Title

Modeling prior information of common genetic variants improves gene discovery for neuroticism

Journal

Human Molecular Genetics, 26(22)

ISSN

0964-6906

Authors

Lo, Min-Tzu
Wang, Yunpeng
Kauppi, Karolina
et al.

Publication Date

2017-11-15

DOI

10.1093/hmg/ddx340

Peer reviewed

Human Molecular Genetics, 2017, Vol. 26, No. 22 4530–4539

doi: 10.1093/hmg/ddx340
Advance Access Publication Date: 30 August 2017
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Modeling prior information of common genetic variants improves gene discovery for neuroticism

Min-Tzu Lo[1,†], Yunpeng Wang[2,3,†], Karolina Kauppi[1,4], Nilotpal Sanyal[1], Chun-Chieh Fan[1,5], Olav B. Smeland[2,6], Andrew Schork[5,7], Dominic Holland[3], David A. Hinds[8], Joyce Y. Tung[8], Ole A. Andreassen[2,6], Anders M. Dale[1,3,9] and Chi-Hua Chen[1,*]

[1]Department of Radiology, Center for Multimodal Imaging and Genetics, University of California, San Diego, La Jolla, CA 92037, USA, [2]NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo 0407, Norway, [3]Department of Neurosciences, University of California, San Diego, La Jolla, CA 92037, USA, [4]Department of Radiation Sciences, Umea University, Umea 90187, Sweden, [5]Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92037, USA, [6]Division of Mental Health and Addiction, Oslo University Hospital, Oslo 0407, Norway, [7]Institute of Biological Psychiatry, Medical Health Center, Sct. Hans, Roskilde, 4000, Denmark, [8]23andMe, Inc., Mountain View, CA 94043, USA and [9]Department of Psychiatry, University of California, San Diego, La Jolla, CA 92037, USA

*To whom correspondence should be addressed. Tel: +1 8588223865; Fax: +1 8585341078; Email: chc101@ucsd.edu

Abstract

Neuroticism reflects emotional instability, and is related to various mental and physical health issues. However, the majority of genetic variants associated with neuroticism remain unclear. Inconsistent genetic variants identified by different genome-wide association studies (GWAS) may be attributable to low statistical power. We proposed a novel framework to improve the power for gene discovery by incorporating prior information of single nucleotide polymorphisms (SNPs) and combining two relevant existing tools, relative enrichment score (RES) and conditional false discovery rate (FDR). Here, SNP's conditional FDR was estimated given its RES based on SNP prior information including linkage disequilibrium (LD)-weighted genic annotation scores, total LD scores and heterozygosity. A known significant locus in chromosome 8p was excluded before estimating FDR due to long-range LD structure. Only one significant LD-independent SNP was detected by analyses of unconditional FDR and traditional GWAS in the discovery sample ($N = 59\,225$), and notably four additional SNPs by conditional FDR. Three of the five SNPs, all identified by conditional FDR, were replicated ($P < 0.05$) in an independent sample ($N = 170\,911$). These three SNPs are located in intronic regions of *CADM2*, *LINGO2* and *EP300* which have been reported to be associated with autism, Parkinson's disease and schizophrenia, respectively. Our approach using a combination of RES and conditional FDR improved power of traditional GWAS for gene discovery providing a useful framework for the analysis of GWAS summary statistics by utilizing SNP prior information, and helping to elucidate the links between neuroticism and complex diseases from a genetic perspective.

[†]These authors contributed equally to this work.

Received: April 11, 2017. **Revised:** July 10, 2017. **Accepted:** August 25, 2017

Introduction

The Big Five personality traits, accounted for by the Five Factor Model (FFM) of personality, were derived from factor analysis of rating scales and have been shown to be robust across all cultures and languages investigated (1,2). The dimensions of FFM are defined as agreeableness, conscientiousness, extraversion, neuroticism and openness to experience, and measure individual differences in behavior and experience (3).

In the FFM, neuroticism is the only dimension characterized by emotional instability involving the tendency to experience negative and distressing emotions, and summarizes the facets of anxiety, angry hostility, depression, self-consciousness, impulsiveness and vulnerability (4,5). There is substantial evidence that neuroticism is correlated with a wider range of mental and physical health problems than other personality traits, including depression, anxiety, substance use and schizophrenia (6–9), as well as cardiovascular diseases and asthma (10–12). Furthermore, neuroticism is referred to as a risk factor or predictor for psychiatric disorders (7,8,13–16) and explains part of the comorbidity among these disorders (6,17,18). In addition, recent studies also suggest that neuroticism is associated with increased risk of Alzheimer's disease (19,20). Neuroticism has been considered to possess robust predictions to mental and physical illnesses, which implies great significance for the public health (21).

Neuroticism is substantially heritable and about equally shaped by genetic and environmental components. Its heritability has been estimated to range from 25 to 56% based on twin and family studies (22–28), and a recent meta-analysis study of heritability, summarizing different study designs, such as twin, adoption and family studies, showed that on average 39% of neuroticism variations can be attributed to genetic variability (29). High genetic correlations between neuroticism and psychiatric disorders were reported, such as 0.43–0.6 for major depression (17,30,31) and about 0.8 for anxiety (17,32), although some overlaps are expected given similarities in questionnaire item criteria between neuroticism and these disorders. Taken together, potential causations and/or comorbid conditions between neuroticism and disorders imply complex and multiple underlying genetic mechanisms (7,21,33,34).

The most studied candidate gene for neuroticism is *SLC6A4*, which encodes the serotonin transporter (5-HTT) (35–38) that is responsible for removing serotonin from the synaptic cleft between two neurons. The short variant of the polymorphism (5-HTTLPR) is not only significantly associated with neuroticism (35–38), but also enhances amygdala activation in response to negative stimuli (39–41). A similar relationship has been reported between 5-HTT, amygdala activity and major depression (42–45). However, 5-HTT was not genome-wide significant in a recent genome-wide association studies (GWAS) of neuroticism (46).

Recently, large-scale GWAS have identified several genes associated with neuroticism, such as *MAGI1* (47), *GRIK3*, *KLHL2*, *CRHR1*, *MAPT*, *CELF4* (48) and *L3MBTL2* (49). Notably, two independent studies (48,49) both found a highly significant association with neuroticism on chromosome 8p23.1 spanning a 4-Mb region of long-range linkage disequilibrium (LD) due to an inversion polymorphism (46,48). Of these significant genes, *MAGI1*, *GRIK3* and *CRHR1* were linked to psychiatric disorders (50–56). Recent studies (49,57) estimated that single nucleotide polymorphism (SNP) heritability (i.e. total additive contribution from all SNPs (58)) is around 12–15% for neuroticism based on additive genetic effects, indicating that a large fraction of variants with small or moderate effect sizes of neuroticism are not identifiable by current GWAS analysis due to low statistical power. Therefore, the current study aimed to improve the statistical power to detect common genetic variants associated with neuroticism by employing prior information of SNPs.

