Genes, Brain and Behavior
Official publication of the International Behavioural and Neural Genetics Society

Genes, Brain and Behavior (2015) ***14:*** *345–356* doi: 10.1111/gbb.12212

Pathway analysis of genome-wide association datasets of personality traits

H.-N. Kim[†], B.-H. Kim[†], J. Cho[‡,§,¶], S. Ryu[‡,], H. Shin[††], J. Sung[‡‡], C. Shin[§§], N. H. Cho[¶¶], Y. A. Sung[***], B.-O. Choi[†††] and H.-L. Kim[†,*]**

[†]Department of Biochemistry, School of Medicine, Ewha Womans University, [‡]Center for Cohort Studies, Total Healthcare Center, Kangbuk Samsung Hospital, School of Medicine, [§]Department of Health Sciences and Technology, SAHIST, Sungkyunkwan University, Seoul, Republic of Korea, [¶]Department of Health, Behavior and Society and Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA, [**]Department of Occupational Medicine, Kangbuk Samsung Hospital, School of Medicine, [††]Department of Family Medicine and Health Screening Center, Kangbuk Samsung Hospital, School of Medicine, Sungkyunkwan University, [‡‡]Complex Disease and Genome Epidemiology Branch, Department of Epidemiology, School of Public Health, Seoul National University, Seoul, [§§]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Korea University Ansan Hospital, Ansan, [¶¶]Department of Preventive Medicine, School of Medicine, Ajou University, Suwon, [***]Department of Internal Medicine, School of Medicine, Ewha Womans University, and [†††]Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
**Corresponding author: H.-L. Kim, Department of Biochemistry, School of Medicine, Ewha Womans University, 1071, Anyangcheon-ro, Yancheon-gu, Seoul 158-710, Republic of Korea. E-mail: hyung@ewha.ac.kr*

Although several genome-wide association (GWA) studies of human personality have been recently published, genetic variants that are highly associated with certain personality traits remain unknown, due to difficulty reproducing results. To further investigate these genetic variants, we assessed biological pathways using GWA datasets. Pathway analysis using GWA data was performed on 1089 Korean women whose personality traits were measured with the Revised NEO Personality Inventory for the 5-factor model of personality. A total of 1042 pathways containing 8297 genes were included in our study. Of these, 14 pathways were highly enriched with association signals that were validated in 1490 independent samples. These pathways include association of: Neuroticism with axon guidance [L1 cell adhesion molecule (L1CAM) interactions]; Extraversion with neuronal system and voltage-gated potassium channels; Agreeableness with L1CAM interaction, neurotransmitter receptor binding and downstream transmission in postsynaptic cells; and Conscientiousness with the interferon-gamma and platelet-derived growth factor receptor beta polypeptide pathways. Several genes that contribute to top-ranked pathways in this study were previously identified in GWA studies or by pathway analysis in schizophrenia or other neuropsychiatric disorders. Here we report the first pathway analysis of all five personality traits. Importantly, our analysis identified novel pathways that contribute to understanding the etiology of personality traits.

Keywords: Axon guidance, behavior, Big Five personality, CAM, five-factor model, genome-wide association study, GWA study, pathway analysis, personality, potassium channel

Received 10 October 2014, revised 18 january 2015, 23 february 2015 and 5 March 2015, accepted for publication 10 March 2015

Personality is an important quantitative trait that affects behavior and lifestyle, and is associated with health and human disease. Recent personality studies have focused on the 'Five-Factor Model (FFM)' of the so-called 'Big Five' elements (Costa & Mccrae 1992; Sutin *et al.* 2010; Terracciano *et al.* 2010). The FFM of personality consists of five broad traits: Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness. These traits encapsulate most of the differences in personality across individuals, and they are linked to behavior, emotion, motivation and cognition (Deyoung *et al.* 2010). The Revised NEO Personality Inventory (NEO-PI-R) is widely used to analyze these personality traits (Costa & Mccrae 1992). The Korean version of the NEO-PI-R has been used in the Korean population with good reliability and validity (Ahn & Chae 1997). This instrument has a robust factor structure that has been replicated in more than 50 other cultures (Mccrae *et al.* 2005) in addition to that of Korea (Ahn & Chae 1997; Piedmont & Chae 1997).

With the development of behavioral genetics, there is growing interest in the genetic effects of personality. Recently, genome-wide association (GWA) studies have been used to identify common personality variants and to broaden interest in the field of personality genetics. However, most of the results from these studies show weak effects, and associations that are not reproduced in independent samples (Terracciano *et al.* 2010). Previously, we identified a novel region on olfactory receptor 1A2 (*OR1A2*) that is associated with Neuroticism in young Korean women, which was reproduced in an independent cohort (Kim *et al.* 2013). Nevertheless, the association of *OR1A2* has not yet been found in other populations but our study.

Despite the use of high-density single-nucleotide polymorphism (SNP) chips and large populations, GWA study

results are unable to adequately explain the genetic architecture of personality (Kim *et al.* 2013; Terracciano *et al.* 2010; Verweij *et al.* 2010). Many of the common variants identified by GWA studies are responsible for only a small portion of genetic variations, making it difficult to link them to the heritability of personality. There is less understanding of the neurobiology of personality traits than of psychiatric disorders. Genetic research on psychiatric disorders has progressed using candidate gene approaches, GWA studies, copy number variation studies, pathway/network studies and exome sequencing. Personality research has followed a similar path. A number of studies suggest that using a collection of weak associations may lead to meaningful results (Braun & Buetow 2011; Segre *et al.* 2010; Zhang *et al.* 2010). In psychiatric disorders, many findings suggest the importance of neurotransmitter-related pathways (Sun *et al.* 2010). Personality is a polygenic trait affected by many genes that individually have small effects. At the stringent genome-wide significance level, many markers that are moderately associated with personality are missed in GWA studies. By examining a large number of genetic variants in a biological pathway, meaningful variants with weak associations can be identified, giving greater understanding of the neurobiology of personality traits.

Pathway analysis offers a complementary approach to interpreting GWA studies by incorporating repositories of expert knowledge found in biological pathway databases. In a GWA study for Neuroticism (Verweij *et al.* 2010), pathway-based approaches were applied to personality traits; however, no significant biological or canonical pathways were identified. Because the variants identified in GWA datasets of East Asian cohorts differed from those of Caucasian cohorts, we performed pathway-based or Gene Set Enrichment Analysis (GSEA) without focusing on specific candidate genes. We hypothesized that personality traits may be determined by an accumulation of genetic variants and the interactions of those variants within their biological pathways. No GWA dataset-based pathway analysis has been performed for the FFM of personality. In this study, we attempted to identify pathways associated with personality traits by applying the modified GSEA method to a GWA dataset.

Materials and methods

GWA datasets: Stage 1

Participants were recruited from the Young Women Cohort in Korea, which was initiated in 2008, and included samples from 2000 Korean women with recorded genotypes. More than 50 traits were extensively examined through physical examinations and laboratory tests. Within this group, 1089 Korean women (18–40 years old) were included in our analysis. We confirmed that none of the subjects had received treatment for psychiatric disorders or taken psychoactive drugs. We used a GWA dataset of personality traits in young Korean women (Kim *et al.* 2013). Personality traits were assessed using the Korean short version of the NEO-PI-R, (Costa & Mccrae 1992), which is a 90-item measure of the five factors of personality (PSI Consulting Corp., Seoul, Korea). The questionnaire consisted of 18 items for each factor: Neuroticism, Extraversion, Openness to experience, Agreeableness and Conscientiousness. The details of subject enrollment and genotyping have been previously reported (Kim *et al.* 2013). Samples were genotyped using the Illumina Human 1 M-Duo DNA Analysis BeadChip kit (Illumina Inc., San Diego, CA, USA). After completing quality control procedures to eliminate ineligible subjects (minor allele frequency <0.05, Hardy Weinberg Equilibrium P-value $<10^{-6}$, and genotyping rate <0.95), 1089 participants were included in our GWA analysis.

