413

AYSU OKBAY

Essays on Genetics and the Social Sciences

AYSU OKBAY - Essays on Genetics and the Social Sciences

ERIM

Essays on Genetics and the Social Sciences

Essays on Genetics and the Social Sciences

Essays over Genetica en de Sociale Wetenschappen

THESIS

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board.

The public defense shall be held on
Thursday, February 9, 2017, at 13.30 hours

by

Aysu Okbay
born in Istanbul, Turkey.

Erasmus University Rotterdam

Doctoral Committee

Doctoral dissertation supervisors:	Prof.dr. A.R. Thurik Prof.dr. P.D. Koellinger Prof.dr. P.J.F. Groenen
Other members:	Prof.dr. M. Johannesson Prof.dr. D. Posthuma Prof.dr. A. Uitterlinden

Erasmus Research Institute of Management – ERIM
The joint research institute of the Rotterdam School of Management (RSM)
and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam
Internet: http://www.erim.eur.nl

ERIM Electronic Series Portal: http://repub.eur.nl/

ERIM PhD Series in Research in Management, 413
ERIM reference number: EPS-2017-413-S&E
ISBN 978-90-5892-477-3
© 2017, Aysu Okbay

Design: PanArt, www.panart.nl

This publication (cover and interior) is printed by Tuijtel on recycled paper, BalanceSilk®
The ink used is produced from renewable resources and alcohol free fountain solution.
Certifications for the paper and the printing production process: Recycle, EU Ecolabel, FSC®, ISO14001.
More info: www.tuijtel.com

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the author.

Contents

CHAPTER 1

Introduction and Conclusion

Abstract

This thesis consists of six chapters that examine the genetic architecture of complex behavioral outcomes. Chapter 1 lists and substantiates the research questions and gives a summary overview of the key findings and their implications. Chapter 2 discusses methodological challenges with traditional candidate gene approaches, which until recently were the dominant paradigm for gene discovery in the social sciences. The chapter explains why well-known problems of low statistical power, population stratification and undisclosed multiple-hypothesis testing render most of the existing evidence from candidate gene studies unreliable. Each potential problem is illustrated with reference to specific studies published by the journal *Emotion*. Recognizing these limitations, most modern gene discovery efforts are genome-wide association studies (GWAS), in which millions of markers are tested for association with the outcome of interest in large samples and with stringent controls for population stratification. Chapters 3 and 4 report the results of two such studies, performed by meta-analyzing summary statistics from a large number of cohorts with combined sample sizes in the hundreds of thousands. The first study (Chapter 3) examines four genetically correlated phenotypes related to well-being (depressive symptoms, life satisfaction, neuroticism and positive affect) and identifies 14 genome-wide significant associations. The second study (Chapter 4) examines educational attainment and identifies 74 genome-wide significant associations. Chapter 5 builds on these findings by exploiting the staggered rollout of an expansion in comprehensive schooling in Sweden to test for gene-by-environment interactions, finding that the impact of the reform on educational outcomes and earnings was heterogeneous by genotype for females: females with higher polygenic scores (derived using the results from Chapter 4) were more likely to obtain a high school degree, which is beyond the new minimum established by the reform. Chapter 6 develops a theoretical multi-study framework relating statistical power and predictive accuracy to cross-study heterogeneity. The chapter shows that in GWAS meta-analyses, including those in Chapters 3 and 4, cross-cohort genetic correlations lower than unity can have a substantial effect on statistical power, and predictive accuracy of polygenic scores derived from meta-analyses of summary statistics from heterogenous cohorts. The methodology developed in Chapter 6 is easily implemented using an online tool that researchers can use to quantify the tradeoff between gains from larger sample sizes and the potential losses in power from pooling measures that are too heterogeneous.

1.1 Motivation

Evidence from behavior-genetic studies of twins, adoptees and other pairs of relatives shows that virtually all human traits, including economic preferences and behaviors, are at least moderately heritable (Polderman et al., 2015; Sacerdote, 2010; Turkheimer, 2000). With more and more genetic data becoming available, it is increasingly feasible to identify specific genetic variants associated with complex outcomes (Visscher, Brown, McCarthy, & Yang, 2012), thus blurring the disciplinary boundaries between the biological and social sciences. The research described here identifies specific genetic variants associated with a suite of complex outcomes – ranging from educational attainment, to subjective well-being and neuroticism. It illustrates some hard-won lessons about the relative merits of various research strategies that have been proposed for efforts to discover genetic associations with complex traits.

There are many ways in which such gene-discovery efforts may prove worthwhile (Beauchamp et al., 2011; Benjamin, Cesarini, Chabris, et al., 2012). My discussion will focus on four promising uses.

The first is that since genetic variation is ubiquitous, unobserved genetic factors are often a plausible confound that researchers interested in the causal impact of some *environmental* variable ought to worry about. With the ever-increasing availability of genetic data, there will be many empirical settings in which it is feasible for researchers to directly control for potential genetic confounds, thus strengthening their ability to credibly make causal inferences. For example, a researcher who finds that a particular policy change was associated with reduced levels of smoking would be able to argue more convincingly that the policy *caused* the reduction if it could be shown that genetic variants associated with smoking cessation had similar distributions among people who were and who were not impacted by the policy. In the years ahead, the work described here suggests that there are many questions in the social sciences that can be more credibly addressed with information about genetic data, even if the research question itself is only related to genetics in the most tangential of ways. In some cases, the ability to control for genetic factors may also prove valuable in settings where it is not well understood why the genes in question are associated with the outcome. For example, in randomized controlled trials, controlling for genetic predictors may substantially boost the precision with which a treatment effect is estimated.

More generally, economists typically study behavior within a framework where people make choices to maximize their utility subject to a budget constraint. They have historically been wary of "preference-based" explanations for differences across individuals on the grounds that such explanations are often difficult to put to a serious test, or indeed, sometimes impossible, in which case the explanation ceases to be scientific (Stigler & Becker, 1977). This

is an understandable concern; to say that Mary eats fish and Bob eats meat because their preferences differ is to say very little unless the explanation has some refutable implications. The discovery of genetic associations will enable researchers to test hypotheses about heterogeneity (in preferences, skills or some other characteristic) that have refutable implications in any world where genetic factors can be measured (Caplan, 2003).

A second contribution is that genetic advances may provide individuals with useful information on which they may wish to take preemptive action. If, for example, genetic markers can be used to predict a child's propensity towards dyslexia with sufficient accuracy, parents will have the option to take precautionary steps such as enrolling the child in supplementary reading programs even before it is possible to make a clinical diagnosis of dyslexia (Benjamin, Cesarini, Chabris, et al., 2012).

A third is that the new data will provide a treasure trove for many scientists who are currently constrained by limited data when trying to test their hypotheses. Many social scientists – i.e. health economists, epidemiologists and sociologists – are fundamentally interested in understanding how policy or environmental factors magnify or dampen genetic risk, but progress on this important question has been dampened by our limited knowledge of the specific genetic factors that influence various complex traits (Duncan & Keller, 2011; Rietveld, Conley, et al., 2014). Another example comes from evolutionary game theory. Evolutionary game theorists formulate models of the evolution of human characteristics (Robson, 1996, 2001; Van Damme, 1994), some of which make predictions that can only be straightforwardly tested with molecular data. As more and more is learnt about the human genome, it is thus likely that an era of empiricism will dawn upon evolutionary game theory.

A fourth contribution lies in the potential of elucidating biological mechanisms. Identification of genes may also help researchers identify new biological systems that point to new theoretical constructs, with clearer biological foundations, as potentially valuable in the study of individual differences (Benjamin, Cesarini, Chabris, et al., 2012). For some outcomes (such as subjective well-being, studied in Chapter 3) genetic discoveries may also inform therapeutic efforts to develop drugs for genetically correlated medical conditions such as major depressive disorder.

The research described in this thesis represents a small step toward the reintegration of the social and biological sciences. Traditional disciplinary boundaries are partly the results of historical accidents, and become reified through a process that insulates researchers within a discipline from the influence of outside forces. Calls for a unification of the social and natural sciences (Wilson, 1998) notwithstanding, such progress has been slow.

1.2 Basic Genetic Concepts

This section, which draws on Beauchamp et al. (2011) and Benjamin, Cesarini, Chabris, et al. (2012), is intended for non-expert readers with little prior knowledge of genetics and defines some basic concepts that will be used throughout the thesis.

The human genome consists of 23 pairs of chromosomes—one of which is inherited from the mother and the other from the father—and each chromosome carries genetic information in the form of deoxyribonucleic acid (DNA). DNA consists of two strands of nucleotide molecules, which are its building blocks. Each nucleotide contains one of four different bases: adenine (A), guanine (G), thymine (T), and cytosine (C). Due to a property called complementarity, nucleotides with the bases C and G are always paired with one another, and nucleotides with the bases G and A are always paired together, forming so-called base pairs and holding the two strands of DNA together. The maternal and paternal sets of 23 chromosomes each contain approximately 3.2 billion nucleotide pairs; human DNA thus consists of 3.2 billion *pairs of nucleotide pairs*.[1]

The vast majority of these nucleotide pairs do not vary across humans (The 1000 Genomes Project Consortium, 2015). The most common type of genetic variation in humans is "single nucleotide polymorphism" (SNP), which is a location in the DNA where individuals differ at a single pair of nucleotide pairs.[2] [3] For most SNPs, only two different nucleotide pairs exist in the population. To define a "genotype" variable for a given SNP, we arbitrarily define one of the two possible nucleotide pairs as the reference; we then code the genotype as "0" if the individual inherited the alternative nucleotide pair from both parents, "1" if he inherited one reference nucleotide pair, and "2" if he inherited two reference nucleotide pairs.

[1] Because the two strands of nucleotides contain the same information due to complementarity, it is customary to use a single strand to summarize this information. For instance, the genotype …AT-GC-AT-TA… can be summarized as …AGAT… or …TCTA…. The information contained in the 3.1 billion *pairs of nucleotide pairs* can thus be summarized as 3.1 billion *pairs of nucleotides*.

[2] Other kinds of variations include, but are not limited to, copy number variations, where a section of the genome is repeated a variable number of times, or inversions where a section of the genome is reversed.

[3] "Genes" are sub-sequences of nucleotide pairs that contain instructions for coding proteins, which in turn regulate bodily functions. The human genome is believed to contain 20,000-25,000 genes (The International Human Genome Sequencing Consortium, 2004), which is only a small portion of the genome. The remaining parts of the genome can still have important functions, such as regulating gene expression.

Genome-wide association studies (GWAS) are hypothesis-free studies that estimate the associations between an outcome of interest and each of a large number of SNPs. For a given trait *y*, a GWAS estimates the following equation for each SNP:

$$y_i = \mu + \beta_j x_{ij} + \boldsymbol{\gamma}\mathbf{C_i} + \epsilon_i \quad (1.1)$$

where *i* indexes individuals, *j* indexes SNPs, x_{ij} is the genotype of individual *i* at SNP *j*, and $\mathbf{C}_i$ is a vector of controls for individual *i*, such as age, gender, and the top principal components (defined below). Because any SNP *j* is correlated with nearby SNPs, β_j captures the causal effects both of SNP *j* and of correlated SNPs. One of the main lessons from the past decade of social genomics research is that behavioral traits such as educational attainment tend to be influenced by a large number of genetic and environmental factors (and interactions between them), and the effect sizes of most if not all SNPs are extremely small (Chabris, Lee, Cesarini, Benjamin, & Laibson, 2015).

When estimating the parameters, it is important to minimize the risk that any association detected is spurious and caused by failure to control for some relevant confound. Population stratification refers to differences in genotypes between sub-populations, coupled with differences in the trait of interest due to non-genetic reasons. It can severely bias estimates of the associations between genetic variants and a trait of interest (Hamer & Sirota, 2000). Restricting the analyses to individuals from the same ethnic group is necessary, but often not sufficient, to address this problem. To further control for sub-population differences, including in the control variables $\mathbf{C}_i$ the top principal components of the covariance matrix of the individuals' genotypic data has emerged as a common practice in genetics research (Price et al., 2009) (the top principal components have been shown to capture population substructures in ethnically homogenous samples).

1.3 Research Questions and Main Results

Five main research questions are addressed in this thesis. The research questions, together with the main results, are described below.

Research question 1: What are the pitfalls of candidate gene studies? How can the reliability of candidate gene study findings be improved? (Chapter 2)

The traditional approach to the investigation of the role of genetic variation in determining behavior in the social sciences is to conduct a 'candidate gene study', in which a limited number of genetic variants are selected based on their hypothesized or known biological

function, and these variants are tested for association with the trait of interest (Benjamin, Cesarini, van der Loos, et al., 2012; Ebstein, Israel, Chew, Zhong, & Knafo, 2010). However, successful replications of candidate gene studies published to date have been alarmingly scarce (Cardon & Palmer, 2003; Duncan & Keller, 2011; Ioannidis, 2005). Chapter 2 describes the main methodological challenges that candidate gene studies face, which, if not properly addressed, can lead to false positive findings, namely: low statistical power due to small sample sizes, lack of replication stage, population stratification and other confounders, and lack of multiple testing correction. To illustrate these challenges, we evaluate all candidate gene and gene × environment studies published in *Emotion* until April 2014. None of the reviewed studies prove to have successfully avoided these pitfalls. The review suggests that some methodological steps should be taken to improve the credibility of candidate gene studies. We recommend adherence to guidelines such as the ones developed by the *Behavior Genetics Association* (Hewitt, 2012) for conducting and evaluating candidate gene studies, which will help to increase the credibility of candidate gene study findings.

Research question 2: What are the common and unique genetic factors underlying subjective well-being, neuroticism and depressive symptoms? What do these findings suggest about the biological mechanisms underlying these traits? (Chapter 3)

Chapter 3 reports a series of separate and joint analyses of subjective well-being (SWB), depressive symptoms (DS), and neuroticism that exploit the strong genetic overlap ($|\hat{\rho}| \approx 0.8$) between the three phenotypes. The primary analysis is a GWAS of SWB based on data from 59 cohorts ($N = 298{,}420$) and identifies three loci associated with subjective well-being at genome-wide significance ($p < 5\times10^{-8}$). Auxiliary GWAS meta-analyses of DS ($N = 180{,}866$) and neuroticism ($N = 170{,}910$) identify two loci associated with DS and eleven with neuroticism, including two inversion polymorphisms. In depression data from an independent sample ($N = 368{,}890$), both DS associations replicate. Joint analyses that exploit the high genetic correlations between the phenotypes allow us to assess the replicability of the associations, strengthening the overall credibility of our findings, and allowing us to identify two additional loci associated with neuroticism and two with both depressive symptoms and neuroticism. Across the three phenotypes, loci regulating expression in central nervous system and adrenal/pancreas tissues are strongly enriched for association. Polygenic scores constructed from all measured SNPs explain a low fraction of variance in independent samples: ~0.9% for subjective well-being, ~0.5% for depressive symptoms, and ~0.7% for neuroticism. The findings suggest that GWAS can successfully identify genetic associations with highly polygenic phenotypes in sufficiently large samples (Hyman, 2014; Sullivan, 2012), and joint analysis of genetically correlated traits can play an important role in boosting statistical power.

Research question 3: What are the genetic variants associated with educational attainment (EA), and what do these variants suggest about the biological mechanisms underlying EA and correlated mental health phenotypes? (Chapter 4)

Educational attainment (EA) is influenced strongly by social and other environmental factors, but genetic factors are also estimated to account for a sizable fraction (> 20%) of the variation across individuals (Rietveld, Medland, et al., 2013). The largest previous genome-wide association study (GWAS) on EA reported three independent genome-wide significant ($p < 5\times10^{-8}$) SNPs associated with EA in a sample of 101,069 individuals (Rietveld, Medland, et al., 2013). Chapter 4 reports results from a GWAS on EA that extends the sample of 101,069 individuals to a discovery sample to 293,723 individuals, and an independent replication sample of 111,349 individuals from the UK Biobank. 74 genome-wide significant loci associated with number of years of schooling are identified in the discovery sample. The effect size estimates range from 0.014 to 0.048 standard deviations per allele (2.7 to 9.0 weeks of schooling), with incremental R^2 in the range 0.01% to 0.035%. LD Score regression (Bulik-Sullivan, Loh, et al., 2015) and within-family analyses performed to quantify the amount of population stratification in the GWAS estimates suggest that stratification effects are small in magnitude. In the U.K. Biobank dataset, 72 out of the 74 lead SNPs replicate with a consistent sign, 52 are significant at the 5% level and 7 reach genome-wide significance, exceeding the corresponding expected numbers assuming each SNP's true effect size is its estimated effect adjusted for the winner's curse. We also find out-of-sample replicability of our overall GWAS results: the genetic correlation between *EduYears* in our meta-analysis sample and in the UKB data is 0.95 (s.e. = 0.021). Across two holdout samples, the mean predictive power of a polygenic score constructed from discovery sample effect sizes of all measured SNPs is 3.2%. SNPs associated with EA are disproportionately found in genomic regions regulating gene expression in the fetal brain. The identified genes are preferentially expressed in neural tissue, especially during the prenatal period, and enriched for biological pathways involved in neural development. The findings demonstrate that, even for a behavioral phenotype that is mostly environmentally determined, a well-powered GWAS identifies replicable associated genetic variants that suggest biologically relevant pathways.

Research question 4: Did Genes Moderate the Effect of the Swedish Comprehensive Schooling Reform?

Gene-by-environment (G×E) effects on behavioral traits are widely believed to be pervasive and there is tremendous excitement about the possibility that advances in molecular genetics

will open up new research possibilities in this area. Some investigators have voiced concerns, however, that a downside to the recognition is that claims of G×E are not scrutinized as critically as findings from other areas of behavioral-genetic research (J. Lee & McGue, 2106). And indeed, it is now accepted that most G×E findings of the last decade have not replicated reliably (Duncan & Keller, 2011; Hewitt, 2012). The reasons for this are numerous, but low statistical power is likely a major culprit.

In the decades following the Second World War, most OECD countries introduced major educational reforms, the stated purpose of which was often to expand equality of opportunity. Though the exact details varied by country, these reforms often increased the number of years of compulsory schooling, introduced national curricula, and delayed the age at which children of different abilities were taught separately (Meghir & Palme, 2005). In this chapter, we ask whether the effect of the Swedish comprehensive schooling reform, which increased increased compulsory schooling from seven to nine years, was heterogeneous by genotype. Specifically, we examine how the reform differentially affected labor market outcomes for individuals with different values of a polygenic score constructed from the meta-analysis results of the study in Chapter 4.

The gradual rollout of the Swedish reform generates quasi-experimental variation that has previously been used to study the causal impact of the reform (Meghir & Palme, 2005). We use a similar identification strategy to ask whether the impact of the reform varied by polygenic score. We find evidence of significant interactions between the polygenic score and the reform for females' educational outcomes and earnings. Specifically, higher ability females were more likely to obtain a high school degree, which is beyond the new minimum established by the reform. This is consistent with a model in which employers screen workers based on their educational credentials, and higher ability females have an incentive to acquire more education to better signal their ability after the reform.

Research question 5: To what extent are the statistical power of a GWAS meta-analysis and the predictive accuracy of polygenic scores attenuated by imperfect cross-cohort heterogeneity? (Chapter 6)

Large-scale GWAS results, such as the ones reported in Chapters 3 and 4, are typically obtained by meta-analyzing GWAS results from multiple studies spanning different regions and/or time periods. This approach averages the estimated effects of individual genetic variants across studies. In case genetic effects are heterogeneous across studies, the statistical power of a GWAS and the predictive accuracy of polygenic scores are attenuated, contributing to the so-called "missing" heritability. In Chapter 6, we develop a theoretical multi-

study framework relating statistical power and predictive accuracy to cross-study heterogeneity. This framework is implemented in an online Meta-GWAS Accuracy and Power calculator that enables to explore to what extent an imperfect cross-study genetic correlation contributes to the missing heritability. Simulation studies show that under a wide range of genetic architectures, the statistical power and predictive accuracy inferred by this calculator are accurate. Using the calculator to assess recent GWAS efforts, it is shown that the effect of cross-study genetic correlation on statistical power and predictive accuracy is substantial. Therefore, *a priori* calculations of statistical power and predictive accuracy, accounting for heterogeneity in genetic effects across studies, are an important tool for adequately inferring whether an intended meta-analysis of GWAS results is likely to yield meaningful outcomes.

1.4 Declaration of Contribution

Chapters 2-6 of this thesis are co-authored. In Chapters 2 and 5, the manuscripts have been reported in their entirety because the author was involved in all aspects of the studies. The studies reported in Chapters 3 and 4 are large-scale genome wide association studies with a multitude of follow-up analyses conducted by a core analyst team. In this thesis, all main results of the respective studies are reported in order to give a full picture of the findings, but the details of the analyses in which the author of this thesis did not play a major role, are omitted. Similarly in Chapter 6, the main results are reported in entirety, but only the analyses to which the author contributed substantially are described in detail (Section 6.5). Below, I declare my contribution to each chapter in detail.

Chapter 2 - On Improving the Credibility of Candidate Gene Studies: A Review of Candidate Gene Studies Published in *Emotion*: The evaluation of the candidate gene articles was conducted by the author of this thesis. The methodological challenges were drafted by both authors. Both authors contributed equally to the writing of the manuscript.

Chapter 3 - Genetic Variants Associated with Subjective Well-being, Depressive Symptoms and Neuroticism Identified through Genome-wide Analyses: The author of this thesis conducted the quality control analyses on the GWAS summary statistics uploaded by cohorts together with Bart M. L. Baselmans and David Cesarini. The subjective well-being, neuroticism and depressive symptoms meta-analyses were conducted by the author and Bart M. L. Baselmans. The post-hoc quality control and meta-analysis of subjective well-being (Section 3.4.3) in cohorts with data imputed to the 1000 Genomes reference panel was conducted by the author, along with the cohort-omitted meta-analyses required for the follow-up analyses reported throughout the study. The quasi-replication analyses were performed by David Cesarini, Patick Turley and the author. The look-up of neuroticism and depressive

symptoms SNPs in an independent depression sample was conducted by the author. The proxy-phenotype analyses were conducted by the author, David Cesarini and Richard Karlsson-Linnér. The polygenic prediction analyses were conducted by the author (in the HRS cohort) and Bart M. L. Baselmans (in the NTR cohort, not reported in this thesis). Sections 3.4.1-3.4.4 were written by the author, Bart M. L. Baselmans, and David Cesarini, Section 3.4.5 by the author, Bart M. L. Baselmans and Meike Bartels, and Sections 3.4.6-3.4.7 by the author and David Cesarini. The remaining analyses reported in Section 3.2 were carried out by the following co-authors: Jonathan P. Beauchamp, Patrick Turley, Tonu Esko, Mark Alan Fontana, Jacob Gratten, James J. Lee, S. Fleur W. Meddens, Michel G. Nivard and Harm-Jan Westra. The study was designed and overseen by Meike Bartels, Daniel J. Benjamin, David Cesarini, Philipp D. Koellinger and Robert F. Krueger. All authors contributed to writing and editing the manuscript. Authors not listed above contributed to the recruitment, genotyping, or data processing for the contributing components of the meta-analysis. For a full list of author contributions, see Supplementary Note Section 11-A of Okbay, Baselmans et al (2016).

Chapter 4 - Genome-wide Association Study Identifies 74 Loci Associated with Educational Attainment: The author of this thesis was responsible for the quality control of the summary statistics uploaded by participating cohorts of this study. Specifically, the SNP filtering of all pooled and sex-stratified GWAS summary statistics on the two phenotypes ("years of education" and "college") provided by 64 cohorts were conducted by the author. Diagnostic plots were analyzed by the author, together with David Cesarini and Tonu Esko. The analyses listed under Section 4.2.5-D were conducted jointly with David Cesarini, Guo-Bo Chen and Mark Alan Fontana. The pooled and sex-stratified "years of education" and pooled "college" meta-analyses were conducted by the author, along with cohort-omitted meta-analyses required for follow-up analyses reported in the paper. The author also performed the within- and out-of-sample replication analyses, and the combined discovery and replication sample meta-analysis. The clumping of meta-analysis results into independent loci was done by the author and Tune H. Pers. The author contributed to writing the parts of the main text (Section 4.1) reporting the aforementioned analyses. Section 4.2 was written by the author and David Cesarini. The remaining analyses reported in Section 4.1 were conducted by the following co-authors: (i) stratification: Patrick Turley, Jonathan P. Beauchamp, Cornelius A. Rietveld, and Jian Yang, (ii) genetic overlap: Jonathan P. Beauchamp, Mark Alan Fontana, and Patrick Turley, (iii) biological annotation: James J. Lee, Tonu Esko, Tune H. Pers, Joseph K. Pickrell, Johannes H. Brandsma, Jonathan P. Beauchamp, Lude Franke, Valur Emilsson, Gerardus A. Meddens, Mark Alan Fontana, S. Fleur W. Meddens, Pascal Timshel, Raymond A. Poot, Ronald de Vlaming and Harm-Jan Westra, (iv) prediction and mediation: Jonathan P. Beauchamp, Mark Alan Fontana and Jian Yang, (v) G×E: Dalton

Conley, Steven F. Lehrer, Karl-Oskar Lindgren., Sven Oskarsson and Kevin Thom. (vi) replication GWAS in UKB: Mark Alan Fontana and Cornelius A. Rietveld. The study was designed and overseen by Daniel J. Benjamin, David Cesarini, Tonu Esko, Magnus Johannesson, Philipp D. Koellinger and Peter M. Visscher. All authors contributed to and critically reviewed the manuscript. Authors not listed above contributed to the recruitment, genotyping, or data processing for the contributing components of the meta-analysis. For a full list of author contributions, see Supplementary Information section 8 of Okbay, Beauchamp et al. (2016).

Chapter 5 - Of Genes and Screens: Educational Reform, Ability, and Labor Market Screening. This paper is joint work with Jonathan Beauchamp, Kevin Thom, Sven Oskarsson and David Cesarini. The author constructed the polygenic sores used in the analyses. Sven Oskarsson and Jonathan P. Beauchamp conducted the empirical analyses. Kevin Thom developed the theoretical framework. David Cesarini contributed to data collection or processing. All authors contributed to the design of the study and to the writing of the manuscript.

Chapter 6 - Meta-GWAS Accuracy and Power (MetaGAP) Calculator Shows that Hiding Heritability is Partially due to Imperfect Genetic Correlations across Studies: The preparation and quality control of the phenotype and genotype data, and the GREML heritability and genetic correlation analyses were conducted by the author of this thesis and Ronald de Vlaming. Ronald de Vlaming developed the theoretical framework, the online power calculator in which the framework is implemented, and conducted the simu-lation analyses. The first draft of the manuscript were written by Ronald de Vlaming, the author of this thesis, and Philipp D. Koellinger. Philipp D. Koellinger and A. Roy Thurik oversaw the study. Authors not listed above contributed to the recruitment, genotyping, or processing of the data used in the empirical analyses. All authors critically reviewed the manuscript.

1.5 Implications and Discussion

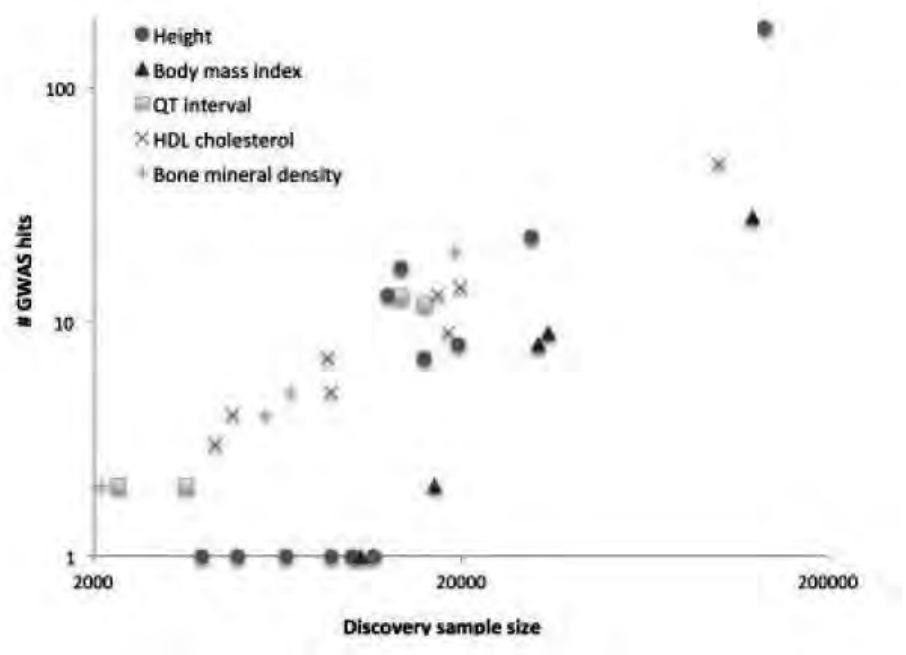
Section 1.1 proposed four broad classes of reasons why the discovery of genetic associations with social-science outcomes may prove scientifically valuable. In this section, I provide an appraisal of the four possible contributions based on the findings of this thesis.

A key “stylized fact” from GWAS studies of medical and anthropometric traits is that as larger and larger discovery samples have become available, the number of genome-wide significant associations has increased steadily (Visscher et al., 2012). This pattern is shown

graphically in Figure 1.1, which plots the number of independent loci identified at genome-wide significance as a function of sample size. Early studies of height identified 10-20 polymorphisms (Gudbjartsson et al., 2008; Weedon et al., 2008). Subsequent studies with larger discovery samples not only confirmed these original associations, but increased the number of identified associations to over 100 (Lango Allen et al., 2010). After the publication of Visscher et al.'s review article, new research by the GIANT consortium (Wood et al., 2014), based on a sample of 250,000 individuals, has identified over 700 SNPs at genome-wide significance. A few years from now, anyone reading this thesis may well marvel at the fact that only 700 height-associated loci were known in 2016.

The findings in chapters 3 and 4 of this thesis suggest that a similar pattern between the size of the discovery sample and the number of identified associations holds for a range social-science traits. To illustrate the point, consider the example of educational attainment. The first genome-wide association study of educational attainment was conducted in a discovery

Figure 1.1. Number of polymorphisms identified as a function of meta-GWAS sample size.



Note: The coordinates are on the logarithmic scale. (Source: Visscher et al. 2012).

sample of 101,069 individuals and identified 3 independent loci at genome-wide significance (Rietveld, Medland, et al., 2013). The follow-up study in Chapter 4 was conducted in a discovery sample of $N = 293{,}723$ individuals and identified 74 loci. In a combined analysis of the discovery and replication samples ($N = 405{,}072$), the number of independent loci further increases to 162. Thus, educational attainment appears to follow a pattern that is qualitatively similar to the medical and anthropometric traits shown in Figure 1.1.

The predictive power of a polygenic score derived from the meta-analysis coefficients has also increased with larger samples. For example, Rietveld, Medland et al. (Rietveld, Medland, et al., 2013) report that a polygenic score derived from their results explains 1.8% of variance in an independent holdout sample from the Swedish Twin Registry (STR). A score based on the discovery stage GWAS in Chapter 4 explains 3.1%, and one based on the pooled discovery and replication samples, about 5.7% of variance in the STR. There is a single unifying principle that can organize many seemingly disparate facts from genetic association studies, including the fact that the number of identified associations and the predictive power of polygenic scores increases with larger samples (Chabris et al., 2015). It is that the heritable variation in most behavioral traits is accounted for by many genetic variants with small effects – a stylized fact that Chabris et al. dub the fourth law of behavioral genetics. The findings in Chapters 3 and 4 strongly suggest this principle applies also to educational attainment and subjective well-being.

I now turn to the four potential contributions listed in Section 1.1.

The first potential contribution was that there are many settings where it is desirable to control for genetic factors. A major limitation to date has been that the explanatory power of a polygenic score based on many SNPs is too small to be useful in many potential applications. An implication of the "fourth law" is that polygenic scores will become increasingly predictive in the years ahead, as larger and larger discovery samples become available. Indeed, it is in fact possible to estimate the total amount of variation explained by common SNPs measured on standard genotyping arrays, and to use such estimates to make predictions about the predictive power of future polygenic scores derived from larger discovery samples (Daetwyler, Villanueva, & Woolliams, 2008). Such projections imply that the predictive power of a polygenic score for educational attainment is likely to exceed 10% in the near future, a level of predictive power at which many new applications become feasible.

Thus, in empirical settings where it does not matter why a polygenic score is predictive (for example, a randomized controlled trial) it seems very likely that in the years ahead, they will increasingly be used by empirical researchers. The analyses in Chapter 6 suggest that additional gains to predictive power will be made possible once association studies can be conducted in very large discovery samples that are environmentally and phenotypically homogenous, thus reducing attenuation in the power of predictive power of the score that is due to imperfect genetic correlations across cohorts. An important caveat is that there are many settings in which the value of the genetic controls does require an understanding of why the polygenic score is predictive and considerable uncertainty remains as to how often such understanding will be possible.

The second potential contribution comes from the possibility of providing individuals with information that may allow them to take preemptive action. Here, my appraisal is considerably more cautious. Though there may be exceptions, it seems unlikely that in the near future, individual risk predictions of behavioral traits based on common variants will have much incremental predictive power beyond factors that are already known, such as family history of risk.

The third potential contribution was that discovery of genetic associations will spur interesting follow-up work, for example on gene-by-environment interactions or behavioral mechanisms that could mediate a relationship between a genetic variant and complex outcome such as education or subjective well-being. As I discuss below, a number of follow-up studies based on the results in Chapters 3 and 4 are already well underway. Indeed, Chapter 5 shows that some applications are already feasible. The chapter systematically evaluates how the effect of the polygenic score (constructed from the results of Chapter 4) interacts with plausibly exogenous variation in the schooling system introduced by the gradual rolling out of the Swedish comprehensive schooling reform of the 1960s. By serendipity, the Swedish genotyped twins (born 1946-1958) come from a dataset in which there is substantial variation in reform status. The study overcomes two primary limitations of the existing G×E literature (Duncan & Keller, 2011), namely that (i) most studies do not rely on exogenous variation in the environmental variable and (ii) many studies suffer from lack statistical power to detect plausible effect sizes. The latter limitation is overcome through the use of a polygenic score.

The fourth potential contribution was that the work could implicate new biological pathways. Bioinformatic analyses of the results in Chapter 4 implicate, with remarkable consistency, a set of gene clusters corresponding to various stages of neural development. These include: the proliferation of neural progenitor cells and their specialization (the cluster

npBAF complex), the migration of new neurons to the different layers of the cortex (forebrain development, abnormal cerebral cortex morphology), the projection of axons from neurons to their signaling targets, the sprouting of dendrites and their spines, and neuronal signaling and synaptic plasticity throughout the lifespan. It is too early to tell whether these and similar discoveries will fundamentally impact how social scientists think about the fundamental dimensions of heterogeneity, but it seems possible that some of the genetic discoveries in Chapters 3 and 4 may contribute to our understanding of the aetiology of neuropsychiatric diseases that are genetically correlated with the outcomes examined here (for example, subjective well-being and major depression). Indeed, one of the genes identified in Chapter 3's analysis of neuroticism is *DRD2*, which encodes the D2 subtype of the dopamine receptor, a target for antipsychotic drugs (Seeman, 2010). The fact that one of the implicated genes operates in a biological pathway targeted by existing drugs suggests that it would be imprudent to dismiss the possibility that some of the other identified genes could provide useful therapeutic leads in the years ahead.

1.6 Conclusion

Though I have emphasized that considerable uncertainty remains about the ultimate value of genetic data in the social sciences, the findings reported here strongly suggest that advances in the years ahead will be rapid. The fourth law implies that more and more variants will be identified with larger discovery samples, and that the predictive power of polygenic scores will continue to increase in the years ahead. In the social sciences, polygenic scores are thus likely to become increasingly valuable for researchers who wish to control genetic confounds in empirical analyses. The research described here also illustrates the importance of assembling strong interdisciplinary teams of researchers when attempting to confront research questions that defy traditional disciplinary boundaries.

Table 1.1 lists the publication status of each chapter. The first three chapters (2-4) have been published in peer-reviewed journals. Of the remaining two, one (Chapter 6) is undergoing revision after a first round of review and the second (Chapter 5) is under preparation for submission.

A large number follow-up studies that utilize the findings from the genome-wide association studies described Chapters 3 and 4 have already been initiated by researchers from a number of disciplines. Many of these studies are still in an early stage, as only a few months have passed between the publication date of the respective chapters and the completion of this

thesis. Table 1.2 provides an overview of the publication status of four such follow-up projects that are in a relatively advanced stage and three other studies that I have contributed to during the writing of this thesis.

Table 1.1. Publication status of chapters.

Chapter	Title	Publication status	Reference
1	Introduction and Conclusion	-	-
2	On Improving the Credibility of Candidate Gene Studies: A Review of Candidate Gene Studies Published in *Emotion*	Published in *Emotion.*	Okbay and Rietveld (2015)
3	Genetic variants associated with subjective well-being, depressive symptoms and neuroticism identified through genome-wide analyses	Published in *Nature Genetics.*	Okbay, Baselmans et al. (2016)
4	Genome-wide association study identifies 74 loci associated with educational attainment	Published in *Nature.*	Okbay, Beauchamp et al. (2016)
5	Of Genes and Screens: Educational Reform, Ability, and Labor Market Screening	Manuscript in preparation	-
6	Meta-GWAS Accuracy and Power (MetaGAP) calculator shows that hiding heritability is partially due to imperfect genetic correlations across studies	Accepted for publication in *PLOS Genetics.*	De Vlaming et al. (2016)

Table 1.2. Publication status of other papers to which I contributed during the writing of this thesis. The papers are ordered alphabetically on the name of first author(s).

Title	Publication status	Reference
Selection against the genetic basis of educational attainment	Accepted for publication in *PNAS.*	Kong et al. (2016)
Genetic variants linked to education predict longevity	Published in *PNAS.*	Marioni et al. (2016)
Molecular genetic contributions to social deprivation and household income in UK Biobank	Published in *Current Biology.*	Hill et al. (2016)
Personality polygenes, positive affect, and life satisfaction	Published in *Twin Research and Human Genetics.*	Weiss et al. (2016)
Meta-analysis on 78,308 individuals identifies 15 novel loci and 36 novel genes for intelligence.	Manuscript under review	Sniekers et al. (2016)
Causal link between education and coronary artery disease.	Manuscript under review	Tillmann et al. (2016)
Genetic heterogeneity in depressive symptoms following the death of a spouse: Polygenic score analysis of the US Health and Retirement Study	Manuscript under review	Domingue et al. (2016)
Bayesian cross-trait meta-analysis.	Manuscript in preparation	Turley et al. (n.d.)
The Relationship Between Genes, Education, and Voting	Manuscript in preparation	Dawes et al. (n.d.)
Y chromosome variation and complex traits: the Ygen consortium.	Manuscript in preparation	Gandin, Joshi, Esko, Wilson, & The Ygen Consortium (n.d.)
Genome-wide association study of 155,439 individuals identifies one locus associated with risk tolerance.	Manuscript in preparation	The Social Science Genetic Associaton Consortium (n.d.)
Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders.	Manuscript in preparation	Weiner et al. (2016)

CHAPTER 2

On Improving the Credibility of Candidate Gene Studies: A Review of Candidate Gene Studies Published in *Emotion*

Based on Okbay and Rietveld (2015).

Abstract

The discovery of genetic variants associated with psychological traits deepens our knowledge about causes and consequences of individual differences. In psychology, the standard approach to identify these variants is the "candidate gene study". In a candidate gene study, a limited set of genetic variants is selected based on their hypothesized or known biological function, and these variants are tested for association with the psychological trait of interest. The successful replication of published candidate gene studies, however, is alarmingly scarce. In this paper we describe the challenges to successfully identifying genetic associations, and review the candidate gene studies published in *Emotion*. We argue that the implementation of four methodological guidelines developed by the Behavior Genetics Association for evaluating candidate gene studies will help to increase the credibility of candidate gene study findings.

2.1 Introduction

The increasing availability of genetic data has fueled the investigation of the role of genetic variation in determining behavior (Benjamin, Cesarini, Chabris, et al., 2012; J. H. Fowler & Schreiber, 2008). The most frequently adopted approach in the social sciences in general and psychology in particular, is to conduct a 'candidate gene study' to link genetic markers to outcomes of interest (Benjamin, Cesarini, van der Loos, et al., 2012; Ebstein et al., 2010). A candidate gene study tests the hypothesized association between a set of pre-specified genes and a phenotype (an observable characteristic or trait), rather than using the full set of genes (Zhu & Zhao, 2007).

Usually, the pre-selection of genetic variants is based on the known biological function of genes and their possible functional relation to the outcome. Such ex ante theories reduce the number of hypotheses being tested in empirical research. However, in the context of behavioral genetics it is difficult to reduce the number of plausible hypotheses purely on theoretical grounds. The majority (~70%) of all human genes (~20,000 in total) is expressed in the brain (Ramsköld, Wang, Burge, & Sandberg, 2009). Together with the fact that the knowledge about the exact functioning of our genes is still very limited, seemingly plausible biological links to behavior could be hypothesized for a large number of genes. Thus, it requires specific attention in candidate gene studies why a specific variant is chosen to study, rather than other genetic variants or the full genome as in genome-wide association studies (Visscher et al., 2012). In the worst case scenario, ignorance of competing hypotheses could lead to the presentation of empirical findings as ex ante theory.

So far, the findings from candidate gene studies of complex traits - medical and behavioral - have rarely been successfully replicated (Cardon & Palmer, 2003; Duncan & Keller, 2011; Ioannidis, 2005). The general non-replicability of findings tempers the enthusiasm surrounding the discovery of genetic causes of psychological traits considerably. Therefore, *Behavior Genetics*, a leading journal in the field of genetic analysis of behavior, issued an editorial policy in 2012 to reduce the number of false positives in candidate gene association and candidate gene × environment (G×E) interaction studies of human behavior (Hewitt, 2012), which was also adopted recently by *The Journal of Abnormal Child Psychology* (Johnston, Lahey, & Matthys, 2013) and Psychological Science. Only studies that meet at least one of the following four criteria are now considered for publication in these journals:

- It is a rigorously conducted, adequately powered, direct replication study of a previously reported result.

- It is an exploratory study or test of a novel hypothesis but with an adequately powered, direct replication study reported in the same paper.
- It is an exploratory analysis or test of a novel hypothesis in the context of an adequately powered study, and the finding meets the statistical criteria for genome wide significance taking into account all sources of multiple testing (e.g., phenotypes, genotypes, environments, covariates, subgroups).
- It is a meta-analysis of several or many studies addressing the same genetic variant and/or environmental variable and the same behavioral outcome.

In this paper, we explain why adherence to these criteria increases the validity of results in candidate gene studies and we provide a review of all candidate gene studies published in Emotion until April 2014 in terms of these guidelines (Table 2.1)[4]. These studies report associations between genetic variants (5-HTTLPR polymorphism in the promoter region of the serotonin transporter gene, the monoamine oxidase A (MAOA) gene, the nicotinic acetylcholine receptor gene (CHRNA4), oxytocin (OXTR) and vasopressin receptor genes (AVPR1A), and the brain-derived neurotrophic factor gene (BDNF)) and a range of psychological phenotypes (emotional reactions to stress, rumination in healthy adults, anger reactivity and control, sensitivity to negative emotional cues, sensitivity to positive and negative affect in marriage, stress in children, emphatic and self-conscious emotional reactivity, social anxiety, positive and negative emotionality, orienting of spatial attention, and selective attention to threat).

Below, we describe the methodological challenges that candidate gene studies face and illustrate these points by examples from the candidate gene studies published in *Emotion*. The paper is structured around three themes: power and credibility, model specification, and statistical significance. The four methodological guidelines are discussed within these three themes.

2.2 Power and Credibility

The aim of a candidate gene study is to find genetic variants associated with the phenotype of interest and to measure the strength of these associations. Genetic variants explaining

[4] Our review also includes studies published before the issue of the Behavior Genetics editorial. We decided to include these as well to provide a complete overview of published candidate gene studies in Emotion. Where our methodological remarks in this paper on candidate gene studies are of general nature, we acknowledge that we give them with hindsight

Table 2.1: Overview of the candidate gene and candidate gene × environment studies published in Emotion until April 2014. The studies are ordered on date of publication.

First author	Year of publication	Sample size	Heterogeneous ethnicity?	Proper control for population stratification?	Tests gene × environment interaction?	Control × gene and control × environment terms included?	Multiple testing correction?
Ford	2014	205	Yes	No	Yes	No	No
Moons	2014	172	Yes	No	Yes	No	No
Papousek	2013	165	No	N/A	No	N/A	No
Haase	2013	125	Yes	No	Yes	No	Yes
Miu	2013	182	No	N/A	No	N/A	No
Gyurak	2013	106 - 163	Yes	No	Yes	No	No
Osinsky	2012	120	Unknown	N/A	Yes	N/A	No
Schoebi	2012	152	Yes	No	Yes	No	No
Carlson	2012	51	Unknown	N/A	Yes	N/A	No
Markett	2011	574	No	N/A	No	N/A	No
Hayden	2010	413	Unknown	N/A	Yes	No	No
Beevers	2009	71	Yes	No	No	N/A	No
Alia-Klein	2009	27	Unknown	N/A	Yes	N/A	No
Osinsky	2008	50	Unknown	N/A	Yes	N/A	No

Note: N/A [Population stratification]: Not applicable, the sample is homogenous. N/A [Control × gene and control × environment terms included?]: Not applicable either because it is not a gene × environment study, or because there are no covariates.

more than 0.3% of an outcome – medical or behavioral - are very rare (Benjamin, Cesarini, Chabris, et al., 2012). For example, the largest genetic discovery study for a medical outcome – human height – with N = 283,288 found that the strongest genetic variant explains only 0.4% of the variance in human height (Lango Allen et al., 2010; Wood et al., 2014), and the largest genetic discovery study on a behavioral outcome – educational attainment – with N = 126,559 found that the strongest associated genetic variant explains only 0.02% of the variance in educational attainment (Rietveld, Medland, et al., 2013). Generally speaking, genetic discovery studies show that for the vast majority of traits, it is not the case that a single gene or genetic variant is responsible for the trait, but rather a large number of genetic variants, all of which have very tiny effects (Visscher et al., 2012)

This finding has consequences for the determination of the required sample size in empirical studies. The sample size should be such that the statistical test has adequate statistical power. The power of a test equals the probability of correctly rejecting the null hypothesis when the null hypothesis is false. Therefore, high statistical power is a prerequisite to get meaningful results from statistical hypothesis testing by reducing the likelihood that the findings are false positives. The power of a test is dependent on the sample size, effect size, and the adopted statistical significance level. In Table 2.2, we present some illustrative power calculations for a true association of $R^2 = 0.1\%$, given certain sample sizes and a significance level of $p = 0.05$. We choose an R^2 of 0.1% because it is very large with respect to the replicable findings for educational attainment (Rietveld, Conley, et al., 2014; Rietveld, Medland, et al., 2013), but moderate as compared to findings for height (Lango Allen et al., 2010; Wood et al., 2014).

Furthermore, we show in Table 2.2 the posterior probability that an association is true as a function of the prior probability and power using Bayes' rule in the spirit of Ioannidis (2005) and Benjamin, Cesarini, Chabris, et al. (2012). In statistical hypothesis testing, one starts with a prior belief in the hypothesis being true. In a Bayesian context, this prior belief is updated by the result of the statistical test. If the statistical result is strongly in favor of the hypothesis, the posterior belief in the hypothesis will be larger than the prior belief. If the statistical results run counter to the prior belief, the posterior belief will be lower than the prior belief. Since many other genes in addition to the chosen candidates can be hypothesized to influence the outcome of interest, the prior belief in the hypotheses may be relatively low in candidate gene studies. Benjamin, Cesarini, Chabris, et al. (2012) argue that for a typical candidate genetic variant the prior belief in the association is likely to be much less than 10%. Taking this into consideration is even more important for candidate gene studies in the field of psychology and behavior, because a large proportion of genes are expressed in the

Table 2.2: Posterior probability of a true association of R^2 = 0.1% that is significant at p = 0.05 as a function of the prior probability and sample size.

Sample Size		100	500	1000	10000	100000
Power		6%	11%	17%	89%	100%
Prior probability of true association	**0.01%**	0.0001	0.0002	0.0003	0.0018	0.002
	1%	0.01	0.02	0.03	0.15	0.17
	10%	0.12	0.19	0.27	0.66	0.69

Note: Power is calculated using the R package 'pwr'. Posterior probabilities are calculated by Bayes' rule: P(true|significant) = (power × prior)/((power × prior) + (0.05 × (1 – prior))).

brain, making it difficult to rule out genes other than the chosen candidates as influencing the trait. If it is difficult to exclude such a possible relation for certain genetic variants beforehand, one should take into account the large amount of competing hypotheses in the calculation of the prior belief. Therefore, in our calculation we evaluate prior beliefs of 0.01%, 1% and 10%.

It is possible to update (upgrade or downgrade) prior beliefs about a genetic association as new insights into the underlying biology of the phenotype or the functioning of the candidate genes are gained. Such knowledge, as long as it was obtained adhering to rigorous scientific standards, can provide a sound basis for the selection of candidate genes for association testing with phenotypes of interest. An empirical approach toward this matter was recently proposed by Rietveld, Conley, et al. (2014). Their 'proxy-phenotype' approach prioritizes certain genetic variants to be associated with cognitive performance due to the association of the same variants with educational attainment. Such a strategy is useful if large genetically informed samples exists for the proxy-phenotype, but not for the actual phenotype of interest, and is likely to be followed more often in behavioral genetics.

Guidelines 1, 2 and 3 from the Behavior Genetics Association editorial all underscore that a candidate gene study should be "adequately powered". The calculations shown in Table 2.2 illustrate that large samples are needed to identify phenotype - genotype associations. In a scenario of an expected effect size of R^2 = 0.1% and a prior belief in the association of 1%, the power of the test equals a respectable 89% if the sample size is 10,000. However, the posterior belief in a significant association at the 5% level is still only 15%. These example calculations make clear that large sample sizes are needed to conduct an adequately powered candidate gene study. The sample sizes in the reviewed articles (Table 2.1) range only between 27 and 574. Therefore, the statistical power in these studies is likely to be very limited.

The underpowered nature of the reviewed papers is sometimes less apparent due to the fact that these papers usually report effect sizes much larger than $R^2 = 0.1\%$. Such effect sizes, if used in an ex post power calculation, imply much higher statistical power than when power is calculated using more conservative, but in our view more reasonable, effect sizes (such as $R^2 = 0.1\%$) in an ex ante power analysis. These large effect size estimates may be correct within the specific (often small) analysis sample, but it should be noted that they usually come together with large standard errors. Moreover, the often limited control for confounders may lead to overestimation of the effect of the genetic variant due to omitted variable bias. Hence, the fact that a test has detected a statistically significant association does not nullify the concern that the test in question is underpowered.

When effect sizes and prior beliefs in the hypotheses are low, replication is of utmost importance to improve the credibility of a statistical finding. The guidelines of the Behavior Genetics Association emphasize that a candidate gene study should either itself be the replication of a previously reported association, or if it is testing a novel hypothesis, should also contain an adequately powered direct replication. A proper replication study demonstrates the robustness of the discovered genetic association across different samples and it improves the power of the statistical test by enlarging the analysis sample. Moreover, it will give a more precise estimate of the genetic effect because the standard errors of the regression coefficient will be smaller in the combined discovery and replication sample. None of the articles that we reviewed (Table 2.1), however, are replication studies themselves, nor do they contain a direct replication of their own findings. The small discovery sample sizes that lead to low statistical power and the lack of replication of the findings imply that the posterior probability that the reported associations are true is rather low in these studies. Adequately powered replications are needed in order to improve the credibility of their findings.

Gathering the behavioral and genetic data needed to conduct a candidate gene study may be laborious and time-consuming, making replication of the findings in an independent sample cumbersome. However, the reality is that genetic effects are small, and the publication of statistically significant findings from studies with small sample sizes lacking the power to detect such effects stunts scientific progress (Hewitt, 2012). In the short run, requiring replication of the findings prior to publication is helpful in avoiding the accumulation of false positives and unreliable scientific knowledge. If the premise is granted that the genetic effects are small, however, the long run response should be to focus on obtaining large sample sizes. This can be achieved by utilizing publically available datasets that contain information on outcomes correlated with the phenotype of interest. Rietveld, Conley, et al. (2014) show that large sample size of a proxy-phenotype can provide higher statistical power for SNP discovery than smaller sample size of the actual phenotype. Hence, conducting a genetic

association study on a proxy-phenotype available in a large dataset can be a practical alternative when extensive data collection for a study on the actual phenotype of interest is not feasible for various reasons. Many such large datasets, such as the Health and Retirement Study (Sonnega et al., 2014) can be accessed through the database of Genotypes and Phenotypes (dbGaP; Mailman et al., 2007). Larger datasets, such as UK Biobank with data on 500,000 individuals, are becoming increasingly available as well, signaling a new era in genetic association studies where the focus is on large sample sizes (Allen, Sudlow, Peakman, & Collins, 2014).

In many applications, researchers are interested in the joint influence of many genetic variants (which can nowadays be relatively inexpensively measured) combined into so-called polygenic scores (Purcell et al., 2009). In such cases, the extensive measures of environmental moderators and behavioral outcomes in many candidate-gene datasets can be immensely useful for learning more about mechanisms and environmental moderators

2.3 Model Specification

A correctly specified model is a prerequisite to find credible associations between genetic variants and outcomes. This is emphasized by the guidelines of the Behavior Genetics Association as “rigorous conduct” of genetic association studies. Our review shows that there are two issues of importance to be discussed in this section. The first issue is incomplete or incorrect correction for potential confounders. The second issue is population stratification, meaning appropriate control for systematic differences in genetic makeup across (sub)populations. We explain these issues in more detail below.

2.3.1 Controlling for confounders in candidate G×E interaction studies

An important aspect of a good empirical model is that it controls for potential confounders to overcome omitted variable bias. Sex, age and ethnicity are well-known examples of such confounders that are often included in models of human behavior. Specific consideration should be given to controlling for confounders in candidate G×E studies. Although the common practice is to include potential confounders as covariates in general linear models, this practice falls short of properly adjusting for confounders in G×E interaction studies because it controls only for the influence that the potential confounders may have on the main effects of the gene and environment. The effects of these confounders on the G×E interaction term stay uncontrolled for unless covariate × environment (C×E) and covariate × gene (C×G) interactions are also entered in the model (Keller, 2014).

Among the studies that we reviewed (Table 2.1), five are candidate G×E interaction studies that control for potential confounders in their models (the other five G×E studies do not control for any covariates). However, none of the five studies control for all necessary C×E and C×G interactions. Furthermore, Haase et al. (2013) and Ford et al. (2014) test three-way interactions in their analyses to establish the generalizability of results in the former and as a main hypothesis in the latter. Following Keller (2014), in models that include three-way interactions, it is necessary to add all the relevant three-way interactions involving covariates into the model in addition to the two-way C×E and C×G interactions in order to be able to interpret the statistical significance of the three-way interaction term. Hence, confounding factors remain uncontrolled for in these five G×E interaction studies, which may result in spurious findings.

2.3.2 Population stratification

The existence of systematic differences in the allele frequencies between different subpopulations is called population stratification (Hamer & Sirota, 2000). Since genetic variation is often correlated with environmental confounders such as culture, population stratification can confound associations reported in genetic discovery studies when the empirical model fails to properly control for it. If differing allele frequencies of the genetic variant across ancestries is coupled with differences in phenotype distribution across ancestries, the detected associations can be spurious as it can simply reflect diverging genetic and social history of populations (Cardon & Palmer, 2003). This can be an issue in ethnically homogenous samples as well, since the frequency of genetic variants can vary across subgroups. Therefore, the common practice in genome-wide association studies is to restrict the sample to an ethnically homogenous group and then control for any remaining population substructure by including principal components of the genetic variance-covariance matrix as covariates in the regressions (Price et al., 2006). However, this method requires genome-wide data which is often not available in candidate gene (× environment) studies. As a result, most studies attempt to solve the population stratification problem by controlling for self-reported ethnicity.

Five of the studies that we reviewed in Emotion do not report any information regarding the ethnicity of the participants, nor do they have any controls for population stratification (Table 2.1). Among the remaining nine studies, six make use of ethnically heterogeneous samples and control for self-reported ethnicity (Table 2.1). Most of these studies (Ford et al., 2014; Gyurak et al., 2013; Haase et al., 2013; Moons et al., 2014) control for ethnicity with a dummy variable for being Caucasian, after pooling different non-Caucasian ethnicities together. Beevers, Wells, and McGeary (2009) do not control for self-reported race, because

they did not find significant differences between ethnicities in terms of allele frequencies or the outcome variable.

Controlling for self-reported ethnicity, however, may fail to properly control for population stratification. The often subtle population stratification within samples is usually not picked up by dummy variables for ethnicity, especially if different ethnicities are pooled together into one category (e.g. being Caucasian versus non-Caucasian). Rietveld, Conley, et al. (2014) illustrate this point by showing that controlling for principal components eliminates a spurious association between educational attainment and a SNP for lactose intolerance that is known to vary in frequency across groups, whereas the association remains significant when only self-reported ethnicity is controlled for. They also show that this SNP is significantly associated with educational attainment within Caucasians, but the association becomes insignificant when principal components are included in the model. Thus, in the absence of whole genomic data and hence the possibility to include principal components as covariates, candidate gene studies may at least want to verify carefully that their samples consist of individuals of the same ethnic origin.

Markett, Montag, and Reuter (2011), Papousek et al. (2013), and Miu, Vulturar, Chis, Ungureanu, and Gross (2013) restrict their samples to ethnically homogenous participants. In the same spirit, Haase et al. (2013) and Gyurak et al. (2013) perform robustness checks by repeating their analyses in the Caucasian subsample, and report that the results remain essentially the same, but some associations were reduced to trend level or their effects were attenuated. Although it is difficult to say whether this is due to the reduced power or the confounding effect of ethnicity being ruled out, we believe that the results that do not hold in the Caucasian subsample should be interpreted carefully.

2.3.3 Statistical significance

Empirical studies generally predefine the statistical significance level (usually 5%) in order to be able to judge how likely it is that the estimated effects are due to mere chance. This probability, however, is defined relative to a single statistical test. With multiple statistical tests, the family-wise false positive rate gains importance. If m statistical tests are performed at a 5% significance level, the probability of reporting a false positive becomes $1 - 0.95^m$, and thus approaches 1 when m is large. Therefore, choosing the appropriate significance level by taking into account the number of tests is a crucial statistical requirement to keep the rate of false positive findings in a study at a reasonable level. Since this is a bigger concern for genome-wide association studies which test millions of independent hypotheses, it tends to be ignored in candidate gene studies where the number of hypotheses is relatively small. Our review of all candidate gene and G×E interaction studies published in *Emotion*

revealed that only one study, Haase et al. (2013), control for multiple testing whereas all studies test multiple hypotheses.

The standard practice in the medical genetics literature to control for the family-wise false positive rate is to adjust the level of significance using Bonferroni correction by dividing the desired family-wise type-I error rate by the number of tests (Benjamini, Drai, Elmer, Kafkafi, & Golani, 2001). This approach is sometimes considered overly conservative since it assumes that all hypotheses being tested are independent. Haase et al. (2013) follow an alternative approach by controlling for the false discovery rate (FDR). The difference between the two notions is that, given a rule for significance, the false positive rate is the rate where truly null features are called significant, whereas the FDR is the rate that significant features are truly null. Unless all null hypotheses are true, the FDR procedure is less strict than controlling for the familywise false positive rate. The exact procedure followed by Haase et al. (2013) is controlling for positive FDR (pFDR), which is defined as the expectation of the proportion of falsely rejected hypotheses conditional on at least one hypothesis having been rejected (Storey & Tibshirani, 2003). This method depends on the assumption that the number of tests is very large (approaching infinite), such as in genome wide association studies. Therefore, controlling for pFDR in a study with small number of tests, may be less appropriate than the Bonferroni correction.

The third guideline of the Behavior Genetics Association emphasizes that findings should meet "the statistical criteria for genome wide significance taking into account all sources of multiple testing (e.g., phenotypes, genotypes, environments, covariates, subgroups)" (Hewitt, 2012). This implies that researchers have to give full account of the analyses (reported and non-reported) performed for the project. A good strategy to ensure this transparency is the publication of an analysis plan in a publicly accessible registry system, prior to the actual conduct of the statistical analysis in a data set. In a data set containing 100 variables, one could find 247 significant bivariate correlations at the 5% level by pure chance. With the potential of so many "false positives" it is useful to register an analysis before conducting it, because this makes it transparent that no selection on results has taken place. Such an analysis plan can also contain follow-up studies on significant findings using more precisely defined phenotypes to shed more light on the mechanisms at play between genetics and behavior.

2.4 Conclusion

Theoretical hypothesis-based research in the genetics of psychology requires caution as it is easy to fall into the trap of ex post rationalization of observed empirical associations between

genetic variants and psychological traits. Low statistical power due to small sample sizes, lack of replication, not properly controlling for population stratification and other confounders, especially in candidate gene × environment studies, and not performing any multiple testing correction on the chosen statistical significance level are some of the most common pitfalls that candidate gene studies fall into. We evaluated all candidate gene and gene × environment studies published in *Emotion* until April 2014 focusing on these issues. Our review shows that all of the studies that we evaluated fall short in at least one of these domains, and most of them in nearly all. We believe that some methodological steps should be taken to improve the credibility of candidate gene studies. The guidelines developed by the Behavior Genetics Association provide useful recommendations in this regard. Adherence to these or similar guidelines will raise the credibility of candidate gene studies.

CHAPTER 3

Genetic Variants Associated with Subjective Well-being, Depressive Symptoms and Neuroticism Identified through Genome-wide Analyses

Based on Okbay, Baselmans, et al. (2016)

Abstract

We conducted genome-wide association studies of three phenotypes: subjective well-being (N = 298,420), depressive symptoms (N = 161,460), and neuroticism (N = 170,910). We identified three variants associated with subjective well-being, two with depressive symptoms, and eleven with neuroticism, including two inversion polymorphisms. The two depressive symptoms loci replicate in an independent depression sample. Joint analyses that exploit the high genetic correlations between the phenotypes ($|\hat{\rho}| \approx 0.8$) strengthen the overall credibility of the findings, and allow us to identify additional variants. Across our phenotypes, loci regulating expression in central nervous system and adrenal/pancreas tissues are strongly enriched for association.

3.1 Introduction

Subjective well-being—as measured by survey questions on life satisfaction, positive affect, or happiness—is a major topic of research within psychology, economics, and epidemiology. Twin studies have found that subjective well-being is genetically correlated with depression (characterized by negative affect, anxiety, low energy, bodily aches and pains, pessimism, and other symptoms) and neuroticism (a personality trait characterized by easily experiencing negative emotions such as anxiety and fear) (Bartels, Cacioppo, van Beijsterveldt, & Boomsma, 2013; Kendler & Myers, 2009; Weiss, Bates, & Luciano, 2008). Depression and neuroticism have received much more attention than subjective well-being in genetic-association studies, but the discovery of associated genetic variants with either of them has proven elusive (De Moor et al., 2015; Hyman, 2014).

In this paper, we report a series of separate and joint analyses of subjective well-being, depressive symptoms, and neuroticism. Our primary analysis is a genome-wide association study (GWAS) of subjective well-being based on data from 59 cohorts (N = 298,420). This GWAS identifies three loci associated with subjective well-being at genome-wide significance ($p < 5\times10^{-8}$). We supplement this primary analysis with auxiliary GWAS meta-analyses of depressive symptoms (N = 180,866) and neuroticism (N = 170,910), performed by combining publicly available summary statistics from published studies with new genome-wide analyses of additional data. In these auxiliary analyses we identify two loci associated with depressive symptoms and eleven with neuroticism, including two inversion polymorphisms. In depression data from an independent sample (N = 368,890), both depressive symptoms associations replicate (p = 0.004 and p = 0.015).

In our two joint analyses, we exploit the high genetic correlation between subjective well-being, depressive symptoms, and neuroticism (i) to evaluate the credibility of the 16 genome-wide significant associations across the three phenotypes, and (ii) to identify novel associations (beyond those identified by the GWAS). For (i), we investigate whether our three subjective well-being-associated SNPs "quasi-replicate" by testing them for association with depressive symptoms and neuroticism. We similarly examine the quasi-replication record of the depressive symptoms and neuroticism loci by testing them for association with subjective well-being. We find that the quasi-replication record closely matches what would be expected given our statistical power if none of the genome-wide significant associations were chance findings. These results strengthen the credibility of (most of) the original associations. For (ii), we use a "proxy phenotype" approach (Rietveld, Esko, et al., 2014): we treat the set of loci associated with subjective well-being at $p < 10^{-4}$ as candidates, and we

test them for association with depressive symptoms and neuroticism. At the Bonferroni-adjusted 0.05 significance threshold, we identify two loci associated with both depressive symptoms and neuroticism and another two associated with neuroticism.

In designing our study, we faced a tradeoff between analyzing a smaller sample with a homogeneous phenotype measure versus attaining a larger sample by jointly analyzing data from multiple cohorts with heterogeneous measures. For example, in our analysis of subjective well-being, we included measures of both life satisfaction and positive affect, even though these constructs are conceptually distinct (Kahneman & Deaton, 2010). In the Supplementary Note and Supplementary Figure 1 of Okbay, Baselmans, et al. (2016), a theoretical framework is presented for evaluating the costs and benefits of pooling heterogeneous measures. In our context, given the high genetic correlation across measures, the framework predicts that pooling increases statistical power to detect variants. This prediction is supported by our results.

3.2 Results

3.2.1 GWAS of subjective well-being

Following a pre-specified analysis plan, we conducted a sample-size-weighted meta-analysis (N = 298,420) of cohort-level GWAS summary statistics. The phenotype measure was life satisfaction, positive affect, or (in some cohorts) a measure combining life satisfaction and positive affect. We confirmed previous findings (Bartels & Boomsma, 2009) of high pairwise genetic correlation between life satisfaction and positive affect using bivariate LD Score regression ($\hat{\rho}$ = 0.981, SE = 0.065; Appendix A - Table A1; Bulik-Sullivan, Loh, et al., 2015). Details on the 59 participating cohorts, their phenotype measures, genotyping, quality-control filters, and association models are provided in Section 3.4.1, Tables Table A2-Table A3 (Appendix A), and Supplementary Tables 2-6 in Okbay, Baselmans, et al. (2016).

As expected under polygenicity (Yang, Weedon, et al., 2011), we observe inflation of the median test statistic (λ_{GC} = 1.206). The estimated intercept from LD Score regression (1.012) suggests that nearly all of the inflation is due to polygenic signal rather than bias. We also performed family-based analyses that similarly suggest minimal confounding due to population stratification. Using a clumping procedure (Section 3.4.2), we identified three approximately independent SNPs reaching genome-wide significance ("lead SNPs"). These three lead SNPs are indicated in the Manhattan plot (Figure 3.1a) and listed in Table 3.1. The SNPs have estimated effects in the range 0.015 to 0.018 standard deviations (SDs) per allele (each $R^2 \approx 0.01\%$).

We also conducted separate meta-analyses of the components of our subjective well-being measure, life satisfaction (N = 166,205) and positive affect (N = 180,281) (Section 3.4.2). Consistent with our theoretical conclusion that pooling heterogeneous measures increased power in our context, the life satisfaction and positive affect analyses yielded fewer signals across a range of p-value thresholds than our meta-analysis of subjective well-being (Table A5).

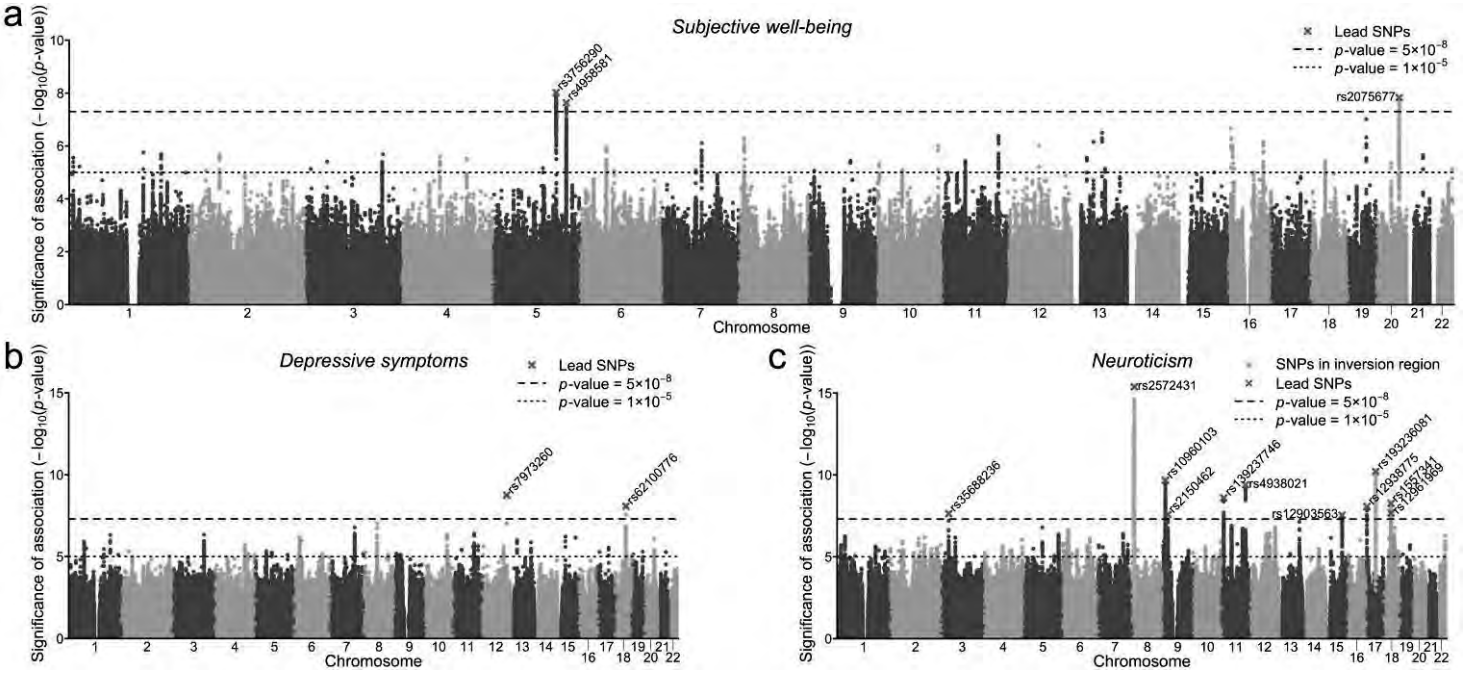
3.2.2 GWAS of depressive symptoms and neuroticism

We conducted auxiliary GWAS of depressive symptoms and neuroticism (see Section 3.4.4, Appendix A - Table A3, Table A6, and Supplementary Tables 8-11 in Okbay, Baselmans, et al., 2016) for details on cohorts, phenotype measures, genotyping, association models, and quality-control filters). For depressive symptoms (N = 180,866), we meta-analyzed publicly available results from a study performed by the Psychiatric Genomics Consortium (PGC; Ripke et al., 2013) together with new results from analyses of the initial release of the UK Biobank data (UKB; Sudlow et al., 2015) and the Resource for Genetic Epidemiology Research on Aging Cohort (GERA; dbGaP, 2015). In UKB (N = 105,739), we constructed a continuous phenotype measure by combining responses to two questions, which ask about the frequency in the past two weeks with which the respondent experienced feelings of unenthusiasm/disinterest and depression/hopelessness. The other cohorts had ascertained case-control data on major depressive disorder (GERA: N_{cases} = 7,231, $N_{controls}$ = 49,316; PGC: N_{cases} = 9,240, $N_{controls}$ = 9,519).

For neuroticism (N = 170,910), we pooled summary statistics from a published study by the Genetics of Personality Consortium (GPC; De Moor et al., 2015) with results from a new analysis of UKB data. The GPC (N = 63,661) harmonized different neuroticism batteries. In UKB (N = 107,245), our measure was the respondent's score on a 12-item version of the Eysenck Personality Inventory Neuroticism scale (Eysenck & Eysenck, 1975).

In both the depressive symptoms and neuroticism GWAS, the heterogeneous phenotypic measures are highly genetically correlated (Table A1). As in our subjective well-being analyses, there is substantial inflation of the median test statistics (λ_{GC} = 1.168 for depressive symptoms, λ_{GC} = 1.317 for neuroticism), but the estimated LD Score intercepts (1.008 and 0.998, respectively) suggest that bias accounts for little or none of the inflation.

Figure 3.1. Manhattan plots of GWAS results.



Note: Results are shown for subjective well-being ($N = 298{,}420$) (a), depressive symptoms ($N = 180{,}866$) (b), and neuroticism ($N = 170{,}911$) (c). The x axis shows chromosomal position, and the y axis shows association significance on a −log10 scale The upper dashed line marks the threshold for genome-wide significance ($p = 5 \times 10^{-8}$), and the lower dashed line marks the threshold for nominal significance ($p = 1 \times 10^{-5}$). Each approximately independent genome-wide significant association (lead SNP) is marked by a red ×. Each lead SNP is the SNP with the lowest p-value within the locus, as defined by our clumping algorithm (Section 3.4.2).

For depressive symptoms, we identified two lead SNPs, indicated in the Manhattan plot (Figure 3.1b). For neuroticism, our meta-analysis yielded 16 loci that are independent according to our locus definition (Figure 3.1c). However, 6 of these reside within a well-known inversion polymorphism (Tian et al., 2008) on chromosome 8. We established that all genome-wide significant signals in the inversion region are attributable to the inversion, and we confirmed that the inversion is associated with neuroticism in both of our neuroticism datasets, the GPC and the UKB (Okbay, Baselmans, et al., 2016, Supplementary Note 5A). In our list of lead SNPs (Table 3.1), we only retain the most strongly associated SNP from these 6 loci to tag the chromosome 8 inversion.

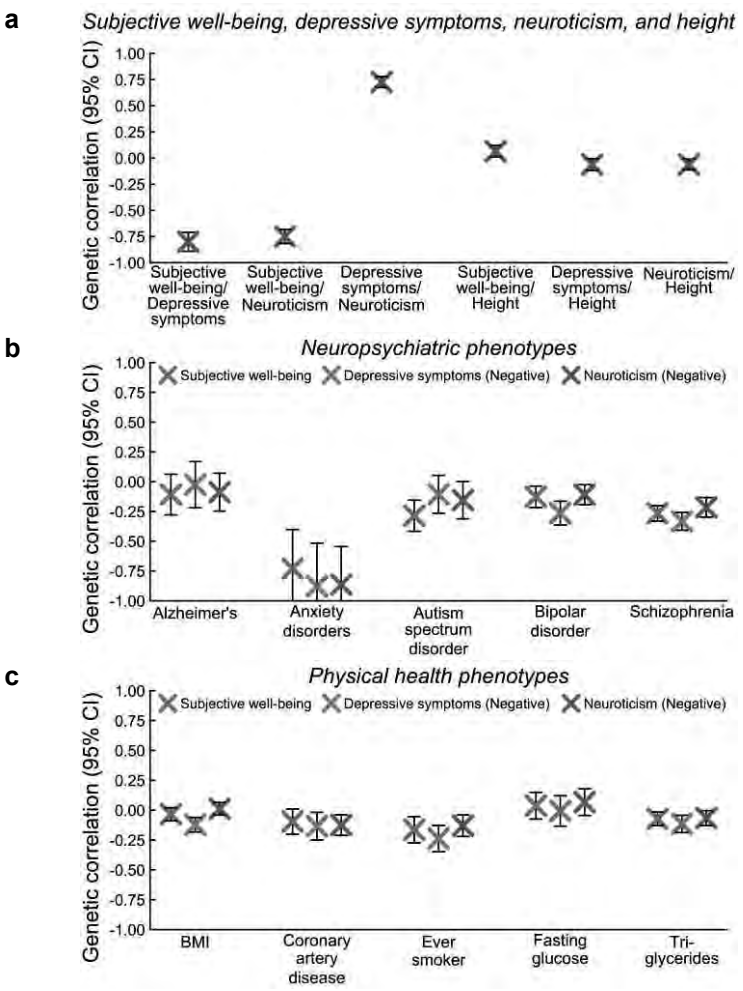
Another lead SNP associated with neuroticism, rs193236081, is located within a well-known inversion polymorphism on chromosome 17. We established that this association is attributable to the inversion polymorphism (Okbay, Baselmans, et al., 2016, Supplementary Note 5B). Because this inversion yields only one significant locus and is genetically complex (Steinberg et al., 2012), we hereafter simply use its lead SNP as its proxy. Our neuroticism GWAS therefore identified 11 lead SNPs, two of which tag inversion polymorphisms. A concurrent neuroticism GWAS using a subset of our sample reports similar findings (Smith et al., 2015).

As shown in Table 3.1, the estimated effects of all lead SNPs associated with depressive symptoms and neuroticism are in the range 0.020 to 0.031 SDs per allele ($R^2 \approx 0.02\%$ to 0.04%). In the UKB cohort we estimated the effect of an additional allele of the chromosome 8 inversion polymorphism itself on neuroticism to be 0.035 SDs (Okbay, Baselmans, et al., 2016, Supplementary Table 13). The inversion explains 0.06% of the variance in neuroticism (roughly the same as the total variance explained jointly by the 6 SNPs in the inversion region).

