Genomic analysis of family data reveals additional genetic effects on intelligence and personality

W. David Hill[1,2]*†, Ruben C. Arslan[3,4]†, Charley Xia†[5], Michelle Luciano[1,2], Carmen Amador[5], Pau Navarro[5], Caroline Hayward[5], Reka Nagy[5], David J. Porteous[1,6,8], Andrew M. McIntosh[1,9], Ian J. Deary[1,2], Chris S. Haley[5,10], and Lars Penke[1,3,4]

1 Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK

2 Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK

3 Georg Elias Müller Institute of Psychology, Georg August University Göttingen, Germany

4 Leibniz ScienceCampus Primate Cognition, Göttingen, Germany

5 MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

6 Generation Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

8 Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh

9 Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh EH10 5HF

10 The Roslin Institute and Royal (Dick) School of Veterinary Sciences, University of Edinburgh, UK

* Corresponding author

† These authors contributed equally

Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK, T: +44 (131) 650 8405, E: David.Hill@ed.ac.uk

Abstract

Pedigree-based analyses of intelligence have reported that genetic differences account for 50-80% of the phenotypic variation. For personality traits, these effects are smaller with 34-48% of the variance being explained by genetic differences. However, molecular genetic studies using unrelated individuals typically report a heritability estimate of around 30% for intelligence and between 0% and 15% for personality variables. Pedigree-based estimates and molecular genetic estimates may differ because current genotyping platforms are poor at tagging causal variants, variants with low minor allele frequency, copy number variants, and structural variants. Using ~20 000 individuals in the Generation Scotland family cohort genotyped for ~520 000 single nucleotide polymorphisms (SNPs), we exploit the high levels of linkage disequilibrium (LD) found in members of the same family to quantify the total effect of genetic variants that are not tagged in GWASs of unrelated individuals. In our models, genetic variants in low LD with genotyped SNPs explain over half of the genetic variance in intelligence, education, and neuroticism. By capturing these additional genetic effects our models closely approximate the heritability estimates from twin studies for intelligence and education, but not for neuroticism and extraversion. From an evolutionary genetic perspective, a substantial contribution of genetic variants that are not common within the population to individual differences in intelligence, education, and neuroticism is consistent with mutation-selection balance.

The scores of different types of cognitive ability tests correlate positively and the variance that is shared between tests is termed general intelligence, general cognitive ability, or *g*.[1] General intelligence typically accounts for around 40% of the overall variance among humans in batteries that contain tests of diverse cognitive abilities. The personality traits of extraversion and neuroticism are two of the five higher-order personality factors that are consistently identified in dimensional models of personality. High levels of extraversion are associated with positive affectivity and a tendency to engage with, and to enjoy, social situations, but also shows phenotypic and genetic associations with mental disorders like attention deficit hyperactivity disorder.[2] High levels of neuroticism are associated with stress sensitivity, as well as mental and physical disorders.[2, 3] All of these traits are partly heritable, but have also been linked to evolutionary fitness. This poses an ostensible paradox, that could be resolved if rare variants, which are less amenable to selection, are found to play a major role in the genetic contribution to variance in these traits. Using a recently-developed analytic design for combined pedigree and genome-wide molecular genetic data, we test whether rare genetic variants, copy number variants (CNVs), and structural variants make an additional contribution to the genetic variance in intelligence, neuroticism, and extraversion.

General intelligence has been found to be heritable, with twin and family studies estimating that 50% to 80%[4] of phenotypic variance is due to additive genetic factors, a proportion that increases with age from childhood to adulthood.[5] Heritability can also be estimated from molecular genetic data. Using the genomic-relatedness-matrix restricted maximum likelihood (GREML) method, additive common single nucleotide polymorphisms (SNPs) are estimated to collectively explain between 20% and 50% of variation in general intelligence,[6, 7] with an estimate of around 30% in the largest studies.[8] General intelligence is also a significant predictor of fitness components including mortality,[9] fertility,[10, 11] higher social status,[12] as well as mental and physical disease,[5] and it is associated with developmental stability,[13, 14] suggesting that general intelligence is not selectively neutral.

As selective pressure on a trait is expected to deplete its genetic variation, the existence of such robust heritability findings seems paradoxical when evolutionary theory is considered.[15] However, mutation-selection balance provides an explanation of how genetic variation can be maintained for quantitative traits that are under directional selective pressure. Mutation-selection balance describes instances where mutations that are deleterious to the phenotype occur within a population at the same rate that they are removed through the effects of selective pressure. Due to the removal of variants with deleterious effects on the phenotype, the existence of common variants of medium to large effect is not expected under mutation-selection balance. This is consistent with the current findings from large GWAS on cognitive phenotypes, including general intelligence and education, where common single nucleotide polymorphisms (SNPs) collectively explain a substantial proportion of phenotypic variance, but the individual effect size of each genome-wide significant SNP discovered so far is around 0.2%.[16, 17]

3

Population genetic simulations show that very rare (Minor allele frequency, MAF < 0.1%) variants explain little of the population variance in traits that are not under selection.[18] However, the contribution made by rare variants increases when their effects on a trait and on fitness are correlated either through pleiotropy, or by the trait directly affecting fitness.[18] The genetically informative evidence that is available tends to show that variants associated with intelligence are also linked to better health,[19, 20] although these effects may be outweighed by a negative effect on fertility,[21, 22] and that the regions of the genome making the greatest contribution to intelligence differences have undergone purifying selection.[23] Whereas this does not necessarily imply that intelligence has been selected for or against across our evolutionary history, it does indicate that genetic variants that are associated with intelligence are also associated with fitness, which suggests that rare genetic variants and hence mutation-selection balance, may act to maintain intelligence differences.[18]

Empirical studies so far have failed to find evidence of a link between intelligence and rare variants.[24] These studies have often been limited in scope, with only copy number variants or exonic regions being considered, or being limited in statistical power because all rare variants were treated as having the same direction of effect through the use of burden tests.[24-27] Where such tests have found an association these have been in small samples and subsequently failed to replicate.[28] However, in large samples, rare variants found within regions of the genome under purifying selection have been found to be associated with educational success,[29] an effect that was greater for genes expressed in the brain. Hence, rare variants found in some genes might have an effect on intelligence.

