From genetic variants to biological pathways in neuropsychiatric traits

Hammerschlag, A.R.

2018

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

Chapter 1

General introduction

Neuropsychiatric disorders like schizophrenia, mood disorders, attention-deficit/hyper-activity disorder (ADHD) and autism spectrum disorder (ASD) place an enormous burden on patients, their family and society. The disorders typically start early in life and are often life-long, influencing the individuals' self-perception, productivity and ability to relate to others[1]. Unfortunately, no major breakthroughs in the treatment of these disorders occurred during the last few decades, and a substantial part of the patient group is resistant to all currently available treatments[1]. This lack of progress points to the complexity of the underlying biological mechanisms that induce these disorders. Improving our understanding of these mechanisms is highly needed to progress treatment development.

Family, twin and adoption studies have shown that neuropsychiatric disorders aggregate in families and are heritable, demonstrating a substantial role of genetic factors in their etiology[2]. Neuropsychiatric disorders fall into the category 'complex traits'. They are influenced by a combination of multiple genetic and environmental risk factors and possibly an interplay between the two. In addition, neuropsychiatric disorders are characterized by a polygenic nature, meaning that they are influenced by many genetic variants that likely belong to multiple biological pathways. These genetic variants have very small effects, which indicates that an individual variant is not sufficient to cause a disorder, but will contribute only for a very small part to the overall risk. Hence, it has been challenging to identify these variants, because the small effects are difficult to detect with standard statistical methods and current sample sizes. However, important advancements in statistical methods have been achieved in recent years, as will be explained in the next section. Apart from the challenge to detect the risk variants with small effects, mental disorders may show clinical heterogeneity. That is, different patients with the same disorder show different sets of symptoms and a different course of the disorder. It is likely that phenotypic heterogeneity reflects genotypic heterogeneity; i.e. quantitatively or qualitatively different genetic risk factors contribute to the disorder in different individuals[3]. Application of a diagnostic definition that is a univariate conceptualization of actually multiple biologically distinct subtypes, often results in loss of statistical power to detect risk variants[4].

In the past two decades, the research field of statistical genetics has developed rapidly with respect to new analysis methodology, and various methods have been applied to identify the genetic factors underlying complex traits[5]. The identification of genetic factors is an initial key step for unraveling the biological causes of neuropsychiatric traits, because identified genes will likely have specific functions and act through specific cellular mechanisms or biological pathways in particular tissue types. This information can highlight impaired biological processes that impact on neuropsychiatric disorders. For example, *CACNA1C* that codes for a calcium voltage-gated channel subunit has been identified in early GWASs of bipolar disorder, schizophrenia and major depressive disorder[6], suggesting involvement of cell-to-cell communication in the etiology of these disorders.

In this thesis I have applied several statistical genetic methods in order to identify genetic variants, genes and biological pathways that are involved in the etiology of neuropsychiatric traits, with the aim to obtain insight in biological causes underpinning these traits. Below, I summarize the different methods in the first section, continued by a section describing the characteristics of the studied disorders and traits and the accompanying chapters that are presented in this thesis.

1. Approaches to identify genetic risk factors and biological mechanisms

1.1 GWAS

During the last decade, Genome-Wide Association Studies (GWAS) have become a promising approach to identify genetic variants for complex traits[5]. GWAS is hypothesis-free, as no prior knowledge on candidate genes is required. In a typical GWAS a high number of genetic variants are genotyped in a large sample of individuals, of which part are affected by the disease or trait of interest (cases) and the other part are not affected (controls), or all included individuals have a continuous score on a trait. The genetic variants investigated in GWAS are single nucleotide polymorphisms (SNPs), which are a result of single point mutations leading to differences between individuals in nucleotides at specific genomic locations. Typically the SNPs tested in GWAS are common, i.e. occur at a frequency of at least 1% in the population, although recent increases in sample sizes have made it feasible to study SNPs of at least 0.1%. Risk variants for polygenic traits can be common because an individual variant will only slightly increase the risk to develop a trait. Therefore, the variant is not subject to selective pressure and can be maintained at a relatively high frequency in the population[7]. GWAS examines the frequency of SNPs between the groups of cases and controls (or the frequency of SNPs over a continuous score). GWAS relies on linkage disequilibrium (LD), a naturally

