Molecular Psychiatry (2008), 1–10

www.nature.com/mp

ORIGINAL ARTICLE

Genome-wide association scan for five major dimensions of personality

A Terracciano[1], S Sanna[2], M Uda[2], B Deiana[2], G Usala[2], F Busonero[2], A Maschio[2], M Scally[1], N Patriciu[1], W-M Chen[3], MA Distel[4], EP Slagboom[5], DI Boomsma[4], S Villafuerte[6,7], E Śliwerska[6], M Burmeister[6,7,8], N Amin[9,10], ACJW Janssens[9], CM van Duijn[9,10], D Schlessinger[1], GR Abecasis[11] and PT Costa Jr[1]

[1]*National Institute on Aging, NIH, Baltimore, MD, USA;* [2]*Istituto di Neurogenetica e Neurofarmacologia, Consiglio Nazionale delle Ricerche, Cagliari, Italy;* [3]*Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA;* [4]*Department of Biological Psychology, VU University, Amsterdam, The Netherlands;* [5]*Molecular Epidemiology Section, Leiden University Medical Centre, Leiden, The Netherlands;* [6]*Molecular & Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI, USA;* [7]*Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA;* [8]*Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA;* [9]*Department of Epidemiology & Biostatistics, Erasmus University Medical Center, Rotterdam, The Netherlands;* [10]*Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands and* [11]*Center 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 362 129 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, $P = 5 \times 10^{-5}$), extraversion with BDNF and two cadherin genes (*CDH13* and *CDH23*; $Ps < 5 \times 10^{-5}$), openness with CNTNAP2 (rs10251794, $P = 3 \times 10^{-5}$), agreeableness with CLOCK (rs6832769, $P = 9 \times 10^{-6}$) and conscientiousness with DYRK1A (rs2835731, $P = 3 \times 10^{-5}$). 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 $P = 2 \times 10^{-5}$). 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.
Molecular Psychiatry advance online publication, 28 October 2008; doi:10.1038/mp.2008.113

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)

Correspondence: Dr A Terracciano or Dr PT Costa Jr, National Institute on Aging, Biomedical Research Center, 251 Bayview Blvd., Baltimore, MD 21224, USA and Dr GR Abecasis, Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA.
E-mail: terraccianoa@mail.nih.gov

Received 11 April 2008; revised 22 September 2008; accepted 1 October 2008

scans, recent studies have reliably identified common variants associated with complex traits and diseases that include body mass index,[8,9] height,[10,11] inflammatory bowel disease,[12,13] and type 1 and type 2 diabetes.[13,14] For personality traits, a GWA study[15] 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.[16] 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 assessed in all cultures tested so far and the five-factor structure can be clearly replicated in most samples.[17] Sex differences and maturational trends are other universal features of personality, with women everywhere generally scoring higher on neuroticism and agreeableness,[18] and with younger people generally scoring higher on neuroticism, extraversion and openness, but lower on agreeableness and conscientiousness in most cultures.[19] Furthermore, the five factors are predictors of important life outcomes,[20,21] including well-being,[22] academic performance,[23] vocational interests,[24] marital stability and satisfaction,[25] health risk behaviors[26,27] and longevity.[28] All five factors are related to personality disorders,[29] 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.[30] More generally, personality traits are thought of as risk factors, diagnostic indicators, and predictors of onset, severity, and outcome for most psychiatric disorders.[31–33] 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.[31,34–36] Personality traits are increasingly recognized as endophenotypes in genetic studies of mental disorders.[7,37]

The present study provides the first GWA results for all five dimensions of personality, as measured by the NEO-PI-R.[38] This study is part of the SardiNIA project,[3] which has targeted a highly interrelated population from the isolated Ogliastra region of Sardinia, Italy.[3] Common variants associated with several complex traits[8,11,39–41] 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 (LD) compared to outbred populations.[42] 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. However, several findings from this Sardinian cohort have been replicated in other populations.[8,11,39–41]

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 other 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.[3] Subjects are native-born, and at least 95% are known to have all grandparents born in the same province.[3] Valid personality data were obtained from 5669 subjects,[43] 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.[38] 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[44] and in more than 50 cultures.[17] Scales have shown longitudinal stability,[45] cross-observer agreement, and convergent and discriminant validity in a large body of studies.[16] 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.[3,43,46]

