Discovering the genetic architecture of the mind
Karlsson Linnér, R.

2019

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

Link to publication in VU Research Portal

citation for published version (APA)

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address: vuresearchportal.ub@vu.nl
Introduction

“Why do people believe that there are dangerous implications of the idea that the mind is a product of the brain, that the brain is organized in part by the genome, and that the genome was shaped by natural selection?”

Steven Pinker

“Everything we do, every thought we’ve ever had, is produced in the human brain. But exactly how it operates remains one of the biggest unsolved mysteries, and it seems the more we probe its secrets, the more surprises we find.”

Neil deGrasse Tyson
Background

Humans have evolved large and complex brains as an adaption to cognitive, environmental, and social challenges. Throughout the life course, our brains develop and that process influences how we think, feel, and act. People are in many ways similar, but also display a range of individual differences in their psychology and behavior. For example, the personal attitude or willingness towards risk taking is a psychological characteristic, studied in this thesis, that varies substantially in humans. In short, risk refers to the possibility of experiencing undesirable outcomes, which is usually accepted in exchange for the prospect of some benefit or reward. Some individuals are more willing to engage in risky activities, such as driving over the speed limit, while others behave more cautiously. Individual differences in risk taking are associated with many other traits, such as health-risk behaviors, financial decision-making, and occupational choice. Overall, it is intriguing to study the determinants of individual differences in mental traits as a way to understand ourselves as humans, and because of their importance for various health and life outcomes.

For the purpose of illustration, similar to the works of other scholars, this introductory chapter refers to psychology and behavior, together with mental traits, interchangeably as a general concept, while the section Introduction to the main phenotypes introduces the primary phenotypes studied in this thesis. Ultimately, mind and consciousness arise from the brain, and observable psychology and behavior come about foremost as a result of the brain’s circuits, processes, and functions. Conceptually, I consider the phenotypes under study to be distinct, but yet related, measures of psychology and behavior that can be quantified, and thus, studied empirically with statistical methods. This thesis is an investigation of how genetic factors influence these particular phenotypes, but it will also treat a few environmental factors. But before we get to that, I will first review what is known about the heritability of psychology and behavior.

The heritability of psychology and behavior

Individual differences in mental traits are the result of both genetic and environmental influences. Decades of research in behavior genetics, the scientific study of the genetic and environmental underpinnings of human psychology and behavior, have convincingly established that practically all mental traits are heritable to some degree. Heritability estimates of psychometric measures of psychology and behavior, a measure of the proportion of phenotypic variability attributable to genetic factors, range from moderate to substantial, and are on average about 50%. Particular psychiatric disorders are highly heritable, such as attention-deficit hyperactivity disorder (ADHD), bipolar disorder, and schizophrenia, and genetic factors explain about 75–80% of the variation in these traits. In contrast, other mental traits are less heritable, such as risk preferences, depression, and educational attainment, and roughly 20–50% of their variation is attributable to genetic factors. Thus, there is overwhelming evidence in support of the premise that psychology and behavior are influenced by a non-negligible genetic component, and it appears as if both genes and environment contribute substantially to individual differences in such traits.

Based on recurring empirical findings, Turkheimer (2000) proposed three general “laws” of behavior genetics. Beyond the first insight that all behaviors are heritable to some degree, the second insight is that the common effect of being raised in the same family is typically weaker than the effect of genetic factors. It is known that mental traits and disorders cluster in families and empirical estimates suggest that this clustering is primarily caused by a shared genetic liability, rather than the shared environment. The third insight is that idiosyncratic environmental factors, unique to each individual, often account for a sizable share of the
phenotypic variation. However, such factors are frequently unobserved in studies of heritability, and it is common to assume that idiosyncratic effects are independent of genetic differences and to model those as part of the stochastic residual variation.

Beyond the three “laws” of behavior genetics, there are two more recent insights that deserve to be mentioned because of their implications for the research strategy and motivation of this thesis. First, converging evidence suggests that the genetic effects on mental traits and disorders are shared across traits, which can be referred to as pleiotropy or genetic overlap. It appears as if the genetic liability that clusters in families not only predisposes family members towards a particular mental disorder, but it also increases the susceptibility towards others. Twin and family studies, as well as more recent investigations based on molecular genetic data, consistently estimate moderate to substantial genetic overlap across many pairs of mental traits and disorders. Yet, most of the particular genetic variants and regions that are responsible for the observed coheritability have not been identified.