Single-nucleotide polymorphism imputation was performed using IMPUTE2 (v2.1.2) software (Howie *et al.* 2009) after pre-phasing the genotype data using SHAPEIT (Delaneau *et al.* 2012) due to different genotyping platforms between discovery and replication sets. Based on NCBI build 36, we used 85 Japanese individuals from Tokyo, Japan and 85 Han Chinese individuals from Beijing, China comprised of 1.39 million SNPs in HapMap3 (www.hapmap.org), and 90 Korean individuals comprised of 1.66 million SNPs in the Korean HapMap (http://www.khapmap.org) as a reference panel. Only SNPs with an imputation quality score (R^2) > 0.7 were retained. After imputation and quality control, 1 172 710 autosomal SNPs were used for the GWA study. Genome-wide association analyses were conducted using PLINK 1.9. Association analyses were performed on imputed SNPs, and standardized residuals were obtained using a linear regression model for each personality trait and age. Genomic control inflation factors (λ) were under 1.008 for all analyses. We did not correct for genomic control in the GWA analyses, as inflation was modest and plots of MDS and PCA suggested that population structure could be disregarded for discovery and replication samples (Figure S1, Supporting Information).

Statement of ethics

The Institutional Review Board of Ewha Womans University Mokdong Hospital approved this study and written informed consent was obtained from all participants. This research study followed all applicable institutional and governmental regulations concerning the ethical use of human volunteers.

Pathway analysis

To discover biological pathways that may be enriched in genes that are moderately associated with personality traits, we applied an adapted GSEA framework to the GWA study data using MAGENTA (Segre *et al.* 2010). This algorithm does not require individual genotypes in the association scans to estimate the significance of gene set enrichment. Detailed information on the software tool was described by Segre *et al.* (2010). Briefly, the steps of MAGENTA analysis were as follows: (1) SNP association P-values and chromosome positions from the GWA studies were used as input; (2) each gene was scored by the most significant P-value among all of SNPs located within the gene or up to ±20 kb from the gene; (3) by applying a step-wise multiple linear regression analysis, gene scores were corrected for six confounders, e.g. gene size (kb), number of SNPs per kb, number of independent SNPs per kb, number of recombination hotspots per kb, linkage disequilibrium (LD) units per kb, and genetic distance; and (4) gene set enrichment P-values were determined by analyzing gene sets enriched with highly ranked gene scores. To minimize biases caused by LD on pathway analysis, SNPs were pruned by PLINK, using the 'indep-pairwise' option with the parameters for window size, step and r^2 set to 50, 5 and 0.2, respectively. About 222 639 SNPs remained after pruning and they were used to correct the LD on pathway analysis. The thresholds for pathway size used a minimum of 10 and maximum of 400 genes, and the human leukocyte antigen (HLA) region was excluded from the analysis. The GSEA algorithm in MAGENTA tested for over-representation of genes in a given gene set above a predetermined gene score rank cutoff. At a given significance threshold (75th or 95th percentile of all gene scores), we calculated the observed number of gene scores with a ranked score above the 75th or 95th percentile. This was compared to the rank that was observed in 10 000 random samples of identically sized gene sets, thus generating a nominal GSEA P-value (P_{gs}) for each pathway. Segre *et al.* (2010) suggested that the 75th percentile cutoff can be used for diseases or traits that are highly polygenic with many associations of weak effects, while the 95th percentile cutoff can detect the enrichment of multiple moderate effects. Hence, we chose the 75th percentile cutoff because of the polygenic nature of the personality traits. Gene sets were defined using the molecular signature database (MSigDB) v4.0 (Subramanian *et al.* 2005)

(downloaded in December 2014) curated from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto 2000), REACTOME (Joshi-Tope *et al.* 2005) and Pathway Interaction Database (PID) (Schaefer *et al.* 2009). We defined a nominal uncorrected significance level of $P_{gs} < 0.01$. To correct for multiple hypothesis testing, the false discovery rate (FDR) was separately used for each database: KEGG (186 pathways), REACTOME (660 pathways) and PID (196 pathways). The threshold of FDR < 0.25 denotes the confidence of 'possible' or 'hypothesis', while the threshold of FDR < 0.05 denotes 'high confidence' or 'statistical significance' (Zhang *et al.* 2010).

Reproducibility of findings: Stage 2

Pathways that provided significant evidence (nominal GSEA *P*-value < 0.01) in Stage 1, were tested in the independent cohort for replicability. A GWA dataset from the Ansung-Ansan cohort of the Korean Genome Epidemiology study (Cho *et al.* 2009) was used to confirm pathways identified in Stage 1. Samples were genotyped using the Affymetrix Genome-Wide Human SNP array 5.0 (Affymetrix, Santa Clara, CA, United States) for 1490 women (46–80 years old). Personality traits were assessed using the Korean short version of the NEO-PI-R, consistent with Stage 1. Quality control procedures for SNPs and samples were performed as in Stage 1. After imputation and quality control, 849 490 autosomal SNPs were used for the GWA study. Genome-wide association analyses were conducted using PLINK 1.9. About 135 404 SNPs remained after pruning, and were used to correct the LD on the pathway analysis. In the pathway analysis using MAGENTA, conditions for the number of gene set permutations, gene boundary and pathway size were applied as in Stage 1. We defined reproducibility as pathways having nominal GSEA *P*-value <0.05 applied 75th percentile cutoff of all gene scores in Stage 2. We summarized the results from Stage 1 and Stage 2. The *P*-values of the two stages were combined using the 'MADAM' program in R (http://www.r-project.org). 'MADAM' is the classical version of Fisher's inverse chi-square method (Fisher 1925), and it consists of computing a combined statistic from the different *P*-values and using this statistic for testing. The FDR was used separately calculated for each database: KEGG (186 pathways), REACTOME (660 pathways) and PID (196 pathways).

Results

We performed pathway analysis to identify the biological function of gene sets with multiple weak associations using data from GWA studies of personality. Overall, 1042 pathways and 8297 genes (186/5267, 660/6023 and 196/2552 pathways/genes from KEGG, REACTOME and PID, respectively) were included in our study. We obtained a nominal GSEA *P*-value by running MAGENTA on 10 000 simulated random gene sets (upper bound = 1 000 000 permutations). For each simulation, *P*-values of the most significant SNP per gene for all genes that ranked above the 75th percentile were extracted. The most significant gene sets, with a cut-off of nominal $P_{gs} < 0.01$ in MAGENTA, are shown in Tables. Biological pathways that were strongly associated with Neuroticism (Table 1), Extraversion (Table 2), Openness (Table 3), Agreeableness (Table 4) and Conscientiousness (Table 5) are shown.