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.[47] 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.[47] 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.[48,49] The remaining SNPs failed quality checks ($\sim$2.9% of SNPs failed checks for data completeness, Hardy–Weinberg equilibrium and Mendelian incompatibilities) or had a minor allele frequency of $<5\%$ ($\sim$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.[8,47] The additive genetic effect was estimated for each SNP in the context of a variance component model that accounts for resemblance among related individuals.[47] We analyzed each of the five-factor scores in turn, including sex, age and age^2 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.[50,51] 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.[52] We checked the genomic control value for our GWA analyses,[52] and carried out principal component analysis of genome-wide SNP data in a subset of unrelated individuals.[53] 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.[3,47] Further, we assumed that average phenotypic values for hemizygous males would be the same as for homozygous females.[3,47]

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.[54,55]

SNPs with *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 *P*-values and a standardized measure of effect size (*Z*), with the sign (+ or −) indicating the direction of the effect.

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.[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.[57,58] Furthermore, abnormalities in the level of SNAP25 have been linked to other psychiatric disorders,[59–61] and genetic variants in SNAP25 have been associated with cognitive ability.[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 *TMEM16D* (rs1849710; $P = 2.29 \times 10^{-5}$), 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.[4] To examine sex-specific effects in the Sardinian sample, we conducted additional association analyses in women and men separately (sex-specific *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 ($P = 4.34 \times 10^{-4}$) and also in men ($P = 4.39 \times 10^{-3}$). No sex-specific effects reached genome-wide significance using the stringent Bonferroni correction.

Extraversion

In an interesting pattern for this factor, multiple independent SNPs within two Cadherin genes (*CDH13* and *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.[63–65] CDH23 is expressed in neurosensory epithelium and linked to some instances of deafness.[66] CDH13 is expressed in the heart and several brain tissues, where it is thought to act as a regulator of neural cell growth.[67] 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 *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 (rs16831315; $P = 8.05 \times 10^{-5}$), 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.

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 rs10251794 ($P = 3.43 \times 10^{-5}$), an intronic SNP in CNTNAP2, which encodes for the member of the neurexin family that has been linked with autism[68–70] and a complex phenotype of schizophrenia, epilepsy and cognitive impairment.[71] Other genes with strong signals and plausible biological relevance are the brain-specific angiogenesis inhibitor 3 (rs9342730; $P = 1.22 \times 10^{-5}$) and the myelin oligodendrocyte glycoprotein (rs16895223; $P = 4.60 \times 10^{-5}$).

Agreeableness

The most notable finding is the association of agreeableness with several SNPs within or close to the *CLOCK* gene (Figure 1). The strongest signal was with rs6832769 ($P = 8.71 \times 10^{-6}$). 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

