Genome-wide association scan for five major dimensions of personality

A Terracciano¹, S Sanna², M Uda², B Deiana², G Usala², F Busonero², A Maschio², M Scally¹, N Patriciu¹, W-M Chen³, MA Distel⁴, EP Slagboom⁵, DI Boomsma⁶, S Villafuerte⁶,⁷, E Śliwerska⁶, M Burmeister⁶,⁷,⁸, N Amin⁹,¹⁰, ACJW Janssens⁹, CM van Duijn⁹,¹⁰, D Schlessinger¹, GR Abecasis¹¹ and PT Costa Jr¹

¹National Institute on Aging, NIH, Baltimore, MD, USA; ²Istituto di Neurogenetica e Neurofarmacologia, Consiglio Nazionale delle Ricerche, Cagliari, Italy; ³Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA; ⁴Department of Biological Psychology, VU University, Amsterdam, The Netherlands; ⁵Molecular Epidemiology Section, Leiden University Medical Centre, Leiden, The Netherlands; ⁶Molecular & Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI, USA; ⁷Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA; ⁸Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA; ⁹Department of Epidemiology & Biostatistics, Erasmus University Medical Center, Rotterdam, The Netherlands; ¹⁰Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands and ¹¹Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA

Personality traits are summarized by five broad dimensions with pervasive influences on major life outcomes, strong links to psychiatric disorders and clear heritable components. To identify genetic variants associated with each of the five dimensions of personality we performed a genome-wide association (GWA) scan of 3972 individuals from a genetically isolated population within Sardinia, Italy. On the basis of the analyses of 362129 single-nucleotide polymorphisms we found several strong signals within or near genes previously implicated in psychiatric disorders. They include the association of neuroticism with SNAP25 (rs362584, P=2×10⁻³), extraversion with BDNF and two cadherin genes (CDH13 and CDH23; P<5×10⁻⁵), openness with CNTNAP2 (rs10251794, P=3×10⁻⁵), agreeableness with CLOCK (rs6832769, P=9×10⁻⁵) and conscientiousness with DYRK1A (rs2835731, P=3×10⁻⁶). Effect sizes were small (less than 1% of variance), and most failed to replicate in the follow-up independent samples (N up to 3903), though the association between agreeableness and CLOCK was supported in two of three replication samples (overall P=2×10⁻⁵). We infer that a large number of loci may influence personality traits and disorders, requiring larger sample sizes for the GWA approach to confidently identify associated genetic variants.

Keywords: personality; genome-wide association; founder population; psychiatry; five-factor model

Introduction

Behavior genetic studies reveal that personality traits, such as psychiatric disorders, have a genetic basis. Twin, adoption and family studies indicate that personality factors are heritable, with about 50% of the variance of the underlying components accounted for by additive and nonadditive genetic factors.¹⁻³ However, identifying the genetic variants associated with personality traits is challenging. A large number of studies have tested several candidate genes, especially for the neuroticism factor, but these studies have produced largely inconclusive results. Similarly, genetic linkage studies, usually based on 400–500 microsatellite markers, have suggested loci for neuroticism,⁴⁻⁷ but only a few genomic regions (for example, 12q) have been reported in multiple studies. Difficulties in identifying specific loci suggest that, as has been observed for a number of quantitative traits, genetic influences on these complex traits are likely attributable to many genes, each with a small effect size. To detect such small genetic effects, there is growing interest in high-throughput genotyping technologies that examine large numbers (for example, 500 000) of single-nucleotide polymorphisms (SNPs) densely mapped across the entire genome. Using these more incisive genome-wide association (GWA)
scans, recent studies have reliably identified common variants associated with complex traits and diseases that include body mass index, height, inflammatory bowel disease, and type 1 and type 2 diabetes. For personality traits, a GWA study used DNA pools from about 2000 individuals with extreme scores, but the results were limited to neuroticism, only one of the five broad dimensions of personality.

