

Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion

AAE Vinkhuyzen^{1,2}, NL Pedersen³, J Yang¹, SH Lee^{1,2}, PKE Magnusson³, WG Iacono⁴, M McGue⁴, PAF Madden⁵, AC Heath⁵, M Luciano⁶, A Payton⁷, M Horan⁸, W Ollier⁷, N Pendleton⁸, IJ Deary⁶, GW Montgomery¹, NG Martin¹, PM Visscher¹ and NR Wray^{1,2}

The personality traits of neuroticism and extraversion are predictive of a number of social and behavioural outcomes and psychiatric disorders. Twin and family studies have reported moderate heritability estimates for both traits. Few associations have been reported between genetic variants and neuroticism/extraversion, but hardly any have been replicated. Moreover, the ones that have been replicated explain only a small proportion of the heritability ($< \sim 2\%$). Using genome-wide single-nucleotide polymorphism (SNP) data from $\sim 12\,000$ unrelated individuals we estimated the proportion of phenotypic variance explained by variants in linkage disequilibrium with common SNPs as 0.06 (s.e. = 0.03) for neuroticism and 0.12 (s.e. = 0.03) for extraversion. In an additional series of analyses in a family-based sample, we show that while for both traits $\sim 45\%$ of the phenotypic variance can be explained by pedigree data (that is, expected genetic similarity) one third of this can be explained by SNP data (that is, realized genetic similarity). A part of the so-called 'missing heritability' has now been accounted for, but some of the reported heritability is still unexplained. Possible explanations for the remaining missing heritability are that: (i) rare variants that are not captured by common SNPs on current genotype platforms make a major contribution; and/ or (ii) the estimates of narrow sense heritability from twin and family studies are biased upwards, for example, by not properly accounting for nonadditive genetic factors and/or (common) environmental factors.

Translational Psychiatry (2012) 2, e102; doi:10.1038/tp.2012.27; published online 17 April 2012

Introduction

The personality traits of neuroticism and extraversion are two higher-order dimensions of personality that are consistently identified in different dimensional models of personality such as the Five Factor Model of personality. Both personality traits are captured in a range of personality inventories (for example, NEO-Five Factor Inventory,¹ Eysenck Personality Questionnaire² and the Multidimensional Personality Questionnaire.^{3,4}

Neuroticism is characterized by a tendency for emotional instability, psychological distress, low self-esteem and negative emotions such as anxiety and depression. Extraversion is characterized by a tendency for high levels of sociability, activity, sensation seeking and positive emotions. Both traits are predictive of a number of social and behavioural outcomes and of anxiety and depressive disorders.^{5–9} For example, high levels of neuroticism are associated with lifetime disorders such as major depressive disorder, generalised anxiety disorder, social phobia, dysthymia and obsessive-compulsive disorder. Low levels of extraversion are associated with social phobia, agoraphobia and dysthymia. Causes of individual differences in both neuroticism and extraversion have been studied extensively using twin and adoption studies, showing

heritability estimates ranging from 13 to 58% for neuroticism and from 34 to 57% for extraversion.^{10–18} Genetic factors that are related to neuroticism and extraversion may also be genetic risk factors for associated mental disorders.^{7,19,20} Therefore, insight into the genetic architecture of neuroticism and extraversion may shed light on biological mechanisms underlying mental disorders that are associated with these traits.

Gene-finding studies for higher-order personality traits have focused mainly on neuroticism, because of its association with anxiety and depression. Identifying genetic variants associated with neuroticism, but also with extraversion, has proven difficult. Suggestive linkage has been reported for several chromosomal regions with only few replications.^{21–27} Candidate studies have reported associations with markers within several genes,^{28–32} but again, replication generally failed.^{33–35} To date, a number of genome-wide association studies (GWAS) for higher-order personality traits have been published.^{36–40} None of them, however, has reported genome-wide significant associations between neuroticism/extraversion and markers. The dilemma of moderate to high heritability estimates of which only a small proportion can be explained by markers has been expressed as 'the case of the missing heritability'.^{41,42}

¹Queensland Institute of Medical Research, Brisbane, Queensland, Australia; ²The University of Queensland, Queensland Brain Institute, Brisbane, Queensland, Australia; ³Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; ⁴Department of Psychology, University of Minnesota, Minneapolis, MN, USA; ⁵Washington University School of Medicine, St Louis, MO, USA; ⁶Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, Edinburgh, UK; ⁷Medical Genetics Section, University of Edinburgh Molecular Medicine Centre, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, UK and ⁸School of Medicine, The University of Manchester, Manchester, UK

Correspondence: Dr AAE Vinkhuyzen, The University of Queensland, Queensland Brain Institute (building no. 79), Brisbane, St Lucia 4072, Queensland, Australia. E-mail: anna.vinkhuyzen@uq.edu.au

Keywords: complex traits; GCTA; genome-wide; polymorphisms; variance

Received 14 February 2012; accepted 3 March 2012

Recently, Yang *et al.*⁴³ showed (i) that 45% of the phenotypic variance (roughly half of the heritability) of human height can be explained by considering all single-nucleotide polymorphism SNPs simultaneously in a linear model analysis, implying that most of the heritability is not missing but is as yet undetected due to effect sizes of individual variants being too small to reach significance in GWAS conducted to date; and (ii) that the remaining heritability can be ascribed to incomplete linkage disequilibrium (LD) between causal variants and genotyped SNPs. Insufficient LD is likely to occur if causal variants have lower minor allele frequency (MAF) than the genotyped SNPs.⁴⁴

The purpose of the present study is to estimate the proportion of the phenotypic variance of neuroticism and extraversion explained by considering all SNPs on currently-used genome-wide arrays simultaneously.^{43,45} The basic idea of this method is to accumulate the effects of all the associated SNPs that might be too small to reach the stringent significance level in single SNP analyses. To this end, we analysed the neuroticism and extraversion personality trait scores, measured in ~12 000 unrelated individuals from combined samples from Australia, Sweden, the United Kingdom and the United States of America.