SNPs in GWAS are not exchangeable, for example, SNPs in or near genes are shown to explain more variation of a trait than SNPs between genes (59). SNPs in some annotation categories are enriched for genetic effects and are more likely to be associated with a given trait (60). In our previous study of Covariate-Modulated Mixture Model (CM3) (61), a relative enrichment score (RES) was constructed for each SNP using the prior information including LD-weighted genic annotation scores (60), total LD score (the sum of pairwise LD r^2) and heterozygosity ($2f(1-f)$, where f is the SNP minor allele frequency, assuming the Hardy-Weinberg equilibrium holds) based on GWAS summary statistics. Using stratification by RES, we re-ranked and classified SNPs into RES ordinal strata and then plotted a quantile-quantile (Q-Q) curve for each RES stratum. We estimated the conditional false discovery rate (FDR) for each SNP on the basis of the stratum-specific Q-Q curve (62,63). Applying this to a GWAS of neuroticism ($N = 59\,225$), significant SNPs (genes) were identified at FDR < 0.05. We also performed a replication analysis in an independent sample ($N = 170\,911$) from a large-scale GWAS (46) including samples of Genetics of Personality Consortium (GPC) (47) and UK Biobank. Compared with traditional GWAS using significance thresholds of P-value (5×10^{-8}) or unconditional FDR (0.05) without any prior information, our framework utilizing SNP prior information (i.e. genic annotation scores, total LD score and heterozygosity) increases the statistical power thus facilitating gene discovery.

Results

Stratified Q-Q plot and enrichment

The stratified Q-Q plot shows different enrichment levels across RES strata, which deviates further away from the null line as RES increases (dotted curves in Fig. 1). The earlier or greater departure from the null line (leftward shift) suggests a larger proportion of true associations for a given nominal P-value. As a result, SNPs with higher RES (i.e. in stratum 4) are more likely to be associated with neuroticism than those with lower RES.

In addition to the model-free Q-Q plot generated by empirical distributions described above, we applied a model-based method to fit the Q-Q curve in each stratum for conditional FDR calculation. For each stratum, the fitted Q-Q plot (solid curves in Fig. 1) is generated from the cumulative distribution function of the corresponding Weibull-chi-square mixture probability distribution using the stratum-specific parameters (see Materials and Methods).

Estimation of conditional FDR from the lookup table

We used a heat map plot to construct the lookup table to visualize variations of conditional FDR across nominal P-values within RES strata shown in Figure 2. We also show variations of unconditional FDR ($-\log_{10}(FDR)$) with lighter colors as nominal P-values decrease (i.e. $-\log_{10}(P)$ increase) in Supplementary Material, Figure S1. In the Figure 2, a gradual decrease of conditional FDR with gradient colors from bottom-left to top-right corners suggests enrichment improved by RES strata as shown a gradual increase of $-\log_{10}(FDR)$. For the FDR threshold of 0.05 (i.e. $\sim$1.3 for $-\log_{10}(FDR)$), the corresponding nominal P-values

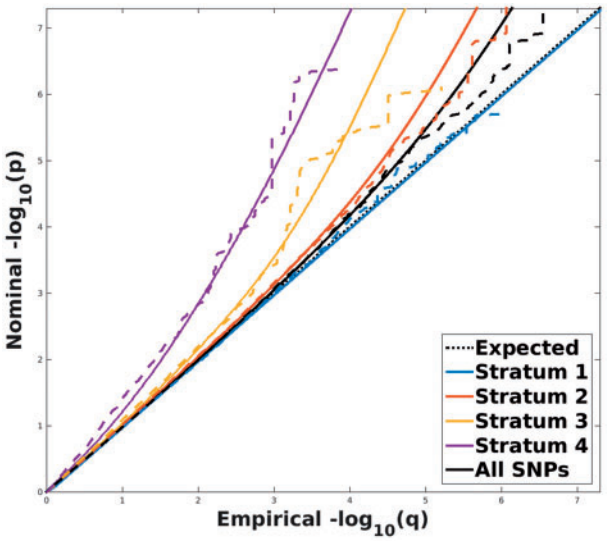

Figure 1. Enrichments in stratified Q-Q plot for neuroticism. The stratified Q-Q plot with differential enrichments across strata (dotted curves) and its predicted lines (solid curves) fitted by using Weibull-chi-square mixture distributions. We re-ranked the SNPs based on their relative enrichment scores (RES) constructed from the SNP prior information, and categorized the SNPs into four strata (e.g. SNPs with higher RES are in the stratum 4).

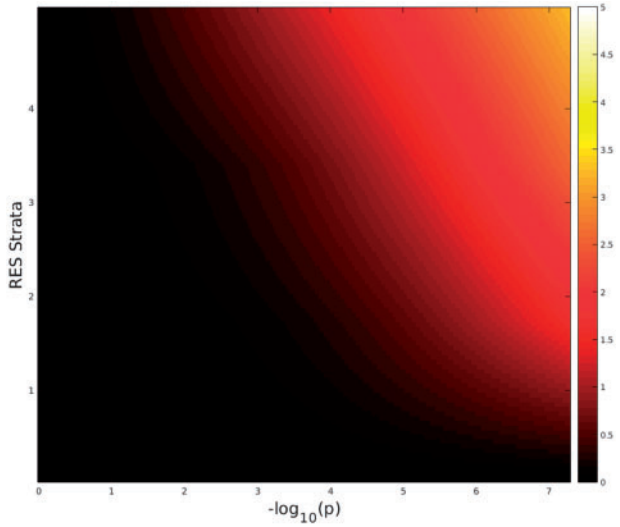

Figure 2. Conditional FDR lookup table. The lookup table for conditional FDR reflecting RES strata against nominal *P*-values illustrated by gradient colors with a color bar showing variations of FDR. Both *P*-value and FDR are scaled as the negative logarithm of base 10.

reduce to around 10^{-4} and 10^{-6} for RES strata 4 and 3, whereas they are about 10^{-7} for RES stratum 2 (Fig. 2) and around 10^{-8} for unconditional FDR (Supplementary Material, Fig. S1).

Significant loci identified by FDR

In our study, SNPs associated with neuroticism were identified by thresholds of *P*-value (5×10^{-8}), unconditional (0.05) and conditional FDR (0.05). To ensure that significant loci are independent, we removed correlated SNPs with LD $r^2 > 0.2$ and retained the SNP with the lowest *P*-value or FDR in each LD block. The retained SNPs are referred to as LD-independent SNP. Given a GWAS threshold of *P*-value $< 5 \times 10^{-8}$ or unconditional FDR < 0.05, only one LD-independent SNP (rs12102100, $P = 6.81 \times 10^{-10}$) located on chromosome 15 was detected (Table 1). Given the threshold of conditional FDR < 0.05, we identified five LD-independent SNPs in four loci located on different chromosomes (3, 9, 15 and 22) (Table 1). Among these five SNPs, one SNP was detected by GWAS *P*-value and unconditional FDR, but it did not reach significance level of *P*-value < 0.05 and had the opposite direction of the effect size (β) in the replication sample. The other four SNPs detected by conditional FDR had the same direction of effect in the replication sample and three of them are significant (*P*-value < 0.05). These four SNPs did not reach GWAS *P*-value threshold (5×10^{-8}) in discovery sample but were uncovered by conditional FDR using SNP prior information. Notably, two SNPs (rs10812851 and rs9611505) are significant (*P*-value $< 5 \times 10^{-8}$) in our combined analysis of discovery and replication samples at GWAS significance level (Table 1). As a result, conditional FDR using SNP prior information to generate RES detected more signals than *P*-value alone and/or unconditional FDR.

We next examined if our results were sensitive to the placement of cut-off points to generate RES strata. For these five LD-independent SNPs, their conditional FDR were robust in the scenarios of different cut-off points of RES for generating strata (Supplementary Material, Table S1) and different thresholds of dichotomized *P*-values in the logistic regression models to calculate RES (Supplementary Material, Table S2). Although the SNP, rs17022974, was borderline significant in some cases, four significant loci were constantly found in all analyses.

Manhattan plot

To visualize SNPs associated with neuroticism, we constructed a Manhattan plot. Four independent loci were identified by conditional FDR (<0.05) (Table 1) where gene symbols for those loci were also shown (Fig. 3). Of these four loci, only one locus was detected by unconditional FDR (<0.05) or GWAS *P*-value ($<5 \times 10^{-8}$). In summary, conditional FDR incorporating SNP prior information, such as annotation categories, total LD score and heterozygosity, is a more powerful method, compared with unconditional FDR and *P*-value, for gene discovery of neuroticism in GWAS.