Personality profiles assessed with the Revised NEO Personality Inventory (NEO-PI-R) can be conveniently summarized by five major dimensions. Neuroticism (N), the tendency to experience negative emotions such as anxiety, anger and depression; extraversion (E), the tendency to be sociable, warm, active, assertive, cheerful, and in search of stimulation; openness to experience (O), the tendency to be imaginative, creative, unconventional, emotionally and artistically sensitive; agreeableness (A), the dimension of interpersonal relations, characterized by altruism, trust, modesty and cooperativeness; and conscientiousness (C), a tendency to be organized, strong-willed, persistent, reliable and a follower of rules and ethical principles. Consistent with their biological basis, these five dimensions can be clearly replicated in most cultures.

Furthermore, the five factors are predictors of important life outcomes, including well-being, academic performance, vocational interests, marital stability and satisfaction, health risk behaviors, and longevity. All five factors are related to personality disorders, and several researchers advocate a dimensional model in the upcoming Diagnostic and Statistical Manual of Mental Disorders (DSM-V) to reorganize the conceptualization and diagnosis of personality disorders. More generally, personality traits are thought of as risk factors, diagnostic indicators, and predictors of onset, severity, and outcome for most psychiatric disorders. These phenotypic links are supported at the genetic level; twin studies indicate that personality traits share a large proportion of genetic variance with depression and other disorders. Personality traits are increasingly recognized as endophenotypes in genetic studies of mental disorders.

The present study provides the first GWA results for all five dimensions of personality, as measured by the NEO-PI-R. This study is part of the SardiNIA project, which has targeted a highly interrelated population from the isolated Ogliastra region of Sardinia, Italy. Common variants associated with several complex traits have been successfully identified in this sample, and the results replicated across diverse populations. The studies of population isolates are advantageous because these groups are highly homogeneous, reducing the risk of spurious associations due to population stratification. Furthermore, population isolates have more extensive stretches of linkage disequilibrium compared to outbred populations. This extended LD should increase the genome-wide coverage of standard SNP arrays in cohorts such as our Sardinian sample. The advantages of having reduced variability in a founder population come at the cost of lower power to replicate the effects in more heterogeneous populations. Several findings from this Sardinian cohort have been replicated in other populations.

Here we report associations of the five personality factors with 362 129 SNPs in 3972 Sardinians. SNPs most strongly associated with each factor were genotyped in independent samples to look at the replicability of the findings in independent populations.

Materials and methods

Sample description

We recruited 6148 individuals, about 62% of the population aged 14–102 years, from a cluster of four towns in the Lanusei Valley. Subjects are native-born, and at least 95% are known to have all grandparents born in the same province. Valid personality data were obtained from 5669 subjects, of which 3972 were genotyped. The sample was composed of 2250 women and 1722 men (43.3%). Age ranged from 14 to 94 years (M = 42.8, s.d. = 17).

Personality assessment

Personality traits were assessed using the Italian version of the NEO-PI-R, a 240-item measure of the five dimensions of personality. The domain scores are computed by summing up the six facets that compose each factor. Items are answered on a five-point Likert scale, from strongly disagree to strongly agree, and scales are roughly balanced to control for the effect of acquiescence. The NEO-PI-R has a robust factor structure that has been replicated in Italy and in more than 50 cultures. Scales have shown longitudinal stability, cross-observer agreement, and convergent and discriminant validity in a large body of studies. Trained Sardinian psychologists administered the tests. In the Sardinian sample, the NEO-PI-R showed good psychometric properties, with internal consistency reliabilities for the five factors ranging from 0.80 to 0.87, and a factor structure that replicated the American normative structure at the phenotypic and the genetic level.

Genotyping and imputation

DNA was extracted from blood samples. In the Sardinian cohort, 3329 and 1412 individuals were genotyped with the Affymetrix 10K and Affymetrix 500K Mapping array set, respectively, with 436 individuals generating an overlapping dataset. We
took advantage of the relatedness among individuals in our sample to reduce study costs. Using a modified Lander–Green algorithm, full genotypes on the 2893 individuals typed with only the 10K panel were imputed based on stretches of shared haplotype, permitting analyses on 4305 individuals, of which 3972 had personality data. For individuals who had genotype data available at the SNP being tested, we coded genotypes as 0, 1 or 2, depending on the number of copies of an arbitrary reference allele for each SNP. For individuals with missing genotype data, we used the Lander–Green algorithm to estimate the number of copies of the allele carried by each individual (based on the genotypes of family members) and assigned each individual a score ranging between 0 and 2. This estimate incorporates allele frequency information, the genotypes of relatives for the SNP of interest, and flanking marker data. For computational efficiency, the Lander–Green algorithm was applied to sub-pedigrees, each including no more than 20–25 individuals, resulting in a dataset where the average analysis unit consisted of a family with 12.3 members and 3.2 generations.