occurring correlation of nearby variants that is introduced by recombination. This correlation is reflected in the association statistics; SNPs in high LD will show similar P values. Hence, an identified variant is not necessarily the causal variant, because the statistical resolution is limited by the LD structure of the genome. By genotyping ~500,000 SNPs in populations with European descent the majority of common variants in the genome is covered[8]. The unobserved variants can be recovered by imputation; haplotypes inferred from multiple genotyped SNPs and haplotypes of a fully sequenced reference panel can together predict the genotypes of unobserved variants. Imputation can increase the SNP pool tested for association to several millions. Because this enormous amount of SNPs is tested together in a GWAS, correction for multiple testing is required in order to keep the type I error low. This correction is substantial ($\alpha = 5 \times 10^{-8}$), which results in the requirement of very large sample sizes in order to achieve adequate statistical power to identify genetic risk variants with small effects. The effect sizes for common genetic variants that have been observed for neuropsychiatric disorders are modest (e.g. OR = 1.06 - 1.20 for schizophrenia[9] and OR = 1.08 - 1.10 for ADHD[10]) and have been detected in studies including large samples (36,989 cases and 113,075 controls for schizophrenia and 20,183 cases and 35,191 controls for ADHD). Because a considerably larger proportion of the heritability of neuropsychiatric disorders can be explained by all common variants[9–11] – including the non-significant variants – it is expected that additional common variants with smaller effects are involved in the risk to develop neuropsychiatric disorders, which can be revealed by increasing sample sizes.

1.2 Gene-set analysis

Several strategies emerged to overcome the limited statistical power of current GWAS to identify the risk variants with small effects on traits. One of these strategies is gene-set analysis, which evaluates the joint effect of multiple SNPs that are present in multiple functionally related genes. It is based on the hypothesis that the numerous genetic variants that contribute to a polygenic trait aggregate in genes that share similar cellular functions. The underlying reasoning is that genes do not work in isolation; they participate in complex networks that regulate biological mechanisms. Because sets of genes are evaluated in one test, multiple testing correction is strongly reduced and statistical power increases compared to the single SNP method used in GWAS. Moreover, the biological relevance of single SNP associations is often difficult to interpret, while the association of a set of genes with a shared function directly generates a hypothesis about causal biological mechanisms[12].

1.3 Rare-variant analysis

Heritability estimates for complex traits that are based on all common variants tested in a GWAS typically explain only a modest fraction of the predicted genetic variance based on twin studies, implying additional contribution of other types of genetic variants. One of the suggested types of variants are rare genetic variants. Compared to common SNPs, rare variants located in exons – the coding parts of genes – may have greater potential to elucidate the biological mechanisms that play a role in complex traits[13]. For example, loss of function variants are exonic variants that result in the loss of function of a protein, and may have greater impact on the trait than other variants with less obvious biological consequences[14]. It is hypothesized that rare disease-associated variants have larger effects on traits as a result of purifying selection for these potentially deleterious mutations which keeps their frequency in the population low[15]. The contribution of rare variants to complex traits and their effect size spectrum is currently poorly understood, yet important to unravel the genetic architecture of complex traits. Because conventional genotyping chips do not capture rare variants, as they are a low-cost strategy to collect genetic information of common variants in large samples, the ExomeChip was designed to study rare variants in an exploratory and cost effective way[16]. The chip contains a dense amount of rare nonsynonymous SNPs. Nonsynonymous SNPs are genetic variants that result in an amino acid change in the protein sequence. Hence, identified variants give direct insight into the mechanism through which they may ascertain their effect on the trait.

It is presumed that rare variants have a larger phenotypic impact than common variants. However, given they are rare in the population (minor allele frequency < 1%), large sample sizes are required to detect them. The statistical power in rare variant analysis is in particular a function of minor allele frequency (MAF). To increase statistical power, a burden test or variance component test can be performed[17]. These tests aggregate the information across multiple rare variants that are present in a gene. Both approaches are promising when causal genes harbour multiple associated rare variants, however they are less benificial when these variants are sparse. Abundance of rare variants will expand when sample size increases.

