bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Educational attainment and personality are genetically intertwined

René Mõttus *
Department of Psychology, University of Edinburgh
Institute of Psychology, University of Tartu, Estonia

Anu Realo
Department of Psychology, University of Warwick
Institute of Psychology, University of Tartu, Estonia

Uku Vainik
Montreal Neurological Institute, McGill University
Institute of Psychology, University of Tartu, Estonia

Jüri Allik
Institute of Psychology, University of Tartu, Estonia
Estonian Academy of Sciences, Tallinn, Estonia

Tõnu Esko
Estonian Genome Centre of University of Tartu, Estonia
Broad Institute, Cambridge, USA

* Corresponding author
7 George Square, EH8 9JZ Edinburgh, UK
Phone: +441316503453
E-mail: rene.mottus@ed.ac.uk

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

It is possible that ubiquitous heritable variance in personality characteristics does not reflect (only) genetic and biological processes specific to personality *per se*. We tested the possibility that Five-Factor Model personality domains, facets and items, as rated by people themselves and their knowledgeable informants, reflect polygenic influences that have been previously associated with educational attainment. In a sample of over 3,000 adult Estonians, polygenic scores for educational attainment, based on small contributions from more than 150,000 genetic variants, were correlated with various personality traits, particularly from the Neuroticism and Openness domains. The correlations of personality traits with phenotypic educational attainment closely mirrored their correlations with educational attainment-related polygenic influences, across facets and individual items. Structural equation modeling of the associations between polygenic risk, personality and educational attainment lent strongest support to the possibility that the same additive genetic influences act independently on both educational attainment and related personality traits.

Keywords: personality; Big Five; education; polygenic; genetic correlation

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Educational attainment and personality are genetically intertwined

Personality trait variance has a substantial genetic component (Vukasović & Bratko, 2015). However, the specific genetic variants responsible for this have largely remained elusive, possibly due to the highly polygenic nature of the traits (Chabris et al., 2013). Collectively, large numbers of common genetic variants explain from nearly zero to about 20% of variance in some personality traits (Power & Pluess, 2015; Smith et al., 2016; van den Berg et al., 2016), but the effect of any one gene is usually too small to be reliably detectable. The same tends to be true for other psychological phenotypes such as intelligence (Davies et al., 2015) or subjective well-being (Okbay, Baselmans, et al., 2016). Slightly more variance has been traced to specific genetic variants for some less-psychological complex phenotypes such as educational attainment (Okbay, Beauchamp, et al., 2016) and body mass index (Locke et al., 2015).

It has also been suggested that personality traits could be conceived of as mostly phenotypic phenomena with limited or even no genetic or biological architecture of their own (Turkheimer, Pettersson, & Horn, 2014). If so, their observed genetic variance may to some or even large extent reflect genetic influences that act broadly across the nervous system or even across the organism more generally (as a general "genetic pull"; Turkheimer et al., 2014) rather than contributing to some systems specifically responsible for what appear as personality traits. In this case, the genetic and resultant biological underpinnings of personality traits should be shared with those of other phenomena that phenotypically relate to these personality traits but fall outside the personality domain (Mõttus, Marioni, & Deary, in press).

Here, we address this possibility by investigating whether phenotypic variability in personality traits is associated with polygenic propensity for educational attainment (henceforth education) as estimated from molecular genetic data. Numerous phenotypes may share genetic influences with personality characteristics. We chose education because it is a broad behavioral phenotype that has a sizable heritable component (Colodro-Conde, Rijsdijk, Tornero-Gómez,

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Sánchez-Romera, & Ordoñana, 2015; Silventoinen, Krueger, Bouchard, Kaprio, & McGue, 2004), is phenotypically correlated with a spectrum of personality traits (Chapman, Fiscella, Kawachi, & Duberstein, 2010; Digman, 1989; Shiner, Masten, & Roberts, 2003) and yet is not part of how the traits are usually operationalized. Also, education has been relatively well characterized in terms of its genomic correlates: its variability is known to be associated with a large number of genetic variants and these specific associations have already been quantified with some level of accuracy (Okbay, Beauchamp, et al., 2016).

Twin studies have revealed that the phenotypic correlations of several personality traits with children's and adolescence academic results can largely be accounted for by shared genetic influences (Hicks, Johnson, Iacono, & McGue, 2008; Rimfeld, Kovas, Dale, & Plomin, 2016). In addition to additive influences of individual genetic variants, these estimates reflect non-additive dominance and epistatis effects due to interactions between and within genetic loci, effects of rare variants and person-environment correlations (Purcell, 2002), and they are possibly confounded with environmental effects that twins share (Vinkhuyzen et al., 2012). Recently, Belsky and colleagues (2016) showed that polygenic variance in education is associated with two childhood personality characteristics, self-control and interpersonal skills; these findings only pertain to the additive genetic effects of common genetic variants included in DNA arrays. Likewise, Okbay and colleagues (2016) showed a negative polygenic correlation between education and Neuroticism. It is not known whether these (additive) polygenic associations generalize to other personality characteristics, including those of the Five-Factor Model (FFM).

Disentangling causality

It seems unlikely that education itself reflects a distinct psychobiological attribute and thereby corresponds to genetic variance that is somehow specific to this phenotype. Instead, its genetic variance is likely to be shared with that of other characteristics such as cognitive and non-cognitive psychological traits. For example, genetic variance in education largely overlaps with that of

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cognitive abilities and (whatever leads to) social deprivation (Marioni et al., 2014). Therefore, when education shows genetic overlap with personality, this may imply different scenarios: they may both be independently influenced by common genetic factors or one may mediate (explain) the genetic effects of the other. In case of mediation, for example, education and thereby its unique genetic influences may phenotypically contribute to certain personality characteristics or the other way around—personality characteristics and thereby their unique genetic influences may explain some of the genetic variance in education (Rimfeld et al., 2016). This means that finding genetic correlations with education would not *inevitably* tell us something on the relative lack of the distinctive etiology of these personality traits. The traits may have their own genetic underpinnings that just happen to bleed into educational level via phenotypic causation.