GWA analysis
Of the combined 500 and 10K Mapping array sets, the association analyses focused on 362 129 SNPs that passed quality control checks. The remaining SNPs failed quality checks (~2.9% of SNPs failed checks for data completeness, Hardy–Weinberg equilibrium and Mendelian incompatibilities) or had a minor allele frequency of <5% (~25.7% of SNPs had low minor allele frequencies). Although the five-factor scores are approximately normally distributed, to avoid inflated type I error rates an inverse normal transformation was applied to all phenotype variables before analysis. Association analyses were carried out as described elsewhere. The additive genetic effect was estimated for each SNP in the context of a variance component model that accounts for resemblance among related individuals. We analyzed each of the five-factor scores in turn, including sex, age and age as covariates.

Our analytical approach considers all observed or estimated genotypes (rather than focusing on alleles transmitted from heterozygous parents) and thus is not immune to effects of population stratification. In homogenous populations, this type of analysis is expected to be more powerful. To adjust for the effects of population structure and cryptic relatedness among sampled individuals, we used the genomic control method to adjust our test statistics for each trait separately. We checked the genomic control value for our GWA analyses and carried out principal component analysis of genome-wide SNP data in a subset of unrelated individuals. Neither analysis suggested evidence for population substructure or genetic outliers in the sample.

To evaluate association on the X chromosome, we modeled a polygenic variance component shared according to an X-linked kinship coefficient in addition to the usual autosomal polygenic variance component. Further, we assumed that average phenotypic values for hemizygous males would be the same as for homozygous females.

Meta-analysis
We use meta-analysis to summarize the results from the Sardinia and replication samples. The overall z-statistic and the corresponding P-value were calculated as a weighted average, where weights were proportional to the square root of the number of individuals examined in each sample and selected such that the squared weights summed to 1.

Replication samples
We attempted to replicate six top signals for each of the five factors in two independent samples. The first replication sample was from Tecumseh, MI, USA. Complete data were obtained for 923 individuals (age: M = 44.3, s.d. = 15.7; 58% women), which consists of 110 unrelated individuals and 813 individuals clustering in 424 families. Personality traits were assessed using the NEO-PI, an earlier version of the NEO-PI-R, which assesses the agreeableness and conscientiousness factors using 18-item scales instead of the NEO-PI-R’s 48-item scales. A total of 25 SNPs were successfully genotyped using TaqMan SNP genotyping assays (Applied Biosystems, Foster City, CA, USA). The second replication sample came from the Netherlands and consisted of participants from the Netherlands Twin Register. The analyses for this study were conducted on 1158 individuals from 418 families (age: M = 45.3, s.d. = 14.6; 61% women). Personality traits were assessed using a validated Dutch version of the NEO-FFI, a short version of the NEO-PI-R that assesses each factor using 12-item scales. SNPs were genotyped using Sequenom technology.

The results from the GWA analyses and the two replication samples prompted us to examine a single SNP in one additional sample. The third follow-up sample consisted of the 1822 participants of the Erasmus Rucphen Family study in whom the SNP had been genotyped and for whom personality data were available (age: M = 48.0, s.d. = 14.5; 58% women). Personality was assessed using the NEO-FFI and the SNP was genotyped using TaqMan.