3.2.3 Genetic overlap across subjective well-being, depressive symptoms, and neuroticism

Figure 3.2a shows that the three pairwise genetic correlations between our phenotypes, estimated using bivariate LD Score regression (Bulik-Sullivan, Loh, et al., 2015), are substantial: -0.81 (SE = 0.046) between subjective well-being and depressive symptoms, -0.75 (SE = 0.034) between subjective well-being and neuroticism, and 0.75 (SE = 0.027) between depressive symptoms and neuroticism. Using height as a negative control, we also examined pairwise genetic correlations between each of our phenotypes and height and, as expected, found all three to be modest, e.g., 0.07 with subjective well-being (Appendix A, Table A1).

Figure 3.2. Genetic correlations with bars representing 95% confidence intervals.



Note: The correlations are estimated using bivariate LD Score (LDSC) regression. (**a**) Genetic correlations between subjective well-being, depressive symptoms, and neuroticism ("our three phenotypes"), as well as between our three phenotypes and height. (**b**) Genetic correlations between our three phenotypes and selected neuropsychiatric phenotypes. (**c**) Genetic correlations between our three phenotypes and selected physical health phenotypes. In (**b**) and (**c**), we report the negative of the estimated correlation with depressive symptoms and neuroticism (but not subjective well-being).

The high genetic correlations between subjective well-being, depressive symptoms, and neuroticism may suggest that the genetic influences on these phenotypes are predominantly related to processes common across the phenotypes, such as mood, rather than being phenotype-specific.

3.2.4 Quasi-replication and Bayesian credibility analyses

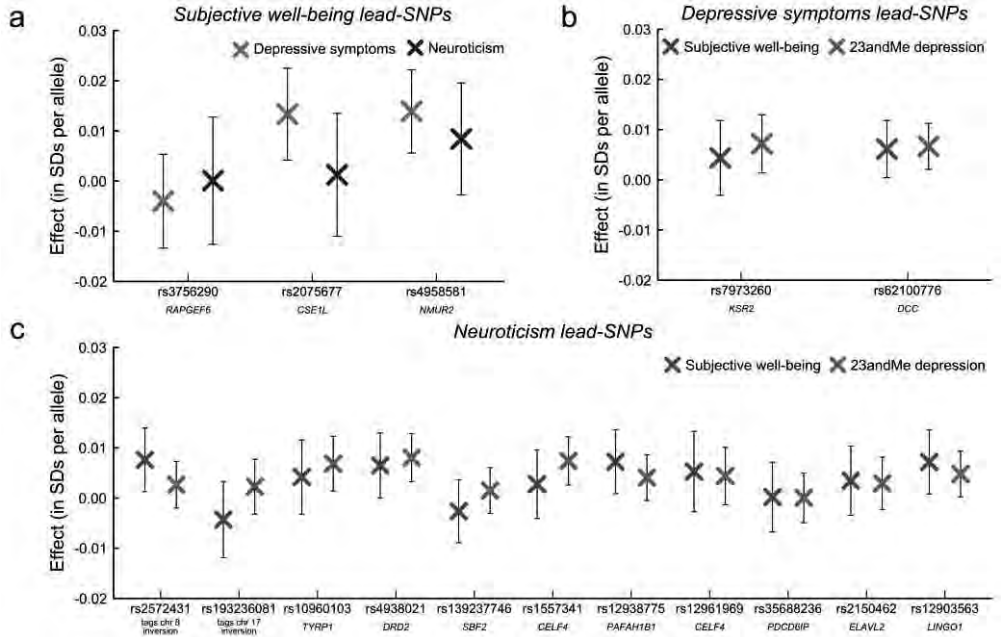
We assessed the credibility of our findings using a standard Bayesian framework (Meuwissen, Hayes, & Goddard, 2001; Vilhjálmsson et al., 2015) in which a positive fraction of SNPs have null effects and a positive fraction have non-null effects. For each phenotype, the non-null effect sizes are assumed to be drawn from a normal distribution whose variance is estimated from the GWAS summary statistics. As a first analysis, for each lead SNP's association with its phenotype, we calculated the posterior probability of null association after having observed the GWAS results. We found that, for any assumption about the fraction of non-null SNPs in the range 1% to 99%, the probability of true association always exceeds 95% for all 16 loci (and always exceeds 98% for 14 of them).

To further probe the credibility of the findings, we performed "quasi-replication" exercises in which we tested the subjective well-being lead-SNPs for association with depressive symptoms and neuroticism. We similarly tested the depressive symptoms lead-SNPs and the neuroticism lead-SNPs for association with subjective well-being. Below, we refer to the phenotype for which the lead SNP was identified as the first-stage phenotype and the phenotype used for the quasi-replication as the second-stage phenotype. To avoid sample overlap, for each quasi-replication analysis we omitted any cohorts that contributed to the GWAS of the first-stage phenotype.

Results of the quasi-replication of the three subjective well-being lead-SNPs are shown in Figure 3.3a. For ease of interpretation, the reference allele for each association in the figure is chosen such that the predicted sign of the second-stage estimate is positive. We find that two out of the three subjective well-being lead-SNPs are significantly associated with depressive symptoms ($p = 0.004$ and $p = 0.001$) in the predicted direction. For neuroticism, where the second-stage sample size ($N = 68{,}201$) is about half as large, the subjective well-being-increasing allele has the predicted sign for all three SNPs, but none reach significance.

Figure 3.3b-c show the results for the depressive symptoms and neuroticism lead-SNPs, respectively. In each panel, the blue crosses depict results from the quasi-replications where subjective well-being is the second-stage phenotype. We find that the two depressive symptoms lead-SNPs have the predicted sign for subjective well-being, and one is nominally significant ($p = 0.04$). Finally, of the eleven neuroticism lead-SNPs, nine have the predicted

Figure 3.3. Quasi-replication and lookup of lead SNPs.



Note: In quasi-replication analyses, we examined whether (**a**) lead SNPs identified in the subjective well-being meta-analyses are associated with depressive symptoms or neuroticism, (**b**) lead SNPs identified in the analyses of depressive symptoms are associated with subjective well-being, and (**c**) lead SNPs identified in the analyses of neuroticism are associated with subjective well-being. The quasi-replication sample is always restricted to non-overlapping cohorts. In a separate lookup exercise, we examined whether lead SNPs for depressive symptoms and neuroticism are associated with depression in an independent sample of 23andMe customers (N = 368,890). The results from this lookup are depicted as green crosses in (**b**) and (**c**). Bars represent 95% CIs (not adjusted for multiple testing). For interpretational ease, we choose the reference allele so that positive coefficients imply that the estimated effect is in the predicted direction. Listed below each lead SNP is the nearest gene.

sign for subjective well-being. Four of the eleven are nominally significantly associated with subjective well-being, all with the predicted sign. One of the four is the SNP tagging the inversion on chromosome 8 (Tian et al., 2008). That SNP's association with neuroticism (and likely with subjective well-being) is driven by its correlation with the inversion (Figure 3.4).

To evaluate what these quasi-replication results imply about the credibility of the 16 GWAS associations, we compared the observed quasi-replication record to the quasi-replication record expected given our statistical power. We calculated statistical power using our Bayesian framework, under the hypothesis that each lead SNP has a non-null effect on both the first- and second-stage phenotypes. Our calculations take into account both the imperfect genetic correlation between the first- and second-stage phenotypes and inflation of the first-stage estimates due to the well-known problem of winner's curse. Of the 19 quasi-replication tests, our calculations imply that 16.7 would be expected to yield the anticipated sign and 6.9 would be significant at the 5% level. The observed numbers are 16 and 7. Our quasi-replication results are thus consistent with the hypothesis that none of the 16 genome-wide significant associations are chance findings, and in fact strengthen the credibility of our GWAS results (Okbay, Baselmans, et al., 2016, Supplementary Note 8, Supplementary Table 14).

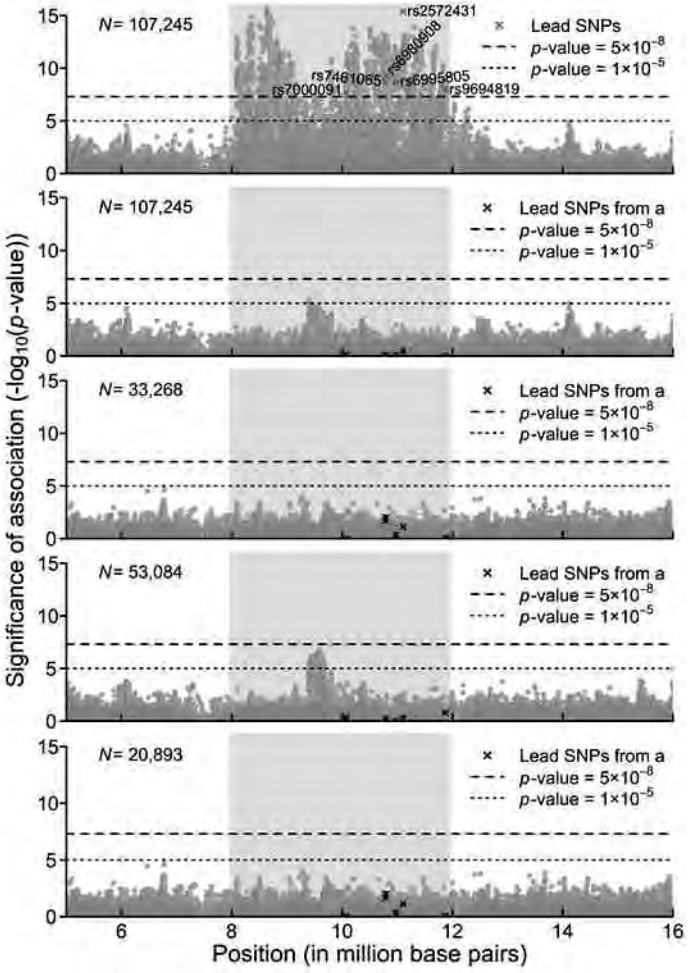
3.2.5 Lookup of depressive symptoms and neuroticism lead-SNPs

Investigators of an ongoing large-scale GWAS of major depressive disorder (N = 368,890; Hyde, n.d.) in the 23andMe cohort shared association results for the loci identified in our depressive symptoms and neuroticism analyses (Section 3.4.7; Okbay, Baselmans, et al., 2016, Supplementary Table 15). Because the depression sample overlaps with our subjective well-being sample, we did not request a lookup of the subjective well-being-associated SNPs.

In Figure 3.3b-c, the results are depicted as green crosses. For interpretational ease, we chose the reference allele so that positive coefficients imply that the estimated effect is in the predicted direction. All 13 associations have the predicted sign. Of the 11 neuroticism polymorphisms, four are significantly associated with depression at the 5% level. Both of the depressive symptoms lead-SNPs replicate (p = 0.004 and p = 0.015), with effect sizes (0.007 and -0.007 SDs per allele), close to those predicted by our Bayesian framework (0.008 and -0.006; Okbay, Baselmans, et al., 2016, Supplementary Tables 14-15).

Panel A of Table 3.1 summarizes the results for the 16 lead SNPs identified across our separate GWA analyses of the three phenotypes. The right-most column summarizes the statistical significance of the quasi-replication and depression lookup analyses of each SNP.

Figure 3.4. Local Manhattan plots of the association between SNPs on chromosome 8 and neuroticism in UKB.



Note: Panel **(a)** without controlling for the inversion-tagging principal component (PC), and **(b)** conditional on the PC. Panels **(c)-(e)** display results without controls for the PC for individuals with **(c)** inversion genotype 0, **(d)** inversion genotype 1, and **(e)** inversion genotype 2. The gray background area indicates the inversion region. Note that the sample sizes differ in panels **(c)-(e)**.

3.2.6 Proxy-phenotype analyses

To identify additional SNPs associated with depressive symptoms, we conducted a two-stage "proxy phenotype" analysis (Section 3.4.6). In the first stage, we ran a new GWAS of sub-

jective well-being to identify a set of candidate SNPs. Specifically, from each locus exhibiting suggestive evidence of association ($p < 10^{-4}$) with subjective well-being, we retained the SNP with the lowest p-value as a candidate. In the second stage, we tested these candidates for association with depressive symptoms at the 5% significance threshold, Bonferroni-adjusted for the number of candidates. We used an analogous two-stage procedure to identify additional SNPs associated with neuroticism. The first-stage subjective well-being sample differs across the two proxy-phenotype analyses (and from the primary subjective well-being GWAS sample) because we assigned cohorts across the first and second stages so as to maximize statistical power for the overall procedure.

For depressive symptoms, there are 163 candidate SNPs. 115 of them (71%) have the predicted direction of effect on depressive symptoms, 20 are significantly associated at the 5% significance level (19 in the predicted direction), and two remain significant after Bonferroni adjustment. For neuroticism, there are 170 candidate SNPs. 129 of them (76%) have the predicted direction of effect, all 28 SNPs significant at the 5% level have the predicted sign, and four of these remain significant after Bonferroni adjustment (Figure 3.5**;** Okbay, Baselmans, et al., 2016, Supplementary Tables 16 and 17). Two of the four are the SNPs identified in the proxy-phenotype analysis for depressive symptoms.

Table 3.1 lists the four SNPs in total identified by the proxy-phenotype analyses.

3.2.7 Biological analyses

To shed some light on possible biological mechanisms underlying our findings, we conducted several analyses.

We began by using bivariate LD Score regression (Bulik-Sullivan, Loh, et al., 2015) to quantify the amount of genetic overlap between each of our three phenotypes and ten neuropsychiatric and physical health phenotypes. Figure 3.2b and c display the estimates for subjective well-being and the *negative* of the estimates for depressive symptoms and neuroticism (since subjective well-being is negatively genetically correlated with depressive symptoms and neuroticism). Subjective well-being, depressive symptoms, and neuroticism have strikingly similar patterns of pairwise genetic correlation with the other phenotypes.

Table 3.1. Summary of polymorphisms identified across analyses

Panel A. Genome-Wide Significant Associations

Subjective Well-Being (SWB, $N = 298,420$)

SNPID	CHR	BP	EA	EAF	Beta (s.e.m.)	R^2 (%)	*P*-value	*N*	Quasi-Repl
rs3756290	5	130,951,750	A	0.24	-0.0177 (0.0031)	1.148	9.55×10^{-9}	286,851	
rs2075677	20	47,701,024	A	0.76	0.0175 (0.0031)	1.111	1.49×10^{-8}	288,454	DS**
rs4958581	5	152,187,729	T	0.66	0.0153 (0.0027)	1.062	2.29×10^{-8}	294,043	DS***

Neuroticism ($N = 170,911$)

SNPID	CHR	BP	EA	EAF	Beta (s.e.m.)	R^2 (%)	*P*-value	*N*	Quasi-Repl
rs2572431[#]	8	11,105,077	T	0.59	0.0283 (0.0035)	0.039	4.20×10^{-16}	170,908	SWB*
rs193236081[##]	17	44,142,332	T	0.77	-0.0284 (0.0043)	0.028	6.26×10^{-11}	151,297	
rs10960103	9	11,699,270	C	0.77	0.0264 (0.0042)	0.024	2.14×10^{-10}	165,380	DS*
rs4938021	11	113,364,803	T	0.66	0.0233 (0.0037)	0.024	4.03×10^{-10}	159,900	DS***, SWB*
rs139237746	11	10,253,183	T	0.51	-0.0204 (0.0034)	0.021	2.55×10^{-9}	170,908	
rs1557341	18	35,127,427	A	0.34	0.0213 (0.0037)	0.021	5.58×10^{-9}	165,579	DS**
rs12938775	17	2,574,821	A	0.53	-0.0202 (0.0035)	0.020	8.54×10^{-9}	163,283	SWB*
rs12961969	18	35,364,098	A	0.20	0.0250 (0.0045)	0.020	2.16×10^{-8}	156,758	
rs35688236	3	34,582,993	A	0.69	0.0213 (0.0038)	0.019	2.35×10^{-8}	161,636	
rs2150462	9	23,316,330	C	0.74	-0.0217 (0.0039)	0.018	2.66×10^{-8}	170,907	
rs12903563	15	78,033,735	T	0.50	0.0198 (0.0036)	0.020	2.86×10^{-8}	157,562	DS*, SWB*

Depressive Symptoms (DS, $N = 180,866$)

SNPID	CHR	BP	EA	EAF	Beta (s.e.m.)	R^2 (%)	*P*-value	*N*	Quasi-Repl
rs7973260	12	118,375,486	A	0.19	0.0306 (0.0051)	0.029	1.78×10^{-9}	124,498	DS*
rs62100776	18	50,754,633	A	0.56	-0.0252 (0.0044)	0.031	8.45×10^{-9}	105,739	DS**, SWB*

Panel B. SNPs Identified via Proxy-Phenotype Analyses of SWB Loci with *P*-value<10^{-4}

Depressive Symptoms in Non-Overlapping Cohorts

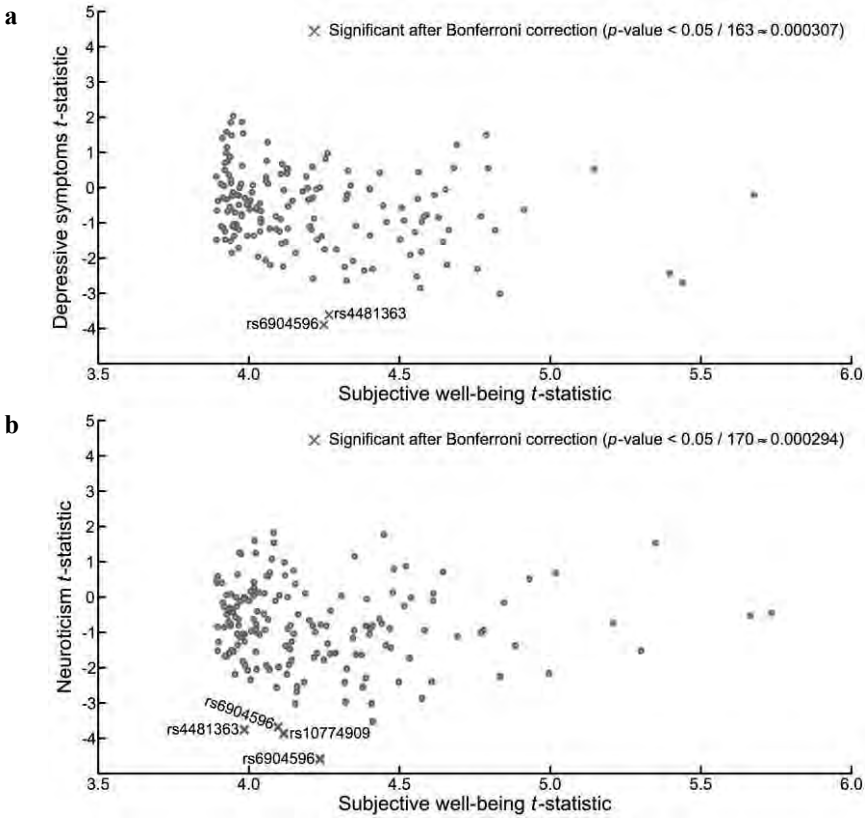
SNPID	CHR	BP	EA	EAF	$Beta_{DS}$ (s.e.m.)	R^2 (%)	P_{DS}	Bonferroni	N_{DS}
rs4346787†	6	27,491,299	A	0.113	-0.023 (0.0059)	0.011	9.79×10^{-5}	0.0160	142,265
rs4481363	5	164,483,794	A	0.524	0.014 (0.0038)	0.009	3.06×10^{-4}	0.0499	142,265

Neuroticism in Non-Overlapping Cohorts

SNPID	CHR	BP	EA	EAF	$Beta_{neuro}$ (s.e.m.)	R^2 (%)	P_{neuro}	Bonferroni	N_{neuro}
rs10838738	11	47,663,049	A	0.49	0.0178 (0.0039)	0.016	5.03×10^{-6}	0.0009	131,864
rs10774909	12	117,674,129	C	0.52	-0.0150 (0.0039)	0.011	1.20×10^{-4}	0.0203	131,235
rs6904596	6	27,491,299	A	0.09	-0.0264 (0.0072)	0.012	2.49×10^{-4}	0.0423	116,335
rs4481363	5	164,474,719	A	0.49	0.0151 (0.0040)	0.011	1.86×10^{-4}	0.0316	122,592

Note: EA: effect allele. EAF: effect allele frequency. All effect sizes are reported in units of SDs per allele. "Quasi-Repl.": phenotypes for which SNP was found to be nominally associated in quasi-replication analyses conducted in independent samples. *significant at the 5%-level, **significant at the 1%-level, ***significant at the 0.1%-level. #inversion-tagging polymorphism on chromosome 8. ##inversion-tagging polymorphism on chromosome 17. †proxy for rs6904596 ($R^2 = 0.98$).

Figure 3.5. Proxy-phenotype analyses: test of SNPs associated with subjective well-being at $p < 1\times10^{-4}$ for association with depressive symptons and neuroticism.



Note: In both analyses, the second-stage sample is restricted to non-overlapping cohorts. For interpretational ease, for each SNP we choose the reference allele for subjective well-being to be the effect-increasing allele. See Supplementary Tables 16 and 17 of Okbay, Baselmans et. al. (2016) for detailed results and Section 3.4.6 for additional details.

Figure 3.2b shows the results for the five neuropsychiatric phenotypes we examined: Alzheimer's disease, anxiety disorders, autism spectrum disorder, bipolar disorder, and schizophrenia. For four of these phenotypes, genetic correlations with depression (but not neuroticism or subjective well-being) were reported in Bulik-Sullivan, Loh, et al. (2015). For schizophrenia and bipolar disorder, our estimated correlations with depressive symptoms, 0.33 and 0.26, are substantially lower than Bulik-Sullivan et al.'s point estimates but contained within their 95% confidence intervals. By far the largest genetic correlations we estimate are with anxiety disorders: −0.73 with subjective well-being, 0.88 with depressive symptoms,

and 0.86 with neuroticism. Genetic correlations estimated from GWAS data have not been previously reported for anxiety disorders.

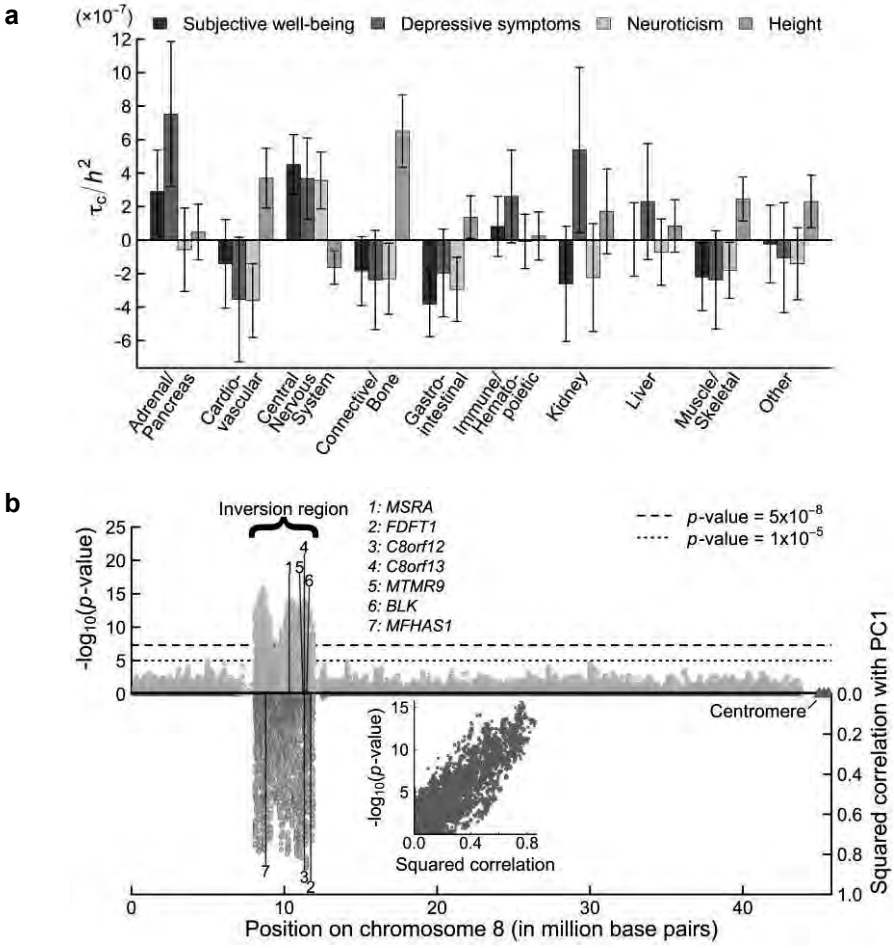
Figure 3.2c shows the results for five physical health phenotypes that are known or believed to be risk factors for various adverse health outcomes: body mass index (BMI), ever-smoker status, coronary artery disease, fasting glucose, and triglycerides. The estimated genetic correlations are all small in magnitude, consistent with earlier work, although the greater precision of our estimates allows us to reject null effects in most cases. The signs are generally consistent with those of the phenotypic correlations reported in earlier work between our phenotypes and outcomes such as obesity (Roberts, Kaplan, Shema, & Strawbridge, 2000), smoking (Glassman et al., 1990; Shahab & West, 2012), and cardiovascular health (Rugulies, 2002).

Next, to investigate whether our GWAS results are enriched in particular functional categories, we applied stratified LD Score regression (Finucane et al., 2015) to our meta-analysis results. In our first analysis, we report estimates for all 53 functional categories included in the "baseline model"; the results for subjective well-being, depressive symptoms, and neuroticism are broadly similar (Okbay, Baselmans, et al., 2016, Supplementary Tables 18-20) and are in line with what has been found for other phenotypes (Finucane et al., 2015). In our second analysis, the categories are groupings of SNPs likely to regulate gene expression in cells of a specific tissue. The estimates for subjective well-being, depressive symptoms, and neuroticism are shown in Figure 3.6a, alongside height, which is again included as a benchmark (Wood et al., 2014; Okbay, Baselmans, et al., 2016, Supplementary Table 21).

We found significant enrichment of CENTRAL NERVOUS SYSTEM for all three phenotypes and, perhaps more surprisingly, enrichment of ADRENAL/PANCREAS for subjective well-being and depressive symptoms. The cause of the ADRENAL/PANCREAS enrichment is unclear, but we note that the adrenal glands produce several hormones, including cortisol, epinephrine, and norepinephrine, known to play important roles in the bodily regulation of mood and stress. It has been robustly found that blood serum levels of cortisol in patients afflicted by depression are elevated relative to controls (Stetler & Miller, 2011).

While the above analyses utilize the genome-wide data, we also conducted three analyses restricted to the 16 GWAS and four proxy-phenotype SNPs in Table 3.1. In brief, we ascertained whether each SNP (or a variant in strong linkage disequilibrium (LD) with it) falls into any of the following three classes: (i) resides in a locus for which genome-wide significant associations with other phenotypes have been reported, (ii) is nonsynonymous, and (iii)

Figure 3.6. Results from selected biological analyses.



Note: (**a**) Estimates of the expected increase in the phenotypic variance accounted for by a SNP due to the SNP's being in a given category (τ_c), divided by the LD Score heritability of the phenotype (h^2). Each estimate of τ_c comes from a separate stratified LD Score regression, controlling for the 52 functional annotation categories in the "baseline model." The bars represent 95% CIs (not adjusted for multiple testing). To benchmark the estimates, we compare them to those obtained from a recent study of height (Wood et al., 2014). (**b**) Inversion polymorphism on chromosome 8 and the 7 genes for which the inversion is a significant *cis*-eQTL at FDR < 0.05. The upper half of the figure shows the Manhattan plot for neuroticism for the inversion and surrounding regions. The bottom half shows the squared correlation between the SNPs and the principal component that captures the inversion. The inlay plots the relationship, for each SNP in the inversion region, between the SNP's significance and its squared correlation with the principal component that captures the inversion.

is an eQTL in blood or in one of 14 other tissues (although the non-blood analyses are based on smaller samples). Here we highlight a few particularly interesting results (Okbay, Baselmans, et al., 2016, Supplementary Tables 22-24).

We found that five of the 20 SNPs are in loci in which genome-wide significant associations have previously been reported. Two of these five are schizophrenia loci. Interestingly, one of them harbors the gene *DRD2*, which encodes the D_2 subtype of the dopamine receptor, a target for antipsychotic drugs (Seeman, 2010) that is also known to play a key role in neural reward pathways (Vallone, Picetti, & Borrelli, 2000). Motivated by these findings, as well as by the modest genetic correlations with schizophrenia reported in Figure 3.2b, we examined whether the SNPs identified in a recent study of schizophrenia (Ripke et al., 2014) are enriched for association with neuroticism in our non-overlapping UKB sample (N = 107,245). We conducted several tests and found strong evidence of such enrichment (Okbay, Baselmans, et al., 2016, Supplementary Note 7E). For example, we found that the p-values of the schizophrenia SNPs tend to be much lower than the p-values of a randomly selected set of SNPs matched on allele frequency ($p = 6.50\times10^{-71}$).

Perhaps the most notable pattern that emerges from our biological analyses is that the inversions on chromosomes 8 and 17 are implicated consistently across all analyses. The inversion-tagging SNP on chromosome 8 is in LD with SNPs that have previously been found to be associated with BMI (Shungin et al., 2015) and triglycerides (Kathiresan et al., 2009; Okbay, Baselmans, et al., 2016, Supplementary Table 22). We also conducted eQTL analyses in blood for the inversion itself and found that it is a significant *cis*-eQTL for 7 genes (Okbay, Baselmans, et al., 2016, Supplementary Table 24). As shown in Figure 3.6b, all 7 genes are positioned in close proximity to the inversion breakpoints, suggesting that the molecular mechanism underlying the inversion's effect on neuroticism could involve the relocation of regulatory sequences. Two of the genes (*MSRA*, *MTMR9*) are known to be highly expressed in tissues and cell types that belong to the nervous system, and two (*BLK*, *MFHAS1*) in the immune system. In the tissue-specific analyses, we found that the SNP tagging the inversion is a significant eQTL for two genes, *AF131215.9* (in tibial nerve and thyroid tissue analyses) and *NEIL2* (tibial nerve tissue), both of which are also located near the inversion breakpoint.

The SNP tagging the chromosome 17 inversion is a significant *cis*-eQTL for five genes in blood and is an eQTL in all 14 other tissues (Okbay, Baselmans, et al., 2016, Supplementary Table 24). It alone accounts for 151 out of the 169 significant associations identified in the 14 tissue-specific analyses. Additionally, the SNP is in near-perfect LD ($R^2 > 0.97$) with 11 missense variants (Okbay, Baselmans, et al., 2016, Supplementary Table 23) in three different genes, one of which is *MAPT*. *MAPT*, which is also implicated in both the blood and the

other tissue-specific analyses, encodes a protein important in the stabilization of microtubules in neurons. Associations have been previously reported between SNPs in *MAPT* (all of which are in strong LD with our inversion-tagging SNP) and neurodegenerative disorders, including Parkinson's disease (Spencer et al., 2011) and progressive supranuclear palsy (Höglinger et al., 2011), a rare disease whose symptoms include depression and apathy.

3.3 Discussion

The discovery of genetic loci associated with subjective well-being, depression, and neuroticism has proven elusive. Our study identified several credible associations for two main reasons. First, our analyses had greater statistical power than prior studies because ours were conducted in larger samples. Our GWAS findings—three loci associated with subjective well-being, two with depressive symptoms, and eleven with neuroticism—support the view that GWAS can successfully identify genetic associations with highly polygenic phenotypes in sufficiently large samples (Hyman, 2014; Sullivan, 2012). A striking finding is that two of our identified associations are with inversion polymorphisms.

Second, our proxy-phenotype analyses further boosted power by exploiting the strong genetic overlap between our three phenotypes. These analyses identified two additional loci associated with neuroticism and two with both depressive symptoms and neuroticism. Through our quasi-replication tests, we also demonstrated how studying genetically overlapping phenotypes in concert can provide evidence on the credibility of GWAS findings. Our direct replication of the two genome-wide significant associations with depressive symptoms in an independent depression sample provides further confirmation of those findings (Figure 3.3b**;** Okbay, Baselmans, et al., 2016, Supplementary Table 15).

We were able to assemble much larger samples than prior work in part because we combined data across heterogeneous phenotype measures. Our results reinforce the conclusions from our theoretical analysis that doing so increased our statistical power, but our strategy also has drawbacks. One is that mixing different measures may make any discovered associations more difficult to interpret. Research studying higher quality measures of the various facets of subjective well-being, depressive symptoms, and neuroticism is a critical next step. Our results can help facilitate such work because if the variants we identify are used as candidates, studies conducted in the smaller samples in which more fine-grained phenotype measures are available can be well powered.

Another limitation of mixing different measures is that doing so may reduce the heritability of the resulting phenotype, if the measures are influenced by different genetic factors. Indeed, our estimates of SNP-based heritability (Bulik-Sullivan, Loh, et al., 2015) for our three phenotypes are quite low: 0.040 (*SE* = 0.002) for subjective well-being, 0.047 (*SE* = 0.004) for depressive symptoms, and 0.091 (*SE* = 0.007) for neuroticism. We correspondingly find that polygenic scores constructed from all measured SNPs explain a low fraction of variance in independent samples: ~0.9% for subjective well-being, ~0.5% for depressive symptoms, and ~0.7% for neuroticism (Section 3.4.5). The low heritabilities imply that even when polygenic scores can be estimated using much larger samples than ours, they are unlikely to attain enough predictive power to be clinically useful.

According to our Bayesian calculations, the true explanatory power (corrected for winner's curse) of the SNP with the largest posterior R^2 is 0.003% for subjective well-being, 0.002% for depressive symptoms, and 0.011% for neuroticism (Okbay, Baselmans, et al., 2016, Supplementary Table 14). These effect sizes imply that in order to account for even a moderate share of the heritability, hundreds or (more likely) thousands of variants will be required. They also imply that our study's power to detect variants of these effect sizes was not high—for example, our statistical power to detect the lead SNP with largest posterior R^2 was only ~13%—which in turn means it is likely that there exist many variants with effect sizes comparable to our identified SNPs that evaded detection. These estimates suggest that many more loci will be found in studies with sample sizes realistically attainable in the near future. Consistent with this projection, when we meta-analyze the 54 SNPs reaching $p < 10^{-5}$ in our analyses of depressive symptoms together with the 23andMe replication sample for depression, the number of genome-wide significant associations rises from 2 to 5 (Appendix A - Table A10).

3.4 Supplementary Methods

3.4.1 Primary GWAS of subjective well-being

There is much interest in the genetic basis of SWB, with twin studies suggesting that genetic factors may account for as much as 40% of the variance in SWB across individuals (Bartels, 2015). A recent study of the "common narrow heritability" (also called "SNP heritability") of SWB estimated that 5-10% of the variance in SWB can be explained by the cumulative additive effects of genetic variants that are common in the population (Rietveld, Cesarini, et al., 2013). Because such common variants are assessed by contemporary genome-wide approaches to genotyping single nucleotide polymorphisms (SNPs), that study concluded that

a GWAS on SWB in a sufficiently large sample of individuals may yield reliably associated SNPs.

While some unsuccessful attempts have been conducted to identify genomic regions of interest for SWB (Bartels & Baselmans, 2015)—including some candidate gene studies (H. Chen et al., 2012; De Neve, Fowler, Frey, & Christakis, 2012) and a genome-wide linkage study (Bartels et al., 2010)—as far as we know, there is no large-scale meta-GWAS effort like the one reported here. From the outset we reasoned that the results of this research would have scientific merit regardless of the outcome. If the meta-analysis succeeded in identifying genetic variants, then such information would be an important first step toward understanding the pathways between genes, SWB, and other phenotypes, as well as the complex interplay between these pathways and the environment. If, on the other hand, no robust associations were uncovered, it would allow us to put a much tighter upper bound on the expected effect sizes for common variants associated with complex traits such as SWB.

A. OVERVIEW OF SUBJECTIVE WELL-BEING ANALYSES

The genome-wide association study (GWAS) of subjective well-being (SWB) is based on summary statistics uploaded by cohort-level analysts to a central server. The summary statistics were subsequently quality controlled and meta-analyzed by a central team of analysts. The lead PI of each cohort affirmed that the results contributed to the study were based on analyses approved by the local Research Ethics Committee and/or Institutional Review Board by signing a collaboration agreement which contained the clause "Each Representative Signing this document on behalf of a particular cohort is responsible for ensuring that the Institutional Review Board (IRB) or ethical committee has approved the analysis of well-being in that sample.". All participants provided written informed consent.

SWB is usually defined broadly to include both positive and negative subjective evaluations. Across the many facets of SWB, a distinction is often made between "positive affect" (PA) and "life satisfaction" (LS; Kahneman & Deaton, 2010; Kahneman & Riis, 2005). PA refers to the frequency and intensity of positive emotions and feeling happy. Typical survey questions used to gauge PA include "During the past week, I was happy?" and "How would you rate your emotional wellbeing at present?" LS refers to a longer-term evaluation of one's life. A typical survey question would be "How satisfied are you with your life as a whole?". The two facets are known to be positively correlated with each other and load on a common genetic factor (Bartels & Boomsma, 2009). For this reason, and to maximize sample size, we decided a priori to make our primary analysis one in which we pool the two measures in

a combined analysis, and to report but treat as secondary, analyses of PA and LS considered separately.

Although LS and PA are phenotypically distinct, multivariate genetic analyses have found that the variance in LS and the variance in PA are explained by the same underlying common genetic factor (Bartels & Boomsma, 2009). For this reason, and to maximize sample size, we decided prior to conducting the study to make our primary analysis one in which we pool the two measures in a combined analysis, and to report, but treat as secondary, analyses of PA and LS considered separately. This approach is in line with modeling SWB as a "hierarchical construct" (Busseri, 2015). By considering the pooled wellbeing measure as the higher order phenotype, we build on the relatedness of the SWB components (LS and PA). At the same time, by also reporting the LS and PA analyses separately, we allow for the possibility of differences in genetic factors underlying these components of SWB, and we can disentangle whether associations we identify with the combined measure are driven by one of these two facets of SWB.

At the beginning of the study, we circulated an analysis plan describing the cohort-level analyses needed (including the exact specification and restriction to European-descent individuals). The plan asked cohort analysts to upload results by April 16, 2012, but this deadline was not strictly enforced. Final results files were uploaded in October. We subsequently performed a meta-analysis of the results ($N \approx 100{,}000$) and failed to find any genome-wide significant hits. The absence of significant results was not surprising in light of a paper published by PIs in 2013 (Rietveld, Cesarini, et al., 2013) that found that the heritability due to common variants of single-question measures of subjective well-being was around 5%. Given this new information, it was decided at the SSGAC meeting on 15 June 2013 to reopen the discovery phase, invite other cohorts to contribute, and integrate newly available data from currently contributing cohorts. Despite the well-known biases in statistical inference from "optional stopping," we felt it was justified to relax the data-freeze deadline due to the near doubling of the potential sample size.

We accordingly formulated an updated analysis plan, posted on https://osf.io/cq2b5/, which stipulated a data freeze of either 31 December 2014, or the day on which the combined sample size of the uploaded results for PA exceeded $N = 150{,}000$, whichever of the two events occurred earlier. By December 2014, we had attained sample sizes of approximately $N = 117{,}000$ for PA, $N = 85{,}000$ for LS, and $N = 152{,}500$ for the combined well-being (WB) phenotype. At that time, the cohort 23andMe, which had contributed 30,000 observations to the original analysis, indicated that they would be willing to upload new results based on a much larger sample ($N = 90{,}000$) that had since become available. A much larger sample

size would also be possible due to the imminent release of the first batch of data from the UK Biobank (Sudlow et al., 2015) with $N \approx 120{,}000$ genotyped European-ancestry individuals, roughly 60,000 of whom have answered a high-quality PA question. In December 2014, we therefore decided to further postpone the data freeze until data from 23andMe and UKB became available. This decision was made prior to running any meta-analyses of results uploaded by December 2014.

B. PARTICIPATING COHORTS

Table A2 in Appendix A provides study-specific details on all results files from the 59 participating cohorts that passed the quality-control analyses described below. Any cohort with acceptable survey measures of either PA or LS was eligible to participate in the study. Some cohorts had measures of both PA and LS, sometimes measured on more than one occasion. Such cohorts were encouraged to upload three results files: one for PA, one for LS, and one for a combined WB measure constructed by combining the LS and PA responses. Cohorts with multiple measures were encouraged to average responses to reduce the amount of transitory variation in responses. The exact construction of the combined measure was typically determined in consultation with the cohort analyst. Of the 59 participating cohorts, 7 uploaded LS results only (hereafter, "LS cohorts"), 28 uploaded PA results only ("PA cohorts"), 12 uploaded results for LS and PA but not WB ("LSPA cohorts"), and 12 uploaded LS, PA, and WB results files ("LSPAWB cohorts"), leaving us with a total of $7 + 28 + 12\times2 + 12\times3 = 95$ results files.

C. STUDY-SPECIFIC MEASURES

Supplementary Table 3 in Okbay, Baselmans, et al. (Okbay, Baselmans, et al., 2016) summarizes the study-specific LS and PA phenotypes. We purposely eschewed limiting the study to a specific questionnaire or survey scale, reasoning that the sample-size gains from an inclusive strategy that permitted some variation in question phrasing would outweigh any loss of power arising due to phenotypic heterogeneity. However, we did not allow questions that asked about happiness or satisfaction in specific domains (e.g., satisfaction with one's health status or financial situation).

Cohorts overwhelmingly used survey questions derived from or adapted from established survey batteries. Of the 31 cohorts with measures of LS, 19 used questions taken or adapted from popular and psychometrically validated life satisfaction or depression scales such as the Satisfaction With Life Scale (Diener, Emmons, Larsen, & Griffin, 1985) or the Geriatric Depression Scale (Yesavage et al., 1983) (which has a sub-item on life satisfaction suitable

for our purposes). Another 7 cohorts used questions adapted from one of the main LS questions of the World Values Survey: "All things considered, how satisfied are you with your life as a whole these days?" Most remaining cohorts asked questions of the general character "How satisfied are you with your life?" Overall, the questions were thus phrased very similarly. The only possible exception of the question used by the 1958 British Birth Cohort ("On balance I look back at my life with a sense of happiness"), which, even though the question contains the word "happiness", we chose to classify as an LS measure because we interpreted the question as an evaluative measure of satisfaction. For LS, there is hence overall little cause for concern that variation in the phrasing of the question introduced substantial phenotypic heterogeneity. The number of response categories varies across questions. However, cohorts with four or fewer response categories account for only 15.4% of the LS sample.

The PA phenotypes analyzed exhibit greater variation and include items such as "During the past week I was happy" and "Do you feel happy most of the time?" Some cohorts also used scores from psychological scales, such as the Subjective Happiness Scale (Lyubomirsky & Lepper, 1999) or the well-being trait scale of the Multidimensional Personality Questionnaire (Tellegen & Waller, 2008). Remaining cohorts used sub-items from a diverse set of psychological questionnaires which includes the Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983), the Scale of Positive and Negative Experience (Diener et al., 2010), and the Positive Affect and Negative Affect Scale (D. Watson, Clark, & Tellegen, 1988). By far the most common strategy was to use items about positive affect from the Center for Epidemiologic Studies Depression ("CES-D") scale (Radloff, 1977), a standard depression battery, which contains a four-item Positive Affect subscale used by some cohorts, whereas others used a single item on this subscale that asks specifically about happiness ("Last week, [how often] were you happy?").

D. GENOTYPING AND IMPUTATION

Genotyping was performed using a range of common, commercially available genotyping arrays. Supplementary Table 4 in Okbay, Baselmans, et al. (2016) provides study-specific details on genotyping platform, pre-imputation quality-control filters applied to the genotype data, subject-level exclusion criteria, imputation software used, and the reference sample used for imputation. Because our study was launched before 1000G-imputation became standard practice, the analysis protocol circulated to analysts recommended uploading results imputed using the HapMap 2 CEU (r22.b36) reference sample (The International HapMap Consortium, 2007). Our analysis plan advised cohorts to exclude from their estimation sample subjects with low overall call rates (< 95%), excess autosomal heterozygosity, or sex mismatch (excessive X-chromosome homozygosity in males). Additionally,

the plan advised family-based cohorts to only include one relative from each pedigree or to report standard errors adjusted for the sample relatedness. We encouraged, but did not require, later enrollees into the study to supply us with 1000G-imputed data. 40 out of 59 cohorts supplied HapMap2-imputed data. Though the cohorts with 1000G data are a minority, they are larger on average, accounting for more than 70% of our combined sample size.

E. ASSOCIATION ANALYSES

Cohorts were asked to estimate the following regression equation for each SNP:

$$Y = \beta_0 + \beta_1 SNP + \mathbf{PC}\,\boldsymbol{\gamma} + \mathbf{B}\,\boldsymbol{\alpha} + \mathbf{X}\,\boldsymbol{\theta} + \epsilon, \tag{3.1}$$

where Y is an unstandardized outcome variable, *SNP* is the allele dose of the SNP; **PC** is a vector of the first four principal components of the variance-covariance matrix of the genotypic data, estimated after the removal of genetic outliers; and **B** is a vector of standardized controls, including sex, age and age squared. Cohorts were also asked to include any study-specific covariates such as study site or batch effects that they considered appropriate. Some cohorts with binary dependent variables uploaded results from a logistic regression model analogous to Equation (3.1).

F. QUALITY-CONTROL PROCEDURES

Generating a Reference File Mapping rsIDs to ChrPosIDs

SNPs imputed using HapMap and 1000G reference panels are ordinarily assigned chromosomal position identifiers ("ChrPosID", a concatenation of a SNP's chromosome number, a colon, and the SNP's base pair position) using different versions of the NCBI build. In several of our analyses, it is desirable to use a harmonized one-to-one mapping from rsID to chromosomal coordinates. We therefore restrict all our SWB analyses to a set of autosomal SNPs with rs identifiers ("rsIDs") that (i) appear in both the HapMap and 1000G reference panels, and (ii) for which a ChrPosID can be generated in build 37 coordinates. To generate this list, we used files that have been made publicly available by the developers of the EasyQC software (Winkler et al., 2014). All combined data from a number of public sources. For details on construction of several of the key files, see pp. 2-3 in the Supplementary Material of Winkler et al. (2014).

In the EasyQC reference files, there are 2,532,578 HapMap SNPs with non-missing information about (i) rsID, (ii) build 36 coordinates, and (iii) European allele frequency in the

HapMap Phase 2 CEU reference sample. From this original set of SNPs, we dropped 66 SNPs because their ChrPosID (in build 36 coordinates) was not unique, 31,657 SNPs whose rsIDs could not be located in the 1000G reference file, and an additional 12,921 SNPs for which the allele frequency is not available in the Europeans-only 1000G reference file (the version which excludes X-chromosome markers and monomorphic SNPs). In the final step, we dropped 719 SNPs due to possible allele misalignment and 2,471 SNPs whose allele frequencies differed by more than 0.25 across the HapMap and 1000G European reference samples. This leaves us with 2,484,798 rsIDs with information about allele frequency in a reference sample of Europeans, and a one-to-one mapping to a ChrPosID expressed in build 37 coordinates. In what follows, we refer to this as our reference file.

Pre-QC Verification of Descriptives

For each cohort, we checked whether (i) they had supplied us with complete descriptive statistics, and (ii) the variable coding was in accordance with the analysis plan (e.g., higher values indicating greater WB, PA, or LS). If not, we contacted the cohort to obtain corrected data. Next, as we explain below, we verified to the extent allowed by our data that the reported information was consistent with the uploaded summary statistics. Many of the QC checks were intended to eliminate problems of reverse coding that, if undetected, can substantially reduce power to detect associations.

EasyQC

We used the software EasyQC (Winkler et al., 2014) to check each uploaded results file for quality-control problems. From each uploaded file, we filtered out SNPs in the following order.

1. We first dropped any SNPs in the uploaded results that could not be identified in the reference file, whose construction we described above. In a few cohorts who had imputed their data against the September or December 2013 releases of the 1000 Genomes Phase 1 haplotypes provided by the software IMPUTE2, we also dropped the 730+199 SNPs whose strands are known to have been incorrectly aligned in these releases[**].

2. We dropped a SNP if neither an effect nor other allele was supplied, or if either of them takes values other than “A”, “C”, “G”, or “T”. We also dropped a SNP if any of the following variables were missing: *p*-value, a coefficient estimate (beta) and

[**] The announcement is available on https://mathgen.stats.ox.ac.uk/impute/impute_v2.html#whats_new

its standard error, effect allele frequency, sample size (*N*), and imputation accuracy (for imputed SNPs). We also dropped SNPs if any of the variables reported for the SNP were outside the permissible range of the variable (for example, *p*-values greater than 1 or negative standard errors). As in Okbay et al. (Okbay, Beauchamp, et al., 2016), we dropped SNPs from cohorts that used likelihood-ratio tests for inference if the reported LR test statistic differed by more than 10% from the squared *t*-statistic constructed by dividing the estimated regression coefficient by the analytical standard error (i.e., the standard error obtained from the information matrix evaluated at the maximum likelihood estimates). This QC screen is based on the observation that both the LR test statistic and the squared *t*-statistic should have the same distribution under the null hypothesis, namely a chi-squared with one degree of freedom. Seven cohorts reported *p*-values from likelihood-ratio (LR) tests.

3. We dropped SNPs with minor allele frequencies below a threshold that varied by sample size. In samples with fewer than 1000 observations, we applied a threshold of 10%. In studies with a sample size between 1000 and 2000, we dropped SNPs with MAF < 5%. In all other samples, we applied a threshold of 3%. Table A3 summarizes these filters and others used in the steps below.

4. We filtered out SNPs with low imputation accuracy. The definition of the imputation accuracy metric varies by imputation software. If the cohort supplied us with the "Rsq" variable generated by MaCH (Y. Li, Willer, Ding, Scheet, & Abecasis, 2010), we dropped SNPs with Rsq < 0.4. If they uploaded the "INFO" variable generated by IMPUTE (Marchini, Howie, Myers, McVean, & Donnelly, 2007), we applied a threshold of 0.5. If PLINK's "info" variable was supplied, we applied a threshold of 0.8.

5. We dropped SNPs with low Hardy-Weinberg Equilibrium (HWE) *p*-value (see Table A3 for exact cutoffs used).

6. We dropped SNPs with call rate below 95%.

7. We dropped duplicated SNPs based on Build 37 base pair positions obtained by mapping the rsIDs in each results file to the ChrPosIDs in the reference file. We also dropped SNPs that could not be successfully aligned due to mismatch with reference alleles.

Having applied all the filters to the cohort-level summary statistics, we examined how many SNPs were dropped in each filtering step. Whenever an unusual number of markers was being dropped, we flagged the cohort as potentially having an error in the uploaded results file. The issue was discussed with the cohort-level analyst and resolved through a new QC iteration. Table A4 shows, for each cohort, the number of SNPs dropped in each filtering step and also shows the estimated genomic control factors. The table contains only the phenotype (LS, PA or WB) included in the main SWB meta-analysis for each cohort. For the full set of SNP filtering results, see Supplementary Table 6 of Okbay, Baselmans, et al. (2016). The genomic control factors in the full set of results are in the range 0.871 to 1.099, with a median value of 1.004.

Visual inspection of diagnostic plots

Having processed the data through these filters, we inspected several diagnostic plots.

i. *Allele Frequency Plots (AF Plots):* We looked for errors in allele frequencies and strand orientations by inspecting a plot of sample allele frequencies against the allele frequency in a European reference sample. Any deviations could indicate a number of problems (failure to exclude ethnic outliers, imputation errors, etc.).

ii. *P-Z Plots:* We checked that reported *p*-values are consistent with the reported coefficient estimates and their *SE*s.

iii. *Q-Q Plots:* We visually inspected the cohort-level Q-Q plots to look for evidence of unaccounted-for stratification.

Potential issues with the plots were always raised with the cohort-level analyst. All analyses are based on results files whose plots did not indicate any quality-control problems.

Additional diagnostics

We supplemented the EasyQC checks with the following diagnostic checks:

i. For cohorts that uploaded results for more than one phenotype, we calculated the correlation between the estimated regression coefficients for the two phenotypes. Since the phenotypes are positively correlated, we expect positively correlated coefficient estimates unless the sample overlap is minimal; negatively correlated betas would be a strong indication of phenotype miscoding or allele misalignment in

one of the results files. We found pairwise correlations in the range 0.2 to 0.9, with correlations in the lower end of this range only in those cohorts where the LS and PA questions were asked several years apart.

ii. We conducted a number of additional diagnostic analyses in a restricted set of SNPs comprising the union of a (i) set of 50,000 randomly sampled HapMap2 SNPs and (ii) SNPs showing suggestive evidence of association in the 23andMe data (henceforth the 50K SNP set). The list of 23andMe SNPs was constructed by selecting the lowest *p*-value SNPs from a set of approximately independent loci. These independent loci were determined using a frequently used (Ripke et al., 2014) iterative clumping procedure implemented in Plink (Chang et al., 2015), as follows. First, the SNP with the smallest *p*-value is identified in the 23andMe results. This SNP is the lead SNP of clump 1. Second, all SNPs whose association *p*-value is lower than 10^{-3} that are within 500 kb of the lead SNP and whose LD with the lead SNP exceeds $R^2 = 0.1$ are identified and assigned to clump 1. We calculate LD using the 1000G phase 1 reference sample composed of Utah Residents (CEPH) with Northern and Western European Ancestry (CEU), Toscani in Italia (TSI), and British in England and Scotland (GBR) (Abecasis et al., 2012). To generate the second clump, the SNP with lowest *p*-value among the SNPs that remain after removal of clump 1 is identified and the same steps are applied to identify the set of SNPs comprising clump 2. The process is repeated until no SNPs with *p*-values below 10^{-3} remain. This process left us with 823 approximately independent loci.

   a. For each cohort, we calculated the degree of sign concordance with the list of 23andMe SNPs. Because of the substantial estimation error expected for individual SNPs in single cohorts, sign concordance below 50% is only suggestive of data problems, except in the case of very large cohorts. The overall tendency is for the signs to align more often than expected by chance and in no case did our 95% CIs allow us to reject a sign concordance below 50%.

   b. We examined how SE's predicted from the *N*'s and *SD*'s supplied in the descriptive statistics compared to the *SE*'s in the results files. Winkler et al. (2014) propose a similar diagnostic (the SE-*N* Plots) which is based on following approximation to the standard error of a coefficient estimated by OLS

$$SE_j \approx \frac{\hat{\sigma}_Y}{\sqrt{N}} \cdot \frac{1}{\sqrt{2\ MAF_j\left(1 - MAF_j\right)}}, \tag{3.2}$$

where $\hat{\sigma}_Y$ is the standard deviation of the dependent variable (equal to 1 in cohorts that reported standardized regression coefficients), MAF_j is the minor allele frequency of SNP j, and N is the sample size. We used Equation (3.2) to generate a predicted standard error for the 50K SNP set, and we then plotted these predicted standard errors against the reported standard errors. We used an analogous equation for cohorts with binary dependent variables that ran logistic regressions. These plots, which we refer to as 50K plots in what follows, were used to check for systematic discrepancies between the predicted and reported standard errors and for outlier *SE*'s. These analyses helped us identify and remedy QC problems or errors in some of the reported summary statistics

iii. To supplement the 23andMe sign tests, we used bivariate LD score regression (Bulik-Sullivan, Finucane, et al., 2015) to estimate the pairwise genetic correlation, r_g, between the 23andMe sample and each cohort with more than 5,000 observations. We also estimated the r_g between the 23andMe sample and several combinations of small samples that were meta-analyzed (each of the individual cohorts, considered individually, is too small to generate informative estimates of r_g). In the 16 out of 23 instances where the estimator converged, the estimate is positive, as expected. We reject a genetic correlation of zero at the 5% significance level in 8 cases.

Again, any anomaly discovered during the course of applying the quality-control steps described above were raised with analysts. In some cases, multiple iterations with analysts were required before the source of the anomaly was identified, problems fixed, and the results files were cleared for inclusion.

3.4.2 Primary meta-analyses of SWB, PA and LS

Our analyses are based exclusively on results files that have passed the diagnostic tests described in the previous section. In our primary analyses of the pooled SWB phenotype, we use the WB variable from LSPAWB cohorts. In LSPA cohorts, we used the LS variable, because our analyses suggested its SNP-based heritability was slightly higher, except in cohorts where the LS variable is binary and the PA variable is not. For remaining cohorts, we use whichever results file is available (LS or PA) in the combined analysis. Table A1 also shows that the constructs are highly genetically correlated.

Though we consider them secondary to the SWB analyses, we also performed separate analyses of PA and LS. We thus ran three meta-analyses:

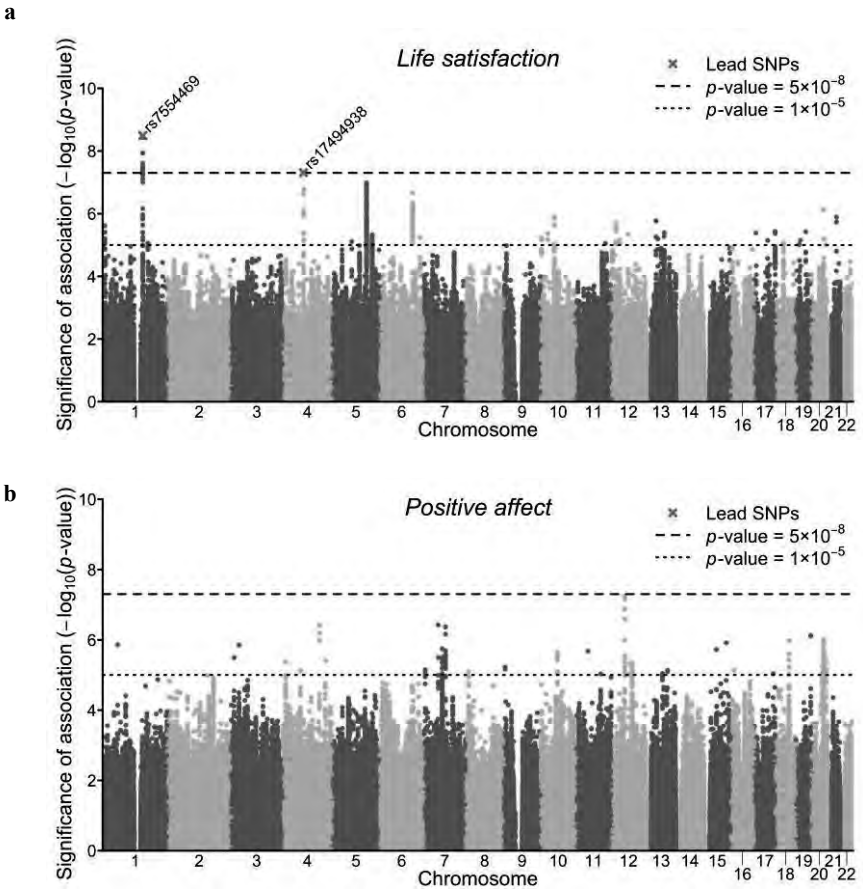
- SWB (N = 298,420)
- PA (N = 180,281)
- LS (N = 166,205)

For each phenotype, we used the software Metal (Willer, Li, & Abecasis, 2010) to perform a sample-size weighted meta-analysis of the cohort-level summary statistics. With the exception of one cohort, described below, we do not apply cohort-level genomic control (Devlin & Roeder, 1999) to adjust the standard errors for non-independence. Instead, we meta-analyze the unadjusted cohort-level summary statistics and subsequently inflate the standard errors from the meta-analysis by the square root of the estimated intercept from an LD score regression (Bulik-Sullivan, Loh, et al., 2015). The exception is deCODE, whose cohort-level regression estimates are not adjusted for the high level of relatedness in the sample: deCODE's standard procedure is to apply genomic control prior to uploading meta-analysis results. Because the LD score regression intercepts are upward biased in cohorts with related individuals unless the relatedness is accounted for in standard error estimation, we estimate the LD score intercept omitting deCODE from the analysis.

The Manhattan plots for the SWB, LS and PA analyses are shown in Figure 3.1a and Figure 3.7a-b, QQ plots are in Figure 3.9a-c. As is expected under polygenicity (Yang, Weedon, et al., 2011), there is more QQ plot inflation of the median test statistic in the combined analysis ($\lambda_{GC} = 1.206$) than in the separate analyses of LS ($\lambda_{GC} = 1.119$) and PA ($\lambda_{GC} = 1.118$). The estimated LD intercepts suggest that the amount of confounding is minimal: the estimates are 1.012 (SWB), 1.011 (PA) and 1.007 (LS). These analyses suggest that nearly all of the observed inflation is due to polygenicity, a conclusion consistent with results from additional analyses reported in Supplementary Note 4 of Okbay, Baselmans, et al. (2016).

Table A5 reports the set of approximately independent SNPs that reached $p < 10^{-6}$ in each of the three meta-analyses (for the SNPs that reached $p < 10^{-5}$, see Okbay, Baselmans, et al., 2016, Supplementary Table 7). To determine these independent loci, we used the following algorithm. First, the SNP with the smallest p-value is identified in the pooled meta-analysis results. This SNP is the lead SNP of clump 1. Second, we identified all SNPs whose LD with the lead SNP exceeds $R^2 = 0.1$ and assigned them to the clump. We calculate LD using the 1000G phase 1 reference sample composed of Utah Residents (CEPH) with Northern and Western European Ancestry (CEU), Toscani in Italia (TSI), and British in England and Scotland (GBR). To generate the second clump, the SNP with lowest p-value among the SNPs that remain after removal of clump 1 is identified and the same steps are applied to identify the set of SNPs comprising clump 2. The process is repeated until no SNPs with p-values below 10^{-5} remain. This process left us with 44 approximately independent loci in the SWB

Figure 3.7. Manhattan plots of GWAS results.



Note: **(a)** Life satisfaction ($N = 166{,}205$), **(b)** Positive affect ($N = 180{,}281$). The x-axis is chromosomal position, and the y-axis is the significance on a $-\log_{10}$ scale. The upper dashed line marks the threshold for genome-wide significance ($p = 5\times10^{-8}$); the lower line marks the threshold for nominal significance ($p = 10^{-5}$). Each approximately independent genome-wide significant association ("lead SNP") is marked by ×. Each lead SNP is the lowest p-value SNP within the locus, as defined by our clumping algorithm.

analysis (3 genome-wide significant), 30 in the LS analysis (2 genome-wide significant), and 30 in the PA analysis (none genome-wide significant). If we instead apply cohort-level genomic, all three SNPs which reach genome-wide significance in the SWB analysis remain genome-wide significant, whereas one of the two LS SNPs remains genome-wide significant.

3.4.3 Post hoc meta-analysis of SWB (1000G)

When this project was launched in 2012, the standard reference panel for imputation in GWAS meta-analyses was HapMap Phase 2 (The International HapMap Consortium, 2007), and our primary analysis of SWB was therefore restricted to testing HapMap2 SNPs for association with SWB. As outlined in the "Genotyping and imputation" section, cohorts with 1000G-imputed data nevertheless account for over 70% of the combined sample. Following the suggestion of a referee, we performed a post hoc GWAS of SWB in which we analyzed 1000G SNPs in all cohorts for which such data were available.

The twin goals of this post hoc analysis were to check if the finer resolution of the 1000G-imputed data would allow us to (i) fine-map the three genome-wide significant associations identified in the primary analysis and (ii) identify any novel associations.

A. PARTICIPATING COHORTS

We included 18 out of the 19 cohorts that used a 1000G reference sample for imputation (listed in Okbay, Baselmans, et al. 2016, Supplementary Table 4). The exception is deCODE; the deCODE GWAS was performed in 1000G-imputed data, but only HapMap2 SNPs were tested for association. To further increase the sample size, we reran the SWB GWAS in four of the cohorts (STR1, MCTFR, RS1, RS2 and RS3) for which only results for HapMap2-imputed variants were available when the primary meta-analysis was performed. These four cohorts were included because the QC team had access to individual-level genotypic and phenotypic data, making it feasible to rerun these analyses in 1000G SNPs. Below, we refer to the 24 (19 + 5) cohorts with GWAS results for 1000G SNPs as the 1000G cohorts. Their combined sample size is N = 229,883, or approximately 77% of the sample in the primary SWB analysis.

B. QUALITY CONTROL

We followed the quality-control procedures similar to those described in the section "Quality-control procedures" with quality-control parameters shown in Panel B of Table A3, and the following additional filters:

1. If the data were imputed against the September or December 2013 releases of the 1000 Genomes Phase 1 haplotypes provided by the software IMPUTE2, we drop the 730+199 SNPs whose strands were incorrectly aligned in these releases.

2. We drop indels and structural variants.

To generate a harmonized one-to-one mapping from rsID to chromosomal coordinates for the 1000G variants, we created a new reference file for use by EasyQC (Winkler et al., 2014). All our quality-control analyses were performed using this file, whose construction is described below.

Our reference file was generated by processing data downloaded on December 22, 2015, from the website of the imputation software MACH (Y. Li et al., 2010).

The first file consists of individual-level data on all European-ancestry (EUR) individuals in 1000G Phase 1 Integrated Release Version 3 Haplotypes (hereafter, 1000G Phase 1) :

File name: phase1_release_v3.20101123.snps_indels_svs.genotypes.refpanel.EUR.vcf
URL: http://csg.sph.umich.edu/abecasis/mach/download/1000G.2012-03-14.html

The second file consists of individual-level data on all individuals in the 1000G Phase3 v5 Reference (hereafter, 1000G Phase 3):

File name: reduced.ALL.phase3_shapeit2_mvncall_integrated_v5.20130502.genotypes.vcf
URL: http://csg.sph.umich.edu/abecasis/mach/download/1000G.Phase3.v5.html

From both data sets, we retain individuals who are members of the CEU (Utah Residents (CEPH) with Northern and Western European Ancestry), TSI (Toscani in Italia), or GBR (British in England and Scotland) populations. Additionally, we used the software Plink to restrict the sample so that it does not include any pairs of individuals whose estimated genomic relatedness exceeds 0.025 (Chang et al., 2015). As in the main analyses, we also restrict all analyses to autosomal, biallelic SNPs.

Imposing these restrictions in the 1000G Phase 1 sample yields a sample of $N = 258$ CEU/GBR/TSI approximately unrelated individuals and 16,001,120 SNPs. Each SNP in the resulting sample is uniquely identified by its ChrPosID, and each ChrPosID in turn maps to a unique rsID. Imposing the same restrictions in the 1000G Phase 3 sample leaves $N = 294$ CEU/GBR/TSI individuals and 43,805,190 SNPs. We drop 514,910 markers with non-unique ChrPosIDs. After this restriction is imposed, each variant is again uniquely identified by its ChrPosID, and each ChrPosID maps to a unique rsID.

We subsequently impose the following restrictions:

1. We drop 28,518,368 SNPs whose minor allele count is zero in the CEU/GBR/TSI subsample.

2. We drop 3,085 SNPs whose two alleles are not consistent across the Phase 1 and Phase 3 releases.

3. We drop 77,964 SNPs because their ChrPosIDs do not map to the same rsID in the Phase 1 and Phase 3 samples.

4. We drop 10,308 SNPs because the absolute value of the difference in allele frequency between the Phase 1 and Phase 3 sample exceeds 0.25 in the CEU/GBR/TSI subsamples.

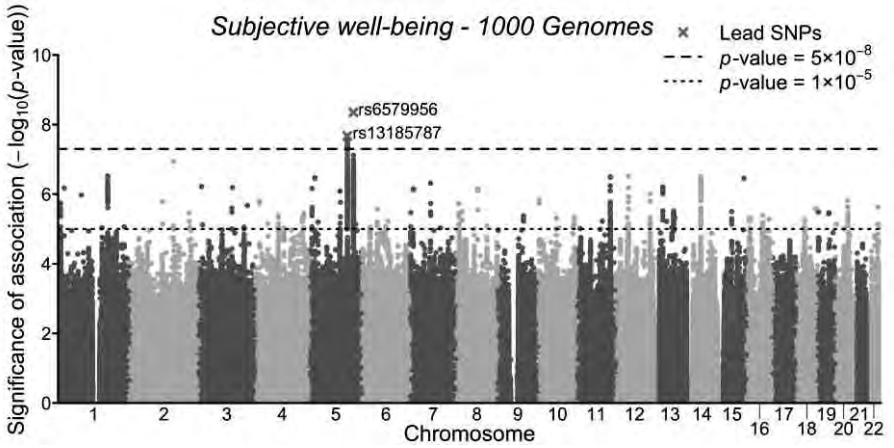
These restrictions leave us with our reference file containing 14,680,555 SNPs, which we use both for quality control and pruning. To maximize comparability with the results from the primary analyses, we closely followed the procedures used in the primary SWB analysis. The results from the SNP filtering are in Panel B of Table A4.

C. RESULTS OF POST HOC ANALYSIS

We sample-size weighted the cohort-level summary statistics using in the software Metal (Willer et al., 2010). As in the main analysis, we adjusted the standard errors using square root of the estimated intercept from an LD score regression (Bulik-Sullivan, Loh, et al., 2015). The association analyses were performed using the same association models as the primary analyses. Restricting the meta-analysis to SNPs with association results for at least $N = 100{,}000$ individuals leaves us with 9.00M SNPs passing all quality control filters.

The Manhattan plot for SWB is displayed in Figure 3.8, QQ plot is in Figure 3.9b. As expected given the smaller sample size, the observed level of inflation (λ_{GC}=1.124) is a little lower than estimated in the primary analyses (λ_{GC} =1.21). And consistent with the results in the primary analysis, the estimated LD intercept (1.008) suggests a minimal amount of confounding. The combined analyses yielded 76 nominally associated at the p-value threshold 10^{-5} (compared to 44 loci in the main SWB analysis). 19 SNPs that reached $p < 10^{-6}$ are reported in Panel D of Table A5 (the full set of 76 SNPs that reached $p < 10^{-5}$ are listed in Okbay, Baselmans, et al. 2016, Supplementary Table 7 – Panel D). Two of these 76 SNPs reached genome-wide significance: rs6579956 (4.44×10^{-9}) and rs13185787 (2.13×10^{-8}). Both SNPs lie in a long-range LD region on chromosome 5 (between 129 and 132 Mb ; Price et al., 2008).

Figure 3.8. Manhattan plot for post hoc subjective well-being analysis using 1000G SNPs



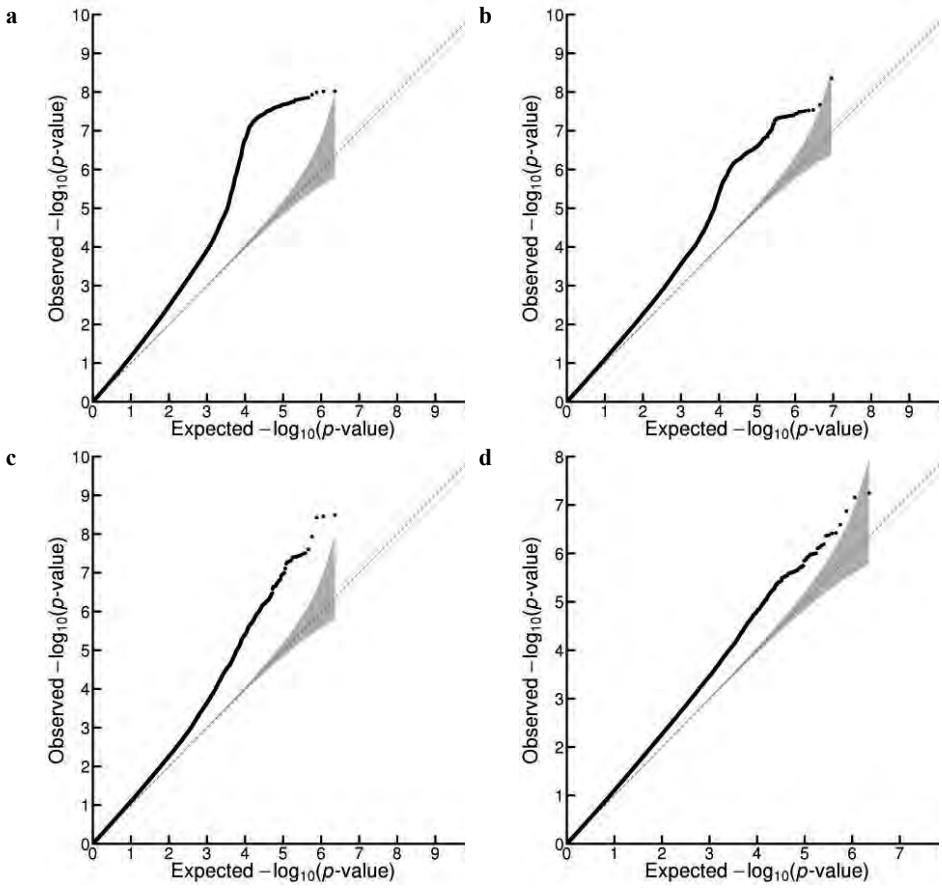
Note: SNPs are plotted on the x-axis according to their position on each chromosome against association with the phenotype on the y-axis (shown as –log10 *p*-value). The solid line indicates the threshold for genome-wide significance ($p = 5\times10^{-8}$) and the dashed line the threshold for suggestive hits ($p = 1\times10^{-5}$). Each independent genome-wide significant association ("lead SNP") is marked by ×.

The first novel SNP, rs6579956, is in linkage disequilibrium with one of the three original SWB-associated SNPs identified in the primary analysis (rs4958581). rs6579956 is positioned 1090kb downstream from rs4958581 and the two SNPs are only in modest linkage disequilibrium ($R^2 = 0.39$). The SNPs lie in a region with long intergenic non-coding RNA (lincRNA). Although lincRNAs are not protein-coding, the majority of lincRNAs are thought to have functional consequences, such as regulation of gene expression or conservation of transcript integrity (Mercer, Dinger, & Mattick, 2009).

Following the methodology described in Supplementary Note 9.C of Okbay, Baselmans, et al. (2016), we used the Genotype-Tissue Expression Portal (www.GTExportal.org) (The GTEx Consortium, 2013) to test both SNPs for association with gene expression levels across 15 tissues. We found that rs6579956 was significantly associated with gene expression of *IK*, *PCDHA1*, *PCDHA3*, *PCDHA4*, *PCDHA7*, *PCDHA10*, *SLC4A9*, and *TMCO6*, whereas we found no evidence that rs4958581 was associated with gene expression levels. We found no evidence that either SNP was in moderate to high LD with any nonsynonymous SNPs.

The second novel hit rs13185787 lies in the intronic region of *CDC42SE2,* and is in high linkage disequilibrium ($R^2 = 0.95$) with the other SWB-associated SNP on chromosome five identified in the primary analysis, rs3756290, located in the intronic region of *RAPGEF6.*

Figure 3.9. Quantile-quantile plots for primary subjective well-being analysis, post hoc subjective well-being analysis using 1000G SNPs, life satisfaction analysis and positive affect analysis.



Note : Panel **(a)** is the primary subjective well-being analysis, **(b)** the post hoc subjective well-being analysis using 1000G SNPs, **(c)** the life satisfaction analysis, and **(d)** the positive affect analysis. The estimated LD score intercepts used to adjust the standard errors are 1.012, 1.008, 1.007, and 1.011, respectively. The gray shaded areas in the Q−Q plots represent the 95% confidence intervals under the null hypothesis.

We conducted analyses identical to those described in the previous paragraph, but found no evidence that either SNP is significantly associated with gene expression levels across 15 tissues tested, nor that either SNP is in moderate to high LD with any nonsynonymous SNPs.

3.4.4 Auxiliary GWAS of depressive symptoms and neuroticism

A. PARTICIPATING COHORTS

Our meta-analysis of depressive symptoms ("DS") and neuroticism were conducted combining summary statistics from published meta-analyses with new genome-wide analysis. In our first auxiliary GWAS of neuroticism, we combine summary statistics from a meta-analysis conducted by the Genetics of Personality Consortium (GPC; De Moor et al., 2015) with our own analyses of UKB data. In our analyses of depressive symptoms ("DS"), we combine summary statistics from a GWAS of major depressive disorder performed by the Psychiatric Genomics Consortium (PGC; Ripke et al., 2013) with own analyses of data from two additional cohorts: the UKB and the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort (dbGaP, 2015).

PGC and GPC studies are conventional meta-analyses performed by pooling data from multiple cohorts of genotyped European-ancestry subjects. The GERA Cohort (dbGaP, 2015) is a genotyped subsample of participants in the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH). GERA cohort members have been linked to administrative health records. Our fourth cohort, UKB, is a British prospective cohort study that targeted individuals aged between 40 and 69 and collected data on them between 2006 and 2010. Of the roughly 500,000 enrollees, data have been released for about 150,000 individuals as part of what is called the interim release (Sudlow et al., 2015). For a a summary overview of the cohorts, their sampling frames, the auxiliary meta-analysis in which the cohort was used, details on how the studies restricted the estimation samples to European-ancestry subjects, and, in the case of publicly available data, the exact data file used in our analyses, see Supplementary Table 8 in Okbay, Baselmans, et al. (2016).

B. PHENOTYPES

Depressive Symptoms

In the UKB, our measure of DS is constructed as the standardized sum of the responses to the following two questions:

1. "Over the past two weeks, how often have you felt down, depressed or hopeless?"
2. "Over the past two weeks, how often have you had little interest or pleasure in doing things?"