1.4 Biological interpretation of GWAS hits

Single SNP associations of GWAS are difficult to translate to biological mechanisms for multiple reasons. First, LD makes inferences about the causal variant difficult, as already discussed above. Second, the identified variants often do not directly point to the causal gene because the majority is located in intergenic regions. Third, the effects of variants will typically be small and interpretation of biological consequences is as such difficult. Traditionally, identified variants for Mendelian diseases (diseases that follow simple mendelian patterns of

inheritance with causal variants that typically have large effects) were subjected to a battery of follow-up functional experiments to investigate the mechanisms through which they act on the disease. However, the many functional experiments are not achievable for the massive amount of associated variants found in GWAS for complex traits to date, and also not always applicable because of small effects or intergenic regions. Large-scale evaluation of functional impact of the GWAS output is therefore required to yield more insight into the underlying biology.

To interpret GWAS output, functional databases on the same whole-genome scale are of high value. Recent development of new molecular technologies enabled the availability of high-throughput sequencing platforms providing omics datasets that can be utilized for functional analyses of risk variants. During the last years, major datasets have been generated by the projects GTEx[18], providing expession quantitative trait loci (eQTLs) in multiple tissue types, and ENCODE[19] and Epigenome RoadMap[20], informing on regulatory elements and chromatin states. Because the majority of GWAS risk variants are noncoding but do affect expression levels by gene regulation[21], these datasets are crucial to map these variants to the genes they regulate and the relevant tissue types. Multiple studies have utilized these datasets and have reported that GWAS hits are highly enriched in regions of active chromatin such as promotors and enhancers in tissue types of interest[22]. In addition to the information from these expression and regulatory maps, coding SNPs can be annotated for functional consequences, for example by mutation prediction algorithms such as PolyPhen[23] or SIFT[24].

Based on information of these databases, genetic variants identified in GWAS can be functionally annotated to prioritize variants and genes. However, it remains challenging to pinpoint a small subset of important variants and genes out of these massive amounts of functional data. To this end, a variety of methods have been developed that merge the functional annotation of multiple databases to enable prioritization of genetic variants and genes, for example CADD (Combined Annotation Dependent Depletion)[25], ANNOVAR (Annotate Variation)[26] and FUMA GWAS (Functional Mapping and Annotation of Genome-Wide Association Studies)[27]. Prioritized genes can in turn be placed in biological contex using data on tissue expression (e.g. GTEx[18]) or biological pathways and cellular mechanisms (e.g. MSigDB (The Molecular Signatures Database)[28]) using tools such as Prixe Fixe[29] and FUMA. Merging the different functional datasets to study the association results of GWAS will aid in identifying potential causal mechanisms contributing to neuropsychiatric disorders.

2. Neuropsychiatric disorders and traits studied in this thesis

2.1 ADHD

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common neurodevelopmental disorders of childhood, characterized by developmentally inappropriate inattentiveness, and/or increased impulsivity and hyperactivity[30]. The world-wide prevalence of ADHD in childhood is ~5%[31,32]. While its onset is during childhood, the symptoms can persist into adulthood with a prevalence of ~3%[33]. High heritability estimates of ~75-90%[34] indicate an important genetic contribution to ADHD risk. However, the identification of genes turned out to be challenging. One of the explanations is the polygenic character of the disorder. The first genome-wide association meta-analysis on ADHD including 896 cases, 2,455 controls and 2,064 trios (child with ADHD and parents) has not yielded genome-wide significant hits[35], implying that the common variants that contribute to ADHD have small effects and larger sample sizes are required to detect them. We applied two approaches with the aim to identify the first genomic loci and biological mechanisms contributing to ADHD risk. These are described in the first two studies of this thesis.

Chapter 2 GWAS meta-analysis of ADHD symptom scores

In the first chapter on ADHD we performed a genome-wide association meta-analysis of ADHD symptoms in population-based pediatric cohorts. Latent class models implied that children diagnosed with ADHD are at the extreme of a continuum of inattentive and hyperactive behaviors[36]. Hence, continuous measures of ADHD symptoms have been suggested to aid in the gene finding for ADHD[37]. Indeed, results from polygenic risk score (PRS) analyses[38–40] and twin studies[41,42] suggested strong overlap between genetic factors for continuous measures of ADHD symptoms in the general population and clinical diagnosis of ADHD. While clinical diagnosed ADHD cohorts with genetic data are not abundant, many population-based pediatric cohorts have collected both continuous ADHD symptom scores and genotype data. Studying continuous ADHD-related behaviors can be a powerful statistical approach in population based studies because it uses the full spectrum of phenotypic variation and hence provides the opportunity to include a much larger sample compared to studying only the extremes of the spectrum. We performed the first genome-wide association meta-analysis on ADHD symptoms in 17,666 children from nine population-based cohorts with the aim to identify genetic variants that contribute to ADHD symptoms. In addition, we investigated if ADHD symptom scores and ADHD diagnoses show indeed a shared genetic background by analyzing the concordance in the effects of SNPs and genes of our meta-analysis with the results of a genome-wide meta-analysis

based on a ADHD case-control study and by estimating their genetic correlation.