Results
We present the results of a GWA scan for five broad personality factors in a founder population from Sardinia, Italy. Although none of the initial results reaches genome-wide significance using the conservative Bonferroni threshold, several interesting candidate genes map near SNPs exhibiting strong evidence of association—prompting us to examine SNPs with the strongest signals in additional samples. Even when none of the SNPs examined reach genome-wide significance in the original scan, we expect that sets of SNPs showing nominally strong
associations will be enriched for truly associated SNPs.\textsuperscript{54,55} SNPs with \textit{P}-values lower or equal to $10^{-5}$ are presented in Supplementary Table A (45 SNPs associated with neuroticism, 54 with extraversion, 59 with openness, 112 with agreeableness and 33 with conscientiousness). Table 1 presents the top six SNPs that we sought to replicate in two independent samples. For each trait, we selected six top non-redundant SNPs and excluded those that were both outside of any gene and not in LD with any other surrounding SNPs. Table 1 provides \textit{P}-values and a standardized measure of effect size (\textit{Z}), with the sign (+ or −) indicating the direction of the effect.

\textbf{Neuroticism}

Of the SNPs that showed the strongest association with neuroticism, rs362584 ($P = 5.03 \times 10^{-5}$) is attractive, because it is within intron 1 of the gene synaptosomal-associated protein of 25 kDa (SNAP25). SNAP25 is important in neurotransmitter release, axonal growth and synaptic plasticity.\textsuperscript{56} Deletion of the region containing the SNAP25 gene in the Coloboma mouse causes neurological abnormalities, including hyperactivity. Such a phenotype is consistent with a role for SNAP25 in attention deficit hyperactivity disorder, which has been tested in a number of studies with multiple positive results.\textsuperscript{57,58} Furthermore, abnormalities in the level of SNAP25 have been linked to other psychiatric disorders,\textsuperscript{59–61} and genetic variants in SNAP25 have been associated with cognitive ability.\textsuperscript{62} In addition to its intrinsic interest for a personality trait, the association between SNAP25 and neuroticism is relevant to several psychiatric disorders for which neuroticism is an intermediate phenotype/endophenotype. In the follow-up samples we found a trend, in the same direction, for the association of SNAP25 with neuroticism ($P = 0.097$). A SNP in the gene \textit{TMEM16D} (rs1849710; $P = 2.29 \times 10^{-3}$), maps in the 12q region relatively close to the marker D12S346, which showed the strongest linkage peak in a extremely discordant and concordant sibling pairs study, particularly among female pairs.\textsuperscript{4} To examine sex-specific effects in the Sardinian sample, we conducted additional association analyses in women and men separately (sex-specific \textit{P}-values for the SNPs in Table 1 are presented in Supplementary Table B). Results indicate that rs1849710 on 12q was associated with neuroticism in women ($P = 4.34 \times 10^{-3}$) and also in men ($P = 4.39 \times 10^{-3}$). No sex-specific effects reached genome-wide significance using the stringent Bonferroni correction.

\textbf{Extraversion}

In an interesting pattern for this factor, multiple independent SNPs within two Cadherin genes (\textit{CDH13} and \textit{CDH23}) have the strongest association with extraversion, though none of these effects were replicated in the two independent samples. The Cadherin genes encode for cell–cell adhesion proteins that form complexes crucial in regulating synapse formation, function and plasticity.\textsuperscript{63–65} CDH23 is expressed in neurosensory epithelium and linked to some instances of deafness.\textsuperscript{66} CDH13 is expressed in the heart and several brain tissues, where it is thought to act as a regulator of neural cell growth.\textsuperscript{57} For rs904208 we found a trend ($P = 0.065$) in the replication samples consistent with the effect found in Sardinia. For rs2813838 there was a consistent effect in the two follow-up samples, but unfortunately the effect in the SardiNIA GWA was in the opposite direction (see signs of \textit{Z}-scores in Table 1). In addition to the SNPs shown in Table 1, a further interesting association is with rs11030064 ($P = 8.05 \times 10^{-5}$), a SNP lying close to the brain-derived neurotrophic factor (BDNF) gene, and an association was also found with the extensively studied variation at Val66Met (rs6265; $P = 0.0016$). Several signals were also seen in the region of RAB3GAP1 (rs16833135; $P = 8.05 \times 10^{-5}$), which encodes a protein implicated in the exocytosis of neurotransmitters and hormones; in GFRA1, the gene for a glial cell line-derived neurotrophic factor (rs4562724; $P = 9.45 \times 10^{-5}$); and in DCAMKL1 (rs17786591; $P = 2.97 \times 10^{-5}$), encoding a doublecortin and CaM kinase-like protein.