In order to tackle causality, we can study people with no "exposure" to the hypothesized mediator (Kippersluis & Rietveld, 2016). If the otherwise present genetic correlation between genetic propensity for education and personality traits is missing in people without formal education (i.e., the mediation pathway is broken), this would support education being the mediator in the genetic associations. To estimate the mediating role of a personality trait, we could hypothetically investigate people without the trait. Alternatively, specific genetic variants with known causal pathways to the hypothesized mediator could be used (Davey Smith, 2010). For example, if some genetic variants have direct causal links with education but are unlikely to be similarly directly causal to a personality trait and yet correlate with the trait, this would support education being phenotypically causal to the personality trait and thereby mediating its genetic influences.

It may be difficult to find people with no exposure to education or, especially, without a specific personality trait, and the genetic variants with clear causal pathways to these phenotypes are currently unknown. Meanwhile, statistical approaches can be used to estimate the plausibility of different causal scenarios. If personality traits and education only share genetic influences with no mediation of one another, the association between the two should be diminished to the extent that

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the common genetic factors are controlled for. Alternatively, we may estimate mediation. For instance, if personality traits appear to account for a relatively larger share of genetic propensity-education associations than education can account for genetic propensity-personality associations, this would be more consistent with personality traits mediating the genetic effects of education than the other way around. If neither appears a relatively stronger mediator of other's genetic effects, this would be consistent with the common genetic factors hypothesis.

The current study

Employing published meta-analytic associations (Okbay, Beauchamp, et al., 2016) between education and single nucleotide polymorphisms (SNP), we created polygenic scores for education (EPS) for 3,061 adult Estonians. We correlated the EPS with the individual FFM domains, facets and items, as well as with a new aggregate personality trait that combined education-related aspects of personality. We tested the extent to which EPS, as a potential common cause, could account for the phenotypic correlation between personality and education. Reflecting alternative causal explanations, we also tested whether either education or its personality correlates accounted for a larger share of other's polygenic influences. The use of both self- and informant-rated personality traits allowed us to generalize the findings across specific assessment methods.

Methods

Sample

The current sample is a subset of the Estonian Biobank cohort (approximately 52,000 individuals), a volunteer-based sample of the Estonian resident adult population (Leitsalu et al., 2014). The participants were recruited randomly by general practitioners (GPs), physicians, or other medical personnel in hospitals or private practices as well as in the recruitment offices of the Estonian Genome Centre of the University of Tartu (EGCUT). Each participant signed an informed consent form, went through a standardized health examination and donated a blood sample for DNA. From among 3,426 individuals for whom both personality and DNA data is available, we

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

selected 3,061 individuals (1,821 women) who were at least 25 years old (mean age 49.54 years, standard deviation 15.49, maximum 91) and had thereby had a chance to complete higher education and obtain a post-graduate degree.

Measures

All but 15 participants completed the Estonian version of the NEO Personality Inventory 3 (NEO PI-3; McCrae & Costa, 2010), which is a slightly modified version of the Revised NEO Personality Inventory. The NEO PI-3 has 240 items that measure 30 personality facets, which are then grouped into the five FFM domains, each including six facets. The items were answered on a five-point scale (0 = *false/strongly disagree* to 4 = *true/strongly agree*). Personality traits of all but 2,904 participants (including the 15 participants with missing self-reports) were also rated by an informant, who was typically spouse/partner, parent/child or friend. For cross-rater correlations, see Mõttus and colleagues (2014).

Education was based on self-reports and quantified on an eight-level scale: without any formal education (N = 6), lower basic (N = 31), basic (N = 207), secondary (N = 550), vocational secondary (N = 956), applied higher (N = 177), higher (N = 967) or post-graduate education (N = 167). The variable was treated as if it was continuous.

Polygenic scores aggregate the small effects of a large number of SNPs on a phenotype: the effect size for each SNP's designated (typically minor) allele, found in an independent sample, is multiplied by the number (0, 1, 2) of the allele for a given individual in the target sample; the sums of these products across all SNPs constitute the individual's score. The EPS were based on a meta-analysis (N = 319,945) that estimated the associations of over 8,000,000 SNPs with the number of years of formal schooling (Okbay, Beauchamp, et al., 2016; the contribution of the Estonian Genome Centre was removed from this meta-analysis for the purpose of the current study). In the current sample, genotyping was completed using different Illumina platforms (CNV370-Duo BeadChip, OmniExpress BeadChips, HumanCoreExome-11 BeadChips and HumanCoreExome-10

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

BeadChips); the genotype data was imputed using the 1000 Genomes Project reference panel. SNPs with a minor allele frequency < 0.01, Hardy Weinberg Equilibrium p-value < .001 and info metric value < .90 were omitted. The genotypes were then linkage disequilibrium-pruned using clumping to obtain SNPs in linkage equilibrium with an $r^2 < 0.25$ within a 250 bp window. Clumping was carried out based on the subsample of 1,377 participants who had been genotyped using HumanCoreExome platforms in such a way that SNPs with lowest p-values in relation to education (in the meta-analysis) were retained as the index SNPs of the clumps. No p-value cutoff was used for retaining SNPs. The EPS were based on from 323,818 to 337,334 alleles (i.e., on over 150,000 SNPs). Ten principal components representing possible population stratification were calculated based on the genotype data: EPS scores were residualized for the scores of these components and the numbers of alleles contributing to EPS. The EPS were calculated using PLINK software (Purcell et al., 2007).