Stratified Q-Q plots based on individual SNP prior information

To assess enrichment improved by genic annotation, total LD score and heterozygosity, we stratified SNPs including those on chromosome 8p by each of these characteristics separately and generated the stratified Q-Q plots. This analysis provided supporting evidence to use SNP prior information to construct RES in our main stratified FDR analysis. For neuroticism, eight genic annotation categories are more enriched than intergenic annotation and all SNPs as shown in the stratified Q-Q plot of the nominal *P*-values (Supplementary Material, Fig. S2A) using LD-weighted genic annotation scores (60). Likewise, similar enrichment patterns are also found in different levels of total LD score and heterozygosity (Supplementary Material, Fig. S2B-C). The results implied that SNP prior information including annotation categories and total LD score in conjunction with heterozygosity can be leveraged to increase power for gene discovery in GWAS of neuroticism.

Table 1. LD-independent SNPs significantly associated with neuroticism at GWAS thresholds of *P*-value ($<5 \times 10^{-8}$) or FDR (<0.05)

SNP	Chr	Closest gene (region)	A1/A2	Frq	Discovery (23andMe), N = 59 225				Replication, N = 170 911		Combined discovery and replication	
					β (se)	*P*-value	FDR[a]	cFDR[b]	β (se)	*P*-value	*P*-value	*N*
P-value[c] < 5 × 10^{-8}												
rs12102100	15	*MEIS2* (intergenic)	A/G	0.42	0.264 (0.043)	6.81×10^{-10}	0.009	0.002	-0.004 (0.004)	0.325	0.0226	230 117
Unconditional FDR < 0.05												
rs12102100	15	*MEIS2* (intergenic)	A/G	0.42	0.264 (0.043)	6.81×10^{-10}	0.009	0.002	-0.004 (0.004)	0.325	0.0226	230 117
Conditional FDR < 0.05												
rs9822731	3	*CADM2* (intron)	T/C	0.78	0.246 (0.050)	7.51×10^{-7}	0.207	0.011	0.012 (0.004)	0.006	1.01×10^{-6}	229 877
rs17022974	3	*CADM2* (intron)	T/C	0.64	-0.197 (0.044)	7.92×10^{-6}	0.431	0.048	-0.006 (0.004)	0.105	2.49×10^{-4}	230 117
rs10812851	9	*LINGO2* (intron)	T/C	0.63	0.223 (0.044)	3.60×10^{-7}	0.157	0.034	0.013 (0.004)	2.86×10^{-4}	1.15×10^{-8}	230 117
rs12102100	15	*MEIS2* (intergenic)	A/G	0.42	0.264 (0.043)	6.81×10^{-10}	0.009	0.002	-0.004 (0.004)	0.325	0.0226	230 117
rs9611505	22	*EP300* (intron)	T/C	0.69	-0.236 (0.046)	3.80×10^{-7}	0.160	0.002	-0.015 (0.004)	5.20×10^{-5}	1.33×10^{-9}	230 117

The following abbreviations are used: Chr, chromosome; A1, effect allele; A2, non-effect allele; Frq, allele frequency of A1.

[a]Unconditional FDR.

[b]Conditional FDR.

[c]GWAS threshold of *P*-value. GWAS hits in chromosome 8p are not shown as they were excluded from our FDR analysis.

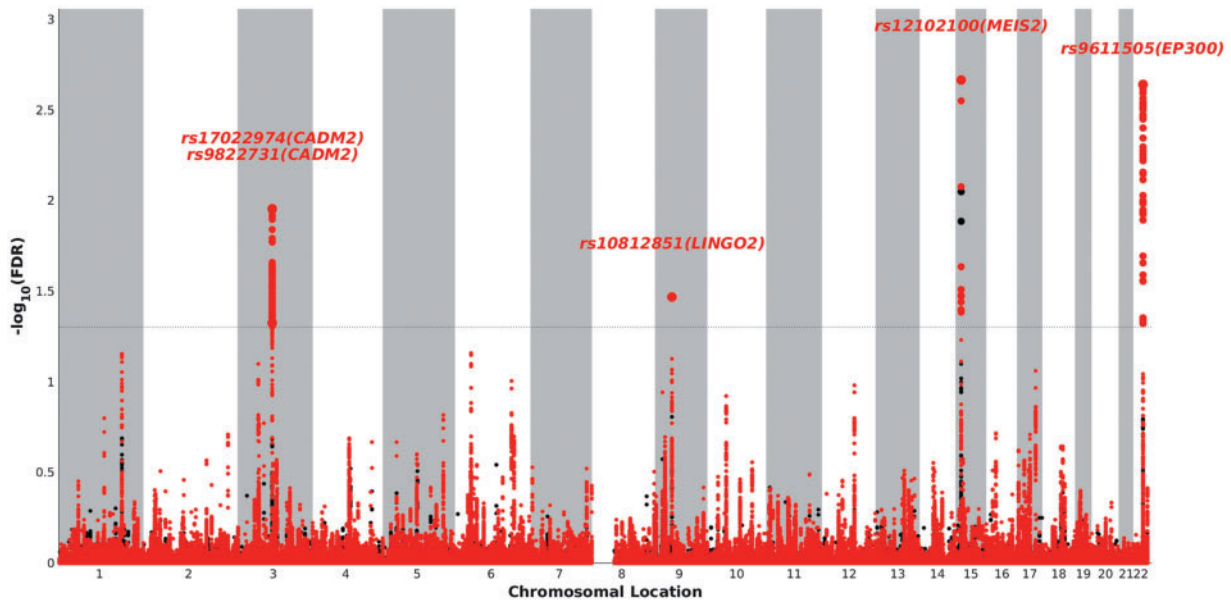

Figure 3. Manhattan plot for neuroticism. The Manhattan plot shows locations of four significant LD-independent loci identified by conditional FDR beyond the given threshold (dotted line, FDR = 0.05 and $-\log_{10}(FDR) \approx 1.3$). The large and small points represent significant (FDR < 0.05) and non-significant SNPs, respectively. Two colors, black and red, denote signals from unconditional and conditional FDR, respectively. The SNP (rs12102100) on chromosome 15 is also detected by GWAS *P*-value ($<5 \times 10^{-8}$) and unconditional FDR (< 0.05), whereas the other three loci containing four SNPs are only detected by conditional FDR.

No enrichment from true null based on permutation tests in an independent dataset

We performed permutation tests in an independent dataset to examine if our method was prone to generate false positives. The ranges and means of the genomic inflation factors (λ) of 10 000-round permutation tests across strata are shown in Supplementary Material, Table S3. The magnitude by which λ departs from 1 is a metric of deviation of null association and reflects over-abundance of low *P*-values compared with the expected null line in a Q-Q plot. Compared with the λ's from the neuroticism GWAS using the discovery sample, the averaged λ in each stratum computed from GWAS of 10 000-round

permutations was smaller, demonstrating that our approach did not increase additional false positive signals.

