

# Psychological Science

<http://pss.sagepub.com/>

---

## Replicability and Robustness of Genome-Wide-Association Studies for Behavioral Traits

Cornelius A. Rietveld, Dalton Conley, Nicholas Eriksson, Tõnu Esko, Sarah E. Medland, Anna A. E. Vinkhuyzen, Jian Yang, Jason D. Boardman, Christopher F. Chabris, Christopher T. Dawes, Benjamin W. Domingue, David A. Hinds, Magnus Johannesson, Amy K. Kiefer, David Laibson, Patrik K. E. Magnusson, Joanna L. Mountain, Sven Oskarsson, Olga Rostapshova, Alexander Teumer, Joyce Y. Tung, Peter M. Visscher, Daniel J. Benjamin, David Cesarini, Philipp D. Koellinger and the Social Science Genetics Association Consortium

*Psychological Science* published online 6 October 2014

DOI: 10.1177/0956797614545132

The online version of this article can be found at:

<http://pss.sagepub.com/content/early/2014/10/06/0956797614545132>

---

Published by:



<http://www.sagepublications.com>

On behalf of:



Association for Psychological Science

**Additional services and information for *Psychological Science* can be found at:**

**Email Alerts:** <http://pss.sagepub.com/cgi/alerts>

**Subscriptions:** <http://pss.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

>> [OnlineFirst Version of Record](#) - Oct 6, 2014

[What is This?](#)

# Replicability and Robustness of Genome-Wide-Association Studies for Behavioral Traits

Psychological Science  
1–12

© The Author(s) 2014

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0956797614545132

pss.sagepub.com



**Cornelius A. Rietveld<sup>1,2</sup>, Dalton Conley<sup>3</sup>, Nicholas Eriksson<sup>4</sup>, Tõnu Esko<sup>5</sup>, Sarah E. Medland<sup>6</sup>, Anna A. E. Vinkhuyzen<sup>7</sup>, Jian Yang<sup>7</sup>, Jason D. Boardman<sup>8,9</sup>, Christopher F. Chabris<sup>10</sup>, Christopher T. Dawes<sup>11</sup>, Benjamin W. Domingue<sup>8</sup>, David A. Hinds<sup>4</sup>, Magnus Johannesson<sup>12</sup>, Amy K. Kiefer<sup>4</sup>, David Laibson<sup>13</sup>, Patrik K. E. Magnusson<sup>14</sup>, Joanna L. Mountain<sup>4</sup>, Sven Oskarsson<sup>15</sup>, Olga Rostapshova<sup>13</sup>, Alexander Teumer<sup>16</sup>, Joyce Y. Tung<sup>4</sup>, Peter M. Visscher<sup>7,17</sup>, Daniel J. Benjamin<sup>18</sup>, David Cesarini<sup>19,20</sup>, Philipp D. Koellinger<sup>1,2,21</sup>, and the Social Science Genetics Association Consortium**

<sup>1</sup>Department of Applied Economics, Erasmus School of Economics, Erasmus University Rotterdam; <sup>2</sup>Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; <sup>3</sup>Department of Sociology, New York University; <sup>4</sup>23andMe, Inc., Mountain View, California; <sup>5</sup>Estonian Genome Center, University of Tartu; <sup>6</sup>Queensland Institute of Medical Research Berghofer Medical Research Institute, Brisbane, Queensland, Australia; <sup>7</sup>Queensland Brain Institute, The University of Queensland, Brisbane; <sup>8</sup>Institute of Behavioral Science, University of Colorado, Boulder; <sup>9</sup>Department of Sociology, University of Colorado, Denver; <sup>10</sup>Department of Psychology, Union College; <sup>11</sup>Wilf Family Department of Politics, New York University; <sup>12</sup>Department of Economics, Stockholm School of Economics; <sup>13</sup>Department of Economics, Harvard University; <sup>14</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet; <sup>15</sup>Department of Government, Uppsala University; <sup>16</sup>Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, Greifswald Medical School; <sup>17</sup>University of Queensland Diamantina Institute, Princess Alexandra Hospital, The University of Queensland, Brisbane; <sup>18</sup>Department of Economics, Cornell University; <sup>19</sup>Center for Experimental Social Science, Department of Economics, New York University; <sup>20</sup>Institute for the Interdisciplinary Study of Decision Making, New York University; and <sup>21</sup>Faculty of Economics and Business, University of Amsterdam

## Abstract

A recent genome-wide-association study of educational attainment identified three single-nucleotide polymorphisms (SNPs) whose associations, despite their small effect sizes (each  $R^2 \approx 0.02\%$ ), reached genome-wide significance ( $p < 5 \times 10^{-8}$ ) in a large discovery sample and were replicated in an independent sample ( $p < .05$ ). The study also reported associations between educational attainment and indices of SNPs called “polygenic scores.” In three studies, we evaluated the robustness of these findings. Study 1 showed that the associations with all three SNPs were replicated in another large ( $N = 34,428$ ) independent sample. We also found that the scores remained predictive ( $R^2 \approx 2\%$ ) in regressions with stringent controls for stratification (Study 2) and in new within-family analyses (Study 3). Our results show that large and therefore well-powered genome-wide-association studies can identify replicable genetic associations with behavioral traits. The small effect sizes of individual SNPs are likely to be a major contributing factor explaining the striking contrast between our results and the disappointing replication record of most candidate-gene studies.

## Keywords

behavior genetics, educational attainment, individual differences, genome-wide association study, population stratification

Received 11/12/13; Revision accepted 6/22/14

The discovery of genetic variants associated with behavioral traits could eventually be transformative for the

## Corresponding Author:

Philipp D. Koellinger, Faculty of Economics and Business, University of Amsterdam, Office M1.20, Plantage Muidergracht 12, 1018 TV Amsterdam, The Netherlands  
E-mail: p.d.koellinger@uva.nl

social sciences, but the first step is identifying the specific genes associated with traits. In psychology, the standard approach for identifying such associations is the candidate-gene study. In a candidate-gene study, a small set of genetic variants (polymorphisms) are selected on the basis of their hypothesized or known biological function, and these polymorphisms are tested for association with a given trait. Most candidate-gene studies have been based on samples of several hundred participants and have applied a significance threshold of .05 (for a review, see Ebstein, Israel, Chew, Zhong, & Knafo, 2010).

Despite the fact that such studies continue to be published in prominent journals, the successful replication of published genetic associations with behavioral traits is the exception, not the rule (Benjamin et al., 2012; Hewitt, 2012). In fact, the situation is so alarming that the editor of the leading field journal *Behavior Genetics* recently issued an editorial policy on candidate-gene studies of behavioral traits that began, “The literature on candidate gene associations is full of reports that have not stood up to rigorous replication,” and went on to say, “it now seems likely that many of the published findings of the last decade are wrong or misleading and have not contributed to real advances in knowledge” (Hewitt, 2012, pp. 1–2). *Psychological Science* has adopted the same strict standards for evaluating candidate-gene studies.