Both questions have five response categories: “Not at all = 1”, “Several days = 2”, “More than half the days = 3”, “Nearly every day = 4” and “Do not know / Prefer not to answer = N/A”. We selected a relatively simple measure for two main reasons. First, we found exploratory LD score regressions (Bulik-Sullivan, Loh, et al., 2015) that our measure’s SNP-based heritability ($h^2 \approx 0.05$) is not appreciably lower than the SNP-based heritabilities of more detailed measures available in UKB. Second, unlike other more detailed mental-health measures available in UKB data (Smith et al., 2013), our measure is available for nearly all “White-British ancestry” respondents ($N = 105{,}739$).

In the GERA cohort, our DS measure is binary and constructed using data on patient encounters at Kaiser Permanente Northern California facilities during January 1, 1995, to March 15, 2013. Participants are classified as having major depressive disorder if they had at least two diagnoses of depression on separate days during this eighteen-year time window according to the ICD 9 CM classification system (dbGaP, 2013; “International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM),” n.d.). Our final estimation sample contains 7,231 cases and 49,137 controls, all of whom are of European ancestry.

Finally, the PGC (Ripke et al., 2013) measure varied across cohorts, but classified as cases those individuals with a diagnosis of lifetime MDD satisfying the criteria in the Diagnostic and Statistical Manual of Mental Disorders (see Table S2 in Ripke et al., 2013). We use summary statistics from the mega-analysis of 9,240 cases and 9,519 controls.

We used bivariate LD score regression (Bulik-Sullivan, Loh, et al., 2015) to estimate the pairwise genetic correlations between the three measures used by the three cohorts. Overall, found high genetic correlations:

- $\hat{r}_g(\text{UKB}, \text{PGC}) = 0.797$ ($SE = 0.108$),
- $\hat{r}_g(\text{UKB}, \text{GERA}) = 0.972$ ($SE = 0.216$)
- $\hat{r}_g(\text{PGC}, \text{GERA}) = 0.588$ ($SE = 0.242$).

Neuroticism

In the UKB, our measure of neuroticism is the respondent’s score on a 12-item version of the Eysenck Personality Inventory Neuroticism scale (Eysenck & Eysenck, 1975). This variable is available for 107,245 individuals. GPC (De Moor et al., 2015) used a combination of neuroticism phenotypes harmonized across 29 discovery cohorts. The neuroticism batter-

ies used by the cohorts varied, and included the NEO Personality Inventory, Eysenck Personality Questionnaire, and International Personality Item Pool inventory. The combined sample size of the contributing cohorts is $N = 63{,}666$. The genetic correlation between UKB and GPC (De Moor et al., 2015) was estimated to be 1.112 ($SE = 0.143$; Table A1).

Supplementary Table 9 in Okbay, Baselmans, et al. (2016) provides details on the phenotypes used in the auxiliary analyses, their distribution, the size of the final estimation sample, and, when applicable, the exact file with summary statistics from previously published studies used.

C. GENOTYPING AND IMPUTATION

Three of the four studies (PGC being the exception) imputed their data against a 1000G reference panel. The genotyping and imputation of the UKB interim release data have been described extensively elsewhere (UK Biobank, 2015a). The GERA genotype data were imputed using 2,504 individuals from the 1000G Phase 3 reference panel (Sudmant et al., 2015). Supplementary Table 10 in Okbay, Baselmans, et al. (2016) summarizes information about the genotyping and imputation procedures in the cohorts included our auxiliary GWAS meta-analyses.

D. ASSOCIATION ANALYSES

In the UKB analyses of DS and neuroticism, we followed the guidelines in the UKB's "exemplar GWAS" (UK Biobank, 2015a). We restrict the estimation sample to the "White-British ancestry" subsample and run linear regressions controlling for 15 principal components, indicator variables for genotyping array, sex, indicator variables for age ranges, and sex-by-age interactions.

The 29 cohorts contributing to the GPC meta-analysis all estimated linear regression models. The exact specifications varied across cohort, see p. 644 in De Moor et al. (2015) for details on cohort-level controls. All cohorts controlled for sex and age and most sought to account for stratification by including controls for principal components. The exact number of PCs used appears to have been left at the discretion of the analyst.

In our analyses of the binary DS indicator in GERA, we ran logistic regressions with controls for four principal components of the genotypic data, sex, and 14 indicator variables for age ranges.

Finally, PGC estimated a logistic regression model that included controls for 5 PCs, sex, age, and cohort fixed effects. For details, see Ripke et al. (2013). Unlike the GPC study, the PGC investigators did not meta-analyze summary statistics but instead asked contributing cohorts to upload individual level phenotypic and genetic data.

E. QUALITY CONTROL

To each results file (including those obtained from publicly available summary statistics), we applied the quality-control filters described in Section 3.4.1-F. For the three cohorts with 1000G data, we use the reference files provided by EasyQC for 1000G data to harmonize the mapping from rsIDs to ChrPosIDs across results files and check for strand issue[††].

For PGC, we use the reference file whose construction was described in Section 3.4.1-F. Table A6 reports the number of SNPs dropped in each filtering step.

F. META-ANALYSES

Our meta-analysis of neuroticism was performed using sample-size weighting. In the meta-analysis of DS, we weight the UKB by sample size but to improve statistical power, we weight the two case-control studies by effective sample size, defined as:

$$N_{eff} = \frac{4}{N_{cases}^{-1} + N_{controls}^{-1}} \tag{3.3}$$

as recommended by Willer et al. (2010). We thus ran the following two meta-analyses:

- DS (N = 180,866; N_{eff} = 149,707)
- Neuroticism (N = 170,911)

With the exception of the weighting scheme used in the DS analyses, these meta-analyses were performed exactly as were the SWB analyses. We used the software program METAL (Willer et al., 2010) to meta-analyze all SNPs that passed the quality-control filters, and we adjusted the standard errors of the resulting meta-analytic estimates by the square root of the estimated LD Score intercept. These adjustment factors were λ = 1.008 for depression and 1 for neuroticism (the actual point estimate being lower than one, λ = 0.9978).

[††]http://homepages.uni-regensburg.de/~wit59712/easyqc/1000g/allelefreq.1000G_EUR_p1v3.impute_legends.no-Mono.noDup.noX.v2.gz. Accessed on 22 June 2015.

G. RESULTS

Depressive Symptoms

The Manhattan and QQ-plots for DS are provided in Figure 3.1b and Figure 3.10a, respectively.

Panel A of Table A7 lists the 15 approximately independent SNPs that reach $p < 10^{-6}$ in our pooled meta-analysis of UKB, PGC and GERA (for the set of 54 SNPs that reach $p < 10^{-5}$, see Supplementary Table 25 – Panel A in Okbay, Baselmans, et al., 2016). Two of these SNPs, *rs7973260* ($p = 1.78\times10^{-9}$) and *rs62100776* ($p = 8.45\times10^{-9}$) reach genome-wide significance. The first SNP (*rs7973260*) is available in UKB and PGC, with consistent signs of the effect. There is no reliable proxy tagging the SNP in GERA (the best available proxy has $R^2 = 0.21$). The second SNP (*rs62100776*) is only available in UKB. For this SNP, a high-LD proxy (*rs8099160*; $R^2 = 0.971$) is available in all three cohorts. This SNP reaches genome-wide significance in the combined meta-analysis ($p = 2.68\times10^{-8}$), though the effect in the GERA cohort is in the opposite direction to the estimated effects in UKB and PGC (though not significantly so).

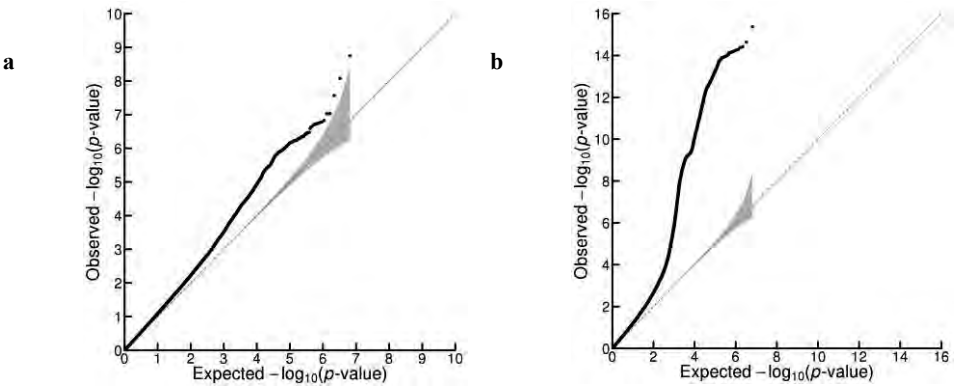
A recent paper reports two genome-wide significant associations with recurrent depression in a case-control study of Chinese women (Cai et al., 2015). The original report showed that the associations do not replicate in the PGC sample. We too were unable to replicate the associations in our larger meta-analysis sample.

Neuroticism

Figure 3.1c and Figure 3.10b display the Manhattan and QQ-plots of the GWAS meta-analysis on neuroticism, respectively.

Panel B of Table A7 shows the 42 approximately independent SNPs reaching p-value $< 10^{-6}$ in the meta-analysis of neuroticism (for the set of 117 SNPs that reach $p < 10^{-5}$, see Supplementary Table 25 – Panel B in Okbay, Baselmans, et al., 2016). Sixteen of these reach genome-wide significance.‡‡ Six of them tag a known inversion on chromosome 8, and one tags a known inversion on chromosome 17 (for additional analyses, see Supplementary Note 5 in Okbay, Baselmans, et al., 2016). None of the 117 SNPs we identify are in linkage dise

Figure 3.10. Quantile–quantile plots for depressive symptoms and neuroticism meta-analyses



Note: (**a**) Depressive symptoms. (**b**) Neuroticism. The estimated LD Score intercept used to adjust the standard errors was 1.008 for depressive symptoms. No adjustment was applied to the neuroticism results because the estimated intercept was below 1 (0.999). The gray-shaded areas in the quantile−quantile plots represent the 95% confidence intervals under the null hypothesis.

quilibrium with the genome-wide significant hit (rs35855737) reported by GPC (De Moor et al., 2015). However, in the independent UKB data, we find that the association with rs35855737 has the same sign and has a *p*-value of 0.06.

Robustness Analysis of Chromosome 18 SNPs

Two of the SNPs on chromosome 18 (rs1557341 and rs12961969) reach genome-wide significance in unconditional analyses and their pairwise linkage disequilibrium is below our cutoff of $R^2 = 0.10$. They therefore satisfy our definition of approximate independence. However, because the SNPs are in close physical proximity and in weak linkage disequilibrium ($R^2 = 0.09$), the evidence that they reflect independent genetic signals is less conclusive than for the other SNPs listed in Table 3.1.

To further investigate the robustness of the evidence that the SNPs reflect distinct genetic signals, we conducted two robustness analyses. In each of these, we sought to estimate the *conditional* effects of the two SNPs; that is, the effect of each SNPs conditioning on the allele count of the other.

In our second robustness check, we performed conditional and joint (COJO) multiple-SNP analyses (Yang et al., 2012). An advantage of COJO is that it is a method that can generate

estimates of conditional effects from summary statistics, thus increasing statistical power to detect independent signals. A potential downside of the approach is that in order for conditional estimates to be reliably estimable from summary statistics, an accurate estimate of the pattern of linkage disequilibrium between the SNPs in the samples that generated the summary statistics is required. In practice, the linkage equilibrium is often estimated from a reference sample, and any differences in LD between the reference sample and the sample used in the GWAS that generated the summary statistics can cause biases.

In our COJO analysis, we used genotype data from the Health and Retirement Study, imputed to the 1000 Genomes Phase I reference panel and converted to best-guess genotypes, to estimate the linkage disequilibrium structure. To generate the sample we used to estimate the linkage equilibrium structure, we began with 8,652 genotyped respondents with European-ancestry individuals. From this sample, we filtered out SNPs with minor allele frequency below 1%, imputation accuracy below 0.7 or with Hardy-Weinberg equilibrium (HWE) *p*-values below 10^{-6}. These filters leave 7,779,969 SNPs. We subsequently estimated pairwise genetic relationships between all individuals using the directly genotyped SNPs, dropping one of each pair individuals with an estimated degree of relatedness above 0.025. This procedure leaves us with our final sample of 8,359 individuals.

Procedurally, COJO is implemented through a stepwise procedure in the software GCTA (Yang, Lee, Goddard, & Visscher, 2011). To convey the intuition, we give a sketch of the key steps of the algorithm here. To initiate the model, the lowest *p*-value SNP is selected from a given list of SNPs. The conditional effects (and *p*-values) of all remaining SNPs in the list are subsequently estimated using the LD structure from the reference sample. COJO then drops the SNP with greatest *p*-value (unless this SNP is genome-wide significant). The conditional effects of remaining SNPs are then estimated again, and the SNP with greatest *p*-value dropped (again, unless this SNP is genome-wide significant). The process is repeated iteratively until only genome-wide significant SNPs remain.

Applying COJO to the chromosome 18 GWAS summary statistics obtained from the neuroticism meta-analysis, we found that rs1557341 remained genome-wide significant ($p = 5.6\times10^{-9}$). In a joint model including rs1557341 and rs12961969, neither SNP reaches genome-wide significance (rs1557341 $p = 1.35\times10^{-5}$; rs12961969 $p = 5.4\times10^{-5}$). Thus the evidence that these SNPs reflect independent genetic signals is weaker than for the remaining SNPs identified in our main analysis (see Table 3.1).

3.4.5 Polygenic prediction

We tested how well a polygenic score for SWB, based on the GWA meta-analysis results, could predict life satisfaction (LS), positive affect (PA) and our composite SWB measure in two independent holdout cohorts: the Health and Retirement Study (HRS; Sonnega et al., 2014) and the Netherlands Twin Register (NTR; van Beijsterveldt et al., 2013; Willemsen et al., 2013). Additionally, we tested how well our polygenic score for SWB could predict different personality traits as measured by the NEO Big Five (Costa & Widiger, 1992): Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Furthermore, because SWB has a strong negative phenotypic correlation with depression, we tested whether the polygenic score for SWB could predict DS. In NTR, DS was measured using the Achenbach System of Empirical Based Assessment (ASEBA) DSM-oriented scale (Achenbach & Rescorla, 2001). In HRS, the CIDI SF assessment was used to measure DS. Finally, for comparison purposes, we also tested how well our polygenic score could predict height, a phenotype that showed practically no phenotypic correlation with SWB but is also highly polygenic.

Here, we report the prediction analysis in HRS. For details on the NTR analysis, see Supplementary Note 6 and Supplementary Table 33 in Okbay, Baselmans, et al. (2016).

A. PHENOTYPE MEASURES

Life satisfaction and positive affect*:* LS and PA questions were administered to 8,248 and 8,285 genotyped participants in four waves. LS was measured using the Satisfaction with Life Scale consisting of five items (e.g., “In most ways my life is close to ideal”), and responses were given on a six-point scale (Diener et al., 1985). The PA measure differs across waves. In 2006, it was measured using eight questions (e.g., “During the past thirty days, how much of the time did you feel...extremely happy?”) from the Midlife Development Inventory (Brim et al., 2003), which was adapted from some well-known instruments such as the Affect Balance Scale (Bradburn, 1969), the University of Michigan’s Composite International Diagnostic Interview (Kessler et al., 1994), the Manifest Anxiety Scale (Taylor, 1953), the Health Opinion Survey (Macmillan, 1957), the General Well-Being Schedule (Fazio, 1977), and the Center for Epidemiological Studies Depression Scale (Radloff, 1977). Responses were given on a five-point scale. In 2008, 2010 and 2012, it was measured using thirteen questions (five-point scale). Eleven of these questions were obtained from Positive and Negative Affect Schedule—Expanded Form (PANAS-X) (David Watson & Clark, 1994), and the remaining two were chosen from two other studies in this area (Carstensen, Pasupathi, Mayr, & Nesselroade, 2000; Ong, Edwards, & Bergeman, 2006). For both LS and

PA, a score was constructed for each time point by taking the mean across items, and the score was set to missing if more than half of the items were unanswered. The final score was constructed for each time point by taking the sum across items, and was set to missing if more than two of the items were unanswered. When subjects have a missing score on one item of the scale, this missing value was replaced by the subject's mean scale score.

Personality: The Big Five personality traits (neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness) were measured in four waves between 2006 and 2012, with 26 items in 2006-2008 and 31 in 2010-2012. The original 26 items in 2006 and 2008 were obtained from the MIDUS survey (Brim et al., 2003). Extraversion (N = 8,271), Agreeableness (N = 8,271), and Conscientiousness (N = 8,268) were measured with five items, Openness to Experience (N = 8,253) with seven, and Neuroticism (N = 8,264) with four. In 2010 and 2012, five items from the International Personality Item Pool (Lachman & Weaver, 1997) were added to the "Conscientiousness" sub-dimension. The responses were given on a four-point scale (Costa & Widiger, 1992). For each trait, scores were constructed by taking the mean of all items in the respective category after recoding opposite-stated items. A score was set to missing if more than half of the items in that category had missing values.

Depressive symptoms: DS was measured in ten waves (1995-2012) for 8,617 individuals using the CIDI-SF questionnaire, which implements the diagnostic criteria of DSM. The CIDI-SF starts with a small number of screen questions that are used to skip out participants least likely to have depressive symptoms. If the screen questions are endorsed for the necessary intensity and duration, participants are asked about seven symptoms. Summary scores equal to the number of symptoms (ranging from zero to seven) are assigned to all respondents that endorse one of the screen questions. If one or more symptom questions are unanswered, we set the summary score to missing. Instead of assigning a diagnosis of depression based on the number of symptoms, we followed the strategy of assigning a probability of "caseness" to each summary score as detailed by Nelson, Kessler, and Mroczek (1998). The probabilities of caseness that correspond to the summary scores are derived from the US National Comorbidity Survey (Kessler et al., 1994) and reflect the probability that a respondent with a certain response profile would meet the diagnostic criteria if given the full CIDI interview. If a respondent was screened out of the interview because of not endorsing any of the screen questions for the necessary duration or intensity, s/he was assigned a probability of caseness equal to zero.

Height: Height data were based on self-report, available for 8,650 HRS participants. All participants were older than eighteen years old.

The final phenotypes for LS, PA, Big Five personality traits, and DS were constructed by first grouping the respondents based on which combination of waves they have responded in, and then taking the standardized residuals from a regression, within each group, of the phenotypic score on sex, age, age^2, and all interactions. For the individuals who have responded in multiple waves, the average phenotypic score was used for obtaining the residuals. The composite SWB measure (N = 8,226) was created by taking the average of the residualized PA and LS scores when both are available, and set to missing if either PA or LS is missing. Height was similarly constructed, but no grouping was made based on number of responses. The first height measurement available was residualized on sex, birth year, birth year squared, and interactions.

B. POLYGENIC SCORES

We ran a meta-analysis of the pooled SWB phenotype excluding HRSs, applying a minimum sample size filter of 100,000 individuals. Using these summary statistics, we constructed two sets of polygenic scores: (1) LDpred polygenic scores (LD-PGS) using LDpred effect sizes (Vilhjálmsson et al., 2015), and (2) linear polygenic scores (Lin-PGS) using the effect sizes from the original meta-analyses (Purcell et al., 2009). LDpred adjusts the effect sizes from the meta-analysis for the effects of linkage disequilibrium (LD) using an external reference panel to estimate the LD structure among SNPs. As the LD reference panel, we used the HRS genotype data imputed to 1000 Genomes Phase 1 reference panel and converted to hard calls. Since HapMap3 SNPs are known to be imputed reliably, the LD-PGS are based on HapMap3 SNPs only. For the results with LD-PGS and Lin-PGS to be comparable, we also restricted the Lin-PGS to HapMap3 SNPs. We constructed all scores in PLINK using allelic dosages of genotypes imputed to 1000G Phase 1. For LD-PGS we set the fraction of causal SNPs to 1, and the Lin-PGS were obtained using all HapMap3 SNPs, without applying a p-value threshold. 1,059,092 SNPs were used to construct the scores.

C. RESULTS

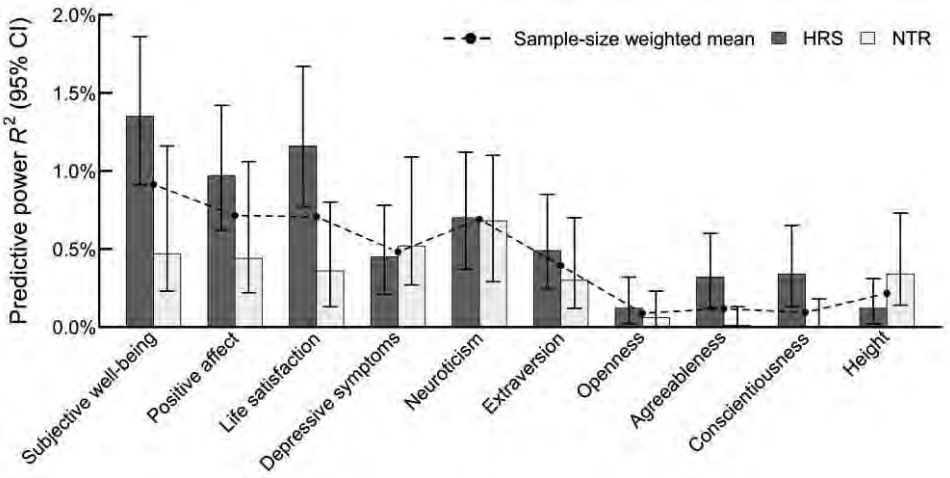
We first regressed each residualized phenotype on 10 principal components, using Ordinary Least Squares (OLS). The principal components were computed using HapMap3 SNPs with a minor allele frequency greater than 0.01, and restricting the sample to individuals of European ancestry. Next, we ran the same regressions adding the score as a covariate and computed the increase in R^2. The incremental R^2's for each phenotype and polygenic score are shown in Table A8. To obtain 95% confidence intervals (CI) around the incremental R^2's, bootstrapping was performed with 1000 repetitions.

The results of the polygenic score analysis in HRS are depicted in Figure 3.11, along with the results in NTR. The predictive power (incremental R^2 after controlling for age, age^2, sex, and PCs) for SWB is 1.35%, in HRS which is statistically distinguishable from zero but small relative to the predictive power attained by polygenic scores estimated in comparable sample sizes for other phenotypes (e.g., educational attainment). Similarly, the mean predictive power for its components, LS and PA, are both 1.16% and 0.97%, respectively.. SWB also significantly predicts DS (0.45%) and the personality traits Neuroticism (0.70%) and Extraversion (0.49%). In contrast, for the other phenotypes, the predictive power is in all cases smaller than 0.35% and in many cases not statistically distinguishable from zero. We compared the LDpred results with the linear approach. The results are similar but with the expected somewhat lower estimates for the linear method.

3.4.6 Proxy-phenotype and genetic overlap analyses

In this section, we describe and report the results from tests of joint enrichment that allow us to formally test if the SNPs showing the strongest evidence of association with one phenotype (for example, SWB), are more strongly associated with another phenotype (for example, DS) than expected by chance. The analyses are motivated by the evidence of the strong genetic correlations between SWB, DS, and neuroticism (Bartels et al., 2013; Kendler & Myers, 2009; Weiss et al., 2008), including the results shown in Table A1.

Figure 3.11. Polygenic score prediction in HRS and NTR



Note: Predictive power of the polygenic score constructed from the subjective well-being GWAS results in two independent holdout cohorts (HRS and NTR). Predictive power is tested for subjective well-being, positive affect, life satisfaction, depressive symptoms, the Big Five personality traits (which include neuroticism), and height.

A. METHODOLOGY FOR PROXY-PHENOTYPE AND CROSS-PHENOTYPE ENRICHMENT ANALYSES

We use a two-stage approach that has been successfully applied in other contexts (Rietveld, Esko, et al., 2014). In the first stage, we conduct a meta-analysis of a first-stage "proxy phenotype" (e.g., SWB). In the second stage, we test the "lead/lead-proxy SNPs"—the SNPs showing strongest evidence of association with the first-stage phenotype—for association with a second-stage phenotype (e.g., DS) in an independent (non-overlapping) sample. Note that in the analyses described in this section, relative to the GWAS on SWB, DS, and neuroticism reported in Sections 3.4.2 and 3.4.4, we omit cohorts from the first-stage or second-stage as needed to ensure that the samples in the two stages are non-overlapping.

In total, we perform three lookup exercises; see Table A9 for a summary overview of the analyses, including cohort restrictions used to eliminate overlap between the stage-one and stage-two samples. In our analysis of DS, we apply the effective sample-size weighting scheme described in Section 3.4.4-F to the two case-control studies (GERA and PGC; dbGaP, 2015; Ripke et al., 2013) and continue to weight UKB by its sample size. As in the main analyses, we perform sample-size-weighted meta-analyses of SWB and neuroticism.

For convenience, in what follows we adopt the convention of naming each lookup analysis in the format "First-stage phenotype → Second-stage phenotype". We conducted three lookup exercises. In our first lookup exercise, the first- and second-stage phenotypes are, respectively, SWB and DS, or simply SWB → DS. Our second lookup is SWB → Neuroticism, and our third lookup is SWB → Height, where we treat Height as a negative control.

We omit from the meta-analysis of the second-stage phenotype SNPs missing from a substantial fraction of individuals; see the notes in Table A9 for details. For example, in the analysis where the second-stage phenotype is DS, we only consider SNPs available in all three DS cohorts (GERA, PGC and UKB; dbGaP, 2015; Ripke et al., 2013; Sudlow et al., 2015b). And in the analysis where the second-stage is neuroticism, we only consider SNPs available in our two neuroticism cohorts, GPC (De Moor et al., 2015) and UKB (Sudlow et al., 2015), with a minimum total sample size of $N = 90{,}000$. Below, we describe the methodology we used to construct the lead SNPs, and the tests of enrichment we performed.

B. GENERATING LEAD SNPS

Throughout, we apply a uniform methodology to define the lead SNPs that are subsequently tested for association, both jointly and individually, with the second-stage phenotype. For brevity, we illustrate the methodology used to construct our list of lead/lead-proxy SNPs

using the example of SWB. However, the procedure used in the other two lookups is nearly identical, as explained in the relevant subsections below.

We began by identifying a set of approximately independent "SWB-associated SNPs" from the first-stage meta-analysis (or more generally, "first-stage-phenotype-associated SNPs"). We applied the clumping methodology described in Section 3.4.2, but with a *p*-value threshold for the index SNPs of 10^{-4}. The more liberal *p*-value threshold was chosen prior to the study based on power calculations. As in our main analyses, we used the 1000G phase 1 reference sample (Abecasis et al., 2012) composed of Utah Residents (CEPH) with Northern and Western European Ancestry (CEU), Toscani in Italia (TSI), and British in England and Scotland (GBR) for clumping and for estimating linkage disequilibrium.

Applying the clumping procedure to the SWB meta-analysis results from the SWB → DS lookup generated 223 approximately independent SWB-associated lead SNPs. Of these, 85 were available in all three DS cohorts used in the second-stage analyses, whereas 148 were not. For each of these 148 SNPs, we examined if there are any SNPs satisfying the following conditions: (i) the SNP is in high LD ($R^2 > 0.8$) with the SWB-associated SNP, and (ii) the SNP is available both in the SWB meta-analysis and in all three cohorts contributing to the meta-analysis of DS. A proxy-lead SNP satisfying these criteria was available for 78 out of 148 SNPs (mean $R^2 = 0.96$, range 0.81 to 1.00). Whenever more than one proxy is available for a SNP, we chose as our proxy the SNP whose R^2 with the SWB-associated SNP was the greatest. Our final list of lead SNPs in the first lookup exercise therefore contains 85+78 = 163 SNPs.

C. TESTING LEAD/PROXY-LEAD SNPS FOR ENRICHMENT

Because SWB, DS, and neuroticism phenotypes are all highly polygenic, it is of limited interest to test the null hypothesis that the *p*-value distribution of the lead/lead-proxy SNPs is uniform. We instead perform a non-parametric test of joint enrichment that probes whether the lead SNPs are more strongly associated with the second-stage phenotype than randomly chosen sets of SNPs with minor allele frequencies within one percentage point of the lead/proxy-lead SNP. To perform our test, we generated 1,000 matched SNPs for each of the Y lead/lead-proxy SNPs (e.g., $Y = 163$ in the SWB → DS analysis).

We then ranked the $Y \times 1000 + Y$ SNPs by *p*-value and conducted a Mann-Whitney test (Nachar, 2008) of the null hypothesis that the *p*-value distribution of the Y lead/lead-proxy SNPs are drawn from the same distribution as the $Y \times 1000$ matched SNPs. To test the individual lead SNPs for experiment-wide significance, we examine whether any of the lead

SNPs (or their high-LD proxies) are significantly associated with the second-stage phenotype at the Bonferroni-corrected significance level of $0.05/Y$. Throughout, we adopt the convention of classifying an effect size as "in the predicted direction" if either (i) the signs are concordant and the two phenotypes are estimated to have a positive genetic correlation, or (ii) the signs are discordant and the phenotypes are estimated to have a negative genetic correlation.

D. RESULTS FROM PROXY-PHENOTYPE AND CROSS-PHENOTYPE ENRICHMENT ANALYSES

Are SWB-Associated SNPs Enriched for Depression?

Figure 3.5a is a two-way scatterplot of the z-statistics of the lead/lead-proxy SNPs in SWB (horizontal axis) against DS (vertical axis). To aid interpretation, we choose the reference allele to be the SWB-increasing variant, so all z-statistics are by construction positive for the first-stage phenotype. On the basis of the negative genetic correlation reported in Table A1 ($\hat{\rho} = -0.81$), we expect plotted points to lie disproportionately below the dashed horizontal line at zero (i.e. negative z-statistics). That is indeed what we find: 116 out of 163 (71%) signs are in the expected direction. Moreover, for 19 out of the 20 SNPs that are nominally significantly associated ($p < 0.05$) in the analysis of DS, the association is in the predicted direction.

Three lead/proxy-lead SNPs reach p-value $< 10^{-7}$ in the SWB meta-analysis. Two of these are nominally associated with DS: rs12517563 ($p = 0.007$) and rs2075677 ($p = 0.0149$). Two other SNPs are significantly associated with depressive symptoms at the Bonferroni-corrected p-value threshold of 0.05/163 = 0.00037. These are rs6904596 ($p = 9.78\times10^{-5}$) and rs4481363 ($p = 3.06\times10^{-4}$). The direction of the association with depressive symptoms is in the predicted direction for all four SNPs (rs12517563, rs2075677, rs6904596, rs4481363): the SWB-increasing allele is estimated to reduce depression risk. Supplementary Table 16 in Okbay, Baselmans, et al. (2016) lists the association results for the lead/proxy-lead SWB-associated SNPs in the first-stage SWB meta-analysis and the second-stage depressive symptoms meta-analysis conducted in an independent sample. The SNPs are ordered by p-value attained in the SWB analysis (from smallest to largest). Among SWB-associated SNPs with p-value $< 10^{-5}$, 80% have signs in the predicted direction. Our test of joint enrichment rejects the null of no enrichment relative to the expected level for a randomly sampled set of SNPs matched on allele frequency ($p = 0.033$).

Are SWB-Associated SNPs Enriched for Neuroticism?

Applying the same clumping algorithm, we identified 170 lead/lead-proxy SNPs from the first-stage analysis of SWB. The results from this lookup analysis are summarized in Figure 3.5b, where the reference allele is again chosen to be the SWB-increasing allele. Given the negative genetic correlation reported in Table A1 ($\hat{\rho} = -0.75$), we expect *z*-statistics disproportionately below the dashed horizontal line. Indeed, 129 out of 170 signs (76%) are in the predicted direction in the neuroticism results. Moreover, all 28 SNPs that are nominally significant in the neuroticism analysis have the predicted sign. None of the three SNPs reaching *p*-value < 10^{-7} in the first-stage analysis are associated with neuroticism. However, four SNPs are significant at the Bonferroni-corrected significance threshold 0.05/173 = 0.00029. These are *rs10838738* ($p = 2.6\times10^{-5}$), *rs6904596* ($p = 4.2\times10^{-5}$), *rs4481363* ($p = 5.7\times10^{-5}$) and *rs10774909* ($p = 7.3\times10^{-5}$). In all four cases, the effects are in the expected direction. For complete results, see Supplementary Table 17 in Okbay, Baselmans, et al. (2016). Finally, our test of joint enrichment rejects the null of no enrichment relative to the expected level for a randomly sampled set of SNPs matched on allele frequency ($p = 10^{-4}$).

Negative-Control Analyses: Are SWB-Associated SNPs Enriched for Height?

For our negative-control analyses, our first-stage analyses of SWB were performed omitting cohorts that contributed to GIANT consortium's yeaommr-2010 study of height (Lango Allen et al., 2010), leaving us with a first-stage discovery sample of N = 229,853. Applying our methodology gives 181 lead/lead-proxy SNPs. Our second-stage lookup is conducted using publicly available summary statistics from the height GWAS (N = 133,859). We find no evidence that the proportion of SNPs for which the allele estimated to increase SWB is also the allele estimated to increase height is statistically distinguishable from 50% (p = 0.373), and the Mann-Whitney test of joint enrichment fails to reject the null hypothesis (p = 0.454).

3.4.7 Lookup of top SNPs in companion study of depression

A. BACKGROUND

We partnered with the investigators of an ongoing large-scale GWAS of major depressive symptoms (N = 368,890) to follow up on the associations identified in the depressive symptoms and neuroticism analyses. The participants of the study were all European-ancestry customers of 23andMe, a personal genomics company, who responded to online survey questions about mental health. The phenotype in this companion study is a binary indicator

equal to 1 if the respondent had experienced depression at least once. This phenotype was measured from survey questions administered to 23andMe customers and had a prevalence of about 25%. For full details on association models, quality-control filters, and the ascertainment of depression status, we refer to the companion study (Hyde et al., 2016).

We obtained association results from the investigators of the companion study for the 54 DS-associated SNPs, and 117 out of the 118 neuroticism-associated SNPs in Supplementary Table 15 in Okbay, Baselmans, et al. (2016) (one neuroticism SNP, rs117893837, was not available in the summary statistics from the companion study). The standard errors in the association statistics have been adjusted for inflation using the square root of the estimated LD-score intercept ($\sqrt{1.059}$; Bulik-Sullivan, Loh, et al., 2015).

Of the 54 DS-associated SNPs, in the 23andMe sample, 40 have the expected sign and ten are associated with DS at 1% level (always with a sign in the anticipated direction). Of the 117 neuroticism-associated SNPs available in the 23andMe sample, 85 have the expected sign, and 16 are significant at the 1% level (again, always with the anticipated signs).

B. META-ANALYSIS OF 54 DS-ASSOCIATED SNPS

We also meta-analyzed the results of our study and the results of the 23andMe cohort for the 54 available SNPs reaching $P < 10^{-5}$ in our analysis of DS. In this meta-analysis, we weight both cohorts by their effective sample size (assuming a prevalence of 25% in the companion study). Results for the SNPs that reach $P < 10^{-6}$ in the meta-analysis of 54 SNPs are shown in Table A10. Five of the 54 SNPs reach genome-wide significance in the weighted meta-analysis. For the full set of results, see Panel C of Supplementary Table 15 in Okbay, Baselmans, et al. (2016).

CHAPTER 4

Genome-wide Association Study Identifies 74 Loci Associated with Educational Attainment

Based on Okbay, Beauchamp et al. (2016)

Abstract

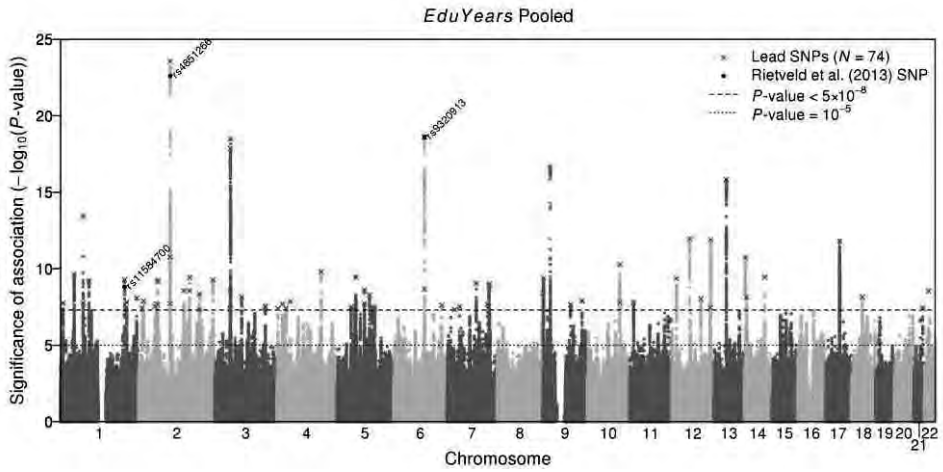
Educational attainment (EA) is strongly influenced by social and other environmental factors, but genetic factors are also estimated to account for at least 20% of the variation across individuals (Rietveld, Medland, et al., 2013). We report the results of a genome-wide association study (GWAS) for EA that extends our earlier discovery sample (Rietveld, Conley, et al., 2014; Rietveld, Medland, et al., 2013) of 101,069 individuals to 293,723 individuals, and a replication in an independent sample of 111,349 individuals from the UK Biobank. We now identify 74 genome-wide significant loci associated with number of years of schooling completed. Single-nucleotide polymorphisms (SNPs) associated with educational attainment are disproportionately found in genomic regions regulating gene expression in the fetal brain. Candidate genes are preferentially expressed in neural tissue, especially during the prenatal period, and enriched for biological pathways involved in neural development. Our findings demonstrate that, even for a behavioral phenotype that is mostly environmentally determined, a well-powered GWAS identifies replicable associated genetic variants that suggest biologically relevant pathways. Because EA is measured in large numbers of individuals, it will continue to be useful as a proxy phenotype in efforts to characterize the genetic influences of related phenotypes, including cognition and neuropsychiatric disease.

4.1 Introduction and Results

We study educational attainment (EA), which is measured in all main analyses as the number of years of schooling completed (*EduYears*, N = 293,723, mean = 14.33, SD = 3.61; see sections 4.2.1-4.2.2 for further detail). All genome-wide association studies (GWAS) were performed at the cohort level in samples restricted to individuals of European descent whose EA was assessed at or above age 30. A uniform set of quality-control (QC) procedures was applied to the cohort-level summary statistics. In our GWAS meta-analysis of ~9.3M SNPs from the 1000 Genomes Project, we used sample-size weighting and applied a single round of genomic control at the cohort level.

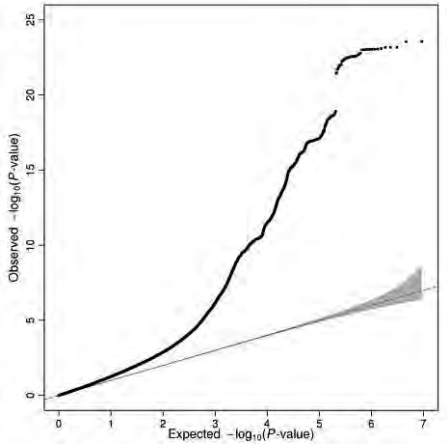
Our meta-analysis identified 74 approximately independent genome-wide significant loci. For each locus, we define the "lead SNP" as the SNP in the genomic region that has the smallest P-value (Section 4.2.6). Figure 4.1 shows a Manhattan plot with the lead SNPs highlighted. The three SNPs that reached genome-wide significance in the discovery stage of our previous GWAS meta-analysis of EA (Rietveld, Medland, et al., 2013) are also highlighted. The quantile-quantile (Q-Q) plot of the meta-analysis (Figure 4.2) exhibits inflation (λ_{GC} = 1.28), as expected under polygenicity (Yang, Weedon, et al., 2011).

Figure 4.1. Manhattan plot for EduYears associations (N = 293,723).



Note: The x-axis is chromosomal position, and the y-axis is the significance on a $-\log_{10}$ scale. The black line shows the genome-wide significance level (5×10^{-8}). The red x's are the 74 approximately independent genome-wide significant associations ("lead SNPs"). The black dots labeled with rs numbers are the 3 Rietveld, Medland, et al. (2013) SNPs.

Figure 4.2. Quantile-quantile plot of the genome-wide association meta-analysis of 64 *EduYears* results files.



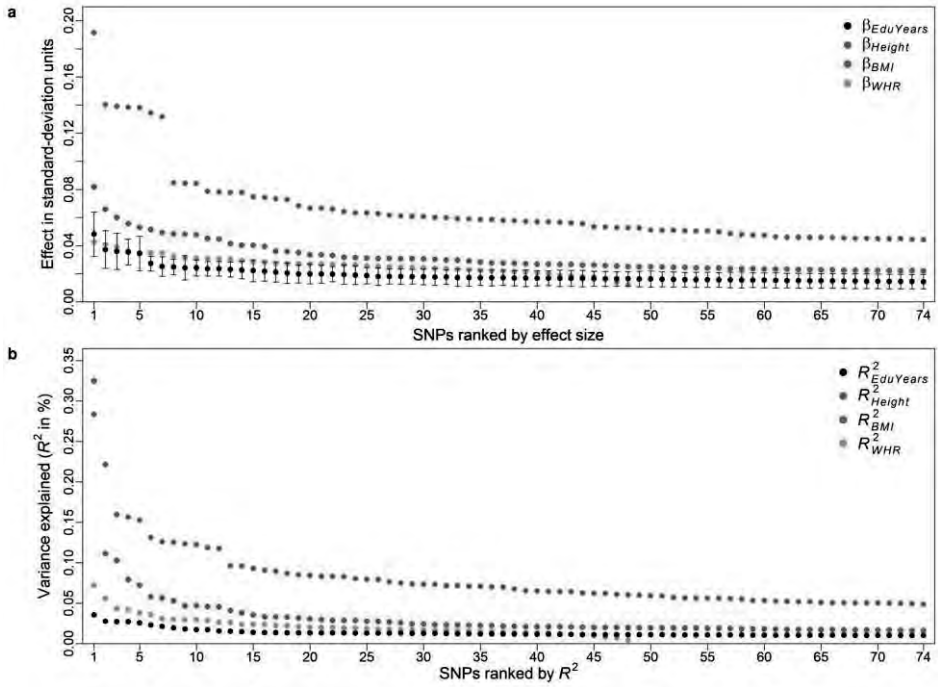
Note: Observed and expected *P*-values are on a $-\log_{10}$ scale. The grey region depicts the 95% confidence interval under the null hypothesis of a uniform *P*-value distribution.

Figure 4.3 shows the estimated effect sizes of the lead SNPs. The estimates range from 0.014 to 0.048 standard deviations per allele (2.7 to 9.0 weeks of schooling), with incremental R^2 in the range 0.01% to 0.035%.

To quantify the amount of population stratification in the GWAS estimates that remains even after the stringent controls used by the cohorts (Section 4.2.4), we used LD Score regression (Bulik-Sullivan, Loh, et al., 2015). The regression results indicate that ~8% of the observed inflation in the mean χ^2 is due to bias rather than polygenic signal (Figure 4.4a), suggesting that stratification effects are small in magnitude. We also found evidence that the genetic association signals taken as a whole replicate reliably in several within-family analyses (Figure 4.4b). See Supplementary Information section 2 of Okbay, Beauchamp, et al. (2016) for further detail on the population stratification analyses.

To further test the robustness of our findings, we examined the within-sample and out-of-sample replicability of SNPs reaching genome-wide significance (Sections 4.2.7-4.2.8). We found that SNPs identified in the previous EA meta-analysis replicated in the new cohorts included here, and conversely, that SNPs reaching genome-wide significance in the new cohorts replicated in the old cohorts. For the out-of-sample replication analyses of our 74 lead SNPs, we used the interim release of the U.K. Biobank (UKB ; $N = 111{,}349$; Sudlow et al., 2015a). As shown in Figure 4.5, 72 out of the 74 lead SNPs have a consistent sign (*P*

Figure 4.3. The distribution of effect sizes of the 74 lead SNPs.



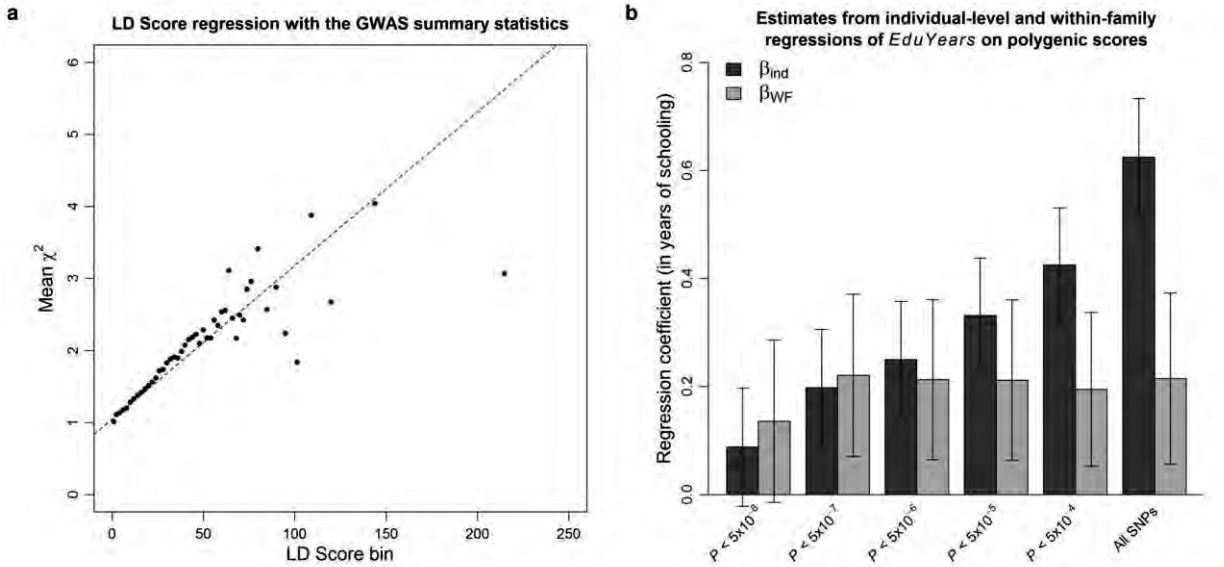
Note: **a.** SNPs ordered by absolute value of the standardized effect of one more copy of the education-increasing allele, with 95% confidence intervals. **b.** SNPs ordered by R^2. Effects on *EduYears* are benchmarked against the top 74 genome-wide significant hits identified in the largest GWAS conducted to date of height and body mass index (BMI), and the 48 associations reported for waist-to-hip ratio adjusted for BMI (WHR). These results are based on the GIANT consortium's publicly available results for pooled analyses restricted to European-ancestry individuals: https://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium.

$= 1.47\times10^{-19}$), 52 are significant at the 5% level ($P = 2.68\times10^{-50}$) and 7 reach genome-wide significance in the U.K. Biobank dataset ($P = 1.41\times10^{-42}$). For comparison, the corresponding expected numbers, assuming each SNP's true effect size is its estimated effect adjusted for the winner's curse, are 71.4, 40.3, and 0.6. (Section 4.2.8-C). We also find out-of-sample replicability of our overall GWAS results: the genetic correlation between *EduYears* in our meta-analysis sample and in the UKB data is 0.95 (s.e. = 0.021)[8].

.

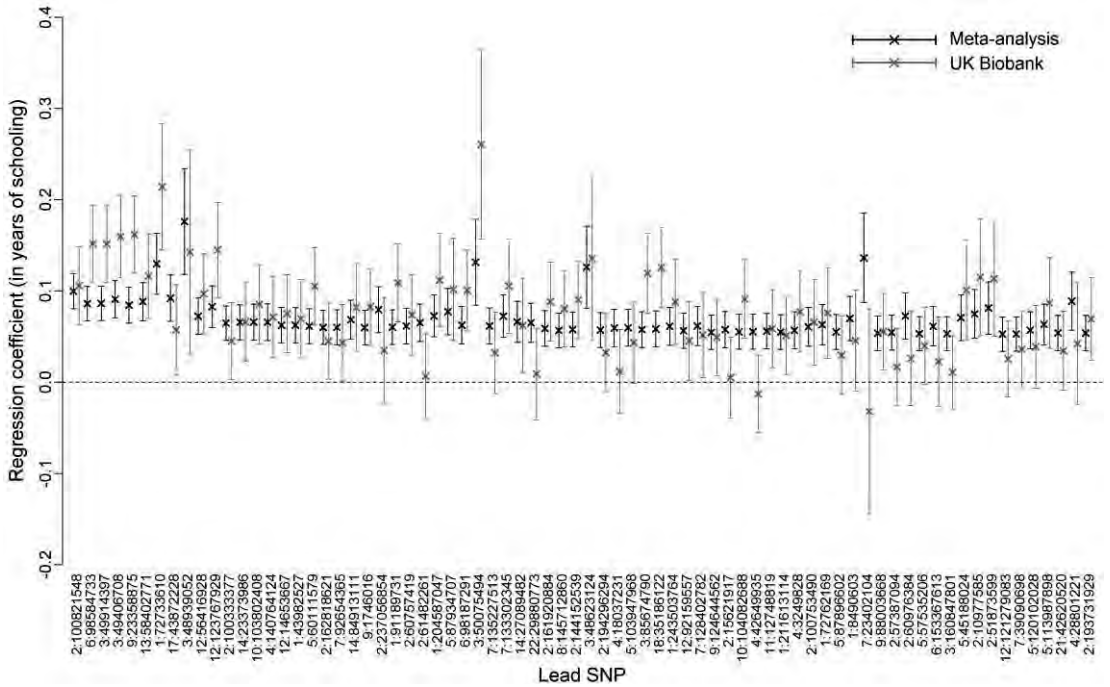
[8] The genetic correlation is obtained using the LDSC python software package and the "eur_w_ld_chr/" files of LD Scores computed by (Finucane et al., 2015). SE is standard error, estimated using a block jackknife over SNPs (by the LDSC software). In the LD Score regression, we include only HapMap3 SNPs with MAF > 0.01.

Figure 4.4. Assessing the extent to which population stratification affects the estimates from the GWAS.



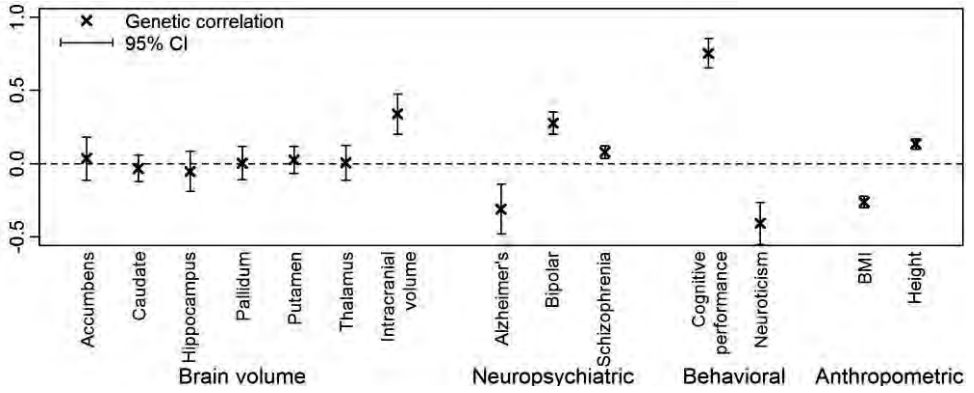
Note: **a,** LD Score regression plot with the summary statistics from the GWAS. Each point represents an LD Score quantile for a chromosome (the *x* and *y* coordinates of the point are the mean LD Score and the mean χ^2 statistic of variants in that quantile). The facts that the intercept is close to one and that the χ^2 statistics increase linearly with the LD Scores suggest that the bulk of the inflation in the χ^2 statistics is due to true polygenic signal and not to population stratification. **b,** Estimates and 95% confidence intervals from individual-level and WF regressions of *EduYears* on polygenic scores, for scores constructed with sets of SNPs meeting different *P*-value thresholds. In addition to the analyses shown here, we conduct a sign concordance test, and we decompose the variance of the polygenic score. Overall, these analyses suggest that population stratification is unlikely to be a major concern for our 74 lead SNPs. See Okbay, Beauchamp, et al. (2016), Supplementary Information section 3 for additional details.

Figure 4.5. Replication of 74 lead SNPs in the UK Biobank data.



Note: Estimated effect sizes (in years of schooling) and 95% confidence intervals of the 74 lead SNPs in the meta-analysis sample (N = 293,723) and the UK Biobank replication sample (N = 111,349). The reference allele is the allele associated with higher values of *EduYears* in the meta-analysis sample. SNPs are in descending order of R^2 in the meta-analysis sample. Of the 74 lead SNPs, 72 have the anticipated sign in the replication sample, 52 replicate at the 0.05 significance level, and 7 replicate at the 5×10^{-8} significance level.

Figure 4.6. Genetic correlations between EduYears and other traits.

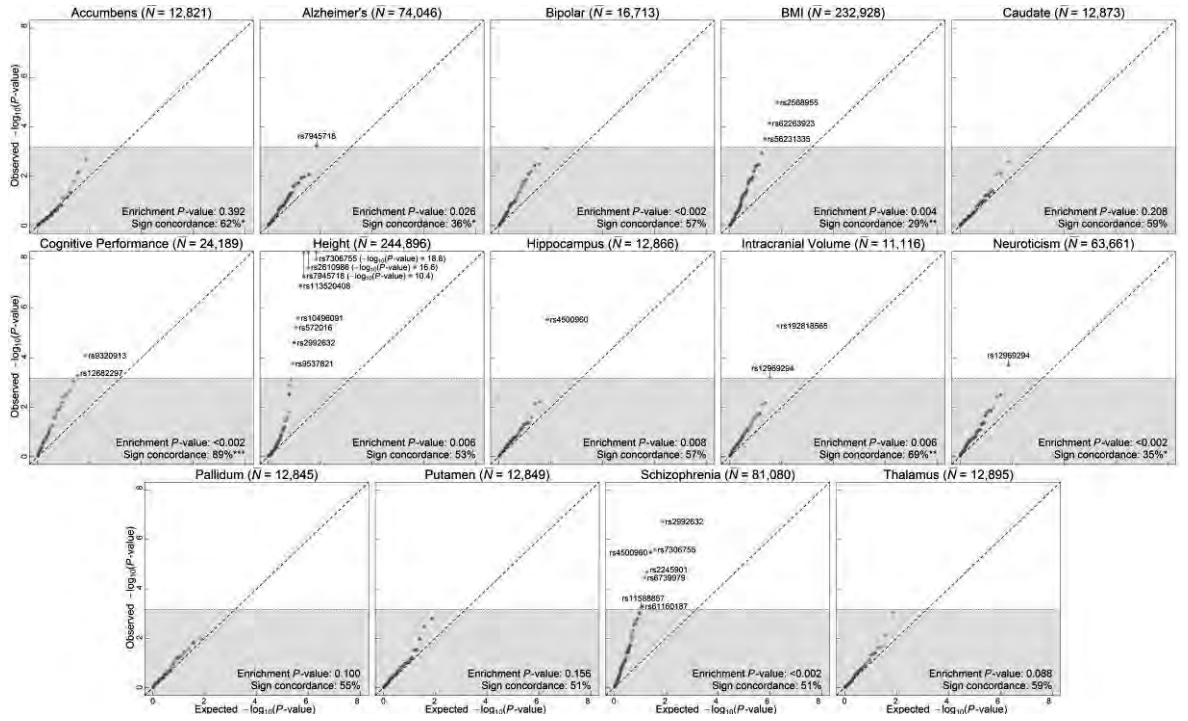


Note: Results from bivariate Linkage-Disequilibrium (LD) Score regressions (Bulik-Sullivan, Finucane, et al., 2015): estimates of genetic correlation with brain volume, neuropsychiatric, behavioral, and anthropometric phenotypes using published GWAS summary statistics. The error bars show the 95% confidence intervals.

It is known that EA, cognitive performance, and many neuropsychiatric phenotypes are phenotypically correlated, and several studies of twins find that the phenotypic correlations partly reflect genetic overlap (T. Fowler, Zammit, Owen, & Rasmussen, 2012; Tambs, Sundet, Magnus, & Berg, 1989; Thompson, Detterman, & Plomin, 1991). Here, we investigate genetic correlation using our GWAS results for *EduYears* and published GWAS results for 14 other phenotypes, using bivariate Linkage-Disequilibrium (LD) Score regression (Bulik-Sullivan, Finucane, et al., 2015). First, we estimated genetic correlations with *EduYears*. As shown in Figure 4.6, on average, alleles associated with greater EA are also associated with increased cognitive performance ($P = 9.9\times10^{-50}$) and intracranial volume ($P = 1.2\times10^{-6}$), increased risk of bipolar disorder ($P = 7\times10^{-13}$), decreased risk of Alzheimer's ($P = 4\times10^{-4}$), and lower neuroticism ($P = 2.8\times10^{-8}$). We also found positive, statistically significant, but very small, genetic correlations with height ($P = 5.2\times10^{-15}$) and risk of schizophrenia ($P = 3.2\times10^{-4}$).

Second, we examined whether our 74 lead SNPs are jointly associated with each phenotype (Figure 4.7; Okbay, Beauchamp, et al., 2016, Supplementary Information section 3.3.1). We reject the null hypothesis of no enrichment at $P < 0.05$ for 10 of the 14 phenotypes (all the exceptions are subcortical brain structures).

Figure 4.7. Q-Q plots for the 74 lead EduYears SNPs (or LD proxies) in published GWAS of other phenotypes.



Note: SNPs with concordant effects on both phenotypes are pink, and SNPs with discordant effects are blue. SNPs outside the gray area pass Bonferroni-corrected significance thresholds that correct for the total number of SNPs we tested ($P < 0.05/74 = 6.8\times10^{-4}$) and are labeled with their rs numbers. Observed and expected *P*-values are on a $-\log_{10}$ scale. For the sign concordance test: * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

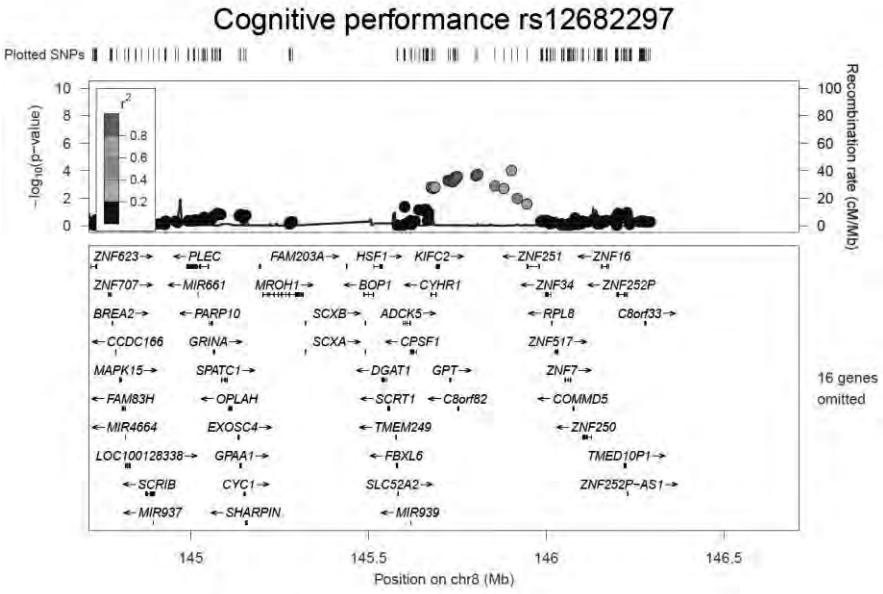
Third, for each phenotype, we tested (in the published GWAS results) each of our 74 lead SNPs or proxy for association at a significance threshold of 0.05/74. We found a total of 25 SNPs meeting this threshold for any of these phenotypes (but only one reaching genome-wide significance). While these results provide suggestive evidence that some of these SNPs may be associated with other phenotypes, further testing of these associations in independent cohorts is required (Figure 4.8; Okbay, Beauchamp, et al., 2016, Supplementary Tables 3.2-3.4).

To consider potential biological pathways, we first tested whether SNPs in particular regions of the genome are implicated by our GWAS results. Unlike what has been found for other phenotypes, SNPs in regions that are DNase I hypersensitive in the fetal brain are more likely to be associated with *EduYears* by a factor of ~5 (95% confidence interval 2.89–7.07; Figure 4.9). Moreover, the 15% of SNPs residing in regions associated with histones marked in the central nervous system (CNS) explain 44% of the heritable variation (Figure 4.10a; Okbay, Beauchamp, et al., 2016, Supplementary Table 4.4.2). This enrichment factor of ~3 for CNS ($P = 2.48\times10^{-16}$) is greater than that of any of the other nine tissue categories in this analysis.Given that our findings disproportionately implicate SNPs in regions regulating brain-specific gene expression, we examined whether genes located near *EduYears*-associated SNPs show elevated expression in neural tissue. We tested this hypothesis using data on mRNA transcript levels in the 37 adult tissues assayed by the Genotype-Tissue Expression Project (GTEx ; Ardlie et al., 2015). Remarkably, the 13 GTEx tissues that are components of the CNS—and only those 13 tissues—show significantly elevated expression levels of genes near *EduYears*-associated SNPs (FDR < 0.05; Figure 4.10b; Okbay, Beauchamp, et al., 2016, Supplementary Table 4.5.2).

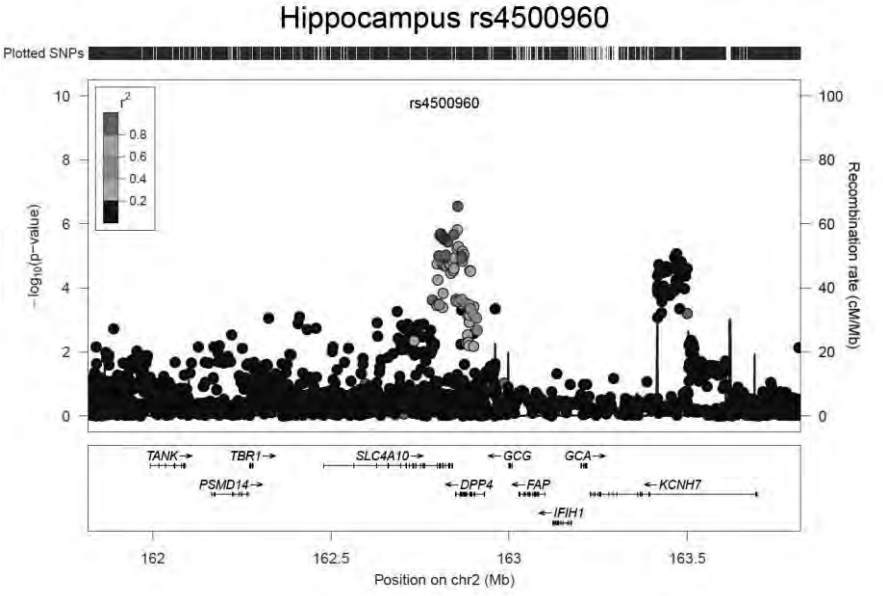
To investigate possible functions of the candidate genes from the GWAS associated loci, we examined the extent of their overlap with groups of genes ("gene sets") whose products are known or predicted to participate in a common biological process (Pers et al., 2015). We found 283 gene sets significantly enriched by the candidate genes identified in our GWAS (FDR < 0.05; Okbay, Beauchamp, et al., 2016, Supplementary Table 4.5.1). To facilitate interpretation, we used a standard procedure (Pers et al., 2015) to group the 283 gene sets into "clusters" defined by degree of gene overlap. The resulting 34 clusters, shown in Figure 4.11, paint a coherent picture, with many clusters corresponding to stages of neural development: the proliferation of neural progenitor cells and their specialization (the *cluster npBAF complex*), the migration of new neurons to the different layers of the cortex (*forebrain development*, *abnormal cerebral cortex morphology*), the projection of axons

Figure 4.8. Regional association plots for four of the ten prioritized SNPs for MHBA phenotypes identified using *EduYears* as a proxy phenotype.

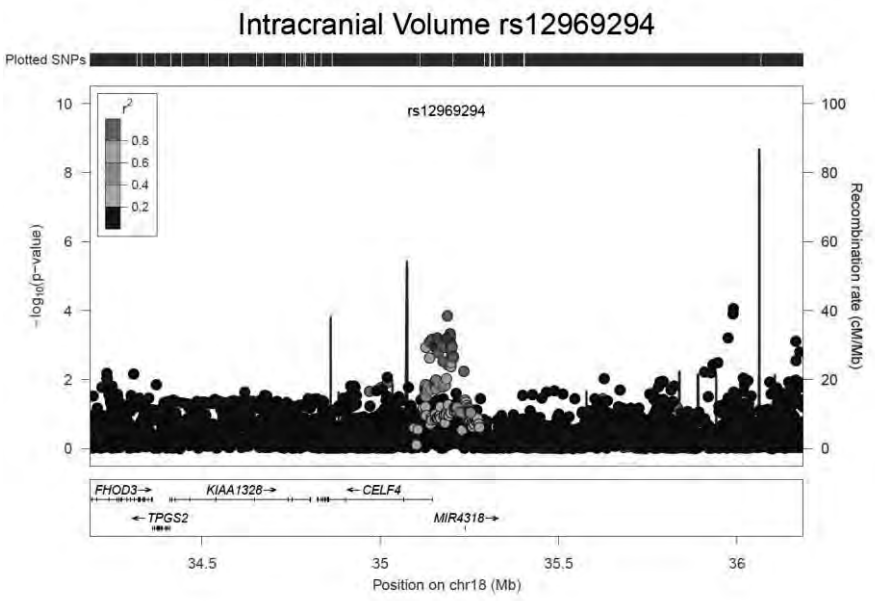
a.



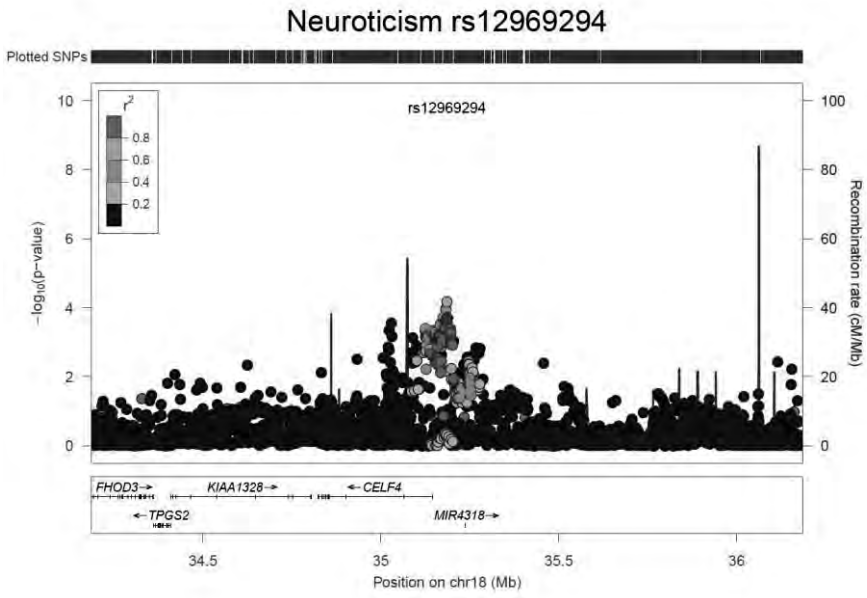
b.



c.

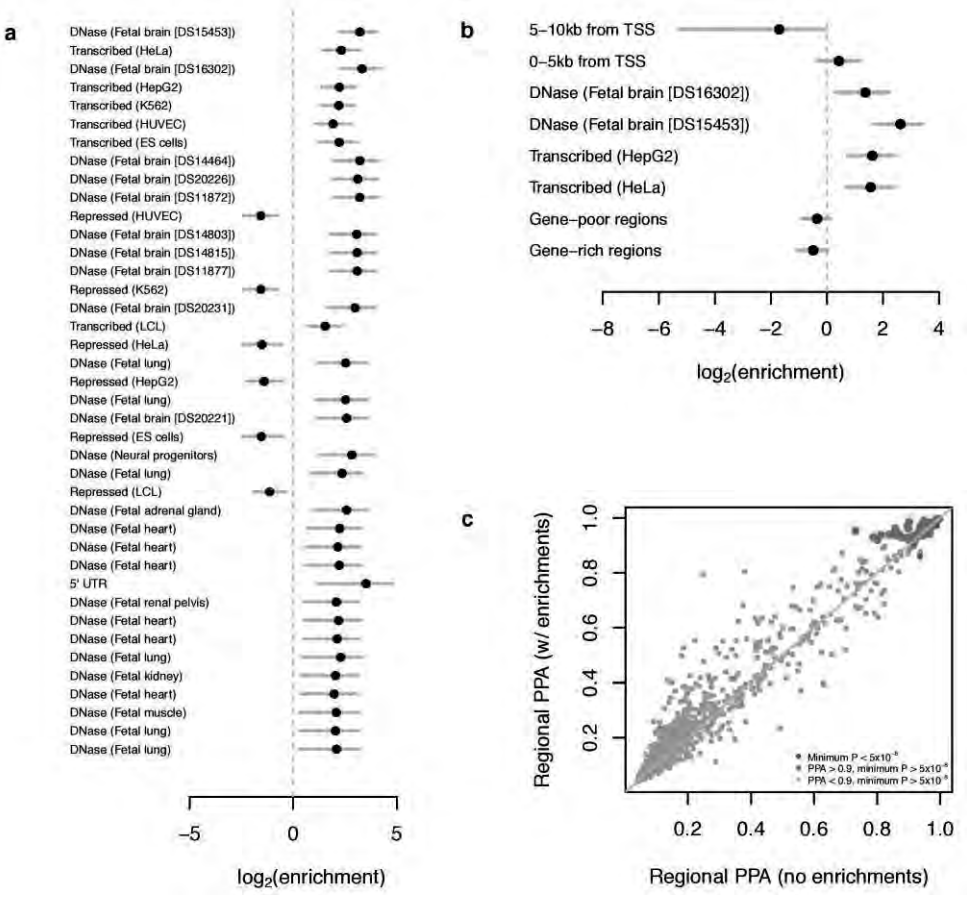


d.



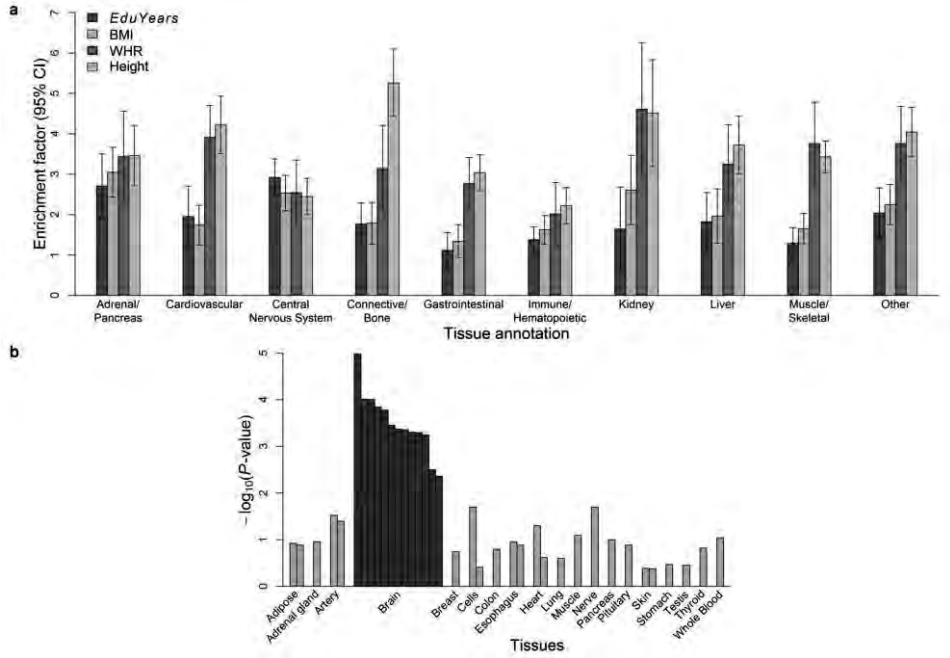
Note: **a,** cognitive performance; **b,** hippocampus; **c,** intracranial volume; **d,** neuroticism. The four were selected because very few genome-wide significant SNPs have been previously reported for these traits. Data sources and methods are described in Supplementary Information section 3. The R^2 values are from the hg19 / 1000 Genomes Nov 2014 EUR references samples. The figures were created with LocusZoom (http://csg.sph.umich.edu/locuszoom/). Mb, megabases.

Figure 4.9. Application of fgwas to EduYears.



Note: a, The results of single-annotation models. "Enrichment" refers to the factor by which the prior odds of association at an LD-defined region must be multiplied if the region bears the given annotation; this factor is estimated using an empirical Bayes method applied to all SNPs in the GWAS meta-analysis regardless of statistical significance. Annotations were derived from ENCODE and a number of other data sources. Plotted are the base-2 logarithms of the enrichments and their 95% confidence intervals. Multiple instances of the same annotation correspond to independent replicates of the same experiment. **b,** The results of combining multiple annotations and applying model selection and cross-validation. Although the maximum-likelihood estimates are plotted, model selection was performed with penalized likelihood. **c,** Reweighting of GWAS loci. Each point represents an LD-defined region of the genome, and shown are the regional posterior probabilities of association (PPAs). The *x*-axis give the PPA calculated from the GWAS summary statistics alone, whereas the *y*-axis gives the PPA upon reweighting on the basis of the annotations in b. The orange points represent genomic regions where the PPA is equivalent to the standard GWAS significance threshold only upon reweighting.

Figure 4.10. Tissue-level biological annotation.

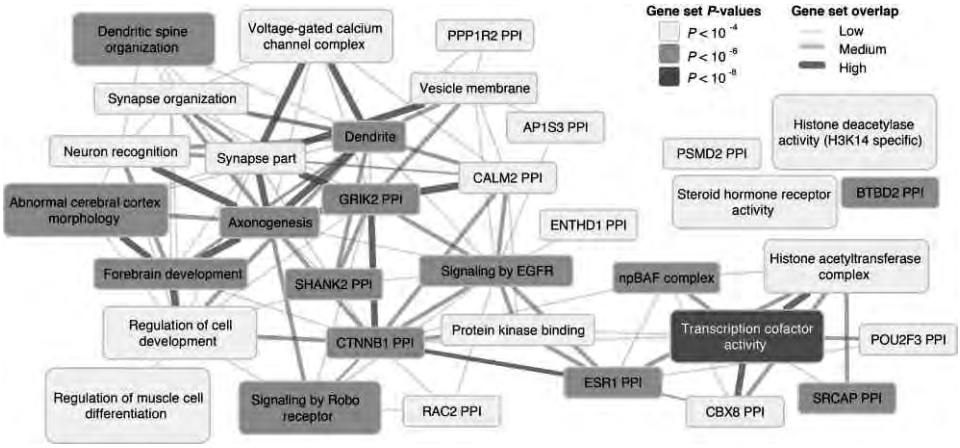


Note: **a,** The enrichment factor for a given tissue type is the ratio of variance explained by SNPs in that group to the overall fraction of SNPs in that group. To benchmark the estimates for *EduYears*, we compare the enrichment factors to those obtained when we use the largest GWAS conducted to date on body mass index, height, and waist-to-hip ratio adjusted for BMI. The estimates were produced with the LDSC python software, using the LD Scores and functional annotations introduced in Finucane et al. (2015) and the HapMap3 SNPs with MAF > 0.05. Each of the 10 enrichment calculations for a particular cell type is performed independently, while each controlling for the 52 functional annotation categories in the full baseline model. The error bars show the 95% confidence intervals. **b,** We took measurements of gene expression by the Genotype-Tissue Expression (GTEx) Consortium and determined whether the genes overlapping *EduYears*-associated loci are significantly overexpressed (relative to genes in random sets of loci matched by gene density) in each of 37 tissue types. These types are grouped in the panel by organ. The colored bars corresponding to tissues where there is significant overexpression. The *y*-axis is the significance on a $-\log_{10}$ scale.

from neurons to their signaling targets (*axonogenesis*, *signaling by Robo receptor*), the sprouting of dendrites and their spines (*dendrite*, *dendritic spine organization*), and neuronal signaling and synaptic plasticity throughout the lifespan (*voltage-gated calcium channel complex*, *synapse part*, *synapse organization*).

Many of our results implicate candidate genes and biological pathways that are active during distinct stages of prenatal brain development. To directly examine how the expression levels of candidate genes identified in our GWAS vary over the course of development, we used

Figure 4.11. Overview of biological annotation.



Note: 34 clusters of significantly enriched gene sets. Each cluster is named after one of its member gene sets. The color represents the *P*-value of the member set exhibiting the most statistically significant enrichment. Overlap between pairs of clusters is represented by an edge. Edge width represents the Pearson correlation ρ between the two vectors of gene membership scores ($\rho < 0.3$, no edge; $0.3 \leq \rho < 0.5$, thin edge; $0.5 \leq \rho < 0.7$, intermediate edge; $\rho \geq 0.7$, thick edge), where each cluster's vector is the vector for the gene set after which the cluster is named.