Discussion

Traditional GWAS have been limited by a stringent significance threshold resulting in insufficient statistical power to detect genetic variants with moderate effects, even in GWAS with large samples. For neuroticism, GWAS have identified several genetic variants but the results are inconsistent, except for the chromosomal region of 8p23.1. Here, we proposed a framework integrating a prior-informed approach (61) with conditional FDR (62,63) for controlling multiple comparisons to increase power in a GWAS of neuroticism. We prioritized SNPs based on their prior information, including genic annotation categories, total LD score and heterozygosity to construct a relative enrichment score (61) for each SNP. All SNPs were stratified into ordinal strata by their RES. Stratum with higher RES level showed greater enrichment. Conditional FDR for each SNP was calculated according to its RES stratum and predicted FDR. Compared with standard GWAS analysis using *P*-value ($<5\times10^{-8}$) or unconditional FDR threshold (<0.05) in which only one LD-independent SNP was identified in the discovery sample ($N=59\,225$), four additional SNPs were identified by conditional FDR and three of them were replicated in an independent GWAS ($N=170\,911$) (46).

RES was an auxiliary measure derived from CM3 to enable a more accurate estimation of replication probabilities in a recent methodological study (61), which used a resampling-based approach and required results of meta-analysis sub-studies of GWAS. In our study, we constructed RES for stratification but omitted the procedures of cross-validation by resampling sub-studies so that our current method can be applied to datasets without sub-samples. In contrast to RES of CM3, we used GWAS summary statistics of height (64) to generate a polygenic-trait RES for each SNP to avoid data overfitting if RES was calculated using the same data for estimating FDR. Furthermore, conditional FDR for detection of pleiotropic effects between two traits was proposed by Andreassen *et al.* (62,63) in which FDR was conditional on GWAS *P*-values of the second trait and here FDR was conditional on RES. To refine estimation of conditional FDR, we used non-overlapping strata stratified by RES and a more stringent pruning threshold ($r^2>0.2$). Here, we combined the features of these studies (61–63), i.e. incorporating RES for each SNP and generating conditional FDR for each SNP given its RES stratum, to increase statistical power for gene discovery. Besides statistical power, false positive findings of RES were also verified in our study (Supplementary Material, Table S3) and it suggested that introduction of RES for stratification will not incur additional false positive signal.

We identified five significant SNPs within four genomic loci by conditional FDR, including *MEIS2* (Meis homeobox 2), *CADM2* (cell adhesion molecule 2), *LINGO2* (leucine rich repeat and Ig domain containing 2) and *EP300* (E1A binding protein p300). Apart from *MEIS2*, the other three loci were novel and only detected by conditional FDR incorporating the SNP prior information. The significant SNP, rs12102100, is located in the downstream intergenic region ~200 kb of *MEIS2*, a gene that has been described to mediate metabolic side effects to antipsychotic drugs (65) and associated with hyperactive-impulsive symptom (66) in a GWAS. Two significant SNPs on chromosome 3 are located in intronic region of *CADM2* associated with persistence of temperament traits (67) and cognitive functions (68) in GWAS and proposed as a candidate gene for autism (69). Different dimensions of temperament have been shown to correlate with FFM of personality such as a positive correlation has been seen between persistence and conscientiousness (70,71) and, our previous study (49) has shown that neuroticism has a negative genetic correlation with conscientiousness. These findings suggested that *CADM2* may have pleiotropic effects on personality traits.

The other two significant SNPs located in intronic regions of *LINGO2* and *EP300*, respectively, were replicated in the independent sample (46) and also significant in our combined analysis (Table 1). *LINGO2* has been found to be expressed in neuronal tissues (72) and linked to neurodegenerative disorders, i.e. essential tremor and Parkinson's disease (73,74). *EP300* regulates transcription as histone acetyltransferase and is involved in the processes of cell proliferation and differentiation (75). Polymorphisms in *EP300* were shown to correlate with schizophrenia in a large-scale GWAS (76). High levels of neuroticism are associated with schizophrenia (16) and recently another *EP300* polymorphism (rs11090039, in LD ($r^2=0.82$) with rs9611505) was identified as a shared risk locus between schizophrenia and neuroticism with concordant directionality of effect in the phenotypes (77). Interestingly, multi-center GWAS have also identified associations of *LINGO2* and *CADM2* with body mass index (BMI) and obesity (78–80), which in turn have been positively correlated with neuroticism (81,82). These genetic and phenotypic findings suggest that neuroticism may share the same behavioral or biological pathways with psychiatric disorders (such as schizophrenia) and BMI and/or play an intermediate role between genes and these traits.

Recently, two independent large GWAS (48,49) both found neuroticism to be associated with 8p23.1, which has been suggested as a potential hub for developmental neuropsychiatric and neurodegenerative disorders (83). In this chromosomal region (spanning about 4 Mb and containing at least 36 genes), abundant significant SNPs were identified in long-range LD block with high pairwise correlations (r^2) between SNPs. A known inversion polymorphism might lead to this long-range LD (84,85) and has been shown to be associated with neuroticism (46). Therefore, we excluded SNPs on chromosome 8p in the process of FDR calculation to prevent underestimation of FDR derived from abundant correlated SNPs with lower *P*-values even after we have performed pruning to remove correlated SNPs with $r^2>0.2$. This finding demonstrated a limitation of our FDR approach when genetic effects are located in high LD regions. Alternatively, a new strategy has been proposed to unveil the complex schizophrenia–MHC (major histocompatibility complex) association using structural variations in terms of haplotypes (86), which might be a good paradigm for studying genetic effects on chromosome 8p.

Given the evidence of phenotypic as well as genetic associations between neuroticism and a wide range of mental and physical diseases, the discovery of new genes linked to neuroticism is important for understanding the genetic mechanisms underlying these diseases. By incorporating SNP prior information into a conditional FDR framework, we increased power for gene discovery for neuroticism. Our study demonstrates that this statistical framework is a promising tool for improving power of gene discovery using existing GWAS summary statistics.

Materials and Methods

Discovery sample

The GWAS summary statistics of neuroticism were obtained from a subset of 23andMe, Inc. research participants ($N=59\,225$)

who showed > 97% European ancestry (details in Supplementary Material). All research participants completed a web-based implementation of the Big Five Inventory (BFI) (87,88), with 44 questions. A score for neuroticism was computed using eight of these items (87). The procedures of genotyping and imputation were described in Supplementary Material. A total of 13 341 935 autosomal SNPs were retained after quality control.

Association tests were performed by regressing neuroticism scores on imputed dosages of SNPs in the 23andMe cohort. Age, gender and the top five principal components (89) for population structure correction were included as covariates and *P*-values were computed using likelihood ratio tests. The original 13 341 935 SNPs were mapped to a high quality LD structure with 2 549 449 SNPs (60) (see Supplementary Material) in our subsequent analyses. Recent GWAS (46,48,49) showed multiple significant SNPs for neuroticism on chromosomal region of 8p23.1 due to the long-range LD (84). Because high correlations between SNPs in the long-range LD may bias FDR estimates (90), we therefore removed SNPs on the 8p region (0–45 Mb) to avoid underestimation of FDR. The remaining 2 484 994 SNPs were then included in the FDR estimation.

The GWAS results of the meta-analysis of 23andMe and Genetics of Personality Consortium data (47) have been published (49). Besides traditional GWAS, in our current study, we only used the 23andMe cohort as the discovery sample in which the analysis framework including RES stratification and conditional FDR was applied to test associations between SNPs and neuroticism.

Replication sample

The GWAS summary statistics of the replication sample were obtained from the Social Science Genetic Association Consortium (SSGAC), which have been published (46) and are available in the public domain (http://www.thessgac.org/#!data/kuzq8). In this study, the sample-size-based meta-analysis of two large GWAS, UK Biobank and Genetics of Personality Consortium data, were performed ($N = 170\,911$) (46). We replicated our findings in this large-scale study and set the significance level as P-value < 0.05 for replication.