Why the replication of findings from candidate-gene studies of complex behaviors has been inconsistent remains an open question, but it is commonly believed that low statistical power is a major contributing factor, and that the problem of low power is further compounded if the reported  $p$  values correct for only a subset of the multiple hypotheses that were tested (Hewitt, 2012; Ioannidis, 2005). Candidate-gene studies also cannot always adequately control for the well-known problem of *population stratification*, in which genotypes covary with unobserved environmental factors (Hamer & Sirota, 2000). For example, individuals with shared genetic ancestry (e.g., individuals from the same ethnic group or from the same ancestral region) may share values, cultural practices, or exposure to other unobserved environmental confounds. Population stratification can give rise to associations driven by the shared environmental factors but spuriously attributed to the shared genotype (Cardon & Palmer, 2003). A finding may be confounded by population stratification even though it is successfully replicated if the population structure that caused a spurious genetic discovery is also present in the replication samples.

Around 2005, as a result of the methodological limitations of candidate-gene studies and the dramatic decline in the cost of genotyping, medical research experienced a paradigm shift, moving away from candidate-gene studies to what are called genome-wide-association (GWA) studies (McCarthy et al., 2008; Pearson & Manolio, 2008;

Visscher, Brown, McCarthy, & Yang, 2012). These are hypothesis-free studies in which researchers test the phenotype of interest for association with all of the (typically millions of) measured single-nucleotide polymorphisms (SNPs). Because of the large number of hypotheses tested, a SNP association is considered established only if it (a) reaches the “genome-wide significance” threshold of  $p < 5 \times 10^{-8}$  and (b) is subsequently successfully replicated in an independent sample at a nominal significance level of .05 (McCarthy et al., 2008).

Advocates of GWA studies argue that they overcome or mitigate many of the limitations of candidate-gene studies. First, the large number of SNPs that are tested for association makes transparent the need to correct for multiple-hypothesis testing, which is achieved by imposing the genome-wide significance threshold of  $p < 5 \times 10^{-8}$  (McCarthy et al., 2008). Moreover, GWA studies, as a practical matter, tend to be based on larger samples (as indeed they must to have any chance of identifying a SNP that reaches genome-wide significance).

Second, Bayes's rule implies that, conditional on observing an association at the genome-wide significance level, the association is likely to be true even if the study had only modest statistical power to detect the association in the first place (see Benjamin et al., 2012, for calculations).

Third, GWA data can be used to mitigate the potential confound of population stratification. In particular, it has become a common practice in GWA studies to (a) estimate the first four principal components of all the genotypes measured by the gene chip (the number 4 having emerged as a convention), (b) drop individuals who are genetic outliers as measured by these principal components, and then (c) include the principal components as control variables in the genetic-association analysis. Intuitively, the principal components capture axes of correlation across the genome that result from common ancestry. The principal components often have a geographic interpretation (Abdellaoui et al., 2013; Price et al., 2009; Price et al., 2006). Controlling for principal components has become standard in GWA studies since Price et al. (2006) showed, through simulation and empirical examples, that doing so can eliminate spurious associations due to population structure. In Section 5 of the Supplemental Material, available online, we illustrate the effectiveness of this method using a simple placebo test. Specifically, we show 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 individuals with different ancestries (Bersaglieri et al., 2004; Campbell et al., 2005). (In contrast, the most common way of addressing population stratification in candidate-gene studies—namely, including controls for self-identified race—does not eliminate the spurious association.)

There are thus many reasons to expect findings from GWA studies to be replicated more consistently than findings from candidate-gene studies. Indeed, in the literature on complex anthropometric and medical traits, GWA findings do in fact have a vastly superior replication record (Visscher et al., 2012). But do positive GWA findings from studies of complex *behavioral* traits similarly identify credible genetic associations that can be replicated consistently? And if the findings can be replicated consistently, is this because what is being observed is a real genetic signal or because population stratification has generated a spurious association in both the discovery sample and the replication sample? If GWA studies do identify credible and replicable genetic associations, then they are a promising response to the nonreplicability problem in gene-discovery research in the social sciences.

Until recently, virtually all GWA studies with positive findings have been studies of anthropometric or medical traits. For this reason, it may be inappropriate to infer from the superior replication record of GWA studies of medical traits that positive findings from GWA studies of behavioral traits will be replicated consistently. If true genetic associations with behavioral traits have smaller effect sizes than true associations with anthropometric and medical traits, then GWA studies on behavioral traits should tend to generate less reliable results because they have lower power to detect true associations. Furthermore, although the convention of controlling for four principal components may be sufficient to minimize population-stratification concerns for anthropometric and medical traits, it might not be sufficient to minimize these concerns for behavioral traits, which may be characterized by more subtle population stratification.

Earlier GWA studies of behavioral traits (Benyamin et al., 2014; De Moor et al., 2012) largely came up empty-handed (probably as a result of a lack of power), but a recent GWA study on educational attainment with a combined sample of more than 100,000 individuals (Rietveld et al., 2013) identified three SNP associations that met the standard criteria for establishing a GWA-study association (i.e., they reached the genome-wide significance threshold of  $p < 5 \times 10^{-8}$  and were replicated in an independent sample at the significance level of .05). The effect sizes of the associations identified by Rietveld et al. are indeed small: The largest effect size corresponds to an  $R^2$  of only approximately 0.02% (equivalent to about 1 month of schooling per allele). This is far smaller than the effect sizes for medical and anthropometric traits; for example, it is less than one tenth the  $R^2$  of the largest associations discovered for height ( $R^2 = 0.4\%$ ; Lango Allen et al., 2010) and body mass index ( $R^2 = 0.3\%$ ; Speliotes et al., 2010). The Rietveld et al. results therefore can serve as a test case for the robustness of the GWA approach to behavioral traits.

We sought to investigate (a) whether the Rietveld et al. (2013) results could be replicated in an independent sample with far more stringent controls for population stratification than are typically applied in GWA studies of medical and anthropometric traits, and (b) whether there is any evidence overall that the meta-analytic results are contaminated by unaccounted-for population stratification.

