidity patterns as observed between psychiatric symptoms and to the broad range of psychiatric symptoms. It is important to keep in mind we cannot draw any conclusions regarding causality as we have not tested this. The increase of 2% in explained variance when leveraging the genetic signal of multiple traits compared to the highest single variable supports a multivariable approach to optimize precision accuracy for general psychopathology. To conclude, our results indicate that a part of the genetic variance influencing a myriad of mental health-related traits also influences psychopathology. A genetic vulnerability for low EA and SI might be predictors for general psychopathology in children and adolescents which can be considered as some of the potential factors preceding the development of psychiatric symptoms. However, causality is not investigated in this thesis and may not be assumed. In addition, a genetic vulnerability for low EA and SI might contribute to specific comorbidity patterns as observed between psychiatric symptoms. Our findings can guide future studies on psychiatric comorbidity, and studies addressing the causal directions between EA, SI and general psychopathology.

In Chapter 4 we zoom in on neurodevelopmental disorders in the sample by including only ASD and ADHD cases and controls. These two disorders are highly comorbid, are genetically correlated and share various symptoms, but they are defined as distinct disorders, and also have unique psychiatric comorbidity patterns (Solberg et al., 2019). In our clinical sample we aimed to explore the association between the PRS for three neurodevelopmental disorders namely ADHD, ASD and for schizophrenia (SCZ), and an ADHD and/or ASD diagnostic status. Based on

the most recent GWAS of ADHD, ASD and SCZ, PRS for each disorder were calculated for our clinical and control sample. We tested for an association with (1) a single diagnosis of ADHD, (2) a single diagnosis of ASD, and (3) combining the two diagnoses, thus subjects with either ASD, ADHD or both. The results revealed a significant association of the ADHD PRS with ADHD diagnostic status (OR 1.6, $P = 1.39 \times 10{-}07$) and with the combined ADHD/ASD status (OR 1.36, $P = 1.211 \times 10{-}05$), but not with ASD diagnostic status (OR 1.14, $P = 1$). We did not find associations for the ASD and SCZ PRS. Concluding, the PRS of ADHD is significantly associated with the combined ADHD/ASD status but this association is likely to be primarily driven by ADHD status.

The objective in Chapter 5 was to identify gene-sets statistically associated with the common genetic variation associated with ASD. This is of interest as despite the large heritability of ASD the corresponding biological mechanisms and their modifications causing ASD are largely unknown. Showing statistical association of gene-sets can help with the formation of hypotheses regarding functional involvement of specific genes or pathways in ASD. To identify ASD associated functional biological pathways we conducted gene-set analyses. The selection of gene-sets was based on previously reported associations with psychiatric disorders but not yet tested in relation to ASD. This resulted in the testing of specific synaptic and glial sets, a glutamate pathway gene-set, mitochondrial gene-sets and gene-sets consisting of fragile X mental retardation protein (FMRP) targets (N 32 gene-sets). FMRP is expressed by the *FMR1* gene located at Xq27.3. Loss of function of this RNA binding protein expressed in the brain results in Fragile X syndrome, a disorder often co-diagnosed with ASD and intellectual disability. We based this study on the largest publicly available ASD GWAS sample at that time (*N*=5305) published by the Psychiatric Genomics Consortium. Our tests resulted in one significantly associated gene-set consisting of FMRP-targeting transcripts. Our results are in line with the literature since FMRP has been linked to ASD and cognitive development in other genetic studies such as copy number variant and whole-exome sequencing studies (A. Jansen et al., 2017). This gene-set has also been linked to SCZ suggesting that FMRP-targeted transcripts might be involved in a general mechanism with a shared genetic etiology between ASD and SCZ. Immunocytochemical research shows an FMRP deficit in neurons in the cortical and subcortical brain structures and increased expression of FMRP in white and gray matter infiltrating astrocytes (Wegiel et al., 2018) in individuals with ASD. The reported shrinkage of FMRP deficient neurons might provide an mechanistic explanation of neuronal, structural and functional changes reported in ASD (Wegiel et al., 2018).

6.2 General discussion and future directions

6.2.1 General discussion

The main aim of the research presented in this thesis was to investigate the shared genetic factors between different psychiatric disorders in children and adolescents. We have tested PRS of multiple neurodevelopmental and psychiatric disorders and psychological traits (Chapter 3

and 4) and ran gene-set analysis using gene-sets previously associated with psychiatric disorders (Chapter 5). The results do not conclusively support the claim that common polygenic risk is indeed shared across disorders but shows a more delicate picture.

We did not find evidence for the previously reported (*DSM 5*, 2013; Grove & et.al., 2017; Kanner, 1965; Zheng et al., 2018) overlap between ASD and SCZ as most of the gene-sets previously associated with SCZ (Duncan et al., 2014; Goudriaan et al., 2013) were not associated with ASD and the SCZ PRS (Pardiñas et al., 2018) was not significantly associated with ASD, ADHD or diagnostic status in the child and adolescent clinical 'Inside-Out' sample. However, we should interpret these results with caution as the samples used in this thesis were all relatively small, reducing the statistical power to detect these associations. In addition, ASD can differ in expression between individuals as the experienced limitations in the social and communication domains and the cognitive functions of individuals vary greatly between individuals. A phenomenon called heterogeneity. This heterogeneity of the ASD phenotype might reflect small genotypic differences in ASD subgroups. Due to selection criteria or institutional biases between samples this may lead to subtle underlying genetic variations between samples resulting in our null findings.

Another major finding is the non-significant association between the ASD PRS and ASD status, ADHD status and psychiatric diagnostic status in the 'Inside-Out' sample. In general, this might be due to different genetic architectures between these disorders. Yet, it remains surprising that the ASD PRS is not associated with ASD status in the 'Inside-Out' sample. This can be a matter of low power in the GWAS with ~ 18,000 cases or point in the direction of a larger role for rare variation instead of common variation.

To summarize the findings regarding ASD:

1) There is no association between the ASD PRS and (1) ASD status, (2) ADHD status and (3) general psychopathology status in the 'Inside-Out' sample.

2) There is no significant association between the ADHD PRS and ASD status in the 'Inside-Out' sample.

3) There is no significant association between the SCZ PRS and ASD status in the 'Inside-Out' sample.

4) The tested SCZ associated gene-sets are not significantly associated with ASD.

These four findings point in the direction of a smaller than expected role of common variation in ASD. There is a vast body of literature reporting ASD to be robustly associated with rare variants (De Rubeis et al., 2014; Iossifov et al., 2014; Sanders et al., 2015; Satterstrom et al., 2020). This is in line with the hypothesis that ASD has a genetic architecture that is more dependent on rare variants than common variants. However, it is possible that there is a still undetected common variant load also contributing to ASD which might be identified with a larger ASD GWAS. We did find genes in a FMRP gene-set to be significantly associated with ASD, more deeply establishing the genetic and molecular link between ASD and FMRP which is thought to be executed through RNA binding and editing (Darnell et al., 2011; Tran et al., 2019).