In the discovery (Stage 1) dataset, 12 gene sets were significantly associated with Neuroticism at the nominal, uncorrected significance level of 0.01, but none at the FDR threshold of 0.05. The strongest overall association in the analysis was between axon guidance gene sets and Neuroticism (FDR = 0.09) (Table 1). The association of axon guidance with Neuroticism was confirmed in the reproducibility sample (Stage 2) (nominal $P_{gs} = 1.8 \times 10^{-3}$) and the combined *P*-value across two Stages showed statistically significance (FDR = 1.6×10^{-3}). Interestingly, three pathways comprising cell adhesion molecules (CAMs), which are known to be involved in schizophrenia, bipolar disorder and personality characteristics such as excitement seeking (O'dushlaine *et al.* 2011; Terracciano *et al.* 2011), had significant enrichment of associated SNPs in Neuroticism. Neural cell adhesion molecule 1 (NCAM1) interaction, NCAM signaling for neurite outgrowth, and L1CAM interaction gene sets showed possible association with Neuroticism (FDR = 0.12, 0.14 and 0.43, respectively). Among them, the association of L1CAM interaction was reproduced in Stage 2 (nominal $P_{gs} = 0.05$). We also found that two pathways involving semaphorin (Sema) from the REACTOME dataset were significantly enriched with Neuroticism-related genes. Although collapsin-response mediator proteins (CRMPs) in Sema3A signaling was showed significant association (FDR = 0.06), it was not reproducible in Stage 2. Further examination of the gene content of these pathways revealed some overlap. The contactin associated protein 1 (*CNTNAP1*) gene was the most strongly associated gene in axon guidance and L1CAM interactions (gene *P*-value = 4.0×10^{-6} and best SNP *P*-value = 1.8×10^{-5}) (Table S1). Axon guidance gene sets from the REACTOME dataset formed a parent pathway of hierarchical sub-pathways including CRMPs in Sema3A signaling, NCAM1 interactions, semaphorin interaction, NCAM signaling for neurite outgrowth, and L1CAM interaction gene sets. The gene netrin 1 (NTN1) was shared by axon guidance from the REACTOME and KEGG databases (gene *P*-value = 1.9×10^{-3} and best SNP *P*-value = 1.1×10^{-4}).

For Extraversion, 33 pathways were significantly enriched with association signals at the nominal $P_{gs} < 0.01$ level (Table 2). Of these, associations of the neuronal system, voltage-gated potassium channels and potassium channels from the REACTOME dataset (FDR = 0.15, 0.09 and 0.14, respectively) were confirmed in replicate samples ($P_{gs} = 0.02$, 4.0×10^{-3}, and 0.02, respectively). Their combined *P*-values of two Stages were also statistically significant or possible (FDR = 0.02, 0.02 and 0.17, respectively). Potassium voltage-gated channel subfamily genes [e.g. potassium voltage-gated channel shaker-related subfamily member 7 (*KCNA7*), *KCNA1*, potassium voltage-gated channel KQT-like subfamily member 2 (*KCNQ2*)] were shared by these three pathways (Table S2).

In the pathway analysis for Openness, purine metabolism from the KEGG dataset was the only gene set that passed the multiple testing significance threshold using MAGENTA (FDR = 0.02) within the five personality traits, but it was not reproducible ($P_{gs} = 0.55$) (Table 3.). Interestingly, circadian clock pathways were associated with Openness, although Terracciano *et al.* reported an association between the *CLOCK* gene and Agreeableness (Terracciano *et al.* 2010). Some studies reported a correlation between eveningness and Openness (Tsaousis 2010). BMAL1: CLOCK/NPAS2 activates circadian expression from REACTOME showing associated signals at the nominal $P_{gs} < 0.01$ level, but narrowly failing to be reproducible ($P_{gs} = 0.60$).

For Agreeableness, the pathways involved in axon guidance (e.g. L1CAM interactions and axon guidance) or neuronal system (e.g. neurotransmitter receptor binding

Table 1: In Neuroticism, the most significant biological pathways or gene sets following the gene set enrichment analysis of personality GWA data

		Stage 1				Stage 2			Combined P ‡	
Neuroticism Database	Biological pathway	Mean gene size (kb)	Expected/ Observed of genes*	Nominal GSEA, P-value (P_{gs})	FDR †	Mean gene size (kb)	Expected/ Observed of genes *	Nominal GSEA, P-value (P_{gs})	P_{gs}	FDR †
REACTOME	**Axon guidance**	**121**	**57/82**	$\mathbf{8.3\times10^{-5}}$	**0.09***	**123**	**56/75**	$\mathbf{1.8\times10^{-3}}$	2.5×10^{-5}	$1.6\times10^{-3**}$
REACTOME	CRMPS in SEMA3A signaling	95	3/9	6.0×10^{-4}	0.06*	95	3/4	0.42	2.3×10^{-3}	0.32
KEGG	**Axon guidance**	**143**	**31/46**	$\mathbf{1.5\times10^{-3}}$	**0.23***	**147**	**30/39**	**0.03**	5.3×10^{-4}	0.10*
REACTOME	NCAM1 interaction	88	9/18	1.6×10^{-3}	0.12*	93	9/12	0.17	2.5×10^{-3}	0.32
REACTOME	NCAM signaling for neurite outgrowth	89	16/27	2.2×10^{-3}	0.14*	93	15/19	0.15	2.9×10^{-3}	0.32
REACTOME	Semaphorin interaction	84	16/26	2.8×10^{-3}	0.20*	85	15/21	0.06	1.6×10^{-3}	0.32
REACTOME	Myogenesis	128	6/13	3.0×10^{-3}	0.20*	127	6/6	0.63	0.01	0.78
PID	ATF2 pathway	66	15/24	4.8×10^{-3}	0.24*	67	14/14	0.55	0.02	0.29
PID	ERBB1 internalization pathway	96	10/18	5.6×10^{-3}	0.29	96	10/10	0.54	0.02	0.29
PID	TCR RAS pathway	115	3/8	6.5×10^{-3}	0.43	115	3/5	0.20	0.01	0.29
REACTOME	**L1CAM interactions**	**114**	**19/29**	$\mathbf{7.3\times10^{-3}}$	**0.43**	**115**	**19/25**	**0.05**	3.5×10^{-3}	0.33
PID	FRA pathway	34	9/16	8.9×10^{-3}	0.29	35	9/7	0.81	0.4	0.29

Nominal gene set enrichment analysis (GSEA) *P*-values and false discovery rates (FDR) were computed for biological gene sets that were taken from three different resources using MAGENTA. Results are presented for the significant pathways ($P<0.01$) that are sorted by nominal GSEA P-value of MAGENTA.

*Gene P-value cutoff was defined as a 75 percentile (top 25%) of all gene P-values in the genome. The number of genes per gene set analyzed by MAGENTA in column four is after removing genes with no SNPs in their extended gene boundaries and after adjusting for physical proximity between subsets of genes in a gene set.

†False discovery rate (FDR) that was separately calculated for each database: KEGG (186 pathways), REACTOME (660 pathways) and PID (196 pathways). (**) specifies a gene set that passes the FDR cutoff <0.05. The asterisk (*) refers to pathways with an FDR < 0.25. Pathways with $P<0.05$ in the replication set are bolded. The gene set size was restricted to 10–400 genes.

‡Combined nominal P-value (P_{gs}) of Stage 1 and Stage 2 were calculated by Fisher's method.

Table 2: In Extraversion, the most significant biological pathways or gene sets following gene set enrichment analysis of personality GWA data