Second, many studies suggest that the genetic influences on complex traits are polygenic, and thus, distributed across many genetic variants. Overall, the aggregate influence of common variants can be substantial, but their individual effects appear to be smaller and distributed across a much larger set of variants than was previously thought, and maybe even more so for psychology and behavior than for many other complex traits. This regularity has been dubbed the fourth “law” of behavior genetics by Chabris et al. (2015). As an example, the largest effect discovered in the first genome-wide association study (GWAS, see below and Chapter 2 – Supplementary Methods) on educational attainment explains only about 0.02% of the phenotypic variation, i.e., a fiftieth of a percent. It is the exception rather than the rule to find common variants with large effects on mental traits. Taken together, these observations conform with an infinitesimal effect-size model. That model assumes that the additive genetic influence is spread over a very large number, and maybe even all, genetic variants of which many have near-zero effects. Recently, it has been proposed that potentially all genetic variants involved in a trait-relevant tissue could exert some small yet important effect even when there is no obvious biological connection. This hypothesis has been named the “omnigenic model” and it is based on the premise that regulatory gene networks are deeply interconnected. In summary, converging evidence suggests that there is much genetic overlap across mental traits, and that such traits are highly polygenic rather than monogenic.

**Genome-wide association studies**

Recent technological advancements have revolutionized the discovery of the genetic architecture of human traits and disorders, defined in the section The genetic architecture of complex traits. Modern genotyping technologies first made it possible to measure particular genetic variants, and later, millions of variants across the genome. When the genetic factor is no longer latent, as it is in the classical twin design, it becomes possible to actually discover which particular genetic variants are responsible for the trait inheritance. The GWAS method is a modern approach used to identify genetic associations, and GWAS typically study the most occurring form of genetic variation, the single-nucleotide polymorphism (SNP; pronounced as “snip”). A SNP is defined as a single base pair in the genetic sequence that varies across individuals. The success of GWAS is strongly dependent on sample size. Theoretically, a sample size of roughly 200,000 individuals is required to attain 80% power to detect an effect of 0.02% at genome-wide significance, similar to the effect mentioned above. Nonetheless, when performed in large samples, GWAS can identify common variants with high posterior probabilities even though their effects are small.

---

b Genome-wide significance is typically declared when the association P-value is less than $5 \times 10^{-8}$. 
Since its introduction, GWAS have successfully identified robust associations with hundreds of genetically complex traits\textsuperscript{43,45,49}. But even so, the vast majority of the variants responsible for the narrow-sense heritability of brain structure, and the resulting variation in observable behavior, have not yet been identified\textsuperscript{30,31,52}. The difference between the narrow-sense heritability and that which can be attributed to genome-wide significant associations is called the “missing heritability”\textsuperscript{30,53}. The missing heritability can be divided into two parts; the “hiding heritability”, which is expected to be uncovered as a function of increasing GWAS sample size, and the “still-missing heritability”, which will not be revealed by traditional GWAS\textsuperscript{53}. Instead, the still-missing heritability is expected to be attributed to other genetic factors that are typically not captured by traditional GWAS, such as rare and structural variation\textsuperscript{40,41,54}. Methods that use millions of common SNPs across the genome to estimate the so-called “SNP heritability” are suggestive of the respective sizes of the hiding and still-missing heritability\textsuperscript{33,55}. As described below, Chapters 2–3 perform large-scale GWAS of a couple of mental traits that, upon publication, had not yet been studied in samples as large as those reported here.

\textbf{Psychosocial experiences and DNA methylation}

Epigenetics is the study of various molecular mechanisms that influence genetic processes, such as the regulation of gene expression\textsuperscript{56–58}. In contrast to the relative constancy of the genetic sequence\textsuperscript{59}, epigenetic states vary across cells and tissues, and are affected by environmental influences\textsuperscript{60}. In behavioral epigenetics, DNA methylation appears to be the most studied mechanism\textsuperscript{61}. Methylation refers to the addition of a methyl group on “top of” a genetic base pair, and it is known to play an important role in cellular development and processes\textsuperscript{56,62,63}. In mammals, methylation occurs almost exclusively at the cytosine nucleotide in cytosine-guanine pairs\textsuperscript{63} connected by a phosphate link, referred to as Cpg probes. Importantly, it is of scientific and societal interest to study methylation because of the repeated observation that it is associated with various medical conditions, such as autoimmune disease, cancers, and neuropsychiatric disorders\textsuperscript{62,64,65}. Microarrays have recently been introduced that measure methylation at hundreds of thousands of Cpg probes across the genome\textsuperscript{66}. A common approach to identify associations with methylation is the epigenome-wide association study (EWAS), an extension of the GWAS method with some additional considerations\textsuperscript{67} (see Chapter 4 – Supplementary Methods).

Interestingly, it has been proposed that environmental factors may shape behavior by influencing gene expression\textsuperscript{12}, and a conceivable mechanism could be the methylation of genes involved in neural development\textsuperscript{56,61,68,69}. However, this is a new and highly active area of research and it has not been concluded what genes and environmental factors have the capability to affect behavior in such a way\textsuperscript{65}. Interestingly, methylation has been offered as a biological explanation of the well-established relationship between adverse psychosocial\textsuperscript{6} experiences in childhood with mental and physical disorders later in life\textsuperscript{70–74}. It has even been proposed that methylation could be an adaptive mechanism that later turns maladaptive\textsuperscript{71}. For ease of discussion, this thesis henceforth refers to this category of viewpoints as “the hypothesis that psychosocial experiences influence disease via methylation”.

