Washington University School of Medicine
Digital Commons@Becker

Open Access Publications

2012

A genome-wide meta-analysis of association studies of Cloninger's Temperament Scales

A. C. Heath
Washington University School of Medicine in St. Louis

P. A. F. Madden
Washington University School of Medicine in St. Louis

et al

Follow this and additional works at: http://digitalcommons.wustl.edu/open_access_pubs

npg

A genome-wide meta-analysis of association studies of Cloninger's Temperament Scales

SK Service[1], KJH Verweij[2,3], J Lahti[4], E Congdon[1], J Ekelund[5,6], M Hintsanen[4,7], K Räikkönen[4], T Lehtimäki[8,9], M Kähönen[9,10], E Widen[11], A Taanila[12], J Veijola[13], AC Heath[14], PAF Madden[14], GW Montgomery[2], C Sabatti[15,16], M-R Järvelin[17,18,19,20], A Palotie[11,21,22,23], O Raitakari[24,25], J Viikari[26,27], NG Martin[2], JG Eriksson[6,28,29,30,31], L Keltikangas-Järvinen[4], NR Wray[2] and NB Freimer[1,32,33]

Temperament has a strongly heritable component, yet multiple independent genome-wide studies have failed to identify significant genetic associations. We have assembled the largest sample to date of persons with genome-wide genotype data, who have been assessed with Cloninger's Temperament and Character Inventory. Sum scores for novelty seeking, harm avoidance, reward dependence and persistence have been measured in over 11 000 persons collected in four different cohorts. Our study had >80% power to identify genome-wide significant loci ($P < 1.25 \times 10^{-8}$, with correction for testing four scales) accounting for $\geqslant 0.4\%$ of the phenotypic variance in temperament scales. Using meta-analysis techniques, gene-based tests and pathway analysis we have tested over 1.2 million single-nucleotide polymorphisms (SNPs) for association to each of the four temperament dimensions. We did not discover any SNPs, genes, or pathways to be significantly related to the four temperament dimensions, after correcting for multiple testing. Less than 1% of the variability in any temperament dimension appears to be accounted for by a risk score derived from the SNPs showing strongest association to the temperament dimensions. Elucidation of genetic loci significantly influencing temperament and personality will require potentially very large samples, and/or a more refined phenotype. Item response theory methodology may be a way to incorporate data from cohorts assessed with multiple personality instruments, and might be a method by which a large sample of a more refined phenotype could be acquired.
Translational Psychiatry (2012) **2**, e116; doi:10.1038/tp.2012.37; published online 15 May 2012

Introduction

Personality and temperament traits are stable representations of emotional, motor and attentional reactivity to stimulation, as manifested by an organized pattern of behavioral responses across a range of contexts.[1] The assessment of personality and temperament measures in human populations is therefore a major component of efforts to correlate higher order behaviors with underlying biology. Two common models for assessing personality include the five factor model of personality (FFM)[2] and the temperament and character inventory (TCI).[3,4]

The big five factors are openness to experience, conscientiousness, extraversion, agreeableness and neuroticism. Openness to experience is reflected in a strong intellectual curiosity and a preference for novelty and variety. Individuals scoring high on conscientiousness are characterized as being disciplined, organized and achievement-oriented. The extraversion dimension characterizes the tendency to be active, seek

[1]Center for Neurobehavioral Genetics, University of California, Los Angeles, CA, USA; [2]Genetic Epidemiology, Molecular Epidemiology and Psychiatric Genetics Laboratories, Queensland Institute of Medical Research, Brisbane, QLD, Australia; [3]School of Psychology, University of Queensland, Brisbane, QLD, Australia; [4]Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland; [5]Department of Psychiatry, University of Helsinki and Finland National Public Health Institute, Helsinki, Finland; [6]Finland Vaasa Hospital District, Vaasa, Finland; [7]Helsinki Collegium for Advanced Studies, University of Helsinki, Helsinki, Finland; [8]Department of Clinical Chemistry, Fimlab Laboratories, Tampere University Hospital, Tampere, Finland; [9]University of Tampere School of Medicine, Tampere, Finland; [10]Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland; [11]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; [12]Institute of Health Sciences, Public Health and General Practice, University of Oulu, Oulu, Finland; [13]Department of Psychiatry, Institute of Clinical Medicine, University of Oulu, Oulu, Finland; [14]Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA; [15]Department of Health and Research Policy, Stanford University, Stanford, CA, USA; [16]Department of Statistics, Stanford University, Stanford, CA, USA; [17]Department of Epidemiology and Biostatistics, School of Public Health, MRC-HPA Centre for Environment and Health, Imperial College London, London, UK; [18]Institute of Health Sciences, University of Oulu, Oulu, Finland; [19]Biocenter Oulu, University of Oulu, Oulu, Finland; [20]Department of Lifecourse and Services, National Institute of Health and Welfare, Oulu Finland; [21]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK; [22]Department of Medical Genetics, University of Helsinki, Helsinki, Finland; [23]University Central Hospital, Helsinki, Finland; [24]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; [25]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; [26]Department of Medicine, Turku University Hospital, Turku, Finland; [27]University of Turku, Turku, Finland; [28]National Institute for Health and Welfare, Helsinki, Finland; [29]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland; [30]Helsinki University Central Hospital, Unit of General Practice, Helsinki, Finland; [31]Folkhalsan Research Centre, Helsinki, Finland; [32]The Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA, USA and [33]Department of Psychiatry, University of California, Los Angeles, Los Angeles, CA, USA

Correspondence: Dr NB Freimer, Center for Neurobehavioral Genetics, University of California, Gonda Center Room 3506, 695 Charles E Young Dr South, Box 951761, Los Angeles, CA 90095, USA.
E-mail: nfreimer@mednet.ucla.edu
Keywords: association; genetics; genome-wide; meta-analysis; personality; temperament
Received 9 December 2011; revised 1 March 2012; accepted 5 April 2012

2

out stimulation and the company of others. Agreeableness evaluates the tendency to be helpful, cooperative and sympathetic towards others. Lastly, emotional stability, impulse control and anxiety are all components of the Neuroticism dimension.