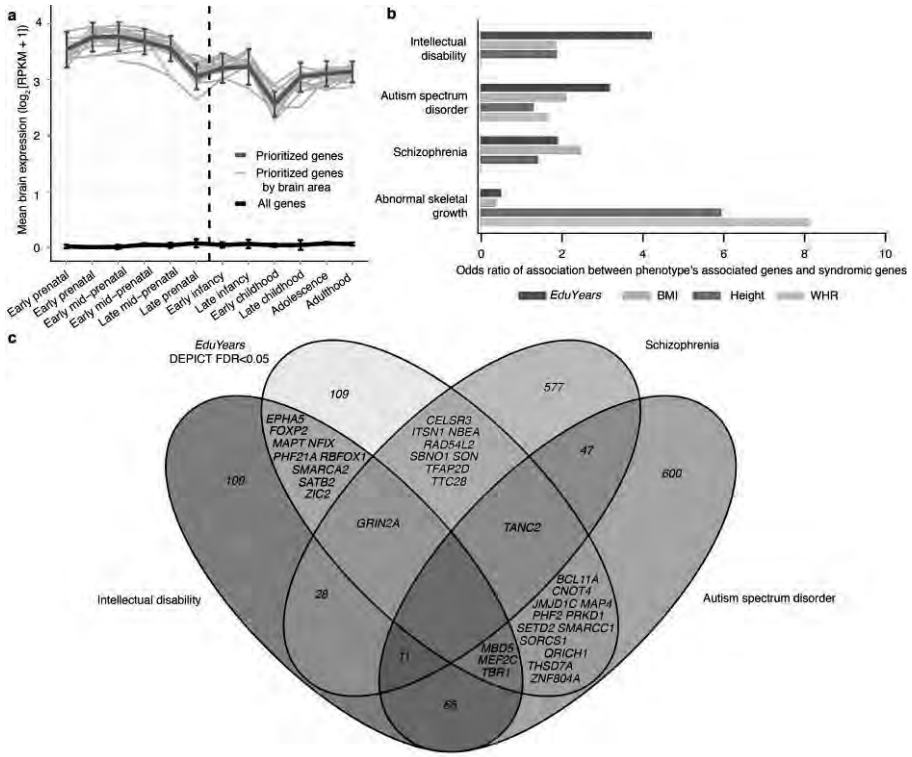
gene expression data from the BrainSpan Developmental Transcriptome (Allen Institute for Brain Science, 2015). As shown in Figure 4.12, these candidate genes exhibit above-baseline expression in the brain throughout life but especially higher expression levels in the brain during prenatal development (1.36 times higher prenatally than postnatally, $P = 6.02 \times 10^{-8}$).

A summary overview of some promising candidate genes for follow-up work is provided in Table 4.1.

We constructed polygenic scores (Purcell et al., 2009) to assess the joint predictive power afforded by the GWAS results (Okbay, Beauchamp, et al., 2016, Supplementary Information section 5.2). Across our two holdout samples, the mean predictive power of a polygenic score constructed from all measured SNPs is 3.2% ($P = 1.18 \times 10^{-39}$; Okbay, Beauchamp, et al., 2016, Supplementary Table 5.2 and Supplementary Information section 5).

Studies of genetic analyses of behavioral phenotypes have been prone to misinterpretation, such as characterizing identified associated variants as "genes for education." Such charac terization is not correct for many reasons: EA is primarily determined by environmental factors, the explanatory power of the individual SNPs is small, the candidate genes may not be

Figure 4.12. Gene-level biological annotation.

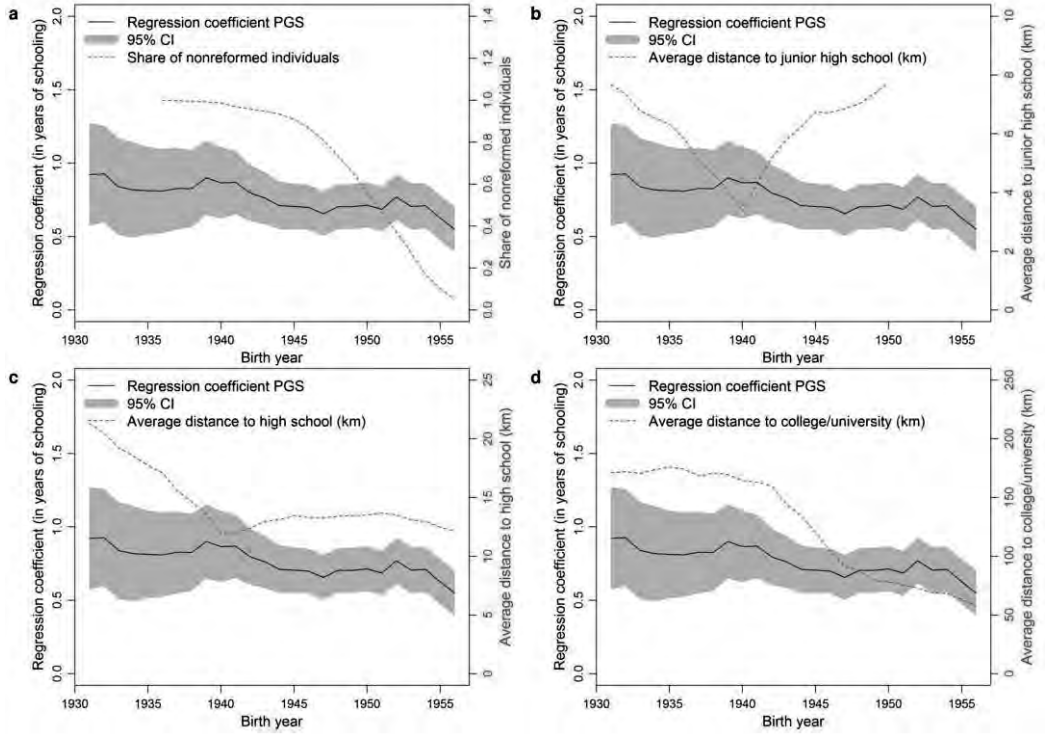


Note: **a,** The DEPICT-prioritized genes for *EduYears* measured in the BrainSpan Developmental Transcriptome data (red curve) are more strongly expressed in the brain prenatally rather than postnatally. The DEPICT-prioritized genes exhibit similar gene-expression levels across different brain regions (gray lines). Analyses were based on $\log_2$-transformed RNA-Seq data. Error bars represent 95% confidence intervals. **b,** For each phenotype and disorder, we calculated the overlap between the phenotype's DEPICT-prioritized genes and genes believed to harbor *de novo* mutations causing the disorder. The bars correspond to odds ratios. *EduYears*, years of education; BMI, body mass index; WHR, waist-to-hip ratio adjusted for BMI. **c,** DEPICT-prioritized genes in *EduYears*-associated loci exhibit substantial overlap with genes previously reported to harbor sites where mutations increase risk of intellectual disability and autism spectrum disorder (Okbay, Beauchamp, et al., 2016, Supplementary Table 4.6.1).

causal, and the genetic associations with EA are mediated by multiple intermediate phenotypes (Krapohl et al., 2014). To illustrate this last point, we studied mediation of the association between the all-SNPs polygenic score and *EduYears* in two of our cohorts. We found that cognitive performance can statistically account for 23-42% of the association ($P <$ 0.001) and the personality trait "openness to experience" for approximately 7% ($P < 0.001$; Okbay, Beauchamp, et al., 2016, Supplementary Information section 6).

It would also be a mistake to infer from our findings that the genetic effects operate independently of environmental factors. Indeed, a recent meta-analysis of twin studies found that genetic influences on EA are heterogeneous across countries and birth cohorts (Branigan, McCallum, & Freese, 2013). We conducted exploratory analyses in the Swedish Twin Registry to illustrate how environmental factors may amplify or dampen the impact of genetic influences (Okbay, Beauchamp, et al., 2016, Supplementary Information section 7). We found that the predictive power of the all-SNPs polygenic score is heterogeneous by birth cohort, with smaller explanatory power in younger cohorts (Figure 4.13**)**. See also Supplementary Information section 7.4 of Okbay, Beauchamp, et al. (2016) for discussion of the contrast between these results and findings from a seminal twin study that estimated EA heritability by birth cohort (Heath et al., 1985).

Figure 4.13. The predictive power of a polygenic score (PGS) varies in Sweden by birth cohort.



Note: Five-year rolling regressions of years of education on the PGS (left axis in all four panels), share of individuals not affected by the comprehensive school reform (**a,** right axis), and average distance to nearest junior high school (**b,** right axis), nearest high school (**c,** right axis) and nearest college/university (**d,** right axis). The shaded area displays the 95% confidence intervals for the PGS effect.

Table 4.1. Selected candidate genes implicated by bioinformatics analyses.

Gene	SNP	Syndromic	Score	Top-ranking gene sets
TBR1	rs4500960	ID, ASD	6	Developmental biology, decreased brain size, abnormal cerebral cortex morphology
MEF2C	rs7277187	ID, ASD	5	ErbB signaling pathway, abnormal sternum ossification, regulation of muscle cell differentiation
ZSWIM6	rs61160187	–	5	Transcription factor binding, negative regulation of signal transduction, PI3K events in ErbB4 signaling
BCL11A	rs2457660	ASD	5	Dendritic spine organization, abnormal hippocampal mossy fiber morphology, SWI/SNF-type complex
CELSR3	rs11712056	SCZ	5	Dendrite morphogenesis, dendrite development, abnormal hippocampal mossy fiber morphology
MAPT	rs192818565	ID	5	Dendrite morphogenesis, abnormal hippocampal mossy fiber morphology, abnormal axon guidance
SBNO1	rs7306755	SCZ	5	Protein serine/threonine phosphatase complex
NBAS	rs12987662	–	5	–
NBEA	rs9544418	SCZ	4	Developmental biology, signaling by Robo receptor, dendritic shaft
SMARCA2	rs1871109	ID	4	–
MAP4	rs11712056	ASD	4	Developmental biology, signaling by Robo receptor, SWI-SNF-type complex
LINC00461	rs10061788	–	4	Decreased brain size, abnormal cerebral cortex morphology, abnormal hippocampal mossy fiber morphology
POU3F2	rs9320913	–	4	Dendrite morphogenesis, developmental biology, decreased brain size
RAD54L2	rs11712056	SCZ	4	Decreased brain size, SWI/SNF-type complex, nBAF complex
PLK2	rs2964197	–	4	Negative regulation of signal transduction, PI3K events in ErbB4 signaling

Note: Fifteen candidate genes implicated most consistently across various analyses. To assemble this list, each gene in a DEPICT-defined locus (Okbay, Beauchamp, et al., 2016, Supplementary Information section 4.5) was assigned a score equal to the number of criteria it satisfies out of ten (see Okbay, Beauchamp, et al., 2016, Supplementary Table 4.1 for details). The DEPICT prioritization *P*-value was used as the tiebreaker. "SNP": the SNP in the gene's locus with the lowest *P*-value in the *EduYears* meta-analysis. "Syndromic": which, if any, of three neuropsychiatric disorders have been linked to *de novo* mutations in the gene (Okbay, Beauchamp, et al., 2016, Supplementary Information section 4.6). "Top-ranking gene sets": DEPICT reconstituted gene sets of which the gene is a top-20 member (Okbay, Beauchamp, et al., 2016, Supplementary Table 4.5.1). The three most significant gene sets are shown if more than three are available. ID, intellectual disability; ASD, autism spectrum disorder; SCZ, schizophrenia.

4.2 Supplementary Methods

4.2.1 Study overview

We examined two phenotypes: a continuous variable measuring the number of years of schooling completed (*EduYears*, $N = 293{,}723$) and an indicator variable for college completion (*College*, $N = 280{,}007$). All analyses were performed at the cohort level according to a pre-specified and publicly archived analysis plan. Summary statistics provided by cohorts were uploaded to a central server and subsequently meta-analyzed. The lead PI of each cohort affirmed that the results contributed to the study were based on analyses approved by the local Research Ethics Committee and/or Institutional Review Board responsible for overseeing research. All participants provided written informed consent. Table B1 provides basic information about the participating cohorts. For additional details, see Supplementary Table 1.1 in Okbay, Beauchamp, et al. (2016).

Our Analysis Plan was preregistered at https://osf.io/paj9m/. With one exception, the analyses reported here follow the original plan. The exception is that the original plan treated *EduYears* and *College* symmetrically whereas throughout the manuscript, we treat *EduYears* as the primary variable and de-emphasize *College*. After circulation of the Analysis Plan to our cohorts, a paper was posted on *bioRxiv* showing that the genetic correlation between the two measures is very high, with the point estimate suggesting a perfect genetic correlation (Bulik-Sullivan, Finucane, et al., 2015). Previously, we had considered as plausible the possibility that *College* would have better power for detecting associations at the upper end of the distribution of *EduYears*. However, since *College* is constructed by dichotomizing *EduYears*, the very high genetic correlation suggests that the *College* phenotype is for all intents and purposes merely a coarsening of the *EduYears* phenotype.

Hence, we reasoned in light of this new evidence that attempts to detect associations with *EduYears* are likely to be better powered, regardless of whether or not the effect is stronger at the upper end of the distribution of *EduYears*. To eliminate (or at least minimize) concerns about data mining, we made the decision to promote *EduYears* to the primary phenotype before quality-control work had begun in earnest. After the decision to make *EduYears* the primary phenotype was made, we performed the quality control sequentially. In the first stage, we completed the quality control of the *EduYears* variable, froze the meta-analysis, and announced to all analysts responsible for follow-up work that their work would be based on the pooled-sex *EduYears* results. We subsequently turned to the *College* quality control.

4.2.2 Phenotype definition

Subjects in our cohorts are heterogeneous in terms of birth cohort and country of birth, and hence they were educated under a diverse set of educational systems. Moreover, the survey questions that were used to evaluate subjects' educational qualifications are not identical across cohorts. To maximize comparability across samples, we use as a standard the 1997 International Standard Classification of Education (ISCED) of the United Nations Educational, Scientific and Cultural Organization (UNESCO, 2006). Specifically, we map each major educational qualification that it is possible to attain in a specific country into one of seven harmonized ISCED categories. To construct our primary outcome variable, EduYears, we impute a years-of-education equivalent for each ISCED category using the mapping shown in Table 4.2. Following Rietveld, Medland, et al. (2013) we also analyzed the binary outcome, College, which takes the value 1 for subjects with an ISCED level equal to 5 or more (and 0 otherwise).

The study-specific phenotype distributions are shown in Table B1. For the exact phenotype measures, see Supplementary Table 1.3 of Okbay, Beauchamp, et al. (2016). With the exceptions of STR and HBCS, whose variables are derived from official register data on educational attainment, the studies relied on surveys to measure educational attainment.

Table 4.2 Mapping from ISCED Level to *EduYears* and *College*.

ISCED Level	Definition	US years of schooling (*EduYears*)	*College*
0	Pre-primary education	1	0
1	Primary education or first stage of basic education	7	0
2	Lower secondary or second stage of basic education	10	0
3	(Upper) secondary education	13	0
4	Post-secondary non-tertiary education	15	0
5	First stage of tertiary education (not leading directly to an advanced research qualification)	19	1
6	Second stage of tertiary education (leading to an advanced research qualification, e.g. a Ph.D.)	22	1

Notes: In some samples the educational attainment measures did not differentiate between levels 5 and 6. In these cases everyone with a tertiary education was coded as ISCED 5, and 20 years of schooling was imputed instead of 19.

4.2.3 Genotyping and imputation

Genotyping was performed using a range of common, commercially available genotyping arrays. Study analysts were encouraged to impute markers from all 23 chromosomes using the 1000 Genomes project (1kGp) March 2012 version 3 release (hereafter, 1000G) as reference panel, the most recently released haplotype version available when the Analysis Plan was circulated. Given the well-known challenges in imputing markers on the X chromosome, cohorts who could only supply results for autosomal markers were also invited to participate. Supplementary Table 1.4 of Okbay, Beauchamp, et al. (2016) provides study-specific details on genotyping platform, pre-imputation quality-control filters applied to the genotype data, subject-level exclusion criteria, imputation software used, the reference sample used for imputation (haplotype release date and whether imputation was done using European-ancestry sample or the full 1000G-sample) and whether the cohort supplied us with results from the X chromosome. As the table shows, the overwhelming majority of cohorts followed the recommendation to impute their data against the March 2012 version 3 release of the 1000G panel. The exceptions are (i) SardiNIA, which used its own reference panel constructed from sequencing data available for about 2000 individuals in their sample (Pistis et al., 2015); (ii) Rush, whose imputation was based on the December 2010 haplotype release; and (iii) a handful of cohorts who began imputation relatively late and used more recent releases that were not available at the time that the Analysis Plan was written and circulated.

4.2.4 Association analyses

A. *EDUYEARS* ANALYSES

Cohorts were asked to estimate this regression equation for each measured SNP (we drop the SNP subscript j here to avoid notational clutter):

$$EduYears = \beta_0 + \beta_1 SNP + \mathbf{PC}\,\boldsymbol{\gamma} + \mathbf{B}\,\boldsymbol{\alpha} + \mathbf{X}\,\boldsymbol{\theta} + \epsilon, \tag{4.1}$$

where *SNP* is the allele dose of the SNP; **PC** is a vector of the first ten principal components of the variance-covariance matrix of the genotypic data, estimated after the removal of genetic outliers; **B** is a vector of standardized controls, including a third-order polynomial in age, an indicator for being female, and their interactions; and **X** is a vector of study-specific controls. Specifically, in **X**, study analysts were encouraged to include dummy variables for major events such as wars or policy changes that may have affected access to education in their specific sample. Mixed-sex cohorts were additionally asked to upload separate regression results for men and women.

B. *COLLEGE* ANALYSES

The *College* specification is analogous to the *EduYears* specification. Cohorts uploaded either coefficient estimates from a linear probability model or from a logistic regression model.

Linear Regression. The linear model can be written as

$$College = \beta_{0,\text{lin}} + \beta_{1,\text{lin}}\, SNP + \mathbf{PC}\, \boldsymbol{\gamma}_{\text{lin}} + \mathbf{B}\, \boldsymbol{\alpha}_{\text{lin}} + \mathbf{X}\, \boldsymbol{\theta}_{\text{lin}} + \epsilon_{\text{lin}}, \tag{4.2}$$

where $College$ is an indicator variable equal to one for individuals who completed college, the other variables are defined as above, and the subscript "lin" indicates that the variables correspond to the linear probability model. The parameter $\beta_{1,\text{lin}}$ is the average change in the fraction of subjects whose value of $College$ is equal to one associated with being endowed with one more copy of the reference allele, after linear adjustment for the covariates.

Logistic Regression. Most participating cohorts uploaded coefficient estimates from the logistic regression model,

$$P(College = 1 | SNP, \mathbf{PC}, \boldsymbol{\alpha}, \mathbf{X}) = \frac{1}{1 + e^{-(\beta_{0,\text{log}} + \beta_{1,\text{log}}\, SNP + \mathbf{PC}\, \boldsymbol{\gamma}_{\mathbf{log}} + \mathbf{B}\, \boldsymbol{\alpha}_{\mathbf{log}} + \mathbf{X}\, \boldsymbol{\theta}_{\mathbf{log}})}}, \tag{4.3}$$

where the subscript "log" is used to label coefficients from the logistic model. In this model, the parameter $\beta_{1,\text{log}}$ can be interpreted as follows: controlling for the covariates, the odds of having completed college is increased by a factor of $e^{\beta_{1,\text{log}}}$ for each increase of one copy of the reference allele.

C. SAMPLE SELECTION CRITERIA

Only individuals satisfying the following criteria were eligible for inclusion in the estimation sample:

a. Educational attainment was measured when the subject was 30 years of age or older.
b. The subject passed the cohort's standard quality controls, which typically include removal of subjects who are genetic outliers (to mitigate stratification concerns) and subjects with poor genotyping rates.

c. The subject is of European ancestry, and the subject's mother tongue is the same as the main language in the country of the cohort.
d. All relevant covariates are available for the subject.

D. STUDY-SPECIFIC DETAILS

The *EduYears* analyses are based on summary statistics from all 64 samples listed in Supplementary Table 1.1 of Okbay, Beauchamp, et al. (2016)**.** Of the 64 samples, whose combined sample size is *N=293,723*, 5 were from single-sex cohorts, and 59 contained pooled results from mixed-sex cohorts (who additionally uploaded separate results for men and women).

The *College* analyses were based on results from 52 of the 64 *EduYears* samples. The combined sample size of these 52 cohorts is *N=280,007*. One small cohort, LBC1921, is excluded because it did not upload *College* results. The cohort analyst determined that the low fraction of college-educated individuals (1-5%) and the small sample would not yield reliable estimates of the standard errors. Indeed, because analytical standard errors may not be reliably estimated in small samples when the dependent variable is rare, we restrict our final analysis to cohorts with a combined sample size (N_{tot}) of at least 500 and at least 100 cases (N_{cases}). We also drop one family-based cohort (ERF) and one isolate (ORCADES) because the estimated standard errors of the logistic regression coefficients did not account for the sample relatedness (in both cases, the standard errors from their *EduYears* did account for relatedness). Column 3 of Supplementary Table 1.5 in Okbay, Beauchamp, et al. (2016) reports if a given sample was included in the *College* analyses and also explains why, in two samples, the *EduYears* sample size is not identical to the *College* sample size.

Column 4 reports whether the cohorts omitted any of the basic control variables recommended in the Analysis Plan in their specification. For example, some cohorts dropped higher-order polynomials in birth year because collinearity was causing problems in model estimation. Column 5 lists extra controls included by the cohorts in the vector **X**, such as controls for cohort-specific events that may have impacted the education system in the cohort.

Several cohorts contain samples with related subjects. The Analysis Plan encouraged cohorts that include related subjects to estimate mixed linear models (MLMs) (Kang et al., 2010; Yang, Zaitlen, Goddard, Visscher, & Price, 2014). To facilitate their implementation, the Analysis Plan contained a supplement with sample code for MLM estimation written for the

software GCTA (Yang, Lee, et al., 2011). Conceptually, the estimation of MLM models involves two steps: (i) the genome-wide data are used to estimate the degree of genetic similarity between each pair of individuals in the sample, and (ii) unlike in standard regression where the covariance of the error term (in an educational attainment regression) between any two individuals is assumed to be zero, the covariance is fitted as an increasing linear function of the individuals' genetic similarity. In other words, to the extent that two individuals are more recently descended from a common ancestor (as very accurately measured by overall genetic similarity)—and thus are more likely to be similar on unobserved environmental factors—these individuals are treated as correlated observations.

Many cohorts that include related subjects have developed strategies for ensuring that the standard errors correctly account for relatedness. Column 6 of Supplementary Table 1.5 (Okbay, Beauchamp, et al., 2016) reports whether the estimated standard errors were adjusted for family relatedness and provides information about the adjustment used. The details vary by software. For example, QIMR estimated a model implemented in the software Merlin Offline (W.-M. Chen & Abecasis, 2007), in which the variance-covariance matrix of the phenotypes of members of the same family is assumed to have a particular structure according to which resemblance between relatives is induced by the additive effects of their shared genes. Some cohorts made no adjustment for non-independence but instead sought to restrict the estimation samples to conventionally unrelated individuals. For example, 23andMe restrict their estimation sample to conventionally unrelated individuals by ensuring that no pair of participants in the final estimation sample share more than 700 centimorgans of their genome identical-by-descent (Eriksson et al., 2010).

4.2.5 Quality control

We closely followed the quality-control protocol used in the GIANT consortium's most recent study of height (Wood et al., 2014). The protocol, implemented by the software *EasyQC*, is described in detail by Winkler et al. (2014). *EasyQC* calculates a range of test statistics that are valuable for identifying possible sources of error in uploaded summary statistics. It also outputs a harmonized set of graphs, described below, that can be visually inspected to identify problems with data or analysis. Below, we describe the quality-control filters that were applied to the uploaded files. We then describe a subset of several additional diagnostic tests that the files were required to pass before being included in the meta-analysis.

From the uploaded files, we filtered out the following markers:

1. If the data were imputed against the September or December 2013 releases of the 1000 Genomes Phase 1 haplotypes provided by the software IMPUTE2, we drop the 730+199 SNPs whose strands were incorrectly aligned in these releases.[9]

2. We drop a marker if neither an effect allele nor other allele is supplied. We also drop a marker if any of the following variables are missing: effect allele frequency, beta, standard error, *P*-value, imputation accuracy (if the marker is imputed), or the imputed/genotyped indicator. For variables that can only take on some restricted range of values, we drop the marker if the value of the variable falls outside the permissible range. For example, *P*-values have to lie within the unit interval, and binary variables can only take on a value of 0 or 1.[10]

3. The analytical standard errors computed by genetic-association software packages are known to be unreliable in small samples, especially for low-frequency variants (Winkler et al., 2014). To guard against spurious associations with low-frequency markers in small samples, we dropped a marker from a cohort if its minor allele count (MAC) was below 25. We also drop markers that explain more than 5% of variance in *EduYears*, two order of magnitudes larger than the effects that should

[9] The announcement is available on https://mathgen.stats.ox.ac.uk/impute/impute_v2.html#whats_new

[10] Four *College* cohorts reported *P*-values from likelihood-ratio (LR) tests in which the test-statistic is defined as $\chi^2_{OBS} = -2ln\left(\frac{L_0}{L_1}\right)$, where L_0 is the log-likelihood of the full model and L_1 is the log-likelihood of a restricted model in which the coefficient for SNP *j* restricted to equal 0. Under the null hypothesis that $\beta_j = 0$, the statistic is approximately distributed $\chi^2(1)$. Remaining cohorts conducted hypotheses-testing using conventional Wald tests in which the *P*-value is derived from the fact that the distribution of the test-statistic $Z_{OBS} = \frac{\hat{\beta}}{\widehat{se}(\hat{\beta})}$ is approximately $N(0,1)$. The two tests are asymptotically equivalent, but may deliver different answers in finite samples. We err on the side of caution by dropping SNPs from the LR-test cohorts that fail to satisfy the inequality $\left|\frac{\chi^2_{OBS}}{Z^2_{OBS}} - 1\right| < 0.1$.

be considered plausible based on the findings in Rietveld et al. (2013).[11, 12] . For each SNP j, we approximate the variance explained by

$$R_j^2 \approx \frac{2\, MAF_j \left(1 - MAF_j\right) \hat{\beta}_j^2}{\hat{\sigma}_y^2} \tag{4.4}$$

4. We drop markers with low imputation-quality metrics. The exact definition of the quality metrics vary by software. In cohorts that supplied us with the Rsq variable generated by the imputation software MaCH (Y. Li et al., 2010), we use a threshold of 0.6. In cohorts that supplied us with the INFO variable generated by the imputation software IMPUTE2 (Marchini et al., 2007), we used a threshold of 0.7. These thresholds are stricter than those that have typically been used in previous studies predating the availability of the 1000G reference panel. We used the stricter thresholds because evaluations have shown that the conventional thresholds (in the range 0.3-0.4) do not filter out all badly imputed rare variants in 1000G data (Pistis et al., 2015). The MACH-Rsq and IMPUTE2-INFO thresholds we use were proposed by Pistis et al. (2015) for variants with minor allele frequency below 1%. For transparency, and to err on the side of conservatism, we apply these thresholds to all mark-

[11] Standard practice is to drop SNPs with estimated betas whose absolute value exceeds some threshold considered to represent an implausibly large effect (Winkler et al., 2014). Rather than select a single β threshold, we decided to apply a more flexible filter that is not sensitive to the measurement scale of the dependent variable and allows the β threshold to vary by allele frequency. The latter is desirable because what constitutes a plausible effect size depends on the allele frequency. To illustrate using the example of height, an effect of 15 cm per allele need not indicate a quality-control problem for very low-frequency variants; in fact rare polymorphisms with effects of that magnitude have been identified (Visscher, Yang, & Goddard, 2010). However, for common variants, effects of that magnitude are impossible (the implied R^2 would exceed 100% for any realistic value of the sample variance of height). To verify that the number of SNPs dropped due to the R^2 filter is not alarmingly high, we reran the filtering of the cohort-level *EduYears* results files with the R^2 filter applied last. We found that the R^2 filter, after applying standard quality-control filters, does not remove *any* SNPs in any of the 44 largest cohorts (combined $N = 278{,}528$). The filter removes a small number of SNPs in ten of the remaining 20 cohorts: LBC1936 (9 SNPs dropped), INGICARL (64), THISEAS (225), H2000 Controls (16), Hypergenes (1), H2000 Cases (2), MoBa (566), OGP (2300), COPSAC2000 (8561). In a logistic regression model, the estimated proportion of variance explained by SNP j is defined as $2\, MAF_j\, (1 - MAF_j)\, \hat{\beta}_{j,\log}^2$.

[12] For cohorts that report marginal effects from linear probability models, it is necessary to transform the estimated linear-probability coefficient $\hat{\beta}_{j,\text{lin}}$ into a quantity that is comparable to $\hat{\beta}_{j,\log}$ as estimated from a logistic model. We use the approximation $\hat{\beta}_{j,\text{lin}} \approx \hat{f}(1-\hat{f})\hat{\beta}_{j,\log}$, where $\hat{f}$ is the fraction of the sample with a college degree. The approximation is accurate for $\hat{\beta}_{j,\text{lin}}$ small. Hence, we drop marker j if $2\, MAF_j \left(1 - MAF_j\right) \hat{\beta}_{j,\log}^2 > 0.05$ (logistic model) or $2\, MAF_j \left(1 - MAF_j\right) \left(\frac{\hat{\beta}_{j,\text{lin}}}{\hat{f}(1-\hat{f})}\right)^2 > 0.05$ (linear probability model).

ers. Finally, for cohorts that supplied us with PLINK's imputation-accuracy measure (info), we follow Winkler et al.'s (2014) recommendation of using a threshold of 0.8.

5. We drop non-autosomal SNPs, indels, or structural variants. We drop the indels and structural variants because they are often poorly imputed and hence difficult to align, and we drop X-chromosome markers because they are analyzed separately.

6. If a cohort supplied us with an rs number, we use the reference file provided by EasyQC[13] to identify the marker's chromosome-position ID (ChrPosID). If a cohort only supplied information about the genetic position (chromosome and base pair) of the SNP, we generate a chromosome-position ID (ChrPosID) by horizontally concatenating the chromosome number and the base pair position. We subsequently drop duplicated markers based on ChrPosID, or markers whose ChrPosID's are unavailable in the 1000 Genomes phase 1 European panel (The 1000 Genomes Project Consortium, 2012) that we use to identify potential strand problems. In this step, SNPs that cannot be successfully aligned due to allele mismatch with the reference panel are also removed.

Having applied filters 1-6 to cohort-level summary statistics, we examined how many SNPs were dropped in each filtering step. Whenever an unusual number of markers were being dropped, we flagged the cohort as potentially having an error in the uploaded results file. The issue was discussed with the cohort-level analyst and resolved through a new QC iteration.

A. EASYQC DIAGNOSTICS

We conducted several additional diagnostic checks after applying the filters described previously. Below, we describe the four most important of these. Winkler et al. (2014) contains a comprehensive discussion of how these four diagnostic tests are useful for identifying a number of potential problems and their possible underlying causes.

Diagnostic Test #1. Allele Frequency Plots (AF Plots)

[13] http://homepages.uni-regensburg.de/~wit59712/easyqc/1000g/rsmid_map.1000G_ALL_p1v3.merged_mach_impute.v1.txt.gz, accessed on 22 June 2015.

We looked for errors in allele frequencies and strand orientations by visually inspecting a plot of the sample allele frequency of filtered SNPs against the frequency in the 1000 Genomes phase 1 version 3 European panel (The 1000 Genomes Project Consortium, 2012).

Diagnostic Test #2. P-value vs Z-score Plots (PZ Plots)
We verified that the reported *P*-values are consistent with the *P*-values implied by the coefficient estimates and standard errors in the results file.

Diagnostic Test #3. Quantile-Quantile Plots (QQ Plots)
We visually inspected the cohort-level QQ plots to look for evidence of unaccounted-for stratification.

Diagnostic Test #4. Predicted vs Reported Standard Error Plots (PRS Plots)
We investigated if the standard errors reported in the *EduYears* files are roughly consistent with the reported sample size, allele frequency, and phenotype distribution. Winkler et al. (2014) propose a similar diagnostic (the *SE-N Plots*), which is based on following approximation to the standard error of a coefficient estimated by OLS

$$(s.e.)_j \approx \frac{\hat{\sigma}_Y}{\sqrt{N}} \cdot \frac{1}{\sqrt{2\,MAF_j\left(1 - MAF_j\right)}}, \tag{4.5}$$

where $\hat{\sigma}_Y$ is the standard deviation of the dependent variable, MAF_j is the minor allele frequency of SNP *j*, and *N* is the sample size. We used Equation (4.5) to generate a predicted standard error for 50,000 randomly sampled SNPs. We then plotted these predicted standard errors against the reported standard errors. Since the assumptions underlying —independent observations, no other controls are included in the regression, and no estimation error that is due to imputation uncertainty—do not hold exactly, the main purpose of the plot is detect substantial discrepancies between the reported and actual size of the estimation sample or errors in phenotype transformation. Specifically, we visually inspected the plot to ensure that the standard errors were of approximately the predicted magnitude and that there were no major outliers.

When examining the standard errors in the *College* files, we proceeded similarly, albeit using an analytical approximation for the standard error of the coefficient from a logistic regression when appropriate. The approximation is

$$(s.e.)_j \approx \frac{1}{\sqrt{N}} \cdot \frac{1}{\sqrt{2\,\hat{f}(1-\hat{f})\,MAF_j\,(1-MAF_j)}}, \tag{4.6}$$

where $\hat{f}$ denotes the fraction of college graduates in the sample.

B. SNP EXCLUSIONS

Our meta-analyses are based on files that have been filtered according to the six QC-filter steps described above and that have passed the four diagnostic tests. Table B2 shows, for each of the cohorts contributing to our pooled *EduYears* analysis, the number of SNPs in the originally uploaded results files, the number of SNP exclusions in each of the six steps, and the number of SNPs remaining after the full set of QC steps were applied. Supplementary Table 1.7 in Okbay, Beauchamp, et al. (2016) shows the analogous numbers for *College*. All subsequent analyses are based on the set of SNPs remaining after these exclusions. All subsequent analyses are based on the set of SNPs remaining after these exclusions.

C. GENOMIC CONTROL FACTORS

The last column of Table B2 shows the genomic control factor, λ_{GC} (Devlin & Roeder, 1999), from each sample in the *EduYears* analyses. With the exception of deCODE, whose standard protocol is to apply genomic control to the standard errors before uploading results, the reported genomic control factors are all computed using untransformed standard errors. For *EduYears*, the unweighted average λ_{GC} is 1.02, with a range from 0.95-1.15 and a median of 1.01. For *College*, the corresponding numbers are 1.01, 0.93-1.13, and 1.01. Table B2 also reports the inflation factor used by deCODE to inflate their standard errors prior to uploading the results. See Supplementary Table 1.7 (Okbay, Beauchamp, et al., 2016) shows the analogous numbers for *College.*

D. ADDITIONAL DIAGNOSTICS

Here, we summarize the results from three additional diagnostic tests of the cleaned results files.

Cohort-Level F_{st} Statistics

F_{st} is a frequently used measure of between-population genetic differentiation. We estimated F_{st} using summary data on cohort-level allele frequencies using an approach described by Weir (1990). For each cohort, we calculated the F_{st} relative to the European-ancestry individuals in the 1000G sample (The 1000 Genomes Project Consortium, 2012). We sampled 30,000 quasi-independent markers with minor allele frequencies greater than 0.05 in the European-ancestry subjects. We computed the F_{st} of each SNP and averaged over the 30,000

markers to get an overall measure of F_{st} in the cohort. Because our reference sample is European, an unusually high level of F_{st} may be an indication that a cohort inadvertently failed to remove genetic outliers or a sign of genotyping or imputation problems.

In Weir (1990), the equation for estimating F_{st} is

$$F_{st} = \frac{\frac{r}{(r-1)\sum_{i=1}^{r} n_i}\left[\sum_{i=1}^{r} n_i\,(p_i - \bar{p})^2\right]}{\bar{p}(1-\bar{p})}, \tag{4.7}$$

where r is the number of populations in the sample, n_i is the number of individuals in the sample from population i, p_i is the sample minor allele frequency of the SNP in the sample in population i, and $\bar{p}$ is the weighted average frequency across populations in the sample. Since in our case $r = 2$, Equation (4.7) specializes to

$$F_{st} = \frac{\frac{2}{N}\left[n_1(p_1 - \bar{p})^2 + n_2(p_2 - \bar{p})^2\right]}{\bar{p}(1-\bar{p})}, \tag{4.8}$$

where $N = n_1 + n_2$, and $\bar{p} = (n_1/N)\,p_1 + (n_2/N)\,p_2$ is the mean allele frequency. For most EA cohorts, the average F_{st} value was below 0.004, which agrees well with previous reports that F_{st} is around 0.004 between European nations (Novembre et al., 2008). The mean F_{st} value across our 64 samples was 0.002 (*SD* = 0.003). The largest F_{st}, a value of 0.02, was observed for the cohort OGP-Talana. It is known that the central-eastern Sardinia region, Ogliastra, has been secluded from the surrounding regions for most of its history. Such isolation is expected to generate an unusually high F_{st} (Pistis et al., 2009). Although the possibility of technical problems for genotype calling or imputation cannot be ruled out, the observed F_{st} values indicate that the quality of the reported genotype data is consistent with observed differences in sample allele frequencies between populations, and there is no evidence that cohorts are derived from non-European ancestry.

λ_{meta} *Test for Genetic Effects for Each Pair of Cohorts*

We computed a second diagnostic summary statistic, λ_{meta}, which can help identify a number of problems, including unknown sample overlap between cohorts (which would violate the assumption of independence underlying the meta-analysis). Given a pair of cohorts and a locus, λ_{meta} is defined as

$$\lambda_{\text{meta}} \equiv \frac{(b_1 - b_2)^2}{\sigma_{b_1}^2 + \sigma_{b_2}^2}, \tag{4.9}$$

where b_i and $\sigma_{b_i}^2$ are the reported allelic effect and sampling variance of the number of minor alleles in cohort $i \in \{1, 2\}$. If the two cohorts are independent and if the genetic correlation of the phenotype across the two cohorts is 1, then the expected value of λ_{meta} across loci is 1. If the cohorts overlap substantially, then the reported effect sizes are too similar, and therefore the numerator is smaller than the denominator, leading to $\lambda_{\text{meta}} < 1$. Conversely, if there is too much heterogeneity in the estimated effect sizes for a pair of cohorts, either because the phenotypes are not the same or because results are not reported for the same allele, then $\lambda_{\text{meta}} > 1$. Hence this statistic is a useful QC metric to detect deviations in the reported summary statistics for a pair of cohorts from the assumed null hypothesis of independence and homogeneity. In our data, the average value of λ_{meta} is only slightly greater than 1 (mean = 1.006, *SD* = 0.023), suggesting no overall deviation from expectation.

Tests of Allele Misalignment

We supplemented our visual inspection of the allele frequency plots with two additional tests of allele misalignment. First, we generated a pruned set of SNPs from the deCODE summary statistics whose *P*-value for the test of association with *EduYears* was smaller than 0.01. For each of our other samples, we calculated the frequency with which the estimated effects had the same sign as in the deCODE results. In all but one of the cohorts with a sample size above 5,000, the fraction of coefficient signs that aligned with deCODE exceeded 50% (see Table B3).

Second, we used LD Score regression (Bulik-Sullivan, Loh, et al., 2015) to estimate the genetic correlation between *EduYears* in each of our samples and *EduYears* in deCODE. The estimator often failed to converge, especially for smaller cohorts, but of the 21 estimates obtained, all but one are in the predicted (positive) direction. The negative estimated genetic correlation is for the cohort Rush-MAP: it is -0.29 but has a large standard error (s.e. = 0.70). Given that Rush-MAP passes all other diagnostics, it is likely that the negative estimate is a chance outcome due to sampling variability. The estimated genetic correlations are shown in Table B3.

4.2.6 Meta-analysis

We used the software program METAL (Willer et al., 2010) to conduct sample-size-weighted meta-analysis of all SNPs that passed the quality-control thresholds. Prior to running the

meta-analyses, we applied a single correction for genomic control to the cohort-level summary statistics. A total of 9,256,490 autosomal SNPs were meta-analyzed using data in the 64 filtered *EduYears* files, and 9,280,749 autosomal SNPs were meta-analyzed using data in the 52 filtered *College* files.[14]

A. *EDUYEARS* (N = 293,723)

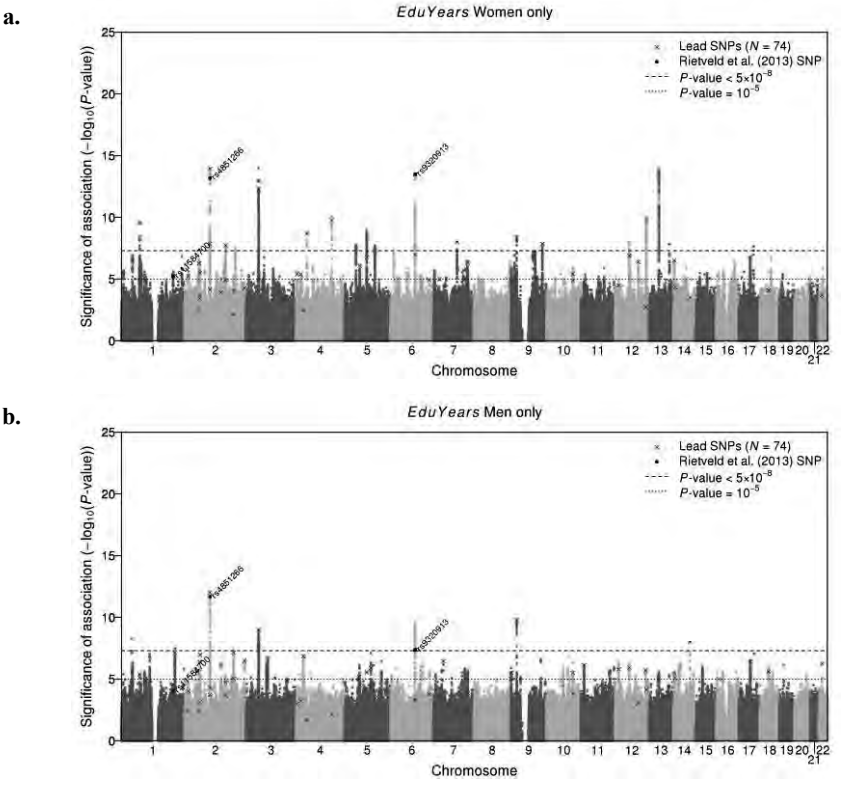
We used sample-size-weighted meta-analysis in our primary analyses because the method is more robust to errors in variable scaling at the cohort level. As a robustness check, we also conducted a secondary meta-analysis of *EduYears* with inverse-variance weighting. Consistent with the results from our many diagnostic tests, the results were highly similar, suggesting that the scale of measurement was successfully harmonized across cohorts. The correlation between the two sets of *P*-values obtained using the two methods was 0.91. We conducted sample-size-weighted sex-stratified meta-analyses of *EduYears* as another robustness check to see whether the results differ for men and women. Figure 4.1 and Figure 4.14a-b provide Manhattan plots for the pooled-sex, women-only, and men-only analyses of *EduYears*.

To select independent genome-wide significant SNPs from our primary *EduYears* results, we first grouped the GWAS results into "clumps" as follows. The SNP with the smallest *P*-value was chosen as the lead SNP in its clump. All SNPs less than 500 kb away from this lead SNP, in LD with it to the extent $r^2 > 0.1$, and with an association *P*-value smaller than 10^{-6} were assigned to this clump. The next clump was greedily formed around the SNP with the next smallest *P*-value not already assigned to the first clump. This process was iterated until no SNPs remained with *P*-value $< 5\times10^{-8}$. The end result was 77 approximately independent clumps, each centered around, and represented by, a genome-wide significant SNP.

Next, we checked the long-range LD between these 77 approximately independent SNPs without imposing any restriction on distance (except for residing on the same chromosome). If the r^2 between two SNPs is greater than 0.5, we merged the corresponding clumps and assigned the SNP with smaller *P*-value to represent that locus. This step resulted in 74 approximately independent loci, each represented by a genome-wide significant SNP. The PLINK tool version 1.9 (Chang et al., 2015) and 1000 Genomes Project phase 1 genotyping data (Abecasis et al., 2012) (from 268 individuals with European ancestry) was used to perform clumping and calculating r^2 between a pair of SNPs. Table B5 shows the *EduYears*

[14] SNPs with a sample size less than 100,000 (3,074,494 SNPs in *EduYears*, and 3,161,722 SNPs in *College*) were excluded from the meta-analyses.

Figure 4.14. Manhattan plots from the sex-stratified analyses of *EduYears*.



Note: In each plot, the *x*-axis is chromosomal position, and the *y*-axis is the *P*-value on a $-\log_{10}$ scale. The black line shows the genome-wide significance level (5×10^{-8}). The red x's are the approximately 74 independent genome-wide significant associations ("lead SNPs") from the *EduYears* pooled results. The black dots labeled with rs numbers are the 3 Rietveld, Medland, et al. (2013) SNPs.

pooled-sex and sex-stratified association results for these 74 approximately-independent genome-wide significant SNPs.

As in the earlier GWAS of EA (Rietveld, Medland, et al., 2013) and other large GWAS of polygenic traits (Locke et al., 2015; Ripke et al., 2014; Wood et al., 2014), the Q-Q plot of the meta-analysis (Figure 4.2) exhibits inflation ($\lambda_{GC} = 1.28$), consistent with a polygenic architecture. Forest plots of the *EduYears*-associated SNPs (not shown) provide little evidence that the estimated effects are driven by a small number of outlier cohorts, cohorts from a given region, or by one of the sexes (see Table B5 for the heterogeneity *P*-values for the lead SNPs).

To help gauge the magnitude of the estimated effects, we used a well-known approximation to compute unstandardized regression coefficients from the METAL output obtained from the sample-size-weighted meta-analysis:

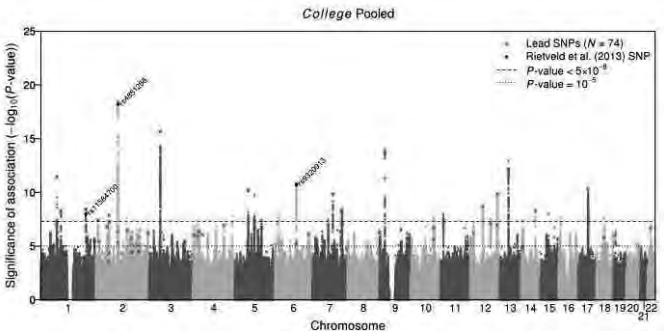
$$\hat{\beta}_j \approx z_j \frac{\hat{\sigma}_Y}{\sqrt{2N_j\, MAF_j\, (1 - MAF_j)}} \tag{4.10}$$

for SNP *j* with minor allele frequency MAF_j, sample size N_j, METAL *z*-statistic z_j, and standard deviation of the phenotype $\hat{\sigma}_Y$. For a derivation, see the SOM in Rietveld, Medland, et al. (2013). Figure 4.3a shows effects in standard-deviation units of the SNP with lowest *P*-value in each of the 74 loci, ordered from largest to smallest. As a benchmark for the magnitudes, the figure also shows corresponding estimates for the three phenotypes studied by the GIANT consortium in similarly large samples to ours (hereafter, the "GIANT phenotypes"): height (Wood et al., 2014), body mass index (Locke et al., 2015) (BMI), and waist-to-hip ratio adjusted for BMI (Shungin et al., 2015) (WHR). Consistent with the findings in Rietveld, Medland, et al. (2013), the *EduYears* estimates are in the range 0.014 to 0.048 standard deviations per allele (2.7 to 9.0 weeks of schooling), with incremental R^2 in the range 0.01% to 0.035%. The *EduYears* effects are smaller than those for height and BMI and more similar to those for WHR. The minor allele frequency of the SNP with the largest effect size in *SD*-units is 0.04.

B. *COLLEGE* (*N* = 280,007)

The Manhattan plot for the *College* analysis is shown in Figure 4.15. Overall, the results are similar to those from the *EduYears* analyses, but with higher *P*-values (consistent with the hypothesis that the *College* variable is a noisier measure of educational attainment than the *EduYears* variable). If we apply the procedure described previously to determine the number of approximately independent SNPs reaching genome-wide significance, we find 34 such SNPs (compared to 74 in the *EduYears* meta-analysis). Of these, 24 reach genome-wide significance in the *EduYears* analyses, and 27 are within 500kb distance and in LD with an *EduYears* lead SNP to the extent $r^2 > 0.1$. Supplementary Table 1.12 in Okbay, Beauchamp, et al. (2016) shows the association results for these 34 approximately independent genome-wide significant SNPs from the *College* meta-analysis and the *EduYears* lead SNPs in the same locus, if any.

Figure 4.15. Manhattan plot from the pooled analysis of the *College* phenotype



Note: The *x*-axis is chromosomal position, and the *y*-axis is the *P*-value on a $-\log_{10}$ scale. The black line shows the genome-wide significance level (5×10^{-8}). The red x's are the approximately 74 independent genome-wide significant associations ("lead SNPs") from the *EduYears* pooled results. The black dots labeled with rs numbers are the 3 Rietveld, Medland, et al. (2013) SNPs.

4.2.7 Within-sample replication

Following the suggestion of a referee, we attempted to replicate the genome-wide associations reported in our previous GWAS of EA (Rietveld, Medland, et al., 2013) in the new cohorts that were added to this study. Conversely, we also examined if the SNPs that reach genome-wide significance in a meta-analysis of the new cohorts replicate in the Rietveld, Medland, et al. (2013) cohorts.

A. COHORT OVERLAP WITH RIETVELD, MEDLAND ET AL. (2013)

The analyses of *EduYears* in Rietveld, Medland, et al. (2013) were based on a discovery sample of 101,069 individuals and a combined sample (discovery + replication) of 126,559 individuals. Some of the cohorts that contributed to the Rietveld, Medland, et al. (2013) study did *not* participate in the present study (N = 13,981). Overall, the combined sample size of the Rietveld, Medland, et al. (2013) cohorts that contributed to our study is N = 126,413 individuals. This number exceeds the difference between 126,559 and 13,981 because some of the original Rietveld, Medland, et al. (2013) cohorts completed additional genotyping since 2013, and were hence able to contribute larger samples to the current study.

B. METHODS IN WITHIN-SAMPLE REPLICATION ANALYSES

Rietveld, Medland, et al. (2013) reported three genome-wide significant SNPs in their discovery sample, all of which replicated in their replication sample. These three SNPs also yielded lower *P*-values in the "combined" (discovery + replication) sample. In a meta-analysis of the combined sample, four additional SNPs reached genome-wide significance. Of these, five were genome-wide significant in the *EduYears* analyses. The remaining two only reached genome-wide significance in the analyses of *College*, but both had *P*-values just shy of genome-wide significance in the combined-sample *EduYears* analysis. Given our decision to make *EduYears* the primary phenotype, and to facilitate comparisons of effect sizes, we attempt to replicate all of the seven original associations in our meta-analyses of the *EduYears* variable. To examine if the seven associations replicate in our new cohorts, we split our overall sample into two subsamples comprising: (1) cohorts that participated in Rietveld, Medland, et al. (2013) and (2) all new cohorts that were added to the current study. In what follows we refer to the former as the "Rietveld Cohorts" and the latter as the "New Cohorts." We refer to the combined-sample meta-analysis results reported by Rietveld, Medland et al. (2013) as the "Rietveld et al. (2013) Cohorts."

C. WITHIN-SAMPLE REPLICATION RESULTS

Table B6 reports the results of the replication analysis. In the upper panel, we report for the seven SNPs, their standardized effect sizes, standard errors, and *P*-values. We report these statistics from three separate meta-analyses of *EduYears* conducted in: (i) the Rietveld et al. (2013) Cohorts (ii) the Rietveld Cohorts, and (iii) the New Cohorts. The reference allele is chosen to be the allele associated with higher values of *EduYears* in Rietveld, Medland, et al.'s analysis (2013).

Given the high degree of overlap between cohorts in the previous EA meta-analysis (Rietveld, Medland, et al., 2013) and the Rietveld Cohorts, the similarity of the effect-size estimates is unsurprising. Reassuringly, the sign of the estimated coefficient in the New Cohorts is always in the predicted direction, and for all but one of the seven SNPs we can reject the null hypothesis of no effect at the 5% significance level (two SNPs, rs4851266 and rs9320913, reach genome-wide significant also in the replication sample). For six of the seven SNPs, the 95% confidence intervals for the estimated effect sizes overlap across the Rietveld Cohorts and the New Cohorts.

To further examine replicability, we examined if SNPs that reach genome-wide significance in a meta-analysis of the New Cohorts replicate in the Rietveld Cohorts. Applying the pruning algorithm described in Section 4.2.6 to meta-analysis results for the New Cohorts resulted in 14 approximately independent SNPs. The results from this replication analyses are reported in Panel B of Table B6. The results are similar to those of the replication of the associations from the Rietveld Cohorts in the New Cohorts: the signs align for all 14 SNPs, and 12 SNP replicate at P-value < 0.05 in the Rietveld Cohorts (none of them at genome-wide significance, but 5 at P-value < 10^{-5}).

In the two replication analyses, the average effects in the replication samples are about 35% smaller than the estimated effect of the genome-wide significant association, roughly consistent with the degree of inflation one would expect from a Winner's Curse correction of the sort described and performed in the Supplementary Information section 1.8.3 of Okbay, Beauchamp, et al., (2016).

4.2.8 Out-of-sample replication

Here, we report the results from a replication analysis of the 74 lead SNPs that emerged from our GWAS meta-analysis of *EduYears* in the first wave of UK Biobank (UKB) data. (Sudlow et al., 2015; UK Biobank, 2015b).

A. METHODS IN OUT-OF-SAMPLE REPLICATION ANALYSES

Our out-of-sample replication analyses uses data from the interim release of the UKB data and closely follows the methodological best practices recommended in the documentation that has been made publicly available through the UKB website (UK Biobank, 2015b). Following the "exemplary GWAS" described in the documentation, we restrict the analysis to the subsample of N = 112,338 conventionally unrelated individuals with "White British" ancestry. Dropping a small number of observations with missing phenotypic data leaves us with our final estimation (N = 111,349). Details on genotyping, pre-imputation quality control, and imputation of the interim release data have been documented extensively elsewhere (UK Biobank, 2015a).

Supplementary Table 1.14 in Okbay, Beauchamp, et al. (2016) provides additional details on the UKB analysis, including information about phenotype construction, sample demographics, association software, and the regression specification we estimate. As recommended by the UKB, we control for genotyping array in all analyses and use the software

SNPTEST with the "–method expected" option specified. We applied exactly the same quality-control filters as in our main analyses to the UKB results.

Because two of the 74 lead SNPs are missing from the quality-controlled UKB results file, we replaced them with nearby proxies. Specifically, we replaced lead SNP rs8005528 with rs8008779 ($r^2 = 0.69$) and lead SNP rs192818565 with rs55943044 ($r^2 = 0.93$). In both cases, the proxy was selected by choosing from the pooled discovery sample the lowest p-value SNP within 500 kb of the original lead SNP, restricting the search to SNPs available in the UKB data.

B. UKB REPLICATION RESULTS

Table B7 and Figure 4.5 report the results. Of the 74 lead SNPs, 72 have the anticipated sign in the replication sample, and 52 replicate at the 5% level (always with an effect size in the anticipated direction). Of the 52 SNPs, 7 reach genome-wide significance in the replication sample.

Under the null model that each of the lead SNPs are null in both the discovery and replication data, we would expect 50% of the SNPs (37 SNPs) to have a concordant sign in the discovery and replication samples, we would expect 5% (3.7 SNPs) to be significant at the 5% level, and we would expect 0.000005% (3.7×10^{-6} SNPs) to be genome-wide significant.

We can construct P-values associated with these results, noting that the number of SNPs that have a concordant sign or that are above a certain significance level is distributed as a Binomial(74, π) where π is the expected fraction of concordant or significant SNPs reported in the previous paragraph. Given that we are specifically interested in an increase in concordance or significance, we use a one-sided test. The P-value associated with the sign concordance is then 1.47×10^{-19}, the P-value associated with the number of SNPs significant at the 5% level is 2.68×10^{-50}, and the P-value associated with the number of genome-wide significant SNPs is 1.41×10^{-42}.

We can additionally measure the replicability of the GWAS estimates generally by assessing the genetic correlation between the discovery and replication samples. We estimate a genetic correlation of 0.946 ($SE = 0.021$) using bivariate LD Score regression [15]. These results, along

[15] We estimate the LD score regression using the "eur_w_ld_chr/" files of LD Scores computed by Finucane et al. (2015). In the LD Score regression, we include only HapMap3 SNPs with MAF > 0.01.

with the *P*-values reported above, suggest that the GWAS coefficients estimated in this paper in general, and the estimates of the 74 lead SNPs in particular, are highly replicable.

C. EXPECTED REPLICATION RECORD

To benchmark this replication record under a natural alternative hypothesis (as opposed to the expected replication under the null hypothesis calculated above), we calculated the expected degree of replication given the meta-analysis results, the sample size in the meta-analysis, and the sample size of the replication sample. To do this, we conducted a Bayesian Winner's Curse correction described in a previous study of cognitive performance (Rietveld, Esko, et al., 2014). Applying this methodology, we find that 71.4 of the 74 SNPs are expected to have matching signs, 40.3 SNPs are expected to be significant at the 5% level, and 0.6 SNPs are expected to be genome-wide significant. The observed numbers are, respectively, 72, 51 and 7. The replication record of the lead SNPs in the UKB is hence somewhat stronger than predicted by the power calculations. Supplementary Information section 1.8.3 in Okbay, Beauchamp, et al. (2016) provides additional details on the analysis.

4.2.9 Combined meta-analysis of discovery and replication cohorts (N = 405,072)

Using procedures identical to those described in Section Meta-analysis4.2.6, we conducted a meta-analysis of the *EduYears* phenotype, combining the results from our discovery cohorts (N = 293,723) and the results from the UKB replication cohort (N = 111,349). Expanding the overall sample size to N = 405,072 increases the number of approximately independent genome-wide significant loci from 74 to 162.

Supplementary Table 1.16 in Okbay, Beauchamp, et al. (2016) provides information about the lead SNPs in each of these loci.

CHAPTER 5

Of Genes and Screens: Educational Reform, Ability, and Labor Market Screening

Abstract

We study the heterogeneous effects of a Swedish educational reform that increased compulsory schooling from seven to nine years. Specifically, we examine how the reform differentially affected labor market outcomes for individuals with different ability levels, as measured by genetic endowments. Recent breakthroughs in genetics (Okbay, Beauchamp, et al., 2016) permit the construction of an index of genetic markers, or a "polygenic score" that credibly and robustly predicts educational attainment. We argue that this polygenic score represents a meaningful measure of labor market ability. The gradual rollout of the Swedish reform generates quasi-experimental variation that can be combined with individual data to estimate interactions between genetic ability and exposure to the reform. We find evidence of significant interactions between genetic ability and the reform for females' educational outcomes and earnings. Specifically, higher ability females were more likely to obtain a high school degree, which is beyond the new minimum established by the reform. This is consistent with a model in which employers screen workers based on their educational credentials, and higher ability females have an incentive to acquire more education to better signal their ability after the reform.

5.1 Introduction

Policymakers widely view education as a key element of the policy response to economic inequality. Indeed, an enormous body of literature studies the impact of educational reforms on schooling decisions, earnings, and other labor market outcomes. Compulsory schooling laws, in particular, have been studied in multiple contexts, including the United States (Acemoglu & Angrist, 2001; Angrist & Keueger, 1991), the United Kingdom (Oreopoulos, 2006), Germany (Pischke & von Wachter, 2008), and Sweden (Meghir & Palme, 2005). If economic inequality is a central motivating concern, then it is particularly important to understand how such policies affect the outcomes of individuals with different ability levels or disparate socioeconomic backgrounds. Existing evidence on the heterogeneous effects of compulsory schooling is mixed. Oreopoulos (2006) finds that an increase in compulsory schooling affecting many individuals in the U.K. had effects that were similar in magnitudes to similar reforms in the U.S. and Canada that affected a much smaller subset of the population. This suggests that there may be limited variation in the treatment effects of extra compulsory schooling – the average treatment effect may be similar to a local average treatment effect. By contrast, Meghir and Palme (2005) provide evidence that compulsory schooling laws may have largely different effects depending on gender, family background, and cognitive performance.

We study the heterogeneous impacts of an educational reform implemented in Sweden over the period 1949-1962. One of the chief provisions of this reform was an increase in compulsory schooling from seven to nine years. However, the reform also involved a nationalized curriculum, and a delay in the sorting of students into academic and vocational tracks. The reform was rolled out gradually throughout municipalities in Sweden. As a result, pupils belonging to the same age cohorts but living in different municipalities, and pupils living in the same municipality but from adjacent age cohorts, were assigned to different school systems. The gradual roll-out of the reform provides a natural experiment that permits the estimation of the effects of the reform that are not confounded by age, cohort, or macro effects.

We ask whether the schooling reform in Sweden led to different outcomes for individuals with different levels of ability. While ability is multi-faceted and arises from multiple sources, labor economists have long acknowledged that genetic factors play an important role in driving individual-level differences in human capital accumulation (Todd & Wolpin, 2003). We measure ability using an index of genetic markers constructed to credibly and robustly predict educational attainment. The index, or "polygenic score," that we use is based on the groundbreaking gene discovery work of Okbay, Beauchamp, et al. (2016). Similar

scores have been shown to predict not only educational attainment, but also later-life economic outcomes conditional on education (Belsky et al., 2016; Papageorge & Thom, 2016). We thus argue that the polygenic score studied here measures previously unobserved genetic components of labor market ability.

The use of genetic endowments to measure ability stands in contrast to an existing literature base that typically relies on cognitive test scores. In an important contribution, Meghir and Palme (2005) study the effect of the Swedish reform on completed education and adult earnings. They find that the reform succeeded in boosting educational attainment (including the probability of completing more than the new compulsory minimum), as well as adult earnings. Moreover, these effects were found to be particularly strong for high ability (high IQ) females from less privileged households (whose fathers had lower levels of education). These results are consistent with the reform exerting differential effects on individuals with different abilities. However, the use of IQ as a proxy for ability suffers from some well-known limitations. Environmental factors such as socio-economic status are known to influence performance in IQ tests (e.g. Turkheimer et al., 2003). That is, IQ performance may be an outcome and not simply a marker of ability. This may be worrisome in the context of the Swedish reform. While Meghir and Palme (2005) use a sixth grade measure of IQ (before the pre-reform compulsory minimum of seven years), the reform also included provisions that standardized the national curriculum and delayed the assignment of students into academic and vocational tracks. These provisions may very well have altered the performance of students before the seventh grade, directly influencing IQ. If this is true, then differential effects by IQ might reflect the actual impact of the reform rather than heterogeneous effects by ability. This highlights one of the advantages of using genetic endowments as a proxy for ability. While the polygenic score that we consider might be correlated with rearing environments (parents pass along their genes and shape environments), variation in the score is not plausibly *caused* by the reform.

We study the heterogeneous impact of the reform on educational attainment, cognitive performance (available only for men), and adult earnings. The reform had an obvious effect on the probability that individuals completed at least nine years of school (the new compulsory minimum). However, we also find that the reform had heterogeneous effects on the females' probability of completing a high school degree (beyond the compulsory minimum). Specifically, we find substantial interactions between the reform and ability as measured by the polygenic score: a one standard deviation increase in the polygenetic score is associated with a 6.8 percentage point increase in the effect of the reform on the likelihood that a female will complete high school. As argued by Lang and Kropp (1986), such results can be rationalized by an economic environment in which employers use educational credentials to screen for

ability. In such an environment, individuals acquire education both to acquire skills and to signal their underlying ability. An increase in the compulsory schooling minimum can thus cause individuals to acquire even more education than the new minimum as a way to signal their ability and distinguish themselves from the mass of individuals at the new minimum. Our results on education are indeed consistent with an account in which higher ability Swedish females acquired at least a high school education to signal their ability. We do not find significant interaction effects of the reform for males, including for cognitive performance.

Turning to adult earnings, we find significant interactions between the reform and genetic ability for females. Among females, a one standard deviation increase in the score is associated with a 0.039 increase in the effect of the reform on log income. These effects appear most strongly during mid-career (ages 33-42), and suggest that the differential gains experienced by high ability females as a result of the reform were also transmitted into differential gains in later-life income. Taken together, our results are quite consistent with those of Meghir and Palme (2005), though we find differential gains for high ability females on average (not just conditioning on those with low father's education).

Our work contributes to a burgeoning literature on the estimation of gene by environment (G×E) interactions. Most of the existing G×E literature suffers from several limitations, including low statistical power (Dick et al., 2015; Duncan & Keller, 2011; Hewitt, 2012) and the use of environmental variables that are not plausibly exogenous and that may thus capture genetic rather than environmental effects. We address the first limitation by employing a sample of several thousands males and females and by using the polygenic score as the genetic variable[16]. We address the second limitation by leveraging the quasi-exogenous variation in exposure to the schooling reform as our environmental variable.

This chapter proceeds as follows. Section 5.2 provides some institutional background on the Swedish reform we study. Section 5.3 describes our theoretical framework and explains the predictions we should expect from different mechanisms linking compulsory schooling laws to human capital accumulation. Section 5.4 describes the survey data used for education and earnings (5.4.1), and the polygenic score education (0). After describing our empirical framework in Section 5.5, we present our results in Section 5.6. Finally, Section 5.7 offers a concluding discussion.

[16] Most existing GxE research uses candidate genetic markers as genetic variables, many of these are now known to be false positive or to have very low R^2 (Beauchamp et al., 2011; Benjamin, Cesarini, Chabris, et al., 2012; Chabris et al., 2015). By contrast, as we show below, our polygenic score explains more than 6% of the variation in educational attainment.

5.2 The Swedish school reform

This section briefly discusses the Swedish compulsory school reform that was gradually rolled out across the country's municipalities during the 1950s and 1960s. A more detailed discussion of the reform is provided by Marklund (1981), Meghir and Palme (2005), Holmlund (2007), Hjalmarsson, Holmlund, and Lindquist (2015) and the references cited therein.

In the pre-reform school system, pupils went through grades one to four or one to six (depending on their municipality) in the "folkskolan" (common basic compulsory school). After grade four or six, more able students were selected based on their marks to attend "realskolan" (five-year or three- to four-year junior secondary school), and the remaining students stayed in the folkskolan until they completed their seven-year compulsory education.[17]

That system was extensively debated throughout the interwar period. In 1948, a parliamentary committee released a report with proposals for the future compulsory school system. The proposals rested on two main objectives: to increase equality of opportunity by postponing tracking and to meet the growing demand for education among the baby boom cohorts of the mid-1940s. The main recommendations were to increase compulsory schooling by two years, from seven to nine years, and to postpone educational tracking so that children with different levels of skills or educational ambition would be kept together in common classes until grade 9.[18]

The committee proposal led to a large-scale nationwide evaluation between 1949 and 1962 (Marklund, 1981), during which the reform was implemented in selected municipalities.[19] As a general rule, for a given municipality, all pupils who were in grades one to five in the year the reform was implemented were exposed to the reform, whereas those in grade six and up were not exposed. Thus, during the evaluation period, pupils belonging to the same birth cohorts but living in different municipalities, and pupils living in the same municipality but from adjacent birth cohorts, were assigned to different school systems.

The selection of municipalities that took part in the evaluation was not random. Municipalities that were interested in taking part in the reform had to report on different characteristics such as population growth, tax revenues, local demand for education, and availability of

[17] In some municipalities, mainly the largest cities, compulsory schooling was extended to eight years before the comprehensive school reform.
[18] The committee also proposed changes to the curriculum, including introducing English in grade 5. However, as Hjalmarsson et al. (2015) show, the changes with respect to tracking and the contents of education should not be exaggerated.
[19] In some large municipalities, the reform was introduced in certain schools only.

teachers and school premises to the central authorities. Based on this information, the National Board of Education selected municipalities for participation from the group of applicants. The main objective for the Board was to obtain a certain amount of variation across municipality types in order to facilitate the ongoing assessment of the reform.

A modest 14 municipalities in 12 different counties were selected for the first year of the evaluation (1949/1950). The number of municipalities joining the evaluation program grew steadily in the subsequent years until 1962, when the parliament decided to implement the reform throughout the country. The municipalities then had until 1969 to implement the new system for all affected cohorts.

5.3 Theoretical Framework

An increase in the level of compulsory schooling can plausibly affect schooling and earnings through multiple mechanisms. In the simplest possible formulation, an increase in the minimum level of schooling places a more stringent constraint on the individual's optimal education choice. Assuming that such laws are rigorously enforced, an increase in the minimum should cause individuals who are currently choosing fewer years of schooling to acquire exactly the new minimum as a new corner solution. If education has a causal effect on productivity (e.g. through the formation of useful skills), then such a reform will then affect earnings by boosting acquired human capital at the bottom end of the educational distribution.

An alternate theoretical account rests on the idea that education serves as a costly signal of ability (Lang & Kropp, 1986). Regardless of whether or not extra years of education actually produce useful skills, it could be the case that the kinds of abilities or characteristics that allow an individual to acquire more schooling (e.g., attention to detail, work ethic) might also independently boost productivity. In such an environment, employers might screen workers for higher levels of education because such credentials reveal valuable information about worker ability. Importantly, this means that relative levels of education may be important — high-ability workers may have an incentive to acquire costly education as a way of distinguishing themselves from lower ability workers who cannot acceptably incur such costs. An increase in the minimum compulsory schooling level may then have important indirect effects. Individuals that were previously choosing above-minimum levels of schooling might choose higher levels in order to distinguish themselves from the now more educated lower tail of the schooling distribution. If this screening story is operative, then an increase in compulsory schooling laws will not only increase educational attainment at the

bottom of the distribution, but it could also ripple through the distribution and cause an increase in the frequency of educational choices above the new minimum. Lang and Kropp (1986) demonstrate that such a pattern is observed in U.S. data.

5.4 Data

5.4.1 Non-genetic data

The non-genetic data we use comes from three main sources: the Swedish Twin registry, Statistics Sweden, and the Military Archives of Sweden. The Swedish Twin Registry (STR) is the world's largest twin registry and it contains all twins born in Sweden from 1926 and onwards (Lichtenstein et al., 2006); Statistics Sweden is our source for administrative data; and the Military Archives of Sweden contains data on cognitive performance.[20]

Swedish children start school the year they turn seven. Thus, the first cohort affected by the reform (who started grade five in 1949) was born in 1938, and the last affected cohort (who started school in 1962 when the parliament decided to permanently introduce the nine-year comprehensive school) was born in 1955.

We use information on home municipality from the census in 1960 to construct the reform status indicator for the individuals in our sample.[21] Holmlund (2007) shows that by 1960, a sizeable fraction of the individuals born between 1938 and 1942 no longer lived with their biological parents, suggesting at least some of them had moved from the municipality in which they attended compulsory school. To avoid miscoding the municipality in which individuals attended compulsory school, and thus the reform indicator, we therefore restrict the sample to individuals born between 1943 and 1955 in our main analyses.

Our analyses focus on three main outcomes: educational attainment, cognitive performance, and income. Data on educational attainment was imputed based on data on educational level and type of education as of 2005 or 2008 from Statistics Sweden administrative data. Educational level was measured according to the three-digit Swedish standard classification of education (SUN 2000) in the registers; following the manual for classifying educational pro-

[20] See Appendix C1 for details on the registers and variables.
[21] We are grateful to Helena Holmlund for sharing the data and code used for creating this indicator.

grammes in OECD countries (ISCED-97), we assigned years of schooling to each classification.[22] We also used the resulting variable to define dummies that indicate the highest educational degree obtained: we defined 9 years of schooling as a junior secondary degree (realskoleexamen), 12 years as a theoretical high school degree, and 15 or more years as a college or graduate degree.

Our income data consists of yearly taxable earnings ("sammanräknad förvärvsinkomst") from 1970, 1975, 1980, 1985, 1990, 1995, 2000 and 2005 as reported by employers to the tax authorities. In the Statistics Sweden administrative registers, taxable earnings are defined as the sum of wage labor income, income from own business, unemployment compensation and pension income. In the income models we use up to 8 observations per individual. We excluded individuals who are likely not to have worked full-time in a given year from our income regressions; specifically, we only included individual-year observations for individuals with an income exceeding SEK 100,000 in 2000 prices (the exchange rate in 2000 was $ 1 ≈ SEK 9.3).

To measure cognitive performance, we used data provided by the Military Archives of Sweden. All males in our sample were required by law to participate in military conscription around the age of 18. The enlistment procedure during the period we consider spanned two days and involved tests of health status, physical fitness, and cognitive and non-cognitive abilities. Among other things, the recruits took four tests intended to measure logical, verbal, spatial, and technical abilities. The results of these tests are transformed to a discrete nine-point Stanine scale with a mean of five and a standard deviation of two (Lindqvist & Vestman, 2011). The resulting measure has been shown to be a good measure of general intelligence (Carlstedt, 2000).

Table 5.1 shows summary statistics for these variables, separately for males and females. To check the representativeness of our sample, columns 3 and 4 display the corresponding population values for the cohorts born between 1943 and 1955 based on data from Statistics Sweden. The individuals in our twin sample are about half a year older than the national

[22] We assigned the following years of schooling to the classifications: (old) primary school (7); (new) compulsory school (9); (old) junior secondary education (9); high school (10-12 depending on the program); short university (13); longer university (14-17 depending on the program); short post-graduate (18); long post-graduate (20).

Table 5.1: Summary statistics and sample representativeness

	Twin sample		Population	
	Males	Females	Males	Females
Birthyear	1948.35	1948.34	1948.84	1948.80
	(3.74)	(3.75)	(3.71)	(3.72)
Share reformed	0.36	0.36	0.40	0.40
	(0.48)	(0.48)	(0.49)	(0.49)
Years of schooling	11.63	11.62	11.34	11.59
	(2.83)	(2.64)	(2.84)	(2.65)
Junior high/	0.46	0.51	0.48	0.51
new compulsory	(0.50)	(0.50)	(0.50)	(0.50)
High school	0.23	0.21	0.21	0.21
	(0.42)	(0.41)	(0.40)	(0.40)
College	0.17	0.18	0.16	0.18
	(0.38)	(0.38)	(0.37)	(0.39)
Income in 2000	312	217	305	218
(in SEK 1000)	(220)	(82)	(199)	(87)
Cognitive perform	5.02	-	-	-
(1943-1955)	(1.53)	-	-	-
Cognitive perform	5.15	-	5.04	-
(1951-1955)	(1.47)	-	(2.15)	-
Sample size	2,725	3,091	607,854	597,902

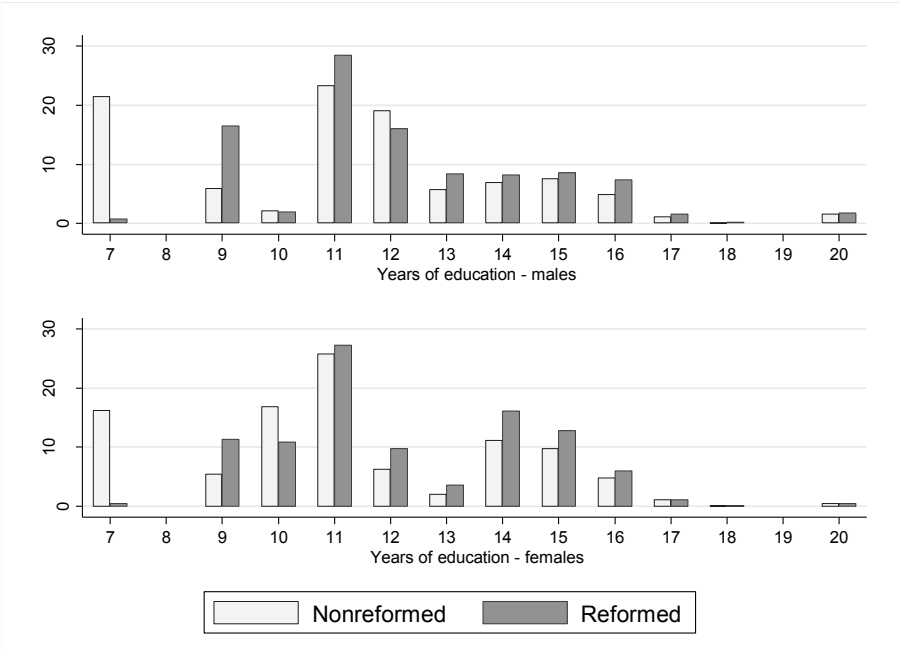
Note: Standard deviations are in parentheses. Data on cognitive performance for males in the population registers is only available from the birth cohort 1951 and onwards; for our main analyses, we standardized cognitive performance separately for each birth cohort.

average; because of that, the share of individuals affected by the school reform is somewhat smaller in our twin sample. Average levels of education, income and cognitive performance are slightly higher in our male twin sample compared to the population values, but the differences are small. Ashenfelter and Krueger (1994) and Ashenfelter and Rouse (1998) report a similar pattern of oversampling of twins with above average income and education. The female twins are very similar to the general population in terms of education and income.

Lastly, there is a slight upward trend in cognitive performance over time in our data. To account for this trend, we standardize the cognitive performance measure by birth cohort, so it has mean zero and unit variance for each birth cohort.

Figure 5.1 displays histograms of the years of schooling variable, separately for individuals who were and were not affected by the reform. Several features of the two distributions are

Figure 5.1: Histogram of years of schooling, before and after the school reform



worth noticing. First, there are virtually no post-reform individuals with less than 9 years of schooling.[23] Second, the reform increased the share of individuals with 9 years of schooling (the post-reform compulsory level) by about 10 percentage points. Third, the reform *decreased* the share of individuals with exactly 10 years of schooling, especially for females. Lastly, the reform had ripple effects through the distribution of schooling, increasing the share of individuals at each discrete value of years of schooling above 10 years of schooling (with the exception of 12 years of schooling for males).