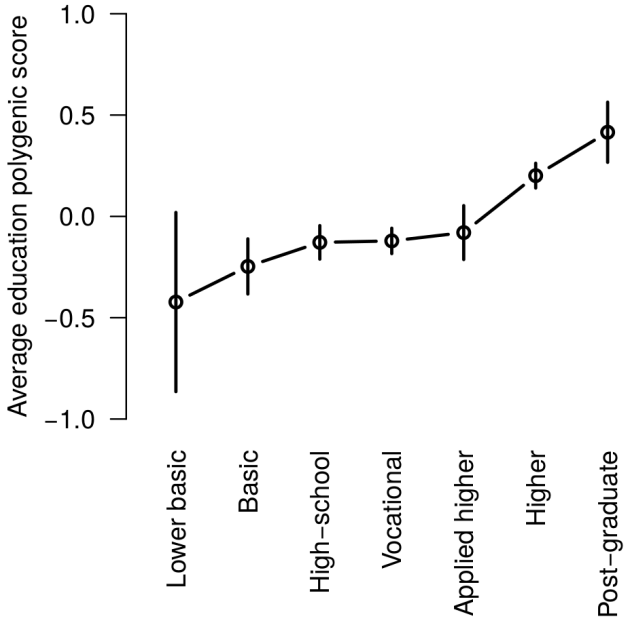
All associations were adjusted for age and sex.

Results

The EPS had a correlation of .18 [95% confidence intervals (CI): .14, .21] with its target phenotype, education. Figure 1 shows that the association was not linear across seven levels of education (the average for the six people with no formal education is not shown in Figure 1), but the average difference between people with lower basic education and a post-graduate degree was substantial (0.84 standard deviation units). Table 1 shows the phenotypic associations of personality traits with both phenotypic education and its polygenic propensity, EPS. The p-values in each column are adjusted for false discovery rate (FDR; Benjamini & Hochberg, 1995) and the associations for which 99% confidence intervals did not span zero are marked in bold.

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1. *Educational levels and average standardized education polygenic scores (EPS; means and 95% confidence intervals).*



Personality and (polygenic propensity for) education

In both self- and informant-ratings, EPS was significantly negatively correlated with the Neuroticism domain and positively correlated with the Openness domain, although the significance did not apply to all of their facets. Specifically, the associations were significant in both rating types for N2: Hostility, O2: Openness to Aesthetics, O4: Openness to Actions, O5: Openness to Ideas and O6: Openness to Values. The associations were also significant in both rating types for the A1: Trust facet of the Agreeableness domain. Some associations were only significant in self-reports; for example, people with higher EPS tended to rate themselves lower on A5: Modesty and A6: Tendermindedness, whereas this was not apparent in informant-ratings.

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. *Associations of personality domains and facets with education and polygenic scores for education.*

	Phenotypic education				Education polygenic scores (EPS)			
	Self-ratings		Informants		Self-ratings		Informants	
	r	p	r	p	r	p	r	p
Neuroticism	**-.16**	<.001	**-.17**	<.001	**-.07**	<.001	**-.05**	.022
Extraversion	**.12**	<.001	**.11**	<.001	.02	.248	.01	.534
Openness	**.25**	<.001	**.19**	<.001	**.11**	<.001	**.09**	<.001
Agreeableness	-.03	.173	.05	.018	-.01	.627	.03	.200
Conscientiousness	**.07**	<.001	**.13**	<.001	-.04	.074	.02	.534
N1: Anxiety	**-.11**	<.001	**-.12**	<.001	-.06	.004	-.03	.256
N2: Hostility	**-.17**	<.001	**-.14**	<.001	**-.08**	<.001	**-.07**	.001
N3: Depression	**-.17**	<.001	**-.15**	<.001	-.07	.001	-.04	.075
N4: Self-Consciousness	**-.09**	<.001	**-.12**	<.001	-.05	.012	-.02	.302
N5: Impulsiveness	-.03	.075	-.09	<.001	-.03	.152	-.04	.079
N6: Vulnerability to Stress	**-.14**	<.001	**-.17**	<.001	-.05	.026	-.03	.232
E1: Warmth	.04	.044	.06	.003	.01	.714	.01	.534
E2: Gregariousness	.06	.003	.04	.023	.00	.843	.00	.899
E3: Assertiveness	**.20**	<.001	**.16**	<.001	.05	.016	.03	.247
E4: Activity	**.12**	<.001	**.14**	<.001	.01	.587	.01	.535
E5: Excitement Seeking	.03	.189	.03	.124	.01	.721	-.01	.761
E6: Positive Emotion	**.07**	<.001	**.06**	.004	.02	.311	.01	.534
O1: Openness to Fantasy	.08	<.001	.04	.052	.04	.029	.03	.199
O2: Openness to Aesthetics	**.17**	<.001	**.13**	<.001	**.06**	.001	**.06**	.007
O3: Openness to Feelings	**.10**	<.001	**.07**	<.001	.03	.216	.00	.899
O4: Openness to Actions	**.23**	<.001	**.18**	<.001	**.09**	<.001	**.07**	.001
O5: Openness to Ideas	**.24**	<.001	**.24**	<.001	**.12**	<.001	**.11**	<.001
O6: Openness to Values	**.20**	<.001	**.09**	<.001	**.11**	<.001	**.06**	.004
A1: Trust	**.22**	<.001	**.16**	<.001	**.09**	<.001	**.06**	.010
A2: Straightforwardness	.01	.610	.07	<.001	.03	.128	.03	.179
A3: Altruism	-.02	.385	.03	.115	.00	.843	.03	.275
A4: Compliance	-.01	.525	.03	.118	.01	.721	.05	.054
A5: Modesty	-.15	<.001	-.04	.024	-.08	<.001	-.02	.432
A6: Tendermindedness	**-.15**	<.001	**-.06**	.001	-.09	<.001	-.02	.534
C1: Competence	**.12**	<.001	**.17**	<.001	.01	.721	.04	.142
C2: Order	.00	.850	.01	.478	-.05	.007	-.03	.190
C3: Dutifulness	**.06**	.003	**.11**	<.001	-.04	.033	.03	.199
C4: Achievement Striving	**.06**	.001	**.15**	<.001	-.02	.248	.00	.880
C5: Self-Discipline	.02	.385	.07	<.001	-.04	.033	.00	.880
C6: Deliberation	**.08**	<.001	**.13**	<.001	.00	.843	.04	.079