Notably, a range of adverse life events, ranging from trauma\textsuperscript{75}; parental attachment\textsuperscript{76}, abuse\textsuperscript{77}, and neglect\textsuperscript{78,79}; to socioeconomic hardship\textsuperscript{70,80–82}, have been found associated with methylation. The notion that biological embedding of psychosocial stress, caused by unfavorable socioeconomic position, would cause behavioral changes and medical conditions via methylation appears to be an appealing and fascinating explanation for observed health

---

\textsuperscript{1} Psychosocial – “Relating to the interrelation of social factors and individual thought and behaviour.”\textsuperscript{181}
Discovering the genetic architecture of the mind

However, it should be noted that studies on this hypothesis are not free from criticism, and great care must be taken in the choice of study design. On the contrary, it has been proposed that chemical influences with known biological effects must first be convincingly ruled out before prioritizing the correlation of psychosocial experiences with methylation as an explanation to medical conditions. Importantly, there are several biologically proximate exposures that have particularly strong and long-lasting influences on both methylation and medical conditions, such as alcohol consumption, diet and obesity, and smoking, in particular if experienced prenatally. These risk factors are known to correlate with socioeconomic adversity and most, if not all, of the aforementioned adverse life experiences, such as child maltreatment. Thus, there is a high risk that the positive findings reported in observational studies of this hypothesis are the result of confounding and/or omitted variable bias. That particular criticism has already been raised because most of the aforementioned studies do not convincingly control for many important confounders. Similarly, it has been observed that there is a tendency to infer causality in observational studies without the use of a causal research design. Notwithstanding these critiques, the number of studies that investigate the hypothesis that psychosocial experiences influence disease via methylation is growing, and concerningly, many are performed in small samples.

In this thesis, Chapter 4 performs a large-scale EWAS on educational attainment, a correlate of cognitive function and personality, which is also a major life experience that occurs day after day over many years of a person’s life. Because education is experienced over a substantial period of childhood, adolescence, and early adulthood it is not unreasonable to think that its variation could be quite strongly associated with methylation in the circumstance that life experiences would indeed influence methylation. In addition, educational attainment could be considered a good test case of that hypothesis because it can be studied in very large samples. On the other hand, because educational attainment, similar to adverse psychosocial experiences, could be considered biologically distal compared to biologically proximate confounders, such as alcohol consumption, smoking, and other health-risk behaviors, it is also conceivable that its association with methylation could be much weaker or non-existent when such factors are accounted for.

Studies of candidate-genes and their interactions with the environment

Prior to GWAS, it was more common to study prioritized candidate genes. In contrast, GWAS test millions of genetic variants across the genome for association, with little to no prioritization. Already in the early 2000s, the candidate-gene approach had received much criticism, primarily because of inadequate statistical power, but also other concerns, such as undisclosed multiple-testing and population stratification. Thus, while GWAS could be criticized for having a high false-negative rate, candidate-gene studies in psychology have instead been criticized for an unacceptable false-positive rate. Similarly, many of the same critiques apply to the popular study of candidate gene-environment interactions, and such studies must be undertaken with great care. Disconcertingly, many candidate gene-environment interaction studies are performed with improper control for potential confounders, and convincing replication of reported interactions is scarce in psychology research. Therefore, to further investigate and raise awareness of the issues surrounding these studies, Chapter 5 performs a critical review and meta-analysis of a particular candidate gene-environment interaction. Specifically, that study investigates the reported literature that has tested whether antisocial behavior is associated with an interaction between stressful life events and 5-HTTLPR, a controversial, yet frequently hypothesized
candidate gene\textsuperscript{34,111}. Although \textit{5-HTTLPR} is viewed with quite some skepticism it still receives much attention in recent gene-environment interaction studies\textsuperscript{112–114}. Remarkably, a couple of notorious candidate genes\textsuperscript{108,110,115,116}, including \textit{5-HTTLPR}, appear to have a renaissance in behavioral epigenetics, see e.g., refs. \textsuperscript{73,76,78,79,100}.

\textbf{Problem statement}

The objective of this thesis is to contribute to the on-going efforts in behavior genetics and related fields that aim to unravel the genetic, biological, and environmental basis of complex traits. The primary measures of psychology and behavior studied in this thesis, henceforth referred to as the “main phenotypes”, are the following:

- **Chapter 2.** General risk tolerance, adventurousness, and the four risky behaviors: (1) automobile speeding propensity, (2) drinks per week, (3) ever smoker, and (4) number of sexual partners and the first principal component (PC) of the four risky behaviors;
- **Chapter 3.** Subjective well-being, depressive symptoms, and neuroticism;
- **Chapter 4.** Educational attainment;
- **Chapter 5.** Antisocial behavior.

The main phenotypes have been extensively studied in different fields with varying research traditions, for example in economics, epidemiology, psychology, and the social sciences. At the outset of this thesis, there were few to no genome-wide significant associations reported for the subset of the main phenotypes under study in Chapters 2 and 3, and larger genetic studies are needed to accelerate the discovery of their genetic architectures\textsuperscript{30,49,52}. Similarly, larger studies could be considered required to further investigate whether life experiences are indeed associated with methylation (Chapter 4), as well as the hypothesis that antisocial behavior is associated with an interaction between stressful life events and \textit{5-HTTLPR} (Chapter 5).