LD-weighted genic annotation

The total of 2 484 994 SNPs analyzed in our study were annotated with a LD-weighted genic annotation score (60). The score was calculated based on the European reference sample provided by the November 2012 release of the Phase I 1000 Genomes Project (1KGP). Specially, each SNP in the 1KGP reference panel was initially assigned to a single mutually exclusive genic annotation category based on its genomic position (UCSC Genome Browser on Human hg19 assembly). Eight genic annotation categories were included: exon, intron, 5′ untranslated region (5′UTR), 3′ untranslated region (3′UTR), 1 and 10 kilo-base pairs upstream of the gene transcription start positions, and 1 and 10 kilo-base pairs downstream of gene transcription end positions. Pairwise LD correlation coefficients (r^2) between SNPs were calculated based on 1KGP. For each SNP, a continuous, non-exclusive LD-weighted category score was assigned as the LD weighted sum of the positional category scores for SNPs tagged in each of the eight categories mentioned above. By incorporating LD information, the annotation of individual SNP reflects the weighted annotation in the context of underlying linkage blocks. For detailed information on SNP annotation, score construction and quality control see Schork *et al.* (60).

Relative enrichment score (RES)

The relative enrichment score (RES) (61) was constructed for each SNP using logistic regression model incorporated with information of annotation categories, total LD score and heterozygosity. In the logistic regression model, dichotomized *P*-value (if *P*-value 10^{-3}, then $y = 1$; otherwise, $y = 0$) for each SNP was regressed on LD-weighted genic annotation scores for eight categories and total LD score (sum of all pair-wise LD r^2) multiplied by heterozygosity (H, where $H = 2f(1-f)$ and f is the SNP minor allele frequency from the 1KGP European reference panel), that is, $\text{logit}(\Pr(y=1)) = \mathbf{X\beta}$, where $\boldsymbol{X}$ is the matrix of regressors including annotation scores and total LD score multiplied by heterozygosity, and $\boldsymbol{\beta}$ is the vector of regression coefficients. For each SNP, RES was defined as the estimated value of the response obtained from the above logistic regression, which was a variant of RES calculation by Wang *et al.* (61). Specifically, for SNP i, the RES is the scalar product $X_i\hat{\boldsymbol{\beta}}$ of the corresponding vector of regressors X_i and the vector of estimated regression coefficients $\hat{\boldsymbol{\beta}}$ calculated based on the logistic regression model. Here, we did not include second trait information for pleiotropy and combined total LD score and heterozygosity into a single variable, which was used in the logistic regression model. Our approach is in contrast to the approach of Wang *et al.* (61), who considered these two variables separately (more details in Supplementary Material). In particular, to avoid overfitting if RES is calculated using identical data of interest, it is preferable to choose an independent set of neuroticism GWAS results if available, or GWAS results from a highly polygenic complex trait. Here, we used GWAS summary statistics of height (from Genetic Investigation of ANthropometric Traits consortium, http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files) (64) to generate a polygenic-trait RES for each SNP, given that height is highly polygenic and its genetic components are relevant to multiple biological mechanisms. In addition, we evaluated different dichotomized *P*-values (such 10^{-2} and 10^{-4}) to ensure robustness of our results (Supplementary Material, Table S2).

Stratified Q-Q plots and enrichment

Stratified Q-Q plots are constructed by grouping SNPs on the basis of levels of an auxiliary measure, i.e. annotation categories, total LD score, heterozygosity and RES, and plotting the Q-Q curve separately for each level. If enrichment is captured by stratification of the auxiliary measure, this is expressed as successive leftward deflections in a stratified Q-Q plot as levels of the auxiliary measure increase (61). For neuroticism, we re-ranked and classified SNPs into four strata determined by percentiles of RES. The intervals for strata are not equally spaced because a large proportion of SNPs are null and only a small proportion of SNPs has effect on the trait for GWAS. In our study, we empirically selected 30, 93 and 99 percentiles of RES as cut-off points to stratify SNPs into four strata. The cut-off selection was generally dependent on the distribution of effect sizes and the LD structure of SNPs in each stratum. We then plotted a Q-Q curve (dotted curves in Fig. 1) for each stratum and examined whether there were differential enrichments, i.e. whether the degree of deflection from the expected null line was dependent on the RES stratum. Specifically, the SNPs with higher RES showed a greater degree of deflection from the expected null line. Furthermore, we used different cut-off points of RES to examine whether they substantially influenced detection of significant loci. We evaluated eight combinations of

four-stratum cut-off points and showed the results in Supplementary Material, Table S1.

Parametric model

The shape of the empirical distributions depicted in the Q-Q plots resembles the shape of the distribution function of a mixture of Weibull and chi-square distributions. So, for each RES stratum we modeled the Q-Q curve with a function that is proportional to the distribution function of a Weibull-chi-square mixture to compute stratum-specific predicted FDR. We assumed different scale parameters for the two component distributions. Further, an exploratory analysis showed that a value of 0.5 is a reliable choice for the shape parameter of the Weibull component. Keeping the shape parameter fixed at 0.5, the unknown parameters of the mixture were estimated by maximizing a cost function using unconstrained nonlinear optimization, where the cost function was proportional to the logarithm of the likelihood function of the parameters given the observed SNP distribution. The solid curve for each RES stratum in Figure 1 shows the predicted Q-Q curve from the cumulative mixture distribution using stratum-specific estimated parameters. Therefore, predicted FDR can be calculated from predicted Q-Q curves estimated by using Weibull-chi-square mixture distributions based on the multiple-bin empirical quantile.

Lookup table

We used lookup tables to interpolate unconditional and conditional FDR for each SNP from predicted FDR. The unconditional FDR lookup table (Supplementary Material, Fig. S1) shows FDR variations (changes by colors) against nominal *P*-values, and the conditional FDR lookup table (Fig. 2) also shows FDR varying by RES strata and nominal *P*-values (see Supplementary Material).

Manhattan plot

To visualize the localization of the genomic loci associated with neuroticism, we constructed a Manhattan plot by plotting all SNPs within an LD block in relation to their chromosomal location. In each LD block, conditional FDR values for SNPs were ranked in ascending order and SNPs that have LD $r^2 > 0.2$ with higher ranking were then removed. Thus, we can retain the most significant SNP associated with neuroticism in each LD block, i.e. LD-independent SNP. As illustrated in Figure 3, the large and small points represent significant (FDR < 0.05) and non-significant SNPs, respectively. Two colors, black and red, denote signals from unconditional and conditional FDR, respectively. LD-independent loci with conditional FDR < 0.05 are also shown by their gene symbols in the plot.

Combined analysis of discovery and replication samples

The combined analysis of discovery and replication samples was performed based on the sample-size based method using METAL (91). We showed the meta-analysis *P*-values and sample sizes of significant SNPs in Table 1.

Permutation tests

To verify validity of RES, we performed permutation tests in an independent dataset to examine whether false positive signals might be detected using RES for gene discovery. We shuffled the disease status in a case-control study including 492 schizophrenia cases and 458 controls from the Denmark cohort of Psychiatric Genomics Consortium (PGC) (76). Theoretically, each randomly permuted case-control dataset satisfies the null hypothesis of no association between the disease and SNPs. For each permutation, we computed *P*-values of association tests for whole-genome SNPs to construct a four-strata Q-Q plot stratified by RES, and then, the genomic inflation factor (λ) in each stratum was calculated. The magnitude by which λ deviates from 1 is a metric of deviation of observed *P*-values from the expected null line in a Q-Q plot. We considered 10 000 permutations, i.e. 10 000 randomly permuted datasets, and for each of them evaluated λ. Note that individual genotype data are required for permutation tests but we only have summary statistics for neuroticism, so we have used the Denmark cohort of PGC to complete this analysis.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

We thank the research participants and employees of 23andMe for making this work possible.