Less is known about the genetics of personality.[30] As with intelligence, extraversion and neuroticism have been found to have higher heritability estimates, of around 34-48%, using quantitative (twin- and family-based) genetic methods[31] compared to molecular genetic estimates (4 – 15% for neuroticism[32] and 0 – 18% for extraversion[2, 33]). Both extraversion and neuroticism are predictive of social and behavioural outcomes as well as anxiety, well-being and fertility.[34-37] Positive genetic correlations have been reported for extraversion with attention deficit hyperactivity disorder and bipolar disorder, and for neuroticism with depression and anorexia nervosa.[2]

In the current study, we quantify the total genetic effect from across the genome on intelligence (including education, which shows strong genetic correlations with general intelligence[38] and is used as a proxy phenotype for it in genetic studies[39]), extraversion, and neuroticism. We are able to include genetic variants not normally captured using GWAS. As our sample included nominally unrelated individuals with varying degrees of genetic similarity, as well as family members who all provided genome-wide SNP data, we were able to decompose two genetic sources of variance corresponding to genetic effects associated with common SNPs in a population (h^2_g), and genetic effects associated with kinship (h^2_{kin}). Among related individuals, linkage disequilibrium is stronger and hence allows us to capture variation not tagged by common SNPs. This includes rare variants, CNVs, and other structural variants. As the inclusion of family members can introduce a confound between shared genetic effects and shared environmental effects,[40] we use the method employed by

4

Xia and colleagues [41] to simultaneously estimate three sources of environmental variance: sibling effects, spouse effects, and family effects. By using information from both nuclear family relationships and the many more distant pedigree relationships in the cohort we analyse, this novel design allows us to estimate kin-specific genetic variation net of common environmental effects.

Materials and Methods

Samples

Data was used from the Generation Scotland: Scottish Family Health Study (GS:SFHS).[42-44] A total of 24 090 individuals (N_{male} = 9927 , N_{female} = 14163 , Age_{mean} = 47.6) were sampled from Glasgow, Tayside, Ayrshire, Arran, and North-East Scotland of whom 23 919 have donated blood or saliva for DNA extraction. These samples were collected, processed, and stored using standard procedures and managed through a laboratory information management system at the Wellcome Trust Clinical Research Facility Genetics Core, Edinburgh.[45] The yield of DNA was measured with a PicoGreen and normalised to 50ng/μl prior to genotyping. Genotype data were generated using an Illumina Human OmniExpressExome -8- v1.0 DNA Analysis BeadChip and Infinium chemistry.[46] We then used an identical quality control procedure as Xia et al.[41] that included removing SNPs in the event that they were not on autosomes or had a minor allele frequency (MAF) of <0.05, a Hardy-Weinberg Equilibrium P-value <10^{-6}, and a missingness of >5%. As per Xia et al.[41] this left 519 729 common SNPs from 22 autosomes. Following quality control, a total of 20 032 individuals (n females = 11 804) were retained; 18 293 of these individuals were a part of 6 578 nuclear or extended families.[47] The mean age of the sample was 47.4 years (SD = 15.0, range 18 to 99 years). As the variance attributable to the shared environment was explicitly modelled here, no relationship cut off (typically, 0.025 is used) was applied to the genetic relationship matrix (GRM).

Ethics

The Tayside Research Ethics Committee (reference 05/S1401/89) provided ethical approval for this study.

Phenotypes

A total of eight phenotypes were examined here. Six of the phenotypes were cognitive in nature and included general intelligence (g), education, the Mill Hill Vocabulary Scale (MHVS),[48] the Wechsler Digit Symbol Substitution Task (DST),[49] Wechsler Logical Memory which measures Verbal declarative memory,[50] and executive function (phonemic Verbal fluency, using letters C, F, L).[51] The general factor of intelligence (g) was derived by extracting the first unrotated principal component from the four cognitive tests. This single component accounted for 42.3% of the variance in the total sample and each of the individual tests used demonstrated strong loadings on the first unrotated component (DST 0.58, Verbal Fluency 0.72, MHVS 0.67, and Verbal declarative memory

0.63). Education was calculated in the GS:SFHS as the years of full time formal education which was recoded into an ordinal scale from 0 to 10 (0: 0 years, 1: 1-4 years, 2: 5-9 years, 3: 10-11 years, 4: 12-13 years, 5: 14-15 years, 6: 16-17 years, 7: 18-19 years, 8: 20-21 years, 9: 22-23 years, 10: > 24 years of education). Education and general intelligence were positively correlated ($r = 0.38$, SE = 0.01, p < 2.20×10^{-16}).

The effects of age, sex and population stratification were controlled for by using regression prior to fitting the models using GREML. Supplementary section Figure 1 shows the number of principal components used to control population stratification for each of the phenotypes used.

Statistical method

Partitioning phenotypic variance into five sources

For each of the phenotypes examined here variance was partitioned into five corresponding effects plus residual variance. This variance components analysis is based on the work of Zaitlen and colleagues[40] who developed a method for estimating h^2_g and h^2_{kin} in a data set with a measured family structure. More recently this method has been extended by Xia and colleagues[41] to include sibling, spouse, and nuclear family environmental effects. The two genetic matrices described by Zaitlan et al. and Xia et al. correspond to those associated with common SNPs (h^2_g) and those associated with pedigree (h^2_{kin}) genetic variants. These two genetic sources of variance were quantified using a genetic relationship matrix derived in the GCTA software.[52] Whereas h^2_g describes variance associated with common SNPs, and those that are in LD with genotyped SNPs on a SNP chip, h^2_{kin} describes variance from the additional genetic effects associated with pedigree.

Matrix construction

Genetic matrices

The contribution made by common SNPs, h^2_g, was quantified using a genomic relationship matrix (GRM_g, or G). This was derived in the manner set out by Yang and colleagues,[52] where the estimated genomic relatedness between each pair of individuals is derived from identity by state SNP relationships and is found in each off diagonal entry in the GRM.

$$\frac{1}{N}\sum_{i=1}^{N}\frac{(x_{ji} - 2p_i)(x_{ki} - 2p_i)}{2p_i(1 - p_i)}$$

6

Minor allele frequency for SNP i is denoted as pi and the allelic dose (x) for individuals j or k at locus i is described as x_{ji} or x_{ki}. N indicates the total number of SNPs.

The kinship relationship matrix, GRM_{kin}, or K, was derived using the method described by Zaitlen et al., (2013)[40] by modifying the GRM_g. Here, values in the GRM_g that were equal to or less than 0.025 were set to 0.

Environmental matrices

Three environmental matrices (ERM) were used to capture the variance associated with specific relationships between individuals. Each ERM was created by deriving an N by N (where N is number of participants) matrix with diagonal entries set to 1and non-diagonal entries set to 1 if the pair of individuals are a part of the environmental relationship described or set to zero otherwise. The three ERMs derived here captured variance associated with the shared environment of spouses, (ERM_{Couple}, or C), siblings ($ERM_{Sibling}$, or S), and nuclear families (ERM_{Family}, or G).