The four temperament dimensions of the TCI are harm avoidance (HA), novelty seeking (NS), reward dependence (RD) and persistence (P). HA is a tendency to respond intensively to signals of aversive stimuli, thereby inhibiting behavior. NS is a tendency to respond with intense excitement to novel stimuli, or cues for potential rewards or potential relief of punishment and thereby activating behavior. RD is a tendency to respond intensely to signals of reward, especially social rewards, thereby maintaining and continuing particular behaviors. P is a tendency to persevere in behaviors that have been associated with reward or relief from punishment.

Although there is some evidence in support of convergence between the TCI and FFM, they differ in how the underlying models were created, with the FFM based on a lexical analysis of trait adjectives and the TCI based on a theoretical model that sought to account for individual differences in personality by integrating neurobiological systems, learning and social influences (for a review, see Stallings et al.[5]). In a sample of 130 individuals that had been administered both the TCI and the NEO-PI-R (representing the FFM[2]), De Fruyt[6] used a multiple regression approach to show that 23–51% of the variance in the TCI scales was explained by the NEO-PI-R scales, and 29–55% of the variance in the NEO-PI-R scales were explained by the TCI scales, indicating a substantial portion of the variance in each of the TCI and NEO-PI-R unaccounted for by the other.

Personality measures from these models show correlations to problem behaviors and psychiatric diagnoses,[7–12] and may serve as useful endophenotypes for the study of genetic components of behavior. An endophenotype is a quantitative, heritable trait, thought to more directly reflect the influence of genetic variation.[13] Personality dimensions assessed in both the FFM and TCI demonstrate broad-sense heritability in excess of 30%,[14] and have been the focus of several genome-wide association studies (GWAS).

The largest GWAS of personality measures has been conducted using the NEO-PI-R. de Moor et al.[15] analyzed the FFM dimensions for association to ~2.4M single-nucleotide polymorphisms (SNPs) in a meta-analysis of 17 375 individuals, and while two dimensions (openness to experience and conscientiousness) were associated at genome-wide significance levels to several SNPs, the associations were not replicated in an independent sample.

Although Cloninger proposed that HA, NS and RD were influenced by the serotonergic system, the dopamine system and the noradrenaline system, respectively,[16] genetic evidence to support this assertion has been mixed, and a recent large meta-analysis of the 5–HTTLPR polymorphism and HA showed no association.[17] Only a single GWAS of the TCI has been reported to date. Verweij et al.[18] tested all four TCI temperament scales for association with ~1.2M SNPs in a sample of 5117 Australians, employing single-marker analyses, gene-based tests and pathway analyses, and did not identify any genetic variants to be genome-wide significant.

Although both de Moor and Verweij failed to find genome-wide association to scales related to personality, there were key differences between these studies. They used different instruments to assess personality, and had very different sample sizes. One cannot rule out that the null results of de Moor et al.[15] were due in part to the instrument used. While both the TCI and FFM dimensions are clearly heritable, at a similar magnitude, the locus-specific heritabilities of dimensions of both instruments are unknown and may differ. That is, if the proportion of variance in the TCI that is not accounted for by the FFM has higher locus-specific heritability than the FFM dimensions themselves, it is possible that the TCI will have greater success in genetic mapping. Although the GWAS of Verweij et al.[18] used the TCI, the sample was not powered to detect genetic variants of small effect size (<1% of the variance explained). The aim of the current study was to employ meta-analytic techniques to evaluate the possibility of small genetic effects on the TCI. By combining four cohorts totaling over 11 000 persons, this sample has 80% power to identify association with an SNP responsible for as little as 0.4% of the variance in the temperament scales, at an alpha level of 1.2×10^{-8} (a genome-wide significance level of 5×10^{-8} corrected for testing four temperament traits). Additionally, two features of the study design should minimize the degree of phenotypic and genotypic heterogeneity across the cohorts. First, temperament in all four cohorts was analyzed using identical items. Second, three of the four cohorts were derived from Finland, a relatively genetically homogeneous population.

Materials and methods

Sample descriptions. The Northern Finland Birth Cohort (NFBC) is a population-based birth cohort comprised of 12 058 individuals born in the northernmost two Finnish provinces in 1966.[19] In 1997, a temperament questionnaire was given to 5999 individuals who participated at the age of 31 in a follow-up assessment. Subjects were asked to complete the questionnaire and return it by mail; 5105 individuals returned the questionnaire,[20] of whom 4508 were genotyped and passed quality control (QC, 55% female).

The cardiovascular risk in young Finns (YFS) study is a stratified random sample of children and adolescents aged 3–18 years from five university cities and surrounding areas with medical schools.[21,22] Subjects were born in years 1962, 1965, 1968, 1971, 1974 and 1977 and followed up every 3–5 years, beginning in 1980. Temperament data used in these analyses were collected in 2001; 2105 persons had valid phenotype data, and of these 1383 were genotyped and passed QC (54% female). The mean age of participants was 32.5 ($\pm$5.1) years.

The Helsinki Birth Cohort Study (HBCS) is a birth cohort sample of individuals born at Helsinki University Central Hospital in 1934–44.[23–25] Temperament data used in these analyses were collected in 2004; 1671 persons had valid phenotype data, and of these 1425 were genotyped and passed QC (60% female). The mean age of participants was 63.4 ($\pm$2.9) years.