## Study 1: Replication of Associations With Specific SNPs in 23andMe Data With Extensive Controls for Stratification

### Method

In Study 1, we sought to replicate the three genome-wide significant SNP associations identified by Rietveld et al. (2013) in a new independent sample. Rietveld et al. tested approximately  $J = 2$  million SNPs for association with educational attainment by running the following regression separately for each SNP  $j \in \{1, 2, \dots, J\}$ :

$$y_i = \mu_j + \beta_j x_{ij} + \gamma_j Z_i + \varepsilon_{ij}, \quad (1)$$

where  $y_i$  is the dependent variable (the phenotype);  $\mu_j$  is a constant term;  $x_{ij}$  is the number of reference alleles (0, 1, or 2) that individual  $i$  is endowed with at SNP  $j$ ;  $\beta_j$  is the coefficient of interest; and  $Z_i$  is a vector of controls, which include age, sex, and the first four principal components of the variance-covariance matrix of the genotypic data. Rietveld et al. studied two dependent variables: EduYears, a measure of the number of years of schooling completed by the individual, and College, a binary variable equal to 1 if the individual had earned a college degree or its equivalent. (The point-biserial correlation between the two measures is roughly .8; see Section 1 in the Supplemental Material.) The tests of association with years of schooling (EduYears) were conducted by running the linear regressions described above, and the tests of association with receipt of a college degree or its equivalent (College) were conducted analogously using logistic regressions.

We sought to replicate the original associations using data provided by 23andMe for a cohort based on a sample of volunteer participants (Eriksson et al., 2010) that was not included in the Rietveld et al. (2013) study. After quality-control filters were applied and the sample was restricted to include only individuals of European descent who responded to a survey question about educational attainment, the sample size was 34,428. Because of the small effects, replication samples of this magnitude are required for adequate power. Given the sample size of 34,428, our power to replicate an association with  $R^2$  of 0.02% at  $p < .05$  was 75% (see Section 6 in the Supplemental Material).

**Table 1.** Replication of Associations Between Single-Nucleotide Polymorphisms and Receipt of College Degree or Equivalent Reported in Rietveld et al. (2013) in 23andMe Data

SNP	23andMe ( $N = 34,428$ )		Rietveld et al. (2013) discovery sample ( $N_s = 95,407$ – $95,419$ )		Rietveld et al. (2013) replication sample ( $N_s = 23,663$ – $23,668$ )	
	OR	$p$	OR	$p$	OR	$p$
rs9320913 or rs12206087	1.035 (0.018)	.046	1.049 (0.011)	$1.06 \times 10^{-5}$	1.042 (0.021)	.022
<b>rs11584700</b>	<b>0.954 (0.020)</b>	<b>.025</b>	<b>0.924 (0.012)</b>	<b><math>2.07 \times 10^{-9}</math></b>	<b>0.923 (0.022)</b>	<b><math>4.86 \times 10^{-4}</math></b>
<b>rs4851266</b>	<b>1.071 (0.019)</b>	<b>.0001</b>	<b>1.069 (0.012)</b>	<b><math>2.20 \times 10^{-9}</math></b>	<b>1.058 (0.022)</b>	<b>.003</b>

Note: Odds ratios (ORs) represent the relative likelihood of attending college for an individual with one more reference allele and were estimated using logistic regression (see discussion after Equation 1 in the main text). Standard errors are shown in parentheses. The Rietveld et al. sample sizes are shown as ranges because different sample sizes are available for the three single-nucleotide polymorphisms (SNPs). Because rs9320913 was unavailable in the 23andMe data, we used rs12206087 as a (very reliable) proxy ( $R^2 = .99$ ; see Section 2.3 in the Supplemental Material). Rows in boldface indicate the SNPs found by Rietveld et al. to be associated with the receipt of a college degree or its equivalent in their discovery sample.

We used the same regression models (see Regression 1) as in Rietveld et al., except that in our analysis, the vector of controls ( $Z$ ) included (in addition to age and sex) the first 25 principal components from the sample's genotype-covariance matrix—compared with only four principal components in Rietveld et al. (2013)—in order to reduce potential population-stratification confounding by partialing out more of the population structure. (For more details on methods, see Section 2 in the Supplemental Material.)

## Results

As shown in Tables 1 and 2, all three SNP associations reported in Rietveld et al. (2013) were replicated at a nominal significance level of .05, in the same direction and with similar effect sizes as in the original report. The replication of effect sizes suggests that the additional controls for population stratification from including more than four principal components made little difference (for related evidence, see Section 5 in the Supplemental Material). As a caveat, we note that because all research

participants were completely anonymous to us, we cannot rule out overlap between the 23andMe sample and the Rietveld et al. discovery or replication samples; in the case of such overlap, the new results would not be fully independent from the results of Rietveld et al. We believe, however, that such potential overlap is likely to be miniscule and is therefore unlikely to have driven our replication findings.

## Study 2: The Robustness of Polygenic Scores in the Swedish Twin Registry and the Queensland Institute of Medical Research Data With Two Distinct Methods of Controlling for Stratification

### Method

Whereas in Study 1, we used a new data set to replicate the three genome-wide significant SNP associations reported by Rietveld et al. (2013), in Study 2, we used some of the same data used in the original report to probe

**Table 2.** Replication of Associations Between Single-Nucleotide Polymorphisms and Years of Schooling Reported in Rietveld et al. (2013) in 23andMe Data

SNP	23andMe ( $N = 34,428$ )		Rietveld et al. (2013) discovery sample ( $N_s = 101,048$ – $101,061$ )		Rietveld et al. (2013) replication sample ( $N_s = 23,523$ – $23,573$ )	
	$\beta$	$p$	$\beta$	$p$	$\beta$	$p$
<b>rs9320913 or rs12206087</b>	<b>0.058 (0.020)</b>	<b>.004</b>	<b>0.106 (0.018)</b>	<b><math>4.19 \times 10^{-9}</math></b>	<b>0.077 (0.034)</b>	<b>.012</b>
rs11584700	–0.053 (0.025)	.032	–0.086 (0.021)	$2.98 \times 10^{-5}$	–0.126 (0.041)	$9.61 \times 10^{-4}$
rs4851266	0.086 (0.021)	$4.28 \times 10^{-5}$	0.076 (0.018)	$3.64 \times 10^{-5}$	0.103 (0.035)	.002

Note: Beta values represent the effect of an increase in one reference allele on years of schooling estimated by the linear regression model (see Equation 1 in the main text). Standard errors are shown in parentheses. The Rietveld et al. sample sizes are shown as ranges because different sample sizes were available for the three single-nucleotide polymorphisms (SNPs). Because rs9320913 was unavailable in the 23andMe data, we used rs12206087 as a (very reliable) proxy ( $R^2 = .99$ ; see Section 2.3 in the Supplemental Material). The row in boldface indicates the SNP found by Rietveld et al. to be associated with years of schooling in their discovery sample.

the robustness of Rietveld et al.'s reported polygenic-score results to potential confounding from population stratification. Following Purcell et al. (2009), polygenic scores are commonly constructed in the GWA literature in order to allow investigators to evaluate the joint predictive power of a large number of SNPs (possibly including SNPs whose effects are too small or estimated too imprecisely to reach genome-wide significance).