NOTE: Informants = informant-ratings. Associations for which 99% confidence intervals did not span zero are marked in bold.

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

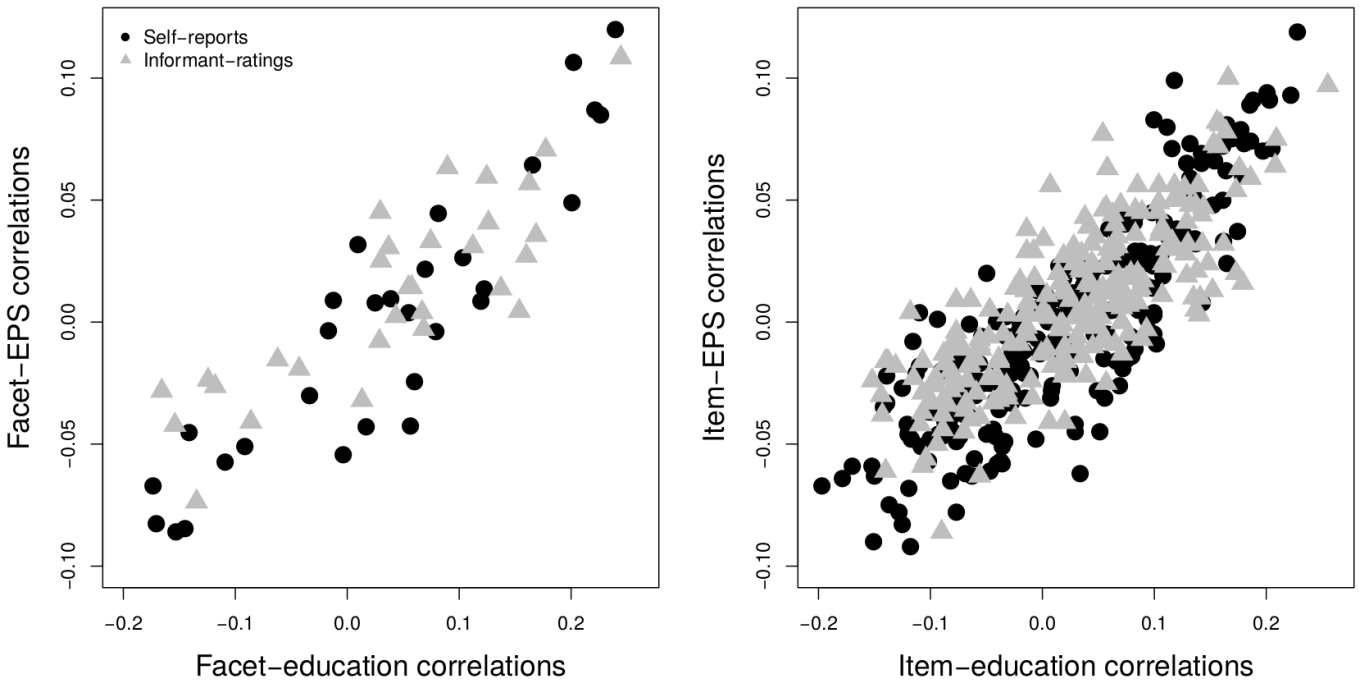
The individual correlations between personality traits and EPS were small in the absolute scale (e.g., .12/.11 and -.08/-.07, respectively for O5: Openness to Ideas and N2: Hostility; self-reports/informant-ratings). However, these effect sizes must be put in the context. First, although the correlation of EPS with the very phenotype it was tailored to was larger (.18), the associations are in the same order of magnitude. Second, studies on the molecular genetic correlates of education have so far been more successful than those on the genetic correlates of personality traits, suggesting that even the correlation .18 is a relatively large effect size in this research context. For example, Okbay and colleagues (2016) reported that meta-analysis-based polygenic scores for subjective well-being, depressive symptoms and Neuroticism explained about 0.9%, 0.5% and 0.7% of variance in their respective traits in independent samples, and similar results were found in another study for Neuroticism (Smith et al., 2016). These estimates translate to correlations below . 10, which is similar to how polygenic scores for a completely different phenotype, education, predicted some personality traits in this study.

Facets' associations with EPS mirrored their associations with phenotypic education. To illustrate this, facet-education and facet-EPS correlations (from Table 1, transformed to z-scores) strongly correlated with each other in both self-reports (r = .91 [CI: .81, .96]) and informant ratings (r = .84 [CI: .69, .92]). As shown in Figure 2, the associations were linear across the spectrum of effect size: neither of these correlations were driven by the few facets that had significant correlations with both education and EPS. These correlations could have been inflated by inter-facet differences in psychometric properties. At the same time, even the *differences between self- and informant-ratings* in facet-level correlations with a) EPS and b) education mirrored each other: the correlation between a) and b) was .80 [CI: .69, .90], suggesting that systematic measurement inaccuracies were not likely to cause the general similarity of the personality-genotype and personality-phenotype associations. This overall pattern suggests that even the weakest (and non-significant) personality trait-education associations could systematically reflect overlapping genetic

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

influences. This in turn suggest that the overlap of education-related genetic influences with those of education-related personality aspects might be stronger than appears from the bi-variate correlations in Table 1. In the next step, we considered this possibility by aggregating the education-related and EPS-related aspects of personality, exactly as polygenic scores aggregate the (small) effects of individual SNPs on a phenotype.

