# Accurate, Large-Scale Genotyping of 5HTTLPR and Flanking Single Nucleotide Polymorphisms in an Association Study of Depression, Anxiety, and Personality Measures

Naomi R. Wray, Michael R. James, Scott D. Gordon, Troy Dumenil, Leanne Ryan, William L. Coventry, Dixie J. Statham, Michele L. Pergadia, Pamela A.F. Madden, Andrew C. Heath, Grant W. Montgomery, and Nicholas G. Martin

**Background:** The length polymorphism repeat in the promoter region of the serotonin transporter gene (5HTTLPR) is one of the most studied polymorphisms for association with a range of psychiatric and personality phenotypes. However, the original 5HTTLPR assay is prone to bias toward short allele calling.

**Methods:** We designed new assays for the 5HTTLPR suitable for large-scale genotyping projects and we genotyped 13 single nucleotide polymorphisms (SNPs) in a 38-kilobase region around the 5HTTLPR, including SNP rs25531, a polymorphism of the 5HTTLPR long allele. Association analysis was conducted for major depression and/or anxiety disorder in unrelated cases ($n = 1161$) and control subjects ($n = 1051$) identified through psychiatric interviews administered to a large population sample of Australian twin families. Participants had been scored for personality traits several years earlier ($n \geq 2643$ unrelated individuals).

**Results:** We identified a two-SNP haplotype proxy for 5HTTLPR; the CA haplotype of SNPs rs4251417 and rs2020934 is coupled with the short allele of 5HTTLPR ($r^2 = .72$). We found evidence for association ($p = .0062$, after accounting for multiple testing) for SLC6A4 SNPs rs6354 and rs2020936 (positioned in a different linkage disequilibrium [LD] block about 15.5 kb from 5HTTLPR) with anxiety and/or depression and neuroticism, with the strongest association for recurrent depression with onset in young adulthood (odds ratio = 1.55, 95% confidence interval = 1.16–2.06).

**Conclusions:** The associated SNPs are in the same LD block as the variable number of tandem repeats serotonin transporter intron 2 marker, for which association has previously been reported.

Serotonergic neurotransmission impacts on a wide range of behaviors, including cognition and emotion (1,2), and drugs targeting serotonin reuptake are clinically effective antidepressants (3). As a result, one of the most studied polymorphisms for association with a broad range of psychiatric and personality phenotypes is the length polymorphism repeat (LPR) in the promoter region of the serotonin transporter gene (5HTT renamed SLC6A4) (5HTTLPR). The 5HTTLPR polymorphism comprises a 43-base pair (bp) (4–8) insertion or deletion (long, "L," with 16 repeat units or short, "S," with 14 repeat units, alleles, respectively). The S allele (frequency in Caucasians ~.45 [9]) reduces transcriptional efficiency, resulting in decreased SLC6A4

expression and function (10). Association studies and subsequent meta-analyses (Table 1) (11–15) have shown conflicting results in support of an association between the S allele and anxiety, depression, and the personality trait neuroticism (a measure of emotional stability that is genetically correlated to both anxiety and depression [16–18]). Conflicting results have dogged candidate gene association studies for many complex disorders, attributable to small sample sizes of both primary and replication studies, heterogenous subject populations, association dependent on environmental conditions such as stressful life events (19), and differing instruments for assessment of phenotypic traits and statistical methods (e.g., [20,21]). However, an additional problem specific to 5HTTLPR relates to the genotyping assay, which has caused considerable bias toward S allele identification (22,23). Furthermore, association may have been compromised by the presence of an A/G single nucleotide polymorphism (SNP), rs25531, that lies within the L allele of 5HTTLPR (5,8); the L allele with the rarer G allele of rs25531 (denoted $L_G$) is functionally equivalent to the S allele because of changes to the activating protein 2 (AP2) transcription factor binding site altered by this SNP (5,8).

The aim of this study was to investigate the association between the 5HTTLPR polymorphism, including the rs25531 polymorphism, and psychiatric and personality phenotypes in a large cohort. As part of our study, we designed new assays for the 5HTTLPR suitable for large-scale genotyping projects and we genotyped tagging SNPs in a 38-kilobase (kb) region around the 5HTTLPR (Figure 1A) to determine whether any SNP or combination of SNPs could be used as proxies for the difficult and

From the Genetic and Molecular Epidemiology Laboratories (NRW, MRJ, SDG, TD, LR, WLC, DJS, GWM, NGM), Queensland Institute of Medical Research, Brisbane, Queensland; Faculty of Arts & Social Sciences (DJS), University of the Sunshine Coast, Maroochydore, Queensland; and School of Behavioural, Cognitive and Social Sciences (WLC), University of New England, Armidale, New South Wales, Australia; and Department of Psychiatry (MLP, PAFM, ACH), Washington University School of Medicine, St. Louis, Missouri.