Deriving the quantity of phenotypic variance explained

For each trait we first fitted the two GRM and the three ERM simultaneously using a linear mixed model (LMM) implemented using the GCTA software.[52, 53] This full model is referred to as the GKFSC model, as it includes the genetic, kinship, family, sibling, and couple matrices.

$$Y = GRM_g + GRM_{kin} + ERM_{Family} + ERM_{Sibling} + ERM_{Couple} + \varepsilon$$

Here, Y is a vector of standardised residuals derived using one of the eight phenotypes examined here. Random genetic effects were explained by fitting the GRM_g and the GRM_{kin}, which captured variants in LD with common SNPs found across a population and the extra genetic effects captured by the increase in LD found between members of the same family, respectively. Random environmental effects that were shared between related pairs of individuals were captured by fitting the ERM_{Family}, $ERM_{Sibling}$, and ERM_{Couple} to quantify the contributions made by environmental similarities between members of a nuclear family, siblings, and couples, respectively.

Restricted maximum likelihood (REML), implemented using the GCTA software,[52] was used to estimate the variance explained by each of the matrices, with statistical significance being examined using a log-likelihood ratio test (LRT) and the Wald test. Model selection began with the full GKFSC model (referred to as the full model) and components were dropped if they were not statistically significant according to both the Wald and the LRT tests (The model that contained only components that explained a significant proportion of variance is referred to as the selected model). If more than one component could be dropped from the model, we dropped the one with the poorer fit

first then tested the significance of the other. The full results of each model can be seen in Supplementary Table 1.

The variance components corresponding to h^2_g (common SNP-associated effects), h^2_{kin} (Pedigree associated genetic effects), e_f^2 (shared family environment effect), e_s^2 (Shared sibling environment effect), and e_c^2 (shared couple environment effect) were estimated (Table 1).

Results

The results of the full GKFSC models (consisting of the GRM_g, GRM_{kin}, ERM_{Family}, $ERM_{Sibling}$, ERM_{Couple}), as well as the results of the selected models, can be seen in Table 1. For general intelligence (g) the final model was the GKSC model, allowing for a significant contribution from additive common genetic effects, additive pedigree-associated genetic variants, shared sibling environment, and a shared couple environment. For g, common SNPs (h^2_g) explained 23% (SE = 2%) of the phenotypic variation. Pedigree-associated genetic variants (h^2_{kin}) added an additional 31% (SE=3%) to the genetic contributions to g, yielding a total contribution of genetic effects on g of 54% (SE=3%). The net contribution of measured environmental factors to phenotypic variance in g was 35%. This was due to two sources of variance, shared sibling environment (e_s^2) and shared couple environment (e_c^2), that accounted for 9 % (SE=1%), and 22% (SE=2%), respectively.

The GKSC model was also the selected model for education, vocabulary, verbal fluency, and digit symbol test. As with general intelligence, pedigree-genetic variants accounted for the majority of the total genetic contribution to phenotypic variation in these traits. Pedigree-associated genetic variants explained between 15% - 30% of the variation, and common SNP effects explained 16% - 26%. The genetic results, i.e. SNP and pedigree contributions combined, for g and education are similar to the heritability estimates derived using the traditional pedigree study design which found a heritability estimate of 54% (SE=2%) for g[54] and 41% (SE=2%) for education using the same data set (Figure 2). This indicates that the genetic variants with the greater estimated cumulative effect on cognitive abilities are those that are poorly tagged on current genotyping platforms.

For logical memory the effect of shared couple environment was non-zero, but not significant, with the final selected model being GKS. Here, common SNP effects explained 12% (SE = 2%) of the variation. As with the other cognitive phenotypes considered here, pedigree-associated variants made a greater estimated contribution to the net genetic effect on logical memory, explaining 20% (SE=3%) of the variation. Sibling effects explained 5% (SE=1%) of the variation in logical memory.

For neuroticism the final model consisted of the G (GRM_g), and K (GRM_{kin}) contributions. Additive common genetic effects explained 11% (SE=2%) of the variance with pedigree-associated variants explaining an additional 19% (SE=3%). Whereas none of the environmental components was statistically significant, the family component accounted for 2% of the variance in the full model and

8

1% in a model that included only the G and the K matrices in addition to F. A lack of power may have occluded this effect.

For extraversion the only detectable source of genetic variation came from GRM_g, where G accounted for 13% (SE=2%), with $ERM_{Familiy}$ explaining a further 9% (SE=1%) of the phenotypic variation. The lack of pedigree-associated genetic effects could be due to low statistical power, as K explained 5% of the variance in the full model and 6% in a GKF model, but with a relatively large SE, estimated at 5%.

In addition to our model selection procedure, we also fit all possible component combinations for all phenotypes, to show a more complete account of the data and to give readers the ability to explore the consequences of including different components for the results, even when some of those components were not significant. The results have been made interactively available at https://rubenarslan.github.io/generation_scotland_pedigree_gcta/.

Discussion

The aim of this study was to use molecular genetic and pedigree data on the same large sample in order to decompose and quantify genetic and environmental sources of variation to intelligence and personality in a novel manner. In doing so, we sought to identify reasons for the gap between pedigree-based and SNP-based estimates of heritability in samples of unrelated individuals, a difference which might be due to genetic variants in poor linkage disequilibrium with SNPs genotyped on current platforms. This permits us to draw inferences about the evolutionary pressures that maintain general intelligence and personality differences. By making use of a large Scottish cohort that consists of close, distant, and spousal relationships, we were able to partition phenotypic variance of cognitive and personality measures into two genetic and three environmental sources of variance. A number of novel findings speak to long-standing questions in behaviour genetics and evolutionary genetics of psychological differences.[15, 30, 55]

Firstly, taken together, the two variance components derived directly from genome-wide molecular genetic data can account for the entire narrow-sense heritability of general intelligence, as estimated in twin and family studies.[54, 56] For all of the cognitive variables measured here, a substantial and significant proportion of the phenotypic variance was found to be explained by pedigree-associated genetic effects (h^2_{kin}). With the exception of the digit symbol test, these pedigree-associated genetic variants accounted for over half of the genetic effects.