SNP	*Gene*	*Chr*	*Position*	*Allele*	*Sardinia*		*Replication: US*		*Replication: Netherlands*		*Combined replication*	*Combined all*
					P	*Z*	*P*	*Z*	*P*	*Z*	*P*	*P*
Neuroticism												
rs6047641	—	20	21739854	G	6.54E−06	+0.197	0.153	−0.096	0.338	−0.032	0.160	2.16E−03
rs1159275	—	1	192380899	T	8.67E−06	−0.175	0.453	−0.007	0.372	+0.018	0.435	4.53E−04
rs7329003	—	13	106571549	A	8.99E−06	−0.156	NA		0.226	+0.037		3.86E−04
rs2039528	*PTPRF*	1	43629623	A	1.60E−05	+0.140	NA		0.227	−0.033		5.81E−04
rs1849710	*TMEM16D*	12	100024546	C	2.29E−05	−0.167	0.018	+0.146	0.179	−0.052	0.239	2.58E−03
rs362584	*SNAP25*	20	10202475	G	5.03E−05	+0.134	0.064	+0.081	0.353	+0.017	0.097	5.22E−05
Extraversion												
rs644148	*ZNF180*	19	49662775	G	8.03E−06	+0.136	[a]		NA			8.03E−06
rs17635977	*CDH23*	10	73016270	A	1.14E−05	+0.137	0.237	+0.033	0.473	−0.003	0.335	1.42E−04
rs4783307	*CDH13*	16	81634135	G	1.70E−05	+0.130	0.473	+0.004	0.244	−0.029	0.318	1.34E−03
rs904208	—	5	143451120	C	1.77E−05	+0.132	0.120	+0.057	0.164	+0.043	0.065	1.27E−05
rs2813838	—	7	23922532	G	2.10E−05	−0.127	0.048	+0.08	0.183	+0.038	0.037	1.63E−02
rs8056579	*CDH13*	16	81380925	G	2.53E−05	−0.212	0.462	+0.007	0.055	+0.087	0.105	7.43E−03
rs928114	*DAPK1*	9	87408360	C	3.83E−05	+0.162	0.262	−0.043	0.362	+0.021	0.436	1.19E−03
Openness												
rs644148	*ZNF180*	19	49662775	G	9.44E−07	+0.145	[a]		NA			9.44E−07
rs6610953	*FUNDC1*	X	44156440	G	1.73E−06	+0.154	0.300	+0.029	NA			5.74E−06
rs17819128	*CREBL2*	12	12652926	G	3.02E−06	+0.173	0.202	+0.052	0.263	−0.036	0.466	1.27E−04
rs9291420	*MIST*	4	10169156	G	3.39E−06	+0.152	0.429	−0.011	0.125	−0.068	0.164	1.42E−03
rs1037791	*TSPAN13*	7	16597902	A	3.94E−06	+0.149	NA		0.440	+0.008		3.59E−05
rs586281	—	1	182931141	G	6.18E−06	+0.178	0.118	+0.072	0.387	−0.015	0.283	6.38E−05
Agreeableness												
rs1380251	—	1	217957570	G	1.64E−06	−0.235	NA		0.375	−0.023		1.25E−05
rs2540226	*THUMPD2*	2	39870711	T	3.85E−06	+0.130	0.120	+0.050	[a]			2.98E−06
rs6832769	*CLOCK*	4	56139122	A	8.71E−06	−0.141	0.050	−0.077	0.110[b]	−0.055	0.022	1.74E−06[c]
rs602041	—	11	59535439	T	1.28E−05	+0.155	0.473	−0.005	0.256	+0.035	0.329	1.48E−04
rs9940706	*CDH13*	16	82256207	C	1.45E−05	−0.290	0.101	+0.088	0.424	−0.009	0.240	1.94E−03
rs7637878	*BFSP2*	3	134677154	G	1.85E−05	+0.148	0.323	+0.025	0.178	−0.045	0.351	1.17E−03
Conscientiousness												
rs11626232	*SMOC1*	14	69557149	C	4.82E−06	−0.175	0.028	−0.134	0.473	+0.004	0.112	9.95E−06
rs10953555	*LAMB1*	7	107175297	C	1.20E−05	+0.140	0.155	+0.050	0.346	−0.017	0.166	3.86E−05
rs17006841[d]	*MRPS18C*	4	84734382	C	1.56E−05	−0.361	NA		NA			1.56E−05
rs2835731	*DYRK1A*	21	37718598	C	2.81E−05	−0.208	0.262	+0.051	0.495	+0.002	0.332	1.70E−03
rs13070781	*EIF4E3*	3	71836676	A	2.82E−05	+0.141	0.201	−0.043	0.070	−0.068	0.048	1.56E−02
rs10945200	*COL19A1*	6	70948461	G	3.05E−05	+0.141	0.174	−0.044	[a]			8.12E−04

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.

[a]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 THUMPD2 rs2540226 we genotyped the nearby rs1861243 (with A: $P=0.39$) and instead of COL19A1 rs10945200 we genotyped the nearby rs3806052 (with C: $P=0.27$).

[b]For rs6832769, in addition to analyzing genotyped individuals, we used 44 flanking microsatellites to impute genotypes for additional family members.

[c]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 $P=8.6\times10^{-5}$.

[d]rs17006841 was merged into rs3182340.