		Stage 1				Stage 2			Combined $P^{\ddagger}$	
Extraversion Database	Biological pathway	Mean gene size (kb)	Expected/ Observed of genes*	Nominal GSEA, *P*-value (P_{gs})	FDR†	Mean gene size (kb)	Expected/ Observed of genes*	Nominal GSEA, *P*-value (P_{gs})	P_{gs}	FDR†
REACTOME	**Neuronal system**	**114**	**66/91**	$\mathbf{1.0\times10^{-4}}$	**0.15***	**116**	**64/79**	**0.02**	3.0×10^{-5}	0.02**
REACTOME	G alpha q signaling events	69	42/61	4.0×10^{-4}	0.16*	70	42/39	0.70	2.6×10^{-3}	0.28
REACTOME	Gastrin CREB signaling pathway via PKC and MAPK	72	47/67	6.0×10^{-4}	0.13*	73	46/44	0.68	3.6×10^{-3}	0.30
REACTOME	NCAM signaling for neurite outgrowth	91	16/27	1.0×10^{-3}	0.11*	92	15/16	0.40	3.5×10^{-3}	0.30
REACTOME	Synthesis of phosphatidylethanolamine	43	3/8	1.1×10^{-3}	0.29	47	3/3	0.47	4.5×10^{-3}	0.32
PID	Estrogen receptor nongenomic pathway	104	10/19	1.4×10^{-3}	0.20*	104	10/4	1.00	0.01	0.24*
REACTOME	**Voltage-gated potassium channels**	**130**	**11/20**	$\mathbf{1.4\times10^{-3}}$	**0.09***	**130**	**11/19**	$\mathbf{4.0\times10^{-3}}$	7.3×10^{-5}	0.02**
KEGG	Small cell lung cancer	99	20/33	1.8×10^{-3}	0.18*	102	19/20	0.44	6.4×10^{-3}	0.37
REACTOME	Axon guidance	121	57/76	2.0×10^{-3}	0.16*	123	56/63	0.13	2.5×10^{-3}	0.28
REACTOME	Signaling by EGFR in cancer	96	26/39	2.1×10^{-3}	0.13*	98	25/26	0.42	7.1×10^{-3}	0.47
PID	Netrin pathway	181	8/16	2.3×10^{-3}	0.11*	186	8/6	0.83	0.01	0.24*
REACTOME	ENOS activation and regulation	42	5/11	2.3×10^{-3}	0.11*	42	5/8	0.07	1.7×10^{-3}	0.27
PID	Endothelin pathway	112	16/26	2.9×10^{-3}	0.13*	113	15/19	0.17	4.2×10^{-3}	0.24*
KEGG	Purine metabolism	90	37/53	3.0×10^{-3}	0.24*	91	37/35	0.67	0.01	0.37
PID	NFAT pathway	98	14/23	3.0×10^{-3}	0.10*	102	13/14	0.39	9.1×10^{-3}	0.24*
REACTOME	**Potassium channels**	**100**	**24/37**	$\mathbf{3.1\times10^{-3}}$	**0.14***	**100**	**24/33**	**0.02**	7.6×10^{-4}	0.17*
REACTOME	O linked glycosylation of mucins	125	13/22	3.2×10^{-3}	0.16*	124	12/9	0.90	0.02	0.76
PID	PDGFRB pathway	85	31/45	3.5×10^{-3}	0.13*	86	31/31	0.49	0.01	0.24*
PID	PI3KCI pathway	63	12/21	4.2×10^{-3}	0.12*	63	12/8	0.92	0.03	0.24*
KEGG	GLIOMA	91	16/26	4.3×10^{-3}	0.18*	94	15/15	0.55	0.02	0.37
PID	KIT pathway	88	13/21	5.8×10^{-3}	0.11*	90	12/15	0.20	8.8×10^{-3}	0.24*
PID	CXCR3 pathway	67	10/18	6.5×10^{-3}	0.11*	67	10/6	0.94	0.04	0.28
KEGG	Nicotinate and nicotinamide metabolism	54	6/12	6.6×10^{-3}	0.18*	54	6/3	0.96	0.04	0.45
KEGG	Axon guidance	144	31/43	7.3×10^{-3}	0.18*	148	30/27	0.73	0.03	0.45
PID	MET pathway	85	20/30	7.7×10^{-3}	0.10*	86	19/20	0.46	0.02	0.24*
PID	Ecadherin stabilization pathway	117	11/18	8.0×10^{-3}	0.09*	117	11/12	0.35	0.02	0.24*
PID	Insulin pathway	100	11/19	8.1×10^{-3}	0.09*	101	11/13	0.30	0.02	0.24*
REACTOME	Signaling by ERBB2	116	24/34	8.4×10^{-3}	0.27	118	23/20	0.79	0.04	0.89
REACTOME	Nephrin interactions	193	5/10	8.5×10^{-3}	0.23*	209	4/3	0.84	0.04	0.89
PID	Androgen receptor nongenomic pathway	90	8/14	9.3×10^{-3}	0.10*	92	7/6	0.77	0.04	0.28
PID	EPHA2 forward pathway	118	5/10	9.3×10^{-3}	0.13*	118	5/6	0.33	0.02	0.24*
REACTOME	Signaling by Rho GTPases	126	24/35	9.3×10^{-3}	0.29	129	24/21	0.79	0.04	0.89
KEGG	Calcium Signaling pathway	121	41/55	9.8×10^{-3}	0.25*	124	40/44	0.26	0.02	0.37

Nominal gene set enrichment analysis (GSEA) *P*-values and false discovery rates (FDR) were computed for biological gene sets that were taken from three different resources using MAGENTA. Results are presented for the significant pathways ($P < 0.01$) that are sorted by nominal GSEA *P*-value of MAGENTA.

*Gene *P*-value cutoff was defined as a 75 percentile (top 25%) of all gene *P*-values in the genome. The number of genes per gene set analyzed by MAGENTA in column four is after removing genes with no SNPs in their extended gene boundaries and after adjusting for physical proximity between subsets of genes in a gene set.

†False discovery rate (FDR) that was separately calculated for each database: KEGG (186 pathways), REACTOME (660 pathways) and PID (196 pathways). (**) specifies a gene set that passes the FDR cutoff <0.05. The asterisk (*) refers to pathways with an FDR < 0.25. Pathways with $P < 0.05$ in the replication set are bolded. The gene set size was restricted to 10–400 genes.

‡Combined nominal *P*-value (P_{gs}) of Stage 1 and Stage 2 were calculated by Fisher's method.

Table 3: In Openness, the most significant biological pathways or gene sets following gene set enrichment analysis of personality GWA data

		Stage 1				Stage 2			Combined $P^{\ddagger}$	
Openness Database	Biological pathway	Mean gene size (kb)	Expected/ Observed of genes *	Nominal GSEA, *P*-value (P_{gs})	FDR†	Mean gene size (kb)	Expected/ Observed of genes*	Nominal GSEA, *P*-value (P_{gs})	P_{gs}	FDR†
KEGG	Purine metabolism	90	37/59	2.0×10^{-4}	0.02**	90	37/37	0.55	1.1×10^{-3}	0.21*
REACTOME	cGMP effects	221	5/12	7.0×10^{-4}	0.13*	221	5/4	0.74	4.4×10^{-3}	0.52
REACTOME	Nitric oxide stimulates guanylate cyclase	195	6/14	7.0×10^{-4}	0.16*	195	6/5	0.79	4.7×10^{-3}	0.52
PID	Wnt signaling pathway	45	7/14	1.3×10^{-3}	0.13*	43	7/7	0.48	5.3×10^{-3}	0.29
REACTOME	Platelet homeostasis	118	19/30	2.2×10^{-3}	0.40	124	18/12	0.96	0.02	0.89
PID	CXCR3 pathway	67	10/19	2.7×10^{-3}	0.15*	67	10/9	0.64	0.01	0.34
PID	CD8 TCR pathway	83	13/21	5.1×10^{-3}	0.28	84	12/13	0.45	0.02	0.34
REACTOME	CRMPs in SEMA3A signaling	95	3/8	5.4×10^{-3}	0.45	95	3/0	1.00	0.03	0.90
KEGG	Hematopoietic cell lineage	35	19/29	5.9×10^{-3}	0.41	36	18/15	0.85	0.03	0.49
REACTOME	BMAL1: CLOCK/NPAS2 activates circadian expression	100	9/16	6.1×10^{-3}	0.55	106	8/8	0.60	0.02	0.89
REACTOME	Circadian clock	87	13/21	6.6×10^{-3}	0.52	91	12/9	0.89	0.04	0.90
REACTOME	NCAM signaling for neurite outgrowth	90	16/25	6.9×10^{-3}	0.59	93	15/18	0.23	0.01	0.89
KEGG	Melanoma	89	17/26	8.7×10^{-3}	0.42	89	16/22	0.07	5.0×10^{-3}	0.31

Nominal gene set enrichment analysis (GSEA) *P*-values and false discovery rates (FDR) were computed for biological gene sets that were taken from three different resources using MAGENTA. Results are presented for the significant pathways ($P<0.01$) that are sorted by nominal GSEA *P*-value of MAGENTA.