Methods

Sample and phenotypes. Information on samples and phenotypes is provided in the Supplementary Information Section 1. Briefly, neuroticism data were available for 17 875 (58% female) individuals (mean age 39.07; s.d. = 16.02, range: 14–86). After estimation of the pairwise genetic similarity using all autosomal markers and deleting one of each pair of individuals with a genetic relatedness > 0.025 (See Supplementary Information Section 1), 11 961 (58% females) individuals were retained (mean age 42.36; s.d. = 16.91, range: 14–86). For extraversion, data from 17 557 individuals (59% female) individuals were available (mean age 39.01; s.d. = 16.18, range: 14–86) of whom 11 786 (58% females) were genetically ‘unrelated’ (mean age 42.36; s.d. = 17.09, range: 14–86).

Analyses. The method applied in the present study is extensively described by Yang *et al.*^{43,45} and is designed to capture variation due to LD between genotyped SNPs and unknown causal variants in the genome. This estimate of variance explained by all SNPs is different from heritability estimates in twin and family studies as the latter include variance explained by all causal variants (an estimate of the latent genetic effect). For more detailed information on the methods see Supplementary Information Section 2.

We first fitted a linear model including all autosomal SNPs to estimate the proportion of the variance explained by all the SNPs. The Study cohort status and the first 20 principal components from a principal-component analysis of the genome-wide genotypes were fitted as covariates in the model to control for the effects attributable to population structure. Analyses were repeated (i) without PC adjustment, (ii) adjusting for imprecise LD between genotyped SNPs and causal variants (See notes Supplementary Information Table

S2 for details), (iii) separately for each study-cohort, (iv) for the total sample excluding one study-cohort at a time, (v) fitting only SNPs on the X chromosome, (vi) separately for men and women and (vii) including a genotype-by-sex interaction term.

Twin and family studies have reported, although inconsistently, sex effects on genetic factors underlying both neuroticism and extraversion.^{3,18,46–50} To investigate possible sex effects on the genetic variance of neuroticism and extraversion in the current design, significance of a genotype-by-sex interaction term was tested. We also estimated genetic variance explained by all the SNPs in sex-specific analyses. We investigated possible sex effects in both the total sample and in the Australian cohort only. Twin and family studies have generally suggested stability of genetic factors underlying both neuroticism and extraversion over age,^{3,18,46–50} others, however, suggested instability over age.⁵⁰ Age effects were not investigated as age is confounded with the different study cohorts in the present study.

To investigate whether the variance explained is proportional to the length of the chromosome, we subsequently partitioned the variance explained into individual autosomes. To this end, all chromosomes were simultaneously fitted in a mixed linear model and the proportion of the variance explained by each of the chromosomes was estimated.⁵¹

In a subsequent series of analyses we investigated possible discrepancy in the percentage of variance explained by SNPs and variance explained by latent additive genetic influences estimated in twin and family studies. We mimicked the conventional AE-model (that is, estimation of additive genetic factors and environmental factors) using the entire Australian sample, including close relatives (for example, MZ/DZ twin pairs, full sibs, parent-offspring pairs and cousins), that is, no cutoff in genetic similarity between individuals was applied. In twin and family studies, estimates of heritability are based on the relationship between phenotypic resemblance and expected genetic similarity based on pedigree information (for example, MZ twin pairs are expected to share 100% of their genetic material, whereas DZ twins, full sibs and parent-offspring pairs are expected to share 50% of their genetic material). SNP data, however, provide us with estimates of the realized genetic similarity, which vary around the expected values from pedigrees (See Supplementary Information Figure S2). The genetic variation around the expected values (for example, 0, 50 or 100%) is not captured by pedigree studies but can be captured by the SNP data. In order to partition the genetic variance into variance explained by pedigree data and variance explained by SNP data, genetic similarity was estimated from pedigree information (expected genetic similarity) and from SNP data (realized genetic similarity) in a series of separate and joint analyses of the pedigree and SNP data. The analyses are described in detail in the Supplementary Information Section 1. Briefly, we (1) estimated variance explained by all SNPs based on ‘unrelated’ individuals using genome-wide SNP data, as above; (2) estimated variance explained by all SNPs based on *all* individuals using genome-wide SNP data; (3) estimated heritability from *all* individuals using expected genetic similarity (pedigree similarity matrix); (4) partitioned the variance into variance captured by the SNP similarity matrix and variance captured by the pedigree similarity matrix from

all individuals; and (5) partitioned the variance further into variance captured by the pedigree similarity matrix from all individuals, the SNP similarity matrix for 'unrelated' individuals and the SNP similarity matrix for 'related' individuals. Note that estimates of variance from analyses including close relatives are increased by the effects of (i) possible environmental factors that render phenotypic similarity between genetically more similar individuals and (ii) by causal variants that are not correlated with genotyped SNPs but are captured by pedigree structure. Close relatives share a substantial proportion of their segregating genes on all chromosomes (for example, ~50% for full siblings), consequently, variance attributable to causal variants on one chromosome will be captured by SNPs on other chromosomes.^{43,52}

Results

A genetic similarity matrix was estimated for all individuals using 849 801 autosomal SNPs that passed quality control. After removing one individual of each pair with a genetic similarity >0.025, a subset of ~12 000 individuals was retained for the analyses of neuroticism and extraversion (See Supplementary Information Table S2 for exact numbers for the combined sample and for specific study cohorts). Estimates of genetic similarities for each pair of 'unrelated'

individuals were normally distributed with mean of -0.0002 and s.d. = 0.0043 (See Supplementary Information Figure S1).