Because underpowered studies have a high chance of finding false positives and inflated effect-sizes estimates it is important that the studies in this thesis are adequately powered\textsuperscript{50,110,117}. To the best of my knowledge, the primary genetic analyses were upon publication the largest for these phenotype-method pairs. In Chapters 2–3, the GWAS samples range from about 160,000 to almost 1 million individuals. In Chapter 4, the EWAS includes 10,767 individuals, which places it among the largest EWAS of any phenotype. Lastly, Chapter 5 is the largest investigation of that particular gene-environment interaction, \( n = 7,680 \), as it is a meta-analysis of the previous literature testing that hypothesis. In summary, the studies reported in this thesis are all performed in large samples with molecular genetic data.

\textbf{Purpose and research questions}

The central purpose of this thesis is to investigate the genetic architecture of the main phenotypes. For that purpose, genetic variants will be tested for association using statistical methods from complex trait genetics applied on molecular genetic data. In more detail, the primary purpose of Chapters 2–3 is to identify associations between SNPs and general risk tolerance, adventurousness, and the four risky behaviors and their first principal component; as well as with subjective well-being, depressive symptoms, and neuroticism. Subsequently, the results will be used in a range of genetic and bioinformatic follow-up analyses with the aim to identify additional aspects of their genetic architectures, as well as biological pathways and mechanisms. The purpose of Chapter 5 is to identify associations between educational attainment and the methylation of CpG probes. The purpose of Chapter 6 is to critically review
and meta-analyze all previous studies that have tested whether antisocial behavior is associated with a candidate gene-environment interaction between adverse life events and 5-HTTLPR.

The thesis aims to answer the following main research questions:

1. a. Do GWAS in hundreds of thousands of individuals identify robustly associated SNPs with the main phenotypes studied in Chapters 2–3?
   b. What do the results reveal about their genetic architectures?
   c. What is the potential to use the results to strengthen inference in empirical research and to perform multi-trait analyses of genetically correlated traits?
   d. What do the results reveal about biological mechanisms and pathways?
   e. What do the results suggest about the biological pathways and candidate genes that have previously been hypothesized to influence risk taking?

2. Is educational attainment associated with CpG methylation and how does the strength of association compare to biologically proximate factors?
   What do the results suggest about the hypothesis that psychosocial experiences influence disease via methylation?

3. What is the overall quality and evidential value of the previous studies that have tested antisocial behavior for association with an interaction between adverse life events and 5-HTTLPR?

**Thesis disposition**

The remainder of this thesis is structured as follows. The next section defines the concept of genetic architecture and motivates the discovery of genetic associations. Thereafter, the main phenotypes are briefly introduced. Lastly, I report my own and my co-authors’ contributions to the large-scale, collaborative efforts upon which this thesis is based. Following this introduction, the thesis consists of four chapters that are based on empirical studies published in international, peer-reviewed journals. **Chapters 2–3** are published in *Nature Genetics* (Karlsson Linnér et al., 2019; Okbay et al., 2016), **Chapter 4** in *Molecular Psychiatry* (Karlsson Linnér et al., 2017), and **Chapter 5** in *the American Journal of Medical Genetics Part B: Neuropsychiatry* (Tielbeek et al., 2016). For completeness, these studies are reported in their entirety though not all results are discussed in this thesis. Lastly, the sixth chapter discusses how a curated selection of the findings answer the main research questions, and it reports a conclusion and the study limitations.

**Statistical methods**

This thesis relies on a large number of statistical methods and procedures, and it is far beyond the scope of this introductory chapter to describe them all. For brevity, the methods to perform a selection of the analyses that I carried out are reported in the **Supplementary Methods** sections. **Chapter 2 – Supplementary Methods** details the method to perform GWAS, quality-control and meta-analysis, estimation of conditional GWAS associations, and three methods to estimate SNP heritability. **Chapter 4 – Supplementary Methods** reports the method to perform EWAS, quality-control and meta-analysis, robustness checks, the comparison of effect-size estimates with biologically proximate factors, among other methods. All other relevant methods and materials are described in great detail either within the thesis chapters, or in the accompanying **Online Supplementary Materials**.
The genetic architecture of complex traits

The genome consists of deoxyribonucleic acid (DNA), which is recombined and transmitted from parents to offspring. DNA is the carrier of genetic information—the instructions required for the growth and development of all known cellular life. The genome is a sequence consisting of four molecular “letters” (base pairs, or nucleotides), and in humans that sequence is roughly 3 billion base pairs long. The base pairs are made out of the molecules adenine (A), cytosine (C), guanine (G), and thymine (T). DNA takes the form of two strands, shaped like a double helix, and because of the complementary property of the nucleotides, that is A is always paired with T and C with G, knowledge of the sequence on one chromosomal strand can be used to infer its complementary strand. In humans, DNA is structured as 22 chromosome pairs (autosomes), as well as a pair of sex chromosomes (allosomes). Because the chromosomes come in pairs, a specific location on the genome (locus) can either be homozygous or heterozygous in terms of the particular nucleotide sequence (alleles) at that locus. For most practical purposes, it is common to assume that an individual’s genetic sequence is stable over the life span, and identical across the trillions of cells and various tissues in the body (with the exception of sex cells, and other special cases).