*Gene *P*-value cutoff was defined as a 75 percentile (top 25%) of all gene *P*-values in the genome. The number of genes per gene set analyzed by MAGENTA in column four is after removing genes with no SNPs in their extended gene boundaries and after adjusting for physical proximity between subsets of genes in a gene set.

†False discovery rate (FDR) that was separately calculated for each database: KEGG (186 pathways), REACTOME (660 pathways) and PID (196 pathways). (**) specifies a gene set that passes the FDR cutoff <0.05. The asterisk (*) refers to pathways with an FDR < 0.25. Pathways with $P<0.05$ in the replication set are bolded. The gene set size was restricted to 10–400 genes.

‡Combined nominal *P*-value (P_{gs}) of Stage 1 and Stage 2 were calculated by Fisher's method.

Table 4: In Agreeableness, the most significant biological pathways or gene sets following gene set enrichment analysis of personality GWA data

		Stage 1				Stage 2			Combined $P^{\ddagger}$	
Agreeableness Database	Biological pathway	Mean gene size (kb)	Expected/ Observed of genes *	Nominal GSEA, *P*-value (P_{gs})	FDR†	Mean gene size (kb)	Expected/ Observed of genes *	Nominal GSEA, *P*-value (P_{gs})	P_{gs}	FDR†
KEGG	Taste transduction	76	11/21	9.0×10^{-4}	0.10*	79	11/13	0.23	2.0×10^{-3}	0.18*
REACTOME	**L1CAM interactions**	**114**	**19/31**	$\mathbf{1.7\times10^{-3}}$	**0.46**	**115**	**19/30**	$\mathbf{2.3\times10^{-3}}$	5.3×10^{-5}	0.02**
REACTOME	Gluconeogenesis	33	7/14	2.2×10^{-3}	0.86	33	7/11	0.07	1.4×10^{-3}	0.16*
PID	mTOR pathway	68	17/27	3.9×10^{-3}	0.82	69	16/17	0.48	0.01	1.00
KEGG	MAPK signaling pathway	85	62/80	4.9×10^{-3}	0.28	86	61/62	0.47	0.02	0.95
REACTOME	**Neurotransmitter receptor binding and downstream transmission in the postsynaptic cell**	**127**	**31/45**	$\mathbf{4.9\times10^{-3}}$	**0.64**	**130**	**30/39**	**0.04**	1.8×10^{-3}	0.17*
REACTOME	**Axon guidance**	**121**	**57/74**	$\mathbf{5.7\times10^{-3}}$	**0.62**	**123**	**56/73**	$\mathbf{5.7\times10^{-3}}$	3.7×10^{-4}	0.06*
KEGG	**Vibrio cholera infection**	**73**	**13/21**	$\mathbf{6.4\times10^{-3}}$	**0.39**	**76**	**12/18**	**0.03**	1.9×10^{-3}	0.18*
REACTOME	Tryptophan catabolism	53	3/7	7.3×10^{-3}	0.65	53	3/4	0.29	0.02	0.57
REACTOME	Glucose metabolism	41	14/23	8.4×10^{-3}	0.57	41	15/18	0.17	0.01	0.56
REACTOME	**Transmission across chemical synapses**	**124**	**43/57**	$\mathbf{8.9\times10^{-3}}$	**0.63**	**128**	**42/60**	$\mathbf{3.0\times10^{-4}}$	3.7×10^{-5}	0.02**
REACTOME	Recycling pathway of L1	98	6/11	9.8×10^{-3}	0.67	98	6/9	0.07	5.6×10^{-3}	0.37

Nominal gene set enrichment analysis (GSEA) *P*-values and false discovery rates (FDR) were computed for biological gene sets that were taken from three different resources using MAGENTA. Results are presented for the significant pathways ($P < 0.01$) that are sorted by nominal GSEA *P*-value of MAGENTA.

*Gene *P*-value cutoff was defined as a 75 percentile (top 25%) of all gene *P*-values in the genome. The number of genes per gene set analyzed by MAGENTA in column four is after removing genes with no SNPs in their extended gene boundaries and after adjusting for physical proximity between subsets of genes in a gene set.

†False discovery rate (FDR) that was separately calculated for each database: KEGG (186 pathways), REACTOME (660 pathways) and PID (196 pathways). (**) specifies a gene set that passes the FDR cutoff <0.05. The asterisk (*) refers to pathways with an FDR < 0.25. Pathways with $P < 0.05$ in the replication set are bolded. The gene set size was restricted to 10–400 genes.

‡Combined nominal *P*-value (P_{gs}) of Stage 1 and Stage 2 were calculated by Fisher's method.

Table 5: In Conscientiousness, the most significant biological pathways or gene sets following gene set enrichment analysis of personality GWA data

		Stage 1				Stage 2			Combined $P^{\ddagger}$	
Conscientiousness Database	Biological pathway	Mean gene size (kb)	Expected/ Observed of genes *	Nominal GSEA, *P*-value (P_{gs})	FDR †	Mean gene size (kb)	Expected/ Observed of genes *	Nominal GSEA, *P*-value (P_{gs})	P_{gs}	FDR †
PID	Hedgehog gli pathway	76	12/22	1.0×10^{-3}	0.12*	80	11/14	0.19	1.8×10^{-3}	0.05**
PID	RAC1 pathway	63	13/23	1.3×10^{-3}	0.07*	64	12/15	0.22	2.6×10^{-3}	0.06*
KEGG	Regulation of actin cytoskeleton	86	50/66	4.5×10^{-3}	0.88	87	48/46	0.67	0.02	0.46
PID	**IFN-γ pathway**	**75**	**10/18**	$\mathbf{3.7\times10^{-3}}$	**0.17***	**75**	**10/21**	$\mathbf{1.0\times10^{-4}}$	7.6×10^{-6}	1.5×10^{-3}**
PID	Integrin CS pathway	82	7/13	4.1×10^{-3}	0.14*	84	6/4	0.91	0.03	0.20*
REACTOME	Sphingolipid de novo biosynthesis	95	8/14	7.0×10^{-3}	1.00	100	7/10	0.15	1.0×10^{-2}	0.53
PID	**TCPTP pathway**	**75**	**11/18**	$\mathbf{9.9\times10^{-3}}$	**0.25***	**74**	**10/18**	$\mathbf{4.2\times10^{-3}}$	4.3×10^{-4}	0.02**
PID	**PDGFRB pathway**	**85**	**32/44**	$\mathbf{8.8\times10^{-3}}$	**0.24***	**85**	**31/45**	$\mathbf{2.9\times10^{-3}}$	3.1×10^{-4}	0.02**
REACTOME	Lipoprotein metabolism	52	7/13	8.8×10^{-3}	0.92	134	3/5	0.21	0.02	0.61

Nominal gene set enrichment analysis (GSEA) *P*-values and false discovery rates (FDR) were computed for biological gene sets that were taken from three different resources using MAGENTA. Results are presented for the significant pathways ($P < 0.01$) that are sorted by nominal GSEA *P*-value of MAGENTA.

*Gene *P*-value cutoff was defined as a 75 percentile (top 25%) of all gene *P*-values in the genome. The number of genes per gene set analyzed by MAGENTA in column four is after removing genes with no SNPs in their extended gene boundaries and after adjusting for physical proximity between subsets of genes in a gene set.