Two studies in my thesis showed that the ADHD PRS is associated with ADHD status, combined ASD/ADHD status and psychiatric diagnostic status in the 'Inside-Out' sample. This is

interesting as the GWAS of ADHD is, compared to other GWAS of psychiatric traits, based on a rather small sample (20,183 cases, 35,191 controls). The fact that the ADHD PRS is significantly associated indicates that in ADHD common genetic variation might play a large role in the development of ADHD. In addition, it is possible that the common genetic variation comprising the ADHD PRS also plays a role in other disorders such and receiving a psychiatric diagnosis in general. This possibility is strengthened by other research with the ADHD PRS also showing association with for example externalizing problems (Ronald et al., 2021). However, as we cannot draw conclusions concerning causality it remains to be investigated whether observed associations are an actual overlap in causal SNPs or if ADHD at a young age is a risk factor for e.g. externalizing problems given impulsive behavior of individuals diagnosed with ADHD, thereby mimicking genetic overlap. Similar questions can be posed for other disorders associated with the ADHD common genetic variation. Regardless, our results implicate a larger overlap in common genetic variation between ADHD and other psychiatric disorders compared to the overlap in common variation between ASD and other psychiatric disorders.

Looking at a broader phenotype, general psychopathology measured as psychiatric diagnostic status (having a diagnosis yes/no), the results are mixed. The testing of six psychological traits (intelligence, neuroticism, smoking initiation, risk taking, insomnia and educational attainment) and eight neurodevelopmental or psychiatric disorders (ASD, ADHD, SCZ, alcohol dependence, major depression, anti-social behavior, bipolar disorder and anxiety) showed that eight PRS (ADHD, educational attainment, smoking initiation, intelligence, neuroticism, alcohol dependence, major depression and anti-social behavior) were significantly associated with diagnostic status. This does not confirm nor reject the general psychopathology factor theory (Smith et al., 2020) described in the introduction of this thesis. The genetic vulnerability for low educational attainment and smoking initiation is associated with general psychopathology and as such they might play a role in the general psychopathology factor. However, before initiating research into this direction the question of causality between educational attainment and smoking initiation and general psychopathology needs to be answered.

My main finding was the highly significant association between smoking initiation and educational attainment and general psychopathology. One explanation for this finding can be the statistical power of these two GWAS. With 1.1 million individuals for educational attainment and 1.2 million for smoking initiation they are among the largest datasets used so far. Next to this statistical explanation we see that associations between educational attainment and psychopathology have been reported before (P. R. Jansen et al., 2017). Additionally, it is important to note that educational attainment is a broad construct with a relatively low (20%) heritability (Okbay et al., 2016). This leaves ample room for environmental influences. Further, it might be beneficial to genetically entangle the most important components this phenotype encompasses such as cognitive abilities (intelligence), non-cognitive abilities (patience, self-control, temperament, motivation, self-discipline, time preference), health endowments and family background (Conti et al., 2010). In addition, educational attainment is strongly correlated to adolescent cognitive ability which is influenced by early-life socioeconomic status (Zhang et

al., 2020). Parental income can be considered as a proxy for socio-economic status with a high parental income representing a higher socio-economic status, which is in turn associated with fewer mental disorders in children (Kinge et al., 2021). This shows that next to the observed genotypic association between educational attainment and childhood psychopathology there is also an observed phenotypic association. With these observed phenotypic and genotypic associations it is of importance to continue the research on the currently unknown underlying mechanisms.

Regarding smoking initiation, phenotypic associations between smoking and general psychopathology have been observed in the adult psychopathology literature (Gfroerer et al., 2013). The causality question is not easily answered as smoking initiation is embedded in a complex interplay of potentially causal relations (Liu et al., 2019) but several hypotheses can be formulated. For example, longitudinal research in a child and adolescent sample points out that smoking at age 17 is preceded by early emergent, ongoing externalizing disorder, starting at the young age of 5 years (Zubrick et al., 2012), an observation congruent with earlier observations showing that psychiatric problems often start earlier than substance disorders (Kessler, 2004). This association might point in the direction of a coping mechanism or may be due to correlated or common factors that are possibly mediated by shared genetic factors (Zubrick et al., 2012). Among individuals with an anxiety or depressive disorder smoking is highly prevalent but the results on causality provide no clear direction between smoking (and by extension smoking initiation) and anxiety or depressive disorders (Fluharty et al., 2017). The current hypotheses for this high concordance rate are 1. the self-medication hypothesis suggesting smoking to come forth from depression or anxiety symptoms, 2. a bidirectional effect with smoking alleviating depressive or anxiety related symptoms short term but worsening them over time or 3. a shared underlying genetic predisposition (Fluharty et al., 2017). Based on animal models, a 4th hypothesis is that smoking leads to depression or anxiety disorders that are associated with increased susceptibility to environmental stressors. These studies indicate dysregulations in the hypothalamic-pituitary-adrenal system, due to prolonged nicotine exposure, resulting in cortisol hypersecretion and altered activity in the monoamine neurotransmitter system which has a main function in regulating reactions to stressors (Fluharty et al., 2017). This alteration seems to normalize after nicotine withdrawal. In addition to this hypothesized neurotransmitter circuit, a large GWAS on smoking initiation reported evidence for pathways involved in nicotinic, dopaminergic and glutamatergic neurotransmission in cortical (inferior temporal pathways, dorsolateral and medial prefrontal cortex) and subcortical (hippocampus, caudatus, striatum) regions in the brain (Liu et al., 2019). Future research could examine if neurotransmitter pathways associated with smoking initiation might partly overlap with pathways associated with childhood psychiatric disorders that genetically correlate with smoking initiation.

Lastly, we found evidence by means of gene-set analyses for a role of FMRP-targeted transcripts in ASD. Next to ASD, FMRP is associated with several psychiatric conditions (Fatemi & Folsom, 2011) and a mutation in the *FMRI* gene. The gene, coding for the RNA interacting with the FMRP-targeted transcripts, is a causal factor for the Fragile-X disorder (Gross et al., 2015). This is a syndrome with a broad range of intellectual disabilities and a high ASD comorbidity rate:

up to 30% (Hagerman et al., 2005) of the children with Fragile-X syndrome are diagnosed with ASD and 2–8% (Hagerman et al., 2005) of children with ASD will have the responsible *FMRI* gene mutation. Our results show that next to the phenotypic comorbidity a genetic association is present. In the overlap between ASD and Fragile-X, FMRP might be an essential component. It is a factor in the regulation of the translation of many other messengers involved with, synaptic functions (plasticity and maturation), axonal guidance and related synaptic pruning. These processes are involved in Fragile-X syndrome, ASD (Hagerman et al., 2005) and general intelligence (Crespi, 2016) and a dysregulation in the processes might be at the core of the overlap between ASD, Fragile-X and their complex relationship with IQ.

6.2.2 Scientific challenges

An important issue for this thesis as well as in the more general field of ASD research, revolves around the entanglement between ASD and intellectual disability (ID). ASD is often co-diagnosed with intellectual disability and a recent study reported percentages as high as 31% of ASD cases having an IQ <70, 25% falling within the 71 – 85 IQ range and the remaining 44% having IQ scores >85 (Christensen et al., 2018). In this thesis, Chapter 3 takes ID into account. As an additional analysis I investigated if the results were driven by the ID subgroup by comparing the diagnostic distribution of the whole sample to the ID part of the sample. Additionally, we ran the univariate analysis for the intelligence and EA PRS on the part of the 'Inside-Out' sample without ID. Both analysis did not show the results to be driven by ID. Interestingly, in a well powered IQ GWAS paper the genetic association between intelligence and ASD is positive (0.25) where it is negative with other psychiatric disorders such as ADHD (-0.36) (Savage et al., 2018). An explanation for the overlap between ASD and IQ alleles may be that ASD can be seen as a disorder involving a dysregulation of several systems regulating cognitive development meaning that ASD is associated with high but imbalanced enhancement across components of intelligence (Crespi, 2016). To elaborate on this, it might be that an overgrowth or to little pruning of neurons during neurodevelopment in individuals with ASD leads to abnormal connectivity in the brain. The genetic correlation in turn might pick up on the neurodevelopmental processes essential for normal cognitive function which are over expressed resulting in a positive correlation. Finally, it is important to realize that the net genetic correlation between ASD and IQ is positive but that the genome-wide correlation is actually a mixture of positive and negative signals. It is likely that the neurodevelopmental genes as a whole have a stronger correlation resulting in the overall positive genetic correlation between ASD and IQ.