The Australian twin registry (QIMR) was initiated in 1978. Temperament questionnaires were sent to two cohorts of Australian twins and their families (parents, children, spouses and siblings), the first in 1988 and the second in 1990. A total of 20 464 individuals had valid phenotypic data, and of these 5117 were genotyped and passed QC (1727 males and 3390 females, from 2567 independent families). The mean age of the participants was 36.2 ($\pm$ 12.1) years. The effective sample size (that is, correcting for non-independence of family members) was calculated to be 4312. Verweij *et al.*[18] used this sample in their GWAS of TCI scales.[19]

Temperament assessment. Temperament in all samples was assessed using Cloninger's TCI.[4] The QIMR sample used a short version (54 dichotomous items[26]) of the Tridimensional Personality Questionnaire (TPQ) subset of the TCI. Although the TPQ originally measured three dimensions, revisions showed that five items contributing to RD should be analyzed as a separate P scale, and that one of the RD items should be assigned to NS. Therefore, the final TPQ measure as obtained in the QIMR sample included 18 HA, 19 NS, 12 RD and 5 P items. For each scale, missing items were replaced with the mean item score. If individuals had >25% of the scales' items missing, their scale score was treated as missing. Scale scores were transformed by taking the arcsine of the square root, corrected for the linear combination of age, age-squared, sex, a sex by age interaction and a sex by age-squared interaction, and standardized separately for each sex to a mean of 0 and a s.d. of 1.

The NFBC used the TPQ subset of the TCI version 9, with 107 binary items. The YFS used the full TCI version 9 with 240 Likert items. HBCS used the TPQ version 4 with 98 binary items. NFBC, YFS and HBCS questionnaires were examined and the subset of questions identical to those administered to the QIMR sample were identified and used in all analyses. Negatively keyed questionnaire items were reverse scored as necessary. Persons missing >25% of data for a scale were set to missing for that scale. Persons missing <25% of data had any missing values imputed by the mean of other persons' responses (in the same study) to that item. The Likert-like scale used in the YFS was converted to a 0–1 measure by mapping $1 = 0$, $2 = 0.25$, $3 = 0.5$, $4 = 0.75$, and $5 = 1.0$. A sum score across all items in a scale was taken as the final measure. The HBCS sum scores were regressed on age and sex, and residuals taken as the phenotype. The NFBC sample was of a uniform age, and the sum score was regressed only on sex, and residuals taken as the final measure. Although the YFS sample varied in age, age was not significantly related to the sum score for any scale, therefore the score was regressed only on sex, and residuals taken as the final measure. Data transformations were not employed for NFBC/YFS analyses, and for HBCS analysis a natural logarithmic (ln) transform was applied to the HA data.

The means of the raw sum scores are very similar for three of the four cohorts (Table 1). HBCS, with mean age ~30 years older than the other three cohorts (63 years vs mid-30's), has lower average NS and P than the other three cohorts.

Genotyping and imputation. Individuals were genotyped on the following platforms, with the respective genotyping centers indicated in parentheses: NFBC—Illumina 370duo Chip (Broad Institute, Cambridge, MA, USA), YFS—Illumina 670K Custom BeadChip, HBCS—Illumina 610K Quad Chip (Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK), QIMR samples—Illumina 317K (Finnish Genome Centre, Helsinki, Finland); Illumina HumanCNV370-Quadv3 (Center for Inherited Disease Research, Baltimore, MD, USA); Illumina Human610-Quad (DeCode, Reykjavik, Iceland) and Affymetrix 6.0. (TGen, Phoenix, AZ, USA). Rather than combining genotype data from different platforms in a joint analysis of all four cohorts, we used meta-analytic techniques to combine results from association analyses performed separately by cohort (see below).

Sample and SNP-level QC in all Finnish cohorts proceeded using the same protocols. Individuals were excluded if they were missing >5% of data, if there was a discrepancy between reported sex and sex determined from the X chromosome, if they were sibs or half-sibs of other subjects or if they withdrew consent. YFS and NFBC subjects were excluded if they had low IQ (<70). Fewer than 5% of subjects were excluded during QC. Genotyped SNPs were excluded with call rate

Table 1 Raw sum scores for each scale, by sex and cohort

Sample	Males N Mean (s.d.)				Females N Mean (s.d.)			
	HA	NS	RD	P	HA	NS	RD	P
NFBC	2007 5.9 (3.9)	2007 8.8 (3.5)	2005 6.0 (2.5)	2009 2.9 (1.2)	2488 6.9 (3.9)	2488 9.4 (3.4)	2483 7.7 (2.3)	2489 2.7 (1.2)
HBCS	567 5.1 (4.1)	574 7.3 (3.6)	566 6.5 (2.6)	570 1.6 (1.2)	851 6.2 (4.3)	853 7.9 (3.8)	850 7.8 (2.4)	854 1.7 (1.3)
YFS	632 6.8 (2.9)	632 8.9 (2.4)	633 6.9 (1.7)	634 2.7 (0.7)	747 6.9 (2.8)	748 9.0 (2.3)	748 7.0 (1.6)	749 2.7 (0.7)
QIMR	1721 6.1 (4.2)	1716 8.5 (3.9)	1721 6.7 (2.7)	1717 3.0 (1.5)	3375 7.9 (4.3)	3371 8.2 (3.7)	3375 8.4 (2.4)	3365 2.9 (1.5)

Abbreviations: HA, harm avoidance; *N*, sample size; NS, novelty seeking; P, persistence; RD, reward dependence.

<95%, *P*-value from an exact test of Hardy–Weinberg Equilibrium (HWE) <10^{-4} and minor allele frequency <1%. Imputation to HapMap2 (HM2) was done at the Wellcome Trust Sanger Institute, separately by cohort, using all samples that passed QC and all genotyped SNPs that passed QC in an individual cohort. Imputation was done using Markov Chain Haplotyper (MaCH),[27] and all data were imputed to the forward/positive strand. The following numbers of SNPs were successfully imputed with $r^2 > 0.30$: NFBC: 2 454 909, YFS: 2 489 350, HBCS: 2 492 667.