The genetic architecture of a particular trait can broadly be defined as the complete set of genetic influences attributable for the heritability. The definition includes, but is not limited to, the set of causal variants; the distribution of their effect sizes; and their number, correlation (linkage), and locality in the genome (e.g., in coding, non-coding, or regulatory regions). A broader definition also includes the occurrence of the alleles (allele frequency) of variants within and across populations; whether the alleles at the same locus interact (dominance), and whether they interact with alleles at other loci (epistasis), or with environmental factors (gene-environment interaction, or G×E). If we would allow an even broader definition, the genetic architecture also covers the states of epigenetic mechanisms, such as methylation, as well as the genetic overlap with other traits. Many of these aspects, but not all, will be studied in the thesis chapters.

Since most genetic variants appear to exert only a tiny influence on psychology and behavior, it is not unreasonable to question whether efforts to identify genetic associations are actually worthwhile. I argue that genetic discovery is both intrinsically valuable, because such efforts may increase our understanding of the etiology of individual differences, which has allured the interest of scholars across time and disciplines, as well as extrinsically valuable, because the findings from genetic discovery have many scientific benefits. The latter can be categorized along three lines. First, genetic associations can in many ways strengthen inference in empirical research. Second, genetic associations with a particular trait can be beneficial for multi-trait analyses of genetically correlated phenotypes. Third, genetic associations offer a window into the biological mechanisms and pathways involved in trait variation. The remainder of this section further explains these three lines of motivation.

**Strengthening inference in empirical research**

Polygenic scores (PGS) index, or summarize, the genetic effects estimated in GWAS across a large number of genetic variants, and PGS can in many ways strengthen inference in empirical research. For example, PGS can be used as control variables to condition the effect of a variable of interest or to increase the statistical power to detect a treatment effect in randomized control trials. In addition, with the use of PGS, the method genetic instrumental variable (GIV) regression can accurately estimate the effect of an exposure in the presence of pleiotropy. Further, gene-environment interaction studies that use PGS, rather than single genetic variants, are likely to have greater statistical power to detect an interaction effect.
Because GWAS summary statistics can be used to construct PGS, among other reasons, it has become common to share these with the wider research community\textsuperscript{43,125}. Thus, genetic discovery could even be considered a scientific public service. Overall, identification of genetic associations has much to offer to observational studies in fields such as epidemiology, psychology, and the social sciences\textsuperscript{126}.

Next, it is of scientific and societal importance to establish causal relationships between environmental risk factors and health outcomes\textsuperscript{127,128}, for example in the effort to reduce health inequalities\textsuperscript{83}. However, concerns have been raised over the unbiasedness of estimates reported in observational studies that lack randomization\textsuperscript{99,117,129}. Because of the possibility that bidirectional and other complex pathways account for the covariation between risk factors and disease there is a risk of simultaneity and other sources of endogeneity\textsuperscript{130,131}. As a remedy, it has been proposed that the random assortment of genes that an offspring inherits, conditional on the parents, could be used as instrumental variables in so-called Mendelian Randomization\textsuperscript{132}. However, using genes as instruments is not without limitations\textsuperscript{127,128,133}.

On the positive side, an individual’s genome is highly stable over time, and in many conceivable cases, exogenous to the variables of interest in the study of risk factors and disease\textsuperscript{59}. Genetic instruments may be particularly useful in research settings where appropriate quasi-randomization is hard to find, and can be applied to a wide range of research questions\textsuperscript{126}. Some have even gone so far as to claim that a genetically-informed research design is a requirement to actually understand environmental influences\textsuperscript{111}. Overall, researchers who aim to disentangle the complex relationships between mental traits, environmental risk factors, and health outcomes, should not neglect the advantages offered by a genetically-informed research design\textsuperscript{126,131}.

**Multi-trait analyses of genetically correlated traits**

Conceptually, the main phenotypes could be considered to be intermediate phenotypes and/or non-clinical measures of mental disorders. The premise that genetic studies of normal-range measures of mental traits can potentially advance our understanding of the disorders of the brain, in contrast to a narrow focus on clinical diagnoses and symptoms, is at the essence of a research framework called the Research Domain Criteria (RDoC)\textsuperscript{134–136}. Thus, genetic association with the main phenotypes could aid the effort to identify genes and biomarkers for mental disorders, which has been declared a grand challenge by an international consortium of clinicians and mental-health researchers\textsuperscript{21}. There are several multi-trait methods available that leverage the genetic overlap across traits, such as the proxy-phenotype method and the multi-trait method MTAG\textsuperscript{137,138}.