Chapter 3 Gene-set analysis of synaptic functions in ADHD

In the second chapter on ADHD we applied gene-set analysis to improve the statistical power to identify genetic factors contributing to ADHD risk. We focused on the synaptic machinery in this study, because its role in ADHD has been suggested by the functions of genes harboring the top hits of the first ADHD GWASs[43–45], and by genes reported in candidate gene studies that are often expressed in the synapse[46]. We performed the analysis on data from the largest clinically diagnosed ADHD sample from the Psychiatric Genomics Consortium at that time, including 2,960 ADHD cases and 4,519 controls.

2.2 Genetic overlap between ADHD, ASD, bipolar disorder, major depressive disorder, and schizophrenia

Psychiatric disorders are classified based on their observed symptoms and disease course. However, symptoms highly overlap, and comorbidity between disorders is the rule rather than the exception. It is therefore debated to what extent the disorders are biologically distinct entities. Indeed, genetic overlap has been found for multiple psychiatric disorders[47,48]. The observation that a genetic variant affects multiple phenotypes is called 'pleiotropy'. The true nature of the pleiotropy is currently unknown. Different scenarios are possible, for example a genetic variant might influence two traits through independent mechanisms, or a causal relationship might exist in which the effect of the variant on the second trait is mediated through its effect on the first trait[49]. Because psychiatric disorders phenotypically overlap, it is likely that the pleiotropy points to shared biological mechanisms that act across psychiatric disorders.

Chapter 4 Gene-set analysis of multiple psychiatric disorders

In this chapter we applied gene-set analysis to investigate shared biological mechanisms across five major psychiatric disorders (ADHD, ASD, bipolar disorder, major depressive disorder, and schizophrenia). A gene-set study from 2015 investigated the contribution of gene sets derived from publicly available databases and identified several gene sets involved in histone methylation, immune and neuronal signaling, and the synapse across the three adult-onset disorders schizophrenia, bipolar disorder and major depressive disorder[50]. We exploited the larger sample sizes that emerged recently for GWAS – which resulted in enormous increases in the identified associated genetic loci for psychiatric disorders – and extended previous research by testing gene sets derived not only from publicly available

databases, but also from expert curation and tissue expression. The expert-curated gene sets are of interest because several studies identified expert-curated gene sets for single psychiatric disorders[51] which makes them excellent candidates for explaining genetic overlap between psychiatric disorders. Furthermore, we aimed to investigate not only overlap in biological mechanisms between adult-onset disorders, but overlap with childhood-onset disorders ADHD and ASD as well.

2.3 Addiction-related behaviors

Both alcohol consumption and tobacco use belong to the world's leading health risks and account together for more morbidity and mortality in Western countries than any other single risk factor or health outcome[52]. Substance use behaviors are strongly linked to neuropsychiatric disorders such as alcohol use disorders and major depressive disorder[53]. Moderate to high heritability estimates have been reported for both behaviors[54], and multiple genetic variants have been identified in candidate gene studies[55] and GWASs[56,57]. Most prominent are the findings of nicotine and alcohol metabolism genes (e.g. *ADH1B, ALDH2, CYP2B6*) and nicotinic receptor genes (e.g., the cluster *CHRNA5-CHRNA3-CHRNB4*). Despite the successes of these studies, the variance in addiction-related behaviors can only partly be explained by the identified genetic loci. One of the explanations for this incomplete genetic explanation is that rare variants affect these behaviors as well. A role for rare variants has been indicated by target sequencing of known loci associated with addiction behaviors, especially within nicotinic receptor gene clusters[58–60] and alcohol metabolism genes[61,62]. The role of rare variants in other genomic loci has yet to be determined.