Initial QC control in the QIMR sample was applied separately to different genotype platform data and different projects. Data were checked for ancestry outliers, Mendelian errors, HWE failure (excluded if $P < 10^{-6}$) and minor allele frequency. After separate QC checks, Illumina and Affymetrix data were imputed separately by MACH using the data from the European HM2. SNPs with an imputation quality score (r^2) >0.3 were retained, resulting in 2 380 486 Illumina and 2 369 130 Affymetrix SNPs. In addition, QC using individuals that were imputed on both the Illumina and the Affymetrix platforms, SNPs were only retained if they had high concordance rates for the most probable genotype, and had a minor allele frequency > 0.01. In total, 1 252 387 SNPs were available for association analyses. More details on QC procedures in the QIMR sample can be found in Verweij *et al.*[18]

Cohort level analysis. The Finnish HM2 imputed data were analyzed separately by cohort and followed identical protocols. A separate analysis was performed for each scale. Data for HA and RD were also analyzed separately by sex, as previous work in twins indicated sex differences in the source of genetic variance for these scales.[28] An additive model was assumed for SNP genotype/dose, and principal components (from a PCA analysis of the genotype IBS matrix between persons) that were significantly related to the phenotype were included as covariates (as per the method of Price *et al.*[29]) to guard against possible stratification. The first two PCs were always included as covariates, as previous work showed they correlated strongly with geographic birth-place and ancestry.[30] The HM2 imputed dosage data were analyzed using probABEL.[31]

For the QIMR sample, the most probable imputed genotype at each SNP was tested for association with the four TPQ scales using a family-based association test,[32] which takes family relationships into account (including identical twins). The additive genetic effect was calculated.

Meta-analysis. Meta-analysis combining results from all four samples was done using METAL (http://www.sph.umich.edu/csg/abecasis/metal/) by calculating a *Z*-statistic that was a weighted average of sample-level statistics, where the weights were proportional to the square root of the number of individuals examined in each sample, and selected so that the squared weights summed to 1. The weight for QIMR reflected only independent individuals. The direction of effect in each study was taken into account in calculating the average. There were 1 252 222 SNPs common to all four data cohorts and scales. The number of individuals analyzed varied by phenotype: HA: 11 597, NS: 11 612, RD: 11 590, P: 11 610.

We also employed the heterogeneity option of METAL. The METAL heterogeneity analysis requires a second pass of analysis to decide whether observed effect sizes (or test statistics) are homogeneous across samples. The resulting heterogeneity statistic has *n*–1 degrees of freedom for *n* samples.

Gene-based tests. To determine whether any genes harbor more associated SNPs than expected by chance, we performed a gene-based test for each personality scale (VEGAS, Versatile Gene-based Association Study[33]). VEGAS tests for association on a per-gene basis, by considering the *P*-value of all SNPs within genes (including +/−50 kb from the 5′ and 3′ UTR), accounting for the number of SNPs per gene, and linkage disequilibrium between the SNPs. As such, the test identifies genes that show more signals of association than expected by chance given their length and linkage disequilibrium between the SNPs. The gene-based test was performed on the meta-analysis association results.

Pathway analysis. Subsequently, all genes from the gene-based test with a *P*-value < 0.01 were included in a pathway analysis using the Ingenuity Pathway analysis program (Ingenuity Systems, Redwood City, CA, USA, release IPA 6.0). By performing these pathway analyses we tried to identify whether the genes most associated with the personality scales were more prevalent in any known biological or canonical pathway than would be expected by chance. The alpha level was set at 0.0125 (0.05/4 personality scales) and significance of individual pathways was corrected for multiple testing by the Benjamini–Hochberg procedure as implemented in Ingenuity. The pathway analysis was performed on the results from the gene-based test of the meta-analysis results.

Prediction. We used the results from a meta-analysis using only the three Finnish cohorts to predict the four TCI scales in the QIMR sample, using the 'score' function in PLINK.[34] We restricted this analysis to the same set of SNPs used in the full meta-analysis, and used only one individual per family. The 'risk score' for individuals in the QIMR sample was constructed by multiplying the number of copies of the effect allele at each SNP by the *Z*-score from the Finnish-only meta-analysis of a given scale, and summing across SNPs. The observed TCI score in the QIMR sample was regressed on this risk score to assess the degree to which variability in the observed phenotype could be explained by variability in the risk score. The risk score was calculated using all SNPs, and also using the top 10, 20, 30, 40 and 50% of SNPs in the Finnish-only meta-analysis.