The proportion of the phenotypic variance explained by all the SNPs (h^2_{SNP}) estimated using all 'unrelated' individuals was 0.06 (s.e. = 0.03, $P < 0.05$) for neuroticism, and 0.12 (s.e. = 0.03, $P < 0.001$) for extraversion (See Figure 1 and Supplementary Information Table S2).

If the analyses were based on only the SNPs that were in common between all the study cohorts, estimates of the heritability were slightly lower (See Figure 1 and Supplementary Information Table S2). This is expected because a less dense SNP set (162 056 versus 871 333) provides lower LD between the genotyped SNPs and the unknown causal variants. Analyses with and without the first 20 principal components included as covariates in the model revealed very similar results, implying that any population stratification that was observed within our European-ancestry sample had a negligible effect on the results. The analyses were repeated for subsets of the sample (Figure 1 and Supplementary Information Table S2). Separate study-cohort analyses showed a fairly inconsistent pattern reflecting larger s.e., but generally estimates were lower for neuroticism than for extraversion. Study-cohort separate analyses based on only the 162 056 SNPs that were in common between all study cohorts showed a similar pattern as the analyses based on all the SNPs, suggesting that different estimates between study

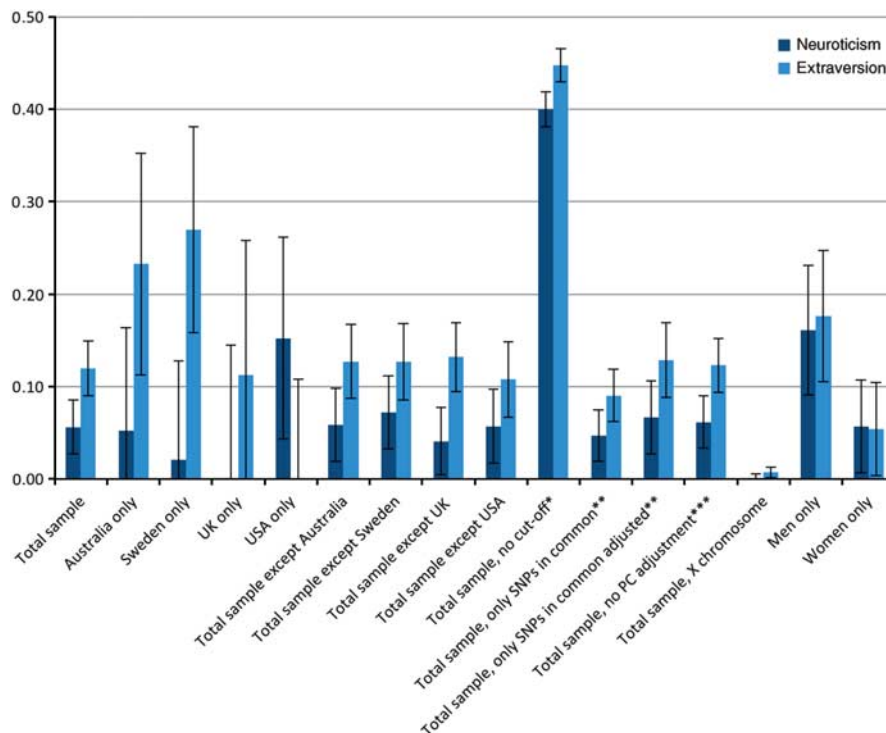


Figure 1 Proportion of variance explained by autosomal single-nucleotide polymorphisms (SNPs) for neuroticism (dark blue) and extraversion (light blue). Notes: all individuals included in the analyses have a pair-wise genetic similarity <0.025 except for the * marked in which no cutoff was made; the first 20 principal components (PC) were included as fixed effects in the model, except for the *** marked; error bars represent s.e.; PC; SNPs; UK and USA.**analyses based on 162 056 SNPs that were in common between all study cohorts; adjusted: estimate adjusted for imprecise LD between genotyped SNPs and causal variants for causal variants within the allelic frequency spectrum as genotyped SNPs, using the regression coefficient β from equation: $\beta = 1 - \frac{(c+1/N)}{\text{var}(A_{jk})}$ (assuming $c = 0$), where A_{jk} is the variance of the off-diagonal elements of the genetic similarity matrix, N is the number of SNPs used to calculate A_{jk} . The value of c depends on the minor allele frequency of the causal variants.⁴⁴ Further details and statistics are provided in Supplementary Table S2. Abbreviations: SNPs, single-nucleotide polymorphisms; UK, United Kingdom; USA, United States of America.

cohorts are not a result of different SNP arrays that were used for genotyping in the different study cohorts. Exclusion of any of the study cohorts provided no evidence that one of them affected the results disproportionately. We also estimated the variance explained by the X chromosome and did not find any significant result for either neuroticism or extraversion.