Address correspondence to Naomi R. Wray, Ph.D., Queensland Institute of Medical Research, 300 Herston Road, Brisbane 4006, Australia; E-mail: naomi.wray@qimr.edu.au.

**Table 1.** Summary of Meta-Analyses of Association Studies of 5HTTLPR in Major Depression, Anxiety Disorders, and Neuroticism

| Phenotype | Comparison | Association (95% CI) | Number of Studies | Description | Reference |
|---|---|---|---|---|---|
| Major Depression | S vs. L | OR = 1.11 (1.04–1.19) | 24 | 3752 cases | (11) |
|  | SS vs. LL | OR = 1.39 (1.21–1.61) |  | 5707 control subjects |  |
| Neuroticism | SS or SL vs. LL | d = −.22 (−.44 to −.01) | 12 | 3247 individuals | (12) |
| Neuroticism | S vs. L | ns: p = .81 | 1[a] | 768 individuals selected from the 5% tails of the neuroticism distribution from a population sample of 88,142 | (13) |
|  |  | ns: p = .47 |  | 4175 selected for extreme neuroticism from population sample of 20,921 |  |
| Panic Disorder | S vs. L | OR = .91 (.80–1.03) | 10 | 1025 cases | (14) |
|  |  |  |  | 1568 control subjects |  |
| OCD | SS vs. LL | OR = 1.21 (1.10, 1.45) | 13 | 1242 cases | (15) |
|  | SL vs. LL | OR = .79 (.64, .98) |  | 2203 control subjects |  |
|  | S vs. L | OR = 1.01 (.88, 1.03) |  |  |  |
| Harm Avoidance | SS or SL vs. LL | d = −.05 (−.66 to .56) | 13 | 2521 individuals | (12) |

5HTTLPR, length polymorphism repeat (LPR) in the promoter region of the serotonin transporter gene; CI, confidence interval; d, standardized mean difference; L, long allele; ns, not significant; OCD, obsessive-compulsive disorder; OR, odds ratio; S, short allele.

[a]Large single study published subsequent to the meta-analysis of neuroticism.

time-consuming 5HTTLPR assay. Our study design allows us to examine association in multiple traits within the same cohort and consistency across independent subcohorts. Within our study sample, we can identify subsets of cases that are predicted to be genetically more homogenous. For example, the relative risk (RR) to first-degree relatives is reported to be higher for recurrent depression (RR ≈ 4–5) (reviewed in [24]) compared with major depression (RR ≈ 2–3). Similarly, although estimates for RR are not available, anxiety comorbid with depression is also considered a genetically more homogeneous group (reviewed in [25]). The optimum balance of sample size versus sample homogeneity cannot be predicted since the true genetic etiology is unknown. However, the availability of a depth of phenotypic information allows us to investigate associations through case subsets.

## Methods and Materials

### Samples

All participants were adult twins and their families recruited through the Australian Twin Registry. They provided informed consent under study protocols approved by the Queensland Institute of Medical Research and Washington University Human Research Ethics Committees.

During the period 1988 to 1990, study participants from two twin birth cohorts (born 1890–1964 and 1964–1971, respectively) were mailed an extensive Health and Lifestyle Questionnaire (1989 Questionnaire Survey [26,27]). This included the shortened, revised 48-item Eysenck Personality Questionnaire (EPQ) (28) and a short-form, 54-item version of the Temperament and Character Inventory (26,29), the Temperament Personality Questionnaire (TPQ). The EPQ measures four dimensions of personality, including neuroticism and extraversion, and the TPQ measures three dimensions of temperament, including harm avoidance. Between 1992 and 1994, twins from the older cohort (n = 5995) (30) were interviewed by telephone using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) instrument. The SSAGA is a comprehensive psychiatric interview that was designed to assess psychiatric disorders in adults (31) according to DSM-III-R but subsequently updated to DSM-IV criteria (32) and modified for use as a telephone survey instrument in Australia (SSAGA-OZ). Of these participants, n = 4597 have subsequently provided a blood sample (or rarely a buccal or saliva sample) for genotyping (33). Over the period