Figure 2. *The associations of facets and items with education and polygenic influences on education (in self- and informant-ratings).*



Associations with aggregated personality

In order to better capture the multi-facet associations between personality and education, we weighed facets by their unique associations with education and then aggregating them into a single composite variable. This resulted in *polyfacet* scores for education. Essentially, this re-casts the idea polygenic scores. In order to calculate the weights for each facet, we used the least absolute shrinkage and selection operator (LASSO) regression (Tibshirani, 2011) with 50-fold cross-validation and a shrinkage parameter lambda that minimized cross-validated error. This method

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

effectively dealt with multi-collinearity among facets. By nature, these polyfacet scores captured as much variance in education as could collectively be predicted by the 30 personality facets, even those that were not significantly correlated with education in the bi-variate analyses described above. The scores could therefore be conceived of as reflecting an education-specific personality trait. We then carried out exactly the same procedure for the EPS, yielding polyfacet scores that were maximally aligned with the polygenic propensity for education. This procedure was done separately for self- and informant-ratings, and all of these polyfacet scores were residualized for age and sex. The correlations between the education polyfacet scores and education itself were .45 [CI: .42, .48] and .39 [CI: .36, .42], respectively for self- and informant-ratings.

The correlations between the education polyfacet scores and EPS were .17 [CI: .14, .21] and .14 [CI: .11, .18], respectively for self- and informant-ratings. This suggests that the association of EPS with the education-related aspects of personality, appropriately aggregated, was nearly of the same magnitude than its correlation with the phenotypic education (.18). Furthermore, the correlations between the polyfacet scores for education and the polyfacet scores for EPS scores were .81 [CI: .79, .82] and .79 [CI: .78, .81], respectively for self- and informant-ratings. These high correlations are consistent with Table 1 and Figure 2, showing that facet-education correlations closely mirrored facet-EPS correlations. Overall, these correlations suggest that although the common genetic influences captured by EPS can only account for a part of personality-education association, the genetic overlap of education and personality is systematic and pervasive across the spectrum of personality traits.

The plausibility of three causal scenarios

In order to assess the plausibility of different causal scenarios, we fitted three structural equation models with the 'lavaan' package (Rosseel, 2012), using maximum likelihood estimator. In the first (common genetic factor) model, both education and the polyfacet scores for education were predicted by EPS. If the genetic influences were directly causal to both phenotypes without any

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mediation of genetic effects between the two, education and the polyfacet scores should become uncorrelated (or locally independent, in latent variable modeling terms), conditional on the genetic influences. However, it is likely that EPS captured only some of the genetic influences on education and thereby also personality. For example, there may be shared non-additive effects or shared effects due to rare variants, which the EPS did not tag, in addition to any common environmental influences. Therefore, we expected that conditioning on EPS would only account for some the correlation between education and personality. We quantified this proportion as the ratio of the indirect pathway between education and personality via EPS to the total association (which was the direct correlation between education and personality, conditional on EPS, plus the indirect effect). Indeed, as shown in the top-left diagram of Figure 3, personality and education were still correlated conditional on EPS in both rating types: EPS could account for about 10% of the total association between them (bottom-right panel of Figure 3). This and the following two models were saturated and therefore fit data perfectly, hence the models could not be compared in terms of fit and no model fit statistics are reported. All reported estimates were significant at $p < .001$.