Another important finding regarding ASD concerns the presence of subgroups due to the large phenotypical heterogeneity. Distinct ASD subgroups have unique SNP heritabilities being respectively 0.097, 0.049 and 0.045 for Asperger's syndrome, autistic disorder and pervasive developmental disorders not otherwise specified (Grove et al., 2019). Additionally, the SNP heritabilities between ASD with and without intellectual disability also differ, being respectively 0.029 and 0.086 (Grove et al., 2019). And as the association with the IQ PRS differs

between ASD subgroups, with the highest IQ PRS load in the subgroup with Asperger's syndrome (Grove et al., 2019), it is even conceivable that the ASD subgroups differ not only in specific causal rare variants but that ASD subtypes are also influenced by different common variants. These are findings that are important to examine as they may indicate differences in underlying etiology.

In addition to the recent insights through common variant research, it is shown that rare de novo mutations play a large role in ASD genetics with most of the mutations playing a role in ASD also decreasing IQ (Iossifov et al., 2014; Sullivan & Geschwind, 2019), results in line with the comorbidity rate between ASD and intellectual disability. This observed pleiotropy needs to be examined further. Follow-up questions might concern the genes with large effect sizes affecting both ASD and intellectual disability. For example: do these genes affect biological processes differently than genes causing ASD without intellectual disability? Further, gene-network analysis suggest molecular processes that distinguish ASD from intellectual disability (Sullivan & Geschwind, 2019) making it important to carefully report on which phenotype is being investigated, intellectual disability and/or ASD. For this issue, in-depth phenotyping can provide solutions. This is however not a clear-cut solution as large scale genetic research is conducted by combining many cohorts and running meta-analyses on these data (Autism Spectrum Disorder Working Group of the Psychiatry Genomics Consortium, Under Review). These cohorts might have used different, all correct, ways of diagnosing patients and might have collected different phenotypic data (such as absence or presence of IQ measures, behavioral measures and comorbid disorders) next to the genetic data. These differences can make it difficult to account for certain traits such as IQ scores. However, the most recent ASD GWAS reported in-depth analyses on ASD subgroups showing that this information is increasingly becoming available and already provides valuable information (Grove et al., 2019). This study shows it is desirable that this in-depth phenotyping takes flight for ASD and other heterogenous disorders.

A final point to address is the fact that currently a lot of the genetic research, including this thesis, is using diagnostic categories to investigate psychiatric disorders. This is a dichotomous measure which likely contains less information than continuous measures. It would be a good addition to the field if the nuance of continuous measures is maintained and tapped into.

6.2.3 Future directions

The main aim of this thesis is to advance the research conducted to benefit patients diagnosed with psychiatric disorders and their families. I highlight the largest steps towards this goal here. The research conducted in this thesis relies heavily on GWAS results based on relatively small GWAS sample sizes such as for ASD and ADHD. Enlarging these GWAS sample sizes to identify more significantly associated SNPs and provide more accurate effect sizes will be beneficial to experiments featuring PRS as conducted in the projects reported here. This will also further the

insights into biological pathways involved in psychiatric disorders. From the early beginnings of the project described in this thesis until now GWAS have been steadily growing in sample size with sample sizes now exceeding the one-million mark for several psychiatric disorders and traits such as insomnia (P. R. Jansen et al., 2019) and neuroticism (Nagel et al., 2018). This milestone gave rise to the idea we have arrived in the post-GWAS era (Dick et al., 2018; Marjoram et al., 2014) which is by no means the case for many psychiatric disorders as the still relatively limited sample sizes (~20,000 cases for ADHD) and consequently low harvest of new risk loci and limited insights in to biological pathways of the most recent GWAS on ASD (Grove et al., 2019) and ADHD (Demontis & Walters, 2017) show. In addition to larger sample sizes, more elaborate captures of the genome by means of whole exome sequencing studies (WES) and whole genome sequencing studies (WGS) are expected to be helpful in elucidating the genetic architecture of disorders of which we expect rare variants to be large contributors to the genetic make-up, such as ASD (A. Jansen et al., 2017). Next to that, WES and WGS will also improve the accuracy of the estimated effect sizes enhancing the value of the PRS for future applications due to improved accuracy. The PRS are already helpful for certain common diseases such as cardiovascular disease to facilitate risk stratification (Khera et al., 2018; Nikpay et al., 2015) and this is an application that might benefit psychiatric disorders in the future as well. These larger WES and WGS datasets do however bring up issues regarding data acquisition, storage, costs and distribution that need to be sorted out (Stephens et al., 2015).

Application of the psychiatric GWAS results to the clinical practice can go beyond PRS and their possibilities for risk stratification in large groups. Another important step forward will be the use of the GWAS results to find novel pharmacological interventions for psychiatric disorders, an area that has not had many recent break-throughs (Breen et al., 2016). However, a step before finding novel pharmaceutical interventions is correct interpretation of GWAS results. In this context several new tools and methods have been developed such as Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) (Watanabe et al., 2017). FUMA can be used to annotate, prioritize, visualize and interpret GWAS results. Variants associated to disorders by large GWAS can exert their effect either via the changed structure of a protein or it may affect the expression level of a gene (so called expression quantitative trail loci/ eQTL). Of these two methods the transcriptional control affecting the expression level of a gene is thought to play the largest role in disorders. Additionally, it has been reported that GWAS hits are enriched in regulatory sequences (Uffelmann & Posthuma, 2021). A way to gain further biological insight is by seeking convergent functions among the identified variants and map these to genes and pathways relevant to disorders (Uffelmann & Posthuma, 2021). This process of functional annotation of GWAS hits uses information from many different, often publicly available resources. For functional annotation the following four resources or methods are important. 1) Positional mapping of genes in coding regions linking them to genes directly. 2) Transcriptomics which consist of RNA signatures as measured by RNAseq, which integrated with genotype data allows for eQTL mapping. 3) Epigenetics which by adding methylation and acetylation data to eQTL data can provide insights into how these eQTL variants regulate gene expression (Uffelmann & Posthuma, 2021). Using this method, it was observed that 9% of

identified eQTL loci mediate their effect through epigenetic modification. 4) Chromatin confirmation which refers to the 3D chromatin organization and the resulting interactions between regions mega bases away (Uffelmann & Posthuma, 2021). After mapping the most likely trait associated variants to genes one can investigate if these variants converge in pre-defined gene-sets by running gene-set analyses which test if the association of the genes in the gene-set are stronger than the associations of genes not in the gene-set (de Leeuw et al., 2015; Uffelmann & Posthuma, 2021).

Still, several caveats need to be resolved. For instance, many of the available datasets used for transcriptomics are not brain specific fine grained single cell data but coarse bulk tissue data, let alone have spatial and temporal information and are, in addition, not disorder specific (most data is on donors without psychiatric disorders). Even when this information will become available in the future, this only allows for hypothesis formulation on specific spatial and temporal stages, possibly of certain cell types, which is too broad for functional in vitro follow-up studies which require specific biomarkers to evaluate (Uffelmann & Posthuma, 2021). Hence, after functional annotation of GWAS hits, one can only start to generate but not confirm hypotheses regarding involved biological processes. By means of computational approaches which take the polygenicity and the effect of the mutations into account these hypotheses can be converted into hypotheses which can be validated. With these hypotheses, one can start with functional in vivo and in vitro follow-up studies (Uffelmann & Posthuma, 2021), an important step which brings new insights into disorder etiologies and new treatment methods a step closer to the patients.