\textbf{Openness}

The most significant effect across the five factors was found between rs644148 and openness ($P = 9.44 \times 10^{-7}$). This same SNP was strongly associated with extraversion as well. Although openness and extraversion are correlated, we observed only a modest overlap between association results for the two traits. For example, among markers that were associated with extraversion at $P < 0.001$, less than 5\% were also associated with openness at a similar significance level (and vice versa). The overlap was even more reduced at more stringent significance levels, indicating that the association between rs644148 and the two traits is quite exceptional. We were unable to type this SNP in the replication samples, but a nearby SNP (rs565819; LD = 1) was genotyped in the US sample and no association was found. Among the other high-ranked associations for openness there is an intriguing association with \textit{CLOCK} gene (Figure 1). The strongest signal was with rs6832769 ($P = 8.71 \times 10^{-9}$). The two follow-up samples showed similar effects that reached statisti-
Table 1 Top association identified in the SardiNIA GWA analyses and tested in two follow-up samples

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Chr</th>
<th>Position</th>
<th>Allele</th>
<th>Sardinia</th>
<th>Replication: US</th>
<th>Replication: Netherlands</th>
<th>Combined replication</th>
<th>Combined all</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs644148</td>
<td>ZNF180</td>
<td>19</td>
<td>49662775</td>
<td>G</td>
<td>8.03E-06 + 0.136</td>
<td>*</td>
<td>NA</td>
<td>9.44E-07</td>
<td></td>
</tr>
<tr>
<td>rs6610953</td>
<td>FUNDIC1</td>
<td>X</td>
<td>44156440</td>
<td>G</td>
<td>1.73E-06 + 0.137</td>
<td>0.237 + 0.033</td>
<td>0.23 + 0.009</td>
<td>NA</td>
<td>5.74E-06</td>
</tr>
<tr>
<td>rs17819128</td>
<td>CREBL2</td>
<td>12</td>
<td>12652926</td>
<td>G</td>
<td>3.02E-06 + 0.173</td>
<td>0.202 + 0.052</td>
<td>0.263 + 0.036</td>
<td>0.466</td>
<td>1.27E-04</td>
</tr>
<tr>
<td>rs1037791</td>
<td>TSPAN13</td>
<td>7</td>
<td>16597902</td>
<td>A</td>
<td>3.94E-06 + 0.149</td>
<td>0.364 + 0.004</td>
<td>0.440 + 0.008</td>
<td>0.359E-05</td>
<td></td>
</tr>
<tr>
<td>rs586281</td>
<td></td>
<td>1</td>
<td>18293114</td>
<td>C</td>
<td>6.18E-06 + 0.178</td>
<td>0.118 + 0.072</td>
<td>0.387 + 0.015</td>
<td>0.283</td>
<td>6.38E-05</td>
</tr>
</tbody>
</table>

Agoraphobia

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Chr</th>
<th>Position</th>
<th>Allele</th>
<th>Sardinia</th>
<th>Replication: US</th>
<th>Replication: Netherlands</th>
<th>Combined replication</th>
<th>Combined all</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1380251</td>
<td></td>
<td>1</td>
<td>21795750</td>
<td>G</td>
<td>1.64E-06 + 0.235</td>
<td>NA</td>
<td>0.375 + 0.023</td>
<td>1.25E-05</td>
<td></td>
</tr>
</tbody>
</table>