Following a common approach in the genetics literature (Purcell et al., 2009; Yang et al., 2012), Rietveld et al. (2013) constructed a polygenic score ( $\hat{g}_i$ ) for each individual ( $i$ ) as equal to a weighted sum of the number of reference alleles (0, 1, or 2) across a set of SNPs, where the weights were derived from the regression coefficients from a GWA study of either years of schooling (EduYears) or receipt of a college degree or its equivalent (College). They then evaluated the predictive power of an individual's score ( $\hat{g}_i$ ) for the individual's educational attainment using two hold-out samples (i.e., samples excluded from the GWA study used for estimating the weights): the Swedish Twin Registry (STR) sample and the Queensland Institute of Medical Research (QIMR) sample. Although the original data sets were from family-based samples, one member from each family was selected at random to be included in the analyses. In each sample (and for scores constructed using GWA studies of each of years of schooling and receipt of a college degree), Rietveld et al. tested four scores constructed from increasingly large sets of SNPs—the sets of SNPs whose GWA-study associations with educational attainment fell below the respective  $p$ -value thresholds:  $5 \times 10^{-8}$  (i.e., only the genome-wide significant SNPs),  $5 \times 10^{-5}$ ,  $5 \times 10^{-3}$ , and 1 (i.e., all SNPs). For each polygenic score ( $\hat{g}_i$ ), Rietveld et al. examined its predictive power by running the following regression:

$$y_i = \mu + \beta \hat{g}_i + \gamma Z_i + \varepsilon_i, \quad (2)$$

where the dependent variable  $y_i$  is always EduYears (never College);  $\mu$  is a constant term;  $\beta$  is the coefficient of interest; and  $Z_i$  is a vector of controls, which include age, sex, and the Age  $\times$  Sex interaction, but no principal components (although principal components were included as controls in the GWA analyses that generated the weights for constructing each  $\hat{g}_i$ ). Rietveld et al. found that the incremental predictive power of the score (i.e., the increase in  $R^2$  from estimating Regression 2 with the score as an independent variable relative to the  $R^2$  without the score) was larger when more SNPs were included in the score. The score that contained all SNPs, which had the largest incremental predictive power, accounted for approximately 2% of the variance across individuals in educational attainment.

To explore the robustness of the original findings, we reran these prediction analyses using two distinct

methods that controlled more stringently for population stratification. In the first, we estimated the same regression model (Regression 2), except we additionally included in the vector of controls the first 20 principal components as control variables. In the second, we estimated mixed linear models (Kang et al., 2010) in place of the regression models. Conceptually, these models involve two steps: (a) The genome-wide data are used to estimate the degree of genetic similarity between each pair of individuals in the sample, and (b) unlike in a standard regression, in which 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, if two individuals are more recently descended from a common ancestor (as can be very accurately measured by overall genetic similarity)—and thus more likely to be similar on unobserved environmental factors—then these individuals are treated as correlated observations on the relationship between educational attainment and the score. (For more details on methods, see Section 3 in the Supplemental Material.)

## Results

The results are shown in Table 3. The upper panel shows the results from the association analyses with the scores constructed using different  $p$ -value thresholds. We separately report results for the STR and QIMR samples and for the scores constructed from weights estimated using the College and EduYears variables. The middle and lower panels show results from regressions with 20 principal components included as controls and from mixed linear models, respectively. Each coefficient is the estimated effect of a 1-standard-deviation increase in the score.

When all SNPs were used to construct the score, it had the predicted sign in all analyses and accounted for approximately 2% of the variance in educational attainment. In the STR sample (the larger and therefore better-powered cohort), the polygenic score was statistically significant in all scenarios, even when only genome-wide significant SNPs were included. The joint effect of the SNP associations with a significance level of  $p < 5 \times 10^{-8}$  explained approximately 0.1% to 0.2% of variance in years of schooling in the STR sample. Because this polygenic score included three SNPs (when constructed using the College variable) or five SNPs (when constructed using the EduYears variable), the results are roughly consistent with Rietveld et al.'s (2013) estimate that each of the most strongly associated SNPs explains approximately 0.02% of variance in years of schooling. Overall, there was no systematic tendency for the predictive power of the scores to change when additional controls for stratification were included.

**Table 3.** Results From Additional Analyses of Polygenic Scores as a Function of Significance Threshold of Genome-Wide-Association-Study Regression Weights Used to Construct the Scores in Queensland Institute of Medical Research and Swedish Twin Registry Data