5.4.2 Polygenic scores of educational attainment

In this study, we use estimates from a recent large-scale GWAS of educational attainment (Okbay, Beauchamp, et al., 2016) to construct "polygenic scores" (PGS) that aggregate the

[23] We dropped those few individuals from our main analyses, as it is likely they were subject to unusual circumstances or that our educational attainment variables or our reform dummy are erroneous for them. Our results are robust to the inclusion of these individuals.

estimated effects of millions of SNPs to partially predict individuals' educational attainment based on their measured genotypes.[24] Specifically, we compute a score $\widehat{PGS}_k = \sum_j \widehat{b_j} x_{kj}$ for all individuals k, where x_{kj} is k's genotype at SNP j and where $\widehat{b_j}$ is an estimate of SNP j's effect on educational attainment.

Polygenic scores are increasingly used in medical genetic research. Although genetic variants individually explain only a tiny fraction of the outcome, polygenic scores can have much larger predictive accuracy. For instance, polygenic scores constructed with the estimates from the most recent GWAS of body height can explain up to 17% of the variation in height (Wood et al., 2014).

The Okbay et al. (2016) GWAS meta-analysis included the STR. To avoid overfitting (Wray et al., 2013), the GWAS meta-analysis was re-run excluding the STR sample, resulting in a GWAS sample size of 390,687 individuals. SNPs with a sample size less than 100,000 individuals were removed from the meta-analysis, leaving 9,888,873 SNPs.

The STR individuals were genotyped using two different chips, the STR-Twingene cohort (9,617 individuals) was genotyped using Illumina HumanOmniExpress-12v1_A (644,556 SNPs) and STR-Salty cohort (5,109 individuals) using Illumina Infinium PsychArray (556,899 SNPs). The genotypes were then imputed to 30,061,897 and 30,941,403 SNPs for STR-Twingene and STR-Salty, respectively, using the March 2012 release of 1000 Genomes Phase 1 haplotypes (The 1000 Genomes Project Consortium, 2012) as reference panel. We converted the imputed genotype probabilities to best-guess format using the GCTA software (Yang, Lee, et al., 2011) and subsequently merged the STR-Twingene and STR-Salty samples, keeping only the SNPs that are present in both data sets. The final genotype data consisted of 28,681,763 SNPs and 14,726 individuals

For our baseline analyses, we used the software LDpred (Vilhjálmsson et al., 2015) to estimate the $\widehat{b_j}$'s. LDpred uses information on the correlation between the SNPs from a reference panel, together with a prior on the SNPs' effect sizes, to adjust the GWAS estimates (the $\hat{\beta}_j$'s) and obtain estimates of the SNPs' causal effects (independent of the effects of the other SNPs).

[24] "Genotyping" refers to the process of determining an individual's genotype at a certain location in the genome. Modern genotyping chips can relatively inexpensively genotype individuals at more than a million SNPs from across the genome. It is then possible to use external reference panels with information on the correlation structure between SNPs to impute the genotypes of up to ~80mil SNPs for individuals genotyped these genotyping chips.

We used an LD reference panel constructed from the STR genotype data. To construct the panel, we first selected the HapMap 3 SNPs from the merged best-guess genotype data described above because HapMap 3 SNPs are in general known to be imputed reliably (The International HapMap 3 Consortium et al., 2010). Since presence of related individuals in the LD reference data can introduce bias in estimates of SNP correlations, we dropped all individuals but one per family. In the remaining sample, we computed the genetic relationship matrix (covariance matrix of the individuals' genotypic data) and removed one individual from each pair with a relatedness greater than 0.025, using the software PLINK (Chang et al., 2015). Finally, we checked for genetic outliers in the sample. Using PLINK, we clustered individuals based on pairwise identity-by-state (IBS) distance. For each individual, we checked whether the individual is less similar to his/her closest neighbor than other individuals are to their closest neighbors by calculating a Z-score for that individual using the sample mean and variance. An extremely low Z-score indicates that the individual is an outlier. Using this procedure with a Z-score cutoff value of -4, we detected no genetic outliers in the sample. The final LD reference panel consisted of 8886 individuals and 1,217,311 SNPs.

The LD-adjusted SNP effects were calculated for the set of 1,179,485 SNPs that were present in both the LD reference data and the GWAS summary statistics. We computed the polygenic scores in PLINK with the merged best-guess genotype data for the STR sample, and the LD-adjusted weights obtained for a range of different priors on the fraction of causal SNPs: 1, 0.3, 0.1, 0.03, 0.01, 0.003, 0.001 and 0.0003. We selected the prior that maximizes the incremental R^2 of adding the score in a regression of educational attainment on sex, birth year, sex and birth year interaction, and the top 10 principal components of the genetic relationship matrix. The best performing score was the score with prior 1, with an incremental R^2 of 6.2% in our sample. To verify the robustness of our analyses, we also used PLINK (Chang et al., 2015) to construct a score with the unadjusted $\hat{\beta}_j$ estimates from the Okbay et al. GWAS. That score has in incremental R^2 of 5.5%, and our main results are robust to the use of that score. Table C1 shows the incremental R^2 of all scores together with 95% (percentile) confidence intervals obtained by bootstrapping with 1000 repetitions.

We interpret the resulting polygenic scores of educational attainment as capturing genetic propensity (or innate ability) to acquire more education, such that individuals with higher scores face lower costs of acquiring more education, *ceteris paribus*.

5.5 Empirical framework

To estimate the average impact of the Swedish comprehensive schooling reform on educational attainment, cognitive performance, and income, we follow Meghir and Palme (2005) and employ differences-in-differences. Our identification strategy leverages the fact that the reform was implemented at different times in different municipalities; it is based on the comparison of cohorts that were schooled before and after the reform within municipalities, and on the comparison of pre- and post-reform municipalities within cohorts.

The point of departure for our empirical framework is the following specification:

$$y_{icm} = \beta_0 + \beta_R R_{cm} + \beta_S S_{icm} + \mathbf{C_{icm}}\boldsymbol{\beta_C} + \varepsilon_{icm}, \tag{5.1}$$

where y_{icm} is some outcome of interest (educational attainment, income or cognitive performance), for individual i in birth cohort c and who went to school in municipality m; R_{cm} is a dummy indicator of reform status for cohort c in municipality m; S_{icm} is the polygenic score of educational attainment for individual i; and $\mathbf{C_{icm}}$ is a vector of control variables for individual i.

The control variables include fixed effects for each birth cohort (i.e., birth year) and a set of municipality clusters. We defined the municipality clusters by grouping together municipalities having the same first birth cohort affected by the reform.[25] They also include the top ten principal components of the genetic-relatedness matrix among the regressors (to control for population stratification, as explained above). We run all regressions separately for males and females, so we do not need to control for gender.

Because the municipalities to a certain extent could self-select into the evaluation program, it is possible that the timing of the reform was related to municipality specific characteristics that also influenced the outcomes of interest. More precisely, our identification strategy relies on the assumption that *changes* in municipality-specific factors are not correlated with

[25] This approach is similar to the setup used in Pekkarinen, Uusitalo, & Kerr (2009) in their study of the effects of a Finnish school reform on intergenerational income mobility. Including a fixed effect for each of the 1,000+ municipalities would substantially reduce the degrees of freedoms of our regressions. Furthermore, in the interaction specification introduced below, the fixed effects for the municipality clusters and the cluster-specific birth year trends are interacted with both the reform dummy and the polygenic score; using fixed effects for each municipality instead of for each municipality cluster would leave us with no degrees of freedom at all. As we show in Section 5.6.4, our main estimates are very similar and remain significant when using municipality fixed effects instead of fixed effects for the municipality clusters (while still using cluster-specific birth year trends).

the exact timing of the reform, conditional on the control variables.[26] To minimize the chances this assumption fails, we also include in our regressions separate birth year trends for each municipality cluster as well as a set of time-varying municipality-level covariates intended to measure demographic and socioeconomic changes.[27] The detailed analyses presented in Hjalmarsson et al. (2015) and Lindgren, Oskarsson and Dawes (n.d.) further corroborate the view that we can treat reform participation as exogenous in our sample.

To estimate possible interaction effects between the polygenic score and the effects of the schooling reform, we augment the above model with terms for the interaction between the reform dummy and the score as well as for interactions between the control variables and the score and between the control variables and the reform dummy[28]:

$$y_{icm} = \beta_0 + \beta_R R_{cm} + \beta_S S_{icm} + \beta_{RS}(R_{cm} \times S_{icm}) + \mathbf{C_{icm}}\boldsymbol{\beta}_{\mathbf{C}} + (\mathbf{C_{icm}} \times S_{icm})\boldsymbol{\beta_{CS}} + (\mathbf{C_{icm}} \times R_{cm})\boldsymbol{\beta_{CR}} + \varepsilon_{icm} \quad (5.2)$$

For the continuous outcomes (years of schooling, cognitive performance, and income), we estimated the above regression by ordinary least squares, clustering at the municipality level. For the binary outcomes (dummies indicating highest degree completed), we estimated linear probability models,[29] also clustering at the municipality level.

For income, we estimated a panel model in which we also controlled for a third degree polynomial of age (in addition to the above control variables), by ordinary least squares and clustering at the municipality level. For the baseline specification, we included all individual-year observations for which an individual was between 25 and 55 years old when his or

[26] Since our regressions include fixed effects for the municipality clusters, time-invariant differences between early and late reformers will not compromise our identification strategy.

[27] The time-varying municipality level covariates include the following variables (each measured the year the individual turned eleven): municipal level voter turnout, vote shares for the largest parties, and size of the electorate. We use political indicators since year-by-year indicators of socioeconomic development at the municipal level are only available for more recent time-periods. However, previous research has shown that aggregate level turnout and party vote shares in Sweden are highly correlated with more direct measures of socioeconomic development (Elinder, 2010). To create the year-by-year indicators, we interpolated turnout, vote shares and electorate size between the election years (1948, 1952, 1956, 1958, 1960, 1964, and 1968).

[28] A common concern in GxE studies is that interaction effects may be driven by confounders; for this reason, it is important to control for interactions between the control variables and the two interacted covariates of interest. For example, suppose that the average polygenic score is higher in wealthier cities and that the reform had a smaller effect in those cities. Under such a scenario, an estimate of the interaction between reform status and the score may be confounded unless we control for the interaction between reform status and municipality wealth (or municipality fixed effects).

[29] As Ai and Norton (2003) show, coefficients on interaction terms are not easy to interpret in probit and logit models. Because we are primarily interested in the coefficient on the interaction between score and reform, we use a linear probability model instead of a logit or probit model for the binary outcomes.

her income was measured (as mentioned above, income was measured every five years). We also ran specifications including all individual-year observations for which an individual was 23 to 32 years old ("early career"); all individual-year observations for which an individual was 33 to 42 years old ("mid career"); and all individual-year observations for which an individual was 43 to 52 years old ("late career"). (Because the income data was measured every five years, most individuals had two individual-year observations in each of the early, mid, and late career specifications.)

In each municipality we exclude the birth cohort preceding the first cohort affected by the school reform. The reason for doing so is that previous studies have shown that the youngest pre-reform cohort was significantly affected by the reform, possibly due to the fact that a substantial share of the pupils born late in a given year started school a year later than they were supposed to (Fredriksson & Öckert, 2014; Hjalmarsson et al., 2015)

5.6 Results

5.6.1 Effect of the schooling reform on educational attainment and cognitive performance

The top panel of Figure 5.2 presents our estimates of the average impact of the reform on years of completed schooling, the probability that one's highest completed degree is junior high school, high school, or college (including graduate school), and cognitive performance. These are estimates from the additive model (5.1) that do not include interactions between the reform and the polygenic score. The estimates suggest that the reform increased mean schooling by about 0.45 years among both males and females. These estimates are similar to (and not statistically different from) those of Meghir and Palme (2005). The reform had a positive impact on the probability that one's highest degree is junior high school—which is not surprising because the reform made this the new mandatory minimum level of education. However, it had no discernible effect at the higher end of the educational distribution nor on cognitive performance for males.

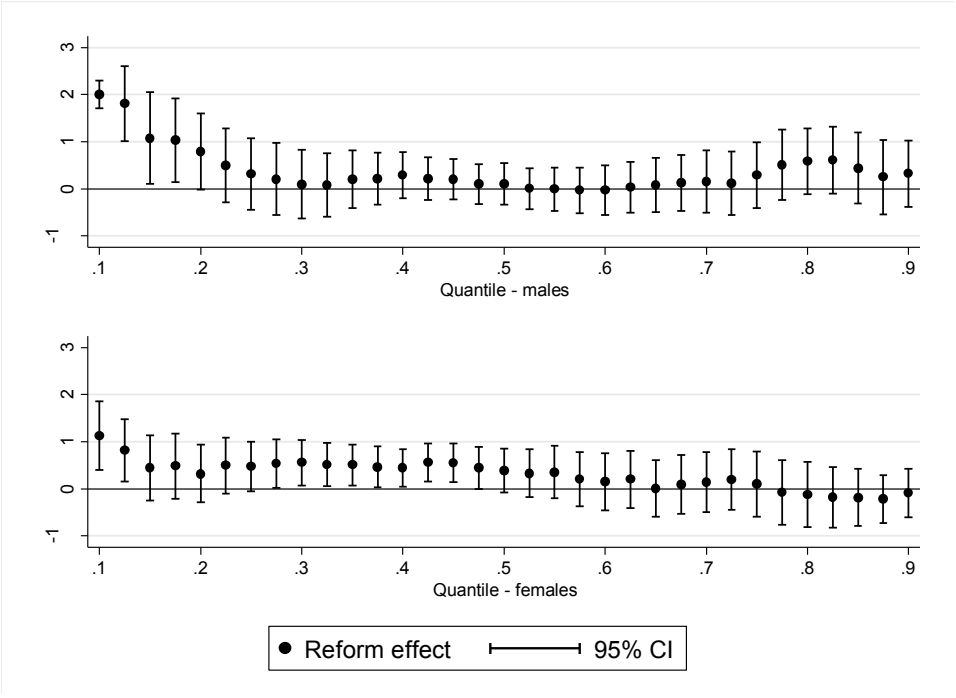
Figure 5.2 plots estimates from quantile regressions explaining years of schooling. The reform had sizeable effects on years of schooling at the lower end of the schooling distribution for both males and females, but effects beyond the 0.2 quantile are not significant. In fact, for males, the reform increased average years of schooling by about two years at the 0.1 quantile, consistent with the fact that it increased the minimum mandatory level of schooling from seven to nine years.

An examination of main effects alone suggest that the reform only impacted the individuals who would otherwise not have completed nine years of schooling, and that it had little or no ripple effects on the distribution of educational attainment. The main effects are thus not consistent with a strong signaling response to the expansion of compulsory education.

5.6.2 Effect of interactions between genes and the reform on educational attainment and cognitive performance

The bottom panel of Table 5.2 reports the estimates on the coefficient β_{RS} on the interaction between the score and the reform dummy in Equation (5.2), for various outcome variables.

Figure 5.2. Quantile regressions: Impact of reform on years of schooling by sex



Note: Each point corresponds to the estimate from a quantile regression of the effect of the reform on years of schooling at a quantile. All quantile regressions include controls for the first 10 principal components, birth year fixed effects, fixed effects for municipality clusters, municipality cluster-specific birth year trends, and municipality level covariates. (The standard errors used to calculate the 95% confidence intervals were not clustered.)

For males, β_{RS} is not significant in any of the regressions, suggesting that the reform did not have markedly larger or smaller effects on males with different abilities.

For females, the interaction between the score and the reform dummy is not significant for years of schooling. However, the estimates suggest that the reform had a larger effect on the probability of completing junior high school as one's highest degree for females with lower scores. This is unsurprising, as the reform increased the mandatory minimum schooling level from seven to nine years, thus directly principally affected the lower end of the educational distribution.

The estimates also suggest that the reform had a larger effect on the probability that high school is one's highest degree for females with a *higher* score. The magnitude of this estimate is striking. A one standard deviation increase in the polygenic score is associated with a reform effect that is larger by 0.068 for completing high school as a terminal degree. This result is consistent with the implications of a signaling model of education. Raising the minimum level of compulsory education from seven to nine years increased the likelihood that some females chose even higher levels of education (high school). Moreover, we find that this behavior is more likely among higher ability females—precisely the group that is incentivized to engage in costly signaling after the reform. Note as well, that this result is not due to a mechanical rightward shift in the educational distribution, since the mean polygenic score for females with a high school degree was about average prior to the reform.

Figures Figure 5.3 and Figure 5.4 further illustrate these results. In Figure 5.3, we present results from linear probability models for the outcome of obtaining exactly x years of schooling for x in the range 9-15. The reform had larger positive effects on the probability of completing exactly nine years of schooling for females with lower scores, and had larger positive effects on the probability of completing exactly 12, 13, or 14 years of schooling for females with higher scores. Figure 5.4 shows the average score by educational degree, before and after the reform. The difference between the mean polygenic score of females whose highest after degree is junior high school and females whose highest degree is high school is much larger after the reform. This suggests that the reform prompted some females with high scores to get a high school degree, consistent with a signaling model.

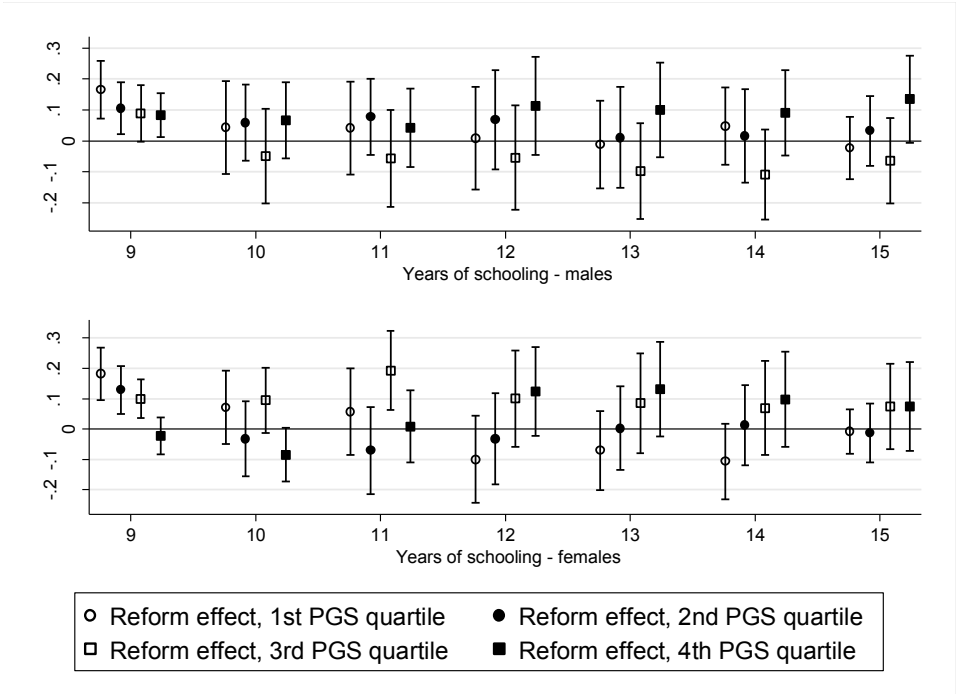
Table 5.2. Impact of the reform on educational attainment and cognitive performance among males and females.

		Years of schooling		Junior high school / New compulsory		High school		College		Cognitive Performance
		(males)	(females)	(males)	(females)	(males)	(females)	(males)	(females)	(males)
Additive models	Reform	0.459**	0.449**	0.097**	0.092**	0.000	-0.001	0.026	0.027	-0.028
		(0.214)	(0.178)	(0.044)	(0.041)	(0.032)	(0.033)	(0.031)	(0.029)	(0.088)
	PGS	0.660***	0.685***	-0.065***	-0.078***	0.033***	0.035***	0.075***	0.080***	0.256***
		(0.045)	(0.052)	(0.009)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)	(0.020)
Interaction models	Reform× PGS	-0.033	0.131	-0.023	-0.164***	-0.035	0.068**	0.031	0.024	0.044
		(0.215)	(0.182)	(0.041)	(0.036)	(0.038)	(0.034)	(0.032)	(0.030)	(0.091)
	Sample size	2,725	3,091	2,725	3,091	2,725	3,091	2,725	3,091	2,605

Note: All regressions include controls for the first 10 principal components, birth year fixed effects, fixed effects for municipality clusters, municipality cluster-specific birth year trends, and municipality level covariates. Standard errors, shown in parentheses, allow for clustering at the municipality level.

***/**/* indicate significance at the 1/5/10% level.

Figure 5.3. Impact of reform by PGS quartiles



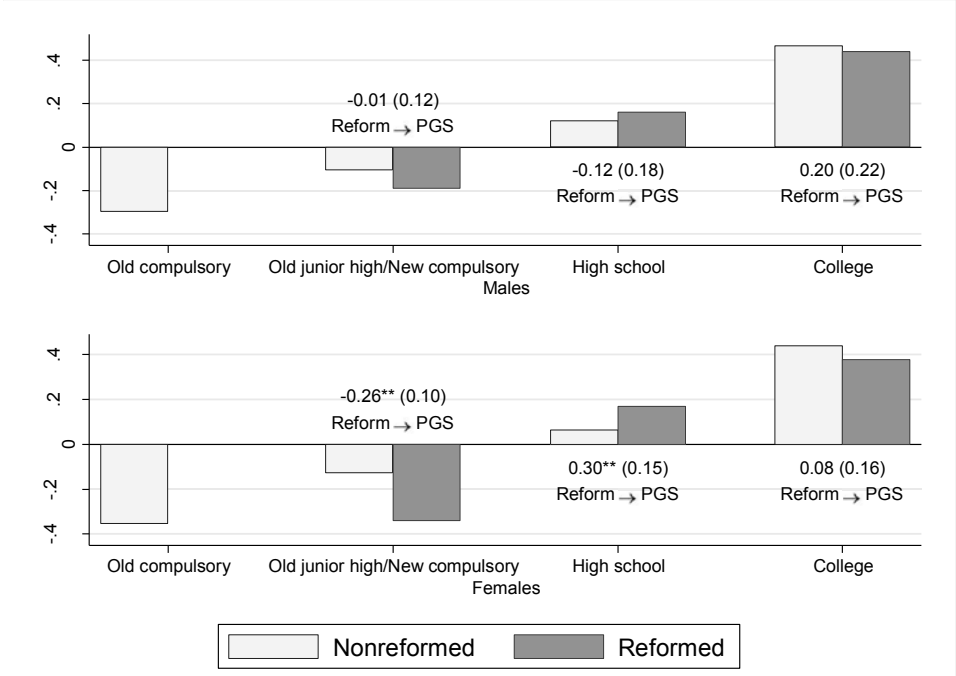
Note: Each point corresponds to the estimated effect of the reform on the probability of having completed exactly x years of schooling, where $x \in \{9, 10, \ldots, 15\}$, for a quartile. Each estimate was obtained from regressions estimating model (1) among individuals in a given PGS quartile, with a dummy indicating if an individual has completed exactly x years of schooling as the dependent variable. The whiskers denote 95% confidence intervals (the standard errors used to calculate these allowed for clustering at the municipality level).

5.6.3 Effects on income

We next turn to adult earnings. The top panel of Table 5.2 shows the estimates of the impact of the reform on earnings from the additive model (5.1) without an interaction between the reform and the polygenic score. For males, the reform had no significant impact on income, consistent with our education results. However for females, the reform had a marginally significant positive effect on income and a significant positive effect when the sample is restricted to mid-career observations (ages 33-42).

The bottom panel of Table 5.2 reports the estimates of the coefficient on the interaction between the reform and the score in our full specification. The estimates imply that the reform had a significantly larger impact on the income of females with higher scores. In the additive model, a one standard deviation increase in the polygenic score is associated with an increase in mid-career log income of approximate 0.045. However, the estimates of the interaction effect suggest that a one-standard deviation increase in genetic ability increased the impact of the reform by about 0.039. That is, the reform appears to have substantially widened the genetic gradient in earnings. This seems to be driven by the higher impact of the reform on the mid-career income of females with higher scores (for early and late career income, the

Figure 5.4. Average score by educational degree, before and after the reform

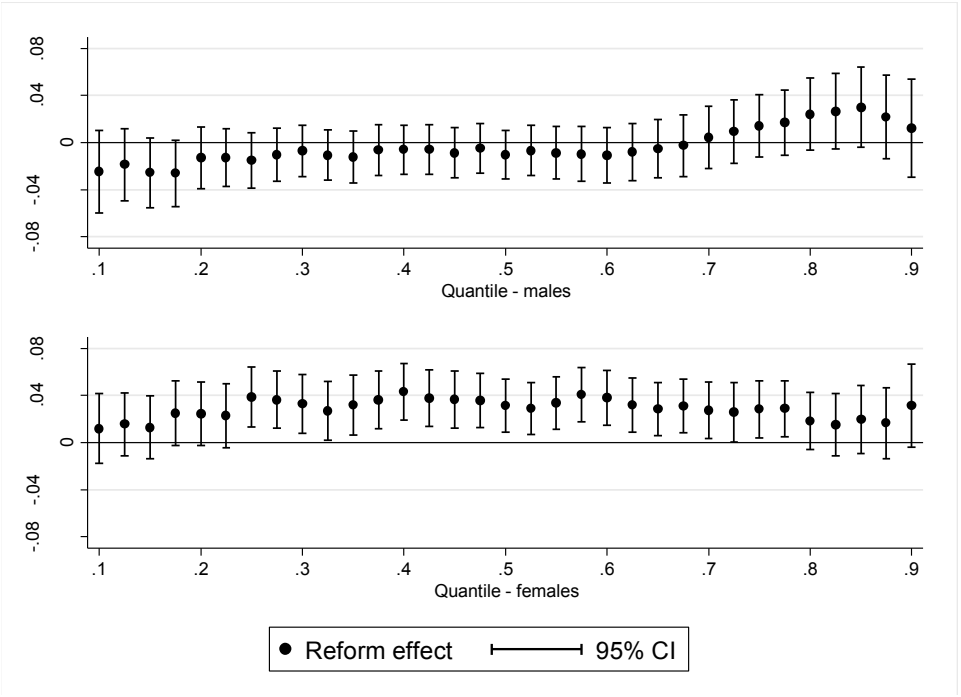


Note: Each number displayed above or below the bars is an estimate of the effect of the school reform on the polygenic score (with standard errors allowing for clustering at the municipality level in parentheses), from a regression of the polygenic score on the reform dummy and on control variables among the sample of individuals whose highest educational degree is the one indicated under the horizontal axis. The control variables included the first 10 principal components, birth year fixed effects, fixed effects for municipality clusters, municipality cluster-specific birth year trends, and municipality level covariates
***/**/* indicate significance at the 1/5/10% level.

coefficients on the interactions are still positive, but the standard errors are large because there are fewer observations.

Figure 5.5 and Figures C1-C3 (Appendix C) present estimates from quantile regressions of the effect of the reform on income, for males and females. These are estimates from additive models that do not include interactions between the reform and the polygenic score. For females, the reform had significant positive effects on income for quantiles between 0.25 and 0.8. This suggests that our estimated earnings effects are coming from differences in the middle of the earnings distribution rather than events in the tails.

Figure 5.5. Quantile regressions: Impact of reform on income by sex



Note: Each point corresponds to the estimate from a quantile regression of the effect of the reform on income at a quantile. All quantile regressions estimated panel models that included all individual-year observations for which an individual was between 25 and 55 years old when his or her income was measured. All quantile regressions include controls for the first 10 principal components, birth year fixed effects, fixed effects for municipality clusters, municipality cluster-specific birth year trends, municipality level covariates, and a third degree polynomial of age. (The standard errors used to calculate the 95% confidence intervals were not clustered.).

5.6.4 Sensitivity analysis

We ran a number of additional regressions to check the robustness of our results. We estimated equations (5.1) and (5.2) for the various outcomes (i) with fixed effects for municipalities (instead of focr municipality clusters); (ii) using all individuals born between 1945 and 1955 and (iii) using all individuals born between 1940 and 1958 (instead of between 1943 and 1955); (iv) using birth municipality to code the reform dummy (instead of using municipality of residence as of the 1960 census); and using the PLINK score instead of the LDpred score. Overall, as can be seen in Tables C2-C7 (Appendix C), the results are robust.

5.7 Conclusion

Our results provide evidence that a major Swedish school reform differentially affected the education and earnings outcomes of individuals with different genetic endowments. These differential effects were highly gender specific. While we find no evidence of interactions for males, we find that the reform boosted the probability of completing exactly a high school degree much more for high-ability females than for low-ability females. Since the reform raised the compulsory schooling minimum to a grade level below high school (grade nine), these results are consistent with a signaling model of education. Higher ability females may have acquired schooling beyond the new minimum as a way to signal their ability in the labor market. These differential effects are also present in earnings results. While the reform boosted the average earnings of females, we find a statistically significant interaction between the reform and our measure of genetic ability in the earnings equation. The reform appears to have increased the earnings of higher ability females more than lower ability females.

The results presented here are largely consistent with results found in Meghir and Palme (2005), who use grade six IQ tests as a measure of ability. The advantage of using genetic data stems from the fact that, unlike cognitive test scores, our measure of genetic endowments cannot be influenced by the school reform under study. Plausible scenarios exist in which reforms may directly impact test-based proxies for ability. In such a world, differential results on the basis of such proxies will not reveal anything about heterogeneous effects. Our results largely match those of Meghir and Palme (2005), which may cast doubt on the existence of a serious endogeneity problem in this instance. However, going forward, our results suggest that genetic information can be used to provide useful measurements of labor market ability and detect the presence of interactions between ability and educational reforms. As

genetic data becomes more available, this approach to the measurement of ability and heterogeneous effects may prove increasingly useful—especially in contexts where ability proxies are either unbailable or highly likely to reflect endogenous factors.

Table 5.3. Impact of the reform on log income among males and females.

		Income		**Income early career**		**Income mid career**		**Income late career**	
		(males)	(females)	(males)	(females)	(males)	(females)	(males)	(females)
Additive models	Reform	-0.006	0.031*	-0.001	0.022	-0.000	0.045**	-0.018	0.029
		(0.020)	(0.018)	(0.017)	(0.022)	(0.022)	(0.023)	(0.028)	(0.022)
	PGS	0.042***	0.045***	0.021***	0.040***	0.046***	0.035***	0.052***	0.051***
		(0.005)	(0.005)	(0.004)	(0.005)	(0.006)	(0.006)	(0.007)	(0.005)
Interaction models	Reform× PGS	0.015	0.039**	0.028*	0.026	0.004	0.059***	0.008	0.021
		(0.019)	(0.016)	(0.016)	(0.021)	(0.022)	(0.022)	(0.028)	(0.021)
	Sample size	15,882	14,512	4,879	3,394	5,226	4,590	5,182	5,691

Note: All regressions estimated panel models that included all individual-year observations for which an individual was in the model's age range when his or her income was measured. All regressions include controls for the first 10 principal components, birth year fixed effects, fixed effects for municipality clusters, municipality cluster-specific birth year trends, municipality level covariates, and a third degree polynomial of age. For models 1 and 2, the age range is 25 to 55 years old. Early, mid, and late career income refer to income when the individual was 23 and 32, 33 to 42, and 43 to 52 years old, respectively. Standard errors, shown in parentheses, allow for clustering at the municipality level.
***/**/* indicate significance at the 1/5/10% level.

CHAPTER 6

Meta-GWAS Accuracy and Power (MetaGAP) Calculator Shows that Hiding Heritability is Partially due to Imperfect Genetic Correlations across Studies

Abstract

Large-scale GWAS results are typically obtained by meta-analyzing GWAS results from multiple studies spanning different regions and/or time periods. This approach averages the estimated effects of individual genetic variants across studies. In case genetic effects are heterogeneous across studies, the statistical power of a GWAS and the predictive accuracy of polygenic scores are attenuated, contributing to the so-called "missing" heritability. However, a theoretical multi-study framework, relating statistical power and predictive accuracy to cross-study heterogeneity, is not available. We address this gap by developing an online Meta-GWAS Accuracy and Power calculator that accounts for the cross-study genetic correlation. This calculator enables to explore to what extent an imperfect cross-study genetic correlation (i.e., less than one) contributes to the missing heritability. By means of simulation studies, we show that under a wide range of genetic architectures, the statistical power and predictive accuracy inferred by this calculator are accurate. We use the calculator to assess recent GWAS efforts and show that the effect of cross-study genetic correlation on statistical power and predictive accuracy is substantial. Hence, cross-study genetic correlation explains a considerable part of the missing heritability. Therefore, *a priori* calculations of statistical power and predictive accuracy, accounting for heterogeneity in genetic effects across studies, are an important tool for adequately inferring whether an intended meta-analysis of GWAS results is likely to yield meaningful outcomes.

6.1 Introduction

Large-scale GWAS efforts are rapidly elucidating the genetic architecture of polygenic traits, including anthropometrics (Locke et al., 2015; Wood et al., 2014), diseases (Eeles et al., 2009; Ehret et al., 2011; Ripke et al., 2014) and behavioral and psychological outcomes (Okbay, Baselmans, et al., 2016; Okbay, Beauchamp, et al., 2016; Rietveld, Cesarini, et al., 2013). These efforts have led to new biological insights, therapeutic targets, and individual-level polygenic scores (PGS), and help to understand the complex interplay between genes and environments in shaping individual outcomes (Benjamin, Cesarini, Chabris, et al., 2012; Okbay, Beauchamp, et al., 2016; Visscher et al., 2012). However, GWAS results for polygenic traits do not yet account for a large part of the heritability (Locke et al., 2015; Okbay, Baselmans, et al., 2016; Okbay, Beauchamp, et al., 2016; Wood et al., 2014). This dissonance, which is referred to as the 'missing' heritability, has received broad attention (Eichler et al., 2010; Maher, 2008; Manolio et al., 2009; Witte, Visscher, & Wray, 2014; Wray et al., 2013; Wray & Maier, 2014; Zuk, Hechter, Sunyaev, & Lander, 2012).

The missing heritability can be split into two parts. The first part, the 'still-missing' heritability (Witte et al., 2014; Wray et al., 2013; Wray & Maier, 2014), is defined as the difference between the estimate of heritability based on family data (h^2) and the SNP-based estimate (h^2_{SNP}). The second part, the 'hiding' heritability (Witte et al., 2014; Wray et al., 2013; Wray & Maier, 2014), is defined as the difference between the h^2SNP and the estimate of heritability based on genetic variants that reach genome-wide significance in a GWAS (h^2_{GWS}). Hence, $h^2 > h^2_{SNP} > h^2_{GWS}$ (Wray et al., 2013).

Four important factors have been proposed to explain the missing heritability. First, conventional genotyping is not sufficiently dense across the whole genome. Therefore, genotyping fails to capture rare variants that explain a non-negligible fraction of trait variation (Yang et al., 2010). Second, gene–gene interactions inflate h^2, creating so-called 'phantom' heritability (Zuk et al., 2012). Third, sample sizes of GWAS efforts are not large enough to fully capture h^2_{SNP} (S. H. Lee, Wray, Goddard, & Visscher, 2011; Yang et al., 2010). Fourth, differences across strata (e.g., studies, ancestry groups, and sexes) in phenotypic measurement, in measurement accuracy, and in genetic effects, can all introduce additional noise and loss of signal (S. H. Lee et al., 2013; Wray, Lee, & Kendler, 2012), and, hence, attenuate statistical power of a GWAS (S. Lee, Teslovich, Boehnke, & Lin, 2013; Sham & Purcell, 2014; Wray & Maier, 2014). The first two factors lead primarily to still-missing heritability (Wray & Maier, 2014; Yang et al., 2010; Zuk et al., 2012), while the latter two contribute foremost

to hiding heritability (S. H. Lee et al., 2011; Sham & Purcell, 2014; Wray & Maier, 2014; Yang et al., 2010).

Recent work has demonstrated the feasibility of denser genotyping (Yang et al., 2015; Zuk et al., 2014) and larger GWAS samples (Dudbridge, 2013; Locke et al., 2015; Ripke et al., 2014; Wood et al., 2014). Hence, these two causes of the missing heritability can be amended. Moreover, the issue of phantom heritability is primarily of importance to the discussion about the still-missing heritability (Wray et al., 2012; Wray & Maier, 2014). In the current study, we focus on one important remaining factor in the hiding heritability discussion: heterogeneity of measures and/or heterogeneity of genetic effects across different strata, and, in particular, across studies.

Large-scale GWAS results are typically obtained by meta-analyzing GWAS results from multiple studies spanning different regions and/or time periods. This approach averages the estimated effects of individual genetic variants across studies. In case genetic effects are heterogeneous across studies (e.g., due to gene–environment interactions and heterogeneity in phenotypic measurement) at least three important quantities are decreased: the estimate of SNP-heritability (Wray et al., 2012; Wray & Maier, 2014), the statistical power of a GWAS (S. Lee et al., 2013; Sham & Purcell, 2014; Wray & Maier, 2014), and the predictive accuracy of the PGS (Dudbridge, 2013). The decrease in these quantities not only explains why heterogeneity contributes to the missing heritability but also shows that heterogeneity decreases the chances of a study to yield meaningful results (S. H. Lee, Yang, Goddard, Visscher, & Wray, 2012; Sham & Purcell, 2014). Therefore, the precise attenuation due to genetic heterogeneity should be well understood, in order to make an informed decision whether to pursue a proposed meta-analysis of GWAS results.

Others have already pointed at the issue of genetic heterogeneity across studies (S. H. Lee et al., 2013; McClellan & King, 2010; Sham & Purcell, 2014). In particular, it has been shown theoretically that misclassification between two diseases tends to deflate heritability estimates and decrease statistical power to detect trait-associated SNPs (Wray et al., 2012). In addition, empirical applications show that SNP-heritability estimates are attenuated when pooling across studies (S. H. Lee et al., 2013; S. H. Lee, DeCandia, et al., 2012). Moreover, simulations have shown that phenotypic and genetic heterogeneity decrease statistical power (S. Lee et al., 2013). Finally, a strong theoretical decrease in statistical power has been shown to exist under genetic heterogeneity of another sort, viz., when different intermediate phenotypes contribute to a single composite phenotype (Wray & Maier, 2014). Finally, a theoretical reduction of PGS predictive accuracy has been shown for a scenario with one discovery study and one study used as hold-out sample for prediction (Dudbridge, 2013).

Overall, findings from simulations, empirical work, and theory suggest attenuation due to genetic and phenotypic heterogeneity. Despite these efforts, a theoretical multi-study framework, relating statistical power and predictive accuracy to cross-study heterogeneity, is still absent.

In the current study, we address the absence of a general multi-study framework by developing a Meta-GWAS Accuracy and Power (MetaGAP) calculator that accounts for the cross-study genetic correlation (CGR). Moreover, by means of simulation studies, we show that under a wide range of genetic architectures, the statistical power and predictive accuracy inferred by this calculator are accurate. The calculator requires users to specify the number of studies, the sample size of each study, the SNP-based heritability per study, and the CGR. From these input parameters, the calculator infers the statistical power to detect associated SNPs and the predictive accuracy of the PGS in a meta-analysis of GWAS results from genetically and phenotypically heterogeneous studies. The MetaGAP calculator enables to explore to what extent an imperfect CGR (i.e., less than one) contributes to the hiding heritability.

As an empirical application of the proposed calculator, we estimate the SNP-based heritability and CGR of several polygenic traits across three distinct studies: the Rotterdam Study (RS), the Swedish Twin Registry (STR), and the Health and Retirement Study (HRS). For height, BMI, years of education, and self-rated health, we obtain point-estimates of CGR between 0.47 and 0.97, suggesting that even extremely large GWAS meta-analyses will fall short of explaining the full h^2_{SNP} for these traits. Using the MetaGAP calculator, we quantify the expected number of hits and the predictive accuracy of the PGS in recent GWAS efforts for these traits. Our theoretical predictions align with empirical observations. Finally, by comparing these figures to the predicted number of hits and PGS accuracy under perfect CGRs, we show that there is considerable attenuation due to imperfect CGRs; even for height (CGR point-estimate of 0.97) the expected relative loss in the number of hits is 8% and the relative loss in PGS R^2 is 6%.

Importantly, the MetaGAP calculator has two desirable properties compared to other calculators. In other calculators one often needs to specify some true value of the SNP effect (Evans & Purcell, 2012) (e.g., by taking the effect estimates of the most significant SNPs from an earlier GWAS, to which one first applies a 'winner's curse' correction; Sham & Purcell, 2014). Instead of requiring the input of an *a priori* unknown effect, our method incorporates a tacit assumption regarding the relation between allele frequency and effect size, such that each trait-affecting SNP has an equal R^2 with respect to the phenotype (e.g., Yang, Lee, et

al., 2011). Therefore, our method merely requires the h^2_{SNP} and the number of independent haplotype blocks harboring trait-affecting variation. The ratio of these two quantities fully specifies the proportion of phenotypic variance which can be explained by a 'representative' associated SNP. In addition, other calculators usually require not only the true effect of a SNP as input parameter but also the allele frequency (Evans & Purcell, 2012; Menashe, Rosenberg, & Chen, 2008; Purcell, Cherny, & Sham, 2003). By focusing on a 'representative' associated SNP, we also eliminate the allele frequency from our power calculator. In our simulations, we show that a violation of this equal-R^2 assumption hardly affects the quality of the predicted statistical power and PGS accuracy.

To summarize, the current study aims to formulate precise relations between genetic heterogeneity across studies on the one hand and statistical power and predictive accuracy, for a meta-analysis of GWAS results, on the other. This aim is achieved, and substantiated in the form of an online calculator, available at www.devlaming.eu, which accounts for the effect of genetic and phenotypic heterogeneity across studies. The calculator does not require *a priori* knowledge about the magnitude of the true association between the SNP and trait of interest. By means of this calculator, it can be shown to what degree CGR affects statistical power and predictive accuracy. By using the calculator to assess recent GWAS efforts, we show that the effect of CGR on statistical power and predictive accuracy is substantial. Hence, CGR explains a considerable part of the hiding heritability. Therefore, *a priori* calculations of statistical power and predictive accuracy, accounting for heterogeneity in genetic effects across studies, are an important tool for assessing whether an intended meta-analysis of GWAS results is likely to yield meaningful outcomes.

6.2 Materials and Methods

6.2.1 Definitions

In our framework, we consider only the SNP-based heritability, as estimated based on the set of SNPs of interest. In line with others, we define the effective number of SNPs, S, as the number of haplotype blocks (i.e., independent chromosome segments; Daetwyler, Villanueva, & Woolliams, 2008), where variation in each block is tagged by precisely one SNP. Hence, in our framework, there are S SNPs contributing to the polygenic score. Due to linkage disequilibrium this number is likely to be substantially lower than the total number of SNPs in the genome (M.-X. Li, Yeung, Cherny, & Sham, 2012), and is inferred to lie between as little as 60,000 (Wray et al., 2013) and as much as 5 million (M.-X. Li et al., 2012). In terms of trait-affecting variants, we consider a subset of M SNPs. Each SNP in this subset tags variation in a segment that bears a causal influence on the phenotype. We refer

to *M* as the associated number of SNPs. We assume that the *M* associated SNPs capture the full SNP-based heritability for the trait of interest.

6.2.2 Power of a GWAS meta-analysis under heterogeneity

Generic expressions for the theoretical distribution of the *Z* statistic, resulting from a meta-analysis of GWAS results under imperfect CGRs, can be found in Supplementary Information Section 1 in De Vlaming et al. (2016). For intuition, we here present the specific case of a meta-analysis of results from two studies with CGR ρ_G, with equal SNP-based heritability h^2_{SNP}, and equal sample sizes (i.e., *N* in Study 1 and *N* in Study 2). Under this scenario, we find that under high polygenicity, the *Z* statistic of an associated SNP *k* is normally distributed with mean zero and the following variance:

$$Var(Z_k) \approx 1 + \left(\frac{h^2_{SNP}}{M} \times N \times (1 + \rho_G) \right) \tag{6.1}$$

The larger the variance in the *Z* statistic, the higher the probability of rejecting the null. The ratio of h^2_{SNP} and *M* can be regarded as the theoretical R^2 of each associated SNP with respect to the phenotype. Equation (6.1) reveals that (i) when sample size increases, power increases, (ii) when h^2_{SNP} increases, the R^2 per associated SNP increases and therefore power increases, (iii) when the number of associated SNPs increases, the R^2 per associated SNP decreases and therefore power decreases, (iv) when the CGR is minus one, the studies perfectly cancel each other's genetic effects, thereby eliminating the power of the meta-analysis and reducing the distribution of the *Z* statistic for an associated SNP to a standard-normal distribution, yielding a strong disadvantage to meta-analyzing in this scenario, (v) when the CGR is zero the power of the meta-analysis is identical to the power obtained in each of the two studies when analyzed separately, yielding no strict advantage to meta-analyzing, and (vi) when the CGR is plus one the additional variance in the *Z* statistic relatively to the variance under the null is twice the additional variance one would have when analyzing the studies separately, yielding a strong advantage to meta-analyzing.

Others have focused on the highly related χ^2 statistics, defined as the squared *Z* statistics. In particular, it has been shown that the χ^2 statistics are influenced by linkage disequilibrium, population stratification, and polygenicity (Bulik-Sullivan, Loh, et al., 2015; Yang, Weedon, et al., 2011; Yang et al., 2014). Although we focus on CGR and how it affects *Z* statistics rather than the χ^2 statistics, the factors that appear in our expressions of the variance of the GWAS *Z* statistics are highly similar to the factors that appear in work aiming to dissect the

expected value of the GWAS χ^2 statistics. As an illustration of the similarity in expressions, consider the scenario where the CGR equals one between two samples of equal size. Based on Equation (6.1), we then have that $Var(Z_k) \approx 1 + (N_{total} \times h^2_{SNP}/M)$ for a trait-affecting haplotype block, where $N_{total} = 2N$. This expressions for the variance of the Z statistic of a trait-affecting haplotype block is completely equivalent to the expected χ^2 statistic from the linear regression analysis for a trait-affecting variant reported in Section 4.2 of the Supplementary Note to Yang et al. (2014) as well as Equation 1 in Bulik-Sullivan, Loh, et al. (2015) when assuming that confounding biases and linkage disequilibrium are absent. However, under a scenario with two or more studies with imperfect CGR, this overlap breaks down.

In order to compute statistical power in a multi-study setting, we first use the generic expression for the variance of the GWAS *Z* statistic derived in Supplementary Information Section 1 in De Vlaming et al. (2016) to characterize the distribution of the *Z* statistic under the alternative hypothesis. We then use the inverse normal cumulative distribution function to quantify the probability of attaining genome-wide significance for an associated SNP. This probability we refer to as the "power per associated SNP". Moreover, given that we use SNPs tagging independent haplotype blocks, we calculate the probability of rejecting the null for at least one of the associated SNPs and the expected number of independent hits as follows:

$$\text{power to detect at least one SNP} = 1 - [1 - (\text{power per associated SNP})]^{M} \quad \text{and}$$

$$\text{E[number of hits]} = M \times (\text{power per associated SNP}).$$

6.2.3 R^2 of a polygenic score under heterogeneity

We derive a generic expression for the theoretical R^2 of a PGS in a hold-out sample, with SNP weights based on a meta-analysis of GWAS results under imperfect CGRs. We consider a PGS that includes all the SNPs that tag independent haplotype blocks (i.e., there is no SNP selection).

The derivations can be found in the Supplementary Information Section 2 in De Vlaming et al. (2016). For intuition, we here present an approximation for prediction in a hold-out sample, with SNP weights based on a GWAS in a single discovery study with sample size *N*, where both studies have SNP-heritability h^2_{SNP}, and with CGR ρ_G, between the studies. Under high polygenicity, the R^2 of the PGS in the hold-out sample is then given by the following expression:

$$R^2 \approx h^2_{SNP} \times \rho_G \times \frac{h^2_{SNP}}{\frac{S}{N} + h^2_{SNP}} \tag{6.2}$$

In case the CGR is one, and we consider the R^2 between the PGS and the genetic value (i.e., the genetic component of the phenotype) instead of the phenotype itself, the first two terms in Equation (6.2) disappear, yielding an expression equivalent to Equation 1 in Daetwyler et al. (2008). Assuming a CGR of one and that all SNPs are associated, Equation (6.2) is equivalent to the expression in Dudbridge (2013) for the R^2 between the PGS and the phenotype in the hold-out sample.

From Equation (6.2), we deduce that (i) as the effective number of SNPs S increases, the R^2 of the PGS deteriorates (since every SNP-effect estimate contains noise, owing to imperfect inferences in finite samples), (ii) given the effective number of SNPs, under a polygenic architecture, the precise fraction of effective SNPs that is associated does not affect the R^2, (iii) R^2 is quadratically proportional to ρ_G, implying a strong sensitivity to CGR, and (iv) as the sample size of the discovery study grows, the upper limit of the R^2 is given by $h^2_{SNP} \times \rho_G$, implying that the full SNP-heritability in the hold-out sample cannot be entirely captured so long as CGR is imperfect.

6.2.4 Online power and R^2 calculator

An online version of the MetaGAP calculator can be found at www.devlaming.eu. This calculator computes the theoretical power per trait-affecting haplotype block, the power to detect at least one of these blocks, and the expected number of independent hits for a meta-analysis of GWAS results from C studies. In addition, it provides the expected R^2 of a PGS for a hold-out sample, including all GWAS SNPs, with SNP-weights based on the meta-analysis of the GWAS results from C studies. Calculations are based on the generic expressions for GWAS power derived in Supplementary Information Section 1 and PGS R^2 derived in Supplementary Information Section 2 in De Vlaming et al. (2016).

The calculator assumes a quantitative trait. Users need to specify either the average sample size per study or the sample size of each study separately. In addition, users need to specify either the average SNP-heritability across studies or the SNP-heritability per study. The SNP-heritability in the hold-out sample also needs to be provided. Users are required to enter the effective number of causal SNPs and the effect number of SNPs in total. The calculator assumes a fixed CGR between all pairs of studies included in the meta-analysis and a fixed

CGR between the hold-out sample and each study in the meta-analysis. Hence, one needs to specify two CGR values: one for the CGR within the set of meta-analysis studies and one to specify the genetic overlap between the hold-out sample and the meta-analysis studies.

Finally, a more general version of the MetaGAP calculator is provided in the form of `MATLAB` code, also available at www.devlaming.eu. This code can be used in case one desires to specify a more versatile genetic-correlation matrix, where the CGR can differ between all pairs of studies. Therefore, this implementation requires the user to specify a full (C+1)-by-(C+1) correlation matrix. Calculations in this code are fully in line with the generic expressions in Supplementary Information Sections 1 and 2 in De Vlaming et al. (2016).

6.2.5 Assessing validity of theoretical power and R^2

We simulate data for a wide range of genetic architectures in order to assess the validity of our theoretical framework. The theoretical expressions we derive for power and R^2 are accurate, even for data generating processes substantially different from the process we assume in our derivations (for details, see Supplementary Information Section 3 in De Vlaming et al., 2016). Our strongest assumption is that SNPs have equal R^2 with respect to the phenotype regardless of allele frequency. When we simulate data where this assumption fails and where allele frequencies are non-uniformly distributed, the root-mean-square prediction error of statistical power lies below 3% and that of PGS R^2 below 0.12%.

6.2.6 Estimating SNP-heritability and CGR

Using 1000 Genomes-imputed (1kG) data from the RS, STR, and HRS, we estimate SNP-based heritability and CGR respectively by means of univariate and bivariate genomic-relatedness-matrix restricted maximum likelihood (GREML) as implemented in `GCTA` (S. H. Lee, Yang, et al., 2012; Yang, Lee, et al., 2011). In our analyses, we consider the subset of HapMap3 SNPs available in the 1kG data. In Section 6.5.1 we report details on the genotype and phenotype data, as well as our quality control (QC) procedure. After QC, we have a data set consisting of ≈ 1 million SNPs and ≈ 20,000 individuals, from which we infer h^2_{SNP} and CGR. In Section 6.5.2 we provide details on the specifications of the models used for GREML estimation.

6.3 Results

6.3.1 Determinants of GWAS power and PGS R^2

Using the MetaGAP calculator, we assessed the theoretical power of a meta-analysis of GWAS results from genetically heterogeneous studies and the theoretical R^2 of the resulting PGS in a hold-out sample, for various numbers of studies and sample sizes, and different values of CGR and h^2_{SNP}.

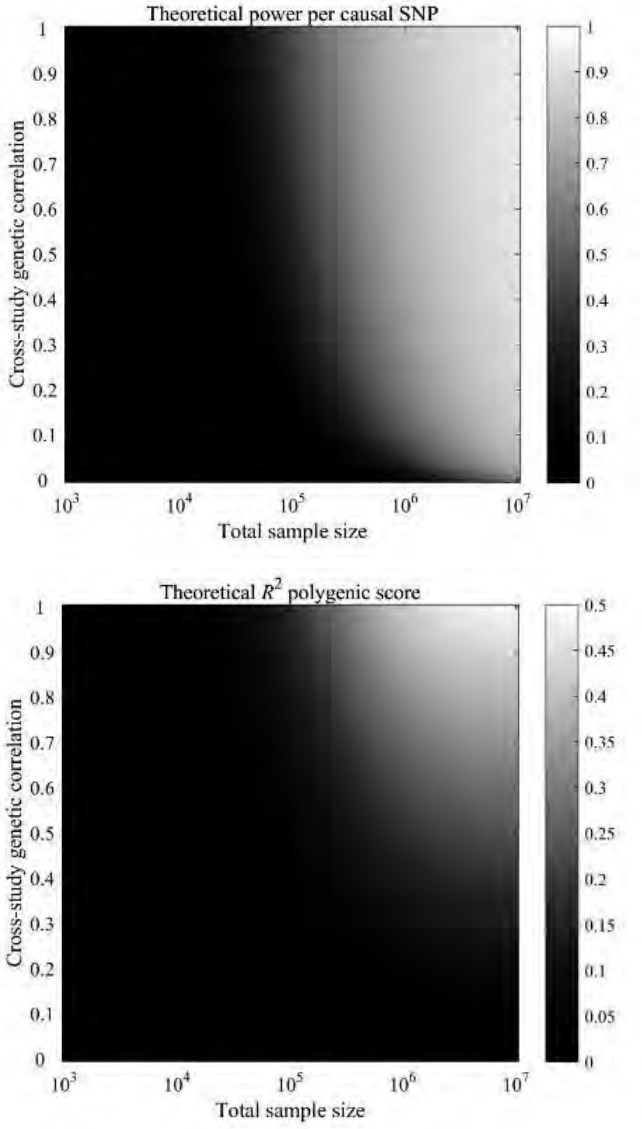
A. SAMPLE SIZE AND CGR

Figure 6.1 shows heat maps for the power per truly associated SNP and R^2 for a setting with 50 studies, for a trait with h^2_{SNP} = 50%, for various combinations of total sample size and CGR. Increasing total sample size enhances both power and R^2. When the CGR is perfect, power and R^2 (relative to h^2_{SNP}) have a near-identical response to sample size. This similarity in response gets distorted when the CGR decreases. For instance, in the scenario of 100k SNPs of which a subset of 1k SNPs is causal with h^2_{SNP} = 50%, in a sample of 50 studies with a total sample size of 10 million individuals, a CGR of one yields 94% power per causal SNP and an R^2 of 49%, which is 98% of the SNP-heritability, whereas for a CGR of 0.2 the power is still 87% per SNP, while the R^2 of the PGS is 8.5%, which is only 17% of h^2_{SNP}. Thus, R^2 is far more sensitive to an imperfect CGR than the meta-analytic power is. This finding is also supported by the approximations of power in Equation (6.1) and of PGS R^2 in Equation (**6.2**); these expressions show that, for two discovery studies, the CGR has a linear effect on the variance of the meta-analysis Z statistic, whereas, for one discovery and one hold-out sample, the PGS R^2 is quadratically proportional to the CGR.

B. SNP-HERITABILITY AND CGR

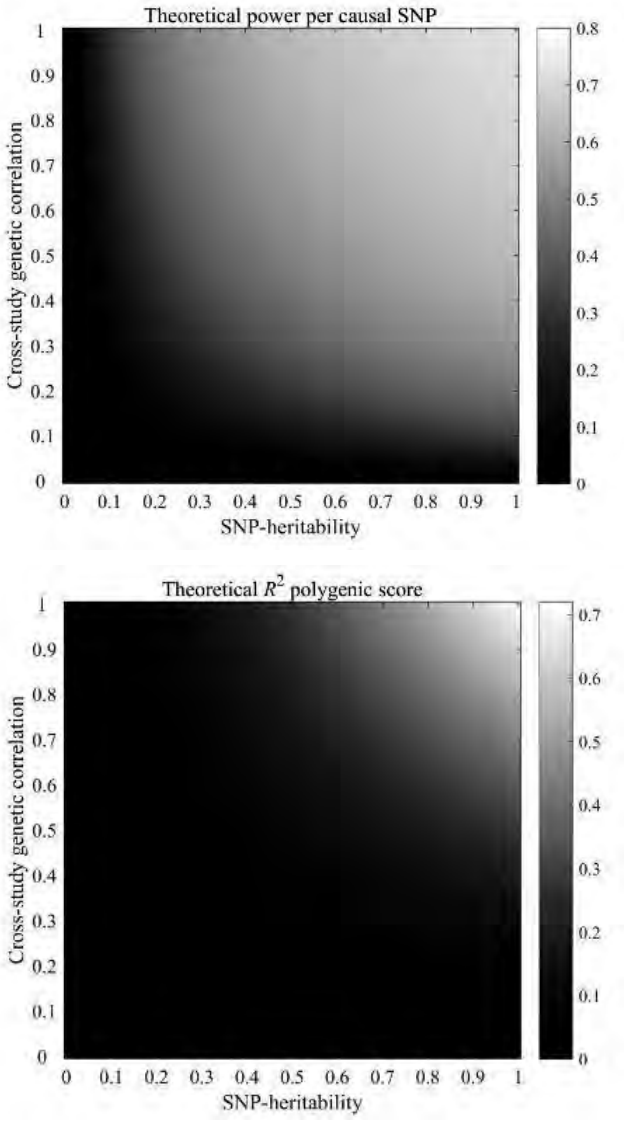
Figure 6.2 shows heat maps for the power per truly associated SNP and R^2 for a setting with 50 studies, with a total sample size of 250,000 individuals, for 1k causal SNPs and 100k SNPs in total, for various combinations of h^2_{SNP} and CGR. The figure shows a symmetric response of both power and R^2 to CGR and h^2_{SNP}. For instance, when h^2_{SNP} = 25% and CGR = 0.5 across all studies, the power is expected to be around 34% and the R^2 around 3.0%. When these numbers are interchanged (i.e., h^2_{SNP} = 50% and CGR = 0.25), similarly, the power is expected to be 35% and the R^2 around 2.9%. Hence, in terms of both R^2 and power,

Figure 6.1. Theoretical predictions of power per causal SNP (upper panel) and out-of-sample R^2 of the PGS (lower panel), as a function of for total sample size and cross-study genetic correlation.



Note: Factor levels are 50 studies, 100k independent SNPs, and heritability $h^2_{SNP} = 50\%$ arising from a subset of 1k independent SNPs.

Figure 6.2. Theoretical predictions of power per causal SNP (upper panel) and out-of-sample R^2 of the PGS (lower panel) as a function of SNP-heritability and cross-study genetic correlation.



Note: Factor levels: 50 studies, sample size 5,000 individuals per study, 100k independent SNPs, and heritability arising from a subset of 1k independent SNPs.

a low heritability can be compensated by a high CGR (e.g., by means of homogeneous measures across studies) and a low CGR can be compensated by high heritability.

When looking at two points with the same power (resp. R^2), any other point on a straight line between these points has a higher power (R^2), than at the end-points of the line. For instance, when both h^2_{SNP} and CGR lie at the midpoint between the 0.25 and 0.5 considered before (i.e., h^2_{SNP} = 37.5% and CGR = 0.375), the expected power is 37% > 35% and the expected R^2 is 3.6% > 3.0%.

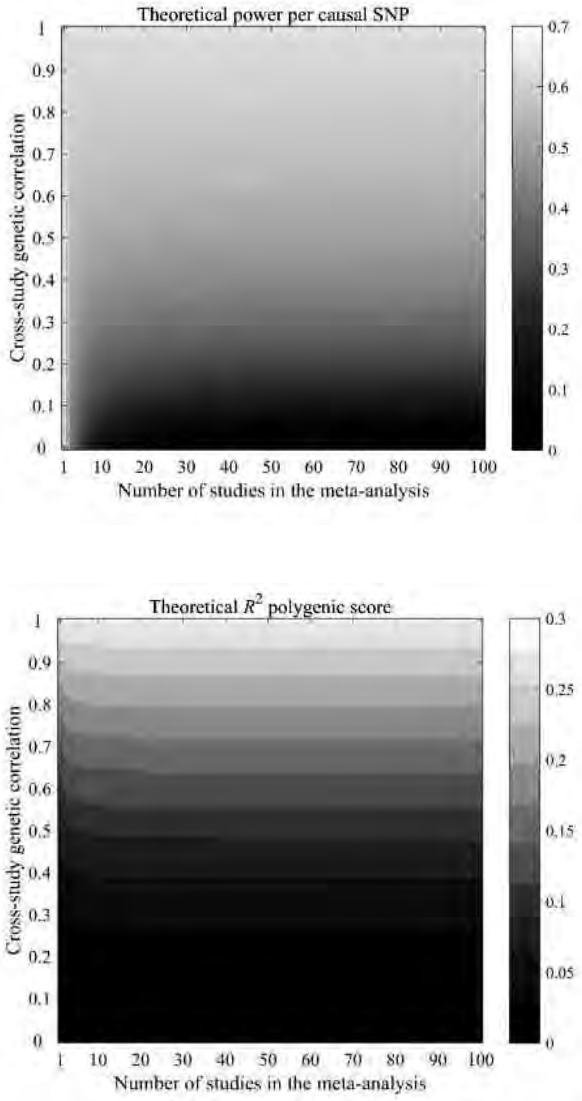
When either CGR or heritability is (close to) zero, both power and R^2 are decimated in the multi-study setting. Hence, least power and R^2 can be found when h^2_{SNP}, CGR, or both are low. However, when both are moderately low but still substantially greater than zero, neither power nor R^2 are completely diminished.

C. NUMBER OF STUDIES AND CGR

Figure 6.3 shows heat maps for the power per truly associated SNP and R^2 for a trait with h^2_{SNP} = 50%, 1k causal SNPs, 100k SNPs in total, and a fixed total sample size of 250,000 individuals. In this figure, various combinations of the number of studies and CGR are considered. A discontinuous color map is used for R^2 to make salient details visible. Logically, when there is just one study for discovery, CGR does not affect power. However, even for two studies, the effect of CGR on power is quite pronounced. For instance, when CGR is a half, the power per causal SNP is 63% for one study, 58% for two studies, 51% for ten studies, and 50% for 100 studies. Thus, when the number of studies is low, increases in the number of studies make the effect of CGR on power more pronounced rapidly. When the number of studies is large, increases in the number of studies hardly make the effect of CGR on power more pronounced.

For a given number of studies, we observed that the effect CGR has on R^2 is stronger than the effect it has on power. This observation is in line with the approximated theoretical R^2 in Equation (**6.2**, indicating that R^2 is quadratically proportional to CGR. However, an interesting observation is that this quadratic relation lessens as the number of studies grows large, despite the total sample size being fixed. For instance, at a CGR of a half, the R^2 in the hold-out sample is expected to be 6.9% when there is only one discovery study. However, the expected R^2 is 8.1% for two discovery studies, 9.3% for ten discovery studies, and 9.6% for 100 discovery studies. The reason for this pattern is that, in case of one discovery study, the PGS is influenced relatively strongly by the study-specific component of the genetic effects.

Figure 6.3. Theoretical predictions of power per causal SNP (upper panel) and out-of-sample R^2 of the PGS (lower panel) as a function of the number of studies in the meta-analysis and cross-study genetic correlation.



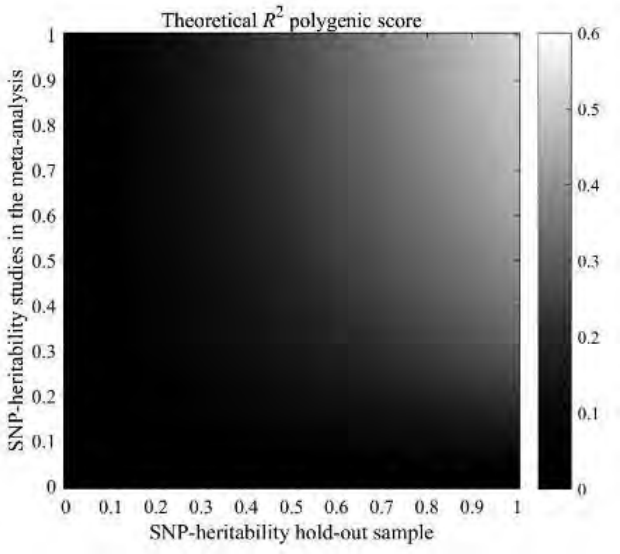
Note: Factor levels: total sample size 250,000 individuals, 100k independent SNPs, and heritability h^2_{SNP}=50% arising from a subset of 1k independent SNPs. For R^2, a discontinuous color map is used to make salient details visible.

This idiosyncrasy is not of relevance for the hold-out sample. As the number of studies increases, even though each study brings its own idiosyncratic contribution, each study also consistently conveys information about the part of the genetic architecture which is common across the studies. Now, since the idiosyncratic contributions from the studies are independent, they tend to average each other out, whereas the common underlying architecture gets more pronounced as the number of studies in the discovery increases, even if the total sample size is fixed.

D. SNP-HERITABILITY IN THE HOLD-OUT SAMPLE

Figure 6.4 shows a heat map for the PGS R^2 based on a meta-analysis of 50 studies with a total sample size of 250,000 individuals, with 1k causal SNPs and 100k SNPs in total, and a CGR of 0.8 between both the discovery studies and the hold-out sample. In the heat maps various combinations of h^2_{SNP} in the discovery samples and h^2_{SNP} in the hold-out sample are considered. The response of PGS R^2 to heritability in the discovery sample and the hold-out sample is quite symmetric, in the sense that a low h^2_{SNP} in the discovery samples and a high h^2_{SNP} in the hold-out sample yield a similar R^2 as a high h2SNP in the discovery sample and a low h^2_{SNP} in the hold-out sample. However, overall R^2 is slightly more sensitive to h^2_{SNP} in

Figure 6.4. Theoretical predictions of out-of-sample R^2 of the PGS as a function of SNP-heritability in the hold-out and discovery samples.



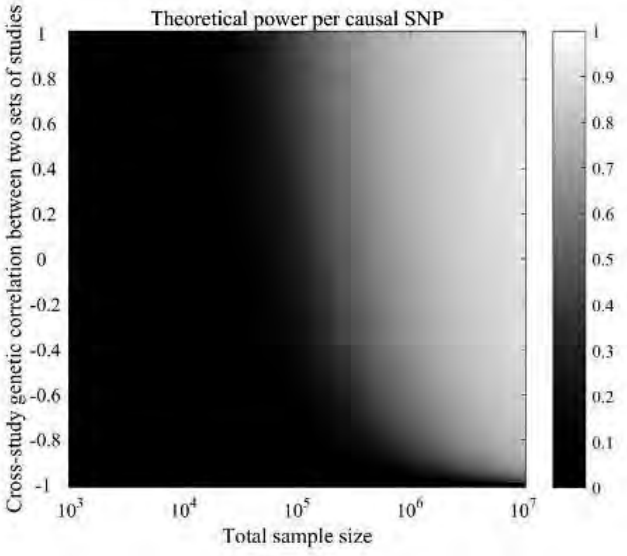
Note: Factor levels: 50 studies, sample size 5,000 individuals per study, cross-study genetic correlation 0.8, 100k independent SNPs, and heritability arising from a subset of 1k independent SNPs.

the hold-out sample than in the discovery samples. For instance, when SNP-heritability in the discovery samples is 50% and 25% in the hold-out sample, the expected R^2 is 10%, whereas in case the SNP-heritability is 25% in the discovery samples and 50% in the hold-out sample, the expected R^2 is 13%.

E. CGR BETWEEN SETS OF STUDIES

Figure 6.5 shows a heat map for the power per truly associated SNP in a setting where there are two sets consisting of 50 studies each. Within each set, the CGR is equal to one, whereas between sets the CGR is imperfect. Consider, for example, a scenario where one wants to meta-analyze GWAS results for height from a combination of two sets of studies; one set of studies consisting primarily of individuals of European ancestry and one set of studies with mostly people of Asian ancestry in it. Now, one would expect CGRs close to one between studies consisting primarily of individuals of European ancestry and the same for the CGRs

Figure 6.5. Theoretical predictions of power per causal SNP as a function of total sample size and CGR between two sets of studies.



Note: Factor levels: 2 sets of 50 studies, CGR equal to 1 within both sets, 100k independent SNPs, and heritability h^2_{SNP}=50% arising from a subset of 1k independent SNPs.

between studies consisting primarily of people of Asian ancestry. However, the CGRs between those two sets of studies might be lower than one, though probably greater than zero.

If CGR between the two sets is zero, meta-analyzing the two sets jointly is sub-optimal; the power of such a meta-analysis lies in between the power obtained by either of these sets when meta-analyzed separately (De Vlaming et al., 2016, Supplementary Information Section 1). Since in Figure 6.5 we considered two equally-powered sets, the power of a meta-analysis using both sets, under zero CGR between sets, is identical to the power obtained when meta-analyzing, for instance, only the first set. However, as CGR between sets increases so does power. For instance, when a total sample size of 250,000 individuals is spread across 2 clusters, each cluster consisting of 50 studies (i.e., sample size of 125,000 individuals per cluster and 2,500 individuals per study), under h^2_{SNP} = 50% due to 1k causal SNPs, a CGR of one within each cluster, and CGR of zero between clusters, the power is expected to be 49%, which is identical to the power of a meta-analysis of either the first or the second cluster. However, if the CGR between clusters is 0.5 instead of zero, the power goes up to 58%. In terms of the expected number of hits, this cross-ancestry meta-analysis yields an expected 82 additional hits, compared to a meta-analysis considering only one ancestry.

Alternatively, one could pool hits from two meta-analyses (e.g., in our example one in the European-ancestry set and one in the Asian-ancestry set). However, this would imply more independent tests being carried out, and, hence, the need for a stronger Bonferroni correction in order to keep the false-positive rate fixed, and, thus, a more stringent genome-wide significance threshold. Therefore, this route is likely to yield less statistical power than a meta-analysis of merely one of the set of two or a joint analysis of both sets.

6.3.2 Empirical results for SNP-based heritability and CGR

In Table 6.1, we report univariate GREML estimates of SNP-heritability and bivariate GREML estimates of genetic correlation for traits that attained a pooled sample size of at least 18,000 individuals, which gave us at least 50% power to detect a genetic correlation near one for a trait that has a SNP-heritability of 10% or more (Visscher et al., 2014). The smallest sample size is N = 19,184 for self-rated health. Details per phenotype (i.e., sample size, univariate estimates of SNP-heritability, and bivariate estimates of genetic correlation, stratified across studies and sexes, as well as cross-study and cross-sex averages) are provided in Table 6.6 - Table **6.7** in Section 6.5.

Table 6.1. GREML estimates of SNP-heritability (h^2_{SNP}) and genetic correlation across studies and sexes.

Phenotype	N	Estimates SNP-heritability			Estimates genetic correlation			
		pooled	study	sexes	RS - STR	RS - HRS	STR – HRS	Females - Males
Height	20,458	43.3% (1.8%)***	44.90%	44.00%	0.976 (0.102)***	0.954 (0.095)***	0.967 (0.106)***	0.981 (0.067)***
BMI	20,449	20.9% (1.7%)***	21.90%	22.80%	1.000 (0.269)***	0.914 (0.172)***	0.847 (0.246)***	0.794 (0.122)***†
EduYears	20,619	16.4% (1.7%)***	18.20%	18.40%	0.690 (0.233)***	0.659 (0.224)***†	1.000 (0.263)***	0.832 (0.162)***
CurrCigt	20,686	18.2% (4.0%)***	19.10%	24.20%	1.000 (0.643)***	0.611 (0.448)*	1.000 (0.607)***	0.543 (0.257)***†
CurrDrinkFreq	20,072	7.00% (2.6%)***	10.30%	8.30%	1.000 (0.666)***	0.298 (0.670)	-0.056 (0.647)	1.000 (2.068)*
Self-rated health	19,184	10.3% (1.8%)***	15.70%	9.50%	0.626 (0.439)**	0.363 (0.223)**††	0.447 (0.278)**	1.000 (0.349)***

Note: Standard errors between parentheses. “pooled”: univariate estimate from pooled data. “study”: sample-size weighted average of univariate estimates across studies. “sexes”: sample-size weighted average of univariate estimates across sexes.
* h^2_{SNP} and/or genetic correlation > 0 at 10% sign. † genetic correlation < 1 at 10% sign.
** h^2_{SNP} and/or genetic correlation > 0 at 5% sign. †† genetic correlation < 1 at 5% sign.
*** h^2_{SNP} and/or genetic correlation > 0 at 1% sign. †††genetic correlation < 1 at 1% sign.

The univariate estimates of SNP-heritability based on the pooled data assume perfect CGRs. Therefore, such estimates of SNP-heritability are downwards biased when based on data from multiple studies with imperfect CGRs. To circumvent this bias, we estimated SNP-heritability in each study separately, and focused on the sample-size-weighted cross-study average estimate of SNP-heritability.

For both height and BMI, we observed genetic correlations close to one across pairs of studies and between females and males. For years of schooling (*EduYears*) we found a CGR around 0.8 when averaged across pairs of studies. Similarly, the genetic correlation for *EduYears* in females and males lies around 0.8. The CGR of self-rated health is substantially below one across the pairs of studies, whilst the genetic correlation between females and males seems to lie around one. The reason for this difference in the genetic correlation between pairs of studies and between females and males may be due to the difference in the questionnaire across studies, discussed in Section 6.5.1. The questionnaire differences can yield a low CGR, while not precluding the remaining genetic overlap for this measure across the three studies, to be highly similar for females and males. For *CurrCigt* and *CurrDrinkFreq*, the estimates of CGR and of genetic correlation between females and males are non-informative. For these two traits the standard errors of the genetic correlations estimates are large, mostly greater than 0.5. In addition, for *CurrDrinkFreq* there is strong volatility in the CGR estimate across pairs of studies.

6.3.3 Attenuation in power and R^2 due to imperfect CGR

Considering only the traits for which we obtained accurate estimates of CGR and SNP-heritability (i.e., with low standard errors), we used the MetaGAP calculator to predict the number of hits in a set of discovery samples and the PGS R^2 in a hold-out sample, in prominent GWAS efforts for these traits.

Since we only had accurate estimates for height, BMI, *EduYears*, and self-rated health, we focused on these four phenotypes. For these traits, we computed sample-size-weighted average CGR estimates across the pairs of studies. Table 6.2 shows the number of hits and PGS R^2 reported in the most comprehensive GWAS efforts to date for the traits of interest, together with predictions from the MetaGAP calculator. We tried several values for the number of independent haplotype blocks (i.e., 100k, 150k, 200k, 250k) and for the number of trait-associated blocks (i.e., 10k, 15k, 20k, 25k). Overall, 250k blocks of which 20k trait-affecting yielded theoretical predictions in best agreement with the empirical observations; we acknowledge the potential for some overfitting (i.e., two free parameters set on the basis of 17 data points; 10 data points for the reported number of hits and 7 for PGS R^2). For height

Table 6.2. Predicted and observed number of genome-wide-significant hits and PGS R^2, for large-scale GWAS efforts to date for height, BMI, EduYears, and self-rated health.

Phenotype	Main studies			Architecture		Number of hits				PGS R2 using all SNPs			
	Study	*N*	C**	h^2_{SNP}	CGR	Study	Theory | CGR <1	CGR =1	Attenua-tion*	Study	Theory | CGR <1	CGR =1	Attenua-tion*
Height	Wood et al. (2014)	253,288	79	44.90%	0.965	697	647.26	700.24	8%	13.50%	13.2%	14.0%	6%
	Allen et al. (2010)	183,727	61	44.90%	0.965	180	292.03	320.77	9%	10.00%	10.5%	11.1%	6%
	Weedon et al. (2008)	13,665	5	44.90%	0.965	7	0.00	0.00	n.a.	***2.9%	1.0%	1.1%	7%
BMI	Locke et al. (2015)	339,224	125	21.90%	0.917	97	188.52	241.07	22%	6.50%	4.3%	5.0%	14%
	Speliotes et al. (2010)	123,865	46	21.90%	0.917	19	5.48	7.64	28%	2.50%	1.8%	2.1%	15%
	Willer et al. (2008)	32,387	15	21.90%	0.917	1	0.01	0.02	65%	n.a.	0.5%	0.6%	16%
EduYears	Okbay et al. (2016)	405,072	65	18.20%	0.783	162	115.28	235.90	51%	n.a.	2.7%	4.1%	36%
	Okbay et al. (2016)	293,723	64	18.20%	0.783	74	39.30	88.93	56%	3.20%	2.0%	3.2%	36%
	Rietveld et al. (2013)	101,069	42	18.20%	0.783	1	0.63	1.64	62%	2.50%	0.8%	1.2%	38%
Self-rated health	Harris et al. (2015)	111,749	1	15.70%	0.468	13	1.35	1.35	0%	n.a.	0.2%	1.0%	78%

Note: Predicted number of genome-wide-significant hits and PGS R^2were calculated using averaged GREML estimates from Table 6.1 as SNP-heritability and CGR and assuming 250k effective SNPs (i.e. independent haplotype blocks) of which 20k trait-affecting. *Attenuation measures the relative loss in expected power and R^2 due to a CGR in accordance with averaged GREML estimates from Table 6.1. **C denotes the number of studies in the meta-analysis; C is slightly subjective (e.g., RS I, II, and III can be considered as one study or as three). ***Based on 20 SNPs.