Results

Meta-analysis. Genomic control lambda parameters[35] estimated from the meta-analysis of 1 252 222 autosomal SNPs indicated minimal inflation of test statistics over the null value of 1.0; HA: 1.01, NS: 1.04, RD: 1.00, P: 1.02 (Figure 1 QQ plots). No SNPs were significant at a genome-wide threshold of 5×10^{-8}. The most significant finding was for rs17608059 on chromosome 17 with scale P, with a *P*-value
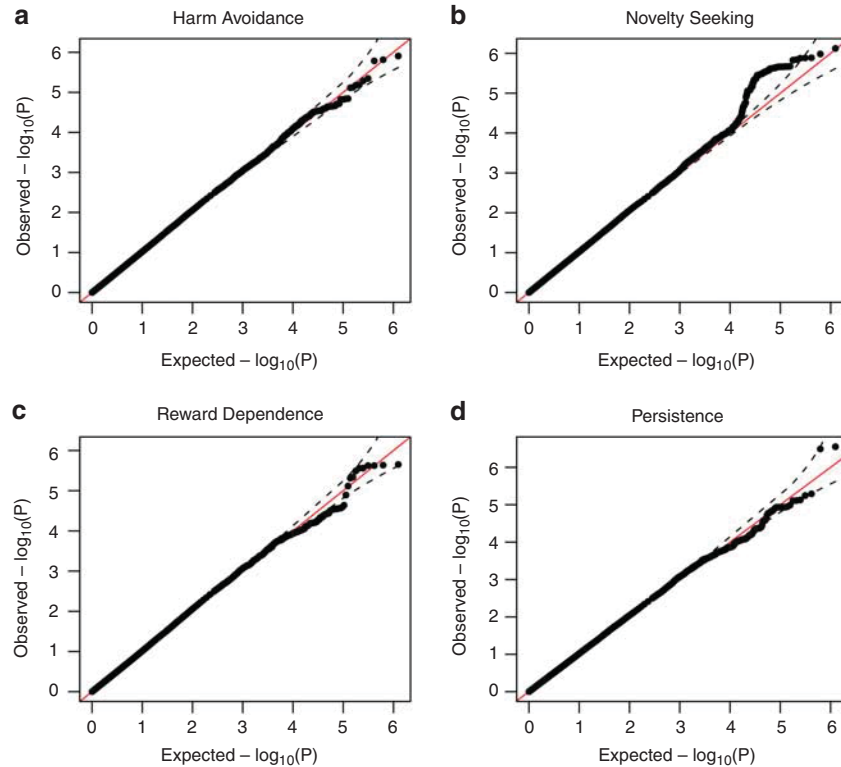
Figure 1 QQ plots of meta-analysis results for each of the four temperament scales. On the *x*-axis is the distribution of −log10 *P*-values expected under the null hypothesis of no association of SNPs to the phenotype. On the *y*-axis is the ordered distribution of observed −log10 *P*-values. Deviation from the 1:1 line in the bulk of the distribution can suggest inflation of test statistics. The 95% confidence bands (dashed lines) are generated assuming the jth order statistic from a uniform (0.1) sample has a beta (j,n−j + 1) distribution, and assuming independence. (**a**) Harm avoidance; (**b**) novelty seeking; (**c**) reward dependence; (**d**) persistence.

of 2.8×10^{-7} (Table 2). There were 83 SNPs from 16 independent genomic locations on 12 chromosomes with $P < 10^{-5}$ (HA: 9 SNPs, NS: 57 SNPs, RD: 10 SNPs, P: 7 SNPs, Supplemental Table S1). Scales HA and RD were also analyzed separately by sex; across all four analyses 73 SNPs from 13 independent genomic locations resulted in $P < 10^{-5}$ but none were significant at a genome-wide level (Supplemental Table S2). Meta-analysis of the three Finnish cohorts alone also did not produce any genome-wide significant results (data not shown), nor did meta-analysis including the heterogeneity option. *A priori*, both QIMR and HBCS might be considered to be cohorts with a heterogeneous signal; QIMR due to population differences and HBCS due to age differences. Among the handful of markers with METAL heterogeneity $P < 10^{-5}$ for one or more scales, it was never true that only the QIMR sample or the HBCS sample had a test result considered to be heterogeneous from the other three cohorts.

de Moor *et al.*[15] found two SNPs on 5q14.3 to be genome-wide significantly associated with openness to experience (rs1477268 and rs2032794) and one SNP on 18q21.1 to be genome-wide significantly associated with conscientiousness (rs2576037). Neither association was replicated in an independent sample. Openness to experience and conscientiousness are not measured in our sample and are only modestly correlated to our phenotypes (correlations of openness to experience/conscientiousness with NS, HA, P and RD are 0.27/−0.36, −0.33/−0.24, 0.03/0.46 and 0.32/0.07,

respectively[6]); however, we still reviewed the association findings for these three SNPs in our results. We find rs1477268 and rs2032794 to be associated to NS ($P = 0.03$). In de Moor *et al.*,[15] the 'T' allele at these markers resulted in a decrease in openness to experience; we find the 'T' allele to result in a decrease in NS. Association of HA, RD and P to these two SNPs all resulted in $P > 0.30$. de Moor *et al.*[15] found the 'T' allele of rs2576037 to result in a decrease in conscientiousness. In our sample the 'T' allele was associated with a decrease in HA ($P = 0.10$) and P ($P = 0.076$). Association of RD and NS to rs2576047 both resulted in $P > 0.27$.

Gene-based tests and pathway analysis. Approximately 17 200 tests were performed as part of the autosomal gene-based analysis. Genomic control parameters indicated minimal inflation of the test statistics over the null value of 1.0; HA: 1.00, NS: 1.04, RD: 0.991, P: 0.963. The percentage of associations to be significant at the 0.05 level range from 4.7% with scale P to 5.5% with scales HA and RD. None of the scales resulted in gene-based associations that survived correction for multiple testing (17 261 tests and four scales, $\alpha = 7.2 \times 10^{-7}$). The top five genes for each scale are presented in Supplemental Table S3.

We then examined, for each scale, all genes with $P < 0.01$ in the gene-based test to see whether they were concentrated in known biological or canonical pathways, using the Ingenuity Pathway analysis program (Ingenuity Systems, release IPA

Table 2 Meta analysis and individual study-level results for the most significant SNPs in each of 16 regions demonstrating association to a temperament phenotype at $P < 10^{-5}$ in the meta-analysis