The SNP-based methods of estimating heritability from unrelated individuals often produce lower heritability estimates than those derived using family-based studies. One reason for this is that population-based SNP methods, such as GREML, rely on LD between genotyped SNPs and causal variants at population level, and are sensitive to the frequency of causal alleles. Should LD between genotyped SNPs and causal variants be low, then the genetic similarity between a pair of individuals at the causal variant will be different to the genetic similarity at genotyped SNPs, resulting in a

reduction in the heritability estimate in the studies group. In within-family and twin studies, relatedness is based on identity by decent (IBD), where segments of DNA have been inherited from a recent common ancestor. Should a region be IBD between a pair of individuals, then all variants except *de novo* mutations within that segment are shared. As population-based SNP methods are sensitive to allele frequency, where IBD methods are blind to such effects, the discrepancy between the heritability estimates is consistent with the idea that causal variants in low LD with genotyped SNPs account for this missing element of the heritability of intelligence differences.

In the current study we investigate if variants in poor LD with genotyped SNPs account for additional heritability, unmeasured in GWAS on unrelated individuals, by using DNA from close family members. Higher genetic relatedness within families leads to an increase in the LD between genotyped SNPs and potentially causal variants and resulted in heritability estimates in our study that are comparable to pedigree-based methods. This provides evidence that, for intelligence, the gap between the heritability estimates derived using IBD methods and those derived using SNP-based population methods is most likely due to causal variants in low LD with genotyped SNPs. In addition, we were able to model this missing variance and separate it from the additive common genetic effects that are estimated in a GREML analysis based on unrelated individuals. The additional source of additive genetic variance from closely related family members, captured here in our kinship matrix (GRM_{kin}), would go unnoticed in a GWAS on unrelated individuals. This shows a need for GWAS on related individuals and for methods such as whole-genome sequencing to capture the individual effects of such variants. Whilst the use of related individuals can result in the confounding of pedigree genetic effects with shared family environmental effects, here we were able to distinguish the contributions made to phenotypic variance by pedigree-associated genetic variants from those by shared environment. Since we modelled three sources of environmental variance alongside the two genetic sources simultaneously, the variance that is due to a shared environment does not contribute towards our estimates of the genetic effects, as would be the case in instances where related individuals are included without adjusting for the shared environmental effects.[41, 52]

Furthermore, despite the level of confounding between the five matrices, we were able to correctly disentangle the contributions of each of the variance components. Simulations conducted by Xia *et al*.[41] show that this method provides robust results due to the dense relationships within the GS:SFHS cohort. The GS:SFHS is a family based cohort and the participants are related to varying degrees, including 1,767, 18,320, 7,851, 4,129, 3,950 and 11,032 pairs of couples, 1st, 2nd, 3rd, 4th and 5th degrees of relatives, respectively. Therefore, what is shared between ERM_{Family} matrix and GRM_{kin} matrix are merely ~18k pairs of entries represented by 1st degree relatives. However, ERM_{Family} holds ~1.8k pairs of unique entries (couple pairs) and GRM_{kin} holds ~23k pairs of unique entries (2nd-5th degree relative pairs of who were greater than 0.025 genetically identical), the unique entries from both matrices result in an increase of power to correctly disentangle the variance from those two different sources. An additional point is that the pedigree-associated genetic effects decay as the

distance of the relationship increases, whereas nuclear family environmental effects do not. Thus, the fact that GS:SFHS consist of different classes of relatives, as well as the unique entries within the GRM_{kin} and ERM_{Family}, helps to capture the property of pedigree-associated genetic variants. This logic extends to separating the variance from each of the environmental matrices. Although ERM_{Couple} and ERM_{Sib} are nested within the ERM_{Family}, there are 9,853 pairs of unique entries (representing parents-offspring) existing within the ERM_{Family}, which helps to separate the environmental effects. Therefore, there are sufficient number of appropriate relationship in GS:SFHS to make sure this method works.

Xia *et al.*[41] showed that our method reliably identifies the major sources of variance that contribute to trait architecture. However, as with any method, with decreasing effect sizes they become harder to measure accurately and more power is needed for the reliable detection of small signals. This means that if one of the matrices here only contributes to a small proportion of the overall phenotypic variance (e.g. less than 5% in GS:SFHS) its contributions will not be estimated reliably and dropped in the model selection procedure. However, any excluded minor component in the final model will have only a limited influence on the estimates of the major components that are retained in the final model. Thus, the major components we detected for each trait should be estimated reliably.

For personality traits, the genetic components can explain slightly more (30%) than the narrow-sense heritability (22%) that was meta-analytically derived from family and adoption studies with heterogeneous measurements of personality.[31] However, it falls short of the broad-sense heritability (47%[31]; 45%[57]). As previous research has suggested,[31, 58] this is consistent with epistasis playing a major role in personality genetics, as a non-additive genetic component is not captured well outside of twin studies. Previous research[58] did not discuss gene-environment correlation (rGE) and interaction (GxE) as a plausible cause for heritability estimates being higher in twin studies than in adoption and family studies, presumably because the shared environment contribution to personality variation was usually estimated not to be different from zero. A more recent meta-analysis gives an estimate of 13%[57] for shared environment that seems to be stable over age, so the difference between twin estimates of heritability and those presented here may also be explained to some extent by gene-by-environment interactions (GxE) and gene-environment correlations (rGE)[30].

The additional variance explained by the GRM_{kin} is unlikely to be due to *common* SNPs that are poorly tagged on current arrays because even with imputed data SNP heritability estimates of general intelligence are around 66% of that explained by twin models in the same sample.[59] Using GREML-LDMS,[60] and data on 43,599 participants with ~17 million SNPs imputed based on the 1000 Genomes reference panel, the heritability for height was found to be 56%, with 8% of this estimate being traced to rare variants with a minor allele frequency between 0.001 and 0.01. By imputing to 1000 Genomes, the same study estimated that 97% of common genetic variation was being captured, but only 68% of rare variation.[60] Using ~500 000 genotyped SNPs, Xia and colleagues[41] found that

common variants explained 43% of the variance for height, whereas 45% of the variance was explained by pedigree-associated variants. These two studies together show there is only a modest increase in heritability by measuring or imputing further common SNPs, and that with 97% of common variants the heritability estimate is lower than what can be found in family studies. As the h^2_g plus h^2_{kin} estimates for height from Xia and colleagues[41] exceed the total contribution from common variants, and closely approximate the estimates from twin studies of $h^2 = 0.89$-0.93[61], it seems more likely that the h^2_{kin} is not mainly driven by common genetic variants that are in low LD with genotyped variants but by rare variants, CNVs, and structural variation.