LD Score regression is a method that is frequently used to estimate genetic correlations\textsuperscript{139,140}, a measure of the extent of overlap in genetic effects between pairs of traits\textsuperscript{39,139,141}. The size of genetic correlations is informative of the possible benefit of multi-trait analysis of genetically correlated traits\textsuperscript{13,138,142}. For example, multi-trait analysis can aid the identification of genetic associations with phenotypes for which there is currently a lack of large samples (such as rare disorders, or phenotypes that for some reason may be difficult to sample)\textsuperscript{137}. When several phenotypes are studied in combination it is possible to increase the statistical power to find associations with a particular trait\textsuperscript{111,138}. Also, genetic associations can be used to quasi-replicate findings from other studies that lack a replication sample\textsuperscript{143}. Overall, the possibility to perform multi-trait analyses of genetically correlated traits is a strong motivation to discover genetic associations.
Identification of biological mechanisms and pathways

Genetic associations can be used to identify biological mechanisms and pathways because genetic variants can be linked to functionally relevant products, such as ribonucleic acids (RNAs), proteins, and gene expression levels\textsuperscript{144–146}. Evidently, biological pathways that were previously thought to be unrelated to particular medical conditions have been discovered with GWAS\textsuperscript{49}. Identification of biological pathways can be considered of additional importance with respect to psychiatric disorders for which the identification of pathophysiology has proven difficult\textsuperscript{11,52,135}. Also, knowledge of biological pathways can aid the study of comorbidity and disease heterogeneity\textsuperscript{19,35}.

However, methods that use genetic associations to discover biological mechanisms and pathways have several limitations. Typically, the strength of such bioinformatic analyses is strongly dependent on the availability of additional data sources, such as information on gene expression, function, and networks. Another limiting factor is the availability of relevant tissue samples. Notwithstanding these limitations, many still consider identification genetic discovery to be an important first step in the effort to reveal the biological underpinnings of human traits and disorders\textsuperscript{30,49}. Since the main phenotypes ultimately stem from the brain, which could be considered an extremely complex organ, this thesis may have limited potential to accurately pinpoint biological mechanisms and pathways. Yet, bioinformatic analyses of the initial GWAS estimates may shed some light on previous findings, and can potentially yield new insights.

Introduction to the main phenotypes

General risk tolerance and risky behaviors

Many activities and decisions entail an element of risk. Semantically, risk can be defined as the possibility of experiencing danger, harm, or loss\textsuperscript{147}. Notably, the inclination to take risks varies substantially across individuals, while it is a relatively stable personal characteristic. It decreases as a function of age and it has been observed that women are less willing to take risks than men\textsuperscript{15–17}. In the literature, the willingness to accept the exposure to risk in exchange for the prospect of a reward is often referred to as risk attitudes, risk aversion, or risk preferences\textsuperscript{13,15,16}. Notably, there is no clear consensus on how this tendency should be defined and measured.

In economics, risky decision-making is often conceptualized based on statistical concepts such as expected values and the variability of outcomes\textsuperscript{15}, and modeled with expected utility theory\textsuperscript{16,23}. As a critique of that theory, Kahneman & Tversky (1979) proposed a different framework—prospect theory—that contests the external validity of many of the assumptions of expected utility theory\textsuperscript{148,149}. In psychology, risk-taking spans across several personality dimensions and is considered similar to sensation seeking, adventurousness, and related personality types and traits\textsuperscript{150–152}. Overall, risk is a somewhat fuzzy concept, but yet an integral part of the decision-making processes of living organisms\textsuperscript{153}.

In Chapter 2, the primary phenotype is general risk tolerance, which is defined here as the self-reported overall willingness to take risks. In addition, that chapter investigates the six supplementary phenotypes adventurousness, automobile speeding propensity, drinks per week, ever smoker, number of lifetime sexual partners, and the first principal component (PC) of the four risky behaviors. The supplementary phenotypes were deliberately chosen to represent real-world risk taking in specific contexts, and the first PC is intended to capture a general component of risk taking across some of those domains. There have been many attempts to identify genes and biological pathways involved in risk tolerance but few have been reliably
performed in large samples\textsuperscript{154}. Overall, there are only two genome-wide significant associations reported in the literature\textsuperscript{155–157}. Therefore, \textbf{Chapter 2} sets out to discover the genetic variants responsible for the missing heritability of risk taking in a much larger sample than has previously been done.