PH	SNP	CH	BP	Gene (kb)	A1	A2	Z score	Pval	YFS (sample size 1379–1383)			NFBC (sample size 4488–4498)			HBCS (sample size 1409–1425)			QIMR (effective sample size = ~4312)		
									Freq1	Beta	Pval	Freq1	Beta	Pval	Freq1	Beta	Pval	Freq1	Beta	Pval
HA	rs10172965	2	195580064	SLC39A10 (649.712)	T	C	−4.544	5.52E-06	0.154	−0.055	7.14E-01	0.139	−0.30	9.56E-03	0.171	−0.066	2.89E-02	0.212	−0.089	8.11E-04
HA	rs2741200	8	134141015	TG/SLA (0)	T	C	4.576	4.73E-06	0.646	0.020	8.59E-01	0.671	0.32	2.55E-04	0.656	0.022	3.52E-01	0.713	0.075	1.74E-03
HA	rs17680945	15	27027109	APBA2	A	G	4.799	1.59E-06	0.018	0.35	3.75E-01	0.020	0.92	1.85E-03	0.019	0.15	9.45E-02	0.024	0.23	1.22E-03
NS	rs11163372	1	82019836	LPHN2 (18.83)	T	C	4.693	2.69E-06	0.192	0.069	5.33E-01	0.188	0.19	4.15E-02	0.188	0.49	5.86E-03	0.151	0.11	2.33E-04
NS	rs6754640	2	50584822	NRXN1 (0)	A	G	−4.884	1.04E-06	0.445	−0.042	6.47E-01	0.452	−0.15	4.51E-02	0.433	−0.32	2.44E-02	0.378	−0.099	1.02E-05
NS	rs987360	4	138447059	PCDH18 (212.5)	T	C	4.730	2.25E-06	0.421	0.20	2.55E-02	0.401	0.27	4.92E-04	0.423	0.37	1.10E-02	0.404	0.033	1.39E-01
NS	rs7387468	8	94427274	FAM92A1 (354.674)	T	G	4.649	3.33E-06	0.580	0.16	8.80E-02	0.587	0.22	2.27E-03	0.561	0.25	7.67E-02	0.525	0.054	1.14E-02
NS	rs11018023	10	133220016	TCERG1L (220.042)	A	G	4.482	7.41E-06	0.153	0.14	2.73E-01	0.175	0.21	3.68E-02	0.167	0.37	4.42E-02	0.210	0.091	5.69E-04
NS	rs1494508	12	64128525	MSRB3 (0)	A	G	−4.767	1.87E-06	0.470	−0.10	2.63E-01	0.479	−0.20	5.84E-03	0.462	−0.12	3.80E-01	0.369	−0.087	1.09E-04
P	rs12494658	3	85957166	CADM2 (0)	T	C	−4.562	5.08E-06	0.749	−0.059	6.41E-02	0.752	−0.078	8.00E-03	0.757	−0.17	3.50E-03	0.755	−0.051	4.01E-02
P	rs10057405	5	127680921	FBN2 (0)	A	C	−4.449	8.63E-06	0.899	−0.076	8.70E-02	0.891	−0.11	4.90E-03	0.912	−0.12	1.77E-01	0.891	−0.091	7.31E-03
P	rs291272	9	10033014	PTPRD (0)	T	G	−4.550	5.37E-06	0.968	−0.12	1.38E-01	0.965	−0.20	4.43E-03	0.966	−0.10	4.61E-01	0.969	−0.20	9.78E-04
P	rs17608059	17	13851274	COX10 (62.169)	T	C	−5.135	2.82E-07	0.504	−0.036	1.82E-01	0.491	−0.094	4.13E-04	0.482	−0.17	2.73E-04	0.466	−0.042	4.84E-02
RD	rs1996720	8	113143272	CSMD3 (161.1)	T	C	4.766	1.88E-06	0.279	0.25	4.49E-04	0.275	0.11	5.41E-02	0.290	0.081	4.32E-01	0.286	0.079	6.35E-04
RD	rs10995356	10	64325919	EGR2 (79.786)	A	G	4.651	3.30E-06	0.380	0.097	1.36E-01	0.390	0.11	4.61E-02	0.415	0.15	1.01E-01	0.511	0.081	1.37E-04
RD	rs2833693	21	32474383	C21orf45 (88.02)	T	C	−4.633	3.60E-06	0.718	−0.14	4.20E-02	0.723	−0.17	4.05E-02	0.714	−0.24	1.88E-02	0.712	−0.050	3.02E-02

Abbreviations: CH, chromosome; HA = harm avoidance, NS = novelty seeking, P = persistence, RD = reward dependence; SNP, single-nucleotide polymorphism. PH is the temperament scale. BP is the physical location in build 36 basepairs. Gene is the closest gene to the SNP. (kb) is the distance from the SNP. A1 is the reference (effect) allele. Freq1 is the frequency of this allele. Beta is the regression coefficient and indicates the change in scale for each copy of the A1 allele. HA was ln-transformed for HBCS, and all scales were arcsin square root transformed for QIMR. *P*-value associated with either the meta-analytic Z-score or the study-level Beta. For YFS, NFBC and HBCS, the sample size varied with phenotype, sample size range is indicated.

6.0). The number of genes included in this analysis is HA: 193 genes, NS: 198 genes, RD: 225 genes, P: 135 genes. Results from the pathway analyses were not significant after correction for multiple testing, indicating that our top genes were not over-represented in known biological or canonical pathways more than one would expect by chance.

Prediction. A risk score calculated from the top 50% of SNPs in the HA Finnish-only meta-analysis accounted for 0.28% of the variance in HA in the QIMR sample ($P = 0.007$); however, this result was not significant after correction for multiple testing (six thresholds and four scales = 24 tests). The top 10% of SNPs in the RD and P Finnish-only meta-analysis accounted for 0.15% ($P = 0.052$) and 0.17% ($P = 0.04$) of the variance in RD and P in the QIMR sample, respectively. Using all SNPs in the Finnish-only meta-analysis of NS accounted for 0.059% ($P = 0.22$) of the variance in NS in the QIMR sample. Other SNP thresholds resulted in less of the variance in the QIMR sample being accounted for by the risk score (Supplemental Table S4).

Discussion

We report here the results of the largest GWAS conducted to date for personality assessed using the TCI. The lack of genome-wide significant associations in our meta-analysis of more than 11 000 subjects, and the lack of replicated associations for personality measured by the NEO-PI-R in an even larger meta-analysis[15] suggest that it will be challenging to identify such associations using standard approaches for studying personality traits. Although we find modest association of the top findings of de Moor *et al.*[15] to some of our phenotypes, the statistical evidence is well below the level required for replication. Additionally, we find no association evidence to support the suggestion of Cloninger[16] that NS, HA and RD would be influenced by genes directly affecting the dopamine, serotoneric or noradrenaline systems, respectively. Two previous studies have identified genome-wide significant linkage to HA[36] and NS.[37] None of our top association signals ($P < 10^{-5}$) for these phenotypes were on the same chromosomal arms as these linkage findings.