Separate analyses on the sexes in the total sample showed different point estimates for men and women with increased point estimates in men for both neuroticism and extraversion and a decreased point estimate in women for extraversion, however, the genotype-by-sex interaction analysis suggested that these differences were not significant (Supplementary Information Table S2). The estimate of genetic variance from the total sample reflects a weighted average of the estimates of variance from each of the sex cohorts and the covariance tagged by SNPs between the sexes. The nonsignificant genotype-by-sex interaction implies no significant differences between the genetic variances for each sex and the covariance between sexes.

Partitioning the genetic variance into the 22 autosomes revealed that the phenotypic variance of extraversion explained by each chromosome is proportional to its length ($R^2 = 0.20$, $P < 0.05$; Figure 2). This corroborates evidence of a polygenic model underlying variation in extraversion. As longer chromosomes harbour more genes, a linear relationship between length of a chromosome and the estimate of variance explained by that chromosome is consistent with the hypothesis that many genes of small effect contribute to genetic variation of the trait. The s.e. of the estimates for the chromosomes were high so that none of the estimates differed significantly from the regression line. This suggests that no individual loci disproportionately affect genetic variation. Genome partitioning for neuroticism did not show a significant relationship between chromosome length and variance explained (Figure 2). This is not unexpected given the smaller genetic variance that can be attributed to common SNPs.

Estimates of the heritability based on the separate and joint analyses of genetic similarity estimated from pedigree and from SNP data based on the Australian sample only are listed in Supplementary Information Table S3. The coefficient from the regression pedigree pairwise similarities on the SNP pairwise similarities is 1.003 (See Supplementary Information Figure S2), as expected.

Heritability estimates from SNP data using only 'unrelated' individuals were substantially lower than estimates reported by twin and family studies (analysis 1 in Supplementary

Information Table S3). Heritability estimates from both SNP and pedigree data using all individuals, however, (analyses 2 and 3 in Supplementary Information Table S3), confirmed a moderate heritability for both neuroticism and extraversion from twin and family studies ($\sim 42\%$ from SNP analyses and $\sim 45\%$ from pedigree analyses). Note that estimates of variance from analyses including close relatives are increased by the effects of all causal variants (rather than only the effects of causal variants in sufficient LD with the genotyped SNPs included in the analyses), but may also be inflated by confounding of closely related individuals and common-environmental factors and nonadditive genetic effects.

The estimates of variance using all data will reflect some weighted average of the estimates from close and distant relatives, with relatively more weight attributed to close relatives because the coefficients of similarity are so much higher. Estimates from SNP data were slightly lower than estimates from pedigree data, reflecting the relatively large number of pairwise combinations of 'unrelateds' in the SNP analyses.

Partitioning the variance into variance explained by pedigree data and variance explained by SNP data showed that for both neuroticism and extraversion one third of the variance (14%) that was explained by pedigree data could be picked up by SNP similarity alone (analysis 4 in Supplementary Information Table S3). Further partitioning showed that the part of the variance that could be explained by SNP similarity was mainly due to variation detected from those individuals that are 'unrelated' in the conventional (pedigree) sense, and much less so to variation in realized genetic similarity of related individuals around the expected (pedigree) values (analyses 5–7 in Supplementary Information Table S3). However, the s.e. of the estimates from realized similarity were large, as expected from theory and empirical studies.^{53,54}

Discussion

The aim of this study was to estimate the proportion of variance for personality dimensions of neuroticism and extraversion explained by all the SNPs in a linear model analysis. Common SNPs explain 6% (s.e. 3%) and 12% (s.e. 3%) of the phenotypic variance for neuroticism and extraversion, respectively.

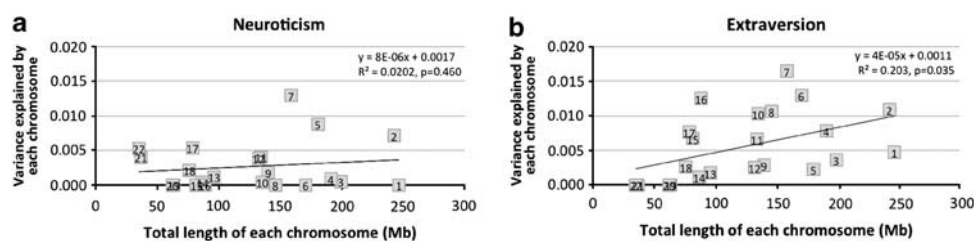


Figure 2 Variance explained by each chromosome for neuroticism (a) and extraversion (b). Notes: equation corresponding to the linear regression line for neuroticism: $y = 8E-06x + 0.0017$ and extraversion: $y = 4E-05x + 0.0011$, where y is the variance explained by each of the chromosomes and x is the total length of the chromosome in mega bases (Mb); R^2 corresponding to the regression equation for neuroticism: 0.0202 ($P = 0.460$) and extraversion: 0.203 ($P = 0.035$).

Estimates of variance explained by all the SNPs for neuroticism and extraversion are similar to estimates reported for the Cloninger Temperament subscales Harm Avoidance and Novelty Seeking in a Finnish sample.⁵⁵ The significant estimates from the present study contrast considerably with results from candidate gene studies and large scale GWAS. Hitherto, only very little variance ($< \sim 2\%$) could be attributed to markers that have been reported in candidate gene studies^{28–32} and no variance could be explained by SNPs detected in large-scale GWAS.^{36–40} Those GWAS were powered to detect genetic variants that explain 1–2% of the phenotypic variance.

Our results suggest a polygenic model in which a large number of variants individually explain a small proportion of the variance. Moreover, we observed a linear relationship between the length of a chromosome and the proportion of the variance explained by that chromosome for extraversion, which also supports a model in which many variants affect the variance of the trait. As none of the chromosomes significantly differed from the regression line, no major gene effects are supported in our analyses. The linear relationship between chromosome length and variance explained was substantially smaller for neuroticism, which was expected given the smaller genetic variance that can be attributed to common SNPs.