\textit{Subjective well-being, depressive symptoms, and neuroticism}

Subjective well-being is defined as an individual’s subjective rating of his or her quality of life\textsuperscript{158–160}. It is typically measured through self-reports, using a combination of survey items intended to capture overall happiness, life satisfaction, and positive affect. In addition, \textbf{Chapter 3} investigates two genetically correlated traits\textsuperscript{161,162}, depressive symptoms and neuroticism. Depression is a debilitating medical condition with a high disease burden, which is hard to treat\textsuperscript{52,163}. Larger genetic studies of depression are motivated to identify druggable targets and to accelerate the development of new treatments. Neuroticism is a personality dimension characterized by an affinity to experience negative feelings, and studies have found that neuroticism is strongly associated with depression\textsuperscript{164–166}. Also, depressive symptoms and neuroticism could largely be considered reverse-coded measures of subjective well-being. Overall, there are few reported genetic associations with these traits in the literature\textsuperscript{52,167}, and the GWAS in large samples reported in \textbf{Chapter 3} may be able to uncover the missing heritability of these traits.

\textit{Educational attainment}

Educational attainment is a measure of an individual’s scholastic achievement. Conceptually, educational attainment is both an outcome that is influenced by personal characteristics and environmental circumstances, and at the same time, education is an environmental exposure and major life experience. It is well-established that variation in education is strongly associated with mental and physical health, as well as longevity\textsuperscript{98,168,169} and higher education is a strong predictor of positive health behaviors and lifestyle choices\textsuperscript{98}. Overall, educational attainment can be considered a crude measure, but nonetheless, it captures a range of important differences across individuals in a population, and it is easy and inexpensive to measure in large samples.

To date, there are three large-scale GWAS of educational attainment published in the literature\textsuperscript{31,170,171}. In these studies, educational attainment is operationalized as a continuous trait measured in years-of-schooling equivalents. The most recent study, which analyzed close to 1.1 million individuals, identified more than twelve hundred near-independent genetic associations\textsuperscript{171}. Thus, GWAS of educational attainment have already been very successful. So, \textbf{Chapter 4} instead investigates the association between educational attainment and CpG methylation. To my knowledge, there are no previous studies on that particular relationship, while there are studies that have investigated its association with methylation by using educational attainment as proxy for, or in combination with, socio-economic status, and a few associations have been reported\textsuperscript{81,172,173}. Notably, our study is several times larger than any previous investigation of the association between educational attainment and methylation.

\textit{Antisocial Behavior}

Antisocial behavior includes various forms of aggressive and violent behavior, violation of social norms, and delinquency\textsuperscript{174,175}. It is associated with many negative health and life outcomes, such as alcohol and substance use, as well as criminal conviction. The study of antisocial behavior is of societal importance because of the high financial and social costs associated with it and its correlates\textsuperscript{176}. It has been estimated that antisocial behavior is highly heritable, but many studies have also identified a multitude of environmental risk factors\textsuperscript{177}. 
For example, antisocial behavior is associated with experiencing various forms of childhood and adolescent adversities, such as child maltreatment and socioeconomic stress. This observation has led to great scientific interest in the question whether the effects of environmental risk factors are contingent on genetic differences. As is applicable for many other traits studied in psychology, it has been suggested that antisocial behavior could be associated with an interaction between environmental adversity and alleles of the gene region 5-HTTLPR. However, few large-scale studies have investigated this hypothesis, and it is necessary to further investigate that hypothesis to corroborate or reject some positive findings reported in the literature.

Individual contributions

The empirical chapters of this thesis are all based on collaborative work, and most include a large number of co-authors that all contributed in various ways, e.g., with data collection, data preparation, and cohort-level analyses. Authors that are not listed in this section contributed primarily to such efforts and details of their contributions can be found in the Online Supplementary Material accompanying the published studies. All authors critically reviewed and commented on the manuscripts prior to publication. Beyond the many contributing authors, the central analyses and writing of the manuscripts were performed by a small, central team of junior and senior authors. A selection of the analyses that I conducted myself or was particularly involved in are reported in the Supplementary Methods in this thesis. To my knowledge, Chapters 2, 3 and 5 are part of other PhD theses. The following paragraphs summarize and report the contributions of the central team of authors, including my own.

Chapter 2 – Genome-wide association analyses of risk tolerance and risky behaviors in over one million individuals identify hundreds of loci and shared genetic influences

In this study, I was the lead analyst and contributed substantially to the analyses and the writing of the manuscript. The study was conceived by Jonathan Beauchamp, Philipp Koellinger, and Daniel Benjamin. I conducted the GWAS analyses in the UKB, and the quality-control and meta-analyses. I performed the conditional analyses of the identified genetic associations, the investigation of long-range LD regions, candidate inversions, and 1000 Genomes structural variants, and I summarized the overlap across the various GWAS. Juan Ramon Gonzalez and Tõnu Esko contributed data to the investigation of candidate inversions. The population stratification, replication, and proxy-phenotype analyses were conducted by Edward Kong. I performed the SNP-based heritability analyses. The genetic correlation analyses were conducted by Robbee Wedow, with assistance from Mark Fontana. I performed polygenic prediction in the STR cohort, and the prediction analyses were led by Pietro Biroli, and Edward Kong, Robbee Wedow, Abdel Abdellaoui, Ronald de Vlaming, Mark Fontana, Michel Nivard also contributed to those analyses. The MTAG analyses were conducted by Pietro Biroli and Christian Zünd. Fleur Meddens led the bioinformatics analyses, and was assisted by Jacob Gratten and Maciej Trzaskowski, Anke Hammerschlag, Gerardus Meddens, and Pascal Timshel. Maël Lebreton conducted the review of the literature attempting to link risk taking to genes and biological pathways, and I performed the SNP-based validation analysis of these previous hypotheses. I prepared the majority of the figures, with assistance from Edward Kong and Stephen Tino. Additional analyses were performed by Aysu Okbay, Niels Rietveld, and Stephen Tino. David Cesarini, Jacob Gratten, and James Lee provided helpful advice and feedback on various aspects of the study design. Jonathan Beauchamp, Pietro Biroli, Edward Kong, Fleur Meddens, Robbee Wedow, and I made especially major contributions to the writing and editing of the manuscript.
All the contributing authors are listed on p. 34. For a complete description of the author contributions, see Supplementary Material section 13 in Karlsson Linnér et al. (2019).