The true genetic architecture underlying variation in personality is of course unknown, but as with other complex polygenic phenotypes, causal loci are likely represented by a mixture of common variants of small effect and rare variants some of which could have larger effect. Height is a classic polygenic phenotype; a recent meta-analysis of ~180K persons has demonstrated that 10.5% of the phenotypic variation in height can be explained by 180 associated loci.[38] Although our study is much smaller, it is worth noting that our prediction analysis accounted for, at most, only 0.28% of the phenotypic variability in temperament scales.

Our study had >80% power to detect loci responsible for 0.4% of the phenotypic variation in temperament scales, at genome-wide significance levels. Failure of our study to detect significant association to what are clearly heritable phenotypes suggests that either the true effect sizes of causal loci are much smaller, and/or that the causal loci are rare and not well tagged by common variation.

GWAS are designed to identify common polymorphisms responsible for variation between individuals, and identification of common loci with small effects on personality traits could be possible by assembling very large sample sizes. For example, GWAS of blood pressure have identified replicated associations at genome-wide significance but only once sample sizes in excess of 60 000 individuals were available for meta-analysis.[39] There are substantial obstacles to amassing a sample of such size for meta-analyses of personality, which do not exist for traits such as blood pressure. Blood pressure is a relatively direct biological measure, which is assessed in an objective and standardized way throughout the world. It is therefore straightforward to combine data across studies. In sharp contrast, personality trait assessment relies on self-report instruments, such as the TCI and NEO-PI-R, which pose two potential problems. First, while test–retest correlation for HA, RD and NS range from 0.58 to 0.84, for measures collected an average of 2 years apart,[26] it is possible that self-report biases and differences in subjective interpretation of questionnaire items may introduce error in the assessment of traits. Second, different instruments reflect different models of personality, highlighting philosophical differences in schools of thought about core components of personality. It therefore remains unclear to what degree the phenotypes in individuals assessed via TCI overlap with those obtained in individuals assessed via NEO-PI-R.

Although combining data from studies using different personality assessments may enable GWAS of personality in sample sizes large enough to detect common loci with small effects on personality, this is a challenging undertaking. A naïve meta-analysis that would use simple sum scores is unlikely to be effective, given the modest correlation between dimensions measured in different instruments (for example, De Fruyt *et al.*[6] showed the maximum correlation between dimensions of the TPQ and the FFM to be 0.54). Alternatively, one might attempt to map in a meta analysis not the sum scores themselves, but the scores from a principal components analysis that would combine information across scales within the same instrument, and potentially account for more of the phenotypic variance. A more sophisticated approach would be to employ item response theory (IRT[40]) to estimate the unmeasured, latent trait thought to be evaluated by personality assessments. van den Berg *et al.*[41] applied IRT to the attention problems subscale of the Young Adults Self Report questionnaire[42] assessed in a sample of individuals from the Netherlands Twin Registry. Heritability of the estimated latent trait was found to be much larger than the heritability of the traditionally used sum score: 73% vs 40%, respectively. Using IRT in samples evaluated with different personality assessments would identify a subset of items in these different instruments that are related to a common, unmeasured latent trait. Samples measured with multiple instruments are needed to identify these items, which are then extracted from samples evaluated using only one instrument. Refinement of personality phenotypes in this manner has the potential to greatly improve power to genetically map these traits.

The results of our meta-analysis, as well as those of de Moor *et al.*,[15] demonstrate that the null GWAS findings are not simply due to the instrument used, and appear to suggest that successful mapping of loci contributing to personality will

require new strategies and methodology. Additionally, next-generation sequencing soon will provide a host of data that may reveal rare variants that, when aggregated in the form of a 'burden analysis',[43] account for variability in personality traits. Understanding the biological processes underlying personality-related traits would be greatly facilitated by discovery of any such associated loci, and such loci may also provide a window for understanding cognitive and behavioral disorders.

Conflict of interest

The authors declare no conflict of interest.

1. Oliver JP, Robins RW, Pervin LA. *Handbook of Personality, Third Edition: Theory and Research*. The Guilford Press: New York, 2008.

2. Costa PT, McCrae RR. *Professional Manual: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor-Inventory (NEO-FFI)*. Psychological Assessment Resources: Odessa, 1992.

3. Cloninger CR. A unified biosocial theory of personality and its role in the development of anxiety states. *Psychiatr Dev* 1986; **4**: 167–226.

4. Cloninger CR, Svrakic DM, Przybeck TR. A psychobiological model of temperament and character. *Arch Gen Psychiatry* 1993; **50**: 975–990.

5. Stallings MC, Hewitt JK, Cloninger CR, Heath AC, Eaves LJ. Genetic and environmental structure of the Tridimensional Personality Questionnaire: three or four temperament dimensions? *J Pers Soc Psychol* 1996; **70**: 127–140.

6. De Fruyt F, Van de Wiele L, Van Heeringen C. Cloninger's psychobiological model of temperament and character and the five-factor model of personality. *Pers Individ Dif* 2000; **29**: 441–452.

7. Cloninger CR, Svrakic DM, Przybeck TR. Can personality assessment predict future depression? A twelve-month follow-up of 631 subjects. *J Affect Disord* 2006; **92**: 35–44.

8. Distel MA, Trull TJ, Willemsen G, Vink JM, Derom CA, Lynskey M *et al.* The five-factor model of personality and borderline personality disorder: a genetic analysis of comorbidity. *Biol Psychiatry* 2009; **66**: 1131–1138.