6.3 Final conclusion

This thesis provides insights into genetic comorbidity between psychiatric disorders and contributes to the rapidly growing body of genetic evidence generating insights into the development of psychiatric disorders.

Interpreting my results with due consideration of the current state of the scientific field lead me to draw three final conclusions.

1) ASD is a unique disorder in the group of psychiatric disorders as it seems to have a different genetic architecture. A conclusion we draw as the ASD PRS is not significantly genetically associated to other psychiatric traits and our ASD sample also shows no association with other psychiatric PRS. This may point towards a limited role for common variation and a larger role for rare variants in the development of ASD. The role for common variation in other neurodevelopmental disorders, for example ADHD, seems larger as the ADHD PRS is significantly associated with psychiatric case-control status, ADHD and combined ASD/ADHD status. For all psychiatric disorders it is important to account for the heterogeneity but for ASD, having this unique genetic signature and previously published differing SNP heritabilities between subtypes, it is essential that in-depth phenotyping becomes the standard.

2) The overlap in common variation between psychiatric disorders is present for certain disorders but not for all. Educational attainment and smoking initiation are genetically associated with general psychopathology, I cannot elaborate on association with specific psychiatric disorders but as the disorders are correlated with each other it is certainly plausible that educational attainment and smoking initiation are correlated to at least some psychiatric disorders individually. There might be a more sophisticated underlying web of correlations on symptom level which needs to be tapped into. Here also, in-depth phenotyping is essential for additional insights.

3) The explained variance of the tested PRS is low but in line with other published data. This leaves ample room for other factors to contribute to the origination of these disorders among which rare variants, currently undiscovered common variation and environmental factors.

The scientific community has by no means disentangled the biological mechanisms through which psychiatric disorders develop but research done on small samples as investigated in this thesis, and the larger collaborations comprising these small samples, is advancing our knowledge at a fast rate and will hopefully in the near future get to a point where the patients in the clinics can benefit from the insights resulting in better quality of life.

References

Autism Spectrum Disorder Working Group of the Psychiatry Genomics Consortium. (Under Review). *High-Resolution Meta-analysis of Genome-wide Association Studies of Autism Spectrum Disorder.*

Breen, G., Li, Q., Roth, B. L., O'Donnell, P., Didriksen, M., Dolmetsch, R., O'Reilly, P. F., Gaspar, H. A., Manji, H., Huebel, C., Kelsoe, J. R., Malhotra, D., Bertolino, A., Posthuma, D., Sklar, P., Kapur, S., Sullivan, P. F., Collier, D. A., & Edenberg, H. J. (2016). Translating genome-wide association findings into new therapeutics for psychiatry. *Nature Neuroscience*, *19*(11), 1392–1396. https://doi.org/10.1038/nn.4411

Christensen, D. L., Braun, K. V. N., Baio, J., Bilder, D., Charles, J., Constantino, J. N., Daniels, J., Durkin, M. S., Fitzgerald, R. T., Kurzius-Spencer, M., Lee, L.-C., Pettygrove, S., Robinson, C., Schulz, E., Wells, C., Wingate, M. S., Zahorodny, W., & Yeargin-Allsopp, M. (2018). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *MMWR Surveillance Summaries*, *65*(13), 1–23. https://doi.org/10.15585/mmwr.ss6513a1

Conti, G., Heckman, J., & Urzua, S. (2010). THE EDUCATION-HEALTH GRADIENT. *The American Economic Review*, *100*(2), 234–238. https://doi.org/10.1257/aer.100.2.234

Costello, E. J., Egger, H., & Angold, A. (2005). 10-Year Research Update Review: The Epidemiology of Child and Adolescent Psychiatric Disorders: I. Methods and Public Health Burden. *Journal*

of the American Academy of Child & Adolescent Psychiatry, *44*(10), 972–986. https://doi.org/10.1097/01.chi.0000172552.41596.6f

Crespi, B. J. (2016). Autism As a Disorder of High Intelligence. *Frontiers in Neuroscience*, *10*, 300. https://doi.org/10.3389/fnins.2016.00300

Darnell, J. C., Van Driesche, S. J., Zhang, C., Hung, K. Y. S., Mele, A., Fraser, C. E., Stone, E. F., Chen, C., Fak, J. J., Chi, S. W., Licatalosi, D. D., Richter, J. D., & Darnell, R. B. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell*, *146*(2), 247–261. https://doi.org/10.1016/j.cell.2011.06.013

de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Computational Biology*, *11*(4), e1004219. https://doi.org/10.1371/journal.pcbi.1004219

De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., Kou, Y., Liu, L., Fromer, M., Walker, S., Singh, T., Klei, L., Kosmicki, J., Shih-Chen, F., Aleksic, B., Biscaldi, M., Bolton, P. F., Brownfeld, J. M., Cai, J., … Buxbaum, J. D. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, *515*(7526), 209–215. https://doi.org/10.1038/nature13772

Demontis, D., & Walters, R. K. (2017). *Discovery Of The First Genome-Wide Significant Risk Loci For ADHD*. https://doi.org/10.1101/145581

Diagnostic and Statistical Manual of mental disorders (Vol. 5). (2013). American Psychiatric Association.

Dick, D. M., Barr, P. B., Cho, S. B., Cooke, M. E., Kuo, S. I.-C., Lewis, T. J., Neale, Z., Salvatore, J. E., Savage, J., & Su, J. (2018). Post-GWAS in Psychiatric Genetics: A Developmental Perspective on the "Other" Next Steps. *Genes, Brain, and Behavior*, *17*(3), e12447. https://doi.org/10.1111/gbb.12447

Duncan, L. E., Holmans, P. A., Lee, P. H., O'Dushlaine, C. T., Kirby, A. W., Smoller, J. W., Öngür, D., & Cohen, B. M. (2014). Pathway analyses implicate glial cells in schizophrenia. *PloS One*, *9*(2), e89441. https://doi.org/10.1371/journal.pone.0089441

Fatemi, S. H., & Folsom, T. D. (2011). The role of fragile X mental retardation protein in major mental disorders. *Neuropharmacology*, *60*(7–8), 1221–1226. https://doi.org/10.1016/j.neuropharm.2010.11.011

Fluharty, M., Taylor, A. E., Grabski, M., & Munafò, M. R. (2017). The Association of Cigarette Smoking With Depression and Anxiety: A Systematic Review. *Nicotine & Tobacco Research: Official Journal of the Society for Research on Nicotine and Tobacco*, *19*(1), 3–13. https://doi.org/10.1093/ntr/ntw140

Gfroerer, J., Dube, S. R., King, B. A., Garrett, B. E., Babb, S., & McAfee, T. (2013). Vital Signs: Current Cigarette Smoking Among Adults Aged ≥18 Years with Mental Illness — United States, 2009–2011. *MMWR. Morbidity and Mortality Weekly Report*, *62*(5), 81–87.