– the trait with the lowest standard error in the estimates of h^2_{SNP} and CGR – the predictions of the number of hits and PGS R^2 for the two largest GWAS efforts are much in line with theoretical predictions. For the smaller GWAS of 13,665 individuals (Weedon et al., 2008), our estimates seem somewhat conservative; 0 hits expected versus the 7 reported.

However, in our framework, we assumed that each causal SNP has the same R^2. Provided there are some differences in R^2 between causal SNPs, especially in smaller samples, the first SNPs that are likely to reach genome-wide significance are the ones with a comparatively large R^2. This view is supported by the fact that a PGS based on merely 20 SNP explains 2.9% of the variation in height.

The notion of a GWAS first picking up the SNPs with a relatively high R^2 is also supported by the predicted and observed number of hits for the reported self-rated-health GWAS (Harris et al., 2015); given a SNP-heritability estimate between 10% and 16% (Table 6.2), according to our theoretical predictions, a GWAS in a sample of around 110k individuals is unlikely to yield even a single genome-wide significant hit. However, this GWAS has yielded 13 independent hits. This finding supports the view that some relatively-high-R^2 SNPs are present in the genome.

For BMI, our predictions of PGS R^2 were quite in line with empirical results. However, for the number of hits, our predictions for the largest efforts seemed overly optimistic. We therefore suspect that the number of independent SNPs associated with BMI is higher than 20k; as a higher number of associated SNPs would reduce the GWAS power, while preserving PGS R^2, yielding good agreement with empirical observation. Nevertheless, given the limited number of data points, this strategy of setting the number of causal SNPs would increase the chance of overfitting.

For *EduYears* we observed that the reported number of hits is in between the expected number of hits when the CGR is set to the averaged GREML estimate of 0.783 and when the CGR is set to one. Given the standard errors in the CGR estimates for *EduYears*, the CGR might very well be somewhat greater than 0.783, which would yield a good fit with the reported number of hits. However, as with the number of truly associated SNPs for BMI, we can make no strong claims about a slightly higher CGR of *EduYears* due to the risk of overfitting.

Overall, our theoretical predictions of the number of hits and PGS R^2 are in moderate agreement with empirical observations, especially when bearing in mind that we are looking at a

limited number of data points, making chance perturbations from expectation likely. In addition, regarding the number of hits, the listed studies are not identical in terms of the procedure to obtain the independent hits. Therefore, the numbers could have been slightly different, had the same pruning procedure been used across all reported studies. Such differences in procedures introduce an additional element of chance.

Regarding attenuation, we observed a substantial spread in the predicted number of hits and PGS R^2 when assuming either a CGR of one, or a CGR in accordance with empirical estimates, with traits with lower CGR suffering from stronger attenuation in power and predictive accuracy. In line with theory, R^2 falls sharply with CGR. For instance, for self-rated health, the estimate CGR of about 0.5, would – in expectation – yield a PGS that retains only 0.5^2=25% of the R^2 it would have had under a CGR of one. This is supported by the reported attenuation of roughly 80%.

Given our CGR estimates, we expect a relative loss in PGS R^2 of 6% for height, 14% for BMI, 36% for *EduYears*, and 78% for self-rated health, compared to the R^2 of a PGS under perfect CGRs (Table 6.2). This loss in R^2 is unlikely to be reduced by larger sample sizes and denser genotyping.

Somewhat contrary to expectation, the number of hits seems to respond even more strongly to CGR than PGS R^2. However, since in each study under consideration the average power per associated SNP is quite small, a small decrease in power per SNP in absolute terms can constitute a substantial decrease in relative terms. For instance, when one has 2% power per truly associated SNP, an absolute decrease of 1% – leaving 1% power – constitutes a relative decrease of 50% of power per causal SNP, and thereby a 50% decrease in the expected number of hits. This strong response shows, for example, in the case of *EduYears*, where the expected number of hits drop by about 37% when going from a CGR of one down to a CGR of 0.783.

6.4 Discussion

In this study we aimed to answer the question whether imperfect cross-study genetic correlations (CGRs) help to explain a part of the 'hiding' heritability for highly polygenic traits such as height. We showed that imperfect CGRs are indeed likely to contribute to the gap between the phenotypic variation accounted for by all SNPs jointly and by the leading GWAS efforts to date. We arrive at this conclusion in five steps.

First, we developed a Meta-GWAS Accuracy and Power (MetaGAP) calculator that accounts for the CGR. This online calculator relates the statistical power to detect associated SNPs and the R^2 of the polygenic score (PGS) in a hold-out sample to the number of studies, sample size and SNP-heritability per study, and the CGR. The underlying theory shows that there is a quadratic response of the PGS R^2 to CGR. Moreover, we showed that the power per associated SNP is also influenced by CGR, although – in absolute terms – not as strongly as the PGS R^2.

Second, we used simulations to demonstrate that our theory is robust to several violations of the assumptions about the underlying data-generating process, regarding the relation between allele frequency and effect size, as well as the distribution of allele frequencies. Further research needs to assess whether out theoretical predictions are also accurate under an even broader set of scenarios (e.g., when studying a binary trait or when studying a trait for which there are relatively many rare variants with relatively small effects).

Third, we used a sample of unrelated individuals from the Rotterdam Study, the Swedish Twin Registry, and the Health and Retirement Study, to estimate SNP-based heritability as well as the CGR for traits such as height and BMI. Although our CGR estimates have considerable standard errors, the estimates make it likely that for many polygenic traits the CGR is positive, albeit smaller than one.

Fourth, based on these empirical estimates of SNP-heritability and CGR for height, BMI, years of education, and self-rated health, we used the MetaGAP calculator to predict the number of expected hits and the expected PGS R^2 for the most prominent studies to date for these traits. We found that our predictions are in good agreement with empirical observations. Although our theory turned out to be somewhat conservative for smaller GWAS samples, for large-scale GWAS efforts our predictions were in line with the outcomes of these efforts.

Fifth, we used our theoretical model to assess statistical power and predictive accuracy for these GWAS efforts, had the CGR been one for the traits under consideration. Our estimates of power and predictive accuracy in this scenario indicated a strong decrease in the PGS R^2 and the expected number of hits, due to imperfect CGRs. Though these observations are in line with expectation for predictive accuracy, for statistical power the effect was larger than we anticipated. This finding can be explained, however, by the fact that though the absolute decrease in power per SNP is small, the relative decrease is large, since the statistical power per associated SNP is often low to begin with.

Overall, our study affirms that although PGS accuracy improves substantially with further increasing sample sizes, in the end PGS R^2 will continue to fall short of the full SNP-based heritability. Hence, this study contributes to the understanding of the hiding heritability reported in the GWAS literature.

Regarding the etiology of imperfect CGRs, the likely reasons are heterogeneous phenotype measures across studies, gene–environment interactions with underlying environmental factors differing across studies, and gene–gene interactions where the average effects differ across studies due to differences in allele frequencies. Our study is not able to disentangle these different causes; by estimating the CGR for different traits we merely quantify the joint effect these three candidates have on the respective traits.

However, in certain situations it is possible to disentangle the etiology of imperfect CGRs to some extent. For instance, in case one considers a specific phenotype that is usually studied by means of a commonly available but relatively heterogeneous and/or noisy measure, while there also exists a less readily available but more accurate and homogeneous measure. If one has access to both these measures in several studies, one can compare the CGR estimates for the more accurate measure and the CGR estimates for the less accurate but more commonly available measure. Such a comparison would help to get some sense of the relatively contribution of phenotype heterogeneity to imperfect CGR in the heterogeneous measure.

In considering how to properly address imperfect CGRs, it is important to note that having a small set of large studies, rather than a large set of small studies, does not by definition abate the problem of imperfect genetic correlations. Despite the fact that having less studies can help to reduce the effects of heterogeneous phenotype measures, larger studies are more likely to sample individuals from different environments. If gene–environment interactions do play a role, strong differences in environment between subsets of individuals in a study lead to imperfect genetic correlations within that study. The attenuation in power and accuracy resulting from the imperfect genetic correlations within studies may prove hard to address.

In addition to studying the reduction in power and predictive accuracy due to CGR, we used our theoretical framework to consider other factors influencing power and accuracy. We found that in terms of power, sample size trumps a lot, even a relatively low CGR. Moreover, we observed – in line with our theoretical framework – that PGS R^2 is far more sensitive to CGR than absolute power per SNP. Also, we found that low CGR can to some extent be leveraged by a high SNP-heritability and vice versa. However, it is better to have both at a

moderate level than one extremely high and the other extremely low; if either is zero the meta-analysis approach will fail.

We observed – given a fixed total sample size – that the substantial effects of CGR on power and predictive accuracy arise even for as few as two studies. Moreover, the CGR-power and CGR-accuracy relations do not change much as the number of underlying studies keeps increasing. This finding is reassuring; given that some power in the meta-analysis is lost due to imperfect CGRs, whether the underlying data is then highly fractured into many small studies or into a few big ones does not really matter for predictive accuracy or statistical power.

For SNP-heritability in the discovery samples and in the hold-out samples, we found that the PGS accuracy is slightly more affected by SNP-heritability in the hold-out sample than in the discovery samples. Hence, when aiming at high PGS accuracy, we recommend to use the study with the highest SNP-heritability and the highest CGR with the discovery samples as hold-out sample.

In addition to the number of studies, sample size and SNP-heritability per study, and CGR, our theoretical model depends on the specification of the following two latent parameters: the number of independent haplotype blocks (i.e., the 'effective number of SNPs') and the number of blocks containing trait-affecting variation (i.e., the number of independent 'causal' SNPs). In our work, setting the independent number of blocks at 250k and the number of trait-affecting blocks at 20k for all traits yielded the most accurate predictions.

Regarding the response of PGS accuracy and statistical power to these two parameters, it is interesting to note that our equations point to strongly opposed responses. Since effect sizes tend to decrease with an increasing number of causal SNPs, the statistical power decreases as the number of causal SNPs increases. The PGS R^2, on the other hand, decreases with the effective number of SNPs, since each SNP in the prediction model contributes some noise. By applying SNP-selection methods in the construction of a PGS, one can reduce the number of SNPs entering the PGS, decreasing the amount of noise and improving R^2. However, such methods may also exclude associated regions, decreasing the amount of signal in the score and attenuating R^2. Hence, SNP-selection methods are only likely to improve PGS R^2 when the selection is based on sufficiently accurate inferences.

Finally, having shown the substantial effect of imperfect CGRs on GWAS power and PGS R^{22}, we believe that the online MetaGAP calculator will prove to be an important tool for

assessing whether an intended meta-analysis of GWAS results from different studies, is likely to yield meaningful outcomes.

6.5 Supplementary Methods

6.5.1 Data and quality control

A. GENOTYPE DATA

In the bivariate and univariate genomic-relatedness-matrix restricted maximum likelihood (GREML) analyses we use genotype data from the Rotterdam Study (RS; Ergo waves 1-4 sample denoted by RS-I, Ergo Plus sample denoted by RS-II, and Ergo Jong sample denoted by RS-III), the Swedish Twin Registry (STR; TwinGene sample), and the Health and Retirement Study (HRS). For each study, details on the genotyping platform, quality control (QC) prior to imputation, the reference sample used for imputation, and imputation software, are listed in Table 6.3.

To increase the overlap of SNPs across studies, we use genotypes imputed on the basis of the 1000 Genomes, Phase 1, Version 3 reference panel (The 1000 Genomes Project Consortium, 2012). We only consider the subset of HapMap3 SNPs available in the 1kG data. By using this subset we substantially reduce the computational burden of the analyses, while preserving overlap between the SNP-sets in the studies and still having a sufficiently dense set of both common and rare SNPs (# SNPs after QC ≈ 1 million).

B. QUALITY CONTROL

Prior to QC, we extract HapMap3 SNPs (source: http://hapmap.ncbi.nlm.nih.gov/downloads/genotypes/hapmap3˙r3/plink˙format/, accessed: December 11, 2014) from the imputed genotype data of each study and convert the allele dosages to best-guess `PLINK` (Chang et al., 2015; Purcell et al., 2007) binary files by rounding dosages using `GCTA` (Yang, Lee, et al., 2011). Subsequently, we perform QC on the best-guess genotypes in two stages. In the first stage, we clean and harmonize the imputed genotype data at the study level. The cleaned and harmonized study genotypes are then merged into a pooled dataset. The second round of QC is aimed at cleaning the pooled dataset, on the basis of the samples for which the phenotype is available. Hence, the first QC stage is phenotype-independent, whereas the second stage depends on the phenotype of interest.

Table 6.3. Genotyping and imputation

Study	Genotyping platform	SNP exclusions			Subject exclusions*	Imputation**
		MAF	Call rate	HWE p-val	Call rate	Software
RS-I	Illumina 550K	0%	97.50%	10^{-7}	97.50%	MaCH/Minimac
RS-II	Illumina 550K	0%	97.50%	10^{-7}	97.50%	MaCH/Minimac
RS-III	Illumina 610K	0%	97.50%	10^{-7}	97.50%	MaCH/Minimac
STR	HumanOmniExpress 12v1A	1%	97.00%	10^{-7}	97.00%	MaCH/Minimac
HRS	Illumina Omni2.5	1%	98.00%	10^{-4}	98.00%	IMPUTE2

Note: *Individuals are also excluded on the basis of sex mismatch, close relatives, duplicates and ancestry outliers (STR excepted), or autosomal heterozygosity outliers (HRS excepted). **All samples have been imputed against the 1000Genomes, Phase 1, Version 3 haplotypes of all ancestries.

In the first QC stage (prior to merging), we filter out the following markers and individuals:

1. SNPs with imputation accuracy below 70%.
2. Non-autosomal SNPs.
3. SNPs with minor-allele frequency below 1%.
4. SNPs with Hardy-Weinberg-Equilibrium *p*-value below 1%.
5. SNPs with missingness greater than 5%.
6. Individuals with missingness greater than 5%.
7. SNPs that are not present in all studies.
8. SNPs whose alleles cannot be aligned across studies.

Prior to the first QC stage, we apply the following two additional steps in HRS:

1. Switch alleles to address a strand-flip error due to incorrect annotation.
2. Drop individuals of non-European ancestry.

After the first round of QC, a set of roughly 1 million overlapping SNPs, available for about 30,000 individuals is left. Panel I in Table 6.4 shows, for each study, the number of SNPs and individuals before and after the first round of QC.

The second QC stage, applied to the pooled data set, comprises the following steps:

1. Keep only individuals for whom the phenotype of interest and all corresponding control variables are available.
2. Drop SNPs with a minor-allele frequency below 1%.

3. Drop SNPs with Hardy-Weinberg-Equilibrium *p*-value below 1%.
4. Drop SNPs with missingness greater than 5%.
5. Drop individuals with missingness greater than 5%.
6. Keep only one individual per pair of individuals with a genomic relatedness greater than 0.025.

Since STR data consist of twins and having highly related individuals can bias estimates of SNP-based heritability due to environment-sharing, we randomly select only one individual per twin pair after Step 1 in the second QC stage.

Panel II in Table 6.4 shows the sample size and the number of SNPs in the pooled dataset for each phenotype. We only consider phenotypes that attain a sample size of at least 18,000 individuals after all QC steps. The lowest sample size after QC is 19,184 for self-rated-health

Table 6.4. Number of individuals and SNPs before and after quality control (QC) at the study level (Panel I) and at the pooled level (Panel II).

Panel I: study-level QC

	N		# SNPs	
Study	pre-QC	post-QC	pre-QC	post-QC
RS-I	6,291	6,291	31,337,615	1,062,589
RS-II	2,157	2,157	31,337,615	1,062,589
RS-III	3,048	3,048	31,337,615	1,062,589
STR	9,617	9,617	31,326,389	1,062,589
HRS	12,454	8,652	21,632,048	1,062,589
Total		29,765		1,062,589

Panel II: pooled-level QC

	N		# SNPs	
Phenotype	pre-QC	post-QC	pre-QC	post-QC
Height	29,765	20,458	1,062,589	1,052,572
BMI	29,765	20,449	1,062,589	1,052,600
EduYears	29,765	20,619	1,062,589	1,052,626
CurrCigt	29,765	20,686	1,062,589	1,052,524
CurrDrinkFreq	29,765	20,072	1,062,589	1,052,958
Self-rated health	29,765	19,184	1,062,589	1,053,190

and the highest is 20,686 for *CurrCigt*. For all phenotypes, the number of SNPs is slightly greater than one million.

C. PHENOTYPE DATA

For HRS, we use the RAND HRS data, version N, to obtain the phenotypes of interest. These data consist of measurements from eleven waves. RS-I consists of four data waves (Ergo 1-4). In both HRS and RS-I, data for some phenotypes are only available in a subset of the waves. RS-II, RS-III and STR do not have multiple measures over time for the phenotypes considered in this study. Table 6.5 describes how the phenotypes are constructed in each of the five studies.

As Table 6.5 shows, height, BMI, *EduYears*, and *CurrCigt* are measured quite consistently across waves. The self-rated health phenotype is also measured quite consistently, although in RS respondents are asked about health compared to members of the same age group, whereas a more absolute question is posed in STR and HRS. The drinking measure *CurrFreqDrink* is also measured somewhat heterogeneously; the threshold for what we treat as 'frequent drinking' is determined solely by how fine-grained the drinking frequency measure is in the respective studies.

6.5.2 GREML estimation

Height, BMI, *EduYears*, and self-rated health are treated as quantitative traits. *CurrCigt* and *CurrDrinkFreq* are treated as binary outcomes. In each study, (after aggregating across waves, if applicable) we regress quantitative phenotypes on age, squared age, sex, and an intercept. The residuals from the regression are standardized to have a sample-mean equal to zero and variance equal to one. For both binary and quantitative traits, the aforementioned covariates are also included in the GREML estimation. In addition, in bivariate GREML and pooled GREML estimation (i.e., considering multiple studies jointly), the intercept is replaced by indicator variables for the respective studies, capturing study-specific fixed effects. Finally, 20 principal components from the phenotype-specific genomic-relatedness matrix are added to the set of control variables in the GREML estimation, in order to correct for population stratification (Price et al., 2006).

6.5.3 GREML results

Details per phenotype on sample size, univariate estimates of SNP-heritability, and bivariate estimates of genetic correlation, stratified across studies, and cross-study averages, are provided in Table 6.6. Results stratified across sexes are listed in Table 6.7.

Table 6.5. Study-level phenotype measures.

Phenotype	Survey instrument				
	RS-I	RS-II	RS-III	STR	HRS
Years of education (*EduYears*)	Constructed in line with (Rietveld, Medland, et al., 2013) in all studies.				
Height	Median height across waves 1-4.	Height	Height	Height	Median height across waves 1-11.
BMI	Median BMI across waves 1-4.	BMI	BMI	BMI	Median BMI across waves 1-11.
Currently smoking cigarettes? (*CurrCigt*)	1 if stated to be a current smoker of cigarettes in the latest available measurement across waves 1-4.	1 if stated to be a current cigarette smoker.	Same as RS-II.	1 if stated to be a current cigarette smoker.	1 if responded positively to "currently smokes cigarettes?" in the latest available measurement across waves 1-11.
Currently drinking frequently (*CurrDrinkFreq*)	1 if indicated to "drink one or more alcoholic beverages per week" in the latest available measurement across waves 1-4.	1 if indicated to "drink one or more alcoholic beverages per week".	1 if indicated to "have drunk at least two alcoholic beverages a month during the past year."	1 if indicated to "have drunk at least two alcoholic beverages in the past month".	1 if indicated to "drink alcohol once per week or more" in the latest available measurement across waves 3-11.
Self-rated health	Only available in wave 1: "How is your general health compared to members of your age group?" Response categories reverse-coded such that 0=worse, 1=same, and 2=better.	Same as RS-I.	n.a.	"Rate your general health". Response categories recoded such that 0=bad, 1=not so good, 2=average, 3=good, 4=excellent.	Mode of the 4-point self-reported health measure across waves 1-11. Responses reverse-coded such that 0=poor, 1=fair, 2=good, 3=very good, and 4=excellent.

Table 6.6. GREML estimates of SNP-heritability (h^2_{SNP}) and genetic correlation (ρ_G) across studies.

Phenotype	*N*				Univariate estimates h^2_{SNP}				Bivariate estimates ρ_G			
	RS	STR	HRS	Total	RS	STR	HRS	Average[1]	RS - STR	RS - HRS	STR - HRS	Average[2]
Height	6,780	5,342	8,336	20,458	48.9% (4.9%)***	50.8% (6.0%)***	37.9% (4.1%)***	44.90%	0.976 (0.102)***	0.954 (0.095)***	0.967 (0.106)***	0.965
BMI	6,775	5,341	8,333	20,449	28.9% (4.9%)***	16.4% (6.1%)***	19.6% (4.1%)***	21.90%	1.000 (0.269)***	0.914 (0.172)***	0.847 (0.246)***	0.917
EduYears	6,735	5,543	8,341	20,619	17.5% (4.8%)***	20.6% (5.8%)***	17.3% (4.0%)***	18.20%	0.690 (0.233)***	0.659 (0.224)***†	1.000 (0.263)***	0.783
CurrCigt	6,803	5,579	8,304	20,686	17.8% (10.1%)**	18.7% (13.8%)*	20.4% (11.2%)**	19.10%	1.000 (0.643)***	0.611 (0.448)*	1.000 (0.607)***	0.858
CurrDrink-Freq	6,172	5,564	8,336	20,072	13.5% (8.7%)*	14.1% (9.5%)*	5.3% (6.3%)	10.30%	1.000 (0.666)***	0.298 (0.670)	-0.056 (0.647)	0.381
Self-rated health	5,264	5,577	8,343	19,184	13.5% (6.2%)**	9.4% (5.7%)**	21.3% (4.0%)***	15.70%	0.626 (0.439)**	0.363 (0.223)**††	0.447 (0.278)**	0.468

Note: Standard errors between parentheses. [1]Sample-size weighted average of univariate estimates across studies. [2]Sample-size weighted average of bivariate estimates across pairs of studies.

*h^2_{SNP} and/or genetic correlation > 0 at 10% sign. †genetic correlation < 1 at 10% sign. ‡genetic correlation < 0 at 10% sign.

**h^2_{SNP} and/or genetic correlation > 0 at 5% sign. ††genetic correlation < 1 at 5% sign. ‡‡genetic correlation < 0 at 5% sign.

***h^2_{SNP} and/or genetic correlation > 0 at 1% sign. †††genetic correlation < 1 at 1% sign. ‡‡‡genetic correlation < 0 at 1% sign.

Table 6.7. GREML estimates of SNP-heritability (h^2_{SNP}) and genetic correlation (ρ_G) across sexes.

Phenotype	*N*			Estimates h^2_{SNP}		Estimate ρ_G	
	Females	Males	Total	Females	Males	Average[1]	Females - Males
Height	11,553	8,905	20,458	43.2% (3.0%)***	45.1% (3.8%)***	44.00%	0.981 (0.067)***
BMI	11,542	8,907	20,449	22.1% (2.9%)***	23.8% (3.8%)***	22.80%	0.794 (0.122)*** †
EduYears	11,653	8,966	20,619	18.1% (2.9%)***	18.9% (3.7%)***	18.40%	0.832 (0.162)***
CurrCigt	11,706	8,980	20,686	22.3% (7.1%)***	26.7% (9.1%)***	24.20%	0.543 (0.257)*** †
CurrDrinkFreq	11,312	8,760	20,072	14.1% (4.6%)***	0.9% (6.0%)	8.30%	1.000 (2.068)*
Self-rated health	10,866	8,318	19,184	8.6% (3.1%)***	10.8% (4.0%)***	9.50%	1.000 (0.349)***

Note: Standard errors between parentheses. [1] Sample-size weighted average of univariate estimates across studies.
$^{*}h^2_{SNP}$ and/or genetic correlation > 0 at 10% sign. †genetic correlation < 1 at 10% sign. ‡genetic correlation < 0 at 10% sign.
$^{**}h^2_{SNP}$ and/or genetic correlation > 0 at 5% sign. ††genetic correlation < 1 at 5% sign. ‡‡genetic correlation < 0 at 5% sign.
$^{***}h^2_{SNP}$ and/or genetic correlation > 0 at 1% sign. †††genetic correlation < 1 at 1% sign. ‡‡‡genetic correlation < 0 at 1% sign.

APPENDIX A

Supplementary Tables to Chapter 3

Table A1. LDSC estimates of pairwise genetic correlations

Panel A. Intra-GWAS Phenotypic Heterogeneity

	SWB				DS				Neuroticism			
	r_g	SE	*p*-value		r_g	SE	*p*-value		r_g	SE	*p*-value
LS/PA	0.981	0.065	$<10^{-20}$	GERA/PGC	0.588	0.242	0.015	GPC/UKB	1.112	0.143	7.38×10^{-15}
WB/LS	0.897	0.017	$<10^{-20}$	GERA/UKB	0.972	0.216	6.57×10^{-6}				
PA/WB	1.031	0.019	$<10^{-20}$	UKB/PGC	0.797	0.108	1.71×10^{-13}				

Panel B. Pairwise Genetic Correlations between Well-being Phenotypes

	Wellbeing Phenotypes				Placebo		
	r_g	SE	*p*-value		r_g	SE	*p*-value
SWB/DS	-0.814	0.046	$<10^{-20}$	Height/SWB	0.065	0.028	0.023
SWB/Neuroticism	-0.749	0.034	$<10^{-20}$	Height/DS	-0.062	0.029	0.034
DS/Neuroticism	0.750	0.027	$<10^{-20}$	Height/Neuroticism	-0.061	0.026	0.017

Note: LS is life satisfaction. PA is positive affect. SWB is the combined well-being measure. DS is depressive symptoms. r_g is the genetic correlation between phenotypes 1 and 2, computed using the LDSC python software package and the "eur_w_ld_chr/" files of LD scores calculated by Finucane et al. (2015). In the LD score regressions, we include only HapMap3 SNPs with MAF > 0.01, and the standard errors of the LD Score regressions are estimated using a block jackknife over SNPs (by the LDSC software). GWAS summary statistics for SWB, DS, and neuroticism come from our main and auxiliary analyses. Panel A results exclude deCODE.

Table A2. Overview of cohorts

Study	Full Name	Country	LS (*N*= 166,210)	PA (*N*= 180,293)	WB (*N*= 35,944)	Cohort Type
Panel A. Cohorts with LS, PA and WB results. Number of cohorts: 12. Combined sample size of WB results: *N* = 35,944						
AGES	Age, Gene/ Environment Susceptibility–Reykjavik Study	Iceland	3,059	3,054	3,044	LSPAWB
ALSPAC	Avon Longitudinal Study of Parents and Children	England	5,649	4,852	5,654	LSPAWB
EGCUTOMNI	Estonian Genome Center, University of Tartu	Estonia	588	588	585	LSPAWB
EGCUT370	Same as above	Estonia	1,112	1,117	1,110	LSPAWB
HNRSoexpr	Heinz Nixdorf Recall Study	Germany	1,332	1,319	1,343	LSPAWB
HNRSomni1	Same as above	Germany	765	759	773	LSPAWB
HRS	Health and Retirement Study	USA	9,938	9,117	9,942	LSPAWB
NTR	Netherlands Twin Register	Netherlands	8,051	6,369	8,060	LSPAWB
RUSHMAP	Rush University Medical Center - Memory and Aging Project	USA	370	370	370	LSPAWB
TEDS	Twins Early Development Study	England and Wales	2,143	2,142	2,148	LSPAWB
TRAILS	Tracking Adolescents' Individual Lives Survey	Netherlands	1,205	1,205	1,204	LSPAWB
YFS	The Cardiovascular Risk in Young Finns Study	Finland	1,720	1,738	1,711	LSPAWB
Panel B. Cohorts with LS and PA (but not WB) results. Number of cohorts: 12. Combined sample size of LS/PA results: *N* = 15,447/17,039.						
1958T1D	1958 British Birth Cohort	UK	2,029	2,034	N/A	LSPA
1958WTC	1958 British Birth Cohort	UK	2,192	2,196	N/A	LSPA
BASE	Berlin Aging Study II	Germany	1,395	1,392	N/A	LSPA
HPFSCHD	Health Professional Follow-up Study	USA	849	854	N/A	LSPA
HPFSKS	Same as above	USA	477	478	N/A	LSPA
HPFST2D	Same as above	USA	1,644	1,653	N/A	LSPA
KORAF3	Kooperative Gesundheitsforschung in der Region Augsburg	Germany	809	827	N/A	LSPA

KORAF4	Same as above	Germany	652	660	N/A	LSPA
NHSBRCA	Nurses' Health Study	USA	1,723	2,268	N/A	LSPA
NHSCHD	Same as above	USA	753	1,116	N/A	LSPA
NHSKS	Same as above	USA	449	490	N/A	LSPA
NHST2D	Same as above	USA	2,475	3,071	N/A	LSPA

Panel C. Cohorts with PA results files only. Number of cohorts: 12. Combined Sample Size of Results Files: $N = 130,624$

ASPS	Austrian Stroke Prevention Study	Austria	N/A	815	N/A	PA
AEGS1	AtheroExpress	Netherlands	N/A	517	N/A	PA
BLSA	Baltimore Longitudinal Study of Aging	USA	N/A	827	N/A	PA
Colaus	Etude Cohorte Lausannoise	Switzerland	N/A	3,530	N/A	PA
CROATIAKOR	Croatia Korčula	Croatia	N/A	876	N/A	PA
CROATIASPL	Croatia Split	Croatia	N/A	495	N/A	PA
CROATIAVIS	Croatia Vis	Croatia	N/A	918	N/A	PA
DECODE	deCODE, Inc.	Iceland	N/A	10,384	N/A	PA
DHS	Dortmund Health Study	Germany	N/A	879	N/A	PA
ERF	Erasmus Rucphen Family Study	Netherlands	N/A	2,129	N/A	PA
FTC	Finnish Twin Cohort	Finland	N/A	685	N/A	PA
FTC2014	Finnish Twin Cohort	Finland	N/A	1,805	N/A	PA
HBCS	Helsinki Birth Cohort Study	Finland	N/A	1,450	N/A	PA
HCS	Hunter Community Study	Australia	N/A	1,906	N/A	PA
LL	The LifeLines Cohort Study	Netherlands	N/A	11,971	N/A	PA
MCTFR	Minnesota Center for Twin and Family Research	USA	N/A	7,007	N/A	PA
MESA	Multi-Ethnic Study of Atherosclerosis	USA	N/A	2,495	N/A	PA
NEO	The Netherlands Epidemiology of Obesity Study	Netherlands	N/A	5,715	N/A	PA
QIMR	Queensland Institute of Medical Research	Australia	N/A	6,572	N/A	PA
QIMR2	Queensland Institute of Medical Research	Australia	N/A	1,235	N/A	PA
RSI	Rotterdam Study Baseline	Netherlands	N/A	3,731	N/A	PA
RSII	Rotterdam Study Extension of Baseline	Netherlands	N/A	2,039	N/A	PA
RSIII	Rotterdam Study Young	Netherlands	N/A	2,027	N/A	PA
SAR	SardiNIA Study of Aging	Italy	N/A	2,542	N/A	PA

STR1	Swedish Twin Registry	Sweden	N/A	9,343	N/A	PA
STR2	Swedish Twin Registry	Sweden	N/A	4,704	N/A	PA
TwinsUK	TwinsUK Adult Twin Registry	UK	N/A	3,484	N/A	PA
UKB	UK Biobank	UK	N/A	40,543	N/A	PA

Panel D. Cohorts with LS results files only. Number of cohorts: 7. Combined Sample Size of Results Files: $N = 114{,}831$

NFBC	Northern Finland Birth Cohort 1966	Finland	5,203	N/A	N/A	LS
23andMe	23andMe, inc	Primarily USA	93,454	N/A	N/A	LS
ARIC	Atherosclerosis Risk in Communities Study	USA	8,716	N/A	N/A	LS
ELSA	The English Longitudinal Study of Ageing	England	5,047	N/A	N/A	LS
GOYA	Genetics of Overweight Young Adults	Denmark	1,179	N/A	N/A	LS
LBC21	Lothian Birth Cohort 1921	Scotland	443	N/A	N/A	LS
LBC36	Lothian Birth Cohort 1936	Scotland	789	N/A	N/A	LS

Note: LS is life satisfaction. PA is positive affect. WB is the combined well-being measure.

Table A3. Quality control filters

Panel A. HapMap 2 Analyses (SWB, LS, PA, DS, Neuroticism)

		Imputation Quality				
	MAF	MaCH	IMPUTE	PLINK	HWE *p*-value	Call rate
$N < 1000$	0.1	0.4	0.5	0.8	10^{-3}	95%
$1000 < N < 2000$	0.05	0.4	0.5	0.8	10^{-4}	95%
$2000 < N > 10000$	0.03	0.4	0.5	0.8	10^{-5}	95%
$N > 10000$	0.03	0.4	0.5	0.8	None	95%

Panel B. 1000G Analyses (SWB)

		Imputation Quality				
	MAC	MaCH	IMPUTE	PLINK	HWE *p*-value	Call rate
$N < 1000$	30	0.6	0.7	0.8	10^{-3}	95%
$1000 < N < 2000$	30	0.6	0.7	0.8	10^{-4}	95%
$2000 < N > 10000$	30	0.6	0.7	0.8	10^{-5}	95%
$N > 10000$	30	0.6	0.7	0.8	None	95%

Note: SWB is the combined well-being measure. LS is life satisfaction. PA is positive affect. DS is depressive symptoms. *N* is the sample size per SNP. *MAF* is minor allele frequency. HWE is Hardy Weinberg equilibrium. HWE *p*-value filter is applied only to genotyped SNPs, with the exception of KORAF3 and KORAF4. Call rate is the minimum fraction of subjects for which the association results for a SNP must be available in order for the SNP to be included. *MAC* is minor allele count.

Table A4. SNP filtering

Panel A. SNP Filtering (HapMap Phase II SNPs)

File	SNPs In	SNP not in reference file	Variable quality	Low imputation accuracy	Low minor allele frequency	Low HWE *p*-value	Low call rate	Duplicate / Allele mismatch	SNPs Out	λ_{GC}
1958WTC	3,856,236	1,376,072	7,404	198,765	27,817	6	0	0	2,246,172	1.002
23andMe	11,972,746	9,504,767	604	176,988	2	0	0	0	2,290,385	1.099
AEGS1	2,543,887	62,845	132,847	539,529	54,107	845	0	0	1,753,714	0.993
AGES	2,593,541	124,358	1	220,641	27,889	0	0	0	2,220,652	1.007
ALSPAC	17,213,858	14,730,670	661	196,028	2,626	10	0	0	2,283,863	1.000
ARIC	28,099,593	25,637,897	84	196,306	1,581	1,404	0	0	2,262,321	1.003
ASPS	2,543,887	62,845	752	653,574	6,517	0	0	0	1,820,199	1.018
BASE	27,213,647	24,748,903	10,252	323,400	18,658	328	0	2	2,112,104	1.003
BLSA	2,543,644	62,656	27	647,568	6,907	539	1,987	0	1,823,960	0.993
Colaus	29,562,715	27,078,585	0	179,739	110,827	67	16,301	0	2,177,196	1.005
CROATIAKOR	2,543,887	62,845	10,069	650,748	25,652	731	0	0	1,793,842	0.919
CROATIASPL	2,543,887	62,845	1,442	655,750	18,161	431	0	0	1,805,258	0.986
CROATIAVIS	2,543,887	62,845	1,042	662,852	18,628	605	0	0	1,797,915	1.004
DECODE	2,490,839	17,901	125	204,898	2,968	0	0	0	2,264,947	0.993
DHS	2,557,252	103,164	62,444	588,506	3,777	967	0	0	1,798,394	1.001
EGCUT370	2,550,547	69,926	15,369	340,266	29,180	44	0	28	2,095,734	1.010
EGCUTOMNI	2,652,572	87,724	23,943	657,759	41,044	378	0	113,418	1,728,306	0.998
ELSA	2,509,150	50,341	23	184,870	4,960	0	47,609	0	2,221,347	1.005
ERF	29,974,334	27,490,074	549	263,206	141,223	0	0	0	2,079,282	1.019
FTC2014	8,981,685	6,597,999	0	269,179	205	59	1,259	0	2,112,984	1.016
FTC	2,068,145	109,915	0	470,420	0	672	0	0	1,487,138	1.004
GOYA	2,823,117	342,051	207	347,305	0	122	0	1	2,133,431	0.988

HBCS	2,543,887	62,845	4,184	353,227	9,246	501	19	0	2,113,865	1.005
HCS	2,679,116	238,800	0	330,119	58,937	10	15	0	2,051,235	1.004
HNRSoexpr	81,610,745	79,209,709	19,065	315,228	515	181	0	611	2,065,436	0.993
HNRSomni1	16,421,060	14,048,760	235	615,567	584	501	0	527	1,754,886	0.995
HPFSCHD	2,543,439	62,692	0	646,987	11,138	1,316	0	0	1,821,306	1.042
HPFSKS	2,543,134	62,602	0	646,737	5,572	608	0	0	1,827,615	1.029
HPFST2D	2,543,778	62,799	0	329,202	16,567	94	0	0	2,135,116	1.018
HRS	17,956,622	15,736,861	0	168,198	865	0	0	0	2,050,698	1.020
KORAF3	2,470,355	103,177	10,472	615,363	65,070	10,378	713	0	1,665,182	1.014
KORAF4	2,669,039	254,348	11,439	622,484	17,436	17,691	25,888	0	1,719,753	0.990
LBC21	2,543,887	62,845	0	659,956	5,885	332	0	0	1,814,869	0.994
LBC36	2,543,887	62,845	0	658,135	5,879	311	0	0	1,816,717	1.006
LL	2,481,385	5	92	192,352	75,947	0	0	7	2,212,982	1.038
RUSHMAP	2,543,887	62,845	63,524	592,627	13,710	0	0	0	1,811,181	0.989
MCTFR	2,622,740	141,673	29	196,124	13,018	39	0	0	2,271,857	1.014
MESA	3,854,680	1,372,706	3,539	183,424	35,017	2,261	163,664	0	2,094,069	1.007
NEO	29,736,423	27,296,081	2,696	195,082	28,483	39	0	0	2,214,042	1.011
NFBC	38,038,268	35,552,706	8,945	214,520	1,770	125	0	1,949	2,258,253	1.014
NHSBRCA	2,543,732	62,767	0	197,660	10,004	174	0	0	2,273,127	0.991
NHSCHD	2,543,570	62,756	0	335,228	14,998	30	0	0	2,130,558	1.000
NHSKS	2,543,092	62,598	0	650,647	5,577	602	0	0	1,823,668	1.018
NHST2D	2,543,847	62,837	0	195,540	20,145	38	0	0	2,265,287	1.008
NTR	6,611,573	4,709,711	0	125,990	1,204	0	0	0	1,774,668	1.019
QIMR2	30,072,738	27,588,323	470	327,958	64,399	0	0	0	2,091,588	1.005
QIMR	2,373,248	37,246	0	139,309	5,986	42	0	0	2,190,665	0.999
RSI	2,543,887	62,845	0	204,816	13,905	0	0	0	2,262,321	1.001
RSII	2,543,887	62,845	0	206,915	14,670	0	0	0	2,259,457	1.005
RSIII	2,543,887	62,845	0	206,203	12,309	0	0	0	2,262,530	0.987
SAR	8,788,395	6,394,510	244	192,682	4,685	0	0	290	2,195,984	0.989
STR1	2,585,286	104,189	37	205,380	19,068	130	0	0	2,256,482	1.009
STR2	10,128,477	7,667,860	278	172,604	46,062	0	0	0	2,241,673	1.003

TEDS	2,731,370	251,825	7,481	194,126	32,370	1,627	154,508	2	2,089,431	1.011
TRAILS	2,565,543	84,502	15,964	328,461	64,149	27	281	0	2,072,159	1.012
TWINSUK	3,044,064	683,931	12,529	274,065	3,156	265	683,812	0	1,386,306	0.995
YFS	2,577,640	96,563	3,214	359,630	9,730	1	0	0	2,108,502	0.991
UKB	11,192,334	8,751,087	0	167,402	116	0	0	9	2,273,720	1.057

Panel B. 1000G SNP Filtering

File	SNPs In	Misaligned SNPs / In-del	SNP not in reference file	Variable quality	Low minor allele count	Low imputation accuracy	Low HWE *p*-value	Low call rate	Duplicate / Allele mismatch	SNPs Out	λ_{GC}
23andMe	11,972,746	908,921	991,338	0	0	672,041	0	0	329	9,400,117	1.072
ALSPAC	17,213,858	1,165,349	336,275	31,114	5,524,672	1,500,033	10	0	4	8,656,401	0.997
ARIC	28,099,593	3,544	635,299	3,206,777	13,584,569	2,396,012	1,626	0	23	8,271,743	1.001
BASE	27,213,647	1,370,583	184,457	14,911,913	3,128,906	1,202,641	1,821	0	38	6,413,288	1.11
Colaus	29,562,715	1,350,158	176,426	0	17,325,950	4,198,035	110	21,331	1	6,490,704	1.008
ERF	29,974,334	1,379,685	131,396	142,229	18,587,189	4,563,486	0	0	0	5,170,349	1.023
FTC2014	8,981,685	78	507,197	0	506,380	366,735	163	30,241	10,401	7,560,490	1.003
HNRSoexpr	81,610,745	3,305,025	129,247	61,181,228	8,744,831	357,655	238	0	21,244	7,871,277	0.994
HNRSomni1	16,421,060	1,542,399	120,605	16,134	7,311,511	238,697	670	0	18,563	7,172,481	0.994
HRS	17,956,622	1,296,997	627,766	0	5,990,027	1,125,113	0	0	5	8,916,714	1.014
LL	28,681,763	0	320,580	43,682	13,648,540	7,836,726	0	0	211	6,832,024	1.03
MCTFR	8,118,911	0	197,419	1,601	5,239	722,817	0	0	0	7,191,835	1.001
NEO	29,736,423	1,378,071	308,690	12,392,604	4,970,395	3,434,518	50	0	7	7,252,088	1.007
NFBC	38,038,268	1,393,605	879,418	21,559,193	4,477,084	1,616,676	154	0	2,126	8,110,012	1.006

NTR	6,611,573	521,986	120,364	0	0	498,274	0	0	1	5,470,948	1.019
QIMR2	30,072,738	1,390,975	158,251	151,445	20,247,666	1,827,798	0	0	0	6,296,603	1.003
RSI	30,072,738	1,390,975	302,233	0	18,259,634	2,371,428	0	0	3	7,748,465	1.007
RSII	29,941,010	1,390,975	208,412	0	19,453,184	1,553,801	0	0	0	7,334,638	1.006
RSIII	30,072,738	1,390,975	211,975	0	19,561,212	1,484,219	0	0	0	7,424,357	0.997
SAR	8,788,395	315,613	341,283	169	445,264	623,903	0	0	2,620	7,059,543	0.986
STR1	10,774,211	921,494	306,594	32,418	0	1,985,292	0	0	0	7,528,413	1.005
STR2	10,128,477	0	289,173	37,002	0	2,647,900	0	0	17,276	7,137,126	1.005
UKB	11,192,334	1,270,169	230,205	16	0	302,553	0	0	485	9,388,906	1.043

Note: For description of reference files, see Section 3.4.1-F (Panel A) or 3.4.3-B (Panel B). A SNP fails the variable quality filter if the following variables are missing or take invalid values: effect allele frequency, Beta, SE, *p*-value, imputation accuracy, or the imputed/genotyped indicator. HWE is Hardy-Weinberg equilibrium. Allele mismatch refers to SNPs whose alleles do not match the alleles in the reference files. deCODE standard errors were adjusted for genomic control prior to uploading results. The original λ_{GC} was 1.121. Misaligned SNPs in imputation reference panel refers to the 730+199 SNPs whose strands were incorrectly aligned in the September or December 2013 releases of the 1000 Genomes Phase 1 haplotypes provided by the software IMPUTE2. For details on QC thresholds used, see Section 3.4.3.

Table A5. Association results for SNPs that reached p-value < 10^{-6} in SWB, LS and PA analyses.

Panel A. SWB meta-analysis results

SNPID	Chr	BP	*EA*	*EAF*	Beta	*SE*	*p*-value	*N*
rs3756290	5	130,951,750	A	0.24	-0.0177	0.0031	9.55×10^{-9}	286,851
rs2075677	20	47,701,024	A	0.76	0.0175	0.0031	1.49×10^{-8}	288,454
rs4958581	5	152,187,729	T	0.66	0.0153	0.0027	2.29×10^{-8}	294,043
rs1654345	19	38,524,288	A	0.74	-0.0158	0.0030	9.74×10^{-8}	295,719
rs6500768	16	6,342,641	T	0.31	-0.0149	0.0029	2.16×10^{-7}	281,538
rs9528554	13	63,282,834	T	0.75	0.0155	0.0030	3.20×10^{-7}	289,384
rs11216168	11	116,741,553	A	0.15	-0.0185	0.0037	4.21×10^{-7}	298,105
rs17783634*	8	11,054,097	A	0.48	-0.0131	0.0026	5.23×10^{-7}	294,036
rs9533975	13	45,373,396	T	0.58	0.0131	0.0026	7.00×10^{-7}	296,585
rs1480570	16	73,676,742	A	0.66	0.0136	0.0027	7.07×10^{-7}	297,064
rs258668	7	81,728,516	T	0.58	-0.0132	0.0027	7.80×10^{-7}	288,125
rs12298541	12	66,306,441	A	0.38	0.0135	0.0027	9.41×10^{-7}	281,216
rs3104705	16	10,069,290	A	0.13	-0.0189	0.0039	9.91×10^{-7}	296,808
rs11245339	10	126,399,930	A	0.36	0.0132	0.0027	9.95×10^{-7}	298,265

Panel B. LS meta-analysis results

SNPID	Chr	BP	*EA*	*EAF*	Beta	*SE*	*p*-value	*N*
rs7554469	1	153,827,552	A	0.49	0.0218	0.0037	3.23×10^{-9}	148,212
rs17494938	4	80,247,342	T	0.63	-0.0196	0.0036	4.90×10^{-8}	166,202
rs7737355	5	130,604,811	A	0.79	0.0226	0.0043	1.03×10^{-7}	166,204
rs6918725	6	126,990,392	T	0.48	-0.0180	0.0035	2.16×10^{-7}	166,202
rs2024595	20	47,517,896	A	0.59	0.0180	0.0036	7.35×10^{-7}	156,262

Panel C. PA meta-analysis results

SNPID	Chr	BP	*EA*	*EAF*	Beta	*SE*	*p*-value	*N*
rs11612312	12	52,349,088	T	0.79	0.0226	0.0042	5.66×10^{-8}	176,411
rs13235506	7	53,785,701	T	0.94	-0.0385	0.0076	3.78×10^{-7}	165,191
rs17374883	4	140,960,502	T	0.71	0.0200	0.0039	3.87×10^{-7}	157,001
rs258677	7	81,733,428	T	0.41	0.0176	0.0035	4.21×10^{-7}	170,810
rs11671324	19	57,515,403	T	0.92	0.0343	0.0069	7.54×10^{-7}	139,224
rs1538482	20	47,705,496	T	0.70	0.0177	0.0036	9.94×10^{-7}	180,259

Panel D. SWB meta-analysis results in cohorts with 1000G-imputed genotypes

SNPID	Chr	BP	*EA*	*EAF*	Beta	*SE*	*p*-value	*N*
rs6579956	5	152,078,663	T	0.47	-0.0173	0.0030	4.44×10^{-9}	229,883
rs13185787	5	130,627,884	A	0.26	-0.0193	0.0034	2.13×10^{-8}	221,823
rs147158068	2	154,006,913	A	0.97	-0.0567	0.0107	1.14×10^{-7}	164,772
rs4606360	1	173,699,007	A	0.63	-0.0157	0.0031	3.02×10^{-7}	229,883
rs75279353	12	45,968,705	A	0.02	-0.0531	0.0104	3.07×10^{-7}	227,341
rs6572548	14	49,595,245	T	0.28	0.0168	0.0033	3.09×10^{-7}	229,883
rs61905145	11	116,718,415	A	0.15	-0.0213	0.0042	3.16×10^{-7}	229,883
rs298555	5	17,330,408	A	0.69	0.0163	0.0032	3.35×10^{-7}	229,883
rs144998313	15	93,434,966	A	0.99	-0.0771	0.0151	3.53×10^{-7}	175,532
rs6955467	7	70,073,157	A	0.63	-0.0156	0.0031	4.85×10^{-7}	221,823
rs1085719	3	9,689,147	A	0.05	0.0417	0.0084	6.04×10^{-7}	163,174
rs4942783	13	31,604,816	C	0.53	0.0150	0.0030	6.24×10^{-7}	221,823
rs186389748	3	117,709,832	C	0.00	0.1617	0.0325	6.46×10^{-7}	116,058
rs12125335	1	21,385,436	T	0.85	-0.0230	0.0046	6.67×10^{-7}	188,526
rs62341806	5	7,301,374	T	0.53	0.0149	0.0030	6.83×10^{-7}	221,823
rs6954274	7	9,881,063	A	0.38	-0.0151	0.0030	7.10×10^{-7}	229,883
rs139049023	8	76,123,783	A	0.99	0.0981	0.0198	7.11×10^{-7}	200,774
rs34256173	5	105,038,633	T	0.10	0.0241	0.0049	8.24×10^{-7}	229,883
rs3213545	12	121,471,337	A	0.29	-0.0159	0.0033	9.71×10^{-7}	229,883

Note: SWB is the combined well-being measure. LS is life satisfaction. PA is positive affect. Genomic coordinates are based on GRCh37. *EA* is effect allele. *EAF* is effect allele frequency. *SE* is standard error. *rs17783634 tags the inversion polymorphism on chromosome 8. λ_{GC}'s for SWB, LS, PA and SWB (1000G) are 1.206, 1.119, 1.118 and 1.124, respectively. λ_{GC}'s are obtained using the full-sample meta-analyses (including deCODE) after inflating the standard errors from the meta-analyses by the square root of the estimated intercept from an LD Score regression. The LD Score regression intercepts used to inflate standard errors, computed using only HapMap3 SNPs with minor allele frequency > 0.01, are 1.012 (SWB), 1.011 (PA), 1.007 (LS) and 1.008 (SWB 1000G).

Table A6. SNP filtering in DS and neuroticism cohorts

File	SNPs In	SNP not in reference file	Variable quality	Low minor allele frequency	Low imputation accuracy	Duplicate / Allele mismatch	Indel	SNPs Out	λ_{GC}
GERA	17,114,733	107,304	102,799	11,795,688	0	10,187	436,882	4,661,873	1.011
GPC	6,949,614	443	0	466,086	0	7	1	6,483,077	1.049
PGC	1,221,666	95,848	39,549	83,428	4,083	5	0	998,753	1.068
UKB - DS	11,192,334	315,074	16	3,321,996	20,276	1,134	988,976	6,544,862	1.098
UKB - Neuroticism	11,192,334	315,027	16	3,322,152	20,284	1,134	988,946	6,544,775	1.214

Note: *DS* is depressive symptoms. For description of the reference file, see Chapter 5, Section 3.4.4-E. A SNP fails the variable quality filter if the following variables are missing or take invalid values: effect allele frequency, Beta, *SE*, *p*-value, imputation accuracy, or the imputed/genotyped indicator. Allele mismatch refers to SNPs whose alleles do not match the reference alleles in the files described in Section 3.4.4-E. For details on QC thresholds used, see Section 3.4.4**.** HWE equilibrium *p*-value and call rate filters have been omitted from the table because none of the results files contained these columns, making the filters inapplicable.

Table A7. Depressive symptoms and neuroticism meta-analysis results

Panel A: Depressive symptoms meta-analysis results

SNPID	Chr	BP	*EA*	*EAF*	Beta	*SE*	*p*-value	*N*
rs7973260	12	118,375,486	A	0.19	0.0306	0.0051	1.78×10^{-9}	124,498
rs62100776	18	50,754,633	A	0.56	-0.0252	0.0044	8.45×10^{-9}	105,739
rs6992714	8	64,628,120	T	0.75	0.0205	0.0038	9.32×10^{-8}	180,866
rs10233018	7	117,523,709	A	0.49	-0.0184	0.0035	1.65×10^{-7}	162,107
rs1690818	11	99,496,554	T	0.69	-0.0193	0.0038	4.05×10^{-7}	162,107
rs7074335	10	106,700,394	T	0.11	0.0351	0.0070	4.57×10^{-7}	105,739
rs9845113	3	146,953,311	C	0.45	0.0178	0.0035	4.64×10^{-7}	162,107
rs9427622	1	196,355,524	T	0.41	-0.0205	0.0041	4.75×10^{-7}	124,498
rs59659806	15	38,919,964	T	0.81	-0.0221	0.0044	6.01×10^{-7}	162,107
rs853679	6	28,296,863	A	0.15	-0.0229	0.0046	6.62×10^{-7}	180,866
rs11636582	15	86,599,160	T	0.07	-0.0423	0.0085	6.98×10^{-7}	105,739
rs10884216	10	107,461,125	T	0.74	-0.0197	0.0040	8.67×10^{-7}	162,107
rs652714	18	53,321,026	A	0.14	-0.0232	0.0047	8.95×10^{-7}	180,866
rs4810896	20	47,535,298	A	0.36	-0.0170	0.0035	9.00×10^{-7}	180,866
rs4942916	13	33,375,996	T	0.65	-0.0223	0.0046	9.97×10^{-7}	105,739

Panel B: Neuroticism meta-analysis results

SNPID	Chr	BP	*EA*	*EAF*	Beta	*SE*	*p*-value	*N*
rs2572431[#]	8	11,105,077	T	0.59	0.0283	0.0035	4.20×10^{-16}	170,908
rs193236081[*]	17	44,142,332	T	0.77	-0.0284	0.0043	6.26×10^{-11}	151,297
rs10960103	9	11,699,270	C	0.77	0.0264	0.0042	2.14×10^{-10}	165,380
rs4938021	11	113,364,803	T	0.66	0.0233	0.0037	4.03×10^{-10}	159,900
rs139237746	11	10,253,183	T	0.51	-0.0204	0.0034	2.55×10^{-9}	170,908
rs1557341	18	35,127,427	A	0.34	0.0213	0.0037	5.58×10^{-9}	165,579
rs12938775	17	2,574,821	A	0.53	-0.0202	0.0035	8.54×10^{-9}	163,283
rs12961969	18	35,364,098	A	0.20	0.0250	0.0045	2.16×10^{-8}	156,758
rs35688236	3	34,582,993	A	0.69	0.0213	0.0038	2.35×10^{-8}	161,636
rs2150462	9	23,316,330	C	0.74	-0.0217	0.0039	2.66×10^{-8}	170,907
rs12903563	15	78,033,735	T	0.50	0.0198	0.0036	2.86×10^{-8}	157,562
rs9584850	13	99,101,426	C	0.27	-0.0213	0.0040	7.29×10^{-8}	161,637
rs10733389	9	23,378,220	A	0.38	0.0189	0.0035	7.86×10^{-8}	170,908
rs7107356	11	47,676,170	A	0.51	-0.0181	0.0034	1.27×10^{-7}	170,911
rs56080343	12	118,876,918	T	0.80	-0.0225	0.0043	1.56×10^{-7}	170,280
rs6882046	5	87,968,864	A	0.71	-0.0197	0.0038	1.60×10^{-7}	170,281
rs1262465	18	52,857,732	A	0.50	-0.0185	0.0035	1.66×10^{-7}	160,013
rs2458167	11	99,500,748	A	0.30	0.0194	0.0037	1.86×10^{-7}	170,907
rs10769123	11	45,193,779	A	0.25	-0.0210	0.0041	2.13×10^{-7}	163,692

rs9468186	6	27,626,631	A	0.79	0.0219	0.0042	2.32×10^{-7}	170,907
rs7973260	12	118,375,486	A	0.20	0.0229	0.0044	2.40×10^{-7}	161,459
rs76659101	3	34,467,077	T	0.95	0.0492	0.0096	2.57×10^{-7}	107,245
rs114304113	6	28,782,363	C	0.05	0.0470	0.0092	3.17×10^{-7}	137,866
rs4761545	12	94,426,468	T	0.60	-0.0178	0.0035	3.54×10^{-7}	170,280
rs10862219	12	81,430,043	T	0.40	0.0182	0.0036	3.78×10^{-7}	161,638
rs10244364	7	117,529,641	T	0.66	-0.0187	0.0037	4.07×10^{-7}	165,342
rs6888114	5	164,450,693	A	0.27	0.0199	0.0039	4.34×10^{-7}	161,325
rs4632195	18	50,746,748	T	0.52	0.0175	0.0035	4.57×10^{-7}	165,579
rs114293326	5	164,471,700	A	0.13	0.0255	0.0050	4.58×10^{-7}	170,102
rs2273085	22	41,615,376	T	0.31	0.0194	0.0039	5.12×10^{-7}	156,935
rs490647	1	37,242,743	A	0.21	0.0215	0.0043	5.61×10^{-7}	161,637
rs72694244	9	11,115,945	A	0.86	0.0246	0.0049	5.65×10^{-7}	170,102
rs932143	12	74,367,283	T	0.56	0.0172	0.0034	5.92×10^{-7}	170,283
rs75225668	2	163,420,424	A	0.95	-0.0411	0.0083	6.46×10^{-7}	148,979
rs35855737	3	65,542,856	T	0.82	0.0222	0.0045	6.72×10^{-7}	170,906
rs4906947	15	25,387,186	C	0.61	0.0177	0.0036	6.98×10^{-7}	165,578
rs61876950	11	9,960,872	T	0.40	0.0173	0.0035	7.13×10^{-7}	170,908
rs6941639	6	131,143,142	A	0.38	0.0180	0.0036	7.53×10^{-7}	161,636
rs116966368	18	26,273,649	A	0.24	0.0207	0.0042	7.53×10^{-7}	156,934
rs6904071	6	27,047,256	A	0.15	-0.0242	0.0049	7.59×10^{-7}	168,513
rs859767	2	135,341,200	A	0.57	0.0178	0.0036	7.83×10^{-7}	156,735
rs17333948	4	183,933,386	T	0.03	-0.0615	0.0125	8.91×10^{-7}	126,253

Note: This table lists the set of approximately independent SNPs with an association *p*-value $<1\times10^{-5}$. Results for depressive symptoms (DS) are in Panel A; and results for neuroticism in Panel B. *EA* is effect allele. *EAF* is effect allele frequency. *SE* is standard error. λ_{GC} is 1.102 for DS and 1.226 for neuroticism. λ_{GC}'s are obtained using the full-sample meta-analyses after inflating the standard errors from the meta-analyses by the square root of the estimated intercept from an LD Score regression. The LD Score regression intercept used to inflate standard errors, computed using only HapMap3 SNPs with minor allele frequency > 0.01, is 1.008 for DS. Because the estimated intercept for neuroticism is below 1 (0.998), the standard errors in the neuroticism GWAS are unadjusted. #inversion-tagging polymorphism on chromosome 8. *inversion-tagging polymorphism on chromosome 17.

Table A8. Predictive power of the polygenic score constructed from the SWB GWAS results in HRS.

Panel A. LDpred Polygenic Scores

	SWB	PA	LS	DS	Neuroticism	Conscientiousness	Extraversion	Agreeableness	Openness	Height
N	8,248	8,269	8,285	8,617	8,264	8,268	8,271	8,271	8,253	8,650
ΔR^2	1.35%	0.97%	1.16%	0.45%	0.70%	0.34%	0.49%	0.32%	0.12%	0.12%
95% CI - low	0.91%	0.62%	0.77%	0.21%	0.37%	0.13%	0.24%	0.12%	0.02%	0.02%
95% CI - high	1.86%	1.42%	1.67%	0.78%	1.12%	0.65%	0.85%	0.60%	0.32%	0.31%

Panel B. Linear Polygenic Scores

	SWB	PA	LS	DS	Neuroticism	Conscientiousness	Extraversion	Agreeableness	Openness	Height
N	8,248	8,269	8,285	8,617	8,264	8,268	8,271	8,271	8,253	8,650
ΔR^2	1.17%	0.89%	0.96%	0.31%	0.63%	0.31%	0.36%	0.24%	0.08%	0.10%
95% CI - low	0.75%	0.54%	0.58%	0.12%	0.35%	0.12%	0.15%	0.07%	0.01%	0.01%
95% CI - high	1.71%	1.33%	1.38%	0.57%	1.00%	0.59%	0.67%	0.50%	0.23%	0.28%

Note: SWB is the combined well-being measure. PA is positive affect. LS is life satisfaction. DS is depressive symptoms. ΔR^2 is the incremental R^2 from including the SWB polygenic score (constructed using all available HapMap3 SNPs) in a regression of the respective phenotype (residualized on sex, age, age^2, and interactions) on the first 10 PCs. For height, birthyear was used instead of age.

Table A9. Summary of cohort restrictions in proxy-phenotype analyses.

Panel A. SWB Cohorts

Cohort Name	Phenotype	*N*/Weight	(SWB → DS)	(SWB → Neuroticism)	(SWB → Height)
AGES	WB	3,044	Yes	Yes	No
ALSPAC	WB	5,654	Yes	No	Yes
EGCUTOMNI	WB	585	Yes	No	No
EGCUT370	WB	1,109	Yes	No	No
HNRSomni1	WB	773	No	Yes	Yes
HNRSoexpr	WB	1,343	No	Yes	Yes
HRS	WB	9,942	Yes	Yes	Yes
NTR	WB	8,060	No	No	No
RUSHMAP	WB	370	Yes	Yes	Yes
TEDS	WB	2,148	Yes	Yes	Yes
TRAILS	WB	1,204	Yes	Yes	Yes
YFS	WB	1,711	Yes	No	Yes
1958T1D	LS	2,029	Yes	Yes	No
1958WTC	LS	2,188	Yes	Yes	No
BASE	LS	1,395	Yes	Yes	Yes
HPFSCHD	PA	854	Yes	Yes	Yes
HPFSKS	PA	478	Yes	Yes	Yes
HPFST2D	PA	1,653	Yes	Yes	Yes
KORAF3	PA	827	No	Yes	No
KORAF4	PA	660	No	Yes	No
NHSBRCA	PA	2,268	Yes	Yes	No
NHSCHD	PA	1,116	Yes	Yes	No
NHSKS	PA	490	Yes	Yes	No
NHST2D	PA	3,071	Yes	Yes	No
ASPS	PA	815	Yes	Yes	Yes
AEGS1	PA	517	Yes	Yes	Yes
BLSA	PA	827	Yes	No	Yes
Colaus	PA	3,530	Yes	Yes	No
CROATIAKOR	PA	876	Yes	No	Yes
CROATIASPL	PA	495	Yes	Yes	Yes
CROATIAVIS	PA	918	Yes	No	No
DECODE	PA	10,383	Yes	Yes	No
DHS	PA	879	Yes	Yes	Yes
ERF	PA	2,129	Yes	No	No
FTC	PA	685	Yes	No	No
FINTWIN2014	PA	1,805	Yes	No	No

HBCS	PA	1,450	Yes	No	Yes
HCS	PA	1,906	Yes	Yes	Yes
LL	PA	11,971	Yes	Yes	Yes
MCTFR	PA	7,007	Yes	No	Yes
MESA	PA	2,495	Yes	Yes	Yes
NEO	PA	5,715	Yes	Yes	Yes
QIMR	PA	6,572	No	No	Yes
QIMR2	PA	1,235	No	No	Yes
RSI	PA	3,731	Yes	Yes	No
RSII	PA	2,039	Yes	Yes	Yes
RSIII	PA	2,027	Yes	Yes	Yes
SAR	PA	2,542	Yes	No	No
STR1	PA	9,343	Yes	No	Yes
STR2	PA	4,704	Yes	No	Yes
TWINSUK	PA	3,484	Yes	Yes	No
UKB	PA	40,543	Yes	Yes	Yes
NFBC	LS	5,203	Yes	Yes	No
23andMe	LS	93,454	Yes	Yes	Yes
ARIC	LS	8,716	Yes	Yes	No
ELSA	LS	5,047	Yes	Yes	Yes
GOYA	LS	1,179	Yes	Yes	Yes
LBC21	LS	443	Yes	No	Yes
LBC36	LS	789	Yes	No	Yes
N			278,956	239,982	229,853
λ			1.015	1.018	1.009
Minimum *N*:			140,000	120,000	115,000

Panel B. DS Cohorts

Cohort Name	Phenotype	N_{cases} / $N_{controls}$	*N*	$N_{effective}$/ Weight	(SWB →DS)
GERA	Major Depression	7,231 / 49,137	56,368	25,214	Yes
PGC	Major Depression	9,240 / 9,519	18,759	18,755	Yes
UKB sans SWB	Continuous Measure		67,138	67,138	Yes
UKB Full	Continuous Measure		105,739	105,739	No
N					142,265
λ					1.010
#Cohorts					3
Minimum *N*:					40,000

Panel C. Neuroticism Cohorts

Cohort Name	Phenotype	*N*/Weight	(SWB → Neuroticism)
GPC	Neuroticism	63,666	Yes
UKB	Neuroticism	107,245	No
UKB sans SWB	Neuroticism	68,201	Yes
N			0
λ			1.002
#Cohorts			2
Minimum *N*:			70,000

Panel D. Height Cohorts

Cohort Name	Phenotype	*N*	(SWB → Height)
Lango Allen	Height	133,859	Yes
N			133,859
λ			NA
#Cohorts			1
Minimum *N*:			100,000

Note: SWB is the combined well-being measure. LS is life satisfaction. PA is positive affect. DS is depressive symptoms. λ is the estimated LD score intercept used to adjust the standard errors. As in the main analyses, the meta-analyses are run without cohort-level genomic control. Minimum *N* is the smallest sample size (effective sample size in Panel B) required for a SNP to be included in the meta-analysis.. # Cohorts: minimum number of cohorts in which SNP must be available to be eligible for inclusion in lookup exercise.

Table A10. SNPs that reached *p*-value< 10^{-6} in the meta-analysis of 54 DS-associated SNPs with results from companion study on major depressive symptoms (Hyde et al., 2016).

		Discovery			Companion			Companion + Discovery		
SNPID	*EA*	*Z*	*p*-value	*N*	*Z*	*p*-value	*N*	*Z*	*p*-value	*N*
rs782212	T	-4.850	1.23×10^{-6}	162,107	-4.544	5.51×10^{-6}	368,890	-6.516	7.19×10^{-11}	530,997
rs4810896	A	-4.912	9.01×10^{-7}	180,866	-3.635	2.78×10^{-4}	368,890	-5.879	4.11×10^{-9}	549,756
rs10809520	T	-4.494	6.98×10^{-6}	162,107	-3.942	8.07×10^{-5}	368,890	-5.820	5.89×10^{-9}	530,997
rs62100776	A	-5.759	8.46×10^{-9}	105,739	-2.848	4.39×10^{-3}	214,692	-5.664	1.47×10^{-8}	320,431
rs7973260	A	6.017	1.77×10^{-9}	124,498	2.424	1.53×10^{-2}	368,890	5.601	2.12×10^{-8}	493,388
rs1690818	T	-5.066	4.06×10^{-7}	162,107	-3.024	2.49×10^{-3}	368,890	-5.405	6.47×10^{-8}	530,997
rs10233018	A	-5.235	1.64×10^{-7}	162,107	-2.831	4.63×10^{-3}	368,890	-5.346	8.97×10^{-8}	530,997
rs853679	A	-4.972	6.62×10^{-7}	180,866	-2.854	4.32×10^{-3}	368,890	-5.298	1.16×10^{-7}	549,756
rs10884216	T	-4.920	8.65×10^{-7}	162,107	-2.624	8.68×10^{-3}	368,890	-4.995	5.89×10^{-7}	530,997
rs1961982	A	-4.730	2.24×10^{-6}	105,739	-2.732	6.30×10^{-3}	368,890	-4.971	6.64×10^{-7}	474,629

Note: DS is depressive symptoms. *EA* is effect allele. Standard errors are adjusted using the estimated LD score intercept, which is 1.060. All effect sizes are reported in units of SDs per allele. Cohorts are weighted by effective sample size; for details, see Section 3.4.4-F.

APPENDIX B

Supplementary Tables to Chapter 4

Table B1. Description of participating cohorts and phenotype distribution.

Cohort	Full name	Country	*EduYears* (Mean/SD)	Fraction *College*	Sample size
ACPRC	Manchester Studies of Cognitive Ageing	England	14.8 (4.0)	0.41	1713
AGES	Age, Gene/ Environment Susceptibility–Reykjavik Study	Iceland	13.6 (4.9)	0.27	3212
ALSPAC	Avon Longitudinal Study of Parents and Children	England	14.3 (4.0)	0.26	2877
ASPS	Austrian Stroke Prevention Study	Austria	11.6 (2.9)	0.09	777
BASE-II	Berlin Aging Study II	Germany	16.8 (3.1)	0.62	1619
CoLaus	Cohorte Lausannoise	Switzerland	15.0 (3.5)	0.33	3269
COPSAC2000	Copenhagen Studies on Asthma in Childhood 2000	Germany	17.3 (3.6)	0.53	318
CROATIA-Korčula	Croatia Korčula	Croatia	12.5 (3.3)	0.11	842
deCODE	deCODE genetics	Iceland	13.4 (3.3)	0.36	46758
DHS	Dortmund Health Study	Germany	13.8 (3.3)	0.18	953
DIL	Wellcome Trust Diabetes and Inflammation Laboratory	England	14.0 (4.5)	0.31	2578
EGCUT1	Estonian Genome Center, University of Tartu	Estonia	14.1 (3.7)	0.22	5597
EGCUT2	Same as above	Estonia	14.9 (3.5)	0.29	1328
EGCUT3	Same as above	Estonia	15.8 (3.0)	0.36	2047
ERF	Erasmus Rucphen Family Study	Netherlands	10.2 (3.4)	0.05	2433
FamHS	Family Heart Study	USA	15.6 (2.8)	0.33	3483
FINRISK	The National FINRISK Study	Finland	13.8 (5.1)	0.36	1685
FTC	Finnish Twin Cohort	Finland	11.9 (3.4)	0.09	2418
GOYA	Genetics of Overweight Young Adults	Denmark	14.4 (4.6)	0.35	1459
GRAPHIC	Genetic Regulation of Arterial Pressure in Humans	England	14.9 (3.1)	0.22	727
GS	Generation Scotland	Scotland	14.0 (4.1)	0.3	8776
H2000 Cases	Health 2000	Finland	12.2 (4.1)	0.11	797
H2000 Controls	Same as above	Finland	12.7 (4.2)	0.15	819
HBCS	Helsinki Birth Cohort Study	Finland	11.7 (4.5)	0.21	1617
HCS	Hunter Community Study	Australia	14.2 (3.3)	0.23	1946
HNRS (CorexB)	Heinz Nixdorf Recall Study	Germany	14.4 (3.5)	0.25	1401
HNRS (Oexpr)	Same as above	Germany	15.2 (3.6)	0.35	1347
HNRS (Omni1)	Same as above	Germany	14.9 (3.6)	0.31	778

HRS	Health and Retirement Study	USA	14.3 (3.2)	0.25	9963
Hypergenes	Hypergenes	Italy/ UK/ Belgium	12.3 (3.7)	0.15	815
INGI-CARL	Italian Network of Genetic Isolates - Carlantino	Italy	9.7 (3.4)	0.07	947
INGI-FVG	Italian Network of Genetic Isolates - Friuli Venezia Giulia	Italy	9.5 (2.6)	0.02	943
KORA S3	Kooperative Gesundheits-forschung in der Region Augsburg	Germany	13.7 (2.6)	0.14	2655
KORA S4	Same as above	Germany	14.1 (2.8)	0.16	2721
LBC1921	Lothian Birth Cohort 1921	Scotland	11.3 (1.8)	0.02	515
LBC1936	Lothian Birth Cohort 1936	Scotland	13.2 (3.7)	0.15	1003
LifeLines	The LifeLines Cohort Study	Netherlands	13.5 (4.2)	0.26	12539
MCTFR	Minnesota Center for Twin and Family Research	USA	15.9 (3.3)	0.37	3819
MGS	Molecular Genetics of Schizophrenia	USA	15.6 (3.2)	0.36	2313
MoBa	Mother and Child Cohort of NIPH	Norway	17.6 (4.0)	0.7	622
NBS	Nijmegen Biomedical Study	Netherlands	12.9 (4.3)	0.29	1808
NESDA	Netherlands Study of Depression and Anxiety	Netherlands	14.2 (4.9)	0.38	1820
NFBC66	Northern Finland Birth Cohort 1966	Finland	13.9 (2.4)	0.15	5297
NTR	Netherlands Twin Register	Netherlands	14.5 (4.9)	0.45	5246
OGP	Ogliastra Genetic Park	Italy	10.0 (3.5)	0.05	370
OGP-Talana	Ogliastra Genetic Park-Talana	Italy	9.1 (3.1)	0.01	544
ORCADES	Orkney Complex Disease Study	Scotland	13.0 (3.3)	0.15	1828
PREVEND	Prevention of Renal and Vascular End-stage Disease	Netherlands	12.0 (4.7)	0.27	3578
QIMR	Queensland Institute of Medical Research	Australia	17.3 (3.2)	0.58	8006
RS-I	Rotterdam Study Baseline	Netherlands	10.4 (3.5)	0.08	6108
RS-II	Rotterdam Study Extension of Baseline	Netherlands	12.5 (3.7)	0.2	1667
RS-III	Rotterdam Study Young	Netherlands	12.9 (4.2)	0.26	3040
Rush-MAP	Rush University Medical Center - Memory and Aging Project	USA	13.3 (2.9)	0.09	887

Rush-ROS	Rush University Medical Center - Religious Orders Study	USA	16.6 (3.3)	0.44	808
SardiNIA	SardiNIA Study of Aging	Italy	10.4 (4.2)	0.09	5616
SHIP	Study of Health in Pomerania	Germany	12.8 (3.7)	0.13	3556
SHIP-TREND	Study of Health in Pomerania	Germany	14.5 (2.7)	0.19	901
STR – Salty	Swedish Twin Registry	Sweden	14.3 (3.7)	0.31	4832
STR – Twingene	Swedish Twin Registry	Sweden	12.9 (4.2)	0.24	9553
THISEAS	The Hellenic Study of Interactions between SNPs & Eating in Atherosclerosis Susceptibility	Greece	13.2 (3.1)	0.17	829
TwinsUK	St Thomas' UK Adult Twin Registry	England	12.9 (4.8)	0.27	4012
WTCCC58C	1958 British Birth Cohort	England	14.1 (4.5)	0.31	2804
YFS	The Cardiovascular Risk in Young Finns Study	Finland	15.8 (3.0)	0.38	2029
23andMe	23andMe, inc	Primarily US	16.0 (3.3)	0.4	76155

Notes: With the exception of 23andMe and FamHS, the sample size used in the *College* analyses is identical to the sample size used in the *EduYears*. The FamHS sample size is smaller (N=1218) because the cohort elected to run the *College* analyses in a subsample of conventionally unrelated individuals. The 23andMe sample size is 75907 in the *College* analyses and 76155 in the *EduYears* analyses because the *College* and *EduYears* variables in this cohort were generated using responses to different survey questions.