Conflicts of Interest statement. D.A.H. and J.Y.T. are employees at and own stock or stock options in 23andMe, Inc. The remaining authors declare no competing financial interests.

Funding

National Institute of Mental Health (R01MH100351 to M.-T.L., N.S., C.-H.C.); NARSAD Young Investigator award (C.-H.C); South-East Norway Regional Health Authority (2016-064) (O.B.S.); Research Council of Norway through a FRIPRO Mobility Grant, contract no. 251134 (Y.W.).

References

1. Digman, J.M. (1990) Personality structure – emergence of the 5-factor model. *Annu. Rev. Psychol.*, **41**, 417–440.
2. Goldberg, L.R. (1993) The structure of phenotypic personality traits. *Am. Psychol.*, **48**, 26–34.
3. Pervin, L.A. and John, O.P. (1999) *Handbook of Personality: Theory and Research*. Guilford Press, New York.
4. Costa, P.T., Jr. and McCrae, R.R. (1987) Neuroticism, somatic complaints, and disease: is the bark worse than the bite? *J. Pers.*, **55**, 299–316.
5. McCrae, R.R., Costa, P.T., Jr. and Martin, T.A. (2005) The NEO-PI-3: a more readable revised NEO Personality Inventory. *J. Pers. Assess.*, **84**, 261–270.
6. Khan, A.A., Jacobson, K.C., Gardner, C.O., Prescott, C.A. and Kendler, K.S. (2005) Personality and comorbidity of common psychiatric disorders. *Br. J. Psychiatry*, **186**, 190–196.
7. Klein, D.N., Kotov, R. and Bufferd, S.J. (2011) Personality and depression: explanatory models and review of the evidence. *Annu. Rev. Clin. Psychol.*, **7**, 269–295.
8. Kotov, R., Gamez, W., Schmidt, F. and Watson, D. (2010) Linking "Big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. *Psychol. Bull.*, **136**, 768–821.

9. Malouff, J.M., Thorsteinsson, E.B. and Schutte, N.S. (2005) The relationship between the five-factor model of personality and symptoms of clinical disorders: a meta-analysis. *J. Psychopathol. Behav. Assess.*, **27**, 101–114.
10. Jokela, M., Pulkki-Raback, L., Elovainio, M. and Kivimaki, M. (2014) Personality traits as risk factors for stroke and coronary heart disease mortality: pooled analysis of three cohort studies. *J. Behav. Med.*, **37**, 881–889.
11. Loerbroks, A., Apfelbacher, C.J., Thayer, J.F., Debling, D. and Sturmer, T. (2009) Neuroticism, extraversion, stressful life events and asthma: a cohort study of middle-aged adults. *Allergy*, **64**, 1444–1450.
12. Shipley, B.A., Weiss, A., Der, G., Taylor, M.D. and Deary, I.J. (2007) Neuroticism, extraversion, and mortality in the UK Health and Lifestyle Survey: a 21-year prospective cohort study. *Psychosom. Med.*, **69**, 923–931.
13. Christensen, M.V. and Kessing, L.V. (2006) Do personality traits predict first onset in depressive and bipolar disorder? *Nord. J. Psychiatry*, **60**, 79–88.
14. Fanous, A.H., Neale, M.C., Aggen, S.H. and Kendler, K.S. (2007) A longitudinal study of personality and major depression in a population-based sample of male twins. *Psychol. Med.*, **37**, 1163–1172.
15. Kendler, K.S., Kuhn, J. and Prescott, C.A. (2004) The interrelationship of neuroticism, sex, and stressful life events in the prediction of episodes of major depression. *Am. J. Psychiatry*, **161**, 631–636.
16. Van Os, J. and Jones, P.B. (2001) Neuroticism as a risk factor for schizophrenia. *Psychol. Med.*, **31**, 1129–1134.
17. Hettema, J.M., Neale, M.C., Myers, J.M., Prescott, C.A. and Kendler, K.S. (2006) A population-based twin study of the relationship between neuroticism and internalizing disorders. *Am. J. Psychiatry*, **163**, 857–864.
18. Weinstock, L.M. and Whisman, M.A. (2006) Neuroticism as a common feature of the depressive and anxiety disorders: A test of the revised integrative hierarchical model in a national sample. *J. Abnorm. Psychol.*, **115**, 68–74.
19. Johansson, L., Guo, X., Duberstein, P.R., Hallstrom, T., Waern, M., Ostling, S. and Skoog, I. (2014) Midlife personality and risk of Alzheimer disease and distress: a 38-year follow-up. *Neurology*, **83**, 1538–1544.
20. Terracciano, A., Sutin, A.R., An, Y., O'Brien, R.J., Ferrucci, L., Zonderman, A.B. and Resnick, S.M. (2014) Personality and risk of Alzheimer's disease: new data and meta-analysis. *Alzheimers Dement.*, **10**, 179–186.
21. Lahey, B.B. (2009) Public health significance of neuroticism. *Am. Psychol.*, **64**, 241–256.
22. Birley, A.J., Gillespie, N.A., Heath, A.C., Sullivan, P.F., Boomsma, D.I. and Martin, N.G. (2006) Heritability and nineteen-year stability of long and short EPQ-R neuroticism scales. *Pers. Individ. Dif.*, **40**, 737–747.
23. Distel, M.A., Trull, T.J., Willemsen, G., Vink, J.M., Derom, C.A., Lynskey, M., Martin, N.G. and Boomsma, D.I. (2009) The five-factor model of personality and borderline personality disorder: a genetic analysis of comorbidity. *Biol. Psychiatry*, **66**, 1131–1138.
24. Keller, M.C., Coventry, W.L., Heath, A.C. and Martin, N.G. (2005) Widespread evidence for non-additive genetic variation in Cloninger's and Eysenck's personality dimensions using a twin plus sibling design. *Behav. Genet.*, **35**, 707–721.
25. Pilia, G., Chen, W.M., Scuteri, A., Orru, M., Albai, G., Dei, M., Lai, S., Usala, G., Lai, M., Loi, P. *et al.* (2006) Heritability of cardiovascular and personality traits in 6, 148 Sardinians. *PLoS Genet.*, **2**, e132.
26. van den Berg, S.M., de Moor, M.H., McGue, M., Pettersson, E., Terracciano, A., Verweij, K.J., Amin, N., Derringer, J., Esko, T., van Grootheest, G. *et al.* (2014) Harmonization of Neuroticism and Extraversion phenotypes across inventories and cohorts in the Genetics of Personality Consortium: an application of Item Response Theory. *Behav. Genet.*, **44**, 295–313.
27. Vernon, P.A., Martin, R.A., Schermer, J.A. and Mackie, A. (2008) A behavioral genetic investigation of humor styles and their correlations with the Big-5 personality dimensions. *Pers. Individ. Dif.*, **44**, 1116–1125.
28. Wray, N.R., Birley, A.J., Sullivan, P.F., Visscher, P.M. and Martin, N.G. (2007) Genetic and phenotypic stability of measures of neuroticism over 22 years. *Twin Res. Hum. Genet.*, **10**, 695–702.
29. Vukasovic, T. and Bratko, D. (2015) Heritability of personality: A meta-analysis of behavior genetic studies. *Psychol. Bull.*, **141**, 769–785.
30. Fanous, A., Gardner, C.O., Prescott, C.A., Cancro, R. and Kendler, K.S. (2002) Neuroticism, major depression and gender: a population-based twin study. *Psychol. Med.*, **32**, 719–728.
31. Kendler, K.S. and Myers, J. (2010) The genetic and environmental relationship between major depression and the five-factor model of personality. *Psychol. Med.*, **40**, 801–806.
32. Hettema, J.M., Prescott, C.A. and Kendler, K.S. (2004) Genetic and environmental sources of covariation between generalized anxiety disorder and neuroticism. *Am. J. Psychiatry*, **161**, 1581–1587.
33. Canli, T. (2008) Toward a neurogenetic theory of neuroticism. *Ann. N. Y. Acad. Sci.*, **1129**, 153–174.
34. Ormel, J., Jeronimus, B.F., Kotov, R., Riese, H., Bos, E.H., Hankin, B., Rosmalen, J.G. and Oldehinkel, A.J. (2013) Neuroticism and common mental disorders: meaning and utility of a complex relationship. *Clin. Psychol. Rev.*, **33**, 686–697.
35. Lesch, K.P., Bengel, D., Heils, A., Sabol, S.Z., Greenberg, B.D., Petri, S., Benjamin, J., Muller, C.R., Hamer, D.H. and Murphy, D.L. (1996) Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science*, **274**, 1527–1531.
36. Munafo, M.R., Freimer, N.B., Ng, W., Ophoff, R., Veijola, J., Miettunen, J., Jarvelin, M.R., Taanila, A. and Flint, J. (2009) 5-HTTLPR genotype and anxiety-related personality traits: a meta-analysis and new data. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **150B**, 271–281.
37. Schinka, J.A., Busch, R.M. and Robichaux-Keene, N. (2004) A meta-analysis of the association between the serotonin transporter gene polymorphism (5-HTTLPR) and trait anxiety. *Mol. Psychiatry*, **9**, 197–202.
38. Sen, S., Burmeister, M. and Ghosh, D. (2004) Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **127b**, 85–89.
39. Drabant, E.M., Ramel, W., Edge, M.D., Hyde, L.W., Kuo, J.R., Goldin, P.R., Hariri, A.R. and Gross, J.J. (2012) Neural mechanisms underlying 5-HTTLPR-related sensitivity to acute stress. *Am. J. Psychiatry*, **169**, 397–405.
40. Hariri, A.R., Mattay, V.S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., Egan, M.F. and Weinberger, D.R. (2002) Serotonin transporter genetic variation and the response of the human amygdala. *Science*, **297**, 400–403.
41. Munafo, M.R., Brown, S.M. and Hariri, A.R. (2008) Serotonin transporter (5-HTTLPR) genotype and amygdala activation: a meta-analysis. *Biol. Psychiatry*, **63**, 852–857.