In our approach, we tried to eliminate possible causes of sample structure (for example, hidden relatedness, population stratification) that may lead to spurious association.⁵⁶ Consequently, phenotypic correlations between individuals are due to genetic similarities so that the estimated genetic variance component entirely reflects LD between the genotyped SNPs and the unknown causal variants. Our estimates are unbiased estimates of the variance in neuroticism and extraversion explained by common SNPs, so that the estimates are not expected to change (within bounds implied by the s.e. and assuming the same genomic coverage by SNPs) as sample sizes increase. These results imply that with larger sample sizes, common associated variants could be detected in single SNP analysis, depending on effect sizes of single variants.

Although our estimates surpass results from GWASs, estimates are still lower than heritability estimates reported in twin and family studies. The discrepancy can be attributed to several factors. First, heritability estimates from twin and family studies include the effects of all causal variants, whereas heritability estimated using the methods employed in this study include only the effects of variants that are in (sufficient) LD with the genotyped SNPs included in the analyses. LD is especially low if the causal variant has a different MAF than the genotyped marker. Consequently, contribution of rare variants ($MAF < 0.5\%$) and variants with low MAF ($0.5\% < MAF < 1\%$) is included in the heritability estimates from twin and family studies but is generally not included in the variance explained by common SNPs. As causal variants are unknown, we cannot estimate LD between genotyped SNPs and causal variants directly. However, assuming that LD between the causal variants and genotyped SNPs is as strong as between the genotyped SNPs (that is, adjustment for prediction error of causal variants that have the same allelic frequency spectrum as genotyped SNPs), we corrected the estimate of variance explained by all the SNPs

that were in common between all study cohorts (162 056 SNPs) for incomplete LD with causal variants. The adjusted point estimates were considerably higher for both neuroticism (7 versus 5%) and extraversion (13 versus 9%) suggesting that with a denser SNP set more variance could be explained (see also Figure 1 and Supplementary Information Table S2).

Similar analyses of a number of other complex traits, for example, human height,⁴³ crystallised and fluid intelligence,⁵⁷ Crohn's Disease, Bipolar Disorder and Type 1 Diabetes,⁵⁸ Body Mass Index and biological measures such as von Willebrand Factor and QT_i⁵¹ have shown that common SNPs explain a substantial part of their reported heritabilities. If rarer variants are more likely to affect variance of neuroticism and extraversion, compared with the traits described above, relatively low heritability estimates would be expected from the current design (that is, genome-wide SNP analysis). A potential reason for a relatively large contribution of rare variants is the possible relation between neuroticism/extraversion and fitness. If personality traits neuroticism and extraversion are correlated with fitness, and decreased reproductive fitness reduces the frequencies of risk variants for less desirable personality traits, one would expect a relatively large contribution of rare variants to standing genetic variation for these traits. As the current data set primarily consists of SNPs with $MAF > 0.01$, rare (causal) variants are generally poorly tagged which, in turn, would explain some of the remaining heritability. Plausible arguments can be made that neuroticism and extraversion are correlated with fitness: mental disorders that are related to, or at the extreme end of, neuroticism and extraversion personality domains are associated with lower fertility, because of, for example, higher mortality rates and reduced mating opportunities.^{59,60} Correlations between fitness and personality domains neuroticism and extraversion, however, are likely to be very low.

A second reason for the lower heritability estimated from SNP data compared with pedigree data is that the latter may be biased upwards through confounding with nonadditive effects while, in contrast, the heritability estimates from SNP data are 'narrow sense' estimates, which reflect purely additive genetic effects. Nonadditive effects have been inconsistently reported for both neuroticism and extraversion with evidence provided from some studies,^{11–13,48,59–61} but not others.^{48,62} However, nonadditive effects are difficult to distinguish from common-environmental effects in the classical twin design.^{48,62,63} Consequently, common-environmental effects may well be masked by nonadditive effects in twin and family studies resulting in inflated estimates of additive genetic effects and deflated estimates of common-environmental effects. In a sample including both twin pairs reared apart and twin pairs reared together, Pedersen *et al.*,¹⁸ showed that common-environmental factors had a modest contribution on the variance of both neuroticism and extraversion while nonadditive genetic effects were absent in neuroticism but significant in extraversion.

Third, and related to the second, heritability estimates from twin and family studies are possibly inflated by common environment of family members because of deviations from assumptions that underpin the classical twin design such as equal common environmental influences for monozygotic and dizygotic twins, although credible arguments against this

special 'MZ twin environment' have been made.⁶⁴ In this study, the variance explained by all SNPs is estimated through such distant genetic relationships that confounding with common environment is unlikely. Furthermore, it is common practice in twin and family studies to 'drop' nonsignificant variance components from the model and subsequently report only the remaining significant variance components. In this way the common environmental term is often dropped but if it truly contributes to variance between individuals and if the reduced most parsimonious model reflects sample size and power, the reported estimates of heritability may be biased upwards with their *s.e.* biased downwards.

From twin and family studies we know that heritability estimates for both neuroticism and extraversion are moderate, with no major differences in estimates between the two dimensions. In the current design, however, we observe a larger estimate of variance explained for extraversion compared with neuroticism. A possible explanation for this discrepancy is that the phenotypic and biological complexity underlying neuroticism, is greater than for extraversion, causing diluted associations between genetic polymorphisms and neuroticism.⁶⁵ The personality dimension of neuroticism includes facets related to both anxiety and depression, which although genetically correlated,⁶⁶ may imply multiple biological dimensions. Other explanations for the lower estimate of variance explained by SNPs for neuroticism compared with extraversion may reflect differential impact of some of the reasons listed above. However, all explanations are speculative as yet.