Chapter 3 – Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses:

This study was the first large-scale GWAS I was involved in, and I made a less prominent contribution than in Chapter 2. The study was designed by Meike Bartels, Daniel Benjamin, David Cesarini, Jan-Emmanuel De Neve, Philipp Koellinger, and Robert Krueger. The quality control and meta-analyses were performed by Aysu Okbay and Bart Baselmans. GWAS analyses in the UKB were conducted by Mark Fontana. Jonathan Beauchamp and Patrick Turley performed tests for population stratification, developed the “quantity-quality tradeoff”, and conducted the Bayesian credibility analyses. Aysu Okbay and Bart Baselmans conducted the polygenic score prediction analyses. Mark Fontana, Patrick Turley, and Jonathan Beauchamp performed the genetic correlation analyses. Aysu Okbay and I conducted the proxy-phenotype and cross-phenotype enrichment analyses. The bioinformatics analyses were performed by Jonathan Beauchamp, Mark Fontana, Fleur Meddens, Michel Nivard, and Tõnu Esko. I prepared the majority of the figures and Fleur Meddens the majority of the tables. The inversion polymorphisms were analyzed by Harm-Jan Westra and Juan Ramon Gonzalez. Daniel Benjamin, Meike Bartels, Jonathan Beauchamp, David Cesarini, Jan-Emmanuel De Neve, Michel Nivard, Philipp Koellinger, Aysu Okbay, and Patrick Turley made especially major contributions to the writing and editing of the manuscript.

All the contributing authors are listed on p. 73. For a complete list of author contributions, see Supplementary Material section 11 in Okbay et al. (2016).

Chapter 4 – An epigenome-wide association study meta-analysis of educational attainment:

In this study, I was the lead analyst together with Riccardo Marioni and Niels Rietveld. I contributed substantially to the analyses and the writing of the manuscript. The study was designed by Daniel Benjamin, Philipp Koellinger, Ian Deary, Niels Rietveld, and Riccardo Marioni. Niels Rietveld and I performed the EWAS quality control and meta-analyses. Riccardo Marioni performed the epigenetic clock analyses, and Niels Rietveld and I assisted those analyses. The polygenic prediction was carried out by Niels Rietveld, Riccardo Marioni, Andrew Simpkin, and Neil Davies. Niels Rietveld and I conducted the robustness analyses of smoking. Niels Rietveld and I designed and performed the enrichment analyses. I designed and performed the tissue-specific methylation analyses. The FUMA expression analysis was performed by Kyoko Watanabe. The methQTL GWAS were performed by Riccardo Marioni and Niels Rietveld, and I performed the subsequent quality control and meta-analysis. Daniel Benjamin, Riccardo Marioni, Niels Rietveld, and I made especially major contributions to the writing and editing of the manuscript.

All the contributing authors are listed on p. 92. The authors not listed above were involved in data collection, data preparation, or cohort-level EWAS analyses.

Chapter 5 – Meta-analysis of the serotonin transporter promoter variant (5-HTTLPR) in relation to adverse environment and antisocial behavior:

In this study, I was the lead analyst together with Jorim Tielbeek. I contributed substantially to the analyses and the writing of the manuscript. The study was designed by Tinca Polderman, Arne Popma, and Danielle Posthuma. Jorim Tielbeek, Kokko Beers, and I performed the data collection. The quality-assessment analysis was performed by Jorim Tielbeek and Kokko Beers, with assistance from Tinca Polderman. Jorim Tielbeek and I performed the meta-analysis, the summary of the results, and the publication-bias analyses. I prepared the majority of the figures.
Jorim Tielbeek and I made especially major contributions to the writing and editing of the manuscript.

Additional publications

During the PhD, I also contributed to the following studies:


References – Preface & Chapter 1


59. Davies, N. M., Holmes, M. V & Smith, G. D. Reading Mendelian randomisation


77. Lutz, P. E. *et al.* Association of a history of child abuse with impaired myelination in


112. Jenness, J. L., Young, J. F. & Hankin, B. L. 5-HTTLPR moderates the association


Discovering the genetic architecture of the mind