42. Gillihan, S.J., Rao, H.Y., Brennan, L., Wang, D.J.J., Detre, J.A., Sankoorikal, G.M.V., Brodkin, E.S. and Farah, M.J. (2011) Serotonin transporter genotype modulates the association between depressive symptoms and amygdala activity among psychiatrically healthy adults. *Psychiatry Res. Neuroimag.*, **193**, 161–167.
43. Gryglewski, G., Lanzenberger, R., Kranz, G.S. and Cumming, P. (2014) Meta-analysis of molecular imaging of serotonin transporters in major depression. *J. Cereb. Blood Flow Metab.*, **34**, 1096–1103.
44. Hamilton, J.P., Siemer, M. and Gotlib, I.H. (2008) Amygdala volume in major depressive disorder: a meta-analysis of magnetic resonance imaging studies. *Mol. Psychiatry*, **13**, 993–1000.
45. Pezawas, L., Meyer-Lindenberg, A., Drabant, E.M., Verchinski, B.A., Munoz, K.E., Kolachana, B.S., Egan, M.F., Mattay, V.S., Hariri, A.R. and Weinberger, D.R. (2005) 5-HTTLPR polymorphism impacts human cingulate-amygdala interactions: a genetic susceptibility mechanism for depression. *Nat. Neurosci.*, **8**, 828–834.
46. Okbay, A., Baselmans, B.M., De Neve, J.E., Turley, P., Nivard, M.G., Fontana, M.A., Meddens, S.F., Linner, R.K., Rietveld, C.A., Derringer, J. *et al.* (2016) Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.*, **48**, 624–633.
47. Genetics of Personality Consortium, de Moor, M.H.M., van den Berg, S.M., Verweij, K.J.H., Krueger, R.F., Luciano, M., Arias Vasquez, A., Matteson, L.K., Derringer, J., Esko, T., Amin, N. *et al.* (2015) Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with major depressive disorder. *JAMA Psychiatry*, **72**, 642–650.
48. Smith, D.J., Escott-Price, V., Davies, G., Bailey, M.E., Colodro-Conde, L., Ward, J., Vedernikov, A., Marioni, R., Cullen, B., Lyall, D. *et al.* (2016) Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Mol. Psychiatry*, **21**, 749–757.
49. Lo, M.T., Hinds, D.A., Tung, J.Y., Franz, C., Fan, C.C., Wang, Y., Smeland, O.B., Schork, A., Holland, D., Kauppi, K. *et al.* (2017) Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat. Genet.*, **49**, 152–156.
50. Etain, B., Mathieu, F., Rietschel, M., Maier, W., Albus, M., McKeon, P., Roche, S., Kealey, C., Blackwood, D., Muir, W. *et al.* (2006) Genome-wide scan for genes involved in bipolar affective disorder in 70 European families ascertained through a bipolar type I early-onset proband: supportive evidence for linkage at 3p14. *Mol. Psychiatry*, **11**, 685–694.
51. Ferentinos, P., Rivera, M., Ising, M., Spain, S.L., Cohen-Woods, S., Butler, A.W., Craddock, N., Owen, M.J., Korszun, A., Jones, L. *et al.* (2014) Investigating the genetic variation underlying episodicity in major depressive disorder: suggestive evidence for a bipolar contribution. *J. Affect. Disord.*, **155**, 81–89.
52. Gray, A.L., Hyde, T.M., Deep-Soboslay, A., Kleinman, J.E. and Sodhi, M.S. (2015) Sex differences in glutamate receptor gene expression in major depression and suicide. *Mol. Psychiatry*, **20**, 1057–1068.
53. Karlsson, R., Graae, L., Lekman, M., Wang, D., Favis, R., Axelsson, T., Galter, D., Belin, A.C. and Paddock, S. (2012) MAGI1 copy number variation in bipolar affective disorder and schizophrenia. *Biol. Psychiatry*, **71**, 922–930.
54. Lee, P.H., Perlis, R.H., Jung, J.Y., Byrne, E.M., Rueckert, E., Siburian, R., Haddad, S., Mayerfeld, C.E., Heath, A.C., Pergadia, M.L. *et al.* (2012) Multi-locus genome-wide association analysis supports the role of glutamatergic synaptic transmission in the etiology of major depressive disorder. *Transl. Psychiatry*, **2**, e184.
55. Schiffer, H.H. and Heinemann, S.F. (2007) Association of the human kainate receptor GluR7 gene (GRIK3) with recurrent major depressive disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **144B**, 20–26.
56. Weber, H., Richter, J., Straube, B., Lueken, U., Domschke, K., Schartner, C., Klauke, B., Baumann, C., Pane-Farre, C., Jacob, C.P. *et al.* (2016) Allelic variation in CRHR1 predisposes to panic disorder: evidence for biased fear processing. *Mol. Psychiatry*, **21**, 813–822.
57. Power, R.A. and Pluess, M. (2015) Heritability estimates of the Big Five personality traits based on common genetic variants. *Transl. Psychiatry*, **5**, e604.
58. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W. *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.*, **42**, 565–569.
59. Yang, J., Manolio, T.A., Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M.G. *et al.* (2011) Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.*, **43**, 519. U544.
60. Schork, A.J., Thompson, W.K., Pham, P., Torkamani, A., Roddey, J.C., Sullivan, P.F., Kelsoe, J.R., O'Donovan, M.C., Furberg, H., Schork, N.J., Tobacco. *et al.* (2013) All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet.*, **9**, e1003449.
61. Wang, Y., Thompson, W.K., Schork, A.J., Holland, D., Chen, C.H., Bettella, F., Desikan, R.S., Li, W., Witoelar, A., Zuber, V. *et al.* (2016) Leveraging Genomic Annotations and Pleiotropic Enrichment for Improved Replication Rates in Schizophrenia GWAS. *PLoS Genet.*, **12**, e1005803.
62. Andreassen, O.A., Thompson, W.K., Schork, A.J., Ripke, S., Mattingsdal, M., Kelsoe, J.R., Kendler, K.S., O'Donovan, M.C., Rujescu, D., Werge, T. *et al.* (2013) Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genet.*, **9**, e1003455.
63. Andreassen, O.A., Djurovic, S., Thompson, W.K., Schork, A.J., Kendler, K. S., O'Donovan, M.C., Rujescu, D., Werge, T., van de Bunt, M., Morris, A.P. *et al.* (2013) Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *Am. J. Hum. Genet.*, **92**, 197–209.
64. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z. *et al.* (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.*, **46**, 1173–1186.
65. Adkins, D.E., Aberg, K., McClay, J.L., Bukszar, J., Zhao, Z., Jia, P., Stroup, T.S., Perkins, D., McEvoy, J.P., Lieberman, J.A. *et al.* (2011) Genomewide pharmacogenomic study of metabolic side effects to antipsychotic drugs. *Mol. Psychiatry*, **16**, 321–332.
66. Lasky-Su, J., Neale, B.M., Franke, B., Anney, R.J., Zhou, K., Maller, J.B., Vasquez, A.A., Chen, W., Asherson, P., Buitelaar, J. *et al.* (2008) Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate. *Gene*