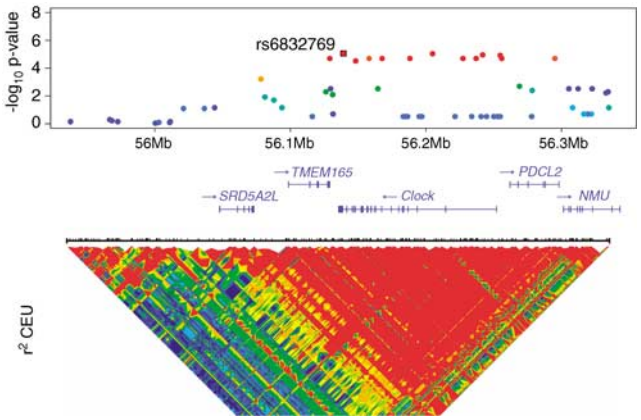

Figure 1 Association with agreeableness and linkage diseqilibrium (LD) pattern in the CLOCK region. Single-nucleotide polymorphisms (SNPs) showing evidence of association with agreeableness in the SardiNIA GWA scan. *P*-values are plotted against genomic position in NCBI Build 35. The SNP (rs6832769) showing the strongest association, and tested in the replication samples, is highlighted. Other SNPs are colored according to their degree of disequilibrium with rs6832769, ranging from high (red), to intermediate (green), to low (blue). Transcripts in the region are indicated at the bottom of the graph, with an arrow indicating the direction of transcription. Bottom panel presents patterns of linkage disequilibrium (r^2) for the *CLOCK* gene region in the HapMap CEPH population.

cal significance when meta-analytically combined ($P=0.022$). On the basis of the initial replication, we genotyped rs6832769 in an additional sample of 1822 individuals. In this third follow-up sample we failed to replicate the association of the rs6832769 CLOCK variant with agreeableness ($P=0.47$). Still, in the combined SardiNIA GWA and the three follow-up samples ($N=7875$) the *P*-value was 2×10^{-5}. CLOCK encodes for a transcription factor that is essential for circadian rhythm,[72,73] which in turn has large influences on human behavior, cognition and emotion. Agreeableness has been linked to morningness preference,[74] and the two have similar maturational trends. Younger adults tend to be more evening type and antagonistic, and older people are more agreeable[75] and more likely to be 'morning types'.[76] Mixed findings have been reported for a CLOCK variant known as 3111 T/C (rs1801260) with numerous traits and disorders, including morningness–eveningness preferences,[77] sleep and mood disorders,[78,79] schizophrenia,[80] and lower body weight among individuals with eating disorders.[81] The 3111 T/C variant is weakly linked to rs6832769 ($r^2=0.212$; physical distance = 3175 bp). To relate our finding to previous studies we used a hidden Markov program[11] to impute results for the 3111 T/C SNP. The imputed genotype was significantly associated with agreeableness (rs1801260, $P=0.0067$, effect = 0.08), with the *T* allele associated with higher agreeableness scores. This is consistent with the findings of the *T* allele associated with morning preference and lower risk of mental disorders found in previous studies.[77–81]

Beyond the SNPs reported in Table 1, other strong signals in genes with plausible biological relevance are OPCML (rs11223249; $P=3.52\times10^{-5}$) an opioid-binding protein; CTNNA2 (rs2861913; $P=6.95\times10^{-5}$), the α-N-catenin that interacts with cadherin proteins in essential brain functions;[63–65] and *IKBKAP* (rs10118853; $P=5.11\times10^{-5}$), a gene that causes familial dysautonomia, a sensory and autonomic neuropathy.

Conscientiousness

The strongest signal for conscientiousness was with a SNP within the gene *SMOC1* (rs11626232; $P=4.82\times10^{-6}$), which also showed a meta-analytic trend in the replication samples ($P=0.11$). Conscientiousness was strongly associated with rs2835731 ($P=2.81\times10^{-5}$) 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,[82] suggest that DYRK1A is involved in mental retardation associated with Down Syndrome.[83] Furthermore, DYRK1A has been associated with Alzheimer disease.[84] 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.[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.[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,[15] 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.[4,7]

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,[89] 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 $P<1\times10^{-5}$). 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.[11,90]

Caution is also required in rejecting the role of a SNP based on a failed replication attempt.[91] 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,[92] 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[17] 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.[11,41]

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[93,94] and other limitations of categorical systems,[95] 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 Ogliastra), 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