Neuroticism

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Chr</th>
<th>Position</th>
<th>Allele</th>
<th>Sardinia</th>
<th>Replication: US</th>
<th>Replication: Netherlands</th>
<th>Combined replication</th>
<th>Combined all</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs644148</td>
<td>ZNF180</td>
<td>19</td>
<td>49662775</td>
<td>G</td>
<td>8.03E-06 + 0.136</td>
<td>*</td>
<td>NA</td>
<td>9.44E-07</td>
<td></td>
</tr>
<tr>
<td>rs6610953</td>
<td>FUNDIC1</td>
<td>X</td>
<td>44156440</td>
<td>G</td>
<td>1.73E-06 + 0.137</td>
<td>0.237 + 0.033</td>
<td>0.23 + 0.009</td>
<td>NA</td>
<td>5.74E-06</td>
</tr>
<tr>
<td>rs17819128</td>
<td>CREBL2</td>
<td>12</td>
<td>12652926</td>
<td>G</td>
<td>3.02E-06 + 0.173</td>
<td>0.202 + 0.052</td>
<td>0.263 + 0.036</td>
<td>0.466</td>
<td>1.27E-04</td>
</tr>
<tr>
<td>rs1037791</td>
<td>TSPAN13</td>
<td>7</td>
<td>16597902</td>
<td>A</td>
<td>3.94E-06 + 0.149</td>
<td>0.364 + 0.004</td>
<td>0.440 + 0.008</td>
<td>0.359E-05</td>
<td></td>
</tr>
<tr>
<td>rs586281</td>
<td></td>
<td>1</td>
<td>18293114</td>
<td>C</td>
<td>6.18E-06 + 0.178</td>
<td>0.118 + 0.072</td>
<td>0.387 + 0.015</td>
<td>0.283</td>
<td>6.38E-05</td>
</tr>
</tbody>
</table>

Conscientiousness

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Chr</th>
<th>Position</th>
<th>Allele</th>
<th>Sardinia</th>
<th>Replication: US</th>
<th>Replication: Netherlands</th>
<th>Combined replication</th>
<th>Combined all</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1362632</td>
<td>SMOC1</td>
<td>14</td>
<td>69557149</td>
<td>C</td>
<td>4.82E-06 + 0.175</td>
<td>0.028 + 0.134</td>
<td>0.473 + 0.004</td>
<td>0.112</td>
<td>9.95E-06</td>
</tr>
<tr>
<td>rs1095355</td>
<td>LAMBI</td>
<td>7</td>
<td>10717529</td>
<td>C</td>
<td>1.20E-05 + 0.140</td>
<td>0.155 + 0.050</td>
<td>0.346 + 0.017</td>
<td>0.166</td>
<td>3.86E-05</td>
</tr>
<tr>
<td>rs1700684</td>
<td>MBP1SC8</td>
<td>4</td>
<td>84734382</td>
<td>C</td>
<td>1.56E-05 + 0.262</td>
<td>0.262 + 0.051</td>
<td>0.495 + 0.002</td>
<td>0.332</td>
<td>1.56E-05</td>
</tr>
<tr>
<td>rs2835731</td>
<td>DYSK1A</td>
<td>21</td>
<td>37735051</td>
<td>T</td>
<td>2.81E-05 + 0.208</td>
<td>0.262 + 0.051</td>
<td>0.495 + 0.002</td>
<td>0.332</td>
<td>1.56E-05</td>
</tr>
<tr>
<td>rs13070781</td>
<td>EPER1</td>
<td>3</td>
<td>71364676</td>
<td>A</td>
<td>2.82E-05 + 0.141</td>
<td>0.201 + 0.043</td>
<td>0.706 + 0.068</td>
<td>0.461</td>
<td>1.56E-02</td>
</tr>
<tr>
<td>rs10945200</td>
<td>COL19A1</td>
<td>6</td>
<td>70948461</td>
<td>C</td>
<td>3.05E-05 + 0.141</td>
<td>0.174 + 0.044</td>
<td>0.141 + 0.017</td>
<td>0.097</td>
<td>8.12E-04</td>
</tr>
</tbody>
</table>

NA, not available.

Sample size: Sardinia N = 3972; United States N = 923; Netherlands N = 1158. Follow-up samples and combined replication P-values are from one tailed tests. All positions refer to May 2004 genome assembly. The effect size ‘Z’ is measured in standard deviation units (so that an effect of 0.1 indicates that each additional copy of the allele increases trait values by 0.1 standard deviations, on average), with the sign (+ or -) indicating the direction of the effect.

*In the US sample, for technical reasons we did not genotype the ZNF180 SNP rs644148 but the nearby rs565819 (with E: P = 0.12; with O: P = 0.44). In the Dutch sample instead of THUMP2 rs2540226 we genotyped the nearby rs1861243 (with A: P = 0.39) and instead of COL19A1 rs10945200 we genotyped the nearby rs3860652 (with C: P = 0.27).