Type of analysis, variable, and statistic	Queensland Institute of Medical Research (N = 3,544)			Swedish Twin Registry (N = 6,770)				
	$p < 5 \times 10^{-8}$	$p < 5 \times 10^{-5}$	$p < 5 \times 10^{-3}$	$p \leq 1$ (all SNPs)	$p < 5 \times 10^{-8}$	$p < 5 \times 10^{-5}$	$p < 5 \times 10^{-3}$	$p \leq 1$ (all SNPs)
No principal-component adjustment								
College								
$\beta$	0.0495 (0.0554)	0.1502 (0.0554)	0.3581 (0.0551)	0.5630 (0.0545)	0.1706 (0.0496)	0.1984 (0.0496)	0.3469 (0.0496)	0.5506 (0.0492)
$p$	.3702	.0067	$9.1 \times 10^{-11}$	$1.4 \times 10^{-24}$	$6.1 \times 10^{-4}$	$6.9 \times 10^{-5}$	$3.1 \times 10^{-12}$	$1.2 \times 10^{-28}$
$\Delta R^2$	.0002	.0021	.0118	.0291	.0017	.0023	.0072	.0180
EduYears								
$\beta$	-0.0221 (0.0554)	0.2478 (0.0551)	0.3326 (0.0551)	0.5544 (0.0548)	0.1353 (0.0496)	0.2481 (0.0496)	0.3194 (0.0496)	0.5617 (0.0492)
$p$	.6894	$7.6 \times 10^{-6}$	$1.7 \times 10^{-9}$	$7.1 \times 10^{-24}$	$6.5 \times 10^{-3}$	$6.4 \times 10^{-7}$	$1.3 \times 10^{-10}$	$1.0 \times 10^{-29}$
$\Delta R^2$	.0000	.0056	.0102	.0282	.0011	.0037	.0061	.0188
Controlling for 20 principal components								
College								
$\beta$	0.0502 (0.5445)	0.1617 (0.0554)	0.3567 (0.0551)	0.5508 (0.0548)	0.1640 (0.0496)	0.1968 (0.0496)	0.3407 (0.0496)	0.5445 (0.0492)
$p$	.3671	$3.5 \times 10^{-3}$	$1.1 \times 10^{-10}$	$1.5 \times 10^{-23}$	$9.9 \times 10^{-3}$	$7.8 \times 10^{-5}$	$7.3 \times 10^{-12}$	$4.9 \times 10^{-28}$
$\Delta R^2$	.0002	.0024	.0117	.0278	.0016	.0023	.0069	.0176
EduYears								
$\beta$	-0.0043 (0.0554)	0.2604 (0.0554)	0.3122 (0.0551)	0.5372 (0.0548)	0.1410 (0.0496)	0.2423 (0.0496)	0.3059 (0.0496)	0.5588 (0.0492)
$p$	.9400	$2.6 \times 10^{-6}$	$1.7 \times 10^{-8}$	$1.8 \times 10^{-22}$	$4.6 \times 10^{-3}$	$1.1 \times 10^{-6}$	$7.8 \times 10^{-10}$	$2.0 \times 10^{-29}$
$\Delta R^2$	.0000	.0062	.0090	.0265	.0012	.0035	.0056	.0186
Mixed linear model analysis								
College								
$\beta$	0.0538 (0.0571)	0.1505 (0.0581)	0.3383 (0.0574)	0.4920 (0.0581)	0.1697 (0.0504)	0.1902 (0.0500)	0.3387 (0.0500)	0.5601 (0.0508)
$p$	.3479	.0096	$3.8 \times 10^{-9}$	$2.5 \times 10^{-17}$	$7.5 \times 10^{-4}$	$1.4 \times 10^{-4}$	$1.1 \times 10^{-11}$	$2.4 \times 10^{-28}$
$\Delta R^2$	.0003	.0021	.0105	.0222	.0017	.0022	.0068	.0187
EduYears								
$\beta$	-0.0083 (0.0581)	0.2303 (0.0584)	0.2666 (0.0561)	0.4947 (0.0587)	0.1410 (0.0513)	0.2403 (0.0504)	0.3104 (0.0500)	0.5662 (0.0504)
$p$	.8863	$7.9 \times 10^{-5}$	$1.9 \times 10^{-6}$	$3.7 \times 10^{-17}$	$6.0 \times 10^{-3}$	$1.9 \times 10^{-6}$	$5.1 \times 10^{-10}$	$2.4 \times 10^{-29}$
$\Delta R^2$	.0000	.0049	.0065	.0225	.0012	.0034	.0057	.0191

Note: Beta values reflect the effect of a 1-SD increase in the polygenic score on years of schooling estimated by the linear regression model (see Regression 2 in the main text) or by the mixed-linear-model analysis. Positive beta values indicate that the polygenic score predicts the years-of-schooling (EduYears) variable in the same direction in the replication sample as in the discovery sample. Standard errors are shown in parentheses. All  $p$  values are for two-sided tests.  $\Delta R^2$  is the increase in  $R^2$  (in units of percentage points) from estimating a model that includes the polygenic score as an independent variable relative to estimating a model that excludes it.

### Study 3: Replication and Within-Family Robustness of Polygenic Scores in the Framingham Heart Study

#### Method

The gold standard for ruling out concerns about population stratification is to show that the association holds within families. The original Rietveld et al. (2013) study reported a within-family analysis using the pooled STR and QIMR samples. For this within-family analysis, the linear polygenic score constructed using all SNPs in the GWA study of years of schooling was strongly associated with educational attainment, and the score constructed using a  $p$ -value threshold of  $5 \times 10^{-3}$  was marginally significant. Power was too low to draw conclusions about the scores constructed using  $p$ -value thresholds of  $5 \times 10^{-5}$  and  $5 \times 10^{-8}$  (which contained information from fewer SNPs). The STR and QIMR analyses were based on sample sizes of 2,774 dizygotic twin pairs and 572 full-sibling pairs, respectively.

In Study 3, we used data from an independent sample, that of the Framingham Heart Study (FHS), to attempt to replicate the within-family analyses of the linear polygenic scores. The FHS is an epidemiological study on three generations of individuals in the Massachusetts town of Framingham; data from the FHS were not included in any of Rietveld et al.'s (2013) analyses (see Section 4 in the Supplemental Material). In the FHS sample, there are 395 families, each with two or more full biological siblings. Fewer SNPs are available in the FHS than in the STR and the QIMR (see Section 4.1 in the Supplemental Material). Consequently, we expected the polygenic scores in Study 3 to have lower explanatory power than the analogous scores from Study 2. Our focus here was on examining, within the FHS data set, how the estimated effect of the score would be affected by restricting the analysis to within-family variation.

Our analyses proceeded in three steps. First, we applied quality controls to the data, pruned the SNPs for linkage disequilibrium, and constructed the polygenic score using the meta-analytic results from Rietveld et al. (2013). Second, we identified all biological full siblings. Finally, we tested the score ( $\hat{g}_i$ ) within families by running regressions of the following form:

$$\text{EduYears}_i = \beta \hat{g}_i + \sum_{k=1}^K \gamma_k X_{ik} + \varepsilon_i, \quad (3)$$

where  $i$  indexes individuals,  $k$  indexes families, and  $X_{ik}$  is an indicator variable assigned a value of 1 if individual  $i$  belongs to family  $k$  and 0 otherwise. Including the *family fixed effect* ( $X_{ik}$ ) is equivalent (except for the resulting  $R^2$ )

to running a regression after both years of schooling ( $\text{EduYears}_i$ ) and the score ( $\hat{g}_i$ ) are demeaned at the family level; hence, the analysis used only the within-family variation in years of schooling and the within-family variation in the score. To account for the nonindependence of the error term among siblings, we clustered the standard errors (Liang & Zeger, 1986) at the level of the family. Because we expected to have less power for this analysis than did Rietveld et al. (2013) as a result of the smaller number of individuals and the smaller number of SNPs, we ran these analyses for only two scores, one constructed from all SNPs in the sample and one using a  $p$ -value threshold of  $5 \times 10^{-3}$ . (We did not conduct within-family tests of the individual SNPs because our statistical power would have been less than 7%—too low to draw a meaningful conclusion regardless of how the analysis turned out; see Section 6 in the Supplemental Material.)

#### Results

Each coefficient in Table 4 is the estimated effect of a 1- $SD$  increase in the score. The two columns on the left show the results from the new within-family analyses using the FHS data: Whether or not we controlled for 20 principal components, both polygenic scores constructed from all SNPs and from SNPs whose associations reached a significance level of  $p < 5 \times 10^{-3}$  were positively and significantly associated with educational attainment. In the two columns on the right, we report analyses analogous to those from Study 2 (i.e., excluding the family fixed effects, thus leveraging both between- and within-family variation in the score). In these analyses, both scores were positively associated with educational attainment, and results were again similar with and without the principal-component controls. The score from SNPs reaching  $p < 5 \times 10^{-3}$  was marginally significant, and the score from all SNPs was highly statistically significant.