Goudriaan, A., de Leeuw, C., Ripke, S., Hultman, C. M., Sklar, P., Sullivan, P. F., Smit, A. B., Posthuma, D., & Verheijen, M. H. G. (2013). Specific Glial Functions Contribute to Schizophrenia Susceptibility. *Schizophrenia Bulletin*. https://doi.org/10.1093/schbul/sbt109

Gross, C., Hoffmann, A., Bassell, G. J., & Berry-Kravis, E. M. (2015). Therapeutic Strategies in Fragile X Syndrome: From Bench to Bedside and Back. *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics*, *12*(3), 584–608. https://doi.org/10.1007/s13311-015-0355-9

Grove, J., & et.al. (2017). *Common risk variants identified in autism spectrum disorder*. http://dx.doi.org/10.1101/224774

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Awashti, S., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Christensen, J. H., … 23andMe Research Team. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, *51*(3), 431–444. https://doi.org/10.1038/s41588-019-0344-8

Hagerman, R. J., Ono, M. Y., & Hagerman, P. J. (2005). Recent advances in fragile X: A model for autism and neurodegeneration. *Current Opinion in Psychiatry*, *18*(5), 490–496. https://doi.org/10.1097/01.yco.0000179485.39520.b0

Iossifov, I., O'Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., Stessman, H. A., Witherspoon, K. T., Vives, L., Patterson, K. E., Smith, J. D., Paeper, B., Nickerson, D. A., Dea, J., Dong, S., Gonzalez, L. E., Mandell, J. D., Mane, S. M., Murtha, M. T., … Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, *515*(7526), 216–221. https://doi.org/10.1038/nature13908

Jansen, A., Dieleman, G. C., Smit, A. B., Verhage, M., Verhulst, F. C., Polderman, T. J. C., & Posthuma, D. (2017). Gene-set analysis shows association between FMRP targets and autism spectrum disorder. *European Journal of Human Genetics: EJHG*. https://doi.org/10.1038/ejhg.2017.55

Jansen, P. R., Polderman, T. J. C., Bolhuis, K., van der Ende, J., Jaddoe, V. W. V., Verhulst, F. C., White, T., Posthuma, D., & Tiemeier, H. (2017). Polygenic scores for schizophrenia and educational attainment are associated with behavioural problems in early childhood in the general population. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*. https://doi.org/10.1111/jcpp.12759

Jansen, P. R., Watanabe, K., Stringer, S., Skene, N., Bryois, J., Hammerschlag, A. R., Leeuw, C. A. de, Benjamins, J. S., Muñoz-Manchado, A. B., Nagel, M., Savage, J. E., Tiemeier, H., White, T., Tung, J. Y., Hinds, D. A., Vacic, V., Wang, X., Sullivan, P. F., Sluis, S. van der, … Posthuma, D. (2019). Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nature Genetics*, *51*(3), 394. https://doi.org/10.1038/s41588-018-0333-3

Kanner, L. (1965). Infantile autism and the schizophrenias. *Behavioral Science*, *10*(4), 412–420.

Kessler, R. C. (2004). The epidemiology of dual diagnosis. *Biological Psychiatry*, *56*(10), 730–737. https://doi.org/10.1016/j.biopsych.2004.06.034

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T., & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, *50*(9), 1219–1224. https://doi.org/10.1038/s41588-018-0183-z

Kinge, J. M., Øverland, S., Flatø, M., Dieleman, J., Røgeberg, O., Magnus, M. C., Evensen, M., Tesli, M., Skrondal, A., Stoltenberg, C., Vollset, S. E., Håberg, S., & Torvik, F. A. (2021). Parental income and mental disorders in children and adolescents: Prospective register-based study. *International Journal of Epidemiology*. https://doi.org/10.1093/ije/dyab066

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., Zhan, X., Choquet, H., Docherty, A. R., Faul, J. D., Foerster, J. R., Fritsche, L. G., Gabrielsen, M. E., Gordon, S. D., Haessler, J., … Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, *51*(2), 237. https://doi.org/10.1038/s41588-018-0307-5

Marjoram, P., Zubair, A., & Nuzhdin, S. V. (2014). Post-GWAS: Where next? More samples, more SNPs or more biology? *Heredity*, *112*(1), 79–88. https://doi.org/10.1038/hdy.2013.52

Nagel, M., Jansen, P. R., Stringer, S., Watanabe, K., de Leeuw, C. A., Bryois, J., Savage, J. E., Hammerschlag, A. R., Skene, N. G., Muñoz-Manchado, A. B., 23andMe Research Team, White, T., Tiemeier, H., Linnarsson, S., Hjerling-Leffler, J., Polderman, T. J. C., Sullivan, P. F., van der Sluis, S., & Posthuma, D. (2018). Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genetics*, *50*(7), 920–927. https://doi.org/10.1038/s41588-018-0151-7

Nikpay, M., Goel, A., Won, H.-H., Hall, L. M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C. P., Hopewell, J. C., Webb, T. R., Zeng, L., Dehghan, A., Alver, M., Armasu, S. M., Auro, K., Bjonnes, A., Chasman, D. I., Chen, S., … Farrall, M. (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics*, *47*(10), 1121–1130. https://doi.org/10.1038/ng.3396

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., Turley, P., Chen, G.-B., Emilsson, V., Meddens, S. F. W., Oskarsson, S., Pickrell, J. K., Thom, K., Timshel, P., de Vlaming, R., Abdellaoui, A., Ahluwalia, T. S., Bacelis, J., Baumbach, C., … Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*(7604), 539–542. https://doi.org/10.1038/nature17671

Pardiñas, A. F., Holmans, P., Pocklington, A. J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S. E., Bishop, S., Cameron, D., Hamshere, M. L., Han, J., Hubbard, L., Lynham, A., Mantripragada, K., Rees, E., MacCabe, J. H., McCarroll, S. A., Baune, B. T., Breen, G., … CRESTAR Consortium. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in

regions under strong background selection. *Nature Genetics*, *50*(3), 381–389. https://doi.org/10.1038/s41588-018-0059-2

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), 702–709. https://doi.org/10.1038/ng.3285

Ronald, A., Bode, N. de, & Polderman, T. J. C. (2021). Systematic Review: How the Attention-Deficit/Hyperactivity Disorder Polygenic Risk Score Adds to Our Understanding of ADHD and Associated Traits. *Journal of the American Academy of Child & Adolescent Psychiatry*, *0*(0). https://doi.org/10.1016/j.jaac.2021.01.019

Sanders, S. J., He, X., Willsey, A. J., Ercan-Sencicek, A. G., Samocha, K. E., Cicek, A. E., Murtha, M. T., Bal, V. H., Bishop, S. L., Dong, S., Goldberg, A. P., Jinlu, C., Keaney, J. F., Klei, L., Mandell, J. D., Moreno-De-Luca, D., Poultney, C. S., Robinson, E. B., Smith, L., … State, M. W. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*, *87*(6), 1215–1233. https://doi.org/10.1016/j.neuron.2015.09.016

Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M. S., Artomov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., … Walters, R. K. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*. https://doi.org/10.1016/j.cell.2019.12.036

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J. R. I., Grasby, K. L., Hammerschlag, A. R., Kaminski, J. A., Karlsson, R., Krapohl, E., Lam, M., Nygaard, M., Reynolds, C. A., Trampush, J. W., … Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, *50*(7), 912–919. https://doi.org/10.1038/s41588-018-0152-6

Schadt, E. E. (2009). Molecular networks as sensors and drivers of common human diseases. *Nature*, *461*(7261), 218–223. https://doi.org/10.1038/nature08454

Smith, G. T., Atkinson, E. A., Davis, H. A., Riley, E. N., & Oltmanns, J. R. (2020). The General Factor of Psychopathology. *Annual Review of Clinical Psychology*, *16*, 75–98. https://doi.org/10.1146/annurev-clinpsy-071119-115848