†False discovery rate (FDR) that was separately calculated for each database: KEGG (186 pathways), REACTOME (660 pathways) and PID (196 pathways). (**) specifies a gene set that passes the FDR cutoff <0.05. The asterisk (*) refers to pathways with an FDR < 0.25. Pathways with $P < 0.05$ in the replication set are bolded. The gene set size was restricted to 10–400 genes.

‡Combined nominal *P*-value (P_{gs}) of Stage 1 and Stage 2 were calculated by Fisher's method.

and downstream transmission in the postsynaptic cell and transmission across chemical synapses from REACTOME database) were significantly enriched with association signals at the nominal $P_{gs} < 0.01$ level and were reproducible with $P_{gs} < 0.05$ in Stage 2 (Table 4). The gene mitogen-activated protein kinase 1 (*MAPK1*) is involved in all four pathways, but does not contribute significantly to their gene enrichment score due to its insignificant association with Agreeableness (gene *P*-value = 0.11). In contrast, two genes [e.g. gamma-aminobutyric acid (*GABA*) A receptor alpha 2 (*GABRA2*) and protein kinase cAMP-dependent catalytic beta (*PRKACB*)] were significant associated with Agreeableness in neurotransmitter receptor binding and downstream in the postsynaptic cell and transmission across chemical synapses gene sets (gene *P*-value = 2.9×10^{-3} and 7.1×10^{-3}, respectively). Solute carrier family 6 (neurotransmitter transporter, GABA), member 11 (*SLC6A11*) significantly enriched transmission across chemical synapses gene sets (gene *P*-value = 2.9×10^{-3}) (Table S4). Taste transduction from the KEGG dataset was most strongly association with Agreeableness with a noteworthy FDR value (FDR = 0.10), and the combined *P*-value of two Stages showed suggestive significance (FDR = 0.18) although it was not reproducible in Stage 2 ($P_{gs} = 0.23$).

In Table 5, the interferon-gamma (IFN-γ), T-cell protein tyrosine phosphatase (TCPTP), and platelet-derived growth factor receptor beta (PDGFRB) pathways from the PID database were possibly associated with Conscientiousness after correcting for multiple testing (FDR = 0.17, 0.25 and 0.24, respectively), and they remained significant in the reproducibility dataset (nominal $P_{gs} = 1.0 \times 10^{-4}$, 4.2×10^{-3}, and 2.9×10^{-3}, respectively). The combined *P*-values of the three pathways across two Stages also showed statistically significance (FDR = 1.5×10^{-3}, 0.02 and 0.02, respectively). Interferon-gamma contributes to aging-associated psychiatric disorders (Oxenkrug 2011). Single-nucleotide polymorphisms of the 5′-upstream region of *PDGFRB* were reported to be associated with schizophrenia in a Korean population (Kim *et al.* 2008). The hedgehog gli and Ras-related C3 botulinum toxin substrate 1 (RAC1) pathways were associated with Conscientiousness with a nominal uncorrected significance, but they were not reproducible. The gene overlap for each pathway pair is shown in Table S5.

Discussion

We have presented the first pathway analysis of all five personality factors. Our results illustrate biological understanding and novel genetic associations to personality using GWA datasets. This study found that personality-specific enriched pathways were related to axon guidance through CAM signaling, synaptic transmission and ion channel activity such as potassium channels. The pathways were validated using discovery and validation datasets. In particular, CAM signaling pathways were associated with four personality dimensions, excluding Conscientiousness. We confirmed the association of L1CAM interaction gene sets in Neuroticism and Agreeableness. Interestingly, over the last few years, a number of studies have reported that schizophrenia and bipolar disorder, as well as other psychiatric disorders, were associated with CAM, which is responsible for synapse formation and normal cell transmission (Corvin 2010; O'dushlaine *et al.* 2011). Besides, neuronal CAM has been associated with drug abuse and personality characteristics such as novelty seeking and reward dependence (Yoo *et al.* 2012). Additionally, the contribution of potassium channels, which were enriched gene sets for Extraversion in this study, has been reported in psychiatric disorders (Judy & Zandi 2013; Zhang *et al.* 2006). Recent studies have reported that common genetic influences are shared among psychiatric disorders including schizophrenia, bipolar and major depressive disorders (Chang *et al.* 2013; Cross-Disorder Group of the Psychiatric Genomics Consortium 2013; Lichtenstein *et al.* 2009). Several genes that contributed to top-ranked pathways in this study had previously been identified by GWA studies or pathway analysis in schizophrenia or other neuropsychiatric disorder phenotypes. Relationships between personality and psychiatric disorders have been reported in previous studies (Bagby *et al.* 1997; Berenbaum & Fujita 1994; Hare *et al.* 2012; Horan *et al.* 2008; Middeldorp *et al.* 2011). Our results support the hypothesis that personality shares common genetic determinants with psychiatric disorders.

Association of the axon guidance pathway was highlighted in Neuroticism. Most of the top-listed pathways were child pathways of axon guidance, although most of them did not pass the threshold of FDR < 0.05, and were not reproducible. The significance of their parent pathway, axon guidance, indicates that the cumulative effect of the sibling pathways may play an important role in the biological function of personality. We identified several enriched genes [e.g. netrin 1 (*NTN1*) (Moore *et al.* 2007), semaphorin (*SEMA3B*, *SEMA4B*, *SEMA3C*) (Kolodkin 1996), nonreceptor tyrosine kinase (*FYN)* (Bashaw & Klein 2010), plexin B2 (*PLXNB2*) (Bashaw & Klein 2010)], which were members of well-characterized axon guidance pathway families. Brain function is based on precise neuronal-network formation during development, which is largely controlled by attractive and repulsive axon guidance molecules (Tessier-Lavigne & Goodman 1996). Many guidance molecules persist in the adult central nervous system, and extensive studies have shown that these factors have roles in maintenance and plasticity of neural circuits (Curinga & Smith 2008). Hence, Neuroticism may be influenced by the regulation of axonal growth and synaptogenesis, as observed in other psychiatric disorders (Lin *et al.* 2009; Wu *et al.* 2013).

The discovery of an association between Extraversion and potassium channels implies that the regulation of action potential and resting membrane potential in neurons may play important role in an extraverted personality. Interestingly, previous genetic studies have implicated potassium channel-related genes in bipolar disorders (Judy & Zandi 2013). A recent study reported that Extraversion was more strongly genetically correlated with bipolar disorder than with other personality dimensions of FFMs (Hare *et al.* 2012). The present findings indicate that potassium channels may be a common link between Extraversion and bipolar disorders.

Finally, an interesting result was the enrichment of synaptic transmission gene sets for Agreeableness. The neurotransmitter receptor binding and downstream transmission

in the postsynaptic cell from REACTOME database included GABA-A receptor α_2 (*GABRA2*) and G protein (*GNAL*). To date, a number of studies identifying the genetic factors of personality have focused on neurotransmitters, but most GWA studies failed to reproduce these findings (Munafo & Flint 2011). The GABA-A receptor was found to be involved in anxiety disorders (Crestani *et al.* 1999) and the pathogenesis of alcoholism (Sander *et al.* 1999). Studies on personality traits are largely lacking, although one group found that a polymorphism in the γ_2 subunit of the GABA-A receptor was associated with alcohol dependence comorbid with antisocial personality disorder (Loh *et al.* 2000; Moeller & Dougherty 2001). Agreeableness reflects the degree to which individuals differ in the development and maintenance of social relationships (Costa & Mccrae 1992). Miller *et al.* reported that lower levels of agreeableness were associated with higher levels of alcohol-related aggressivity (Miller *et al.* 2009). Agreeableness was a significant predictor of alcohol behaviors in a meta-analysis and a family study (Chassin *et al.* 2004; Malouff *et al.* 2007). Our results may bring us one step closer to understanding the genetic influence on alcohol behavior.