Our variance analyses are blind to the direction of effects and the number of variants involved in each genetic component. If, as we would predict, future work finds that variants with the lowest minor allele frequency tend to have the greatest negative effects, it would imply a coupling between the selection coefficient of alleles and their effect on intelligence, as selective pressure would act to minimise the frequency of highly deleterious variants. If this coupling were strong,[62] future work might infer that selection on intelligence was important in the past, even though current selective pressure may differ. If the impact of intelligence on fitness were limited to instances of pleiotropy with, for example, health, as some initial research suggests,[19, 20] the coupling between the selection coefficients of alleles and their effect sizes would be expected to be weaker. Selective pressure would act on the health-linked variants and intelligence-linked variants would only be selected to the extent of their pleiotropic effects on health. This would de-couple the selection coefficient of an allele and its effect on intelligence.

Future work can use the SNPs known to affect intelligence and personality[16, 17] to empirically quantify the coupling between allele frequency (indicating selection strength) and effect size to test this explanation, as has been demonstrated for height and BMI.[60] Targeted re-sequencing of enriched genetic regions[23, 63, 64] might be necessary to find very rare genetic variants associated with intelligence and personality, as has proven fruitful for example in prostate cancer research.[65] Future studies should test directly whether rarer SNPs have stronger negative effects on intelligence and personality, as has been shown for height and BMI.[60] This could test whether selection has acted to minimise the frequency of variants that negatively affect these traits.

The sibling component, which was retained in all models of intelligence, tracks the meta-analytic estimate of shared environmental variance (11%) from twin studies almost exactly. However, in our study the sibling component might also include a quarter of the dominance variation in intelligence that siblings share, because siblings are the only relationship in this study where dominance plays a significant role.[41] In the classical twin design, dominance variation (making dizygotic twins more dissimilar than half the similarity of monozygotic twins) can be obscured by shared environment effects (making DZs more similar). There is some evidence from other approaches that dominance only plays a minor role in intelligence differences.[66-69]

The family component was only retained in the model for Extraversion, although the point estimate was non-zero in the full Neuroticism model as well. This is consistent with meta-analytic estimates of shared environment for adults,[57] but may also be due to some level of confounding between K and F, where the association between extraversion and the GRM_{Family} is due to contributions of the genetic factors accounted for by the GRM_{kin}.

The couple component is somewhat complex to interpret. For intelligence[70] and education[71], there is evidence of assortative mating, which will increase both the genetic and environmental similarity between couples. The spousal similarity could explain still uncaptured genetic and environmental variance. If recent research on similarity in genetic propensities for education[71] is a good guide, spousal similarity in intelligence may be mostly explained by genetic similarity. Apart from this, the spouse's trait value may also serve as a good aggregate indicator of any effects the current environment has on a person, so that the couple component would also reflect recent environmental influences. The importance of shared environment with siblings appears to decline from childhood to adulthood,[72] as individuals pick their environmental individual niches (i.e., active gene-environment correlation). It may be that the *current* environment, now shared with a spouse, still has causal influences. We find no couple component for personality, which is consistent with much weaker assortative mating on personality, especially neuroticism and extraversion.[73-75]

In the current study we were able to exploit the high LD found between members of the same family to measure the contribution of genetic effects that are normally missed in GREML analyses of GWAS data. We also simultaneously modelled the effect of the family, sibling, and couple environment to avoid potential environmental confounds inflating our estimates. For intelligence and education, we find that genetic variants poorly tagged on current genotyping platforms explained a substantial and significant proportion of the phenotypic variance, raising our heritability estimates to match those derived using pedigree-based methods. Such variants can include CNVs, structural variants, and rare variants. We find similar effects for neuroticism, a dimension of personality genetically correlated with many fitness traits,[76] where pedigree-associated genetic variants explained 19% of variation. For extraversion, pedigree-associated variants appear to play a smaller role in phenotypic variation. These results suggest that mutation-selection balance has maintained heritable variation in intelligence and neuroticism, explaining why differences in these traits persist to this day despite selection. Future work should directly measure rare variants, as well as CNVs and structural variants, and test the direction of their effects.

Table 1

Results of the variance components analysis for cognitive abilities using the full model and the final model selected from a stepwise selection procedure.

Phenotype	N	Model	Variance components	GRM_g h^2_g %(S.E)	GRM_{kin} h^2_{kin} % (S.E)	ERM_{Family} e_f^2 % (S.E)	$ERM_{Sibling}$ e_s^2 % (S.E)	ERM_{Couple} e_c^2 % (S.E)
Cognitive								
g	19 036	Full	GKFSC	21.1 (2.0)	41.5 (4.8)	1.0×10^{-4} (2.2)	8.9 (1.3)	26.4 (2.6)
	19 036	Selected	GKSC	22.7 (2.1)	31.3 (2.9)		9.2 (1.3)	22.1 (2.0)
Education	18 528	Full	GKFSC	13.3 (2.0)	39.4 (5.1)	1.0×10^{-4} (2.4)	10.9 (1.4)	36.1 (2.7)
	18 528	Selected	GKSC	15.6 (2.1)	28.1 (3.0)		11.4 (1.4)	31.3 (2.8)
Vocabulary	19 269	Full	GKFSC	23.4 (0.2)	39.4 (4.7)	1.0×10^{-4} (2.2)	6.5 (1.3)	30.7 (2.6)
	19 269	Selected	GKSC	25.6 (2.0)	30.1 (2.8)		7.3 (1.3)	27.4 (1.9)
Verbal Fluency	19 380	Full	GKFSC	18.3 (2.0)	30.7 (5.2)	1.0×10^{-4} (2.4)	4.5 (1.3)	16.2 (3.1)
	19 380	Selected	GKSC	18.9 (2.1)	27.1 (2.9)		0.5 (1.3)	14.7 (2.1)
Digit Symbol Test	19 385	Full	GKFSC	20.2 (2.1)	22.7 (5.1)	1.0×10^{-4} (2.4)	8.0 (1.4)	17.3 (3.1)
	19 385	Selected	GKSC	21.4 (2.1)	14.7 (2.8)		8.1 (1.3)	13.2 (2.3)
Logical Memory	19 365	Full	GKFSC	11.4 (2.0)	24.0 (5.2)	1.0×10^{-4} (2.5)	5.1 (1.4)	5.3 (3.2)
	19 365	Selected	GKS	11.9 (2.0)	20.3 (2.8)		5.4 (1.4)	
Personality								
Neuroticism	19 494	Full	GKFSC	10.7 (2.0)	14.9 (5.1)	2.3 (2.5)	1.0×10^{-4} (1.4)	1.0×10^{-4} (3.4)
		Selected	GK	10.8 (2.0)	19.2 (2.5)			
Extraversion	19 487	Full	GKFSC	11.3 (0.2)	4.9 (5.1)	7.3 (2.5)	1.0×10^{-4} (1.4)	1.0×10^{-4} (3.3)
	19 487	Selected	GF	13.0 (1.7)		9.0 (1.1)		

Figure legends

Figure 1. Selected models plotted for each of the phenotypes included. The proportion of variance explained is on the x-axis with each of the phenotypes used on the y-axis. Each component from the selected model is plotted individually, with the stacked bar plot showing the total proportion of the variance explained by the selected models. Error bars indicate standard error.