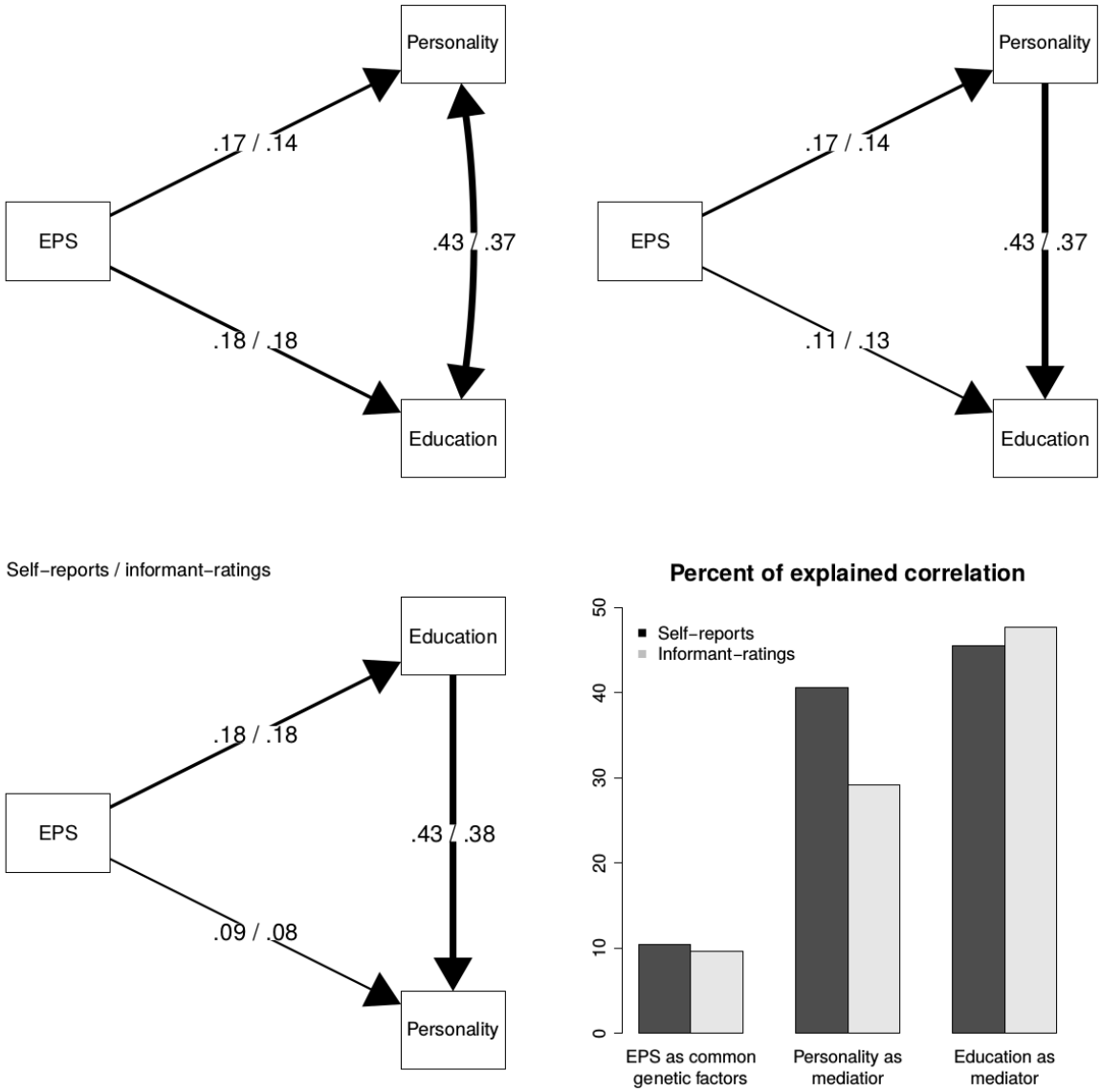
In the second and third model, mediations via personality to education and education to personality were tested, respectively (top-right and bottom-left panels of Figure 3). As reported above, the correlations of EPS with education and its polyfacet scores were very similar in self-ratings (respectively .18 and .17). As a result, the mediation effects inevitably had to be quite similar in both directions. Indeed, properly quantified, personality could mediate 41% of the polygenic effects on education, whereas mediation via education could account for 46% of the polygenic influences on education-related personality traits (bottom-right panel of Figure 2). For informant-ratings, the respective estimates were 29% and 48%: education appeared to mediate somewhat more of the effect of EPS on personality than the other way around. Both indirect pathways were significant at $p < .001$.

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Overall, these results lend little support to personality mediating the genetic effects on education and the evidence for education mediating genetic effects on personality was also inconsistent. As a result, the most plausible interpretation of the current findings is that polygenic propensity, as captured by EPS, is more or less independently related to personality traits and education. Of course, the mediation could work in both directions. However, such a bidirectional mediation amounts to person-environment transactions, whereby pre-existing personality traits predispose to educational experiences, which then reinforce these traits: the results of these transactions would be indistinguishable from additive genetic effects (Purcell, 2002).

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3. *Three causal models. Education polygenic scores (EPS) independently contributing to education and the personality traits related to it, operationalized as polyfacet scores for education (top-left panel). Personality mediating the polygenic influences on education (top-right panel). Education mediating the polygenic influences on personality (bottom-left panel). Bottom-right panel shows the proportions of overall personality-education associations accounted for by education-related polygenic influences, personality-mediated effects of polygenic influences on education and education-mediated polygenic influences on personality.*



bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Item-level analyses

We have previously argued that facet- and domain-level analyses should be supplemented with item-level analyses and where there is evidence for item-specificity in the correlations, the associations should not be generalized to aggregate traits (Mõttus, 2016; Vainik et al., 2015). Supplemental Online Material reports the correlations of single items with EPS and education.

For some facets (e.g., O5: Openness to Ideas), the associations with EPS seemed to generalize across all of their items, whereas for some facets (e.g., O6: Openness to Values and A1: Trust) the correlations appeared to be largely driven by a subset of their items. Occasionally, items of facets that had not been significantly correlated with EPS displayed such associations. For example, although the A2: Straightforwardness facet was not significantly correlated with EPS, its item (A2.4) referring to the belief that honesty is the best policy had a highly significant correlation with polygenic propensity for education (.08 and .07, respectively in self- and informant-ratings). Overall, for 19 of the 240 items, the associations with EPS were significant in both self- and informant-ratings ($p < .05$, adjusted for false discovery rate; Benjamini & Hochberg, 1995) and these associations were always in the same direction. Item-education associations were often in-consistent in size and even direction within the facets.

However, as with facets, even non-significant item-EPS associations might in fact have reflected the genetic overlap. Consistently with this, the (z-transformed) 240 item-EPS correlations were correlated with (z-transformed) 240 item-education correlations .85 [CI: .81, .88] and .77 [CI: .72, .82], respectively in self- and informant-reports. Scatterplots revealed that similarly to facets, the correlations were not driven by items with stronger correlations, but generalized across the spectrum of effect size (Figure 2).

Furthermore, when education-related personality was operationalized based on 240 items instead of 30 facets (using similar procedure to how polyfacet scores had been created), the *polyitem* scores correlated with EPS .19 [CI: .16, .23] and .17 [CI: .14, .21], respectively in self- and

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

informant-ratings. These correlations were somewhat higher than those based on polyfacet scores (.14 and .17; the differences were marginally significant, respectively p = .052 and .025). We re-calculated the associations of EPS, personality and education presented in Figure 3 based on the polyitem scores, yielding mediation results that were generally similar to those from the polyfacet scores-based analyses (Supplemental Online Material).