Associations. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **147B**, 1345–1354.
67. Service, S.K., Verweij, K.J., Lahti, J., Congdon, E., Ekelund, J., Hintsanen, M., Raikkonen, K., Lehtimaki, T., Kahonen, M., Widen, E. *et al.* (2012) A genome-wide meta-analysis of association studies of Cloninger's Temperament Scales. *Transl. Psychiatry*, **2**, e116.
68. Ibrahim-Verbaas, C.A., Bressler, J., Debette, S., Schuur, M., Smith, A.V., Bis, J.C., Davies, G., Trompet, S., Smith, J.A., Wolf, C. *et al.* (2016) GWAS for executive function and processing speed suggests involvement of the CADM2 gene. *Mol. Psychiatry*, **21**, 189–197.
69. Casey, J.P., Magalhaes, T., Conroy, J.M., Regan, R., Shah, N., Anney, R., Shields, D.C., Abrahams, B.S., Almeida, J., Bacchelli, E. *et al.* (2012) A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Hum. Genet.*, **131**, 565–579.
70. MacDonald, D.A. and Holland, D. (2002) Examination of relations between the NEO Personality Inventory-Revised and the Temperament and Character Inventory. *Psychol. Rep.*, **91**, 921–930.
71. De Fruyt, F., Van de Wiele, L. and Van Heeringen, C. (2000) Cloninger's psychobiological model of temperament and character and the five-factor model of personality. *Pers. Individ. Dif.*, **29**, 441–452.
72. Homma, S., Shimada, T., Hikake, T. and Yaginuma, H. (2009) Expression pattern of LRR and Ig domain-containing protein (LRRIG protein) in the early mouse embryo. *Gene Expr. Patterns*, **9**, 1–26.
73. Vilarino-Guell, C., Wider, C., Ross, O.A., Jasinska-Myga, B., Kachergus, J., Cobb, S.A., Soto-Ortolaza, A.I., Behrouz, B., Heckman, M.G., Diehl, N.N. *et al.* (2010) LINGO1 and LINGO2 variants are associated with essential tremor and Parkinson disease. *Neurogenetics*, **11**, 401–408.
74. Wu, Y.W., Prakash, K.M., Rong, T.Y., Li, H.H., Xiao, Q., Tan, L.C., Au, W.L., Ding, J.Q., Chen, S.D. and Tan, E.K. (2011) Lingo2 variants associated with essential tremor and Parkinson's disease. *Hum. Genet.*, **129**, 611–615.
75. Gayther, S.A., Batley, S.J., Linger, L., Bannister, A., Thorpe, K., Chin, S.F., Daigo, Y., Russell, P., Wilson, A., Sowter, H.M. *et al.* (2000) Mutations truncating the EP300 acetylase in human cancers. *Nat. Genet.*, **24**, 300–303.
76. Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**, 421–427.
77. Smeland, O.B., Wang, Y., Lo, M.T., Li, W., Frei, O., Witoelar, A., Tesli, M., Hinds, D.A., Tung, J.Y., Djurovic, S. *et al.* (2017) Identification of genetic loci shared between schizophrenia and the Big Five personality traits. *Sci. Rep.*, **7**, 2222.
78. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Lango Allen, H., Lindgren, C.M., Luan, J., Magi, R. *et al.* (2010) Association analyses of 249, 796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.*, **42**, 937–948.
79. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J. *et al.* (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature*, **518**, 197–206.
80. Berndt, S.I., Gustafsson, S., Magi, R., Ganna, A., Wheeler, E., Feitosa, M.F., Justice, A.E., Monda, K.L., Croteau-Chonka, D.C., Day, F.R. *et al.* (2013) Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat. Genet.*, **45**, 501–512.
81. Armon, G., Melamed, S., Shirom, A., Shapira, I. and Berliner, S. (2013) Personality Traits and Body Weight Measures: Concurrent and Across-Time Associations. *Eur. J. Pers.*, **27**, 398–408.
82. Sutin, A.R. and Terracciano, A. (2016) Personality traits and body mass index: Modifiers and mechanisms. *Psychol. Health*, **31**, 259–275.
83. Tabares-Seisdedos, R. and Rubenstein, J.L.R. (2009) Chromosome 8p as a potential hub for developmental neuropsychiatric disorders: implications for schizophrenia, autism and cancer. *Mol. Psychiatry*, **14**, 563–589.
84. Price, A.L., Weale, M.E., Patterson, N., Myers, S.R., Need, A.C., Shianna, K.V., Ge, D., Rotter, J.I., Torres, E., Taylor, K.D. *et al.* (2008) Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.*, **83**, 132–135; author reply 135. 139.
85. Tian, C., Plenge, R.M., Ransom, M., Lee, A., Villoslada, P., Selmi, C., Klareskog, L., Pulver, A.E., Qi, L., Gregersen, P.K. *et al.* (2008) Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet.*, **4**, e4.
86. Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V. *et al.* (2016) Schizophrenia risk from complex variation of complement component 4. *Nature*, **530**, 177–183.
87. John, O.P., Donahue, E.M. and Kentle, R. L. (1991) *The Big Five Inventory–Versions 4a and 54. Berkeley, CA: University of California, Berkeley*, Institute of Personality and Social Research.
88. Soto, C.J. and John, O.P. (2009) Ten facet scales for the Big Five Inventory: Convergence with NEO PI-R facets, self-peer agreement, and discriminant validity. *J. Res. Pers.*, **43**, 84–90.
89. Lehoucq, R.B., Sorensen, D.C., Yang, C. and Mathematics, SfIaA. (1998), In *Software, environments, tools 6*. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), Philadelphia, Pa., pp. 1 electronic text (xv, 142 p).
90. Schwartzman, A. and Lin, X. (2011) The effect of correlation in false discovery rate estimation. *Biometrika*, **98**, 199–214.
91. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.