Figure 2. Bar plots showing the proportion of variance explained using family based methods and using molecular genetic data. Both of these analyses were performed using the same GS:SFHS data (n = 20 522, Education n = 22 406, current manuscript, g n = 19 036, Education n = 18 5280. Estimates shown in red were derived in the current study using GREML and show two sources of genetic variance. Bright red being common genetic effects captured by the GRM_g matrix and dark red shows the additional genetic effects captured by exploiting the higher level of linkage disequilibrium between family members using the GRM_{kin} matrix. The estimates in blue are taken from Marioni and colleagues[54] and show the total genetic effects using ASReml-R when relatedness is inferred using identity by descent.

References

1.   Spearman C. "General Intelligence," Objectively Determined and Measured. *Am J Psychol* 1904; **15**(2)**:** 201-292.

2.   Lo M-T, Hinds DA, Tung JY, Franz C, Fan C-C, Wang Y *et al.* Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat Genet* 2016; **advance online publication**.

3.   Lahey BB. Public health significance of neuroticism. *American Psychologist* 2009; **64**(4)**:** 241.

4.   Posthuma D, De Geus EJC, Boomsma DI. Perceptual speed and IQ are associated through common genetic factors. *Behav Genet* 2001; **31**(6)**:** 593-602.

5.   Deary IJ. Intelligence. *Annu Rev Psychol* 2012; **63:** 453-482.

6.   Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D *et al.* Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatry* 2011; **16**(10)**:** 996-1005.

7.   Kirkpatrick RM, McGue M, Iacono WG, Miller MB, Basu S. Results of a "GWAS Plus:" General Cognitive Ability Is Substantially Heritable and Massively Polygenic. *PloS one* 2014; **9**(11)**:** e112390.

8.   Davies G, Marioni RE, Liewald DC, Hill WD, Hagenaars SP, Harris SE *et al.* Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). *Mol Psychiatry* 2016.

9.   Batty GD, Deary IJ, Gottfredson LS. Premorbid (early life) IQ and later mortality risk: systematic review. *Ann Epidemiol* 2007; **17**(4)**:** 278-288.

10.   Barban N, Jansen R, de Vlaming R, Vaez A, Mandemakers JJ, Tropf FC *et al.* Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat Genet* 2016; **48**(12)**:** 1462-1472.

11.   Day FR, Helgason H, Chasman DI, Rose LM, Loh P-R, Scott RA *et al.* Physical and neurobehavioral determinants of reproductive onset and success. *Nat Genet* 2016; **48**(6)**:** 617-623.

12.   Deary IJ, Strand S, Smith P, Fernandes C. Intelligence and educational achievement. *Intelligence* 2007; **35**(1)**:** 13-21.

13.   Banks GC, Batchelor JH, McDaniel MA. Smarter people are (a bit) more symmetrical: A meta-analysis of the relationship between intelligence and fluctuating asymmetry. *Intelligence* 2010; **38**(4)**:** 393-401.

14.   Van Dongen S, Gangestad SW. Human fluctuating asymmetry in relation to health and quality: a meta-analysis. *Evolution and Human behavior* 2011; **32**(6)**:** 380-398.

15.   Penke L, Denissen JJ, Miller GF. The evolutionary genetics of personality. *European Journal of Personality* 2007; **21**(5)**:** 549-587.

16.   Davies G, Armstrong N, Bis JC, Bressler J, Chouraki V, Giddaluru S *et al.* Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N = 53 949). *Mol Psychiatry* 2015; **20**(2)**:** 183-192.

17.   Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016; **533**(7604)**:** 539-542.

18.   Simons YB, Turchin MC, Pritchard JK, Sella G. The deleterious mutation load is insensitive to recent population history. *Nat Genet* 2014; **46**(3)**:** 220.

19.   Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**(11)**:** 12236 - 11241.

20.   Hill WD, Davies G, The CHARGE Cognitive Working Group, Liewald DC, McIntosh AM, Deary IJ. Age-dependent pleiotropy between general cognitive function and major psychiatric disorders. *Biol Psychiatry* 2015; **80**(4)**:** 266-273.

21.   Beauchamp JP. Genetic evidence for natural selection in humans in the contemporary United States. *Proceedings of the National Academy of Sciences* 2016; **113:** 7774–7779.

22.   Conley D, Domingue B. The Bell Curve Revisited: Testing Controversial Hypotheses with Molecular Genetic Data. *Sociological Science* 2016; **3:** 520-539.

23.   Hill WD, Davies G, Harris SE, Hagenaars SP, Liewald D, Penke L *et al.* Molecular genetic aetiology of general cognitive function is enriched in evolutionarily conserved regions. *Translational psychiatry* 2016; **6**(e980).

24.   Marioni RE, Penke L, Davies G, Huffman JE, Hayward C, Deary IJ. The total burden of rare, non-synonymous exome genetic variants is not associated with childhood or late-life cognitive ability. *Proceedings of the Royal Society B: Biological Sciences* 2014; **281**(1781)**:** 20140117.

18

25.   MacLeod AK, Davies G, Payton A, Tenesa A, Harris SE, Liewald D *et al.* Genetic copy number variation and general cognitive ability. *PloS one* 2012; **7**(12)**:** e37385.

26.   McRae AF, Wright MJ, Hansell NK, Montgomery GW, Martin NG. No association between general cognitive ability and rare copy number variation. *Behav Genet* 2013; **43**(3)**:** 202-207.

27.   Kirkpatrick RM, McGue M, Iacono WG, Miller MB, Basu S, Pankratz N. Low-frequency copy-number variants and general cognitive ability: No evidence of association. *Intelligence* 2014; **42:** 98-106.

28.   Yeo RA, Gangestad SW, Liu J, Calhoun VD, Hutchison KE. Rare copy number deletions predict individual variation in intelligence. *PloS one* 2011; **6**(1)**:** e16339.