Solberg, B. S., Zayats, T., Posserud, M.-B., Halmøy, A., Engeland, A., Haavik, J., & Klungsøyr, K. (2019). Patterns of Psychiatric Comorbidity and Genetic Correlations Provide New Insights Into Differences Between Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder. *Biological Psychiatry*, *86*(8), 587–598. https://doi.org/10.1016/j.biopsych.2019.04.021

Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., Iyer, R., Schatz, M. C., Sinha, S., & Robinson, G. E. (2015). Big Data: Astronomical or Genomical? *PLoS Biology*, *13*(7), e1002195. https://doi.org/10.1371/journal.pbio.1002195

Sullivan, P. F., & Geschwind, D. H. (2019). Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell*, *177*(1), 162–183. https://doi.org/10.1016/j.cell.2019.01.015

Tran, S. S., Jun, H.-I., Bahn, J. H., Azghadi, A., Ramaswami, G., Van Nostrand, E. L., Nguyen, T. B., Hsiao, Y.-H. E., Lee, C., Pratt, G. A., Martínez-Cerdeño, V., Hagerman, R. J., Yeo, G. W., Geschwind, D. H., & Xiao, X. (2019). Widespread RNA editing dysregulation in brains from autistic individuals. *Nature Neuroscience*, *22*(1), 25–36. https://doi.org/10.1038/s41593-018-0287-x

Uffelmann, E., & Posthuma, D. (2021). Emerging Methods and Resources for Biological Interrogation of Neuropsychiatric Polygenic Signal. *Biological Psychiatry*, *89*(1), 41–53. https://doi.org/10.1016/j.biopsych.2020.05.022

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*, *101*(1), 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, *8*(1), 1826. https://doi.org/10.1038/s41467-017-01261-5

Wegiel, J., Brown, W. T., Fauci, G. L., Adayev, T., Kascsak, R., Kascsak, R., Flory, M., Kaczmarski, W., Kuchna, I., Nowicki, K., Martinez-Cerdeno, V., Wisniewski, T., & Wegiel, J. (2018). The role of reduced expression of fragile X mental retardation protein in neurons and increased expression in astrocytes in idiopathic and syndromic autism (duplications 15q11.2-q13). *Autism Research*, *11*(10), 1316–1331. https://doi.org/10.1002/aur.2003

Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A. E., Dudbridge, F., & Middeldorp, C. M. (2014). Research review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *55*(10), 1068–1087. https://doi.org/10.1111/jcpp.12295

Wray, N. R., Lin, T., Austin, J., McGrath, J. J., Hickie, I. B., Murray, G. K., & Visscher, P. M. (2020). From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2020.3049

Zhang, Z., Liu, H., & Choi, S.-W. (2020). Early-life socioeconomic status, adolescent cognitive ability, and cognition in late midlife: Evidence from the Wisconsin Longitudinal Study. *Social Science & Medicine (1982)*, *244*, 112575. https://doi.org/10.1016/j.socscimed.2019.112575

Zheng, Z., Zheng, P., & Zou, X. (2018). Association between schizophrenia and autism spectrum disorder: A systematic review and meta-analysis. *Autism Research: Official Journal of the International Society for Autism Research*, *11*(8), 1110–1119. https://doi.org/10.1002/aur.1977

Zubrick, S. R., Lawrence, D., Mitrou, F., Christensen, D., & Taylor, C. L. (2012). Early mental health morbidity and later smoking at age 17 years. *Psychological Medicine*, *42*(5), 1103–1115. https://doi.org/10.1017/S0033291711002182

Mia xx

Addendum

Nederlandse samenvatting

Dit proefschift gaat over de genetische basis van psychiatrische aandoeningen. Eerder onderzoek heeft aangetoond dat deze aandoeningen vaak in meer of mindere mate erfelijk zijn. Een hoge mate van erfelijkheid betekent dat we aannemen dat de verschillen tussen mensen vooral door verschillen in genetische make-up kunnen worden verklaard. Ik heb verschillende statistische methoden gebruikt om tot verdere inzichten over de onderliggende *veel voorkomende genetische varianten* te komen, zowel aandoeningsspecifiek als overkoepelend tussen aandoeningen. Als wetenschappelijk vakgebied zijn we op het punt dat we voor een deel van de aandoeningen een eerste inschatting kunnen maken over welke *veel voorkomende genetische varianten* (ook wel single nucleotide polymorphisms, SNPs in het kort, genoemd) een rol spelen bij de ontwikkeling van een aandoening. We hebben echter nog geen helder beeld over hoe onderliggende cellulaire en moleculaire processen, waarop we eventueel kunnen ingrijpen met medicatie, samenhangen met psychiatrische aandoeningen. Met dit proefschrift dragen we bij aan het ontrafelen van de genetische puzzel door te kijken naar de effecten van veel voorkomende genetische varianten. We onderzoeken de effecten van veel voorkomende genetische variatie op dezelfde aandoening maar ook tussen aandoeningen. Door genetische verbanden tussen aandoeningen aan te tonen kunnen we paden ontdekken die bij meerdere aandoeningen afwijken van de standaard. Hierop kan na nader onderzoek mogelijk ingegrepen worden om symptomen te verlichten.

Om de leesbaarheid ten goede te komen worden de *schuingedrukte technische termen* in de woordenlijst aan het eind van dit hoofdstuk beknopt uitgelegd.

In hoofdstuk 3 doe ik onderzoek naar het *fenotype* “psychiatrische diagnostische status”. Ik vergelijk een groep kinderen met verschillenden psychiatrische diagnoses (o.a. autisme spectrum stoornis (ASS), aandachtstekort en hyperactiviteit stoornis (ADHD), angststoornissen, depressie en eetstoornissen) met gezonde controles. Daarbij wil ik ontdekken welke *polygene risico scores* (PRS) het verschil tussen wel of geen psychiatrische diagnose kunnen voorspellen. Ik heb PRS van 14 kenmerken of aandoeningen getest: ASS, ADHD, angststoornis, depressie, schizofrenie, bipolaire stoornis, alcoholverslaving, neuroticisme, slapeloosheid/slaapproblemen, antisociaal gedrag, risicovol gedrag, beginnen met roken, intelligentie en aantal opleidingsjaren. Na het individueel testen van deze PRS heb ik de significante PRS in één model getest om te kijken of en zo ja in welke mate de *verklaarde variantie* vergroot. De analyses wijzen uit dat aantal opleidingsjaren, beginnen met roken, intelligentie, neuroticisme, alcoholverslaving, ADHD, depressie en antisociaal gedrag *significante voorspellers* zijn. Van deze acht significante eigenschappen zijn aantal opleidingsjaren en beginnen met roken de hoofdbevindingen omdat daze verreweg de sterkste verbanden hebben met het hebben van een psychiatrische diagnostische status wat we zien aan de lage *P-waarde*. Dit betekent dat genetische varianten die geassocieerd zijn met deze twee eigenschappen ook geassocieerd zijn met het al dan niet krijgen van een psychiatrische diagnose. Dit kan een aanknopingspunt zijn voor verder onderzoek naar de

samenhang tussen aantal opleidingsjaren en/of beginnen met roken en psychiatrische aandoeningen. Er kan bijvoorbeeld onderzocht worden of het een het ander veroorzaakt (causaliteit), of dat er wellicht een andere onderliggende eigenschap (of genetische variant) is die zowel aantal opleidingsjaren, beginnen met roken en het krijgen van een psychiatrische diagnose kan verklaren. Ook zou onderzocht kunnen worden of het aantal opleidingsjaren (bij kinderen gezien als schooluitval) en/of beginnen met roken een voorspellende rol kunnen spelen in het vroeg ontdekken van kinderen die een verhoogd risico lopen op het ontwikkelen van psychische klachten. Wat betreft de tweede onderzoeksvraag bleek dat meerdere PRS in één model de verklaarde variantie aanzienlijk doet stijgen wat betekent dat het interessant kan zijn in vervolgonderzoek te kijken naar welke PRS zinvol bijdragen aan zo'n model en hoe je tot de hoogste verklaarde variantie kan komen. Zo'n model zou in de toekomst mogelijk nuttig kunnen zijn als onderdeel van een testbatterij bij mensen met milde psychische klachten.