The observed difference in estimates of variance explained based on a sample of 'unrelateds' (that is, not affected by common environmental influences and/or pedigree structure) and of variance explained based on a sample of 'relateds' (that is, possibly affected by common environmental influences and/or pedigree structure), especially for neuroticism, informs on the poor outcomes reported to date from GWAS of major depressive disorder.⁶⁷ Future research may consider possible environmental factors that dilute associations between genetic polymorphisms and the trait in the classic GWAS design.

In summary, a significant proportion of the phenotypic variance of neuroticism and extraversion can be explained by unknown causal variants in LD with common SNPs. This means that some of the so-called missing heritability of neuroticism and extraversion can be found in many common variants of small effects. The remaining discrepancy between our estimates from all the SNPs and the reported estimates of narrow sense heritability from pedigree studies might be attributed to (i) a large (*er*) role of rare variants, (ii) nonadditive effects (dominance and epistasis), and/or (iii) environmental influences that are included in the heritability estimate from pedigree studies but not in the heritability based on SNP data. Large samples with genome-wide sequencing data that are likely to become available with current developments in sequencing technologies may reveal the contribution of rare variants to the heritability of neuroticism and extraversion. Applying the current method to sequencing data could not only allow quantification of the contribution of rare variants, it should also be more conclusive about the total contribution of additive genetic effects to phenotypic variation in personality traits.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements. We would like to acknowledge the contribution of all the study participants without whom this work would not be possible. This work was supported by grants from the Australian National Health and Medical Research Council (NHMRC): 496688, 613608, the Australian Research Council (ARC): FT0991360. Genotyping of the UK cohorts was supported by the UK's Biotechnology and Biological Sciences Research Council (BBSRC). Phenotype collection in the Lothian Birth Cohort 1921 was supported by the BBSRC, The Royal Society and The Chief Scientist Office of the Scottish Government. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Research Into Ageing (which continues as part of Age UK's The Disconnected Mind project). Phenotype collection in the Aberdeen Birth Cohort 1936 was supported by BBSRC, the Wellcome Trust and the Alzheimer's Research Trust. Phenotype collection in the Manchester and Newcastle Longitudinal Studies of Cognitive Aging cohorts was supported by Social Science Research Council, Medical Research Council, Economic and Social Research Council, Research Into Ageing, the Wellcome Trust and Unilever plc. ML is a Royal Society of Edinburgh/Lloyds TSB Foundation for Scotland Personal Research Fellow. Data collection and genotyping for the Minnesota Center for Twin and Family Research samples was supported by the following grants from the United States Public Health Service: R01 DA05147, R01 AA09367, R01 AA11886, R01 DA13240 and U01 DA024417. Genotyping for the Swedish data was supported in part by grants from the United States NIH: U01 DK066134 and the Swedish Ministry of Higher Education.

- Costa PM. *RR. Professional Manual: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor-Inventory (NEO-FFI)*. Psychological Assessment Resources: Odessa FL, 1992.
- Eysenck HJ, Eysenck SBG. *Manual of the Eysenck Personality Questionnaire*. Hodder & Stoughton: London, 1975.
- Finkel D, McGue M. Sex differences and nonadditivity in heritability of the Multidimensional Personality Questionnaire Scales. *J Pers Soc Psychol* 1997; **72**: 929–938.
- Tellegen A, Waller NG. Exploring personality through test construction: development of the multidimensional personality questionnaire. In: Boyle GJ, Matthews G, Saklofske DH (eds). *The SAGE Handbook of Personality Theory and Assessment: Vol. 2. Personality Measurement and Testing*. Sage: London, 2008, pp 261–292.
- Saulsman LM, Page AC. The five-factor model and personality disorder empirical literature: A meta-analytic review. *Clin Psychol Rev* 2004; **23**: 1055–1085.
- Samuel DB, Widiger TA. A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: a facet level analysis. *Clin Psychol Rev* 2008; **28**: 1326–1342.
- Hettema JM, Neale MC, Myers JM, Prescott CA, Kendler KS. A population-based twin study of the relationship between neuroticism and internalizing disorders. *Am J Psychiatry* 2006; **163**: 857–864.
- 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. *Depress Anxiety* 2004; **20**: 92–97.
- Terracciano A, Lockenhoff CE, Zonderman AB, Ferrucci L, Costa Jr PT. Personality predictors of longevity: activity, emotional stability, and conscientiousness. *Psychosom Med* 2008; **70**: 621–627.
- Distel MA, Trull TJ, Willemssen G, Vink JM, Derom CA, Lynskey M *et al*. The five-factor model of personality and borderline personality disorder: a genetic analysis of comorbidity. *Biol Psychiatry* 2009; **66**: 1131–1138.
- Keller MC, Coventry WL, Heath AC, Martin NG. Widespread evidence for non-additive genetic variation in Cloninger's and Eysenck's personality dimensions using a twin plus sibling design. *Behav Genet* 2005; **35**: 707–721.
- Lake RL, Eaves LJ, Maes HH, Heath AC, Martin NG. Further evidence against the environmental transmission of individual differences in neuroticism from a collaborative study of 45 850 twins and relatives on two continents. *Behav Genet* 2000; **30**: 223–233.
- Eaves L, Heath A, Martin N, Maes H, Neale M, Kendler K *et al*. Comparing the biological and cultural inheritance of personality and social attitudes in the Virginia 30 000 study of twins and their relatives. *Twin Res* 1999; **2**: 62–80.
- Loehlin J, McCrae RR, Costa PT, John OP. Heritabilities of common and measure-specific components of the big five personality factors. *J Res Pers* 1988; **32**: 431–453.
- Birley AJ, Gillespie NA, Heath AC, Sullivan PF, Boomsma DI, Martin NG. Heritability and nineteen-year stability of long and short EPQ-R Neuroticism scales. *Pers Individ Dif* 2006; **40**: 737–747.