1 Bouchard TJ, Loehlin JC. Genes, evolution, and personality. *Behav Genet* 2001; **31**: 243–273.
2 Jang KL, McCrae RR, Angleitner A, Riemann R, Livesley WJ. Heritability of facet-level traits in a cross-cultural twin sample: support for a hierarchical model of personality. *J Pers Soc Psychol* 1998; **74**: 1556–1565.
3 Pilia G, Chen WM, Scuteri A, Orrú M, Albai G, Dei M *et al.* Heritability of cardiovascular and personality traits in 6148 sardinians. *PloS Genet* 2006; **2**: e132.
4 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.
5 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.
6 Neale BM, Sullivan PF, Kendler KS. A genome scan of neuroticism in nicotine dependent smokers. *Am J Med Genet B Neuropsychiatr Genet* 2005; **132**: 65–69.
7 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; **144**: 463–468.
8 Scuteri A, Sanna S, Chen WM, Uda M, Albai G, Strait J *et al.* Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet* 2007; **3**: e115B.
9 Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007; **316**: 889–894.
10 Weedon MN, Lettre G, Freathy RM, Lindgren CM, Voight BF, Perry JR *et al.* A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat Genet* 2007; **39**: 1245–1250.
11 Sanna S, Jackson AU, Nagaraja R, Willer CJ, Chen WM, Bonnycastle LL *et al.* Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat Genet* 2008; **40**: 198–203.
12 Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006; **314**: 1461–1463.
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.
14 Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J *et al.* Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 2006; **38**: 320–323.
15 Shifman S, Bhomra A, Smiley S, Wray NR, James MR, Martin NG *et al.* A whole genome association study of neuroticism using DNA pooling. *Mol Psychiatry* 2008; **13**: 302–312.
16 McCrae RR, Costa Jr PT. *Personality in Adulthood: A Five-Factor Theory Perspective*, 2nd edn. Guilford Press: New York, 2003.
17 McCrae RR, Terracciano A. 78 members of the personality profiles of cultures Project. Universal features of personality traits from the observer's perspective: data from 50 cultures. *J Pers Soc Psychol* 2005; **88**: 547–561.
18 Costa Jr PT, Terracciano A, McCrae RR. Gender differences in personality traits across cultures: robust and surprising findings. *J Pers Soc Psychol* 2001; **81**: 322–331.
19 McCrae RR, Costa Jr PT, de Lima MP, Simões A, Ostendorf F, Angleitner A *et al.* Age differences in personality across the adult life span: parallels in five cultures. *Dev Psychol* 1999; **35**: 466–477.
20 Paunonen SV. Big five factors of personality and replicated predictions of behavior. *J Pers Soc Psychol* 2003; **84**: 411–424.
21 Ozer DJ, Benet-Martinez V. Personality and the prediction of consequential outcomes. *Annu Rev Psychol* 2006; **57**: 8.1–8.21.
22 Costa Jr PT, McCrae RR. Influence of extraversion and neuroticism on subjective well-being: happy and unhappy people. *J Pers Soc Psychol* 1980; **38**: 668–678.
23 Chamorro-Premuzic T, Furnham A. Personality traits and academic examination performance. *Eur J Pers* 2003; **17**: 237–250.
24 Gottfredson GD, Jones EM, Holland JL. Personality and vocational interests: the relation of Holland's six interest dimensions to the five robust dimensions of personality. *J Couns Psychol* 1993; **40**: 518–524.
25 Kelly EL, Conley JJ. Personality and compatibility: a prospective analysis of marital stability and marital satisfaction. *J Pers Soc Psychol* 1987; **52**: 27–40.
26 Terracciano A, Lockenhoff CE, Crum RM, Bienvenu OJ, Costa Jr PT. Five-factor model personality profiles of drug users. *BMC Psychiatry* 2008; **8**: 22.
27 Trobst KK, Herbst JH, Masters III HL, Costa Jr PT. Personality pathways to unsafe sex: personality, condom use, and HIV risk behaviors. *J Res Pers* 2002; **36**: 117–133.
28 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.
29 Widiger TA, Costa Jr PT, McCrae RR. A proposal for Axis II: diagnosing personality disorders using the Five-Factor Model. In: Costa Jr PT, Widiger TA (eds). *Personality Disorders and the Five-Factor Model of Personality*, 2nd edn. American Psychological Association: Washington, DC, 2002, pp 431–456.
30 Widiger TA, Trull TJ. Plate tectonics in the classification of personality disorder: shifting to a dimensional model. *Am Psychol* 2007; **62**: 71–83.
31 Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. A longitudinal twin study of personality and major depression in women. *Arch Gen Psychiatry* 1993; **50**: 853–862.
32 Dinzeo TJ, Docherty NM. Normal personality characteristics in schizophrenia: a review of the literature involving the FFM. *J Nerv Ment Dis* 2007; **195**: 421–429.
33 Bagby RM, Rector NA, Bindseil K, Dickens SE, Levitan RD, Kennedy SH. Self-report ratings and informant ratings of personalities of depressed outpatients. *Am J Psychiatry* 1998; **155**: 437–438.
34 Kendler KS, Gatz M, Gardner CO, Pedersen NL. Personality and major depression: a Swedish longitudinal, population-based twin study. *Arch Gen Psychiatry* 2006; **63**: 1113–1120.
35 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.
36 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.
37 Savitz JB, Cupido CL, Ramesar RS. Trends in suicidology: personality as an endophenotype for molecular genetic investigations. *PLoS Med* 2006; **3**: e107.
38 Costa Jr PT, McCrae RR. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual*. Psychological Assessment Resources: Odessa, FL, 1992.
39 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