9. Hintsanen M, Pulkki-Raback L, Juonala M, Viikari JS, Raitakari OT, Keltikangas-Jarvinen L. Cloninger's temperament traits and preclinical atherosclerosis: the Cardiovascular Risk in Young Finns Study. *J Psychosom Res* 2009; **67**: 77–84.

10. Khan AA, Jacobson KC, Gardner CO, Prescott CA, Kendler KS. Personality and comorbidity of common psychiatric disorders. *Br J Psychiatry* 2005; **186**: 190–196.

11. Ono Y, Ando J, Onoda N, Yoshimura K, Momose T, Hirano M *et al.* Dimensions of temperament as vulnerability factors in depression. *Mol Psychiatry* 2002; **7**: 948–953.

12. Sovio U, King V, Miettunen J, Ek E, Laitinen J, Joukamaa M *et al.* Cloninger's Temperament dimensions, socio-economic and lifestyle factors and metabolic syndrome markers at age 31 years in the Northern Finland Birth Cohort 1966. *J Health Psychol* 2007; **12**: 371–382.

13. Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry* 2003; **160**: 636–645.

14. Johnson AM, Vernon PA, Feiler AR. Behavioral genetic studies of personality: An introduction and review of the results of 5-+ years of research. *Handbook of personality theory and assessment*. Sage: London, 2008; 145–173.

15. de Moor MH, Costa PT, Terracciano A, Krueger RF, de Geus EJ, Toshiko T *et al.* Meta-analysis of genome-wide association studies for personality. *Mol Psychiatry* 2012; **17**: 337–349.

16. Cloninger CR. A systematic method for clinical description and classification of personality variants. A proposal. *Arch Gen Psychiatry* 1987; **44**: 573–588.

17. Munafo MR, Freimer N, Ng W, Ophoff R, Veijola J, Miettunen J *et al.* 5-HTTLPR genotype and anxiety-related personality traits: A meta-analysis and new data. *Am J Med Genet B* 2008; **150b**: 271–281.

18. Verweij KJ, Zietsch BP, Medland SE, Gordon SD, Benyamin B, Nyholt DR *et al.* A genome-wide association study of Cloninger's temperament scales: implications for the evolutionary genetics of personality. *Biol Psychol* 2010; **85**: 306–317.

19. Rantakallio P. Groups at risk in low birth weight infants and perinatal mortality. *Acta Paediatr Scand Suppl* 1969; **193**: 43.

20. Miettunen J, Kantojarvi L, Ekelund J, Veijola J, Karvonen JT, Peltonen L *et al.* A large population cohort provides normative data for investigation of temperament. *Acta Psychiatr Scand* 2004; **110**: 150–157.

21. Akerblom HK, Uhari M, Pesonen E, Dahl M, Kaprio EA, Nuutinen EM *et al.* Cardiovascular risk in young Finns. *Ann Med* 1991; **23**: 35–39.

22. Raitakari OT, Juonala M, Ronnemaa T, Keltikangas-Jarvinen L, Rasanen L, Pietikainen M *et al.* Cohort profile: the cardiovascular risk in Young Finns Study. *Int J Epidemiol* 2008; **37**: 1220–1226.

23. Barker DJ, Osmond C, Forsen TJ, Kajantie E, Eriksson JG. Trajectories of growth among children who have coronary events as adults. *N Engl J Med* 2005; **353**: 1802–1809.

24. Eriksson JG, Osmond C, Kajantie E, Forsen TJ, Barker DJ. Patterns of growth among children who later develop type 2 diabetes or its risk factors. *Diabetologia* 2006; **49**: 2853–2858.

25. Raikkonen K, Pesonen A-K, Heinonen K, Lahti J, Kajantie E, Forsen TJ *et al.* Infant growth and hostility in adult life. *Psychosom Med* 2008; **70**: 306–313.

26. Heath AC, Cloninger CR, Martin NG. Testing a model for the genetic structure of personality: a comparison of the personality systems of Cloninger and Eysenck. *J Pers Soc Psychol* 1994; **66**: 762–775.

27. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2009; **34**: 816–834.

28. Keller MC, Coventry WL, Heath AC, Martin NG. Widespread evidence for non-additive genetic variation in Cloninger's and Eysenck's personality dimensions using a twin plus sibling design. *Behav Genet* 2005; **35**: 707–721.

29. 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.

30. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* 2009; **41**: 35–46.

31. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinform* 2010; **11**: 134.

32. Chen WM, Abecasis GR. Family-based association tests for genomewide association scans. *Am J Hum Genet* 2007; **81**: 913–926.

33. Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM *et al.* A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* 2010; **87**: 139–145.

34. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.

35. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; **55**: 997–1004.

36. Cloninger CR, Van Eerdewegh P, Goate A, Edenberg HJ, Blangero J, Hesselbrock V *et al.* Anxiety proneness linked to epistatic loci in genome scan of human personality traits. *Am J Med Genet* 1998; **81**: 313–317.

37. Curtis D. Re-analysis of collaborative study on the genetics of alcoholism pedigrees suggests the presence of loci influencing novelty-seeking near D12S391 and D17S1299. *Psychiatr Genet* 2004; **14**: 151–155.

38. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010; **467**: 832–838.

39. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A *et al.* Genome-wide association study of blood pressure and hypertension. *Nat Genet* 2009; **41**: 677–687.

40. Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of Item Repsonse Theory.* Sage Press: Newbury Park, 1991.

41. van den Berg SM, Glas CA, Boomsma DI. Variance decomposition using an IRT measurement model. *Behav Genet* 2007; **37**: 604–616.

42. Achenbach TM. *Manual for the young adult self-report and young adult behavior checklist.* University of Vermont: Burlington, 1997.

43. Bansal V, Libiger O, Torkamani A, Schork NJ. Statistical analysis strategies for association studies involving rare variants. *Nat Rev Genet* 2010; **11**: 773–785.

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