Overall, these findings suggest that sometimes the polygenic propensity of education was more likely linked with item's unique variance, or personality nuances (Mõttus, Kandler, Bleidorn, Riemann, & McCrae, in press), than with whatever the items of the same facets share, and that sometimes facets were only associated with "education genes" (and thereby education itself) because some of their items were. This points to etiological heterogeneity within facets, which is consistent with findings such as item-specific developmental trends (Mõttus et al., 2015).

Discussion

The findings showed pervasive genetic overlap between education and personality. Polygenic scores tailored to capture additive genetic variance in educational attainment (EPS) correlated with several self- and informant-rated personality traits, especially those belonging to the Openness domain. The magnitude of these correlations was sometimes comparable to how polygenic scores tailored to personality traits correlate with their target traits (Okbay, Baselmans, et al., 2016). When personality traits were aggregated as per their association with education, the correlations of EPS with these education-related personality scores were comparable to its correlation with education itself. Thus, to the extent that the polygenic scores captured genetic variance, they did it similarly in both education and its related personality traits.

One explanation for these findings is that personality traits partially mediate genetic variance in education (Rimfeld et al., 2016). Some traits may predispose people to seek out more schooling and thereby their genetic influences can account for some of the genetic variance in education, alongside any down-stream consequences this important life-outcome may have. Alternatively,

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

experiences related to education may be causal to personality traits and therefore genetic influences on education can account for some of the genetic variance in these traits. For example, certain genetic variants may predispose people to completing more years of schooling (e.g., via faster information processing or better physical health that allows for more school attendance), which in turn may enhance people's interest in aesthetic and intellectual experiences or contribute to disapproval of dishonesty. The third explanation for the genetic overlap between personality and education is that same genetic influences independently act on both. Indeed, there was no clear and consistent statistical evidence for stronger mediation in one or another direction, leaving the common genetic factors hypothesis the most plausible explanation.

The polygenic scores could only explain 10% of the education-personality covariance, but these scores are unlikely to capture full genetic variance in education and thereby also in personality characteristics related to it. For example, the heritability of education has been estimated at more than 20% based on alternative procedures (Marioni et al., 2014), whereas EPS could account for only 3%. Thus, the 10% of explained covariance might be an underestimate. What may be more revealing is how systematically, across facets and items, EPS traced associations with phenotypic education.

The findings are consistent with the possibility that at least some of the genetic variance in education-related personality traits does not reflect distinctive genetic and thereby biological mechanisms for these traits (Turkheimer et al., 2014). Of course, it will ultimately require knowing relevant genetic and biological mechanisms of the phenotypes, or studying groups who have had no exposure to possible mediators, to appropriate disentangle the causal pathways. But if our findings hold, attempts to delineate the specific genetic underpinnings of education (Okbay, Beauchamp, et al., 2016) may incidentally reveal the genetic mechanisms of phenotypically related personality characteristics. Also, education could be used as a proxy to narrow the range of potentially personality-related genetic variants (Rietveld et al., 2014).

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The genetic overlap will also need to be factored into attempts to interpret the phenotypic associations between personality traits and education. Turkheimer and colleagues (2014) argue that when associations of personality traits with other variables are investigated "our scientific hypotheses are usually phenotypic in nature" (p. 533): one phenotype causes the other. To the extent that phenotypic associations reflect genetic overlap, there is no such phenotypic causation. Naturally, the implications of our findings stretch beyond the associations between personality traits and education: genetic overlaps should be considered for *any* phenomenon that is hypothesized to be either causal to personality traits or among their downstream consequences. For example, personality traits are phenotypically associated with obesity (Sutin, Ferrucci, Zonderman, & Terracciano, 2011), but these links may at least to some extent reflect genetic overlaps.

With molecular genetic data becoming widely accessible, researchers will be increasingly interested in using them to decompose phenotypic associations into genetic and non-genetic components. The present study highlights one possible methodology for doing this. Although other techniques that allow for estimating genetic correlations from molecular genetic data are available (Bulik-Sullivan et al., 2015; Yang, Lee, Goddard, & Visscher, 2011), they typically require even large samples that the use of polygenic scores. This approach requires SNP-outcome associations from an independent large sample but, in the era of genome-wide association studies, such information is becoming available for an ever-increasing number of phenotypes.

In sum, the current study systematically examined polygenic overlap between education and personality traits, and found clear evidence for this. The findings are consistent with the possibility that genetic influences on personality may not necessarily pertain to some personality-specific neurobiological structures and that genetic studies on education could also provide useful insights into the genetic underpinnings of personality variability. Finally, personality-outcome associations may not always be phenotypically causal, but there are means for testing this.

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

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

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., … Neale. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. https://doi.org/10.1038/ng.3406

Chabris, C. F., Lee, J. J., Benjamin, D. J., Beauchamp, J. P., Glaeser, E. L., Borst, G., … Laibson, D. I. (2013). Why it is hard to find genes associated with social science traits: Theoretical and empirical considerations. *American Journal of Public Health*, *103*(S1), S152–S166. https://doi.org/10.2105/AJPH.2013.301327

Chapman, B. P., Fiscella, K., Kawachi, I., & Duberstein, P. R. (2010). Personality, socioeconomic status, and all-cause mortality in the United States. *American Journal of Epidemiology*, *171*(1), 83–92. https://doi.org/10.1093/aje/kwp323

Colodro-Conde, L., Rijsdijk, F., Tornero-Gómez, M. J., Sánchez-Romera, J. F., & Ordoñana, J. R. (2015). Equality in educational policy and the heritability of educational attainment. *PLOS ONE*, *10*(11), e0143796. https://doi.org/10.1371/journal.pone.0143796

Davey Smith, G. (2010). Mendelian Randomization for strengthening causal inference in observational studies: Application to gene × environment interactions. *Perspectives on Psychological Science*, *5*(5), 527–545. https://doi.org/10.1177/1745691610383505

Davies, G., Armstrong, N., Bis, J. C., Bressler, J., Chouraki, V., Giddaluru, S., … Deary, I. J. (2015). Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). *Molecular Psychiatry*, *20*(2), 183–192. https://doi.org/10.1038/mp.2014.188

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Digman, J. M. (1989). Five robust trait dimensions: development, stability, and utility. *Journal of Personality*, *57*(2), 195–214.