The MAGENTA algorithm adjusts for confounders of gene scores such as gene size, number of SNP, number of independent SNPs, number of recombination hotspots, LD and genetic distance. Previously, we reported a significant association between *OR1A2* and Neuroticism (Kim *et al.* 2013). We could not find gene enrichments in the olfactory signaling pathway using MAGENTA, although *OR1A2* showed a significant association with Neuroticism in gene-levels. Effects of correcting confounders may cause inconsistent results. In pathway analysis, LD must be accounted for to prevent highly correlated SNPs from biasing gene-level significance (Ramanan *et al.* 2012).

We observed some degree of pathway overlap in our results. We did not restrict the analysis of the databases to certain levels in the hierarchy. Several pathways from the REACTOME database in particular had a hierarchical structure. Therefore, the property of the database could obscure the significance after correction for multiple testing. Although correction for multiple comparisons must be applied to pathway *P*-values to control for false positives, most methods seem too conservative for pathway analyses because of dependence across pathways. These approaches to bias are best complemented by the reproducibility of pathway analysis findings in independent datasets (Ramanan *et al.* 2012). In this study, most reproducible pathways passed the threshold of FDR < 0.25 and their combined results across two Stages showed the FDR < 0.05. It is meaningful that the possible pathways were replicated independent samples.

As the first report on the biological etiology of personality traits using pathway analysis, this study will inspire further research on several levels. Our pathway analyses for personality implicate several gene sets involved in neuronal cell adhesion, neuronal ion channel functioning and synaptic transmission. The most notable finding is the significant convergence on key molecules in these pathways which have been broadly implicated in psychiatric disorders, including schizophrenia and bipolar disorders. By understanding the relationship between personality and psychiatric disorders, it might be possible to identify individuals at risk for developing the diseases as well as to provide behavioral intervention. Therefore, personality traits may be useful endophenotypes for psychiatric disorders. Further studies are required to reproduce these results in other cohorts and ethnicities. If these findings can be reproduced in other ethnicities, it will play an important role in comprehending the biological system of personality and the extent to which they may contribute to the genetic overlap between personality traits and psychiatric disorders. Our results provide an opportunity to look at how genetic architecture regulates the formation and maintenance of personality.

References

Ahn, C. & Chae, J. (1997) Standardization of the Korean version of the revised NEO personality inventory. *Kor J Counsel Psychother* **9**, 443–473.

Bagby, R.M., Bindseil, K.D., Schuller, D.R., Rector, N.A., Young, L.T., Cooke, R.G., Seeman, M.V., McCay, E.A. & Joffe, R.T. (1997) Relationship between the five-factor model of personality and unipolar, bipolar and schizophrenic patients. *Psychiatry Res* **70**, 83–94.

Bashaw, G.J. & Klein, R. (2010) Signaling from axon guidance receptors. *Cold Spring Harb Perspect Biol* **2**, a001941.

Berenbaum, H. & Fujita, F. (1994) Schizophrenia and personality: exploring the boundaries and connections between vulnerability and outcome. *J Abnorm Psychol* **103**, 148–158.

Braun, R. & Buetow, K. (2011) Pathways of distinction analysis: a new technique for multi-SNP analysis of GWAS data. *PLoS Genet* **7**, e1002101.

Chang, S.H., Gao, L., Li, Z., Zhang, W.N., Du, Y. & Wang, J. (2013) BDgene: a genetic database for bipolar disorder and its overlap with schizophrenia and major depressive disorder. *Biol Psychiatry* **74**, 727–733.

Chassin, L., Fora, D.B. & King, K.M. (2004) Trajectories of alcohol and drug use and dependence from adolescence to adulthood: the effects of familial alcoholism and personality. *J Abnorm Psychol* **113**, 483–498.

Cho, Y.S., Go, M.J., Kim, Y.J. *et al.* (2009) A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat Genet* **41**, 527–534.

Corvin, A.P. (2010) Neuronal cell adhesion genes: key players in risk for schizophrenia, bipolar disorder and other neurodevelopmental brain disorders? *Cell Adh Migr* **4**, 511–514.

Costa, P. & McCrae, R. (1992) *Revised NEO-Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (FFI) Manual*, Odessa, FL. Psychological Assessment Resources, Inc.

Crestani, F., Lorez, M., Baer, K., Essrich, C., Benke, D., Laurent, J.P., Belzung, C., Fritschy, J.M., Luscher, B. & Mohler, H. (1999) Decreased GABAA-receptor clustering results in enhanced anxiety and a bias for threat cues. *Nat Neurosci* **2**, 833–839.

Cross-Disorder Group of the Psychiatric Genomics Consortium (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* **45**, 984–994.

Curinga, G. & Smith, G.M. (2008) Molecular/genetic manipulation of extrinsic axon guidance factors for CNS repair and regeneration. *Exp Neurol* **209**, 333–342.

Delaneau, O., Marchini, J. & Zagury, J.F. (2012) A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179–181.

DeYoung, C.G., Hirsh, J.B., Shane, M.S., Papademetris, X., Rajeevan, N. & Gray, J.R. (2010) Testing predictions from personality neuroscience. Brain structure and the big five. *Psychol Sci* **21**, 820–828.

Fisher, R.A. (1925) *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh; London.

Hare, E., Contreras, J., Raventos, H., Flores, D., Jerez, A., Nicolini, H., Ontiveros, A., Almasy, L. & Escamilla, M. (2012) Genetic structure