#### Summary

To summarize, in Study 1, we replicated in an independent sample the associations between educational attainment and Rietveld et al.'s (2013) three genome-wide-significant SNPs, using more stringent controls for population stratification than is typical in the GWA literature; the next two studies showed that polygenic scores were robustly replicated in regressions with controls for population stratification and in within-family analyses.

To facilitate a comparison of the effect sizes across Studies 1 through 3 and Rietveld et al.'s (2013) analyses, we present 95% confidence intervals for the effect on years of schooling ( $\text{EduYears}$ ) in Figure 1. An effect size

**Table 4.** Results From Analyses of Polygenic Scores for Years of Schooling in the Framingham Heart Study Sample

Threshold for inclusion of SNPs in polygenic score and statistic	Within-family variation only		Between- and within-family variation	
No principal-component adjustment	$p < 5 \times 10^{-3}$	$p < 1$ (all SNPs)	$p < 5 \times 10^{-3}$	$p < 1$ (all SNPs)
$\beta$	0.2386 (0.0934)	0.2677 (0.1034)	0.1142 (0.0637)	0.2448 (0.0614)
$p$	.011	.010	.073	$7.44 \times 10^{-5}$
$\Delta R^2$	.0036	.0037	.0031	.0141
Controlling for 20 principal components				
$\beta$	0.2308 (0.0947)	0.2642 (0.1044)	0.1173 (0.0634)	0.2534 (0.0621)
$p$	.015	.012	.065	$5.13 \times 10^{-5}$
$\Delta R^2$	.0033	.0036	.0031	.0140

Note: Beta values represent the effect of a 1-SD increase in the polygenic score on years of schooling estimated by Linear Regression Model 3 (for within-family variation; see Equation 3 in the main text) or by Linear Regression Model 2 (for between- and within-family variation; see Equation 2 in the main text). Positive beta values indicate that the score predicts the years-of-schooling (EduYears) variable in the same direction in the replication sample as in the discovery sample. Standard errors are shown in parentheses. All  $p$  values are for two-tailed tests.  $\Delta R^2$  is the increase in  $R^2$  (in units of percentage points) from a model that includes the polygenic score as an independent variable relative to a model that excludes it. All analyses were conducted with data from 1,256 individuals (395 families with two or more children). SNP = single-nucleotide polymorphism.

of 0.1, for example, is equivalent to approximately 1 month of schooling. For the individual SNPs, the average effect of changing the number of reference alleles by 1 is shown, and for the polygenic score containing all SNPs, the average effect of increasing the score by 1 standard deviation is shown. Figure 1a shows that the effect sizes of the genome-wide significant SNPs are comparable across data sets. The effect sizes of the polygenic scores shown in Figure 1b are also similar across the QIMR and STR data sets and methods of controlling for population stratification. The effect sizes of the polygenic score in the FHS (Study 3) are not comparable to those from the QIMR and the STR; the effect sizes are attenuated in the FHS because the scores were constructed from the smaller number of SNPs available in this sample (see Section 4 in the Supplemental Material). Within the FHS, the effect sizes remained similar across different methods of controlling for population stratification, including the within-family analyses.

Although these results are encouraging, we also note a potential limitation of this study. Our evidence, especially the finding that the polygenic score is significantly associated with educational attainment in within-family analyses, suggests that it is extremely unlikely that the findings of Rietveld et al. (2013) were largely an artifact of stratification. However, biases due to very subtle population stratification may still account for some of the observed relationships between educational attainment and some of the individual SNPs. This possibility cannot be *conclusively* ruled out until large enough family samples (e.g.,  $N = 47,000$  sibling pairs; see Section 6 in the Supplemental Material) are available to enable adequately powered within-family tests of association with individual SNPs. This potential limitation applies to all GWA studies. Our findings suggest, however, that the

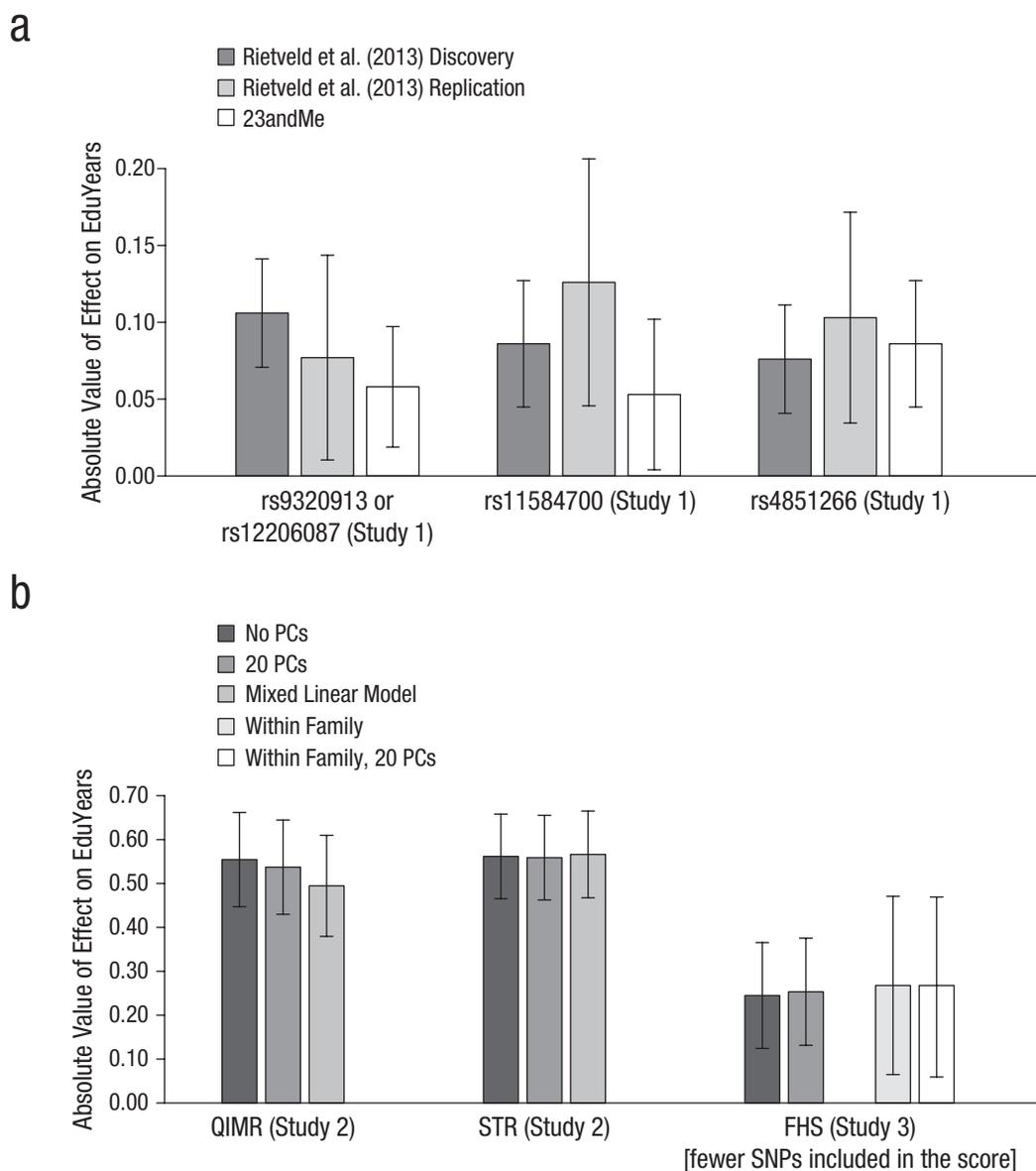
individual SNPs' associations with educational attainment were robust even when we included substantially more stringent controls than are standard in medical genetics.