For rs6832769, in addition to analyzing genotypic individuals, we used 44 flanking microsatellites to impute genotypes for additional family members. The ‘Combined all’ value in the table includes the SardiNIA GWA and the two follow-up samples. When the third follow-up sample is combined (N = 7875), the association is reduced to P = 8.6 x 10^-5.

rs17006841 was merged into rs3182340.
Beyond the SNPs reported in Table 1, other strong signals in genes with plausible biological relevance are OPCML (rs11223249; \( P = 3.52 \times 10^{-6} \)) an opioid-binding protein; CTNNA2 (rs2861913; \( P = 6.95 \times 10^{-3} \)), the \( \alpha \)-N-catenin that interacts with cadherin proteins in essential brain functions;\textsuperscript{63–65} and IKBAP (rs10118853; \( P = 5.11 \times 10^{-6} \)), a gene that causes familial dysautonomia, a sensory and autonomic neuropathy.

**Conscientiousness**

The strongest signal for conscientiousness was with a SNP within the gene SMOCI (rs11626232; \( P = 4.82 \times 10^{-6} \)), which also showed a meta-analytic trend in the replication samples (\( P = 0.11 \)). Conscientiousness was strongly associated with rs2835731 (\( P = 2.81 \times 10^{-5} \)) and other SNPs within the gene DYRK1A, which is thought to have an effect on brain development. However, the association was not supported in the follow-up samples. DYRK1A maps to the Down Syndrome critical region on chromosome 21, and several other lines of evidence, including observations in a transgenic mouse model,\textsuperscript{85} suggest that DYRK1A is involved in mental retardation associated with Down Syndrome.\textsuperscript{86} Furthermore, DYRK1A has been associated with Alzheimer disease.\textsuperscript{87} Being persistent, organized and self-controlled are central traits of conscientiousness, and deficiencies along these dimensions are clinical features associated with the neurodegenerative diseases for which DYRK1A has been implicated. There are several studies that support the links of conscientiousness with Alzheimer disease.\textsuperscript{85,86}

**Discussion**

We have presented the results of this first GWA study of all five major dimensions of personality assessed with the NEO-PI-R, a comprehensive, reliable, and widely used measure of the five-factor model. Compared with the existing literature (and especially with the candidate gene studies), a major strength is the sample size of about 4000 individuals included in the GWA analyses. A further contribution to increased signal/noise ratio is provided by the relative homogeneity of this sample from a founder population in Sardinia. The GWA approach offers a new opportunity for a systematic search of the genetic underpinnings of personality traits. These data might become even more valuable with the accumulation of GWA results from multiple samples, as for other traits and diseases.\textsuperscript{11,41,87,88} To provide some initial evidence of association beyond our Sardinian cohort, a selected number of top signals from the GWA analyses were typed in two independent samples. Although replication attempts clearly failed to find consistent effects for most SNPs we examined, a few provided more convincing evidence. In particular, the association of agreeableness with the CLOCK gene variant was consistent across the Sardinia and the two follow-up samples, but it was not in a third follow-up
sample. For neuroticism, the top SNPs in Sardinia were not among the top signals identified in the previous GWA scan, but we found some evidence for a role of SNAP25, and another SNP, rs1849710, maps in the 12q region where one of the most convincing linkage peaks has been reported. Given the large number of statistical tests performed, the likelihood of false positives is high. To address the problem of multiple testing we evaluated the significance level against the overly stringent Bonferroni threshold (it is questionable whether the Bonferroni correction is appropriate, and whether the number of tests are truly independent, given that many SNPs are in high LD). Using the Bonferroni threshold, none of the associations between the five personality factors and the individual SNPs tested reached the genome-wide significance. Nevertheless, we believe our results provide useful insights into the genetic architecture of personality traits. For example, when we used simulations that took into account the specific structure of the SardiNIA pedigrees, the availability of phenotype data and the pattern individuals genotyped with the 10 and 500K arrays in each family to evaluate the power of experimental design, we estimated ~88% power to detect alleles that account for 1.5% or more of the variance in one of the five main NEO-PI-R personality dimensions (at \( P < 1 \times 10^{-5} \)). We expect that we would have nearly 100% power to replicate these associations in samples of >2000 unrelated individuals. Unfortunately, we had only low power (~50% or less) to replicate smaller effects even in our larger follow-up panel of about 2000 individuals. As we did not observe any association signals in the SardiNIA sample that replicate consistently in the follow-up samples, it seems unlikely that alleles with large effects on personality exist. Researchers interested in the genetics of personality should not be discouraged because for many other quantitative traits with a definite genetic basis, such as height, it is now clear that most associated common alleles have only modest effects.