of {beta}-thalassemia. *Proc Natl Acad Sci USA* 2008; **105**: 1620–1625.
40 Li S, Sanna S, Maschio A, Busonero F, Usala G, Mulas A *et al.* The GLUT9 gene is associated with serum uric acid levels in sardinia and chianti cohorts. *PLoS Genet* 2007; **3**: e194.
41 Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R *et al.* Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008; **40**: 161–169.
42 Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorious H, Bedoya G *et al.* Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet* 2006; **38**: 556–560.
43 Costa Jr PT, Terracciano A, Uda M, Vacca L, Mameli C, Pilia G *et al.* Personality traits in Sardinia: testing founder population effects on trait means and variances. *Behav Genet* 2007; **37**: 376–387.
44 Terracciano A. The Italian version of the NEO PI-R: conceptual and empirical support for the use of targeted rotation. *Pers Indiv Differ* 2003; **35**: 1859–1872.
45 Terracciano A, Costa PTJ, McCrae RR. Personality plasticity after age 30. *Pers Soc Psychol Bull* 2006; **32**: 999–1009.
46 Yamagata S, Suzuki A, Ando J, Ono Y, Kijima N, Yoshimura K *et al.* Is the genetic structure of human personality universal? A cross-cultural twin study from North America, Europe, and Asia. *J Pers Soc Psychol* 2006; **90**: 987–998.
47 Chen WM, Abecasis GR. Family-based association tests for genomewide association scans. *Am J Hum Genet* 2007; **81**: 913–926.
48 Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy–Weinberg equilibrium. *Am J Hum Genet* 2005; **76**: 887–893.
49 Wigginton JE, Abecasis GR. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* 2005; **21**: 3445–3447.
50 Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 2000; **66**: 279–292.
51 Fulker DW, Cherny SS, Sham PC, Hewitt JK. Combined linkage and association sib-pair analysis for quantitative traits. *Am J Hum Genet* 1999; **64**: 259–267.
52 Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; **55**: 997–1004.
53 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–909.
54 Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006; **38**: 209–213.
55 Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B *et al.* Genomewide association analysis of coronary artery disease. *N Engl J Med* 2007; **357**: 443–453.
56 Osen-Sand A, Catsicas M, Staple JK, Jones KA, Ayala G, Knowles J *et al.* Inhibition of axonal growth by SNAP-25 antisense oligonucleotides *in vitro* and *in vivo*. *Nature* 1993; **364**: 445–448.
57 Kustanovich V, Merriman B, McGough J, McCracken JT, Smalley SL, Nelson SF. Biased paternal transmission of SNAP-25 risk alleles in attention-deficit hyperactivity disorder. *Mol Psychiatry* 2003; **8**: 309–315.
58 Feng Y, Crosbie J, Wigg K, Pathare T, Ickowicz A, Schachar R *et al.* The SNAP25 gene as a susceptibility gene contributing to attention-deficit hyperactivity disorder. *Mol Psychiatry* 2005; **10**: 998–1005, 1973.
59 Honer WG, Falkai P, Bayer TA, Xie J, Hu L, Li HY *et al.* Abnormalities of SNARE mechanism proteins in anterior frontal cortex in severe mental illness. *Cereb Cortex* 2002; **12**: 349–356.
60 Fatemi SH, Earle JA, Stary JM, Lee S, Sedgewick J. Altered levels of the synaptosomal associated protein SNAP-25 in hippocampus of subjects with mood disorders and schizophrenia. *Neuroreport* 2001; **12**: 3257–3262.
61 Scarr E, Gray L, Keriakous D, Robinson PJ, Dean B. Increased levels of SNAP-25 and synaptophysin in the dorsolateral prefrontal cortex in bipolar I disorder. *Bipolar Disord* 2006; **8**: 133–143.
62 Gosso MF, de Geus EJ, van Belzen MJ, Polderman TJ, Heutink P, Boomsma DI *et al.* The SNAP-25 gene is associated with cognitive ability: evidence from a family-based study in two independent Dutch cohorts. *Mol Psychiatry* 2006; **11**: 878–886.
63 Togashi H, Abe K, Mizoguchi A, Takaoka K, Chisaka O, Takeichi M. Cadherin regulates dendritic spine morphogenesis. *Neuron* 2002; **35**: 77–89.