## Discussion

The contrast between the robustness of our findings and the disappointing replication record of most candidate-gene studies of behavioral traits is striking. To draw the appropriate methodological conclusions, it is necessary to understand the causes of this difference.

A first major contributing factor is that the Rietveld et al. (2013) analyses were based on a sample size that was unprecedentedly large by the standards of social-science genetics. If, as now seems likely, the effects of individual genetic variants on most behavioral traits are small, then much larger samples than are generally used are required to produce credible findings. This is a methodological lesson that applies to all studies, whether they are GWA studies or not. However, in practice, candidate-gene studies tend to be based on much smaller samples. Though it seems clear that much larger samples are needed, it is important to recognize that statistical power also depends on the reliability of the available phenotypic measure. Researchers will sometimes face a trade-off between studying a cruder variable available in a larger sample (e.g., educational attainment) or studying more proximal variables available in a smaller sample (e.g., cognitive ability). Rietveld et al. (2013, Supplementary Online Material, Section 7) provided a framework for quantifying this trade-off.

A second contributing factor is that some of the seeming discipline underlying the hypothesis-based research of existing candidate-gene studies is illusory: Because a vast majority of genes are expressed in the brain (Ramsköld, Wang, Burge, & Sandberg, 2009), it is usually



**Fig. 1.** Absolute value of the effect on years of schooling (EduYears) of a change in one reference allele for each of the three individual single nucleotide polymorphisms (SNPs) as a function of SNP and sample (a) and absolute value of the effect on years of schooling of a 1-SD change in polygenic score (including all SNPs) as a function of study and type of analysis (b). Because rs9320913 was unavailable in the 23andMe data, we used rs12206087 as a (very reliable) proxy ( $R^2 = .99$ ; see Section 2.3 in the Supplemental Material). Error bars show 95% confidence intervals. QIMR = Queensland Institute of Medical Research; STR = Swedish Twin Registry; FHS = Framingham Heart Study; PC = principal component.

possible to create an ex post rationalization for an observed association between a candidate gene and a behavioral trait that sounds at least superficially biologically plausible. Therefore, the main advantage of the candidate-gene approach—namely, the theoretical discipline that it imposes on the investigator—may be exaggerated.

We believe that there are two key implications of our findings for research on genetics of behavioral traits. First, our results suggest that standard GWA-study protocols

from epidemiological research can indeed be successfully applied to the study of behavioral traits and may therefore offer a way to avoid the replication failures that are plaguing much research on the genetics of complex behavior. Second, even if (given the current state of biological knowledge) current candidate-gene approaches are not bearing fruit, this does not rule out an eventual “come-back” for hypothesis-based research on the genetics of behavioral traits. In fact, we envision that as the number

of credibly established associations from GWA studies rises, these discoveries will usher in a new era of “empirical” candidate-gene studies in which the candidates are drawn from among the SNPs identified by GWA studies of related phenotypes. For example, the SNPs associated with educational attainment could be used as candidates to study cognitive and personality traits that may be part of the causal pathway. Such follow-up studies will of course need to be adequately powered to produce robust results, but because the GWA-study results restrict the number of SNPs that are subsequently tested for association, the  $p$ -value threshold can be set much more liberally than the level of genome-wide significance.

What does the finding of small effect sizes—reported by Rietveld et al. (2013) and replicated in our three studies—imply about how genetics research in psychology should be conducted and what its payoffs will be for the field? An immediate implication is that current research using genotypic data in laboratory experiments is almost certainly underpowered, and that psychology should therefore accelerate its move away from such methods, as they are unlikely to yield robust findings. A more subtle implication of the small effect sizes is that—as Turkheimer (2012) has persuasively argued—exuberant forecasts that the availability of genetic data will quickly transform the social sciences should be viewed skeptically. In principle, a genetic variant identified in an association study can explain a tiny part of the variation in the phenotype and yet point to an interesting biological system (this has happened several times in medical genetics). In practice, it seems likely that SNPs with smaller effect sizes, on average, are more likely to operate on the phenotype through distal causal pathways involving a large number and many layers of mediating environmental factors. Therefore, it is conceivable that the identification of SNPs with very small effects will not lead to a useful psychological theory of the phenotype.

At present, the extent to which the identification of individual SNPs will reveal new biological and psychological insights about highly polygenic behavioral traits remains an open question. But we believe it is likely that genetic-association research will benefit psychology in the long run, for at least two other reasons. First, even if genetic associations can be discovered only in samples of many tens of thousands of individuals, once the genetic variants to focus on have been identified, large-but-attainable samples of a few thousand individuals will provide sufficient statistical power to address interesting research questions, such as questions about the nature and magnitude of gene-environment interactions.

Second, even though individual genetic variants have very small effects, polygenic scores can have large

enough effects to be usable even in relatively small samples. The polygenic score explored here has modest explanatory power ( $R^2 \approx 2\%$ ), but when the weights for constructing the score are estimated in larger samples, the explanatory power will be much greater. For example, Rietveld et al. (2013, Supplementary Online Material, Table S26) estimated that a polygenic score constructed using results from a discovery sample of 500,000 individuals will have an  $R^2$  of approximately 12%. We anticipate that such sample sizes will be attainable in the next few years, making it possible to construct such a score. Once a polygenic score with an  $R^2$  of 12% can be calculated for each genotyped participant in a study, a sample of only 62 participants will be needed for 80% power to detect its effect.