Chapter 5 Exome chip meta-analysis of alcohol consumption and tobacco use

This chapter describes an exome chip meta-analysis to explore the role of rare exonic variants in alcohol consumption and tobacco use. Compared to common SNPs used in GWAS, rare exonic variants may have greater potential to increase our understanding of the biological mechanisms involved in addiction[13]. As described above, exonic variants that result in the loss of normal functioning of the protein (loss of function variants) may have greater impact on the trait than other variants with less clear biological consequences. The analysis of rare variants presents analytical challenges because statistical power is a function of MAF, and rare variants usually have a MAF < 0.01. Therefore large samples are essential. We aimed to increase statistical power by meta-analyzing eight Dutch cohorts which resulted in 12,466 and 7,432 individuals for alcohol consumption and tobacco use, respectively. In addition, we applied the burden test to increase statistical power to identify causal genes.

2.4 Insomnia Disorder

Insomnia Disorder is one of the most common mental disorders with prevalence estimates ranging from 10% in adults to 22% in the elderly[63,64]. It is characterized by lasting problems falling asleep or waking up in the night or early morning, that have subjective repercussions for daytime functioning. Not only are the direct costs of insomnia extremely high[65], it is also the primary risk factor for depression[66] and contributes to the risks of cardiovascular disease[67–71] and obesity[72]. A recent longitudinal twin study replicated a moderate heritability for males (38%) and a significantly higher heritability for females (59%)[73], suggesting that (partly) different genetic factors contribute to insomnia in males and females. Identifying the responsible genetic risk variants could provide important clues to the biological pathways involved in insomnia vulnerability. However, as compared to other complex genetic traits, strikingly few genetic studies have been conducted to identify susceptibility genes for insomnia. The available GWASs for insomnia have not identified any consistent genetic risk variants, most likely due to small sample sizes (N < 5,000)[74]. In the last study described in this thesis we aim to identify the genetic risk factors and biological pathways contributing to the etiology of insomnia.

Chapter 6 GWAS of insomnia complaints

In this chapter we used GWAS to identify genomic loci associated with insomnia complaints in a large population-based cohort, the UK Biobank, including 113,006 individuals with phenotype and genotype data. Because population-based cohorts generally do not have information on clinical diagnosis of Insomnia Disorder, we used a question on experiencing trouble falling asleep or waking up in the middle of the night, which we validated as a good proxy for Insomnia Disorder. This gave us the opportunity to substantially increase sample size compared to clinically diagnosed cohorts. Because the single SNP results of a GWAS are difficult to translate to biological mechanisms, further analyses of the GWAS output were applied to yield more insight into the biological underpinnings of Insomnia Disorder. These additional analyses involved gene-based and gene-set analysis, functional annotations, tissue enrichment tests and network analysis. In addition, we investigated if the overlap with other diseases and traits seen at the phenotypic level could be explained by shared genetic risk factors.

References

1. Akil, H. *et al.* The Future of Psychiatric Research: Genomes and Neural Circuits. *Science* **327,** 1580–1581 (2010).
2. Kendler, K. S. What psychiatric genetics has taught us about the nature of psychiatric illness and what is left to learn. *Mol. Psychiatry* **18,** 1058–1066 (2013).
3. McClellan, J. & King, M.-C. Genetic heterogeneity in human disease. *Cell* **141,** 210–217 (2010).
4. van der Sluis, S., Posthuma, D. & Dolan, C. V. TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genet.* **9,** e1003235 (2013).
5. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* **101,** 5–22 (2017).
6. Green, E. K. *et al.* The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia. *Mol. Psychiatry* **15,** 1016–1022 (2010).
7. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273,** 1516–1517 (1996).
8. Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6,** 95–108 (2005).
9. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).
10. Demontis, D. *et al.* Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. *bioRxiv* (2017). doi:https://doi.org/10.1101/145581
11. The Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. 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 (2017).
12. Schadt, E. E. Molecular networks as sensors and drivers of common human diseases. *Nature* **461,** 218–223 (2009).
13. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).
14. Sveinbjornsson, G. *et al.* Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat. Genet.* **48,** 314–317 (2016).
15. Kiezun, A. *et al.* Exome sequencing and the genetic basis of complex traits. *Nat. Genet.* **44,** 623–630 (2012).
16. Grove, M. L. *et al.* Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One* **8,** e68095 (2013).
17. Bomba, L., Walter, K. & Soranzo, N. The impact of rare and low-frequency genetic variants in common disease. *Genome Biol.* **18,** 77 (2017).
18. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348,** 648–660 (2015).
19. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
20. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015).
21. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94,** 559–573 (2014).
22. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169,** 1177–1186 (2017).
23. Ramensky, V., Bork, P. & Sunyaev, S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res. 30, 3894–3900 (2002).
24. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31,** 3812–3814 (2003).
25. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46,** 310–315 (2014).
26. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).

27. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. FUMA: Functional mapping and annotation of genetic associations. *bioRxiv* (2017). doi:https://doi.org/10.1101/110023
28. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
29. Tasan, M. *et al.* Selecting causal genes from genome-wide association studies via functionally coherent subnetworks. *Nat. Methods* **12,** 154–159 (2015).
30. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (5th ed.)*. (American Psychiatric Publishing, 2013).
31. Polanczyk, G., de Lima, M. S., Horta, B. L., Biederman, J. & Rohde, L. A. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *Am. J. Psychiatry* 164, 942–948 (2007).
32. Polanczyk, G. V, Willcutt, E. G., Salum, G. A., Kieling, C. & Rohde, L. A. ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. *Int. J. Epidemiol.* **43,** 434–442 (2014).
33. Simon, V., Czobor, P., Balint, S., Meszaros, A. & Bitter, I. Prevalence and correlates of adult attention-deficit hyperactivity disorder: meta-analysis. *Br. J. Psychiatry* **194,** 204–211 (2009).
34. Lichtenstein, P., Carlstrom, E., Rastam, M., Gillberg, C. & Anckarsater, H. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. *Am. J. Psychiatry* **167,** 1357–1363 (2010).
35. Neale, B. M. *et al.* Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. J. Am. Acad. Child Adolesc. *Psychiatry* **49,** 884–897 (2010).
36. Lubke, G. H., Hudziak, J. J., Derks, E. M., van Bijsterveldt, T. C. E. M. & Boomsma, D. I. Maternal ratings of attention problems in ADHD: evidence for the existence of a continuum. *J. Am. Acad. Child Adolesc. Psychiatry* **48,** 1085–1093 (2009).
37. Hawi, Z. *et al.* The molecular genetic architecture of attention deficit hyperactivity disorder. *Mol. Psychiatry* **20,** 289–297 (2015).
38. Groen-Blokhuis, M. M. *et al.* Attention-deficit/hyperactivity disorder polygenic risk scores predict attention problems in a population-based sample of children. *J. Am. Acad. Child Adolesc. Psychiatry* **53,** 1123–1129.e6 (2014).
39. Martin, J., Hamshere, M. L., Stergiakouli, E., O'Donovan, M. C. & Thapar, A. Genetic risk for attention-deficit/hyperactivity disorder contributes to neurodevelopmental traits in the general population. *Biol. Psychiatry* **76,** 664–671 (2014).
40. Stergiakouli, E. *et al.* Shared genetic influences between attention-deficit/hyperactivity disorder (ADHD) traits in children and clinical ADHD. *J. Am. Acad. Child Adolesc. Psychiatry* **54,** 322–327 (2015).
41. Polderman, T. J. C. *et al.* Across the continuum of attention skills: a twin study of the SWAN ADHD rating scale. *J. Child Psychol. Psychiatry* **48,** 1080–1087 (2007).
42. Larsson, H., Anckarsater, H., Rastam, M., Chang, Z. & Lichtenstein, P. Childhood attention-deficit hyperactivity disorder as an extreme of a continuous trait: a quantitative genetic study of 8,500 twin pairs. *J. Child Psychol. Psychiatry.* **53,** 73–80 (2012).
43. Lesch, K. P. *et al.* Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. *J. Neural Transm.* **115,** 1573–1585 (2008).
44. Franke, B., Neale, B. M. & Faraone, S. V. Genome-wide association studies in ADHD. *Hum. Genet.* **126,** 13–50 (2009).
45. Poelmans, G., Pauls, D. L., Buitelaar, J. K. & Franke, B. Integrated genome-wide association study findings: identification of a neurodevelopmental network for attention deficit hyperactivity disorder. *Am. J. Psychiatry* **168,** 365–377 (2011).
46. Gizer, I. R., Ficks, C. & Waldman, I. D. Candidate gene studies of ADHD: a meta-analytic review. *Hum. Genet.* **126,** 51–90 (2009).
47. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45,** 984–994 (2013).

48. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381,** 1371–1379 (2013).
49. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* **48,** 709–717 (2016).
50. The Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18,** 199–209 (2015).
51. Sullivan, P. F. & Posthuma, D. Biological pathways and networks implicated in psychiatric disorders. *Curr. Opin. Behav. Sci.* **2,** 58–68 (2015).
52. Ezzati, M., Lopez, A. D., Rodgers, A., Vander Hoorn, S. & Murray, C. J. L. Selected major risk factors and global and regional burden of disease. *Lancet (London, England)* **360,** 1347–1360 (2002).
53. Shield, K. D., Parry, C. & Rehm, J. Chronic Diseases and Conditions Related to Alcohol Use. *Alcohol Res.* **35,** 155–171 (2014).
54. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47,** 702–709 (2015).
55. Eng, M. Y., Luczak, S. E. & Wall, T. L. ALDH2, ADH1B, and ADH1C genotypes in Asians: a literature review. Alcohol Res. *Health* **30,** 22–27 (2007).
56. Tabacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42,** 441–447 (2010).
57. Saccone, N. L. *et al.* Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet.* **6,** e1001053 (2010).
58. Yang, J. *et al.* The contribution of rare and common variants in 30 genes to risk nicotine dependence. *Mol. Psychiatry* **20,** 1467–1478 (2015).
59. Thorgeirsson, T. E. *et al.* A rare missense mutation in CHRNA4 associates with smoking behavior and its consequences. *Mol. Psychiatry* **21,** 594–600 (2016).
60. Olfson, E. *et al.* Rare, low frequency and common coding variants in CHRNA5 and their contribution to nicotine dependence in European and African Americans. *Mol. Psychiatry* **21,** 601–607 (2016).
61. Way, M. *et al.* Genetic variants in or near ADH1B and ADH1C affect susceptibility to alcohol dependence in a British and Irish population. *Addict. Biol.* **20,** 594–604 (2015).
62. Zuo, L. *et al.* Rare ADH variant constellations are specific for alcohol dependence. *Alcohol Alcohol* **48,** 9–14 (2013).
63. Zhang, B. & Wing, Y.-K. Sex differences in insomnia: a meta-analysis. *Sleep* **29,** 85–93 (2006).
64. Morin, C. M. *et al.* Insomnia disorder. *Nat. Rev. Dis. Prim.* **1,** 15026 (2015).
65. Gustavsson, A. *et al.* Cost of disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* **21,** 718–779 (2011).
66. Baglioni, C. *et al.* Insomnia as a predictor of depression: A meta-analytic evaluation of longitudinal epidemiological studies. *Journal of Affective Disorders* **135,** 10–19 (2011).
67. Palagini, L. *et al.* Sleep Loss and Hypertension: A Systematic Review. *Curr. Pharm. Des.* **19,** 2409–2419 (2013).
68. Mallon, L., Broman, J. E. & Hetta, J. Sleep complaints predict coronary artery disease mortality in males: A 12-year follow-up study of a middle-aged Swedish population. *J. Intern. Med.* **251,** 207–216 (2002).
69. Nilsson, P. M., Nilsson, J. Å., Hedblad, B. & Berglund, G. Sleep disturbance in association with elevated pulse rate for prediction of mortality - Consequences of mental strain? *J. Intern. Med.* **250,** 521–529 (2001).
70. Schwartz, S. W. *et al.* Are sleep complaints an independent risk factor for Myocardial infarction? *Ann. Epidemiol.* **8,** 384–392 (1998).
71. Clark, A., Lange, T., Hallqvist, J., Jennum, P. & Rod, N. H. Sleep impairment and prognosis of acute myocardial infarction: a prospective cohort study. *Sleep* **37,** 851–858 (2014).
72. Hargens, T. A., Kaleth, A. S., Edwards, E. S. & Butner, K. L. Association between sleep disorders, obesity, and exercise: A review. *Nature and Science of Sleep* **5,** 27–35 (2013).

73. Lind, M. J., Aggen, S. H., Kirkpatrick, R. M., Kendler, K. S. & Amstadter, A. B. A Longitudinal Twin Study of Insomnia Symptoms in Adults. *Sleep* **38,** 1423–1430 (2015).

74. Byrne, E. M. et *al.* A genome-wide association study of sleep habits and insomnia. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **162B,** 439–451 (2013).