64 Bozdagi O, Valcin M, Poskanzer K, Tanaka H, Benson DL. Temporally distinct demands for classic cadherins in synapse formation and maturation. *Mol Cell Neurosci* 2004; **27**: 509–521.
65 Hirano S, Suzuki ST, Redies C. The cadherin superfamily in neural development: diversity, function and interaction with other molecules. *Front Biosci* 2003; **8**: d306–d355.
66 Siemens J, Kazmierczak P, Reynolds A, Sticker M, Littlewood-Evans A, Muller U. The Usher syndrome proteins cadherin 23 and harmonin form a complex by means of PDZ-domain interactions. *Proc Natl Acad Sci USA* 2002; **99**: 14946–14951.
67 Takeuchi T, Misaki A, Liang SB, Tachibana A, Hayashi N, Sonobe H *et al.* Expression of T-cadherin (CDH13, H-Cadherin) in human brain and its characteristics as a negative growth regulator of epidermal growth factor in neuroblastoma cells. *J Neurochem* 2000; **74**: 1489–1497.
68 Bakkaloglu B, O'Roak BJ, Louvi A, Gupta AR, Abelson JF, Morgan TM *et al.* Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am J Hum Genet* 2008; **82**: 165–173.
69 Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, Ikeda M *et al.* A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. *Am J Hum Genet* 2008; **82**: 160–164.
70 Alarcon M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, Bomar JM *et al.* Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am J Hum Genet* 2008; **82**: 150–159.
71 Friedman JI, Vrijenhoek T, Markx S, Janssen IM, van der Vliet WA, Faas BH *et al.* CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. *Mol Psychiatry* 2008; **13**: 261–266.
72 King DP, Zhao Y, Sangoram AM, Wilsbacher LD, Tanaka M, Antoch MP *et al.* Positional cloning of the mouse circadian clock gene. *Cell* 1997; **89**: 641–653.
73 Steeves TD, King DP, Zhao Y, Sangoram AM, Du F, Bowcock AM *et al.* Molecular cloning and characterization of the human CLOCK gene: expression in the suprachiasmatic nuclei. *Genomics* 1999; **57**: 189–200.
74 DeYoung CG, Hasher L, Djikic M, Criger B, Peterson JB. Morning people are stable people: circadian rhythm and the higher-order factors of the Big Five. *Pers Indiv Differ* 2007; **43**: 267–276.
75 Terracciano A, McCrae RR, Brant LJ, Costa Jr PT. Hierarchical linear modeling analyses of NEO-PI-R scales in the Baltimore Longitudinal Study of Aging. *Psychol Aging* 2005; **20**: 493–506.
76 Paine SJ, Gander PH, Travier N. The epidemiology of morningness/eveningness: influence of age, gender, ethnicity, and socioeconomic factors in adults (30–49 years). *J Biol Rhythms* 2006; **21**: 68–76.
77 Mishima K, Tozawa T, Satoh K, Saitoh H, Mishima Y. The 3111T/C polymorphism of hClock is associated with evening preference and delayed sleep timing in a Japanese population sample. *Am J Med Genet B Neuropsychiatr Genet* 2005; **133**: 101–104.
78 Benedetti F, Dallaspezia S, Fulgosi MC, Lorenzi C, Serretti A, Barbini B *et al.* Actimetric evidence that CLOCK 3111 T/C SNP influences sleep and activity patterns in patients affected by bipolar depression. *Am J Med Genet B Neuropsychiatr Genet* 2007; **144**: 631–635.
79 Benedetti F, Serretti A, Colombo C, Barbini B, Lorenzi C, Campori E *et al.* Influence of CLOCK gene polymorphism on circadian mood fluctuation and illness recurrence in bipolar depression. *Am J Med Genet B Neuropsychiatr Genet* 2003; **123**: 23–26.
80 Takao T, Tachikawa H, Kawanishi Y, Mizukami K, Asada T. CLOCK gene T3111C polymorphism is associated with Japanese schizophrenics: a preliminary study. *Eur Neuropsychopharmacol* 2007; **17**: 273–276.
81 Tortorella A, Monteleone P, Martiadis V, Perris F, Maj M. The 3111T/C polymorphism of the CLOCK gene confers a predisposition to a lifetime lower body weight in patients with anorexia nervosa and bulimia nervosa: a preliminary study. *Am J Med Genet B Neuropsychiatr Genet* 2007; **144**: 992–995.