16. Wray NR, Birmaher BJ, Sullivan PF, Visscher PM, Martin NG. Genetic and phenotypic stability of measures of neuroticism over 22 years. *Twin Res Hum Genet* 2007; **10**: 695–702.
17. Johnson AM, Vernon PA, Feiler AR. Behavioral genetic studies of personality: An introduction and review of the results of 50+ years of research. In: Boyle G, Matthews G, Saklofske D (eds). *Handbook of personality theory and assessment*. Sage: London, 2008 pp 145–173.
18. Pedersen NL, Plomin R, McClearn GE, Friberg L. Neuroticism, extraversion, and related traits in adult twins reared apart and reared together. *J Pers Soc Psychol* 1988; **55**: 950–957.
19. Dick DM, Aliev F, Wang JC, Gruzca RA, Schuckit M, Kuperman S et al. Using dimensional models of externalizing psychopathology to aid in gene identification. *Arch Gen Psychiatry* 2008; **65**: 310–318.
20. Fanous A, Gardner CO, Prescott CA, Cancro R, Kendler KS. Neuroticism, major depression and gender: a population-based twin study. *Psychol Med* 2002; **32**: 719–728.
21. Wray NR, Middeldorp CM, Birmaher BJ, Gordon SD, Sullivan PF, Visscher PM et al. Genome-wide linkage analysis of multiple measures of neuroticism of 2 large cohorts from Australia and the Netherlands. *Arch Gen Psychiatry* 2008; **65**: 649–658.
22. Fullerton J, Cubin M, Tiwari H, Wang C, Bomhra A, Davidson S et al. Linkage analysis of extremely discordant and concordant sibling pairs identifies quantitative-trait loci that influence variation in the human personality trait neuroticism. *Am J Hum Genet* 2003; **72**: 879–890.
23. Nash MW, Huezo-Diaz P, Williamson RJ, Sterne A, Purcell S, Hoda F et al. Genome-wide linkage analysis of a composite index of neuroticism and mood-related scales in extreme selected sibships. *Hum Mol Genet* 2004; **13**: 2173–2182.
24. Kuo PH, Neale MC, Riley BP, Patterson DG, Walsh D, Prescott CA et al. A genome-wide linkage analysis for the personality trait neuroticism in the Irish affected sib-pair study of alcohol dependence. *Am J Med Genet B Neuropsychiatr Genet* 2007; **144B**: 463–468.
25. Neale BM, Sullivan PF, Kendler KS. A genome scan of neuroticism in nicotine dependent smokers. *Am J Med Genet B Neuropsychiatr Genet* 2005; **132B**: 65–69.
26. Gillespie NA, Zhu G, Evans DM, Medland SE, Wright MJ, Martin NG. A genome-wide scan for Eysenckian personality dimensions in adolescent twin sibships: psychoticism, extraversion, neuroticism, and lie. *J Pers* 2008; **76**: 1415–1446.
27. Amin N, Schuur M, Gusareva ES, Isaacs A, Aulchenko YS, Kirichenko AV et al. A genome-wide linkage study of individuals with high scores on NEO personality traits. *Mol Psychiatry* 2011; doi:10.1038/mp.2011.97.
28. Loehlin JC, Medland SE, Montgomery GW, Martin NG. Eysencks Psychoticism and the X-linked androgen receptor gene CAG polymorphism in additional Australian samples. *Pers Individ Dif* 2005; **39**: 7.
29. Sen S, Burmeister M, Ghosh D. Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits. *Am J Med Genet B Neuropsychiatr Genet* 2004; **127B**: 85–89.
30. Schinka JA, Busch RM, Robichaux-Keene N. A meta-analysis of the association between the serotonin transporter gene polymorphism (5-HTTLPR) and trait anxiety. *Mol Psychiatry* 2004; **9**: 197–202.
31. Juhász G, Chase D, Pegg E, Downey D, Toth ZG, Stones K et al. CNR1 gene is associated with high neuroticism and low agreeableness and interacts with recent negative life events to predict current depressive symptoms. *Neuropsychopharmacology* 2009; **34**: 2019–2027.
32. Zhou Z, Zhu G, Hariri AR, Enoch MA, Scott D, Sinha R et al. Genetic variation in human NPY expression affects stress response and emotion. *Nature* 2008; **452**: 997–1001.
33. Cotton CH, Flint J, Campbell TG. Is there an association between NPY and neuroticism? *Nature* 2009; **458**: E6, discussion E7.
34. Paterson AD, Sunohara GA, Kennedy JL. Dopamine D4 receptor gene: novelty or nonsense? *Neuropsychopharmacology* 1999; **21**: 3–16.
35. Munafò MR, Clark TG, Moore LR, Payne E, Walton R, Flint J. Genetic polymorphisms and personality in healthy adults: a systematic review and meta-analysis. *Mol Psychiatry* 2003; **8**: 471–484.
36. de Moor MH, Costa PT, Terracciano A, Krueger RF, de Geus EJ, Toshiko T et al. Meta-analysis of genome-wide association studies for personality. *Mol Psychiatry* 2010; **17**: 337–349.
37. Terracciano A, Sanna S, Uda M, Deiana B, Usala G, Busonero F et al. Genome-wide association scan for five major dimensions of personality. *Mol Psychiatry* 2010; **15**: 647–656.
38. van den Oord EJ, Kuo PH, Hartmann AM, Webb BT, Moller HJ, Hettema JM et al. Genome-wide association analysis followed by a replication study implicates a novel candidate gene for neuroticism. *Arch Gen Psychiatry* 2008; **65**: 1062–1071.