In summary, our results suggest that in psychology, a shift away from candidate-gene studies and toward GWA studies is likely to be fruitful. However, before the potential payoffs can be realized, the focus of much research on the genetics of behavioral traits will need to be reoriented, and new research infrastructures will need to be created—for example, to build much larger sample sizes than most GWA studies of behavioral traits have had access to. Nevertheless, we believe that this investment is worth making, because it may lead to the accumulation of reliable and replicable knowledge about the genetics of behavioral traits.

### Author Contributions

All authors contributed to the conception and design of the studies. N. Eriksson, D. A. Hinds, A. K. Kiefer, J. L. Mountain, and J. Y. Tung performed the analyses for Study 1. S. E. Medland, A. A. E. Vinkhuyzen, J. Yang, and P. M. Visscher performed the analyses for Study 2. D. Conley, J. D. Boardman, C. T. Dawes, and B. W. Domingue performed the analyses for Study 3. C. A. Rietveld, D. J. Benjamin, D. Cesarini, and P. D. Koellinger all wrote substantial portions of the article and the Supplemental Material. D. Conley, N. Eriksson, T. Esko, C. F. Chabris, M. Johannesson, D. Laibson, P. K. E. Magnusson, S. Oskarsson, O. Rostapshova, and A. Teumer critically reviewed and edited the manuscript. P. M. Visscher, D. J. Benjamin, D. Cesarini, and P. D. Koellinger contributed equally and should be considered as joint senior authors. The Social Science Genetics Association Consortium provided results of meta-analyses estimated using a subset of the data sets used by Rietveld et al. (2013); see Section 7 of the Supplemental Material, available online, for details.

### Acknowledgments

This research was carried out under the auspices of the Social Science Genetics Association Consortium, a cooperative enterprise among medical researchers and social scientists that coordinates genetic-association studies for social-science variables.

## Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

## Funding

This research was funded primarily by the Ragnar Söderberg Foundation (E9/11) and also by the Swedish Council for Working Life and Social Research (2006-1623) and by the National Institute on Aging and the National Institutes of Health (P01-AG005842, P01-AG005842-2052, P30-AG012810, and T32-AG000186-23). The formation of the Social Science Genetics Association Consortium was made possible by an EAGER grant from the National Science Foundation and a supplemental grant from the National Institutes of Health Office of Behavioral and Social Sciences Research (SES-1064089).

## Supplemental Material

Additional supporting information can be found at <http://pss.sagepub.com/content/by/supplemental-data>

## Open Practices

The meta-analytic data from the genome-wide-association study on educational attainment (Rietveld et al., 2013) are publicly available for follow-up research and replication purposes at <http://ssgac.org/Data.php>. Rietveld et al. (2013) was published before we began the analyses whose results we report in the present article. Further details about materials and methods used in Rietveld et al. (2013) can be found in the Supplementary Material for Rietveld et al. (2013). The complete Open Practices Disclosure can be found at <http://pss.sagepub.com/content/by/supplemental-data>.

## References

- Abdellaoui, A., Hottenga, J.-J., de Knijff, P., Nivard, M. G., Xiao, X., Scheet, P., . . . Boomsma, D. I. (2013). Population structure, migration, and diversifying selection in the Netherlands. *European Journal of Human Genetics*, *21*, 1277–1285. doi:10.1038/ejhg.2013.48
- Benjamin, D. J., Cesarini, D., Chabris, C. F., Glaeser, E. L., Laibson, D. I., Gudnason, V., . . . Lichtenstein, P. (2012). The promises and pitfalls of gene economics. *Annual Review of Economics*, *4*, 627–662. doi:10.1146/annurev-economics-080511-110939
- Benyamin, B., Pourcain, B., Davis, O. S., Davies, G., Hansell, N. K., Brion, M.-J., . . . Haworth, C. M. A. (2014). Childhood intelligence is heritable, highly polygenic and associated with FBNP1L. *Molecular Psychiatry*, *19*, 253–258.
- Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., . . . Hirschhorn, J. N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics*, *74*, 1111–1120. doi:10.1086/421051
- Campbell, C. D., Ogburn, E. L., Lunetta, K. L., Lyon, H. N., Freedman, M. L., Groop, L. C., . . . Hirschhorn, J. N. (2005). Demonstrating stratification in a European American population. *Nature Genetics*, *37*, 868–872. doi:10.1038/ng1607
- Cardon, L. R., & Palmer, L. J. (2003). Population stratification and spurious allelic association. *The Lancet*, *361*, 598–604. doi:10.1016/S0140-6736(03)12520-2
- De Moor, M. H. M., Costa, P. T., Terracciano, A., Krueger, R. F., de Geus, E. J. C., Toshiko, T., . . . Boomsma, D. I. (2012). Meta-analysis of genome-wide association studies for personality. *Molecular Psychiatry*, *17*, 337–349. doi:10.1038/mp.2010.128
- Ebstein, R. P., Israel, S., Chew, S. H., Zhong, S., & Knafo, A. (2010). Genetics of human social behavior. *Neuron*, *65*, 831–844.
- 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. doi:10.1371/journal.pgen.1000993. Retrieved from <http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1000993>
- Hamer, D., & Sirota, L. (2000). Beware the chopsticks gene. *Molecular Psychiatry*, *5*, 11–13. doi:10.1038/sj.mp.4000662
- Hewitt, J. K. (2012). Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behavior Genetics*, *42*, 1–2. doi:10.1007/s10519-011-9504-z
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. Retrieved from <http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.0020124>
- 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*, 348–354. doi:10.1038/ng.548
- Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., . . . O'Connell, J. R. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, *467*, 832–838. doi:10.1038/nature09410
- Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13–22. doi:10.1093/biomet/73.1.13
- McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nature Reviews Genetics*, *9*, 356–369. doi:10.1038/nrg2344
- Pearson, T. A., & Manolio, T. A. (2008). How to interpret a genome-wide association study. *The Journal of the American Medical Association*, *299*, 1335–1344. doi:10.1001/jama.299.11.1335
- 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. Retrieved from <http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.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*, 904–909. doi:10.1038/ng1847
- Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., . . . Moran, J. L. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*, 748–752. doi:10.1038/nature08185
- 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. Retrieved from <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000598>
- 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*, 1467–1471. doi:10.1126/science.1235488
- Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., . . . Loos, R. J. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics*, *42*, 937–948. doi:10.1038/ng.686
- Turkheimer, E. (2012). Genome wide association studies of behavior are social science. In K. S. Plaisance & T. A. C. Reydon (Eds.), *Philosophy of behavioral biology* (pp. 43–64). New York, NY: Springer.
- Visscher, P. M. M., Brown, M. A. A., McCarthy, M. I. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, *90*, 7–24. doi:10.1016/j.ajhg.2011.11.029
- 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*, 369–375. doi:10.1038/ng.2213