of personality factors and bipolar disorder in families segregating bipolar disorder. *J Affect Disord* **136**, 1027–1033.
Horan, W.P., Blanchard, J.J., Clark, L.A. & Green, M.F. (2008) Affective traits in schizophrenia and schizotypy. *Schizophr Bull* **34**, 856–874.
Howie, B.N., Donnelly, P. & Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529.
Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L., Lewis, S., Birney, E. & Stein, L. (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res* **33**, D428–432.
Judy, J.T. & Zandi, P.P. (2013) A review of potassium channels in bipolar disorder. *Front Genet* **4**, 105.
Kanehisa, M. & Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30.
Kim, H.J., Kim, M.H., Choe, B.K., Kim, J.W., Park, J.K., Cho, A.R., Bae, H., Shin, D.H., Yim, S.V., Kwack, K., Kwon, Y.K. & Chung, J.H. (2008) Genetic association between 5′-upstream single-nucleotide polymorphisms of PDGFRB and schizophrenia in a Korean population. *Schizophr Res* **103**, 201–208.
Kim, H.N., Roh, S.J., Sung, Y.A., Chung, H.W., Lee, J.Y., Cho, J., Shin, H. & Kim, H.L. (2013) Genome-wide association study of the five-factor model of personality in young Korean women. *J Hum Genet* **58**, 667–674.
Kolodkin, A.L. (1996) Semaphorins: mediators of repulsive growth cone guidance. *Trends Cell Biol* **6**, 15–22.
Lichtenstein, P., Yip, B.H., Bjork, C., Pawitan, Y., Cannon, T.D., Sullivan, P.F. & Hultman, C.M. (2009) Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234–239.
Lin, L., Lesnick, T.G., Maraganore, D.M. & Isacson, O. (2009) Axon guidance and synaptic maintenance: preclinical markers for neurodegenerative disease and therapeutics. *Trends Neurosci* **32**, 142–149.
Loh, E.W., Higuchi, S., Matsushita, S., Murray, R., Chen, C.K. & Ball, D. (2000) Association analysis of the GABA(A) receptor subunit genes cluster on 5q33-34 and alcohol dependence in a Japanese population. *Mol Psychiatry* **5**, 301–307.
Malouff, J.M., Thorsteinsson, E.B., Rooke, S.E. & Schutte, N.S. (2007) Alcohol involvement and the Five-Factor model of personality: a meta-analysis. *J Drug Educ* **37**, 277–294.
McCrae, R.R., Terracciano, A. & Personality Profiles of Cultures Project (2005) Universal features of personality traits from the observer's perspective: data from 50 cultures. *J Pers Soc Psychol* **88**, 547–561.
Middeldorp, C.M., de Moor, M.H., McGrath, L.M. *et al.* (2011) The genetic association between personality and major depression or bipolar disorder. A polygenic score analysis using genome-wide association data. *Transl Psychiatry* **1**, e50.
Miller, C.A., Parrott, D.J. & Giancola, P.R. (2009) Agreeableness and alcohol-related aggression: the mediating effect of trait aggressivity. *Exp Clin Psychopharmacol* **17**, 445–455.
Moeller, F.G. & Dougherty, D.M. (2001) Antisocial personality disorder, alcohol, and aggression. *Alcohol Res Health* **25**, 5–11.
Moore, S.W., Tessier-Lavigne, M. & Kennedy, T.E. (2007) Netrins and their receptors. *Adv Exp Med Biol* **621**, 17–31.
Munafo, M.R. & Flint, J. (2011) Dissecting the genetic architecture of human personality. *Trends Cogn Sci* **15**, 395–400.
O'Dushlaine, C., Kenny, E., Heron, E., Donohoe, G., Gill, M., Morris, D., International Schizophrenia Consortium & Corvin, A. (2011) Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. *Mol Psychiatry* **16**, 286–292.
Oxenkrug, G. (2011) Interferon-gamma – inducible inflammation: contribution to aging and aging-associated psychiatric disorders. *Aging Dis* **2**, 474–486.
Piedmont, R. & Chae, J. (1997) Cross-cultural generalizability of the Five-Factor Model of personality: development and validation of the NEO-PI-R for Koreans. *J Cross Cul Psychol* **28**, 131–155.
Ramanan, V.K., Shen, L., Moore, J.H. & Saykin, A.J. (2012) Pathway analysis of genomic data: concepts, methods, and prospects for future development. *Trends Genet* **28**, 323–332.
Sander, T., Ball, D., Murray, R., Patel, J., Samochowiec, J., Winterer, G., Rommelspacher, H., Schmidt, L.G. & Loh, E.W. (1999) Association analysis of sequence variants of GABA(A) alpha6, beta2, and gamma2 gene cluster and alcohol dependence. *Alcohol Clin Exp Res* **23**, 427–431.
Schaefer, C.F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T. & Buetow, K.H. (2009) PID: the pathway interaction database. *Nucleic Acids Res* **37**, D674–679.
Segre, A.V., DIAGRAM Consortium, MAGIC investigators, Groop, L., Mootha, V.K., Daly, M.J. & Altshuler, D. (2010) Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* **6**, e1001058.
Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. & Mesirov, J.P. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–15550.
Sun, J., Jia, P., Fanous, A.H., van den Oord, E., Chen, X., Riley, B.P., Amdur, R.L., Kendler, K.S. & Zhao, Z. (2010) Schizophrenia gene networks and pathways and their applications for novel candidate gene selection. *PLoS One* **5**, e11351.
Sutin, A.R., Terracciano, A., Deiana, B., Uda, M., Schlessinger, D., Lakatta, E.G. & Costa, P.T. Jr. (2010) Cholesterol, triglycerides, and the Five-Factor Model of personality. *Biol Psychol* **84**, 186–191.
Terracciano, A., Esko, T., Sutin, A.R. *et al.* (2011) Meta-analysis of genome-wide association studies identifies common variants in CTNNA2 associated with excitement-seeking. *Transl Psychiatry* **1**, e49.
Terracciano, A., Sanna, S., Uda, M. *et al.* (2010) Genome-wide association scan for five major dimensions of personality. *Mol Psychiatry* **15**, 647–656.
Tessier-Lavigne, M. & Goodman, C.S. (1996) The molecular biology of axon guidance. *Science* **274**, 1123–1133.
Tsaousis, I. (2010) Circadian preferences and personality traits: a meta-analysis. *Eur J Personality* **24**, 356–373.
Verweij, K.J., Zietsch, B.P., Medland, S.E., Gordon, S.D., Benyamin, B., Nyholt, D.R., McEvoy, B.P., Sullivan, P.F., Heath, A.C., Madden, P.A., Henders, A.K., Montgomery, G.W., Martin, N.G. & Wray, N.R. (2010) A genome-wide association study of Cloninger's temperament scales: implications for the evolutionary genetics of personality. *Biol Psychol* **85**, 306–317.
Wu, L., Huang, Y., Li, J., Zhao, H., Du, H., Jin, Q., Zhao, X., Ma, H. & Zhu, G. (2013) Association study of the Fyn gene with schizophrenia in the Chinese-Han population. *Psychiatr Genet* **23**, 39–40.
Yoo, B.K., Shim, J.C., Lee, B.D., Kim, C., Chung, Y.I., Park, J.M., Kim, S.G., Kim, J.H., Lee, Y.M., Moon, E.S. & Kwon do, H. (2012) Association of the Neuronal Cell Adhesion Molecule (NrCAM) gene variants with personality traits and addictive symptoms in methamphetamine use disorder. *Psychiatry Investig* **9**, 400–407.
Zhang, K., Cui, S., Chang, S., Zhang, L. & Wang, J. (2010) i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res* **38**, W90–95.
Zhang, L., Li, X., Zhou, R. & Xing, G. (2006) Possible role of potassium channel, big K in etiology of schizophrenia. *Med Hypotheses* **67**, 41–43.

Acknowledgments

This research was supported by the National Research Foundation of Korea, funded by the Ministry of Education

(NRF-2010-0026606 and NRF-2013R1A1A2062702). The genotype data were gratefully made available by the Center for Genome Science, Korea National Institute of Health, Korea Centers for Disease Control and Prevention.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site:

Table S1: Gene/Gene set overlap matrix. Results are presented for the observed genes (>75th percentile cutoff) responsible for the enrichment signal in personality traits. Gene sets with $P < 0.05$ in the replication set are bolded. The yellow colors correspond to genes with gene *P*-value <0.05. Neuroticism.

Table S2: Gene/Gene set overlap matrix. Results are presented for the observed genes (>75th percentile cutoff) responsible for the enrichment signal in personality traits. Gene sets with $P < 0.05$ in the replication set are bolded. The yellow colors correspond to genes with gene *P*-value <0.05. Extraversion.

Table S3: Gene/Gene set overlap matrix. Results are presented for the observed genes (>75th percentile cutoff) responsible for the enrichment signal in personality traits. Gene sets with $P < 0.05$ in the replication set are bolded. The yellow colors correspond to genes with gene *P*-value <0.05. Openness.

Table S4: Gene/Gene set overlap matrix. Results are presented for the observed genes (>75th percentile cutoff) responsible for the enrichment signal in personality traits. Gene sets with $P < 0.05$ in the replication set are bolded. The yellow colors correspond to genes with gene *P*-value <0.05. Agreeableness.

Table S5: Gene/Gene set overlap matrix. Results are presented for the observed genes (>75th percentile cutoff) responsible for the enrichment signal in personality traits. Gene sets with $P < 0.05$ in the replication set are bolded. The yellow colors correspond to genes with gene *P*-value <0.05. Conscientiousness.

Figure S1: Multidimensional scaling (MDS) plot and principal component analysis (PCA) plot of samples of Stage 1 and Stage 2 in our study. CEU: Utah residents with ancestry from Northern and Western Europe, YRI: Yoruba in Ibadan, Nigeria, CHB: Han Chinese South, China, JPT: Japanese in Tokyo, Japan, KOR1: Korean, South Korea (Stage 1), KOR2: Korean, South Korea (Stage 2).