Hoofdstuk 4 gaat over zogeheten *ontwikkelingsstoornissen* waartoe ASS, ADHD en schizofrenie behoren. In het deel van onze testpopulatie wat een ASS en/of ADHD diagnose heeft test ik aan de hand van case-control onderzoek of de ASS, ADHD of schizofrenie PRS is geassocieerd met de diagnostische status (zijnde het wel of niet hebben van de betreffende aandoening) in de groep met alleen ASS, met alleen ADHD en in de gemengde ASS/ADHD groep. We zien dat de ASS en schizofrenie PRS niet geassocieerd zijn maar de ADHD PRS wel. De resultaten wijzen in de richting van ADHD specifieke genetische variatie en geven geen indicatie dat er overlap is in genetische variatie tussen de geteste aandoeningen. Interessant is dat de ASS PRS de ASS status ook niet kan voorspellen, dat zou kunnen betekenen dat in deze aandoening de veel voorkomende genetische varianten misschien een kleinere rol spelen dan in andere aandoeningen en dat ASS in sterkere mate wordt bepaald door *zeldzame varianten* of "*de novo" varianten*. Het kan ook zo zijn dat de ASS PRS simpelweg nog te weinig variantie verklaard, oftewel te weinig *statistische power* heeft.

In hoofdstuk 5 zoom ik verder in op ASS. ASS wordt al decennia lang geassocieerd met, of een kind-vorm van schizofrenie genoemd. Recent genetisch onderzoek laat een gemengd beeld zien. Sommige onderzoeken ondersteunen de hypothese dat ASS en schizofrenie verwant zijn, andere niet. Aan de hand van openbare genetische datasets onderzoek ik of ASS beïnvloed wordt door genetische varianten die zich in genetische paden bevinden die in eerder onderzoek met schizofrenie zijn geassocieerd. Gensets zijn groepen met genen die een rol spelen in hetzelfde biologische proces. De *genset analyses* die ik hiervoor heb uitgevoerd wijzen globaal niet in die richting. Van de 32 geteste *gensets* is er één die geassocieerd blijkt met ASS, de 'FMRP targeted transcripts' genset. Deze bevinding is in overeenstemming met de literatuur waarin FMRP, veroorzaker van het Fragiele X syndroom, eerder met ASS is geassocieerd. Onze resultaten tonen aan dat de FMRP targeted transcripts, welke in Fragile X syndroom ontregeld zijn, mogelijk ook een rol spelen in het ontwikkelen van ASS.

De onderzoeken in dit proefschrift geven een tweeledig beeld van de overlap in onderliggende genetische variatie. Er wordt aangetoond dat de overlap in onderliggende genetische variatie voor specifieke psychiatrische aandoeningen geldt. Voor de eerder in de literatuur gerapporteerde overlap tussen ASS en schizofrenie worden geen aanwijzingen gevonden. De gensets die eerder met schizofrenie geassocieerd zijn geven geen significante resultaten en de schizofrenie en ASS PRS kunnen in de gebruikte onderzoekspopulatie geen cases van controles onderscheiden. Buiten dat de overlap tussen ASS en schizofrenie niet aangetoond wordt, zijn ASS cases niet door de ADHD PRS van controles te onderscheiden. Dit, terwijl eerder onderzoek wel een genetische samenhang tussen beide aandoeningen aantoonde. De rol van veel voorkomende genetische varianten lijkt hierdoor in ADHD groter dan in ASS. ASS zou sterker kunnen afhangen van zeldzame varianten of “de novo” varianten. ASS wordt wel geassocieerd met de FMRP targeted genes wat de aanname dat er een genetische overlap tussen deze twee aandoeningen is ondersteunt. De twee belangrijkste eigenschappen die geassocieerd zijn met het al dan niet krijgen van een psychiatrische diagnose zijn schoolopleiding in jaren en beginnen met roken. Omdat de gedane testen geen uitspraak kunnen doen over causaliteit is het verder onderzoeken hiervan een belangrijke vervolgstap. Ook het verder uitpluizen van welke componenten van de veelzijdige eigenschap schoolopleiding in jaren verantwoordelijk zijn voor de associatie met psychiatrische diagnose is belangrijk.

Als kanttekening bij dit proefschrift wil ik de beperkte grootte van de data set noemen. We hebben een voor de genetica als klein bestempelde data set gebruikt (N < 2.000). De resultaten kunnen hierdoor beïnvloed worden, wat in de praktijk betekent dat we verbanden die er wel zijn over het hoofd kunnen zien. De kans dat we verbanden noemen die er in werkelijkheid niet zijn is kleiner. Het blijft dus van groot belang om (inter)nationaal samen te werken om data sets samen te voegen tot grote data sets met voldoende slagkracht om meer geassocieerde SNPs te vinden en de huidige kennis optimaal te benutten en te vergroten.

Alfabetische woordenlijst

Veel voorkomende genetische varianten:	Variatie in het DNA van een individu welke in meer dan 1% van de bevolking voorkomt.
De novo varianten:	Variatie in het DNA van een individu welke de ouders niet met zich meedragen en dus “nieuw” is in het kind, niet doorgegeven door de ouders.
Fenotype:	Uiterlijke verschijning van het onderliggende *genotype.*
Genetische varianten:	Het genetische materiaal van de mens, het DNA, bestaat uit een hele lange reeks nucleotiden, afgekort met de letters A, C, T en G. De nucleotiden die tussen mensen kunnen verschillen noemen we een genetische variant ook wel een single nucleotide polymorphisme (SNP) genoemd.
Genset:	Samngestelde set van genen welke een onderlinge samenhang hebben met biologie relevantie.
Genset analyses:	Statistische test welke bepaald of de genen in een genset meer met de aandoening samen hangen dan genen niet in de genset.
Genotype:	De genetische opmaak, het DNA, van een individu.
Ontwikkelingsstoornissen :	Aandoeningen welke zich in een vroeg levensstadium openbaren door verschillen in de ontwikkeling van hersenen en zenuwstelsel.
Polygene risico score (PRS):	Optelsom van alle genetische risico factoren die een individu in zijn DNA met zich meedraagt.
P-waarde:	Grens waartegen de uitslag van een statistische toets wordt gelegd. Deze is in regulier onderzoek 0.05. Een P-waarde onder 0.05 is significant (er is een verband aangetoond), een P-waarde boven 0.05 niet (er is geen verband aangetoond). In genetisch onderzoek is de P-waarde grens lager, namelijk $5x10^{-8}$, omdat er veel meer tests worden uitgevoerd wat de kans op toevallige of onjuiste toetsuitslagen vergroot.
Zeldzame varianten:	Variatie in het DNA van een individu welke in minder dan 1% van de bevolking voorkomt.