Hicks, B. M., Johnson, W., Iacono, W. G., & McGue, M. (2008). Moderating effects of personality on the genetic and environmental influences of school grades helps to explain sex differences in scholastic achievement. *European Journal of Personality*, *22*(3), 247–268.

Kippersluis, H. van, & Rietveld, C. A. (2016). Pleiotropy-robust Mendelian Randomization. *bioRxiv*, 72603. https://doi.org/10.1101/072603

Leitsalu, L., Haller, T., Esko, T., Tammesoo, M.-L., Alavere, H., Snieder, H., … Metspalu, A. (2014). Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *International Journal of Epidemiology*, *44*, 1137–1147. https://doi.org/10.1093/ije/dyt268

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

Marioni, R. E., Davies, G., Hayward, C., Liewald, D., Kerr, S. M., Campbell, A., … Deary, I. J. (2014). Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence*, *44*, 26–32. https://doi.org/10.1016/j.intell.2014.02.006

McCrae, R. R., & Costa, P. T. (2010). *NEO Inventories professional manual*. Odessa, FL: Psychological Assessment Resources.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

Mõttus, R. (2016). Towards more rigorous personality trait–outcome research. *European Journal of Personality*, *30*(4), 292–303. https://doi.org/10.1002/per.2041

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (in press). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspp0000100

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Mõttus, R., Marioni, R., & Deary, I. J. (in press). Markers of psychological differences and social and health inequalities: Possible genetic and phenotypic overlaps. *Journal of Personality*. https://doi.org/10.1111/jopy.12220

Mõttus, R., McCrae, R. R., Allik, J., & Realo, A. (2014). Cross-rater agreement on common and specific variance of personality scales and items. *Journal of Research in Personality*, *52*, 47–54. https://doi.org/10.1016/j.jrp.2014.07.005

Mõttus, R., Realo, A., Allik, J., Esko, T., Metspalu, A., & Johnson, W. (2015). Within-trait heterogeneity in age group differences in personality domains and facets: Implications for the development and coherence of personality traits. *PLoS ONE*, *10*(3), e0119667. https://doi.org/10.1371/journal.pone.0119667

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

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

Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research*, *5*(6), 554–571. https://doi.org/10.1375/136905202762342026

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

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

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

proxy-phenotype method. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(38), 13790–13794. https://doi.org/10.1073/pnas.1404623111

Rimfeld, K., Kovas, Y., Dale, P. S., & Plomin, R. (2016). True grit and genetics: Predicting academic achievement from personality. *Journal of Personality and Social Psychology*, No Pagination Specified. https://doi.org/10.1037/pspp0000089

Rosseel, I. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*, 1–36.

Shiner, R. L., Masten, A. S., & Roberts, J. M. (2003). Childhood personality foreshadows adult personality and life outcomes two decades later. *Journal of Personality*, *71*(6), 1145–1170. https://doi.org/10.1111/1467-6494.7106010

Silventoinen, K., Krueger, R. F., Bouchard, T. J., Kaprio, J., & McGue, M. (2004). Heritability of body height and educational attainment in an international context: Comparison of adult twins in Minnesota and Finland. *American Journal of Human Biology*, *16*(5), 544–555. https://doi.org/10.1002/ajhb.20060

Smith, D. J., Escott-Price, V., Davies, G., Bailey, M. E., Colodro Conde, L., Ward, J., … O'Donovan, M. (2016). Genome-wide analysis of over 106,000 individuals identifies 9 neuroticism-associated loci. *bioRxiv*. https://doi.org/10.1101/032417

Sutin, A. R., Ferrucci, L., Zonderman, A. B., & Terracciano, A. (2011). Personality and obesity across the adult life span. *Journal of Personality and Social Psychology*, *101*(3), 579–592. https://doi.org/10.1037/a0024286

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(3), 273–282. https://doi.org/10.1111/j.1467-9868.2011.00771.x

Turkheimer, E., Pettersson, E., & Horn, E. E. (2014). A phenotypic null hypothesis for the genetics of personality. *Annual Review of Psychology*, *65*(1), 515–540. https://doi.org/10.1146/annurev-psych-113011-143752

Vainik, U., Mõttus, R., Allik, J., Esko, T., & Realo, A. (2015). Are trait–outcome associations caused by scales or particular items? Example analysis of personality facets and BMI. *European Journal of Personality*, *29*, 622–634. https://doi.org/10.1002/per.2009

bioRxiv preprint doi: https://doi.org/10.1101/078014; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

van den Berg, S. M., de Moor, M. H. M., Verweij, K. J. H., Krueger, R. F., Luciano, M., Arias Vasquez, A., … Boomsma, D. I. (2016). Meta-analysis of genome-wide association studies for Extraversion: Findings from the Genetics of Personality Consortium. *Behavior Genetics*, *46*, 170–182. https://doi.org/10.1007/s10519-015-9735-5

Vinkhuyzen, A. A. E., Pedersen, N. L., Yang, J., Lee, S. H., Magnusson, P. K. E., Iacono, W. G., … Wray, N. R. (2012). Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. *Translational Psychiatry*, *2*(4), e102. https://doi.org/10.1038/tp.2012.27

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

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011