39. Calboli FC, Tozzi F, Galwey NW, Antoniadis A, Mooser V, Preisig M et al. A genome-wide association study of neuroticism in a population-based sample. *PLoS One* 2010; **5**: e11504.
40. Verweij KJ, Zietsch BP, Medland SE, Gordon SD, Benjamin B, Nyholt DR et al. A genome-wide association study of Cloninger's temperament scales: implications for the evolutionary genetics of personality. *Biol Psychol* 2010; **85**: 306–317.
41. Maher B. Personal genomes: The case of the missing heritability. *Nature* 2008; **456**: 18–21.
42. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ et al. Finding the missing heritability of complex diseases. *Nature* 2009; **461**: 747–753.
43. 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**: 565–569.
44. Wray NR. Allele frequencies and the r^2 measure of linkage disequilibrium: impact on design and interpretation of association studies. *Twin Res Hum Genet* 2005; **8**: 87–94.
45. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**: 76–82.
46. Rettew DC, Vink JM, Willemsen G, Doyle A, Hudziak JJ, Boomsma DI. The genetic architecture of neuroticism in 3301 Dutch adolescent twins as a function of age and sex: a study from the Dutch twin register. *Twin Res Hum Genet* 2006; **9**: 24–29.
47. Eaves L, Eysenck H. The nature of extraversion: a genetic analysis. *J Pers Soc Psychol* 1975; **32**: 102–112.
48. Floderus-Myrhed B, Pedersen N, Rasmuson I. Assessment of heritability for personality, based on a short-form of the Eysenck Personality Inventory: a study of 12 898 twin pairs. *Behav Genet* 1980; **10**: 153–162.
49. Eaves LJ, Heath AC, Neale MC, Hewitt JK, Martin NG. Sex differences and non-additivity in the effects of genes on personality. *Twin Res* 1998; **1**: 131–137.
50. Bouchard Jr TJ, Loehlin JC. Genes, evolution, and personality. *Behav Genet* 2001; **31**: 243–273.
51. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 2011; **43**: 519–525.
52. Visscher PM, Yang J, Goddard ME. A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang et al. (2010). *Twin Res Hum Genet* 2010; **13**: 517–524.
53. Visscher PM, Medland SE, Ferreira MA, Morley KI, Zhu G, Cornes BK et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet* 2006; **2**: e41.
54. Visscher PM, Macgregor S, Benjamin B, Zhu G, Gordon S, Medland S et al. Genome partitioning of genetic variation for height from 11 214 sibling pairs. *Am J Hum Genet* 2007; **81**: 1104–1110.
55. Verweij KJH, Yang J, Lahti J, Veijola J, Hintsanen M, Pulkki-Råback L et al. Maintenance of genetic variation in personality: Estimating heritability due to common causal variants and investigating the effect of inbreeding. *Evolution* 2012 (in press).
56. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB et al. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 2010; **42**: 348–354.
57. 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**: 996–1005.
58. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* 2011; **88**: 294–305.
59. Rettew DC, Rebollo-Mesa I, Hudziak JJ, Willemsen G, Boomsma DI. Non-additive and additive genetic effects on extraversion in 3314 Dutch adolescent twins and their parents. *Behav Genet* 2008; **38**: 223–233.
60. Eaves LJ, Martin NG, Heath AC, Hewitt JK, Neale MC. Personality and reproductive fitness. *Behav Genet* 1990; **20**: 563–568.
61. Alvergne A, Jokela M, Lummaa V. Personality and reproductive success in a high-fertility human population. *Proc Natl Acad Sci USA* 2010; **107**: 11745–11750.
62. Hill WG, Goddard ME, Visscher PM. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* 2008; **4**: e1000008.
63. Gillespie NA, Cloninger CR, Heath AC, Martin NG. The genetic and environmental relationship between Cloninger's dimensions of temperament and character. *Pers Individ Dif* 2003; **35**: 15.
64. Tellegen A, Lykken DT, Bouchard Jr TJ, Wilcox KJ, Segal NL, Rich S. Personality similarity in twins reared apart and together. *J Pers Soc Psychol* 1988; **54**: 1031–1039.
65. van der Sluis S, Verhage M, Posthuma D, Dolan CV. Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. *PLoS One* 2010; **5**: e13929.
66. Middeldorp CM, Cath DC, Van Dyck R, Boomsma DI. The co-morbidity of anxiety and depression in the perspective of genetic epidemiology. A review of twin and family studies. *Psychol Med* 2005; **35**: 611–624.
67. Wray NR, Pergadia ML, Blackwood DH, Penninx BW, Gordon SD, Nyholt DR et al. Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol Psychiatry* 2012; **17**: 13.



Translational Psychiatry is an open-access journal published by Nature Publishing Group. This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

Supplementary Information accompanies the paper on the Translational Psychiatry website (<http://www.nature.com/tp>)