Table B2. Description of SNP filtering in *EduYears* analyses

Cohort	SNPs before QC	#1 Misaligned SNPs in reference panel	#2 Variable quality	#3 MAC<25 / $R^2 > 0.05$	#4 Imputation accuracy	#5 Non-autosomal marker/ INDEL	#6 Invalid or duplicated position / Allele mismatch	SNPs after QC	λ_{GC}
ACPRC	5,762,631	0	357	130,931	487	0	112	5,630,744	1.015
AGES	30,061,896	0	13,704,096	5,137,569	2,971,618	659,350	23,390	7,565,873	1.048
ALSPAC	27,449,291	653	11,830,580	5,105,051	1,110,248	759,991	12,602	8,630,166	1.037
ASPS	9,356,818	0	550	842,853	1,111,039	611,370	10	6,790,996	1.006
BASE-II	30,072,216	0	16,638,032	3,646,376	2,425,263	591,550	1,507	6,769,488	1.007
CoLaus	29,356,512	0	8,238,659	9,266,534	4,549,189	572,443	19,169	6,710,518	1.026
COPSAC2000	30,061,896	0	16,702,811	6,562,711	422,351	555,758	0	5,818,265	0.978
CROATIA-Korčula	37,944,789	0	19,646,773	9,653,841	984,755	652,688	5,868	7,000,864	1.003
deCODE	19,543,940	0	244,133	2,236,281	6,451,732	788,727	353,904	9,469,163	0.975
DHS	30,077,782	0	20,683,359	506,092	400,482	731,812	916	7,755,121	1.007
DIL	17,694,930	0	1,361	6,908,043	1,590,809	742,731	20,593	8,431,393	1.013
EGCUT1	16,380,884	0	0	4,241,713	2,313,202	776,099	56,689	8,993,181	1.026
EGCUT2	14,330,257	0	0	4,883,973	1,116,291	702,202	4,208	7,623,583	1.012
EGCUT3	16,225,699	0	0	5,905,295	2,138,630	670,958	14,849	7,495,967	1.012
ERF	29,974,334	0	9,555,783	8,747,357	5,864,074	455,812	35,672	5,315,636	1.051
FamHS	36,545,314	0	13,088,451	11,837,664	3,035,700	684,140	15,211	7,884,148	1.026
FINRISK	14,687,799	0	1,746	4,807,684	701,824	772,641	13,082	8,390,822	1.001
FTC	8,373,047	0	0	4,344	232,721	19,027	21,631	8,095,324	1.001
GOYA	30,059,051	0	15,511,478	5,632,336	761,499	678,575	1,458	7,473,705	1.004
GRAPHIC	13,888,387	0	0	5,517,234	568,758	675,000	1,362	7,126,033	1.003

GS	38,073,166	658	13,604,576	10,839,285	3,336,122	781,496	124,833	9,386,196	1.017
H2000 Cases	9,483,122	0	2,070	3,219,565	281,257	525,769	2,150	5,452,311	0.992
H2000 Controls	13,162,729	0	3,336	4,520,095	372,652	723,080	3,246	7,540,320	1.007
HBCS	39,282,667	0	24,148,312	4,968,565	773,076	1,012,715	14,692	8,365,307	1.025
HCS	9,056,792	0	135,903	89,932	1,336,452	425,526	0	7,068,979	1.016
HNRS (CorexB)	20,050,734	0	10,286	9,966,217	1,154,779	1,009,278	337,467	7,572,707	1.002
HNRS (Oexpr)	18,953,852	0	3,878	9,086,498	526,945	1,068,691	362,359	7,905,481	1.009
HNRS (Omni1)	16,449,949	0	2,454	7,552,352	353,377	1,009,504	316,798	7,215,464	1.007
HRS	17,917,583	0	0	6,067,296	1,531,716	836,747	84,625	9,397,199	1.035
Hypergenes	30,061,059	0	12,719,615	8,818,467	770,179	661,281	113	7,091,404	1.003
INGI-CARL	9,259,987	0	0	501,036	1,140,542	636,650	37,218	6,944,541	0.946
INGI-FVG	8,483,800	0	0	161,768	729,091	644,278	6,710	6,941,953	0.995
KORA S3	16,499,238	0	0	5,636,956	1,578,135	746,929	13,798	8,523,420	1.013
KORA S4	16,906,023	0	0	5,954,725	1,793,551	734,500	15,427	8,407,820	1.011
LBC1921	38,050,692	0	20,539,333	9,833,468	825,407	587,759	9	6,264,716	1.013
LBC1936	38,050,692	0	21,283,428	7,949,405	1,267,485	634,950	453	6,914,971	1.008
LifeLines	8,167,165	0	0	0	2,841,687	0	313	5,325,165	1.072
MCTFR	10,011,985	0	76,811	264,741	2,717,097	0	0	6,953,336	1.023
MGS	30,019,281	0	14,553,477	4,939,447	2,614,382	635,047	2,712	7,274,216	0.999
MoBa	13,322,167	0	0	5,279,405	481,898	659,112	196	6,901,556	1.001
NBS	38,037,370	0	21,548,646	6,407,437	1,615,645	691,958	7,309	7,766,375	1.070
NESDA	29,840,204	0	16,386,959	3,678,982	2,161,659	619,062	23,004	6,970,538	1.010
NFBC66	38,031,723	0	21,200,557	4,973,640	1,899,225	793,564	82,510	9,082,227	1.060
NTR	10,952,826	0	0	1,164,319	1,867,654	630,799	36,665	7,253,389	1.053
OGP	29,418,125	645	13,754,730	8,596,219	968,886	521,788	1,938	5,573,919	0.976
OGP-Talana	25,293,773	633	11,437,371	6,581,764	1,183,154	535,093	40,396	5,515,362	0.963
ORCADES	37,868,226	0	17,341,695	10,460,004	2,026,300	644,987	18,873	7,376,367	1.011
PREVEND	17,692,306	627	0	6,820,200	1,685,999	738,103	30,633	8,416,744	1.055

QIMR	8,448,373	0	22	0	1,172,502	602,674	301	6,672,874	1.024
RS-I	30,071,840	0	8,678,058	8,441,043	3,864,741	710,136	36,641	8,341,221	1.025
RS-II	30,071,857	0	12,614,895	7,655,729	1,547,612	681,192	2,729	7,569,700	1.010
RS-III	30,071,835	0	11,365,727	7,612,705	2,320,243	702,377	12,726	8,058,057	1.045
Rush-MAP	37,426,733	0	29,621,356	293,621	1,386,742	0	66,874	6,058,140	0.986
Rush-ROS	37,426,733	0	29,635,764	389,857	1,346,914	0	66,092	5,988,106	1.014
SardiNIA	8,340,764	0	0	0	35,364	0	636,468	7,668,932	1.000
SHIP	17,533,349	0	863,428	4,976,161	2,086,772	995,906	9,722	8,601,360	1.024
SHIP-TREND	17,585,496	0	2,234,356	6,255,705	403,696	975,317	42,628	7,673,794	1.009
STR – Salty	11,077,616	0	100,736	50,994	2,839,827	651,122	4,289	7,430,648	1.000
STR – Twingene	10,771,271	0	64,604	30,003	1,657,235	725,813	7,467	8,286,149	1.007
THISEAS	30,697,962	0	15,769,553	6,113,168	804,441	881,401	1,832	7,127,567	1.015
TwinsUK	15,369,515	0	41,514	4,459,413	1,373,740	771,908	55,075	8,667,865	1.006
WTCCC58C	18,291,939	0	2,541	6,952,894	1,422,530	1,049,105	36,790	8,828,079	1.012
YFS	15,289,292	0	0	4,778,499	873,552	1,034,859	23,563	8,578,819	0.984
23andMe	11,972,722	604	2	130	714,121	1,146,282	325,731	9,785,852	1.146

Note: "Misaligned SNPs in reference panel" refers to the 730+199 SNPs whose strands were incorrectly aligned in the September or December 2013 releases of the 1000 Genomes Phase 1 haplotypes provided by the software IMPUTE2. "Variable quality" refers to missing or invalid values for effect allele frequency, Beta, SE, *P*-value, imputation accuracy, or the imputed/genotyped indicator. R^2 refers to the variance explained in *EduYears*. The imputation accuracy thresholds used are 0.6 for MaCH, 0.7 for IMPUTE2, and 0.8 for PLINK. deCODE standard errors were adjusted for genomic control prior to uploading results. The original λ_{GC} was 1.93.

Table B3. Sign concordance of effects with deCODE sample in *EduYears* analyses

Cohort	# SNPs checked	# SNPs aligned	Percent aligned	N
deCODE	1066	1066	100.0	46758
23andMe	968	536	55.4	76155
ACPRC	395	198	50.1	1713
AGES	755	560	74.2	3212
ALSPAC	939	483	51.4	2877
ASPS	647	291	45.0	777
BASE-II	562	286	50.9	1619
CoLaus	541	272	50.3	3269
COPSAC2000	549	281	51.2	318
CROATIA-Korčula	700	359	51.3	842
DHS	890	444	49.9	953
DIL	929	472	50.8	2578
EGCUT1	933	495	53.1	5597
EGCUT2	810	411	50.7	1328
EGCUT3	737	387	52.5	2047
ERF	389	196	50.4	2433
FamHS	794	415	52.3	3483
FINRISK	903	453	50.2	1685
FTC	856	415	48.5	2418
GOYA	795	397	49.9	1459
GRAPHIC	782	397	50.8	727
GS	962	485	50.4	8776
H2000 Cases	588	302	51.4	797
H2000 Controls	808	395	48.9	819
HBCS	911	461	50.6	1617
HCS	734	377	51.4	1946
HNRS (CorexB)	803	416	51.8	1401
HNRS (Oexpr)	885	419	47.3	1347
HNRS (Omni1)	801	433	54.1	778
HRS	989	511	51.7	9963
Hypergenes	734	351	47.8	815
INGI-CARL	646	327	50.6	947
INGI-FVG	695	337	48.5	943
KORA S3	938	476	50.7	2655
KORA S4	899	429	47.7	2721
LBC1921	605	298	49.3	515
LBC1936	701	365	52.1	1003
LifeLines	356	196	55.1	12539
MCTFR	530	278	52.5	3819
MGS	666	364	54.7	2313
MoBa	761	406	53.4	622
NBS	830	441	53.1	1808

NESDA	672	354	52.7	1820
NFBC66	935	486	52.0	5297
NTR	674	342	50.7	5246
OGP	496	257	51.8	370
OGP-Talana	489	266	54.4	544
ORCADES	756	375	49.6	1828
PREVEND	868	437	50.3	3578
QIMR	622	342	55.0	8006
RS-I	826	422	51.1	6108
RS-II	798	390	48.9	1667
RS-III	835	424	50.8	3040
Rush-MAP	406	202	49.8	887
Rush-ROS	405	202	49.9	808
SardiNIA	784	385	49.1	5616
SHIP	894	467	52.2	3556
SHIP-TREND	869	444	51.1	901
STR – Salty	685	340	49.6	4832
STR – Twingene	896	466	52.0	9553
THISEAS	744	363	48.8	829
TwinsUK	920	469	51.0	4012
WTCCC58C	967	490	50.7	2804
YFS	917	472	51.5	2029

Note: A pruned set of 1066 SNPs from the deCODE results file whose *P*-value for the test of association with *EduYears* was smaller than 0.01 were used to check for sign concordance.

Table B4. Estimated genetic correlation with deCODE sample in *EduYears*.

Cohort	ρ_G	S.E.	*Z*-score	*P*-value	*N*
23andMe	0.802	0.066	12.246	<0.0001	76155
ASPS	0.397	0.947	0.419	0.68	777
DHS	0.505	0.485	1.042	0.3	953
EGCUT1	0.917	0.356	2.577	<0.01	5597
EGCUT2	0.616	0.221	2.789	<0.005	1328
EGCUT3	0.337	0.177	1.907	0.06	2047
FamHS	0.224	0.135	1.666	0.1	3483
FVG	0.212	0.276	0.767	0.44	943
HBCS	0.805	0.255	3.159	0.002	1617
HNRS (Oexpr)	0.494	0.273	1.809	0.07	1347
HRS	0.933	0.184	5.058	<0.0001	9963
KORA3	0.769	0.444	1.732	0.08	2655
LBC1921	0.812	1.295	0.627	0.53	515
LBC1936	0.943	2.649	0.356	0.72	1003
LifeLines	0.820	0.173	4.734	<0.0001	12539
Rush-MAP	-0.293	0.700	-0.419	0.68	887
MCTFR	0.815	0.297	2.743	0.01	3819
NBS	0.494	0.240	2.059	0.04	1808
NTR	0.682	0.171	3.994	<0.0001	5246
OGPTALANA	0.038	0.335	0.114	0.91	544
Prevend	0.559	0.167	3.345	0.0008	3578

Note: ρ_G is the genetic correlation between deCODE and the respective cohort, computed using the LDSC python software package and the "eur_w_ld_chr/" files of LD scores calculated by Finucane et al. (2015). In the LD score regressions, we include only HapMap3 SNPs with MAF > 0.01, and the standard errors of the LD Score regressions are estimated using a block jackknife over SNPs (by the LDSC software).

Table B5. Association results for the 74 independent SNPs that reached genome-wide significance ($P < 5\times10^{-8}$) in the pooled-sex EduYears meta-analysis

				Pooled				Male		Female	
SNP	Chr	Position	*EA*	*EAF*	Beta	*P*-value	Heterogeneity *P*-value	Beta	*P*-value	Beta	*P*-value
rs301800	1	8490603	T	0.18	0.019	1.79E-08	0.516	0.022	2.19E-06	0.022	2.19E-06
rs11210860	1	43982527	A	0.37	0.017	2.36E-10	0.796	0.019	2.83E-07	0.019	2.83E-07
rs34305371	1	72733610	A	0.09	0.035	3.76E-14	0.071	0.04	2.85E-10	0.04	2.85E-10
rs2568955	1	72762169	T	0.24	-0.017	1.80E-08	0.841	-0.02	1.11E-06	-0.02	1.11E-06
rs1008078	1	91189731	T	0.41	-0.016	6.01E-10	0.07	-0.016	7.50E-06	-0.016	7.50E-06
rs11588857	1	204587047	A	0.21	0.02	5.27E-10	0.459	0.019	5.37E-06	0.019	5.37E-06
rs1777827	1	211613114	A	0.59	0.015	1.55E-08	0.4	0.011	1.32E-03	0.011	1.32E-03
rs2992632	1	243503764	A	0.72	0.017	8.23E-09	0.78	0.017	2.00E-05	0.017	2.00E-05
rs76076331	2	10977585	T	0.15	0.02	3.63E-08	0.735	0.024	1.76E-06	0.024	1.76E-06
rs11689269	2	15621917	C	0.33	0.016	1.28E-08	0.484	0.017	5.78E-06	0.017	5.78E-06
rs1606974	2	51873599	A	0.12	0.022	2.80E-08	0.092	0.016	2.66E-03	0.016	2.66E-03
rs11690172	2	57387094	A	0.59	0.015	1.99E-08	0.141	0.018	5.33E-07	0.018	5.33E-07
rs2457660	2	60757419	T	0.64	-0.017	7.11E-10	0.821	-0.013	3.68E-04	-0.013	3.68E-04
rs114598875	2	60976384	A	0.84	-0.02	2.41E-08	0.991	-0.022	3.14E-06	-0.022	3.14E-06
rs10496091	2	61482261	A	0.29	-0.018	5.62E-10	0.252	-0.014	2.42E-04	-0.014	2.42E-04
rs13402908	2	100333377	T	0.46	-0.018	1.70E-11	0.992	-0.02	1.37E-08	-0.02	1.37E-08
rs4851251	2	100753490	T	0.27	-0.017	1.91E-08	0.46	-0.016	6.85E-05	-0.016	6.85E-05
rs12987662	2	100821548	A	0.39	0.027	2.69E-24	0.09	0.028	1.17E-14	0.028	1.17E-14
rs17824247	2	144152539	T	0.59	-0.016	2.77E-09	0.571	-0.014	1.20E-04	-0.014	1.20E-04
rs16845580	2	161920884	T	0.63	0.016	2.65E-09	0.872	0.016	1.18E-05	0.016	1.18E-05
rs4500960	2	162818621	T	0.46	-0.016	3.75E-10	0.591	-0.02	1.89E-08	-0.02	1.89E-08

rs6739979	2	193731929	T	0.63	-0.015	4.70E-08	0.586	-0.01	7.02E-03	-0.01	7.02E-03
rs2245901	2	194296294	A	0.4	-0.016	4.54E-09	0.863	-0.014	8.63E-05	-0.014	8.63E-05
rs55830725	2	237056854	A	0.17	-0.022	5.37E-10	0.403	-0.019	5.21E-05	-0.019	5.21E-05
rs35761247	3	48623124	A	0.05	0.034	3.82E-08	0.488	0.033	7.86E-05	0.033	7.86E-05
rs62259535	3	48939052	A	0.96	0.048	2.63E-09	0.415	0.048	1.11E-05	0.048	1.11E-05
rs148734725	3	49406708	A	0.32	0.025	1.36E-18	0.391	0.028	1.12E-13	0.028	1.12E-13
rs11712056	3	49914397	T	0.57	0.024	3.30E-19	0.001	0.025	6.02E-13	0.025	6.02E-13
rs112634398	3	50075494	A	0.95	0.036	4.61E-08	0.878	0.041	3.96E-06	0.041	3.96E-06
rs62263923	3	85674790	A	0.64	-0.016	7.01E-09	0.154	-0.012	6.62E-04	-0.012	6.62E-04
rs6799130	3	160847801	C	0.52	-0.015	2.82E-08	0.041	-0.013	3.00E-04	-0.013	3.00E-04
rs12646808	4	3249828	T	0.66	0.016	4.00E-08	0.148	0.018	4.06E-06	0.018	4.06E-06
rs2610986	4	18037231	T	0.67	-0.016	2.01E-08	0.398	-0.018	4.17E-06	-0.018	4.17E-06
rs34072092	4	28801221	T	0.9	0.024	3.91E-08	0.364	0.017	3.28E-03	0.017	3.28E-03
rs3101246	4	42649935	T	0.6	-0.015	1.43E-08	0.354	-0.022	1.96E-09	-0.022	1.96E-09
rs4863692	4	140764124	T	0.31	0.018	1.56E-10	0.371	0.024	1.66E-10	0.024	1.66E-10
rs4493682	5	45188024	C	0.17	0.019	3.32E-08	0.959	0.026	3.71E-08	0.026	3.71E-08
rs2964197	5	57535206	T	0.5	0.015	3.02E-08	0.329	0.012	4.19E-04	0.012	4.19E-04
rs61160187	5	60111579	A	0.61	-0.017	3.49E-10	0.129	-0.018	8.38E-07	-0.018	8.38E-07
rs324886	5	87896602	T	0.39	-0.015	1.91E-08	0.485	-0.019	1.77E-07	-0.019	1.77E-07
rs10061788	5	87934707	A	0.18	0.021	2.46E-09	0.545	0.02	1.73E-05	0.02	1.73E-05
rs2431108	5	103947968	T	0.68	0.016	5.27E-09	0.534	0.014	1.98E-04	0.014	1.98E-04
rs1402025	5	113987898	T	0.78	0.017	3.42E-08	0.227	0.015	4.75E-04	0.015	4.75E-04
rs62379838	5	120102028	T	0.69	0.016	3.30E-08	0.811	0.021	4.83E-08	0.021	4.83E-08
rs56231335	6	98187291	T	0.67	-0.017	2.07E-09	0.387	-0.02	1.11E-07	-0.02	1.11E-07
rs9320913	6	98584733	A	0.48	0.024	2.46E-19	0.717	0.027	3.25E-14	0.027	3.25E-14
rs7767938	6	153367613	T	0.75	0.017	2.44E-08	0.662	0.018	1.25E-05	0.018	1.25E-05
rs2615691	7	23402104	A	0.04	-0.037	4.71E-08	0.961	-0.041	1.04E-05	-0.041	1.04E-05
rs12531458	7	39090698	A	0.51	0.014	3.11E-08	0.3	0.011	1.34E-03	0.011	1.34E-03
rs12671937	7	92654365	A	0.53	0.016	9.15E-10	0.763	0.021	9.51E-09	0.021	9.51E-09
rs113520408	7	128402782	A	0.27	0.017	1.97E-08	0.27	0.019	1.60E-06	0.019	1.60E-06

rs17167170	7	133302345	A	0.8	0.02	1.14E-09	0.818	0.022	4.25E-07	0.022	4.25E-07
rs11768238	7	135227513	A	0.34	-0.017	9.90E-10	0.802	-0.019	5.12E-07	-0.019	5.12E-07
rs12682297	8	145712860	A	0.46	-0.016	3.93E-09	0.372	-0.015	1.25E-05	-0.015	1.25E-05
rs1871109	9	1746016	T	0.55	-0.016	4.35E-10	0.335	-0.015	1.48E-05	-0.015	1.48E-05
rs13294439	9	23358875	A	0.59	-0.023	2.20E-17	0.328	-0.022	3.93E-09	-0.022	3.93E-09
rs895606	9	88003668	A	0.45	0.015	2.25E-08	0.176	0.018	1.44E-07	0.018	1.44E-07
rs7854982	9	124644562	T	0.46	-0.015	1.29E-08	0.902	-0.02	1.42E-08	-0.02	1.42E-08
rs11191193	10	103802408	A	0.66	0.018	5.44E-11	0.976	0.017	3.77E-06	0.017	3.77E-06
rs12772375	10	104082688	T	0.4	-0.015	1.56E-08	0.178	-0.016	1.31E-05	-0.016	1.31E-05
rs7945718	11	12748819	A	0.63	0.015	1.54E-08	0.638	0.012	9.47E-04	0.012	9.47E-04
rs7955289	12	14653667	A	0.61	0.017	4.49E-10	0.816	0.015	3.31E-05	0.015	3.31E-05
rs2456973	12	56416928	A	0.67	-0.02	1.06E-12	0.239	-0.02	1.26E-07	-0.02	1.26E-07
rs7131944	12	92159557	A	0.62	0.015	9.02E-09	0.262	0.018	4.10E-07	0.018	4.10E-07
rs572016	12	121279083	A	0.51	0.014	3.46E-08	0.587	0.011	1.83E-03	0.011	1.83E-03
rs7306755	12	123767929	A	0.21	0.023	1.26E-12	0.668	0.027	1.76E-10	0.027	1.76E-10
rs9537821	13	58402771	A	0.72	0.024	1.50E-16	0.993	0.03	1.48E-14	0.03	1.48E-14
rs1043209	14	23373986	A	0.61	0.018	1.82E-11	0.721	0.018	3.15E-07	0.018	3.15E-07
rs8005528	14	27098611	A	0.75	-0.018	7.19E-09	0.631	-0.017	4.58E-05	-0.017	4.58E-05
rs17119973	14	84913111	A	0.26	-0.019	3.55E-10	0.013	-0.014	3.23E-04	-0.014	3.23E-04
rs192818565	17	43991515	T	0.81	0.025	1.47E-12	0.174	0.025	1.88E-07	0.025	1.88E-07
rs12969294	18	35186122	A	0.34	-0.016	7.24E-09	0.317	-0.015	8.24E-05	-0.015	8.24E-05
rs2837992	21	42620520	T	0.39	0.015	3.80E-08	0.565	0.015	3.41E-05	0.015	3.41E-05
rs165633	22	29880773	A	0.74	-0.018	2.86E-09	0.287	-0.015	1.92E-04	-0.015	1.92E-04

Note: *EA* is effect allele. *EAF* is effect allele frequency.

Table B6. Within-sample replication analyses

Panel A

				Rietveld et al. (2013)			Rietveld Cohorts			New Cohorts		
SNP	Chr	Position	A1	Beta	S.E.	*P*-value	Beta	S.E.	*P*-value	Beta	SE	*P*-value
rs11584700	1	204576983	G	0.024	0.005	3.25E-07	0.028	0.005	4.3E-09	0.012	0.004	3.8E-03
rs4851266	2	100818479	T	0.020	0.004	5.61E-07	0.021	0.004	1.8E-07	0.031	0.004	4.8E-18
rs7309	2	162092640	G	0.022	0.004	3.60E-08	0.019	0.004	1.2E-06	0.011	0.003	1.1E-03
rs11687170	2	237058144	T	0.027	0.005	3.25E-08	0.024	0.005	3.8E-06	0.019	0.005	5.9E-05
rs13401104	2	237105518	G	0.027	0.005	4.74E-08	0.027	0.005	2.2E-07	0.011	0.005	1.7E-02
rs1056667	6	26510564	T	0.023	0.004	1.86E-08	0.017	0.004	1.7E-05	0.005	0.003	1.9E-01
rs9320913	6	98584733	A	0.025	0.004	3.50E-10	0.021	0.004	1.1E-07	0.025	0.003	3.0E-13

Panel B

				New Cohorts (*N*=167,310)			Rietveld Cohorts (*N*=126,413)		
SNP	Chr	Position	A1	Beta	S.E.	*P*-value	Beta	S.E.	*P*-value
rs34305371	1	72733610	A	0.038	0.006	2.4E-10	0.032	0.007	2.4E-05
rs10496091	2	61482261	G	0.024	0.004	3.2E-10	0.010	0.004	2.7E-02
rs13018640	2	100821545	C	0.031	0.004	1.3E-18	0.022	0.004	7.8E-08
rs62262514	3	49125477	C	0.050	0.008	2.1E-09	0.018	0.012	1.2E-01
rs11711536	3	49391240	T	0.028	0.004	3.7E-14	0.018	0.004	1.9E-05
rs11712056	3	49914397	T	0.028	0.003	2.7E-15	0.018	0.004	5.0E-06
rs12493563	3	85682087	T	0.020	0.004	3.1E-08	0.009	0.004	2.9E-02
rs61160187	5	60111579	G	0.020	0.004	1.2E-08	0.012	0.004	2.6E-03
rs9401593	6	98549801	C	0.025	0.003	2.3E-13	0.021	0.004	1.7E-07
rs11794152	9	23345347	G	0.024	0.004	2.1E-11	0.021	0.004	6.8E-07
rs12428841	13	58391491	T	0.024	0.004	7.3E-10	0.024	0.004	5.7E-08
rs10146424	14	23438980	T	0.020	0.004	1.0E-08	0.013	0.004	1.5E-03
rs78889595	16	7913629	G	0.029	0.005	3.1E-08	0.004	0.006	5.4E-01
rs916888	17	44863133	T	0.022	0.004	4.8E-08	0.018	0.005	2.5E-04

Note: Panel A shows the genome-wide significant SNPs identified by Rietveld, Medland, et al. (2013) and the corresponding effect sizes in the Rietveld Cohorts and the New Cohorts. Panel B shows the genome-wide significant SNPs identified in a meta-analysis of the New Cohorts and the corresponding effect sizes in the Rietveld Cohorts. A1 is allele 1.

Table B7. Out-of-sample replication analyses of lead SNPs in UK Biobank

				EduYears (*N*=293,723)			UKB (*N*=111,349)		
SNP	Chr	Position	Allele 1	Beta	S.E.	*P*-value	Beta	S.E.	*P*-value
rs301800	1	8490603	T	0.0191	0.0034	1.8E-08	0.0089	0.0055	1.1E-01
rs11210860	1	43982527	A	0.0171	0.0027	2.4E-10	0.0136	0.0043	1.5E-03
rs34305371	1	72733610	A	0.0355	0.0047	3.8E-14	0.0418	0.0069	1.3E-09
rs2568955	1	72762169	C	0.0173	0.0031	1.8E-08	0.0148	0.0049	2.7E-03
rs1008078	1	91189731	C	0.0165	0.0027	6.0E-10	0.0213	0.0043	6.6E-07
rs11588857	1	204587047	A	0.0198	0.0032	5.3E-10	0.0219	0.0051	1.9E-05
rs1777827	1	211613114	A	0.0150	0.0027	1.5E-08	0.0100	0.0043	1.9E-02
rs2992632	1	243503764	A	0.0168	0.0029	8.2E-09	0.0172	0.0046	2.2E-04
rs76076331	2	10977585	T	0.0205	0.0037	3.6E-08	0.0225	0.0064	4.2E-04
rs11689269	2	15621917	C	0.0158	0.0028	1.3E-08	0.0010	0.0044	8.3E-01
rs1606974	2	51873599	A	0.0222	0.0040	2.8E-08	0.0221	0.0063	4.2E-04
rs11690172	2	57387094	A	0.0149	0.0027	2.0E-08	0.0033	0.0042	4.4E-01
rs2457660	2	60757419	C	0.0168	0.0027	7.1E-10	0.0144	0.0044	1.0E-03
rs114598875	2	60976384	G	0.0198	0.0036	2.4E-08	0.0051	0.0052	3.2E-01
rs10496091	2	61482261	G	0.0178	0.0029	5.6E-10	0.0013	0.0046	7.7E-01
rs13402908	2	100333377	C	0.0177	0.0026	1.7E-11	0.0088	0.0042	3.6E-02
rs4851251	2	100753490	C	0.0166	0.0030	1.9E-08	0.0129	0.0047	6.2E-03
rs12987662	2	100821548	A	0.0273	0.0027	2.7E-24	0.0207	0.0043	1.2E-06
rs17824247	2	144152539	C	0.0158	0.0027	2.8E-09	0.0176	0.0042	3.1E-05
rs16845580	2	161920884	T	0.0161	0.0027	2.7E-09	0.0172	0.0043	6.5E-05
rs4500960	2	162818621	C	0.0164	0.0026	3.8E-10	0.0088	0.0042	3.4E-02
rs6739979	2	193731929	C	0.0147	0.0027	4.7E-08	0.0135	0.0045	2.6E-03
rs2245901	2	194296294	G	0.0157	0.0027	4.5E-09	0.0063	0.0042	1.4E-01
rs55830725	2	237056854	T	0.0217	0.0035	5.4E-10	0.0068	0.0058	2.4E-01
rs35761247	3	48623124	A	0.0345	0.0063	3.8E-08	0.0265	0.0088	2.7E-03

rs62259535	3	48939052	A	0.0481	0.0081	2.6E-09	0.0278	0.0112	1.3E-02
rs148734725	3	49406708	A	0.0249	0.0028	1.4E-18	0.0312	0.0045	4.8E-12
rs11712056	3	49914397	T	0.0236	0.0026	3.3E-19	0.0296	0.0042	1.7E-12
rs112634398	3	50075494	A	0.0360	0.0066	4.6E-08	0.0509	0.0103	8.5E-07
rs62263923	3	85674790	G	0.0158	0.0027	7.0E-09	0.0233	0.0044	9.1E-08
rs6799130	3	160847801	G	0.0145	0.0026	2.8E-08	0.0021	0.0040	5.9E-01
rs12646808	4	3249828	T	0.0155	0.0028	4.0E-08	0.0151	0.0044	6.5E-04
rs2610986	4	18037231	C	0.0162	0.0029	2.0E-08	0.0023	0.0046	6.2E-01
rs34072092	4	28801221	T	0.0243	0.0044	3.9E-08	0.0083	0.0066	2.1E-01
rs3101246	4	42649935	G	0.0152	0.0027	1.4E-08	-0.0025	0.0043	5.5E-01
rs4863692	4	140764124	T	0.0180	0.0028	1.6E-10	0.0140	0.0044	1.7E-03
rs4493682	5	45188024	C	0.0193	0.0035	3.3E-08	0.0197	0.0054	2.8E-04
rs2964197	5	57535206	T	0.0145	0.0026	3.0E-08	0.0078	0.0042	6.2E-02
rs61160187	5	60111579	G	0.0168	0.0027	3.5E-10	0.0205	0.0042	1.2E-06
rs324886	5	87896602	C	0.0151	0.0027	1.9E-08	0.0058	0.0043	1.7E-01
rs10061788	5	87934707	A	0.0212	0.0035	2.5E-09	0.0199	0.0056	3.6E-04
rs2431108	5	103947968	T	0.0163	0.0028	5.3E-09	0.0085	0.0044	5.6E-02
rs1402025	5	113987898	T	0.0173	0.0031	3.4E-08	0.0170	0.0049	6.0E-04
rs62379838	5	120102028	T	0.0156	0.0028	3.3E-08	0.0076	0.0045	9.3E-02
rs56231335	6	98187291	C	0.0171	0.0029	2.1E-09	0.0197	0.0044	9.1E-06
rs9320913	6	98584733	A	0.0236	0.0026	2.5E-19	0.0297	0.0042	1.1E-12
rs7767938	6	153367613	T	0.0168	0.0030	2.4E-08	0.0044	0.0049	3.6E-01
rs2615691	7	23402104	G	0.0373	0.0068	4.7E-08	-0.0062	0.0112	5.8E-01
rs12531458	7	39090698	A	0.0144	0.0026	3.1E-08	0.0070	0.0042	9.2E-02
rs12671937	7	92654365	A	0.0164	0.0027	9.2E-10	0.0084	0.0042	4.4E-02

rs113520408	7	128402782	A	0.0169	0.0030	2.0E-08	0.0101	0.0046	2.9E-02
rs17167170	7	133302345	A	0.0198	0.0033	1.1E-09	0.0206	0.0051	6.4E-05
rs11768238	7	135227513	G	0.0169	0.0028	9.9E-10	0.0063	0.0044	1.6E-01
rs12682297	8	145712860	T	0.0155	0.0026	3.9E-09	0.0158	0.0042	1.5E-04
rs1871109	9	1746016	G	0.0164	0.0026	4.3E-10	0.0161	0.0042	1.2E-04
rs13294439	9	23358875	C	0.0231	0.0027	2.2E-17	0.0316	0.0042	7.1E-14
rs895606	9	88003668	A	0.0147	0.0026	2.2E-08	0.0108	0.0042	9.3E-03
rs7854982	9	124644562	C	0.0149	0.0026	1.3E-08	0.0096	0.0042	2.1E-02
rs11191193	10	103802408	A	0.0180	0.0028	5.4E-11	0.0167	0.0043	1.3E-04
rs12772375	10	104082688	G	0.0151	0.0027	1.6E-08	0.0178	0.0043	3.3E-05
rs7945718	11	12748819	A	0.0153	0.0027	1.5E-08	0.0115	0.0043	7.1E-03
rs7955289	12	14653667	A	0.0171	0.0027	4.5E-10	0.0147	0.0043	5.8E-04
rs2456973	12	56416928	C	0.0198	0.0028	1.1E-12	0.0190	0.0044	1.4E-05
rs7131944	12	92159557	A	0.0154	0.0027	9.0E-09	0.0089	0.0043	3.9E-02
rs572016	12	121279083	A	0.0144	0.0026	3.5E-08	0.0050	0.0041	2.3E-01
rs7306755	12	123767929	A	0.0226	0.0032	1.3E-12	0.0283	0.0052	4.8E-08
rs9537821	13	58402771	A	0.0242	0.0029	1.5E-16	0.0227	0.0046	9.0E-07
rs1043209	14	23373986	A	0.0180	0.0027	1.8E-11	0.0129	0.0043	2.5E-03
rs8008779 (proxy for rs8005528)	14	27089482	C	0.0185	0.0032	1.1E-08	0.0122	0.0051	1.8E-02
rs17119973	14	84913111	G	0.0187	0.0030	3.6E-10	0.0160	0.0048	8.2E-04
rs55943044 (proxy for rs192818565)	17	43872228	G	0.0229	0.0034	1.6E-11	0.0112	0.0049	2.3E-02
rs12969294	18	35186122	G	0.0159	0.0028	7.2E-09	0.0245	0.0044	2.4E-08
rs2837992	21	42620520	T	0.0148	0.0027	3.8E-08	0.0068	0.0043	1.2E-01
rs165633	22	29880773	G	0.0178	0.0030	2.9E-09	0.0018	0.0050	7.2E-01

APPENDIX C

Supplementary Material to Chapter 5

C1. Details on data and measures

C1.1. Administrative Data

Sex — Equal to 1 if male, 2 if female. Information is retrieved from the Swedish Population Register.

Birth year — Information is retrieved from the Swedish Population Register.

Municipality of residence — Code for municipality of residence in 1960. Information is retrieved from the 1960 census.

Birth municipality — Code for municipality of birth according to the codes used in the 1960 census. Information is retrieved from the Multi-Generation Registry.

Years of schooling — Educational attainment according to the three-digit Swedish standard classification of education (SUN 2000). Following the manual for classifying educational programmes in OECD countries (ISCED-97), we assigned the following years of schooling to each category: (old) primary school (7); (new) compulsory school (9); (old) junior secondary education (9); high school (10-12 depending on the program); short university (13); longer university (14-17 depending on the program); short post-graduate (18); long post-graduate (20). The information on educational attainment is retrieved from the Longitudinal integration database for health insurance and labour market studies (LISA by Swedish acronym).

Income in 1970-2005 — The data relates to individuals 16 or over. The variables measure yearly taxable earnings ("sammanräknad förvärvsinkomst") defined as the sum of wage labor income, income from own business, unemployment compensation and pension income from 1970, 1975, 1980, 1985, 1990, 1995, 2000 and 2005 as reported by employers to the tax authorities. Information is retrieved from the Income and Taxation Registry.

Cognitive performance — Scores from a cognitive performance test taken during military conscription. All males in our sample were required by law to participate in military conscription around the age of 18. The test of cognitive performance used by the Swedish military consists of four subtests: logical, verbal, spatial, and technical. The first subtest about logical ability measures the ability to understand complicated instructions. In the second subtest about verbal ability, the subjects have to pick out one out of five words that differs from the four other words. The third test is a test of spatial ability where the subjects are asked to see which pieces fit with a specific figure. In the final test, the subjects answer

questions about technical problems with the guidance of graphs. Information is retrieved from the Military Archives of Sweden.

C1.2. Municipality Data

Social democratic party vote share — Share of total votes going to the Social democratic party. Data are from the Election Data Archive (Valdataarkivet) 1948-1970 obtained from the Swedish National Data Service (SND).

Communist party vote share — Share of total votes going to the Communist party. Data are from the Election Data Archive (Valdataarkivet) 1948-1970 obtained from the Swedish National Data Service (SND).

Agrarian party vote share — Share of total votes going to the Agrarian party. Data are from the Election Data Archive (Valdataarkivet) 1948-1970 obtained from the Swedish National Data Service (SND).

Liberal party vote share — Share of total votes going to the Liberal party. Data are from the Election Data Archive (Valdataarkivet) 1948-1970 obtained from the Swedish National Data Service (SND).

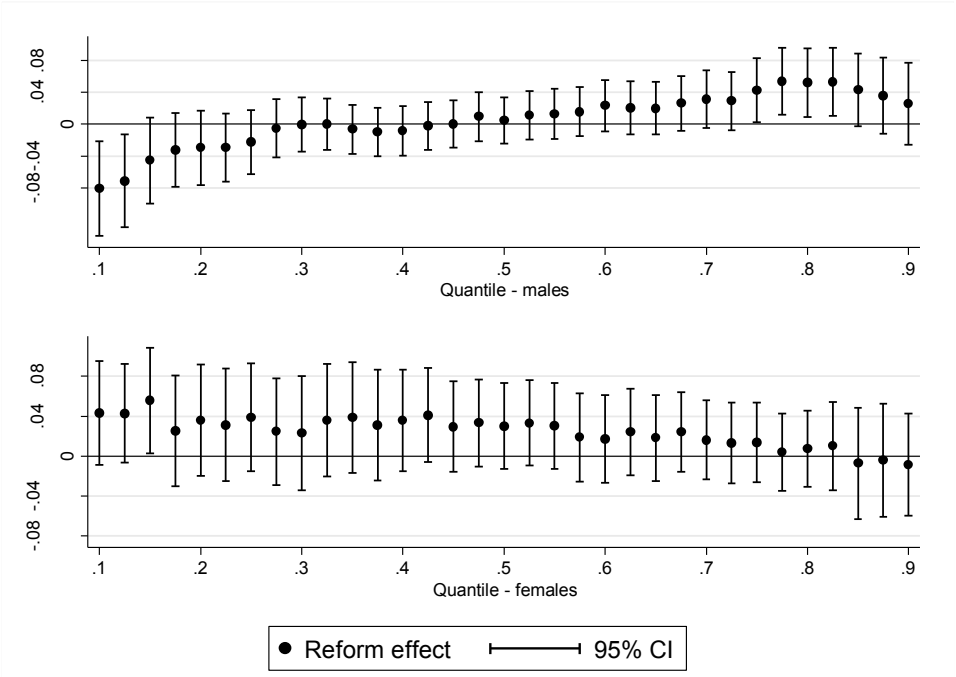
Conservative party vote share — Share of total votes going to the Conservative party. Data are from the Election Data Archive (Valdataarkivet) 1948-1970 obtained from the Swedish National Data Service (SND).

Electorate — Total number of eligible voters. Data are from the Election Data Archive (Valdataarkivet) 1948-1970 obtained from the Swedish National Data Service (SND).

Turnout — Total number of votes divided by the number of eligible voters. Data are from the Election Data Archive (Valdataarkivet) 1948-1970 obtained from the Swedish National Data Service (SND).

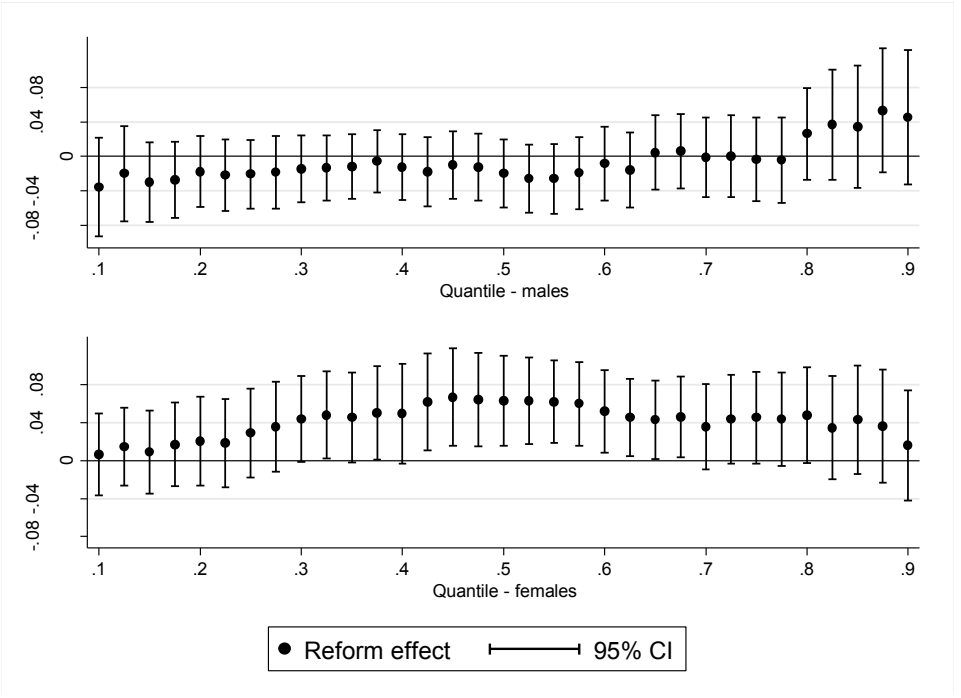
C2. Supplementary Figures

Figure C1.Quantile regressions: Impact of reform on early career income by sex



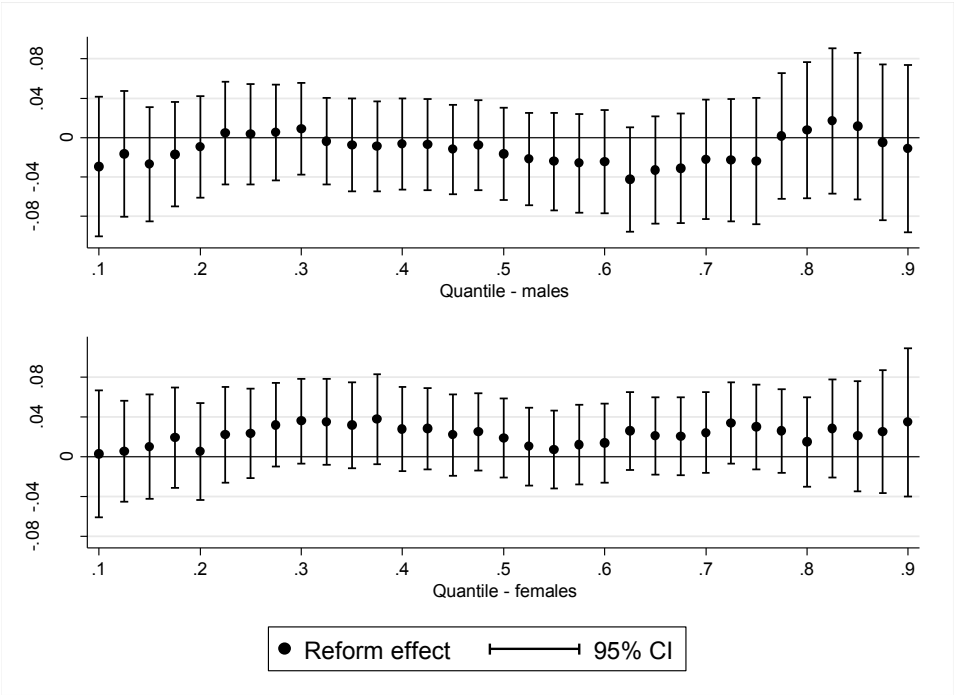
Note: Each point corresponds to the estimate from a quantile regression of the effect of the reform on income at a quantile. All quantile regressions estimated panel models that included all individual-year observations for which an individual was between 23 and 32 years old when his or her income was measured. All quantile regressions include controls for the first 10 principal components, birth year fixed effects, fixed effects for municipality clusters, municipality cluster-specific birth year trends, municipality level covariates, and a third degree polynomial of age. (The standard errors used to calculate the 95% confidence intervals were not clustered.)

Figure C2. Quantile regressions: Impact of reform on mid career income by sex.



Note: Each point corresponds to the estimate from a quantile regression of the effect of the reform on income at a quantile. All quantile regressions estimated panel models that included all individual-year observations for which an individual was between 33 and 42 years old when his or her income was measured. All quantile regressions include controls for the first 10 principal components, birth year fixed effects, fixed effects for municipality clusters, municipality cluster-specific birth year trends, municipality level covariates, and a third degree polynomial of age. (The standard errors used to calculate the 95% confidence intervals were not clustered.)

Figure C3. Quantile regressions: Impact of reform on late career income by sex



Note: Each point corresponds to the estimate from a quantile regression of the effect of the reform on income at a quantile. All quantile regressions estimated panel models that included all individual-year observations for which an individual was between 43 and 52 years old when his or her income was measured. All quantile regressions include controls for the first 10 principal components, birth year fixed effects, fixed effects for municipality clusters, municipality cluster-specific birth year trends, municipality level covariates, and a third degree polynomial of age. (The standard errors used to calculate the 95% confidence intervals were not clustered.)

C3. Supplementary Tables

Table C1. Predictive power of the polygenic score in the analysis sample.

Score	ΔR^2	*P*-value	95% CI - low	95% CI - high
LD-adjusted, q=1	6.15%	0	5.06%	7.38%
LD-adjusted, q=0.3	5.81%	0	4.78%	7.10%
LD-adjusted, q=0.1	4.57%	0	3.63%	5.67%
LD-adjusted, q=0.03	2.67%	0	1.90%	3.48%
LD-adjusted, q=0.01	2.24%	0	1.57%	3.06%
LD-adjusted, q=0.003	1.03%	0	0.55%	1.59%
LD-adjusted, q=0.001	0.89%	0	0.49%	1.39%
LD-adjusted, q=0.0003	0.16%	0.137	0.02%	0.43%
Unadjusted	5.49%	0	4.44%	6.63%

Note: "LD-adjusted" refers to scores constructed with LD-adjusted effect sizes using the LDpred software package (Vilhjálmsson et al., 2015). "q" is the assumed fraction of causal SNPs. "Unadjusted" refers to the score constructed using unadjusted effect sizes of the same set of SNPs included in LD-adjusted scores. ΔR^2 is the incremental R^2 from including the polygenic score in a regression of years of education on sex, age, sex × age, and the first 10 PCs. CI is confidence interval.

Table C2. Robustness checks – fixed effects for municipality in 1960

		Years of schooling		Junior high school/ New compulsory		High school		College		Cognitive Ability
		(males)	(females)	(males)	(females)	(males)	(females)	(males)	(females)	(males)
Additive models	Reform	0.332	0.541**	0.080	0.082	0.014	0.017	0.011	0.038	-0.047
		(0.299)	(0.240)	(0.065)	(0.056)	(0.049)	(0.047)	(0.045)	(0.039)	(0.122)
	PGS	0.672***	0.695***	-0.068***	-0.091***	0.033***	0.041***	0.076***	0.081***	0.253***
		(0.057)	(0.067)	(0.012)	(0.010)	(0.011)	(0.010)	(0.009)	(0.009)	(0.026)
Interaction models	Reform× PGS	0.067	0.220	-0.031	-0.168***	-0.037	0.084**	0.043	0.028	-0.005
		(0.303)	(0.239)	(0.060)	(0.046)	(0.058)	(0.042)	(0.042)	(0.040)	(0.130)
	Sample size	2,725	3,091	2,725	3,091	2,725	3,091	2,725	3,091	2,605

Note: All regressions include controls for the first 10 principal components, birth year fixed effects, fixed effects for municipalities in 1960 (instead of for municipality clusters), municipality cluster-specific birth year trends, and municipality level covariates. Standard errors, shown in parentheses, allow for clustering at the municipality level.

***/**/* indicate significance at the 1/5/10% level.

Table C3. Robustness checks – born 1945-1955 (and first cohort affected 1947-1955)

		Years of schooling		Junior high school/ New compulsory		High school		College		Cognitive Ability
		(males)	(females)	(males)	(females)	(males)	(females)	(males)	(females)	(males)
Additive models	Reform	0.416*	0.267	0.106**	0.067	-0.006	-0.005	0.022	0.023	-0.082
		(0.250)	(0.190)	(0.050)	(0.046)	(0.037)	(0.038)	(0.035)	(0.033)	(0.099)
	PGS	0.674***	0.656***	-0.076***	-0.080***	0.037***	0.034***	0.079***	0.081***	0.246***
		(0.055)	(0.050)	(0.011)	(0.010)	(0.009)	(0.009)	(0.008)	(0.009)	(0.023)
Interaction models	Reform× PGS	0.119	0.398**	-0.018	-0.145***	-0.033	0.069	0.037	0.049	0.062
		(0.260)	(0.200)	(0.048)	(0.043)	(0.042)	(0.042)	(0.036)	(0.034)	(0.107)
	Sample size	2,077	2,356	2,077	2,356	2,077	2,356	2,077	2,356	1,971

Note: All regressions include controls for the first 10 principal components, birth year fixed effects, fixed effects for municipality clusters, municipality cluster-specific birth year trends, and municipality level covariates. Standard errors, shown in parentheses, allow for clustering at the municipality level.
***/**/* indicate significance at the 1/5/10% level.

Table C4. Robustness checks – born 1938-1958 (and first cohort affected 1940-1958)

		Years of schooling		Junior high school/ New compulsory		High school		College		Cognitive Ability
		(males)	(females)	(males)	(females)	(males)	(females)	(males)	(females)	(males)
Additive models	Reform	0.690***	0.394**	0.115***	0.074**	0.026	0.026	0.028	0.010	0.078
		(0.185)	(0.167)	(0.035)	(0.032)	(0.028)	(0.027)	(0.025)	(0.025)	(0.074)
	PGS	0.678***	0.662***	-0.065***	-0.077***	0.039***	0.036***	0.079***	0.075***	0.255***
		(0.036)	(0.041)	(0.007)	(0.007)	(0.006)	(0.006)	(0.008)	(0.006)	(0.015)
Interaction models	Reform× PGS	-0.021	0.158	-0.032	-0.136***	-0.030	0.081***	0.030	0.006	0.029
		(0.177)	(0.145)	(0.033)	(0.030)	(0.030)	(0.028)	(0.026)	(0.025)	(0.069)
	Sample size	4,324	4,944	4,324	4,944	4,324	4,944	4,324	4,944	4,157

Note: All regressions include controls for the first 10 principal components, birth year fixed effects, fixed effects for municipality clusters, municipality cluster-specific birth year trends, and municipality level covariates. Standard errors, shown in parentheses, allow for clustering at the municipality level.
***/**/* indicate significance at the 1/5/10% level.

Table C5. Robustness checks – birth municipality (born 1938-1955 and first cohort affected 1940-1955)

		Years of schooling		Junior high school/ New compulsory		High school		College		Cognitive Ability
		(males)	(females)	(males)	(females)	(males)	(females)	(males)	(females)	(males)
Additive models	Reform	0.535***	0.669***	0.135***	0.018	-0.016	0.021	0.035	0.065***	0.080
		(0.184)	(0.145)	(0.034)	(0.034)	(0.025)	(0.024)	(0.025)	(0.024)	(0.078)
	PGS	0.678***	0.661***	-0.068***	-0.080***	0.039***	0.036***	0.075***	0.077***	0.251***
		(0.037)	(0.038)	(0.008)	(0.007)	(0.007)	(0.005)	(0.007)	(0.006)	(0.014)
Interaction models	Reform×	-0.178	0.158	0.026	-0.105***	-0.043	0.038	-0.006	0.031	-0.022
	PGS	(0.164)	(0.146)	(0.030)	(0.028)	(0.027)	(0.024)	(0.023)	(0.024)	(0.065)
	Sample size	4,293	4,957	4,293	4,957	4,293	4,957	4,293	4,957	4,123

Note: All regressions include controls for the first 10 principal components, birth year fixed effects, fixed effects for municipality clusters, municipality cluster-specific birth year trends, and municipality level covariates. Standard errors, shown in parentheses, allow for clustering at the municipality level.
***/**/* indicate significance at the 1/5/10% level.

Table C6. Robustness checks – with the PLINK score instead of the LDpred score

		Years of schooling		Junior high school/ New compulsory		High school		College		Cognitive Ability
		(males)	(females)	(males)	(females)	(males)	(females)	(males)	(females)	(males)
Additive models	Reform	0.535***	0.669***	0.135***	0.018	-0.016	0.021	0.035	0.065***	0.080
		(0.184)	(0.145)	(0.034)	(0.034)	(0.025)	(0.024)	(0.025)	(0.024)	(0.078)
	PGS	0.678***	0.661***	-0.068***	-0.080***	0.039***	0.036***	0.075***	0.077***	0.251***
		(0.037)	(0.038)	(0.008)	(0.007)	(0.007)	(0.005)	(0.007)	(0.006)	(0.014)
Interaction models	Reform×PGS	-0.178	0.158	0.026	-0.105***	-0.043	0.038	-0.006	0.031	-0.022
		(0.164)	(0.146)	(0.030)	(0.028)	(0.027)	(0.024)	(0.023)	(0.024)	(0.065)
	Sample size	4,293	4,957	4,293	4,957	4,293	4,957	4,293	4,957	4,123

Note: All regressions include controls for the first 10 principal components, birth year fixed effects, fixed effects for municipality clusters, municipality cluster-specific birth year trends, and municipality level covariates. Standard errors, shown in parentheses, allow for clustering at the municipality level.
***/**/* indicate significance at the 1/5/10% level.

Table C7. Robustness checks – income models

		Municipality fixed effects		Born 1945-1955		Born 1938-1958		Birth municipality		PLINK score	
		(males)	(females)	(males)	(females)	(males)	(females)	(males)	(females)	(males)	(females)
Additive models	Reform	-0.031	0.041**	-0.014	0.032	0.007	0.032**	0.015	0.026*	-0.007	0.031*
		(0.025)	(0.020)	(0.023)	(0.019)	(0.015)	(0.015)	(0.019)	(0.016)	(0.021)	(0.018)
	PGS	0.045***	0.043***	0.045***	0.042***	0.045***	0.041***	0.045***	0.046***	0.040***	0.043***
		(0.006)	(0.005)	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)
Interaction models	Reform×PGS	0.015	0.051***	0.023	0.059***	0.014	0.031**	-0.016	0.035*	0.012	0.058***
		(0.024)	(0.018)	(0.021)	(0.018)	(0.016)	(0.013)	(0.017)	(0.019)	(0.019)	(0.016)
	Observations	15,882	14,512	12,206	11,495	24,258	22,128	20,575	18,352	15,582	14,512

Note: All regressions estimated panel models that included all individual-year observations for which an individual was between 25 and 55 years old. All regressions include controls for the first 10 principal components, birth year fixed effects, fixed effects for municipality clusters (except for the models in columns 1 and 2), municipality cluster-specific birth year trends, municipality level covariates, and a third degree polynomial of age. The models in columns 1 and 2 include municipality fixed effects instead of fixed effects for municipality clusters. In models 3 and 4, the sample is restricted to individuals born between 1945 and 1955. In models 5 and 6, all individuals born between 1938 and 1958 are included in the regression. In models 7 and 8, birth municipality is used to code the reform indicator (instead of municipality according to the 1960 census). In models 9 and 10 the PLINK score is used instead of the LDpred score. Standard errors, shown in parentheses, allow for clustering at the municipality level.

***/**/* indicate significance at the 1/5/10% level.

Summary

This thesis explores questions at the intersection of economics and biology, and thus contributes to an emerging field of research so young that researchers have yet to agree on its name. Without attempting an exhaustive laundry list of names, some commonly used ones are: genoeconomics, social-science genetics, biosocial science and biological economics.

Chapter 1 is an introductory chapter that seeks to lay out the terrain by motivating and summarizing research on the molecular genetic architecture of complex social-scientific outcomes.

Each of the five remaining essays contributes in a small way toward the integration of the social and biological sciences.

Chapter 2 explains the limitations of traditional gene-discovery approaches, which include low power and undisclosed testing of multiple hypotheses. The chapter provides some historical context to, and motivation for, the genome-wide association studies reported in Chapters 3 (subjective well-being, neuroticism and depressive symptoms) and 4 (educational attainment), which report positive findings that replicate reliably and implicate interesting biology. Chapter 5 shows how robust findings from genome-wide association studies can be used, in conjunction with quasi-experimental research designs from economics, to conduct rigorous and well-powered investigations of the interactions between genes and environment. Finally, Chapter 6 provides a framework for quantifying tradeoffs between heterogeneity in sample-size in gene-discovery efforts.

Though considerable uncertainty remains about the ultimate value of genetic data in the social sciences, the results reported here strongly suggest that there will be at least some settings in which genetic data will prove valuable. As more genetic data become available, the number of identified associations in genome-wide association studies of complex behavioral is likely to continue to grow. In addition to broadening our knowledge about the biological mehanisms underlying behavioral outcomes, these advances will provide researchers from many disciplines with increasingly powerful tools that can be used to understand, for example, the interplay between genes and environment.

Samenvatting (Summary in Dutch)

Dit proefschrift omvat onderzoek dat grenst aan zowel de economische als biologische wetenschappen, en draagt daardoor bij aan een relatief nieuw onderzoeksveld. Dit onderzoeksveld is bekend onder verschillende namen zoals: ‘geneconomics’, ‘social-science genetics’, ‘biosocial science’ en ‘biological economics’ .

Hoofdstuk 1 geeft de motivatie voor dit proefschrift weer. Daarnaast geeft het een korte samenvatting van de verschillende onderzoeksvragen en belangrijkste bevindingen.

De overige vijf hoofdstukken leveren elk een kleine bijdrage aan de integratie van de sociale en biologische wetenschappen.

Hoofdstuk 2 geeft een overzicht van de limitaties van conventionele gen-ontdekkings methoden. Enkele voorbeelden hiervan zijn: het ontbreken van power om genetische varianten te vinden en het niet vermelden van het testen van meerdere hypothesen. Daarnaast geeft het hoofdstuk ook de geschiedenis en motivatie weer om genoomwijde-associatie studies (GWAS) uit te voeren, die besproken worden in hoofdstuk 3 (subjectief welbevinden, depressie en neuroticisme) en hoofdstuk 4 (opleidingsniveau). Deze twee hoofdstukken laten op betrouwbare en replicieerbare wijze zien, dat specifieke genetische varianten betrokken zijn bij complexe eigenschappen en teven interessante implicaties hebben vanuit biologisch perspectief. Hoofdstuk 5 laat zien hoe robuuste GWAS bevindingen samen met quasi-experimenteel economisch onderzoek gebruikt kunnen worden om op een betrouwbare wijze gen-omgeving interacties aan te tonen. Tenslotte, biedt hoofdstuk 6 een framewerk voor het kwantificeren van afwegingen in de keuze tussen heterogeniteit van je sample en samplegrootte om genetische varianten te ontdekken met voldoende power.

Hoewel onzekerheid over de uiteindelijke waarde van genetische data in de sociale wetenschapen blijft bestaan, laten de resultaten, zoals beschreven in de verschillende hoofdstukken, zien dat er een aantal scenario’s bestaan waarin genetische informatie waardevol kan zijn. Naarmate er meer genetische data beschikbaar komt, zal het aantal associaties met complex gedrag groeien, met al gevolg dat onze kennis over de onderliggende biologische mechanismes zal toenemen. Daarnaast zal deze vooruitgang, wetenschappers uit verschillende disciplines beter in staat stellen om bijvoorbeeld de complexe interactie tussen genetische en omgevingsfactoren te bestuderen.

References

Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., … McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 56–65. https://doi.org/10.1038/nature11632

Acemoglu, D., & Angrist, J. (2001). How large are human-capital externalities? Evidence from compulsory-schooling laws. In *NBER Macroeconomics Annual 2000, Volume 15* (pp. 9–74). MIT Press. Retrieved from http://www.nber.org/chapters/c11054

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School - Age Forms & Profiles*. Burlington: University of Vermont.

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, *80*(1), 123–129. https://doi.org/10.1016/S0165-1765(03)00032-6

Alia-Klein, N., Goldstein, R. Z., Tomasi, D., Woicik, P. a, Moeller, S. J., Williams, B., … Volkow, N. D. (2009). Neural mechanisms of anger regulation as a function of genetic risk for violence. *Emotion (Washington, D.C.)*, *9*(3), 385–96. https://doi.org/10.1037/a0015904

Allen, N. E., Sudlow, C., Peakman, T., & Collins, R. (2014). UK biobank data: come and get it. *Science Translational Medicine*, *6*(224), 224ed4. https://doi.org/10.1126/scitranslmed.3008601

Allen Institute for Brain Science. (2015). BrainSpan atlas of the developing human brain.

Angrist, J. D., & Keueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, *106*(4), 979–1014. https://doi.org/10.2307/2937954

Ardlie, K. G., Deluca, D. S., Segre, A. V., Sullivan, T. J., Young, T. R., Gelfand, E. T., … Lockhart, N. C. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, *348*(6235), 648–660. https://doi.org/10.1126/science.1262110

Ashenfelter, O., & Krueger, A. (1994). Estimating the returns to schooling using a new sample of twins. *The American Economic Review*, *84*(5), 1157–1173.

Ashenfelter, O., & Rouse, C. (1998). Income, schooling, and ability: Evidence from a new sample of identical twins. *The Quarterly Journal of Economics*, *113*(1), 253–284.

Bartels, M. (2015). Genetics of wellbeing and its components satisfaction with life, happiness, and quality of life: a review and meta-analysis of heritability studies. *Behavior Genetics*, *45*(2), 137–156. https://doi.org/10.1007/s10519-015-9713-y

Bartels, M., & Baselmans, B. M. L. (2015). Molecular genetics of psychological well-being. In M. Pluess (Ed.), *Genetics of Psychological Well-Being: The role of heritability and genetics in Positive Psychology* (1st ed., pp. 163–176). Oxford

University Press.

Bartels, M., & Boomsma, D. I. (2009). Born to be happy? The etiology of subjective well-being. *Behavior Genetics*, *39*(6), 605–615. https://doi.org/10.1007/s10519-009-9294-8

Bartels, M., Cacioppo, J. T., van Beijsterveldt, T. C. E. M., & Boomsma, D. I. (2013). Exploring the association between well-being and psychopathology in adolescents. *Behavior Genetics*, *43*(3), 177–190. https://doi.org/10.1007/s10519-013-9589-7

Bartels, M., Saviouk, V., de Moor, M. H. M., Willemsen, G., Beijsterveldt, T. C. E. M. van, & Hottenga, J.-J. (2010). Heritability and genome-wide linkage scan of subjective happiness. *Twin Research and Human Genetics*, *13*, 135–142.

Beauchamp, J., Cesarini, D., Johannesson, M., van der Loos, M. J. H. M., Koellinger, P. D., Groenen, P. J. F., … Christakis, N. A. (2011). Molecular genetics and economics. *Journal of Economic Perspectives*, *25*(4), 57–82.

Beevers, C. G., Wells, T. T., & McGeary, J. E. (2009). The BDNF Val66Met polymorphism is associated with rumination in healthy adults. *Emotion (Washington, D.C.)*, *9*(4), 579–84. https://doi.org/10.1037/a0016189

Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B., Harrington, H., Hogan, S., … Caspi, A. (2016). The Genetics of Success: How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development. *Psychological Science*, 956797616643070. https://doi.org/10.1177/0956797616643070

Benjamin, D. J., Cesarini, D., Chabris, C. F., Glaeser, E. L., Laibson, D. I., Gudnason, V., … Lichtenstein, P. (2012). The promises and pitfalls of genoeconomics. *Annual Review of Economics*, *4*(1), 627–662. https://doi.org/10.1146/annurev-economics-080511-110939

Benjamin, D. J., Cesarini, D., van der Loos, M. J. H. M., Dawes, C. T., Koellinger, P. D., Magnusson, P. K. E., … Visscher, P. M. (2012). The genetic architecture of economic and political preferences. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(21), 8026–31. https://doi.org/10.1073/pnas.1120666109

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., & Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behavioural Brain Research*, *125*(1–2), 279–84. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11682119

Bradburn, N. M. (1969). *The Structure of Psychological Well-Being*. *Chicago: Aldine.* https://doi.org/10.1037/0022-3514.69.4.719

Branigan, A. R., McCallum, K. J., & Freese, J. (2013). Variation in the heritability of educational attainment: An international meta-analysis. *Northwestern University Institute for Policy Research Working Paper*, *92*, 109–140.

Brim, O. G., Baltes, P. B., Bumpass, L. L., Cleary, P. D., Featherman, D. L., Hazzard, W.

R., … Shweder, R. A. (2003). National survey of midlife development in the United States (MIDUS), 1995-1996. https://doi.org/10.3886/ICPSR02760

Bulik-Sullivan, B. K., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., … Perry. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. https://doi.org/10.1038/ng.3406

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, … Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. https://doi.org/doi:10.1038/ng.3211

Busseri, M. A. (2015). Toward a resolution of the tripartite structure of subjective well-being. *Journal of Personality*, *83*, 413–428. https://doi.org/10.1111/jopy.12116

Cai, N., Bigdeli, T. B., Kretzschmar, W., Li, Y., Liang, J., Song, L., … Flint, J. (2015). Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*, *523*(7562), 588–591. https://doi.org/10.1038/nature14659

Caplan, B. (2003). Stigler-Becker versus Myers-Briggs: Why preference-based explanations are scientifically meaningful and emperically important. *Journal of Economic Behavior and Organization*, *50*(4), 391–405.

Cardon, L. R., & Palmer, L. J. (2003). Population stratification and spurious allelic association. *The Lancet*, *361*(9357), 598–604. https://doi.org/10.1016/S0140-6736(03)12520-2

Carlson, J. M., Mujica-Parodi, L. R., Harmon-Jones, E., & Hajcak, G. (2012). The orienting of spatial attention to backward masked fearful faces is associated with variation in the serotonin transporter gene. *Emotion*, *12*(2), 203–207. https://doi.org/10.1037/a0025170

Carlstedt, B. (2000). *Cognitive abilities - aspects of structure, process and measurement*. University of Gothenburg.

Carstensen, L. L., Pasupathi, M., Mayr, U., & Nesselroade, J. R. (2000). Emotional experience in everyday life across the adult life span. *Journal of Personality and Social Psychology*, *79*(4), 644–655. https://doi.org/10.1037//0022-3514.79.4.644

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science*, *24*(4), 304–312. https://doi.org/10.1177/0963721415580430

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Chen, H., Pine, D. S., Ernst, M., Gorodetsky, E., Kasen, S., Gordon, K., … Cohen, P. (2012). The MAOA gene predicts happiness in women. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *40*, 122–125.

Chen, W.-M., & Abecasis, G. R. (2007). Family-based association tests for genomewide association scans. *American Journal of Human Genetics*, *81*(5), 913–926.

Costa, P. T., & Widiger, T. A. (1992). The 5-factor model of personality and its relevance to personality-disorders. *Journal of Personality Disorders*, *6*(343). https://doi.org/10.1037/10423-001

Daetwyler, H. D., Villanueva, B., & Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLOS ONE*, *3*, e3395.

Dawes, C. T., Lindgren, K.-O., Oskarsson, S., Okbay, A., Cesarini, D., Johannesson, M., & Magnusson, P. K. E. (n.d.). *The relationship between genes, education, and voting*.

dbGaP. (2013). Phenotype submission based on ICD-9-CM as of March 15, 2013.

dbGaP. (2015). Resource for genetic epidemiology research on adult health and aging (GERA).

De Moor, M. H. M., van den Berg, S. M., Verweij, K. J. H., Krueger, R. F., Luciano, M., Arias Vasquez, A., … Boomsma, D. I. (2015). Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with major depressive disorder. *JAMA Psychiatry*, *72*(7), 642–650. https://doi.org/10.1001/jamapsychiatry.2015.0554

De Neve, J.-E., Fowler, J. H., Frey, B. S., & Christakis, N. A. (2012). Genes, economics, and happiness. *Journal of Neuroscience, Psychology, and Economics*, *5*, 193–211.

De Vlaming, R., Okbay, A., Rietveld, C. A., Johannesson, M., Magnusson, P. K. E., Uitterlinden, A. G., … Koellinger, P. D. (2016). Meta-GWAS Accuracy and Power (MetaGAP) calculator shows that hiding heritability is partially due to imperfect genetic correlations across studies. *PLoS Genet.*, *Forthcomin*. https://doi.org/10.1101/048322

Devlin, B., & Roeder, K. (1999). Genomic control for association studies. *Biometrics*, *55*(4), 997–1004.

Dick, D. M., Agrawal, A., Keller, M. C., Adkins, A., Aliev, F., Monroe, S., … Sher, K. J. (2015). Candidate gene-environment interaction research: Reflections and recommendations. *Perspectives on Psychological Science*, *10*(1), 37–59. https://doi.org/10.1177/1745691614556682

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*, 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.-W., & Oishi, S. (2010). New measures of well-being: flourishing and positive and negative feelings. *Social Indicators Research*, *97*, 143–156.

Domingue, B. W., Liu, H., Okbay, A., & Belsky, D. W. (2016). *Genetic heterogeneity in depressive symptoms following the death of a spouse: Polygenic score analysis of the*

US Health and Retirement Study. *bioRxiv*. https://doi.org/10.1101/065847

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLOS Genetics*, *9*, e1003348.

Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *The American Journal of Psychiatry*, *168*(10), 1041–9. https://doi.org/10.1176/appi.ajp.2011.11020191

Ebstein, R. P., Israel, S., Chew, S. H., Zhong, S., & Knafo, A. (2010). Genetics of human social behavior. *Neuron*, *65*(6), 831–44. https://doi.org/10.1016/j.neuron.2010.02.020

Eeles, R. A., Kote-Jarai, Z., Al Olama, A. A., Giles, G. G., Guy, M., Severi, G., … others. (2009). Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nature Genetics*, *41*, 1116–1121.

Ehret, G. B., Munroe, P. B., Rice, K. M., Bochud, M., Johnson, A. D., Chasman, D. I., … others. (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*, *478*, 103–109.

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*, 446–450.

Elinder, M. (2010). Local economies and general elections: The influence of municipal and regional economic conditions on voting in Sweden 1985–2002. *European Journal of Political Economy*, *26*(2), 279–292. https://doi.org/10.1016/j.ejpoleco.2010.01.003

Eriksson, N., Macpherson, J. M., Tung, J. Y., Hon, L. S., Naughton, B., Saxonov, S., … Mountain, J. (2010). Web-based, participant-driven studies yield novel genetic associations for common traits. *PLOS Genetics*, *6*(6), e1000993–e1000993. https://doi.org/10.1371/journal.pgen.1000993

Evans, D. M., & Purcell, S. M. (2012). Power calculations in genetic studies. *Cold Spring Harb. Protoc.*, *2012*, pdb.top069559.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. London: Hodder and Stroughton.

Fazio, A. (1977). *A Concurrent Validational Study of the NCHS General Well-Being Schedule.* Hyattsville, MD: National Center for Health Statistics.

Finucane, H. K., Bulik-Sullivan, B. K., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., … Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, *47*(11), 1228–1235. https://doi.org/10.1038/ng.3404

Ford, B. Q., Mauss, I. B., Troy, A. S., Smolen, A., & Hankin, B. (2014). Emotion regulation moderates the risk associated with the 5-HTT gene and stress in children. *Emotion (Washington, D.C.)*, *14*(5), 930–9. https://doi.org/10.1037/a0036835

Fowler, J. H., & Schreiber, D. (2008). Biology, politics, and the emerging science of human nature. *Science (New York, N.Y.)*, *322*(5903), 912–4. https://doi.org/10.1126/science.1158188

Fowler, T., Zammit, S., Owen, M. J., & Rasmussen, F. (2012). A population-based study of shared genetic variation between premorbid IQ and psychosis among male twin pairs and sibling pairs from Sweden. *Archives of General Psychiatry*, *69*(5), 460–466.

Fredriksson, P., & Öckert, B. (2014). Life-cycle effects of age at school start. *The Economic Journal*, *124*(579), 977–1004. https://doi.org/10.1111/ecoj.12047

Gandin, I., Joshi, P. K., Esko, T., Wilson, J. F., & The Ygen Consortium. (n.d.). *Ygen: the first systematic assessment of the influence of human Y chromosome variation on a wide range of health-related traits*.

Glassman, A. J., Helzer, J. E., Covey, L. S., Cottler, L. B., Sterner, F., Tipp, J. E., & Johnson, J. (1990). Smoking, smoking cessation, and major depression. *Journal of the American Medical Association*, *264*, 1546–1549.

Gudbjartsson, D. F., Walters, G. B., Thorleifsson, G., Stefansson, H., Halldorsson, B. V, Zusmanovich, P., … Stefansson, K. (2008). Many sequence variants affecting diversity of adult human height. *Nature Genetics*, *40*(5), 609–15. https://doi.org/10.1038/ng.122

Gyurak, A., Haase, C. M., Sze, J., Goodkind, M. S., Coppola, G., Lane, J., … Levenson, R. W. (2013). The effect of the serotonin transporter polymorphism (5-HTTLPR) on empathic and self-conscious emotional reactivity. *Emotion (Washington, D.C.)*, *13*(1), 25–35. https://doi.org/10.1037/a0029616

Haase, C. M., Saslow, L. R., Bloch, L., Saturn, S. R., Casey, J. J., Seider, B. H., … Levenson, R. W. (2013). The 5-HTTLPR polymorphism in the serotonin transporter gene moderates the association between emotional behavior and changes in marital satisfaction over time. *Emotion (Washington, D.C.)*, *13*(6), 1068–79. https://doi.org/10.1037/a0033761

Hamer, D., & Sirota, L. (2000). Beware the chopsticks gene. *Molecular Psychiatry*, *5*(1), 11–13. https://doi.org/10.1038/sj.mp.4000662

Harris, S. E., Hagenaars, S. P., Davies, G., Hill, W. D., Liewald, D. C. M., Ritchie, S. J., … others. (2015). Molecular genetic contributions to self-rated health. *bioRxiv*, 29504.

Hayden, E. P., Klein, D. N., Sheikh, H. I., Olino, T. M., Dougherty, L. R., Dyson, M. W., … Singh, S. M. (2010). The serotonin transporter promoter polymorphism and childhood positive and negative emotionality. *Emotion*, *10*(5), 696–702. https://doi.org/10.1037/a0019527.The

Heath, A. C., Berg, K., Eaves, L. J., Solaas, M. H., Corey, L. A., Sundet, J., … Nance, W. E. (1985). Education policy and the heritability of educational attainment. *Nature*, *314*(6013), 734–736.

Hewitt, J. K. (2012). Editorial policy on candidate gene association and candidate gene-by-

environment interaction studies of complex traits. *Behavior Genetics*, *42*(1), 1–2. https://doi.org/10.1007/s10519-011-9504-z

Hill, D. W., Hagenaars, S. P., Marioni, R. E., Harris, S. E., Liewald, D. C. M., Davies, G., … Deary, I. J. (2016). Molecular genetic contributions to social deprivation and household income in UK Biobank. *Current Biology*, *26*(22), 3083–3089. https://doi.org/10.1016/j.cub.2016.09.035

Hjalmarsson, R., Holmlund, H., & Lindquist, M. J. (2015). The effect of education on criminal convictions and incarceration: Causal evidence from micro-data. *The Economic Journal*, *125*(587), 1290–1326. https://doi.org/10.1111/ecoj.12204

Höglinger, G. U., Melhem, N. M., Dickson, D. W., Sleiman, P. M. A., Wang, L.-S., Klei, L., … Schellenberg, G. D. (2011). Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nature Genetics*, *43*(7), 699–705.

Holmlund, H. (2007). A Researchers Guide to the Swedish Compulsory School Reform. *CEE Discussion Papers*.

Hyde, C. L., Nagle, M. W., Tian, C., Chen, X., Paciga, S. A., & Al., E. (2016). Common genetic variants associated with major depressive disorder among individuals of European descent. *Nature Genetics*.

Hyman, S. (2014). Mental health: Depression needs large human-genetics studies. *Nature*, *515*(7526), 189–191.

International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). (n.d.).

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Johnston, C., Lahey, B. B., & Matthys, W. (2013). Editorial policy for candidate gene studies. *Journal of Abnormal Child Psychology*, *41*(4), 511–514. https://doi.org/10.1007/s10802-013-9741-0

Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(38), 16489–16493.

Kahneman, D., & Riis, J. (2005). Living and thinking about it: Two Perspectives on life. In F. Uppter, N. Baylis, & B. Keverne (Eds.), *The Science of Well-Being* (pp. 285–301). Oxford: Oxford University Press.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-Y., Freimer, N. B., … Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, *42*(4), 348–54. https://doi.org/10.1038/ng.548

Kathiresan, S., Willer, C. J., Peloso, G. M., Demissie, S., Musunuru, K., Schadt, E. E., … Cupples, L. A. (2009). Common variants at 30 loci contribute to polygenic

dyslipidemia. *Nature Genetics*, *41*(1), 56–65.

Keller, M. C. (2014). Gene × environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biological Psychiatry*, *75*(1), 18–24. https://doi.org/10.1016/j.biopsych.2013.09.006

Kendler, K. S., & Myers, J. (2009). The genetic and environmental relationship between major depression and the five-factor model of personality. *Psychological Medicine*, *40*(5), 801.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., … Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Archives of General Psychiatry*. https://doi.org/10.1001/archpsyc.1994.03950010008002

Kong, A., Frigge, M. L., Thorleifsson, G., Stefánsson, H., Young, A. I., Zink, F., … Stefansson, K. (2016). *Selection against the genetic basis of educational attainment*. *PNAS* (Vol. Forthcomin).

Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J.-B., … Plomin, R. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(42), 15273–15278.

Lachman, M. E., & Weaver, S. L. (1997). The Midlife Development Inventory (MIDI) personality scales: Scale construction and scoring. *Waltham, MA: Brandeis University*, (July), 1–9.

Lang, K., & Kropp, D. (1986). Human capital versus sorting: the effects of compulsory attendance laws. *The Quarterly Journal of Economics*, *101*(3), 609–624. https://doi.org/10.1080/02724980343000242

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., … Hirschhorn, J. N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, *467*(7317), 832–838. https://doi.org/10.1038/nature09410

Lee, J., & McGue, M. (2106). Why behavioral genetics matters: Comment on Plomin et al. (2016). *Perspectives on Psychological Science*, *11*(1), 29–30.

Lee, S. H., DeCandia, T. R., Ripke, S., Yang, J., Sullivan, P. F., Goddard, M. E., … others. (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nature Genetics*, *44*, 247–250.

Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V, Purcell, S. M., Perlis, R. H., … others. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics*, *45*, 984–994.

Lee, S. H., Wray, N. R., Goddard, M. E., & Visscher, P. M. (2011). Estimating missing heritability for disease from genome-wide association studies. *American Journal of*

Human Genetics, *88*, 294–305.

Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*, *28*, 2540–2542.

Lee, S., Teslovich, T. M., Boehnke, M., & Lin, X. (2013). General framework for meta-analysis of rare variants in sequencing association studies. *American Journal of Human Genetics*, *93*, 42–53.

Li, M.-X., Yeung, J. M. Y., Cherny, S. S., & Sham, P. C. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human Genetics*, *131*, 747–756.

Li, Y., Willer, C. J., Ding, J., Scheet, P., & Abecasis, G. R. (2010). MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology*, *34*(8), 816–834.

Lichtenstein, P., Sullivan, P. F., Cnattingius, S., Gatz, M., Johansson, S., Carlström, E., … Pedersen, N. L. (2006). The Swedish Twin Registry in the third millennium: an update. *Twin Research and Human Genetics : The Official Journal of the International Society for Twin Studies*, *9*(6), 875–82. https://doi.org/10.1375/183242706779462444

Lindgren, K.-O., Oskarsson, S., & Dawes, C. T. (n.d.). Can political inequalities be educated away? Evidence from a large-scale reform. *American Journal of Political Science*.

Lindqvist, E., & Vestman, R. (2011). The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment. *American Economic Journal: Applied Economics*, *3*(1), 101–128. https://doi.org/10.1257/app.3.1.101

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., … Speliotes, E. K. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, *518*(7538), 197–206. https://doi.org/10.1038/nature14177

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: preliminary reliability and construct validation. *Social Indicators Research*, *46*, 137–155. https://doi.org/10.1023/A:1006824100041

Macmillan, A. M. (1957). The health opinion survey: technique for estimating prevalence of psychoneurotic and related types of disorders in communities. *Psychological Reports*, *3*, 325–339.

Maher, B. (2008). The case of the missing heritability. *Nature*, *456*, 18–21.

Mailman, M. D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., … Sherry, S. T. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nature Genetics*, *39*(10), 1181–6. https://doi.org/10.1038/ng1007-1181

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., … others. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*, 747–753.

Marchini, J., Howie, B., Myers, S., McVean, G., & Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics*, *39*(7), 906–13. https://doi.org/10.1038/ng2088

Marioni, R. E., Ritchie, S. J., Joshi, P. K., Hagenaars, S. P., Okbay, A., Fischer, K., … Deary, I. J. (2016). *Genetic variants linked to education predict longevity*. *PNAS* (Vol. 113). https://doi.org/10.1073/pnas.1605334113

Markett, S., Montag, C., & Reuter, M. (2011). The nicotinic acetylcholine receptor gene CHRNA4 is associated with negative emotionality. *Emotion (Washington, D.C.)*, *11*(2), 450–455. https://doi.org/10.1037/a0021784

Marklund, S. (1981). *Skolsverige 1950-1975: Del 2 Försöksverksamheten*. Stockholm: Liber.

McClellan, J., & King, M.-C. (2010). Genetic heterogeneity in human disease. *Cell*, *141*, 210–217.

Meghir, C., & Palme, M. (2005). Educational reform, ability, and family background. *American Economic Review*, *95*(1), 414–424. https://doi.org/10.1257/0002828053828671

Menashe, I., Rosenberg, P. S., & Chen, B. E. (2008). PGA: power calculator for case-control genetic association analyses. *BMC Genetics*, *9*, 36.

Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nature Reviews Genetics*, *10*(3), 155–9.

Meuwissen, T. H., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819–1829.

Miu, A. C., Vulturar, R., Chiş, A., Ungureanu, L., & Gross, J. J. (2013). Reappraisal as a mediator in the link between 5-HTTLPR and social anxiety symptoms. *Emotion*, *13*(6), 1012–22. https://doi.org/10.1037/a0033383

Moons, W. G., Way, B. M., & Taylor, S. E. (2014). Oxytocin and vasopressin receptor polymorphisms interact with circulating neuropeptides to predict human emotional reactions to stress. *Emotion*, *14*(3), 562–72. https://doi.org/10.1037/a0035503

Nachar, N. (2008). The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. *Quantitative Methods for Psychology*, *4*(1), 13–20.

Nelson, C. B., Kessler, R. C., & Mroczek, D. (1998). Scoring the World Health Organization's Composite International Diagnostic Interview Short Form (CIDI-SF; v1.0 NOV98). https://doi.org/10.1002/mpr.47

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., … Bustamante,

C. D. (2008). Genes mirror geography within Europe. *Nature*, *456*(7218), 98–101.

Okbay, A., Baselmans, B. M. L., De Neve, J.-E., Turley, P., Nivard, M. G., Fontana, M. A., … Cesarini, D. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics*, *48*(6), 624–633. https://doi.org/10.1038/ng.3552

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., … 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

Okbay, A., & Rietveld, C. A. (2015). On improving the credibility of candidate gene studies: A review of candidate gene studies published in Emotion. *Emotion*. American Psychological Association. https://doi.org/10.1037/emo0000076

Ong, A. D., Edwards, L. M., & Bergeman, C. S. (2006). Hope as a source of resilience in later adulthood. *Personality and Individual Differences*, *41*(7), 1263–1273. https://doi.org/10.1016/j.paid.2006.03.028

Oreopoulos, P. (2006). Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter. *American Economic Review*, *96*(1), 152–175. https://doi.org/10.1257/000282806776157641

Osinsky, R., Lösch, A., Hennig, J., Alexander, N., & MacLeod, C. (2012). Attentional bias to negative information and 5-HTTLPR genotype interactively predict students' emotional reactivity to first university semester. *Emotion*, *12*(3), 460–469. https://doi.org/10.1037/a0026674

Osinsky, R., Reuter, M., Kupper, Y., Schmitz, A., Kozyra, E., Alexander, N., … Hennig, J. J. (2008). Variation in the serotonin transporter gene modulates selective attention to threat. *Emotion*, *8*(4), 584–588. https://doi.org/10.1037/a0012826

Papageorge, N., & Thom, K. (2016). What genetic information can tell us about economic inequality. Retrieved July 12, 2016, from http://www.brookings.edu/blogs/brown-center-chalkboard/posts/2016/05/11-genes-education-economic-inequality-papageorge-thom

Papousek, I., Reiser, E. M., Schulter, G., Fink, A., Holmes, E. a, Nagl, S., … Niederstätter, H. (2013). Serotonin transporter genotype (5-HTTLPR) and electrocortical responses indicating the sensitivity to negative emotional cues. *Emotion (Washington, D.C.)*, *13*(6), 1173–81. https://doi.org/10.1037/a0033997

Pekkarinen, T., Uusitalo, R., & Kerr, S. (2009). School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics*, *93*(7–8), 965–973. https://doi.org/10.1016/j.jpubeco.2009.04.006

Pers, T. H., Karjalainen, J. M., Chan, Y., Westra, H.-J., Wood, A. R., Yang, J., … Franke, L. (2015). Biological interpretation of genome-wide association studies using predicted

gene functions. *Nature Communications*, *6*, 5890. https://doi.org/10.1038/ncomms6890

Pischke, J.-S., & von Wachter, T. (2008). Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation. *Review of Economics and Statistics*, *90*(3), 592–598. https://doi.org/10.1162/rest.90.3.592

Pistis, G., Piras, I., Pirastu, N., Persico, I., Sassu, A., Picciau, A., … Angius, A. (2009). High differentiation among eight villages in a secluded area of Sardinia revealed by genome-wide high density SNPs analysis. *PLoS One*, *4*(2), e4654. https://doi.org/10.1371/journal.pone.0004654

Pistis, G., Porcu, E., Vrieze, S. I., Sidore, C., Steri, M., Danjou, F., … Sanna, S. (2015). Rare variant genotype imputation with thousands of study-specific whole-genome sequences: implications for cost-effective study designs. *European Journal of Human Genetics*, *23*(7), 975–983. https://doi.org/doi:10.1038/ejhg.2014.216

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

Price, A. L., Helgason, A., Palsson, S., Stefansson, H., St Clair, D., Andreassen, O. A., … Stefansson, K. (2009). The impact of divergence time on the nature of population structure: an example from Iceland. *PLoS Genetics*, *5*(6), e1000505. https://doi.org/10.1371/journal.pgen.1000505

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*(8), 904–9. https://doi.org/10.1038/ng1847

Price, A. L., Weale, M. E., Patterson, N., Myers, S. R., Need, A. C., Shianna, K. V, … Reich, D. (2008). Long-range LD can confound genome scans in admixed populations. *American Journal of Human Genetics*, *83*(1), 132-135-139.

Purcell, S. M., Cherny, S. S., & Sham, P. C. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*, *19*, 149–150.

Purcell, S. M., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … others. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*, 559–575.

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., … Consortium, T. I. S. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*(7256), 748–52. https://doi.org/10.1038/nature08185

Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401.

https://doi.org/10.1177/014662167700100306

Ramsköld, D., Wang, E. T., Burge, C. B., & Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Computational Biology*, *5*(12), e1000598. https://doi.org/10.1371/journal.pcbi.1000598

Rietveld, C. A., Cesarini, D., Benjamin, D. J., Koellinger, P. D., De Neve, J.-E., Tiemeier, H., … Bartels, M. (2013). Molecular genetics and subjective well-being. *Proceedings of the National Academy of Sciences*, *110*(24), 9692–9697. https://doi.org/10.1073/pnas.1222171110

Rietveld, C. A., Conley, D. C., Eriksson, N., Esko, T., Medland, S. E., Vinkhuyzen, A. A. E. E., … Koellinger, P. D. (2014). Replicability and robustness of genome-wide-association studies for behavioral traits. *Psychological Science*, *25*(11), 1975–86. https://doi.org/10.1177/0956797614545132

Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Turley, P., Benyamin, B., … Koellinger, P. D. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(38), 13790–13794.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., … Koellinger, P. D. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, *340*(6139), 1467–1471. https://doi.org/10.1126/science.1235488

Ripke, S., Neale, B. M., Corvin, A., Walters, J. T. R., Farh, K.-H., Holmans, P. A., … O'Donovan, M. C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427. https://doi.org/10.1038/nature13595

Ripke, S., Wray, N. R., Lewis, C. M., Hamilton, S. P., Weissman, M. M., Breen, G., … Sullivan, P. F. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular Psychiatry*, *18*(4), 497–511. https://doi.org/10.1038/mp.2012.21.A

Roberts, R. E., Kaplan, G. a, Shema, S. J., & Strawbridge, W. J. (2000). Are the obese at greater risk for depression? *American Journal of Epidemiology*, *152*(2), 163–170. https://doi.org/10.1093/aje/152.2.163

Robson, A. J. (1996). A biological basis for expected and non-expected utility. *Journal of Economic Theory*, *68*, 397–424.

Robson, A. J. (2001). The biological basis of economic behavior. *Journal of Economic Literature*, *XXXIX*(1), 11–33.

Rugulies, R. (2002). Depression as a predictor for coronary heart disease: A review and meta-analysis. *American Journal of Preventive Medicine*, *23*(1), 51–61.

Sacerdote, B. (2010). Nature and nurture effects on children's outcomes: What have we learned from studies of twins and adoptees? In J. Benhabib, M. Jackson, & A. Bisin

(Eds.), *Handbook of Social Economics*. North Holland, Amsterdam.

Schoebi, D., Way, B. M., Karney, B. R., & Bradbury, T. N. (2012). Genetic moderation of sensitivity to positive and negative affect in marriage. *Emotion*, *12*(2), 208–212. https://doi.org/10.1037/a0026067

Seeman, P. (2010). Dopamine D2 receptors as treatment targets in schizophrenia. *Clinical Schizophrenia & Related Psychoses*, *4*(1), 56–73. https://doi.org/10.3371/CSRP.4.1.5

Shahab, L., & West, R. (2012). Differences in happiness between smokers, ex-smokers and never smokers: Cross-sectional findings from a national household survey. *Drug and Alcohol Dependence*, *121*(1–2), 38–44.

Sham, P. C., & Purcell, S. M. (2014). Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics*, *15*, 335–346.

Shungin, D., Winkler, T. W., Croteau-Chonka, D. C., Ferreira, T., Locke, A. E., Mägi, R., … Mohlke, K. L. (2015). New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, *518*(7538), 187–196. https://doi.org/10.1038/nature14132

Smith, D. J., Escott-Price, V., Davies, G., Bailey, M. E., Colodro Conde, L., Ward, J., … O'Donovan, M. (2015). Genome-wide analysis of over 106,000 individuals identifies 9 neuroticism-associated loci. *bioRxiv*, 32417. https://doi.org/10.1101/032417

Smith, D. J., Nicholl, B. I., Cullen, B., Martin, D., Ul-Haq, Z., Evans, J., … Pell, J. P. (2013). Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PloS One*, *8*(11), e75362.

Sniekers, S., Stringer, S., Watanabe, K., Taskesen, E., Okbay, A., Amin, N., … Posthuma, D. (2016). Meta-analysis on 78,308 individuals identifies 15 novel loci and 36 novel genes for intelligence.

Sonnega, A., Faul, J. D., Ofstedal, M. B., Langa, K. M., Phillips, J. W. R., & Weir, D. R. (2014). Cohort profile: the Health and Retirement Study (HRS). *International Journal of Epidemiology*, *43*(2), 576–585. https://doi.org/10.1093/ije/dyu067

Spencer, C. C. A., Plagnol, V., Strange, A., Gardner, M., Paisan-Ruiz, C., Band, G., … Wood, N. W. (2011). Dissection of the genetics of Parkinson's disease identifies an additional association 5' of SNCA and multiple associated haplotypes at 17q21. *Human Molecular Genetics*, *20*(2), 345–353.

Steinberg, K. M., Antonacci, F., Sudmant, P. H., Kidd, J. M., Campbell, C. D., Vives, L., … Eichler, E. E. (2012). Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nature Genetics*, *44*(8), 872–80. https://doi.org/10.1038/ng.2335

Stetler, C., & Miller, G. E. (2011). Depression and hypothalamic-pituitary-adrenal activation: a quantitative summary of four decades of research. *Psychosomatic Medicine*, *73*(2), 114–126.

Stigler, G. J., & Becker, G. S. (1977). De gustibus non est disputandum. *American Economic Review*, *67*(2), 76–90.

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(16), 9440–5. https://doi.org/10.1073/pnas.1530509100

Sudlow, C., Gallacher, J., Allen, N. E., Beral, V., Burton, P., Danesh, J., … Collins, R. (2015). UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., … Korbel, J. O. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, *526*(7571), 75–81. https://doi.org/10.1038/nature15394

Sullivan, P. F. (2012). Don't give up on GWAS. *Molecular Psychiatry*, *17*(1), 2–3.

Tambs, K., Sundet, J. M., Magnus, P., & Berg, K. (1989). Genetic and environmental contributions to the covariance between occupational status, educational attainment, and IQ: a study of twins. *Behavior Genetics*, *19*(2), 209–222.

Taylor, J. A. (1953). A personality scale of manifest anxiety. *Journal of Abnormal Psychology*, *48*(2), 285–290. https://doi.org/10.1037/h0056264

Tellegen, A., & Waller, N. (2008). Exploring personality through test construction: Development of the multidimensional personality questionnaire. In G. J. Boyle, G. Matthews, & S. D. H. (Eds.), *The SAGE Handbook of Personality Theory and Assessment* (pp. 261–272). SAGE Publications.

The 1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*, 56–65.

The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

The GTEx Consortium. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, *45*(6), 580–5.

The International HapMap 3 Consortium, Altshuler, D. M., Gibbs, R. A., Peltonen, L., Dermitzakis, E., Schaffner, S. F., … McEwen, J. E. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, *467*(7311), 52–8. https://doi.org/10.1038/nature09298

The International HapMap Consortium. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, *449*(7164), 851–861.

The International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931–45. https://doi.org/10.1038/nature03001

Thompson, L. A., Detterman, D. K., & Plomin, R. (1991). Associations between cognitive

abilities and scholastic achievement: Genetic overlap but environmental differences. *Psychological Science*, *2*(3), 158–165.

Tian, C., Plenge, R. M., Ransom, M., Lee, A., Villoslada, P., Selmi, C., … Seldin, M. F. (2008). Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genetics*, *4*(1), e4.

Tillmann, T., Vaucher, J., Okbay, A., Pikhart, H., Peasey, A., Kubinova, R., … Holmes, M. V. (2016). Causal link between education and coronary artery disease.

Todd, P. E., & Wolpin, K. I. (2003). On The Specification and Estimation of The Production Function for Cognitive Achievement. *Economic Journal*, *113*(485).

Turkheimer, E. (2000). Three laws of Behavior Genetics and what they mean. *Current Directions in Psychological Science*, *9*(5), 160–164.

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*, *14*(6), 623–628. https://doi.org/10.1046/j.0956-7976.2003.psci_1475.x

Turley, P., Maghzian, O., Okbay, A., Lee, J. J., Cesarini, D., & Benjamin, D. J. (n.d.). *Bayesian cross-trait meta-analysis*.

UK Biobank. (2015a). *Genotype Imputation and Genetic Association Studies of UK Biobank: Interim Data Release*.

UK Biobank. (2015b). *Genotyping and Quality Control of UK Biobank, a Large-Scale, Extensively Phenotyped Prospective Resource: Information for Researchers. Interim Data Release.*

UNESCO. (2006). *International Standard Classification of Education ISCED 1997*.

Vallone, D., Picetti, R., & Borrelli, E. (2000). Structure and function of dopamine receptors. *Neuroscience & Biobehavioral Reviews*, *24*(1), 125–132. https://doi.org/10.1016/S0149-7634(99)00063-9

Van Beijsterveldt, C. E. M., Groen-Blokhuis, M., Hottenga, J. J., Franić, S., Hudziak, J. J., Lamb, D., … Boomsma, D. I. (2013). The Young Netherlands Twin Register (YNTR): Longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics*, *16*(1), 252–267. https://doi.org/10.1017/thg.2012.118

Van Damme, E. (1994). Evolutionary game theory. *European Economic Review*, *38*(3–4), 847–858.

Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., … De Jager, P. L. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *The American Journal of Human Genetics*, *97*(4), 576–592.

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., Hemani, G., Vinkhuyzen, A. A. E., Chen, G.-B., Lee, S. H., Wray, N. R., … Yang, J. (2014). Statistical power to detect genetic (co) variance of complex traits using SNP data in unrelated samples. *PLOS Genetics*, *10*, e1004269.

Visscher, P. M., Yang, J., & Goddard, M. E. (2010). A commentary on "common SNPs explain a large proportion of the heritability for human height" by Yang et al. (2010). *Twin Research and Human Genetics*, *13*(6), 517–524.

Watson, D., & Clark, L. A. (1994). *The PANAS-X manual for the positive and negative affect schedule - expanded form*. *Iowa Research Online*.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070.

Weedon, M. N., Lango, H., Lindgren, C. M., Wallace, C., Evans, D. M., Mangino, M., … others. (2008). Genome-wide association analysis identifies 20 loci that influence adult height. *Nature Genetics*, *40*, 575–583.

Weiner, D. J., Wigdor, E. M., Ripke, S., Walters, R. K., Kosmicki, J. A., Grove, J., … Robinson, E. B. (2016). *Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders*. *bioRxiv*. https://doi.org/10.1101/089342

Weir, B. S. (1990). *Genetic Data Analysis*. Sinauer.

Weiss, A., Baselmans, B. M. L., Hofer, E., Yang, J., Okbay, A., Lind, P. A., … Luciano, M. (2016). Personality polygenes, positive affect, and life satisfaction. *Twin Research and Human Genetics*, *19*(5), 407–417. https://doi.org/10.1017/thg.2016.65

Weiss, A., Bates, T. C., & Luciano, M. (2008). Happiness is a personal(ity) thing: The genetics of personality and well-being in a representative sample. *Psychological Science*, *19*(3), 205–210.

Willemsen, G., Vink, J. M., Abdellaoui, A., den Braber, A., van Beek, J. H. D. a., Draisma, H. H. M., … Boomsma, D. I. (2013). The Adult Netherlands Twin Register: Twenty-five years of survey and biological data collection. *Twin Research and Human Genetics*, *16*(1), 271–281. https://doi.org/10.1017/thg.2012.140

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, *26*(17), 2190–2191.

Wilson, E. O. (1998). *Consilience: The Unity of Knowledge* (Vol. 15). New York: Random House.

Winkler, T. W., Day, F. R., Croteau-Chonka, D. C., Wood, A. R., Locke, A. E., Mägi, R., … Loos, R. J. F. (2014). Quality control and conduct of genome-wide association meta-analyses. *Nature Protocols*, *9*(5), 1192–212.

Witte, J. S., Visscher, P. M., & Wray, N. R. (2014). The contribution of genetic variants to disease depends on the ruler. *Nature Reviews Genetics*, *15*, 765–776.

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., … Frayling, T. M. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, *46*(11), 1173–86. https://doi.org/10.1038/ng.3097

Wray, N. R., Lee, S. H., & Kendler, K. S. (2012). Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. *European Journal of Human Genetics*, *20*, 668–674.

Wray, N. R., & Maier, R. (2014). Genetic basis of complex genetic disease: the contribution of disease heterogeneity to missing heritability. *Current Epidemiology Reports*, *1*, 220–227.

Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics*, *14*(7), 507–515. https://doi.org/10.1038/nrg3457

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A. A. E., Lee, S. H., … others. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics*, *47*, 1114–1120.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., … others. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*, 565–569.

Yang, J., Ferreira, T., Morris, A. P., Medland, S. E., Madden, P. A. F., Heath, A. C., … Visscher, P. M. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics*, *44*(4), 369–375.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yang, J., Weedon, M. N., Purcell, S., Lettre, G., Estrada, K., Willer, C. J., … Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2011). Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics*, *19*(7), 807–812. https://doi.org/doi:10.1038/ng.3211

Yang, J., Zaitlen, N. a, Goddard, M. E., Visscher, P. M., & Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics*, *46*(2), 100–6.

Yesavage, J., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1983). Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of Psychiatric Research*, *17*, 37–49.

Zhu, M., & Zhao, S. (2007). Candidate gene identification approach: progress and challenges. *International Journal of Biological Sciences*, *3*(7), 420–7. Retrieved from

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2043166&tool=pmcentrez&rendertype=abstract

Zigmond, S., & Snaith, R. P. (1983). The hospital anxiety and depression scale (HADS). *Acta Psychiatrica Scandinavica*, *67*, 361–370.

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*, 1193–1198.

Zuk, O., Schaffner, S. F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., … Lander, E. S. (2014). Searching for missing heritability: Designing rare variant association studies. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, E455--E464.

Acknowledgments

The work presented in this thesis was made possible with the support and scientific contributions of many people. I would like to thank all my co-authors for their contribution. It would unfortunately be infeasible to thank everyone personally here as some of the chapters are co-authored by more than a hundred researchers, but I would like to mention a few individuals by name.

I want to start by thanking my supervisors Philipp Koellinger, Roy Thurik and Patrick Groenen for their support and guidance. Philipp, thank you for introducing me to behavior genetics and the SSGAC, it has been and continuing to be a great journey. Also, thank you for always leaving your cats with me when you are travelling, they have been great morale. Leo wanted me to add the following, "5r44444444z<ar2111111111111111111111111y66666666666666666666666666666666677777777777777777" – or at least I think that is what he intends to say by sitting on my laptop while I am writing the acknowledgments. I think it might be the meaning of life. 42 was too short anyway.

Thanks to the small committee – Magnus Johannesson, Danielle Posthuma, Andre Uitterlinden for reading and evaluating my thesis. Thanks to Stephan Ripke, Cornelia van Duijn and Dan Benjamin for agreeing to be in my PhD defense committee.

Thanks to David Cesarini and Dan Benjamin for showing me the ropes. Thanks, David, for seeing EA2 and SWB through with me, always being available for advice and not hesitating to get your hands dirty when I needed help. Without your dedication, the work presented in this thesis would not have been possible.

Patrick Turley, Mark Fontana, Tonu Esko, Kevin Thom, Jonathan Beauchamp and Sven Oskarsson, I truly enjoyed working with you and the times we spent together in various SSGAC adventures. Patrick and Mark, you were a pleasure to work with. Thank you for collaborating on Chapters 3 and 4. Tonu, thank you for your support and encouragement throughout EA2. To quote Isaac Newton, if I have seen further it is by standing on the shoulders of GIANT's (pun intended). Kevin, Jonathan and Sven, thank you for collaborating on the fifth chapter, and the effort you have put into meeting my thesis deadline.

Bart, it was a blast to work with you on Chapter 3. Thank you for patiently listening to my rants, helping with the Dutch translations, and your support while I was writing this thesis.

Against all logic, I enjoyed being reminded every other day that I need to stop procrastinating and get it over with.

Thanks to the EURIBEB team - Ronald, Niels, Fleur, and Richard - for our good time together. Looking forward to more of it! Ronald, thank you for your support throughout the four years, and the random fun chats on topics varying from Cholesky decompositions to Dutch royalty. It has been great to have your company throughout the four years.

I would like to thank my parents and my brother Doruk for their love and support, and for being understanding about my shortage of time and attention.

Roxana, thank you for blowing off steam with me at our self-declared happy hours, and feeding me chocolate every time I drop by your office. And the great job I am sure you will do as my paranymph.

My heartfelt thanks, Firat, for tolerating my extreme working hours (albeit with some complaining) and your support and love through thick and thin in the past four years, which I imagine were not always particularly enjoyable. And last but not least, I would like to thank Memik and Lucy, my best feline and canine friends, for all the unconditional love.

About the Author

Aysu Okbay was born on June 17, 1987 in Istanbul, Turkey. She completed her high school education in 2006 at the Istanbul High School in Istanbul. She graduated summa cum laude from Sabanci University, Istanbul, in 2010, with a major in Economics and minor in Mathematics. She continued her studies in Economics at Sabanci University with a focus in Game Theory, and received her Masters degree in 2012. She started her PhD research in August 2012 at the department of Applied Economics in Erasmus University Rotterdam under the supervision of professors Roy Thurik, Philipp Koellinger and Patrick Groenen.

Aysu's research focuses on questions at the intersection of economics and biology. Her main research interest lies in exploring the genetic architecture of complex behavioral outcomes. She has presented her work at various international conferences, such as the annual meetings of Behavior Genetics Association and the CHARGE Consortium. Her research has been published in leading peer-reviewed journals including Nature, Nature Genetics, and Emotion. Aysu is currently continuing her career as a postdoctoral researcher at the Complex Trait Genetics Lab in Vrije Universiteit Amsterdam, and a core researcher in the Social Science Genetic Association Consortium (SSGAC).

PhD Portfolio

Educational background

2012-2016	**PhD in Applied Economics**, Erasmus School of Economics, Erasmus University Rotterdam, the Netherlands. Project title: "*Genoeconomics - The molecular genetics of economic preferences and outcomes"*. Dissertation title: "*Essays on genetics and the social sciences*".
2010-2012	**Master of Arts in Economics**, Sabanci University, Istanbul, Turkey. GPA: 3.95 / 4.00. Dissertation title: "*One–memory in multiperson bargaining*".
2006-2010	**Bachelor of Arts in Economics**, Sabanci University, Istanbul, Turkey. Summa cum laude, overall GPA: 3.94 / 4.00. **Minor in Mathematics**, GPA: 3.96 / 4.00.

Publications

- De Vlaming, R., Okbay, A., Rietveld, C.A., Johannesson, M., Magnusson, P.K.E., Uitterlinden, A.G., … Koellinger, P.D. (2016). Meta-GWAS Accuracy and Power (MetaGAP) calculator shows that hiding heritability is partially due to imperfect genetic correlations across studies. Forthcoming in *PLOS Genetics*.
- Kong, A., Frigge, M.L., Thorleifsson, G., Stefansson, H., Young, A.I., Zink, F., Okbay, A., … Stefansson, K. (2016). Selection against the genetic basis of educational attainment. Forthcoming in *PNAS*.
- Hill, D., Hagenaars, S.P., Marioni, R.E., Harris, S.E., Liewald, D.C.M, Davies, G., Okbay, A., … Deary, I.J. (2016). Molecular genetic contributions to social deprivation and household income in UK Biobank. *Current Biology*, 26(22), 3083-3089. doi:10.1016/j.cub.2016.09.035.
- Marioni, R.E., Ritchie, S.J., Joshi, P.K., Hagenaars, S.P., Okbay, A., Fischer, K., … Deary, I.J. (2016). Genetic variants linked to education predict longevity. *PNAS*, 113(47), 13366-13371. doi:10.1073/pnas.1605334113.

- Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.J., Pers, T.H., Rietveld, C.A., … Benjamin, D.J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature,* 533, 539–542. doi:10.1038/nature17671.
- Okbay, A., Baselmans, B.M.L, De Neve, J.E., Turley, P., Nivard, M.G, Fontana, M.A., … Cesarini, D. (2016). Genetic variants associated with subjective well-being, depressive symptoms and neuroticism identified through genome-wide analyses. *Nature Genetics*, 48, 624633. doi:10.1038/ng.3552.
- Weiss, A., Baselmans, B. M. L., Hofer, E., Yang, J., Okbay, A., Lind, P. A., … Luciano, M. (2016). Personality polygenes, positive affect, and life satisfaction. Twin Research and Human Genetics, 19(5), 407–417. https://doi.org/10.1017/thg.2016.65
- Okbay, A., Rietveld, C.A. (2015). On improving the credibility of candidate gene studies: A review of candidate gene studies published in Emotion. *Emotion*, 15(4), 531–537. doi:10.1037/emo0000076.

Manuscripts under review

- Sniekers, S., Stringer, S., Watanabe, K., Taskesen, E., Okbay, A., Amin, N., … Posthuma, D. (2016). Meta-analysis on 78,308 individuals identifies 15 novel loci and 36 novel genes for intelligence. Under review at *Nature*.
- Domingue, B.W., Liu, H., Okbay, A., Belsky, D.W. (2016). Genetic heterogeneity in depressive symptoms following the death of a spouse: Polygenic score analysis of the US Health and Retirement Study. Under review at *American Journal of Psychiatry*.
- Tillmann, T., Vaucher, J., Okbay, A., Pikhart, J., Peasey, A., Kubinova, R., … Holmes, M.V. (2016). Causal link between education and coronary artery disease. Under review at *New England Journal of Medicine.*

Manuscripts in preparation

- Turley, P., Maghzian, O., Okbay, A., Fontana, M., Benjamin, D.J., Cesarini, D. *Bayesian cross-trait meta-analysis.*

- Beauchamp, J.P., Oskarsson, S., Okbay, A., Thom, K., Cesarini, D. *Of genes and screens: Educational reform, ability, and labor market screening.*
- Dawes, C.T., Lindgren, K.-O., Oskarsson, S., Okbay, A., Cesarini, D., Johannesson, M., & Magnusson, P.K.E. *The relationship between genes, education, and voter turnout.*
- Social Science Genetic Association Consortium. *Genome-wide association study of 155,439 individuals identifies one locus associated with risk tolerance.*
- Weiner, D. J., Wigdor, E. M., Ripke, S., Walters, R. K., Kosmicki, J. A., Grove, J., … Robinson, E. B. (2016). *Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. bioRxiv.* https://doi.org/10.1101/089342

Conference and seminar presentations

- Behavior Genetics Association, 21-23 June 2016, Brisbane, Australia.
- Erasmus Happiness Economics Research Organisation (EHERO) Seminar, 29 January 2016, Rotterdam, the Netherlands.
- Social Science Genetic Association Consortium Working Group Meeting, 2015 CHARGE Investigator Meeting, 1 July 2015, Jackson, USA.
- Behavior Genetics Association, 17-20 June 2015, San Diego, USA
- Social Science Genetic Association Consortium Working Group Meeting, 2014 CHARGE Investigator Meeting, 13 November 2014, Washington DC, USA.
- 2nd SSGAC Internal Workshop, 14-16 August 2014, Bro, Sweden.
- OSE Seminar, 21 May 2014, Erasmus School of Economics, Rotterdam, the Netherlands.
- Social Science Genetic Association Consortium Working Group Meeting in 2014 CHARGE Investigator Meeting, 23 January 2014, Los Angeles, USA.
- 1st SSGAC Internal Workshop, 18-22 August 2013, Ösmo, Sweden.

Refereeing

Economics and Biology, Psychological Methods, Psychoneuroendocrinology, Molecular Psychiatry

Teaching activities

2016	Guest lecture - Genome-Wide Data Analysis Summer School, Tinbergen Institute.
2013 - 2016	Supervision – 2 graduate theses, 5 undergraduate theses.
2014, 2016	Seminar in Small Business and Entrepreneurship (FEM11055)
2013, 2014	Seminar in Innovation and Entrepreneurship (FEM11063)
2013, 2014	Teaching Assistant - Economics of Innovation (FEM11011).

(Selected) PhD courses

- The 2014 International Workshop on Statistical Genetic Methods for Human Complex Traits
- The SNP Course IX, Erasmus Medical Center
- 2012 Erasmus Medical Center Summer Programme
- Programming
- Data Analysis in R
- Statistical Methods
- Behavioral Decision Theory
- The New Statistics – Estimation for better research: Effect sizes, confidence intervals, and meta-analysis.

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT (ERIM)

The ERIM PhD Series

The ERIM PhD Series contains PhD dissertations in the field of Research in Management defended at Erasmus University Rotterdam and supervised by senior researchers affiliated to the Erasmus Research Institute of Management (ERIM). All dissertations in the ERIM PhD Series are available in full text through the ERIM Electronic Series Portal: http://repub.eur.nl/pub. ERIM is the joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics at the Erasmus University Rotterdam (EUR).

Dissertations in the last five years

Abbink, E.J., *Crew Management in Passenger Rail Transport*, Promotors: Prof. L.G. Kroon & Prof. A.P.M. Wagelmans, EPS-2014-325-LIS, http://repub.eur.nl/pub/76927

Acar, O.A., *Crowdsourcing for Innovation: Unpacking Motivational, Knowledge and Relational Mechanisms of Innovative Behavior in Crowdsourcing Platforms*, Promotor: Prof. J.C.M. van den Ende, EPS-2014-321-LIS, http://repub.eur.nl/pub/76076

Akin Ates, M., *Purchasing and Supply Management at the Purchase Category Level: strategy, structure and performance*, Promotors: Prof. J.Y.F. Wynstra & Dr E.M. van Raaij, EPS-2014-300-LIS, http://repub.eur.nl/pub/50283

Akpinar, E., *Consumer Information Sharing*, Promotor: Prof. A. Smidts, EPS-2013-297-MKT, http://repub.eur.nl/pub/50140

Alexander, L., *People, Politics, and Innovation: A Process Perspective*, Promotors: Prof. H.G. Barkema & Prof. D.L. van Knippenberg, EPS-2014-331-S&E, http://repub.eur.nl/pub/77209

Alexiou, A. *Management of Emerging Technologies and the Learning Organization : Lessons from the Cloud and Serious Games Technology*, Promotors: Prof. S.J. Magala, Prof. M.C. Schippers and Dr. I. Oshri, EPS-2016-404-ORG, http://repub.eur.nl/pub/93818

Almeida e Santos Nogueira, R.J. de, *Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty*, Promotors: Prof. U. Kaymak & Prof. J.M.C. Sousa, EPS-2014-310-LIS, http://repub.eur.nl/pub/51560

Bannouh, K., *Measuring and Forecasting Financial Market Volatility using High-frequency Data*, Promotor: Prof. D.J.C. van Dijk, EPS-2013-273-F&A, http://repub.eur.nl/pub/38240

Ben-Menahem, S.M., *Strategic Timing and Proactiveness of Organizations*, Promotors:

Prof. H.W. Volberda & Prof. F.A.J. van den Bosch, EPS-2013-278-S&E, http://repub.eur.nl/pub/39128

Benning, T.M., *A Consumer Perspective on Flexibility in Health Care: Priority Access Pricing and Customized Care*, Promotor: Prof. B.G.C. Dellaert, EPS-2011-241-MKT, http://repub.eur.nl/pub/23670
Benschop, N, *Biases in Project Escalation: Names, frames & construal levels*, Promotors: Prof. K.I.M. Rhode, Prof. H.R. Commandeur, Prof. M.Keil & Dr A.L.P. Nuijten, EPS-2015-375-S&E, hdl.handle.net/1765/79408

Berg, W.E. van den, *Understanding Salesforce Behavior using Genetic Association Studies*, Promotor: Prof. W.J.M.I. Verbeke, EPS-2014-311-MKT, http://repub.eur.nl/pub/51440

Betancourt, N.E., *Typical Atypicality: Formal and Informal Institutional Conformity, Deviance, and Dynamics*, Promotor: Prof. B. Krug, EPS-2012-262-ORG, http://repub.eur.nl/pub/32345

Beusichem, H.C. van, *Firms and Financial Markets: Empirical Studies on the Informational Value of Dividends, Governance and Financial Reporting*, Promotors: Prof. A. de Jong & Dr. G. Westerhuis, EPS-2016-378-F&A, http://repub.eur.nl/pub/93079

Bliek, R. de, *Empirical Studies on the Economic Impact of Trust*, Promotor: Prof. J. Veenman & Prof. Ph.H.B.F. Franses, EPS-2015-324-ORG, http://repub.eur.nl/pub/78159

Blitz, D.C., *Benchmarking Benchmarks*, Promotors: Prof. A.G.Z. Kemna & Prof. W.F.C. Verschoor, EPS-2011-225-F&A, http://repub.eur.nl/pub/22624

Boons, M., *Working Together Alone in the Online Crowd: The Effects of Social Motivationsand Individual Knowledge Backgrounds on the Participation and Performance of Members of Online Crowdsourcing Platforms*, Promotors: Prof. H.G. Barkema & Dr D.A. Stam, EPS-2014-306-S&E, http://repub.eur.nl/pub/50711

Brazys, J., *Aggregated Marcoeconomic News and Price Discovery*, Promotor: Prof. W.F.C. Verschoor, EPS-2015-351-F&A, http://repub.eur.nl/pub/78243

Burger, M.J., *Structure and Cooptition in Urban Networks*, Promotors: Prof. G.A. van der Knaap & Prof. H.R. Commandeur, EPS-2011-243-ORG, http://repub.eur.nl/pub/26178

Byington, E., *Exploring Coworker Relationships: Antecedents and Dimensions of Interpersonal Fit,Coworker Satisfaction, and Relational Models*, Promotor: Prof. D.L. van Knippenberg, EPS-2013-292-ORG, http://repub.eur.nl/pub/41508

Camacho, N.M., *Health and Marketing: Essays on Physician and Patient Decision-Making*, Promotor: Prof. S. Stremersch, EPS-2011-237-MKT, http://repub.eur.nl/pub/23604

Cancurtaran, P., *Essays on Accelerated Product Development*, Promotors: Prof. F. Langerak & Prof.G.H. van Bruggen, EPS-2014-317-MKT, http://repub.eur.nl/pub/76074

Caron, E.A.M., *Explanation of Exceptional Values in Multi-dimensional Business Databases*, Promotors: Prof. H.A.M. Daniels & Prof. G.W.J. Hendrikse, EPS-2013-296-LIS, http://repub.eur.nl/pub/50005

Carvalho, L. de, *Knowledge Locations in Cities: Emergence and Development Dynamics*, Promotor: Prof. L. Berg, EPS-2013-274-S&E, http://repub.eur.nl/pub/38449

Cranenburgh, K.C. van, *Money or Ethics: Multinational corporations and religious organisations operating in an era of corporate responsibility*, Prof. L.C.P.M. Meijs, Prof. R.J.M. van Tulder & Dr D. Arenas, EPS-2016-385-ORG, http://repub.eur.nl/pub/93104

Consiglio, I., *Others: Essays on Interpersonal and Consumer Behavior*, Promotor: Prof. S.M.J. van Osselaer, EPS-2016-366-MKT, http://repub.eur.nl/pub/79820

Cox, R.H.G.M., *To Own, To Finance, and To Insure - Residential Real Estate Revealed*, Promotor: Prof. D. Brounen, EPS-2013-290-F&A, http://repub.eur.nl/pub/40964

Darnihamedani, P. *Individual Characteristics, Contextual Factors and Entrepreneurial Behavior*, Promotors: Prof. A.R. Thurik & S.J.A. Hessels, EPS-2016-360-S&E, http://repub.eur.nl/pub/93280

Deichmann, D., *Idea Management: Perspectives from Leadership, Learning, and Network Theory*, Promotor: Prof. J.C.M. van den Ende, EPS-2012-255-ORG, http://repub.eur.nl/pub/31174

Deng, W., *Social Capital and Diversification of Cooperatives*, Promotor: Prof. G.W.J. Hendrikse, EPS-2015-341-ORG, http://repub.eur.nl/pub/77449

Depecik, B.E., *Revitalizing brands and brand: Essays on Brand and Brand Portfolio Management Strategies*, Promotors: Prof. G.H. van Bruggen, dr Y.M. van Everdingen and Dr M.B. Ataman, EPS-2016406-MKT, http://repub.eur.nl/pub/93507

Desmet, P.T.M., *In Money we Trust? Trust Repair and the Psychology of Financial Compensations*, Promotor: Prof. D. de Cremer, EPS-2011-232-ORG, http://repub.eur.nl/pub/23268

Dollevoet, T.A.B., *Delay Management and Dispatching in Railways*, Promotor: Prof. A.P.M. Wagelmans, EPS-2013-272-LIS, http://repub.eur.nl/pub/38241

Doorn, S. van, *Managing Entrepreneurial Orientation*, Promotors: Prof. J.J.P. Jansen, Prof. F.A.J. van den Bosch, & Prof. H.W. Volberda, EPS-2012-258-STR, http://repub.eur.nl/pub/32166

Douwens-Zonneveld, M.G., *Animal Spirits and Extreme Confidence: No Guts, No Glory?*

Promotor: Prof. W.F.C. Verschoor, EPS-2012-257-F&A, http://repub.eur.nl/pub/31914

Duca, E., *The Impact of Investor Demand on Security Offerings*, Promotor: Prof. A. de Jong, EPS-2011-240-F&A, http://repub.eur.nl/pub/26041

Duyvesteyn, J.G. *Empirical Studies on Sovereign Fixed Income Markets,* Promotors: Prof. P.Verwijmeren & Prof. M.P.E. Martens, EPS-2015-361-F&A, hdl.handle.net/1765/79033

Duursema, H., *Strategic Leadership: Moving Beyond the Leader-Follower Dyad*, Promotor: Prof. R.J.M. van Tulder, EPS-2013-279-ORG, http://repub.eur.nl/pub/39129

Eck, N.J. van, *Methodological Advances in Bibliometric Mapping of Science*, Promotor: Prof. R. Dekker, EPS-2011-247-LIS, http://repub.eur.nl/pub/26509

Elemes, A, *Studies on Determinants and Consequences of Financial Reporting Quality,* Promotor: Prof. E.Peek, EPS-2015-354-F&A, http://hdl.handle.net/1765/79037

Ellen, S. ter, *Measurement, Dynamics, and Implications of Heterogeneous Beliefs in Financial Markets*, Promotor: Prof. W.F.C. Verschoor, EPS-2015-343-F&A, http://repub.eur.nl/pub/78191

Erlemann, C., *Gender and Leadership Aspiration: The Impact of the Organizational Environment*, Promotor: Prof. D.L. van Knippenberg, EPS-2016-376-ORG, http://repub.eur.nl/pub/79409

Eskenazi, P.I., *The Accountable Animal*, Promotor: Prof. F.G.H. Hartmann, EPS-2015-355-F&A, http://repub.eur.nl/pub/78300

Essen, M. van, *An Institution-Based View of Ownership*, Promotors: Prof. J. van Oosterhout & Prof. G.M.H. Mertens, EPS-2011-226-ORG, http://repub.eur.nl/pub/22643

Evangelidis, I., *Preference Construction under Prominence*, Promotor: Prof. S.M.J. van Osselaer, EPS-2015-340-MKT, http://repub.eur.nl/pub/78202

Faber, N., *Structuring Warehouse Management*, Promotors: Prof. MB.M. de Koster & Prof. Ale Smidts, EPS-2015-336-LIS, http://repub.eur.nl/pub/78603

Fernald, K., *The Waves of Biotechnological Innovation in Medicine: Interfirm Cooperation Effects and a Venture Capital Perspective,* Promotors: Prof. E.Claassen, Prof. H.P.G. Pennings & Prof. H.R. Commandeur, EPS-2015-371-S&E, http://hdl.handle.net/1765/79120

Fliers, P.T., *Essays on Financing and Performance: The role of firms, banks and board*, Promotor: Prof. A. de Jong & Prof P.G.J. Roosenboom, EPS-2016-388-F&A, http://repub.eur.nl/pub/93019

Fourne, S.P., *Managing Organizational Tensions: A Multi-Level Perspective on Exploration, Exploitation and Ambidexterity*, Promotors: Prof. J.J.P. Jansen & Prof. S.J. Magala, EPS-2014-318-S&E, http://repub.eur.nl/pub/76075

Gaast, J.P. van der, *Stochastic Models for Order Picking Systems*, Promotors: Prof. M.B.M de Koster & Prof. I.J.B.F. Adan, EPS-2016-398-LIS, http://repub.eur.nl/pub/93222

Gharehgozli, A.H., *Developing New Methods for Efficient Container Stacking Operations*, Promotor: Prof. M.B.M. de Koster, EPS-2012-269-LIS, http://repub.eur.nl/pub/37779

Gils, S. van, *Morality in Interactions: On the Display of Moral Behavior by Leaders and Employees*, Promotor: Prof. D.L. van Knippenberg, EPS-2012-270-ORG, http://repub.eur.nl/pub/38027

Ginkel-Bieshaar, M.N.G. van, *The Impact of Abstract versus Concrete Product Communications on Consumer Decision-making Processes*, Promotor: Prof. B.G.C. Dellaert, EPS-2012-256-MKT, http://repub.eur.nl/pub/31913

Gkougkousi, X., *Empirical Studies in Financial Accounting*, Promotors: Prof. G.M.H. Mertens & Prof. E. Peek, EPS-2012-264-F&A, http://repub.eur.nl/pub/37170

Glorie, K.M., *Clearing Barter Exchange Markets: Kidney Exchange and Beyond*, Promotors: Prof. A.P.M. Wagelmans & Prof. J.J. van de Klundert, EPS-2014-329-LIS, http://repub.eur.nl/pub/77183

Hekimoglu, M., *Spare Parts Management of Aging Capital Products,* Promotor: Prof. R. Dekker, EPS-2015-368-LIS, http://hdl.handle.net/1765/79092

Heij, C.V., *Innovating beyond Technology. Studies on how management innovation, co-creation and business model innovation contribute to firm's (innovation) performance*, Promotors: Prof. F.A.J. van den Bosch & Prof. H.W. Volberda, EPS-2012-370-STR, http://repub.eur.nl/pub/78651

Heyde Fernandes, D. von der, *The Functions and Dysfunctions of Reminders*, Promotor: Prof. S.M.J. van Osselaer, EPS-2013-295-MKT, http://repub.eur.nl/pub/41514

Heyden, M.L.M., *Essays on Upper Echelons & Strategic Renewal: A Multilevel Contingency Approach*, Promotors: Prof. F.A.J. van den Bosch & Prof. H.W. Volberda, EPS-2012-259-STR, http://repub.eur.nl/pub/32167

Hoever, I.J., *Diversity and Creativity*, Promotor: Prof. D.L. van Knippenberg, EPS-2012-267-ORG, http://repub.eur.nl/pub/37392

Hogenboom, A.C., *Sentiment Analysis of Text Guided by Semantics and Structure,* Promotors: Prof. U. Kaymak & Prof. F.M.G. de Jong, EPS-2015-369-LIS, http://hdl.handle.net/1765/79034

Hogenboom, F.P., *Automated Detection of Financial Events in News Text*, Promotors: Prof. U. Kaymak & Prof. F.M.G. de Jong, EPS-2014-326-LIS, http://repub.eur.nl/pub/77237

Hollen, R.M.A., *Exploratory Studies into Strategies to Enhance Innovation-Driven International Competitiveness in a Port Context: Toward Ambidextrous Ports,* Promotors: Prof. F.A.J. Van Den Bosch & Prof. H.W.Volberda, EPS-2015-372-S&E, hdl.handle.net/1765/78881

Hoogendoorn, B., *Social Entrepreneurship in the Modern Economy: Warm Glow, Cold Feet*, Promotors: Prof. H.P.G. Pennings & Prof. A.R. Thurik, EPS-2011-246-STR, http://repub.eur.nl/pub/26447

Hoogervorst, N., *On The Psychology of Displaying Ethical Leadership: A Behavioral Ethics Approach*, Promotors: Prof. D. de Cremer & Dr M. van Dijke, EPS-2011-244-ORG, http://repub.eur.nl/pub/26228

Hout, D.H. van, *Measuring Meaningful Differences: Sensory Testing Based Decision Making in an Industrial Context; Applications of Signal Detection Theory and Thurstonian Modelling*, Promotors: Prof. P.J.F. Groenen & Prof. G.B. Dijksterhuis, EPS-2014-304-MKT, http://repub.eur.nl/pub/50387

Houwelingen, G.G. van, *Something To Rely On*, Promotors: Prof. D. de Cremer & Prof. M.H. van Dijke, EPS-2014-335-ORG, http://repub.eur.nl/pub/77320

Hurk, E. van der, *Passengers, Information, and Disruptions*, Promotors: Prof. L.G. Kroon & Prof. P.H.M. Vervest, EPS-2015-345-LIS, http://repub.eur.nl/pub/78275

Hytonen, K.A., *Context Effects in Valuation, Judgment and Choice: A Neuroscientific Approach*, Promotor: Prof. A. Smidts, EPS-2011-252-MKT, http://repub.eur.nl/pub/30668

Iseger, P. den, *Fourier and Laplace Transform Inversion with Applications in Finance*, Promotor: Prof. R. Dekker, EPS-2014-322-LIS, http://repub.eur.nl/pub/76954

Jaarsveld, W.L. van, *Maintenance Centered Service Parts Inventory Control*, Promotor: Prof. R. Dekker, EPS-2013-288-LIS, http://repub.eur.nl/pub/39933

Jalil, M.N., *Customer Information Driven After Sales Service Management: Lessons from Spare Parts Logistics*, Promotor: Prof. L.G. Kroon, EPS-2011-222-LIS, http://repub.eur.nl/pub/22156

Kappe, E.R., *The Effectiveness of Pharmaceutical Marketing*, Promotor: Prof. S. Stremersch, EPS-2011-239-MKT, http://repub.eur.nl/pub/23610

Karreman, B., *Financial Services and Emerging Markets*, Promotors: Prof. G.A.

van der Knaap & Prof. H.P.G. Pennings, EPS-2011-223-ORG, http://re-pub.eur.nl/pub/22280

Khanagha, S., *Dynamic Capabilities for Managing Emerging Technologies*, Promotor: Prof. H.W. Volberda, EPS-2014-339-S&E, http://repub.eur.nl/pub/77319

Kil, J., *Acquisitions Through a Behavioral and Real Options Lens*, Promotor: Prof. H.T.J. Smit, EPS-2013-298-F&A, http://repub.eur.nl/pub/50142

Klooster, E. van 't, *Travel to Learn: the Influence of Cultural Distance on Competence Development in Educational Travel*, Promotors: Prof. F.M. Go & Prof. P.J. van Baalen, EPS-2014-312-MKT, http://repub.eur.nl/pub/51462

Koendjbiharie, S.R., *The Information-Based View on Business Network Performance: Revealing the Performance of Interorganizational Networks*, Promotors: Prof. H.W.G.M. van Heck & Prof. P.H.M. Vervest, EPS-2014-315-LIS, http://re-pub.eur.nl/pub/51751

Koning, M., *The Financial Reporting Environment: The Role of the Media, Regulators and Auditors*, Promotors: Prof. G.M.H. Mertens & Prof. P.G.J. Roosenboom, EPS-2014-330-F&A, http://repub.eur.nl/pub/77154

Konter, D.J., *Crossing Borders with HRM: An Inquiry of the Influence of Contextual Differences in the Adoption and Effectiveness of HRM*, Promotors: Prof. J. Paauwe & Dr L.H. Hoeksema, EPS-2014-305-ORG, http://repub.eur.nl/pub/50388

Korkmaz, E., *Bridging Models and Business: Understanding Heterogeneity in Hidden Drivers of Customer Purchase Behavior*, Promotors: Prof. S.L. van de Velde & Prof. D. Fok, EPS-2014-316-LIS, http://repub.eur.nl/pub/76008

Kroezen, J.J., *The Renewal of Mature Industries: An Examination of the Revival of the Dutch Beer Brewing Industry*, Promotor: Prof. P.P.M.A.R. Heugens, EPS-2014-333-S&E, http://repub.eur.nl/pub/77042

Kysucky, V., *Access to Finance in a Cros-Country Context*, Promotor: Prof.dr. L. Norden, EPS-2015-350-F&A, http://repub.eur.nl/pub/78225

Lam, K.Y., *Reliability and Rankings*, Promotor: Prof. Ph.H.B.F. Franses, EPS-2011-230-MKT, http://repub.eur.nl/pub/22977

Lander, M.W., *Profits or Professionalism? On Designing Professional Service Firms*, Promotors: Prof. J. van Oosterhout & Prof. P.P.M.A.R. Heugens, EPS-2012-253-ORG, http://repub.eur.nl/pub/30682

Langhe, B. de, *Contingencies: Learning Numerical and Emotional Associations in an Uncertain World*, Promotors: Prof. B. Wierenga & Prof. S.M.J. van Osselaer,

EPS-2011-236-MKT, http://repub.eur.nl/pub/23504

Lee, C.I.S.G, *Big Data in Management Research: Exploring New Avenues*, Promotors: Prof. S.J. Magala & Dr W.A. Felps, EPS-2016-365-ORG, http://repub.eur.nl/pub/79818

Legault-Tremblay, P.O., Corporate Governance During Market Transition: Heterogeneous responses to Institution Tensions in China, Promotor: Prof. B. Krug, EPS-2015-362-ORG, http://repub.eur.nl/pub/78649

Lenoir, A.S. *Are You Talking to Me? Addressing Consumers in a Globalised World,* Promotors: Prof. S. Puntoni & Prof. S.M.J. van Osselaer, EPS-2015-363-MKT, http://hdl.handle.net/1765/79036

Leunissen, J.M., *All Apologies: On the Willingness of Perpetrators to Apologize*, Promotors: Prof. D. de Cremer & Dr. M. van Dijke, EPS-2014-301-ORG, http://repub.eur.nl/pub/50318

Li, D., *Supply Chain Contracting for After-sales Service and Product Support*, Promotor: Prof. M.B.M. de Koster, EPS-2015-347-LIS, http://repub.eur.nl/pub/78526

Li, Z., *Irrationality: What, Why and How*, Promotor(s): Prof. H. Bleichrodt, Prof. P.P. Wakker, & Prof. K.I.M. Rohde, EPS-2014-338-MKT, http://repub.eur.nl/pub/77205

Liang, Q.X., *Governance, CEO Identity, and Quality Provision of Farmer Cooperatives*, Promotor: Prof. G.W.J. Hendrikse, EPS-2013-281-ORG, http://repub.eur.nl/pub/39253

Liket, K., *Why 'Doing Good' is not Good Enough: Essays on Social Impact Measurement*, Promotors: Prof. H.R. Commandeur & Dr K.E.H. Maas, EPS-2014-307-STR, http://repub.eur.nl/pub/51130

Loos, M.J.H.M. van der, *Molecular Genetics and Hormones: New Frontiers in Entrepreneurship Research*, Promotors: Prof. A.R. Thurik, Prof. P.J.F. Groenen, & Prof. A. Hofman, EPS-2013-287-S&E, http://repub.eur.nl/pub/40081

Lovric, M., *Behavioral Finance and Agent-Based Artificial Markets*, Promotors: Prof. J. Spronk & Prof. U. Kaymak, EPS-2011-229-F&A, http://repub.eur.nl/pub/22814

Lu, Y., *Data-Driven Decision Making in Auction Markets*, Promotors: Prof. H.W.G.M. van Heck & Prof. W. Ketter, EPS-2014-314-LIS, http://repub.eur.nl/pub/51543

Ma, Y., *The Use of Advanced Transportation Monitoring Data for Official Statistics,* Promotors: Prof. L.G. Kroon and Dr Jan van Dalen, EPS-2016-391-LIS, hdl.handle.net/1765/80174

Manders, B., *Implementation and Impact of ISO 9001*, Promotor: Prof. K. Blind, EPS-2014-337-LIS, http://repub.eur.nl/pub/77412

Markwat, T.D., *Extreme Dependence in Asset Markets Around the Globe*, Promotor: Prof. D.J.C. van Dijk, EPS-2011-227-F&A, http://repub.eur.nl/pub/22744

Mees, H., *Changing Fortunes: How China's Boom Caused the Financial Crisis*, Promotor: Prof. Ph.H.B.F. Franses, EPS-2012-266-MKT, http://repub.eur.nl/pub/34930

Mell, J.N., *Connecting Minds: On The Role of Metaknowledge in Knowledge Coordination*, Promotor: Prof. D.L. van Knippenberg, EPS-2015-359-ORG, http://hdl.handle.net/1765/78951

Meuer, J., *Configurations of Inter-firm Relations in Management Innovation: A Study in China's Biopharmaceutical Industry*, Promotor: Prof. B. Krug, EPS-2011-228-ORG, http://repub.eur.nl/pub/22745

Micheli, M.R., *Business Model Innovation: A Journey across Managers' Attention and Inter-Organizational Networks*, Promotor: Prof. J.J.P. Jansen, EPS-2015-344-S&E, http://repub.eur.nl/pub/78241

Mihalache, O.R., *Stimulating Firm Innovativeness: Probing the Interrelations between Managerial and Organizational Determinants*, Promotors: Prof. J.J.P. Jansen, Prof. F.A.J. van den Bosch, & Prof. H.W. Volberda, EPS-2012-260-S&E, http://repub.eur.nl/pub/32343

Milea, V., *News Analytics for Financial Decision Support*, Promotor: Prof. U. Kaymak, EPS-2013-275-LIS, http://repub.eur.nl/pub/38673

Moniz, A, *Textual Analysis of Intangible Information*, Promotor(s): Prof C.B.M. van Riel, Prof. F.M.G de Jong & Dr G.A.J.M. Berens, EPS-2016-393-ORG, http://repub.eur.nl/pub/93001

Mulder, J. *Network design and robust scheduling in liner shipping*, Promotors: Prof. R. Dekker & Dr W.L. van Jaarsveld, Eps-2016-384-LIS, http://repub.eur.nl/pub/80258

Naumovska, I., *Socially Situated Financial Markets: A Neo-Behavioral Perspective on Firms, Investors and Practices*, Promotors: Prof. P.P.M.A.R. Heugens & Prof. A. de Jong, EPS-2014-319-S&E, http://repub.eur.nl/pub/76084

Neerijnen, P., *The Adaptive Organization: the socio-cognitive antecedents of ambidexterity and individual exploration*, Promotors: Prof. J.J.P. Jansen, P.P.M.A.R. Heugens & Dr T.J.M. Mom, EPS-2016-358-S&E, http://repub.eur.nl/pub/93274

Nielsen, L.K., *Rolling Stock Rescheduling in Passenger Railways: Applications in short term planning and in disruption management*, Promotor: Prof. L.G. Kroon, EPS-2011-224-LIS, http://repub.eur.nl/pub/22444

Nuijten, A.L.P., *Deaf Effect for Risk Warnings: A Causal Examination applied to Information Systems Projects*, Promotors: Prof. G.J. van der Pijl, Prof. H.R. Commandeur & Prof. M. Keil, EPS-2012-263-S&E, http://repub.eur.nl/pub/34928

Oord, J.A. van, *Essays on Momentum Strategies in Finance,* Promotor: Prof. H.K. van Dijk, EPS-2016-380-F&A, hdl.handle.net/1765/80036

Osadchiy, S.E., *The Dynamics of Formal Organization: Essays on bureaucracy and formal rules*, Promotor: Prof. P.P.M.A.R. Heugens, EPS-2011-231-ORG, http://repub.eur.nl/pub/23250

Ozdemir, M.N., *Project-level Governance, Monetary Incentives, and Performance in Strategic R&D Alliances*, Promotor: Prof. J.C.M. van den Ende, EPS-2011-235-LIS, http://repub.eur.nl/pub/23550

Peers, Y., *Econometric Advances in Diffusion Models*, Promotor: Prof. Ph.H.B.F. Franses, EPS-2011-251-MKT, http://repub.eur.nl/pub/30586

Peters, M., *Machine Learning Algorithms for Smart Electricity Markets*, Promotor: Prof. W. Ketter, EPS-2014-332-LIS, http://repub.eur.nl/pub/77413

Porck, J., *No Team is an Island: An Integrative View of Strategic Consensus between Groups*, Promotors: Prof. P.J.F. Groenen & Prof. D.L. van Knippenberg, EPS-2013-299-ORG, http://repub.eur.nl/pub/50141

Porras Prado, M., *The Long and Short Side of Real Estate, Real Estate Stocks, and Equity*, Promotor: Prof. M.J.C.M. Verbeek, EPS-2012-254-F&A, http://repub.eur.nl/pub/30848

Poruthiyil, P.V., *Steering Through: How organizations negotiate permanent uncertainty and unresolvable choices*, Promotors: Prof. P.P.M.A.R. Heugens & Prof. S.J. Magala, EPS-2011-245-ORG, http://repub.eur.nl/pub/26392

Pourakbar, M., *End-of-Life Inventory Decisions of Service Parts*, Promotor: Prof. R. Dekker, EPS-2011-249-LIS, http://repub.eur.nl/pub/30584

Pronker, E.S., *Innovation Paradox in Vaccine Target Selection*, Promotors: Prof. H.J.H.M. Claassen & Prof. H.R. Commandeur, EPS-2013-282-S&E, http://repub.eur.nl/pub/39654

Protzner, S. *Mind the gap between demand and supply: A behavioral perspective on demand forecasting,* Promotors: Prof. S.L. van de Velde & Dr L. Rook, EPS-2015-364-LIS, *http*://repub.eur.nl/pub/79355

Pruijssers, J.K., *An Organizational Perspective on Auditor Conduct*, Promotors: Prof. J. van Oosterhout & Prof. P.P.M.A.R. Heugens, EPS-2015-342-S&E, http://repub.eur.nl/pub/78192

Retel Helmrich, M.J., *Green Lot-Sizing*, Promotor: Prof. A.P.M. Wagelmans, EPS-2013-291-LIS, http://repub.eur.nl/pub/41330

Rietdijk, W.J.R. *The Use of Cognitive Factors for Explaining Entrepreneurship,* Promotors: Prof. A.R. Thurik & Prof. I.H.A. Franken, EPS-2015-356-S&E, http://repub.eur.nl/pub/79817

Rietveld, N., *Essays on the Intersection of Economics and Biology*, Promotors: Prof. A.R. Thurik, Prof. Ph.D. Koellinger, Prof. P.J.F. Groenen, & Prof. A. Hofman, EPS-2014-320-S&E, http://repub.eur.nl/pub/76907

Rijsenbilt, J.A., *CEO Narcissism: Measurement and Impact*, Promotors: Prof. A.G.Z. Kemna & Prof. H.R. Commandeur, EPS-2011-238-STR, http://repub.eur.nl/pub/23554

Rösch, D. *Market Efficiency and Liquidity,* Promotor: Prof. M.A. van Dijk, EPS-2015-353-F&A, http://hdl.handle.net/1765/79121

Roza, L., *Employee Engagement In Corporate Social Responsibility: A collection of essays*, Promotor: L.C.P.M. Meijs, EPS-2016-396-ORG, http://repub.eur.nl/pub/93254

Roza-van Vuren, M.W., *The Relationship between Offshoring Strategies and Firm Performance: Impact of innovation, absorptive capacity and firm size*, Promotors: Prof. H.W. Volberda & Prof. F.A.J. van den Bosch, EPS-2011-214-STR, http://repub.eur.nl/pub/22155

Rubbaniy, G., *Investment Behaviour of Institutional Investors*, Promotor: Prof. W.F.C. Verschoor, EPS-2013-284-F&A, http://repub.eur.nl/pub/40068

Schoonees, P. *Methods for Modelling Response Styles*, Promotor: Prof.dr P.J.F. Groenen, EPS-2015-348-MKT, http://repub.eur.nl/pub/79327

Schouten, M.E., *The Ups and Downs of Hierarchy: the causes and consequences of hierarchy struggles and positional loss*, Promotors; Prof. D.L. van Knippenberg & Dr L.L. Greer, EPS-2016-386-ORG, http://repub.eur.nl/pub/80059

Shahzad, K., *Credit Rating Agencies, Financial Regulations and the Capital Markets*, Promotor: Prof. G.M.H. Mertens, EPS-2013-283-F&A, http://repub.eur.nl/pub/39655

Smit, J. *Unlocking Business Model Innovation: A look through the keyhole at the inner workings of Business Model Innovation*, Promotor: H.G. Barkema, EPS-2016-399-S&E, http://repub.eur.nl/pub/93211

Sousa, M.J.C. de, *Servant Leadership to the Test: New Perspectives and Insights*, Promotors: Prof. D.L. van Knippenberg & Dr D. van Dierendonck, EPS-2014-313-ORG, http://repub.eur.nl/pub/51537

Spliet, R., *Vehicle Routing with Uncertain Demand*, Promotor: Prof. R. Dekker, EPS-2013-293-LIS, http://repub.eur.nl/pub/41513

Staadt, J.L., *Leading Public Housing Organisation in a Problematic Situation: A Critical Soft Systems Methodology Approach*, Promotor: Prof. S.J. Magala, EPS-2014-308-ORG, http://repub.eur.nl/pub/50712

Stallen, M., *Social Context Effects on Decision-Making: A Neurobiological Approach*, Promotor: Prof. A. Smidts, EPS-2013-285-MKT, http://repub.eur.nl/pub/39931

Tarakci, M., *Behavioral Strategy: Strategic Consensus, Power and Networks*, Promotor: Prof. D.L. van Knippenberg & Prof. P.J.F. Groenen, EPS-2013-280-ORG, http://repub.eur.nl/pub/39130

Teixeira de Vasconcelos, M., *Agency Costs, Firm Value, and Corporate Investment*, Promotor: Prof. P.G.J. Roosenboom, EPS-2012-265-F&A, http://repub.eur.nl/pub/37265

Troster, C., *Nationality Heterogeneity and Interpersonal Relationships at Work*, Promotor: Prof. D.L. van Knippenberg, EPS-2011-233-ORG, http://repub.eur.nl/pub/23298

Tsekouras, D., *No Pain No Gain: The Beneficial Role of Consumer Effort in Decision-Making*, Promotor: Prof. B.G.C. Dellaert, EPS-2012-268-MKT, http://repub.eur.nl/pub/37542

Tuijl, E. van, *Upgrading across Organisational and Geographical Configurations*, Promotor: Prof. L. van den Berg, EPS-2015-349-S&E, http://repub.eur.nl/pub/78224

Tuncdogan, A., *Decision Making and Behavioral Strategy: The Role of Regulatory Focus in Corporate Innovation Processes*, Promotors: Prof. F.A.J. van den Bosch, Prof. H.W. Volberda, & Prof. T.J.M. Mom, EPS-2014-334-S&E, http://repub.eur.nl/pub/76978

Uijl, S. den, *The Emergence of De-facto Standards*, Promotor: Prof. K. Blind, EPS-2014-328-LIS, http://repub.eur.nl/pub/77382

Vagias, D., *Liquidity, Investors and International Capital Markets*, Promotor: Prof. M.A. van Dijk, EPS-2013-294-F&A, http://repub.eur.nl/pub/41511

Valogianni, K. *Sustainable Electric Vehicle Management using Coordinated Machine Learning*, Promotors: Prof. H.W.G.M. van Heck & Prof. W. Ketter, Eps-2016-387-LIS, http://repub.eur.nl/pub/93018

Veelenturf, L.P., *Disruption Management in Passenger Railways: Models for Timetable, Rolling Stock and Crew Rescheduling*, Promotor: Prof. L.G. Kroon, EPS-2014-327-LIS, http://repub.eur.nl/pub/77155

Venus, M., *Demystifying Visionary Leadership: In search of the essence of effective vision communication*, Promotor: Prof. D.L. van Knippenberg, EPS-2013-289-ORG, http://repub.eur.nl/pub/40079

Vermeer, W., *Propagation in Networks:The impact of information processing at the actor level on system-wide propagation dynamics,* Promotor: Prof. P.H.M.Vervest, EPS-2015-373-LIS, http://repub.eur.nl/pub/79325

Versluis, I., *Prevention of the Portion Size Effect*, Promotors: Prof. Ph.H.B.F. Franses & Dr E.K. Papies, EPS-2016-382-MKT, http://repub.eur.nl/pub/79880

Vishwanathan, P., *Governing for Stakeholders: How Organizations May Create or Destroy Value for their Stakeholders*, Promotors: Prof. J. van Oosterhout & Prof. L.C.P. M. Meijs, EPS-2016-377-ORG, http://repub.eur.nl/pub/93016

Visser, V.A., *Leader Affect and Leadership Effectiveness: How leader affective displays influence follower outcomes*, Promotor: Prof. D.L. van Knippenberg, EPS-2013-286-ORG, http://repub.eur.nl/pub/40076

Vlam, A.J., *Customer First? The Relationship between Advisors and Consumers of Financial Products*, Promotor: Prof. Ph.H.B.F. Franses, EPS-2011-250-MKT, http://repub.eur.nl/pub/30585

Vries, J. de, *Behavioral Operations in Logistics,* Promotors: Prof. M.B.M de Koster & Prof. D.A. Stam, EPS-2015-374-LIS, http://repub.eur.nl/pub/79705

Wagenaar, J.C., *Practice Oriented Algorithmic Disruption Management in Passenger Railways*, Prof. L.G. Kroon & Prof. A.P.M. Wagelmans, EPS-2016-390-LIS, http://repub.eur.nl/pub/93177

Waltman, L., *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality*, Promotors: Prof. R. Dekker & Prof. U. Kaymak, EPS-2011-248-LIS, http://repub.eur.nl/pub/26564

Wang, P., *Innovations, status, and networks*, Promotors: Prof. J.J.P. Jansen & Dr V.J.A. van de Vrande, EPS-2016-381-S&E, http://repub.eur.nl/pub/93176

Wang, T., *Essays in Banking and Corporate Finance*, Promotors: Prof. L. Norden & Prof. P.G.J. Roosenboom, EPS-2015-352-F&A, http://repub.eur.nl/pub/78301

Wang, Y., *Information Content of Mutual Fund Portfolio Disclosure*, Promotor: Prof. M.J.C.M. Verbeek, EPS-2011-242-F&A, http://repub.eur.nl/pub/26066

Wang, Y., *Corporate Reputation Management: Reaching Out to Financial Stakeholders*, Promotor: Prof. C.B.M. van Riel, EPS-2013-271-ORG, http://repub.eur.nl/pub/38675

Weenen, T.C., *On the Origin and Development of the Medical Nutrition Industry*, Promotors: Prof. H.R. Commandeur & Prof. H.J.H.M. Claassen, EPS-2014-309-S&E, http://repub.eur.nl/pub/51134

Wolfswinkel, M., *Corporate Governance, Firm Risk and Shareholder Value*, Promotor: Prof. A. de Jong, EPS-2013-277-F&A, http://repub.eur.nl/pub/39127

Yang, S., *Information Aggregation Efficiency of Prediction Markets*, Promotor: Prof. H.W.G.M. van Heck, EPS-2014-323-LIS, http://repub.eur.nl/pub/77184

Yuferova, D. *Price Discovery, Liquidity Provision, and Low-Latency Trading*, Promotors: Prof. M.A. van Dijk & Dr D.G.J. Bongaerts, EPS-2016-379-F&A, http://repub.eur.nl/pub/93017

Zaerpour, N., *Efficient Management of Compact Storage Systems*, Promotor: Prof. M.B.M. de Koster, EPS-2013-276-LIS, http://repub.eur.nl/pub/38766

Zhang, D., *Essays in Executive Compensation*, Promotor: Prof. I. Dittmann, EPS-2012-261-F&A, http://repub.eur.nl/pub/32344

Zwan, P.W. van der, *The Entrepreneurial Process: An International Analysis of Entry and Exit*, Promotors: Prof. A.R. Thurik & Prof. P.J.F. Groenen, EPS-2011-234-ORG, http://repub.eur.nl/pub/23422

Propositions

I. With few exceptions, traditional candidate gene studies of complex behavioral traits did not yield any real insights into the genetics of behavioral phenotypes because of small sample sizes leading to low statistical power, and an inability to address methodological challenges such as population stratification and undisclosed multiple-hypothesis testing (Chapter 2).

II. Even for behavioral phenotypes that are mostly environmentally determined, a well-powered GWAS can successfully identify replicable genetic associations (Chapters 3 and 4).

III. Because educational attainment is measured in large numbers of individuals, it will continue to be useful as a proxy phenotype in efforts to characterize the genetic influences of related phenotypes, including cognition and neuropsychiatric disease (Chapter 4).

IV. The number of credibly established genetic associations identified via genome-wide association studies, and the predictive power of polygenic scores derived from the findings of such studies, is rapidly increasing and will continue to increase in the years ahead (Chapters 3 and 4).

V. Polygenic scores and quasi-experimental methods for causal inference will usher in an era of more credible research on gene-by-environment interactions (Chapter 5).

VI. The tradeoff between sample size and cross-cohort heterogeneity is quantifiable and should therefore be accounted for in the design phase of studies (Chapter 6).

VII. Bad quality control is detrimental to statistical power in genome-wide association studies.

VIII. Science should not be conceived as a race to be the first on the finish line, but as a collaborative effort to advance the research frontier.

IX. Any entity capable of intelligently designing something as improbable as a Dutchman's Pipe would have to be even more improbable than a Dutchman's Pipe. (*R. Dawkins*)

X. If you trust in yourself… and believe in your dreams… and follow your star… you'll still get beaten by people who spent their time working hard and learning things and weren't so lazy. (*T. Pratchett*)

XI. Somewhere, something incredible is waiting to be known. (*C. Sagan*)

Evidence from behavior-genetic studies of twins, adoptees and other pairs of relatives shows that virtually all human traits, including economic preferences and behaviors, are at least moderately heritable. With increasing availability of genetic data, it is now becoming feasible to identify specific genetic variants associated with complex outcomes, thus blurring the disciplinary boundaries between biological and social sciences. Building on these developments, this thesis explores questions at the intersection of economics and biology and thus contributes to an emerging field of research commonly referred to as: genoeconomics, social-science genetics, or biological economics.

The research described here identifies specific genetic variants robustly associated with a suite of complex outcomes – ranging from educational attainment, to subjective well-being, neuroticism and depression – and shows how these findings can be used, in conjunction with quasi-experimental research designs from economics, to conduct rigorous and well-powered investigations of the interactions between genes and environment. It illustrates some hard-won lessons about the relative merits of various research strategies that have been proposed for efforts to discover genetic associations with complex traits. Furthermore, it provides a framework for quantifying tradeoffs between outcome heterogeneity and sample-size in gene-discovery efforts.

The results reported in this thesis strongly suggest that there will be many settings in which genetic data will prove valuable to social sciences. In addition to broadening our knowledge about the biological mechanisms underlying behavioral outcomes, these advances will provide researchers from many disciplines with increasingly powerful tools that can be used to integrate genetic factors into a wide range of empirical models.

ERiM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are the Rotterdam School of Management (RSM), and the Erasmus School of Economics (ESE). ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focused on the management of the firm in its environment, its intra- and interfirm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced doctoral programme in Research in Management. Within ERIM, over three hundred senior researchers and PhD candidates are active in the different research programmes. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.

ERiM

ERIM PhD Series
Research in Management

Erasmus University Rotterdam (EUR)
Erasmus Research Institute of Management
Mandeville (T) Building
Burgemeester Oudlaan 50
3062 PA Rotterdam, The Netherlands

P.O. Box 1738
3000 DR Rotterdam, The Netherlands
T +31 10 408 1182
E info@erim.eur.nl
W www.erim.eur.nl