29.   Ganna A, Genovese G, Howrigan DP, Byrnes A, Kurki MI, Zekavat SM *et al.* Ultra-rare disruptive and damaging mutations influence educational attainment in the general population. *Nat Neurosci* 2016; **19**(12)**:** 1563-1565.

30.   Penke L, Jokela M. The evolutionary genetics of personality revisited. *Current Opinion in Psychology* 2016; **7:** 104-109.

31.   Vukasović T, Bratko D. Heritability of personality: A meta-analysis of behavior genetic studies. *Psychological bulletin* 2015; **141**(4)**:** 769.

32.   Smith DJ, Escott-Price V, Davies G, Bailey ME, Colodro-Conde L, Ward J *et al.* Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Mol Psychiatry* 2016.

33.   van Den Berg SM, de Moor MH, Verweij KJ, Krueger RF, Luciano M, Vasquez AA *et al.* Meta-analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium. *Behav Genet* 2016; **46**(2)**:** 170-182.

34.   Terracciano A, Löckenhoff CE, Zonderman AB, Ferrucci L, Costa Jr PT. Personality predictors of longevity: activity, emotional stability, and conscientiousness. *Psychosom Med* 2008; **70**(6)**:** 621.

35.   Bienvenu OJ, Samuels JF, Costa PT, Reti IM, Eaton WW, Nestadt G. Anxiety and depressive disorders and the five-factor model of personality: A higher-and lower-order personality trait investigation in a community sample. *Depression and anxiety* 2004; **20**(2)**:** 92-97.

36.   Jokela M. Birth-cohort effects in the association between personality and fertility. *Psychol Sci* 2012; **23**(8)**:** 835-841.

37.   Roberts RO, Geda YE, Knopman DS, Cha RH, Pankratz VS, Boeve BF *et al.* The incidence of MCI differs by subtype and is higher in men The Mayo Clinic Study of Aging. *Neurology* 2012; **78**(5)**:** 342-351.

38.   Krapohl E, Euesden J, Zabaneh D, Pingault J, Rimfeld K, Von Stumm S *et al.* Phenome-wide analysis of genome-wide polygenic scores. *Mol Psychiatry* 2015; (doi: 10.1038/mp.2015.126).

39.   Rietveld CA, Esko T, Davies G, Pers TH, Turley P, Benyamin B *et al.* Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc Natl Acad Sci* 2014; **111**(38)**:** 13790-13794.

40.   Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S *et al.* Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet* 2013; **9**(5)**:** e1003520.

41.   Xia C, Amador C, Huffman J, Trochet H, Campbell A, Porteous D *et al.* Pedigree-and SNP-Associated Genetics and Recent Environment are the Major Contributors to Anthropometric and Cardiometabolic Trait Variation. *PLoS Genet* 2016; **12**(2)**:** e1005804.

42.   Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM *et al.* Cohort profile: Generation Scotland: Scottish Family Health Study (GS: SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol* 2012; **42**(3)**:** 689-700.

43.   Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ *et al.* Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Medical Genetics* 2006; **7**(1)**:** 9.

44.   Nagy R, Boutin TS, Marten J, Huffman JE, Kerr SM, Campbell A *et al.* Exploration of Haplotype Research Consortium Imputation for Genome-Wide Association Studies in > 20,000 Generation Scotland Participants. *Genome Medicine* in press.

45.   Kerr SM, Campbell A, Murphy L, Hayward C, Jackson C, Wain LV *et al.* Pedigree and genotyping quality analyses of over 10,000 DNA samples from the Generation Scotland: Scottish Family Health Study. *BMC medical genetics* 2013; **14**(1)**:** 38.

46.   Gunderson KL. Whole-genome genotyping on bead arrays. In: Dufva M (ed). *DNA Microarrays for Biomedical Research*. Humana Press.2009, pp pp. 197-213.

47.   Amador C, Huffman J, Trochet H, Campbell A, Porteous D, Wilson JF *et al.* Recent genomic heritage in Scotland. *BMC Genomics* 2015; **16**(1)**:** 1.

48.   Raven JC, Court JH, Raven J. *Manual for Raven's Progressive Matrices and Vocabulary Scales*. H.K. Lewis: London, 1977.

49.   Wechsler D. *Wechsler Adult Intelligence Scale - third edition*. The Psychological Corporation: London, 1998.

50.   Wechsler D. *Wechsler Memory Scale III UK*. Psychological Corporation: London, 1998.

51.   Lezak M, Howieson DB, Loring DW. *Neuropsychological Assessment*. 4 edn. Oxford University Press, USA: New York, 2004.

52.   Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* 2011; **88**(1)**:** 7.

53.   Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010; **42**(7)**:** 4.

54.   Marioni RE, Davies G, Hayward C, Liewald D, Kerr SM, Campbell A *et al.* Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence* 2014; **44:** 26-32.

55.   Arslan RC, Penke L. Zeroing in on the Genetics of Intelligence. *Journal of Intelligence* 2015; **3**(2)**:** 41-45.

56.   Chipuer HM, Rovine MJ, Plomin R. LISREL modeling: Genetic and environmental influences on IQ revisited. *Intelligence* 1990; **14**(1)**:** 11-29.

57.   Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* 2015; **47**(7)**:** 702–709.

58.   Plomin R, Corley R, Caspi A, Fulker DW, DeFries J. Adoption results for self-reported personality: Evidence for nonadditive genetic effects? *Journal of Personality and Social Psychology* 1998; **75**(1)**:** 211.

59.   Plomin R, Haworth CM, Meaburn EL, Price TS, Davis OS. Common DNA markers can account for more than half of the genetic influence on cognitive abilities. *Psychol Sci* 2013; **24**(4)**:** 6.

60.   Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AA, Lee SH *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet* 2015; **47**(10)**:** 1114-1120.

61.   Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C *et al.* Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin research* 2003; **6**(05)**:** 399-408.

62.   Eyre-Walker A. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proceedings of the National Academy of Sciences* 2010; **107**(suppl 1)**:** 1752-1756.

63.   Hill WD, Davies G, van de Lagemaat LN, Christoforou A, Marioni RE, Fernandes CPD *et al.* Human cognitive ability is influenced by genetic variation in components of postsynaptic signalling complexes assembled by NMDA receptors and MAGUK proteins. *Translational psychiatry* 2014; **4**(1)**:** e341.

64.   Johnson MR, Shkura K, Langley SR, Delahaye-Duriez A, Srivastava P, Hill WD *et al.* Systems genetics identifies a convergent gene network for cognition and neurodevelopmental disease. *Nat Neurosci* 2015.