82 Altafaj X, Dierssen M, Baamonde C, Marti E, Visa J, Guimera J *et al.* Neurodevelopmental delay, motor abnormalities and cognitive deficits in transgenic mice overexpressing Dyrk1A (minibrain), a murine model of Down's syndrome. *Hum Mol Genet* 2001; **10**: 1915–1923.
83 Dowjat WK, Adayev T, Kuchna I, Nowicki K, Palminiello S, Hwang YW *et al.* Trisomy-driven overexpression of DYRK1A kinase in the brain of subjects with Down syndrome. *Neurosci Lett* 2007; **413**: 77–81.
84 Kimura R, Kamino K, Yamamoto M, Nuripa A, Kida T, Kazui H *et al.* The DYRK1A gene, encoded in chromosome 21 Down syndrome critical region, bridges between beta-amyloid production and tau phosphorylation in Alzheimer disease. *Hum Mol Genet* 2007; **16**: 15–23.
85 Dawson DV, Welsh-Bohmer KA, Siegler IC. Premorbid personality predicts level of rated personality change in patients with Alzheimer disease. *Alzheimer Dis Assoc Disord* 2000; **14**: 11–19.
86 Wilson RS, Schneider JA, Arnold SE, Bienias JL, Bennett DA. Conscientiousness and the incidence of Alzheimer disease and mild cognitive impairment. *Arch Gen Psychiatry* 2007; **64**: 1204–1212.
87 Sullivan PF, Lin D, Tzeng JY, van den Oord E, Perkins D, Stroup TS *et al.* Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry* 2008; **13**: 570–584.
88 Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K *et al.* Whole-genome association study of bipolar disorder. *Mol Psychiatry* 2008; **13**: 558–569.
89 Perneger TV. What's wrong with Bonferroni adjustments. *BMJ* 1998; **316**: 1236–1238.
90 Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S *et al.* Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet* 2008; **40**: 584–591.
91 Moonesinghe R, Khoury MJ, Liu T, Ioannidis JP. Required sample size and nonreplicability thresholds for heterogeneous genetic associations. *Proc Natl Acad Sci USA* 2008; **105**: 617–622.
92 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; **127**: 85–89.
93 Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S *et al.* Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Arch Gen Psychiatry* 1994; **51**: 8–19.
94 Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005; **62**: 617–627.
95 Costa Jr PT, Widiger TA (eds) *Personality Disorders and the Five-Factor Model of Personality.* American Psychological Association: Washington, DC, 2002.

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