Significantie:	Bij een significante voorspeller is het aannemelijk dat de bevinding niet op toeval berust. De 0.05 grens is de algemeen aanvaarde wetenschappelijke grens welke zegt dat de kans dat het gevonden effect op toeval berust kleiner is dan 5%.
Statistische power:	Mate van accuratesse waarmee statistische toetsen geinterpreteerd kunnen worden. Hoge power wil zeggen dat er weinig kan is dat er een effect gevonden wordt wat er niet is of een effect wat er wel is over het hoofd wordt gezien.
Verklaarde variantie:	Het deel van de verschillen tussen individuen wat door de geteste variabele verklaard kan worden.

Publications

What twin studies tell us about the heritability of brain development, morphology, and function: a review.
Jansen AG, Mous SE, White T, Posthuma D, Polderman TJ.
Neuropsychol Rev. 2015 Mar;25(1):27-46. doi: 10.1007/s11065-015-9278-9. Epub 2015 Feb 12. PMID: 25672928; PMCID: PMC4412550.

Gene-set analysis shows association between FMRP targets and autism spectrum disorder.
Jansen A, Dieleman GC, Smit AB, Verhage M, Verhulst FC, Polderman TJC, Posthuma D.
Eur J Hum Genet. 2017 Jun;25(7):863-868. doi: 10.1038/ejhg.2017.55. Epub 2017 Apr 19. PMID: 28422133; PMCID: PMC5520067.

Polygenic Risk Scores as Predictor for Attention Deficit/Hyperactivity Disorder and Autism Spectrum Disorder in a Clinical Child and Adolescent Sample.
Jansen AG, Dieleman GC, Jansen PR, Verhulst FC, Posthuma D, Polderman TJC.
Psychiatric Behav Genet. 2020 Jul;50(4):203-212. doi: 10.1007/s10519-019-09965-8. Epub 2019 Jul 25. PMID: 31346826; PMCID: PMC7355275.

The predictive capacity of psychiatric and psychological polygenic risk scores for distinguishing cases in a child and adolescent psychiatric sample from controls
Jansen AG, Jansen PR, Savage JE, Kraft J, Skarabis N, Polderman TJC, Dieleman GC
J Child Psychol Psychiatry. 2021 Sep; 62(9): 1079–1089. Published online 2021 Apr 6. doi: 10.1111/jcpp.13370

Dankwoord

One does not simply complete a PhD project… veel dank voor iedereen die mij hier de afgelopen jaren in heeft gesteund. In het bijzonder:

Mijn promotoren Danielle en Frank; bedankt voor de vele kansen, lessen en kennis die ik van jullie heb gekregen. Frank, jouw bijdragen zijn vaak meer vanaf de zijlijn geweest maar daardoor niet minder waardevol en sturend voor het hele project. Bedankt dat je van begin tot eind zorg hebt gedragen voor het welzijn van de Binnenste Buiten data en de daaruit voortvloeiende publicaties. Danielle, dat ik onder jouw leiding mijn eerste stappen in mijn wetenschappelijke carrière heb kunnen zetten vind ik een groot voorrecht. Je interesse in de wetenschap en vooruitstrevende blik hebben mij geleerd altijd het onderste uit de kan te halen om de data zoveel mogelijk geheimen prijs te laten geven. Niet altijd makkelijk maar een waardevolle skill binnen en buiten de wetenschap.

Mijn copromotor Tinca; vanaf het eerste moment dat we contact hebben gehad in de zomer 2014 (waar blijft de tijd?) ben je echt altijd positief, ondersteunend en optimistisch geweest. Ik ben enorm dankbaar dat jij mijn dagelijkse begeleiding en daarmee een groot deel van mijn wetenschappelijk vorming op je hebt genomen. Ik vind je een geweldige en integere wetenschapper en een super fijn mens. Ik had me geen betere begeleider kunnen wensen al die jaren. Bedankt dat je in mooie én in moeilijke tijden een sturend lichtje bent geweest.

Mijn copromotor Gwen; de eerste jaren van mijn traject heb ik vooral in Amsterdam doorgebracht, daarin hadden wij iets minder contact. Ik ben blij dat dit in de tweede helft van het traject is veranderd. Je fijne manier van communiceren, aandacht voor de mens achter de student en natuurlijk je enorme medisch wetenschappelijke kennis hebben mijn publicaties en algemene kennis naar een hoger nivo gebracht. Bedankt dat je mij hebt begeleid op een af en toe hobbelige weg.

My promotion committee; I find it fortunate that I can finish my PhD project with inspiring scientists like you all. Thank you for reading and the discussion.

Mijn collega's aan het CTG en aan de afdeling kinder- en jeugdpsychiatrie van het Erasmus MC; bedankt voor de fijne gesprekken, gezellige uitjes en de algemeen plezierige kadans van de dagen die ik op de afdelingen heb gesleten. Een bijzondere shout-out naar *Anke H.*, de eerste weken, en alle weken erna, was ik niet zonder jou doorgekomen. *Philip J.*, bedankt voor het cleanen, filteren, nog een keer cleanen en filteren etc. etc. van de Binnenste Buiten data, het maken van mooie grafieken en het intikken van code op de command line wanneer ik het even niet meer wist, en dat alles zonder morren. Tenslotte een grote bedankt aan *Jeanne S.,* je scherpe blik heeft vele malen punten, comma's of scripting foutjes van een grotere aard weten bloot te leggen wat je altijd op een positieve manier wist te verwoorden, bedankt!

Mijn familie; jullie zijn mijn kracht en mijn zwakte. Dat we dit nog lang voor elkaar mogen zijn.

Mijn lieve vriendinnen; Meike, Charissa, Rosanne en Lisette. Bedankt dat jullie altijd een luisterend oor zijn geweest als de wetenschappelijke perikkelen me hoog zaten en dat jullie de mooie momenten met me hebben gevierd. Ik hoop dat jullie, wat mijn nieuwe avontuur ook moge worden, naast mij zullen blijven staan. Jullie zijn de chocolade sprinkels van het leven.

Rogier; wat fijn toch, dat met Marco, ook jij en de andere boebes mijn leven in zijn gekomen. Bedankt voor de gezellige familie dates en je positieve en vooral professionele blik op het leven.

My friend Trish; even though your contribution to this thesis is of a completely different nature, you deserve a shout-out! Thank you for walking the dance path beside me all those years, and for continuing to stand behind me when dance fell away. Your friendship and generosity have helped me become the confident woman I needed to be start this PhD journey.

Suhaila; thank you for teaching me to suck it up and put on my big girl panties. Dance and science are not so different so these and all other lesson are not lost nor forgotten.

Mijn schone ouders; bedankt dat jullie mij met raad en daad terzijde hebben gestaan. Jullie relativerende blik en hulp met onze lieve Mia wanneer ik ging werken hebben mijn leven enorm verrijkt. En, niet onbelangrijk, natuurlijk ook voor het opvoeden van een fijne zoon die inmiddels mijn fijne man is. Merci beaucoup!

Marco; lief, als ik in dit dankwoord alles moet beschrijven wat je hebt toegevoegd aan dit project en aan mijn leven wordt het echt een LTDR. Om het beknopt samen te vatten, ik ben dankbaar dat we al bijna 22 jaar(!) lief en leed delen, dat je mijn soms ontbrekende technisch en IT gerelateerde inzicht aanvult (mede binnen dit project, LOL, met onze eigen pizza sessies), en dat we de genetica uit de theorie naar de praktijk hebben gehaald, misschien wel ons mooiste kado aan elkaar. Jij bent mijn geluk.

Mia; lieve schat, alles wat ik doe is voor jou. Een hele grote dank-je-wel dat je mijn leven vol liefde, kleur en gezelligheid stopt. Het is een feest het leven met jou en papa te delen. Enta omri.

Dinosaurussen; jullie zijn geweldig!!


CNCR
CTGLAB