Caution is also required in rejecting the role of a SNP based on a failed replication attempt. In our case, the use of the relatively homogeneous founder population might have facilitated the detection of associations, but some of the identified SNPs may be particularly difficult to replicate in more heterogeneous populations. For example, the SNPs we identified might be in LD with the functional variant in the Sardinia but not in other populations. Some of the associations we identified might also depend on population-specific genetic or environmental background (gene–gene and gene–environment interactions). The replication samples also differed in recruitment strategies and used shorter and slightly less reliable phenotypic measures, which might have reduced power and contributed to differences across samples. In fact, heterogeneity among candidate gene studies has been linked to differences in the instrument used, but when we scored the SardiNIA data for the shorter questionnaires versions we found little differences in the results. Furthermore, given the universality of the five factors and the commonality of genetic factors for each of a variety of other complex traits that have been studied in many populations, the major limitations are almost certainly the numbers of individuals studied in relation to the small effect sizes observed. Increasing power would likely require much larger initial and follow-up scans. As with other quantitative traits, meta-analysis of genome-wide association scans may provide an effective means to dissect these small effects.

In addition to the five broad dimensions, future research should examine the specific facets (or lower-order traits) that compose each factor. The facets describe more specific and narrow phenotypes, which might be more easily linked to genetic variants. In fact, although facets tend to covary, a high score on the broad factor can result from the effect of different facets across individuals. For example, among those who score high on neuroticism, some might score high on anxiety but not depression, whereas other might score high on depression but not anxiety. This phenotypic variability increases noise and reduces the likelihood of identifying genetic variants. There is less phenotypic variability associated with the narrower phenotype assessed by facets, which should provide more power for GWA scans. However, analyses of multiple facets increase the number of tests and the risk of false positives.

Our results are consistent with most GWA studies of other quantitative traits in identifying SNPs that explain very small amounts of variance, generally less than 1%, and even these estimates for any particular trait/SNP are likely to be inflated (the ‘winner’s curse’). However, even a SNP that explains a very small amount of variance can guide our understanding of the biological underpinning of complex phenotypes and diseases. Indeed, genetic association tests across the five personality factors point to several SNPs within genes known for their functions in the brain and their effects on behavior and mental disorders (for example, SNAP25, CDH13, CDH23, BDNF, CNTNAP2, CLOCK, CTNNA2, IKBKAP, DYRK1A). These findings seem to reflect the phenotypic links between personality and psychiatric disorders. If confirmed in future studies, these findings might also advance our understanding of the continuum between normal and abnormal personality phenotypes. Given the high degree of comorbidity and other limitations of categorical systems, a dimensional approach to molecular psychiatry might well provide greater power to detect genetic variants associated with psychiatric disorders, and also provide possible points for eventual pharmaceutical intervention.

Acknowledgments

We thank the individuals who participated in this study; the SardiNIA team thanks Monsignore Piseddu (Bishop of Ogliastro), the mayors of the four Sardinian towns (Lanusei, Ilbono, Arzana and Elini), and the head of the Public Health Unit ASL4 for cooperation. We thank Professor Antonio Cao for his leadership of the SardiNIA project. This research was supported by
the Intramural Research Program of the NIH, National Institute on Aging. Dr Margit Burmeister’s laboratory was supported by the National Institute of Health NIH—NIMH R21 MH070793 and the National Alliance for Research on Schizophrenia and Depression (NARSAD), Independent Investigator Award (MB) and Young Investigator Award (SV).

Conflict of interest
Paul T Costa Jr receives royalties from the Revised NEO Personality Inventory. The authors declare that they have no other competing interests.

References

13 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. Nature 2007; 447: 661–678.
37 Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype


Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)