65.   Mancuso N, Rohland N, Rand KA, Tandon A, Allen A, Quinque D *et al.* The contribution of rare variation to prostate cancer heritability. *Nat Genet* 2015.

66.   Bashi J. Effects of inbreeding on cognitive performance. *Nature* 1977; **266:** 440 - 442.

67.   Gamsiz ED, Viscidi EW, Frederick AM, Nagpal S, Sanders SJ, Murtha MT *et al.* Intellectual disability is associated with increased runs of homozygosity in simplex autism. *The American Journal of Human Genetics* 2013; **93**(1)**:** 103-109.

68.   Joshi PK, Esko T, Mattsson H, Eklund N, Gandin I, Nutile T *et al.* Directional dominance on stature and cognition in diverse human populations. *Nature* 2015; **523**(7561)**:** 459-462.

69.   Power AR, Nagoshi C, DeFries JC, Wellcome Trust Case Control Consortium, Plomin R. Genome-wide estimates of inbreeding in unrelated individuals and their association with cognitive ability. *Eur J Hum Genet* 2014; **22:** 386-390.

70.   Jensen AR. *The g factor: The science of mental ability.* CT: Praeger: Westport, 1998.

71.   Robinson MR, Kleinman A, Graff M, Vinkhuyzen AAE, Couper D, Miller MB *et al.* Genetic evidence of assortative mating in humans. *Nature Human Behaviour* 2017; **1:** 0016.

72.   Plomin R, Deary IJ. Genetics and intelligence differences: five special findings. *Mol Psychiatry* 2014; **20:** 98-108.

73.   Zietsch BP, Verweij KJ, Heath AC, Martin NG. Variation in human mate choice: simultaneously investigating heritability, parental influence, sexual imprinting, and assortative mating. *The American Naturalist* 2011; **177**(5)**:** 605-616.

74.   Dyrenforth PS, Kashy DA, Donnellan MB, Lucas RE. Predicting relationship and life satisfaction from personality in nationally representative samples from three countries: the

relative importance of actor, partner, and similarity effects. *Journal of personality and social psychology* 2010; **99**(4)**:** 690.

75.   Rammstedt B, Spinath FM, Richter D, Schupp J. Partnership longevity and personality congruence in couples. *Personality and Individual Differences* 2013; **54**(7)**:** 832-835.

76.   Gale C, Hagenaars SP, Davies G, Hill WD, Liewald DC, Cullen B *et al.* Pleiotropy between neuroticism and physical and mental health: findings from 108 038 men and women in UK Biobank. *Translational psychiatry* 2016; **6**(4)**:** e791.

Genetic and environmental contributions to phenotypic variance
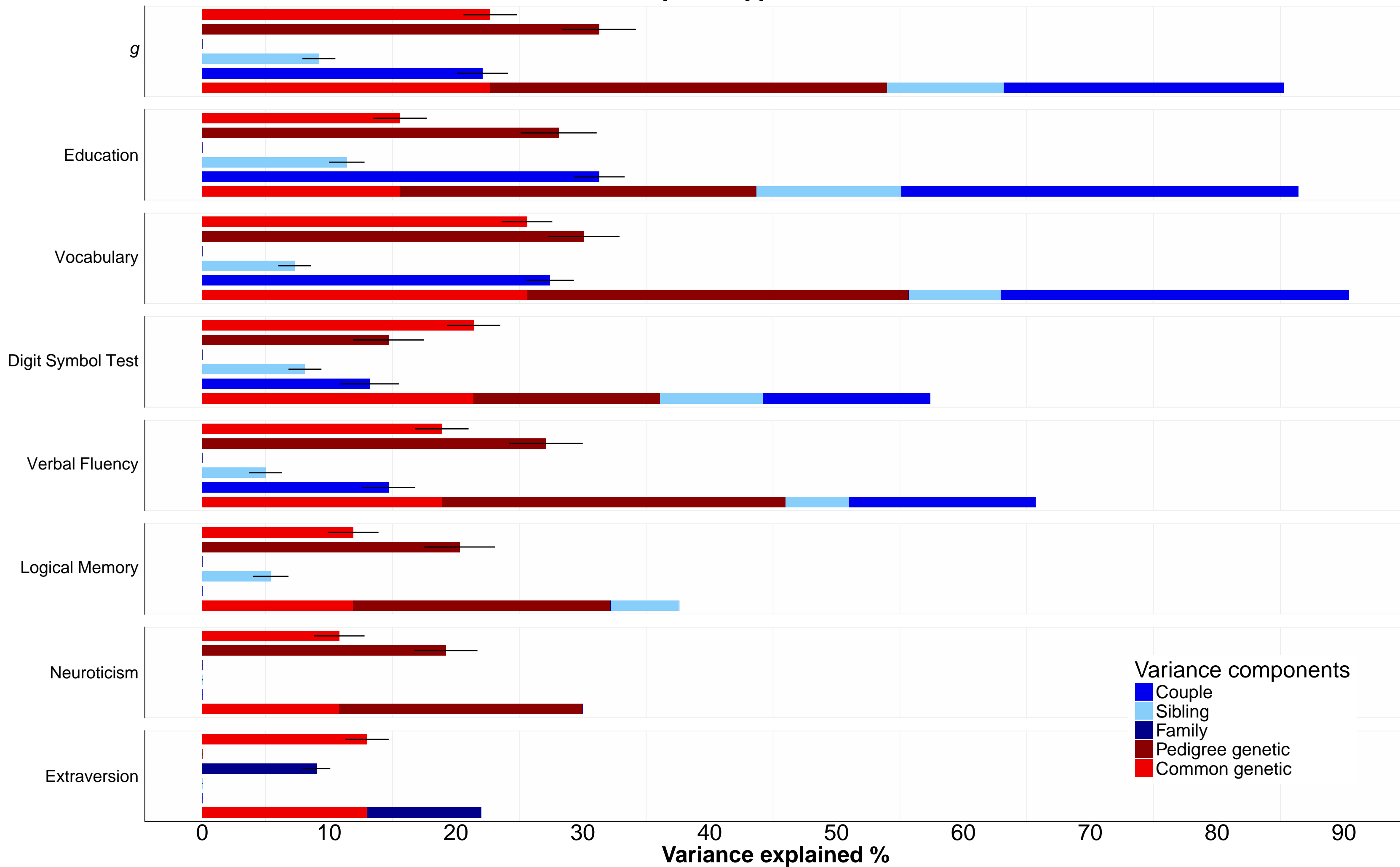
Comparison to the traditional pedigree estimate