1996 to 1969, sibling pairs participating in the 1989 Questionnaire Survey that were either concordant or discordant for extreme EPQ neuroticism scores (one sibling in the top or bottom decile, the other sibling in the top or bottom quintile, and allowing inclusion of multiple siblings) were recruited (35–37). Participants (n = 2456) completed the shortened Composite International Diagnostic Interview (CIDI) (34), which provides DSM-IV (35) lifetime diagnoses of depression (including recurrent major depression) and anxiety disorders. This extreme discordant and concordant (36) design is a cost-efficient strategy for obtaining an informative dataset for genetic studies (37). Of these participants, 2213 provided DNA samples, plus 837 parents. Full details of the recruitment procedure for the study, including response rates and incidence of DSM-IV diagnoses for anxiety- and depression-related disorders, are given elsewhere (38–40). Finally, between 2001 and 2005, some participants from the earlier studies were re-interviewed using an adaptation of the SSAGA (see [41,42] for details). Participants reported ancestry of all four grandparents, and a small number with two or more grandparents with known non-European ancestry were excluded from the analysis. Selection of individuals used in the association analyses are described in the Association Analysis section below.

### Genotyping

Genomic DNA was extracted from blood (or buccal) samples using standard protocols (43). Samples were plated in 384-well plates in two study sample sets comprising 1) monozygotic (MZ) and dizygotic (DZ) twins from the 1992 to 2000 SSAGA interview studies and 2) participants and their parents from the CIDI interview. Some 764 DNA samples are included in both sets, providing an opportunity for extensive quality control checking.

### 5HTTLPR Assay

The original assay (44) for the 5HTTLPR used polymerase chain reaction (PCR) primers in the nonrepetitive sequences that flanked the 16 repeat elements, which are each comprised of between 19 bp and 23 bp (10,44) (Figure 1B). This assay proved less than ideal: the PCR is difficult because of the very high guanine-cytosine (GC) content and the long length of the PCR products. In pilot studies, we found that these difficulties cause considerable bias toward S allele identification; the L allele signal is weak, so heterozygotes are frequently misscored as S homozygotes (e.g., see Figure 1D of [44]). This observation has also been

made by others (e.g., [22,23]). We improved on the assay by redesigning the PCR primers; full details are provided in Supplement 1. The final assay comprised a multiplexing scheme of PCR primers L3 + R3 and L3 + R4 located within the repeat elements (Figure 1B). This choice of primers provided two PCR products that are semi-independent (L3 is shared) and that could be analyzed together on a gel. The primer products were pooled (one DNA sample from each primer product) and run on a gel together (twofold multiplexing). Further, fourfold multiplexing was achieved by loading a second pair of samples onto the same gel (Figure 1C). In operation, a 384-well plate of DNAs can be amplified for both products and pooled in a way to allow duplicate scoring for each individual on two gels. Our pilot experiments showed that this was an efficient and robust system to genotype this difficult polymorphism.

**SNP and STin2 Genotyping.** The genotyping assay for SNP rs25531 (A/G) is described in Supplement 1. DNA samples were

genotyped for 13 additional SNPs (Figure 1) in the 38-kb region surrounding the 5HTTLPR. Our motivation was first, to determine whether there is an SNP-based proxy for 5HTTLPR that would be easier to genotype in large sample sets and second, to see if any SNPs in this region show stronger association with any of the phenotypic measures than 5HTTLPR. Single nucleotide polymorphisms were selected to represent the linkage disequilibrium landscape of the region bounded by rs140700 (intron 6) and rs7214991 (promoter), and preference was given to SNPs used in association studies of autism (45,46). All nongenotyped HapMap (build 35) (www.hapmap.org) (47) SNPs with minor allele frequency >.1 were represented by single SNP genotyped proxies with $r^2 > .8$. Three nonsynonymous SNPs were also included: rs28914832 (isoleucine/valine exon 10, also known as serotonin transporter [SERT] Ile425Val), rs6355 (glycine/alanine exon 3, also known as Gly56Ala), and rs1050565 (isoleucine/valine exon 12 of bleomycin hydrolase [BLMH] 13 kb from 5HTTLPR).

In pilot experiments, we attempted to genotype the 17-bp variable number tandem repeat (VNTR) of SLC6A4 (Figure 1A), known as serotonin transporter intron 2 (STin2). However, genotyping results did not pass our quality control checks (including Mendelian error checking) and so STin2 was not genotyped on the full study sample; problems with the assay have been found by others (e.g., [22]).

## Genotype Analysis

Our study design allowed us to undertake quality control checks often not possible in association studies (Supplement 1). As the rare $L_G$ alleles are functionally equivalent to S alleles (4,8), we constructed a 5HTTLPR + rs25531 (marker 12) variable with three levels used in the analysis, SS and $SL_G$, $SL_A$ and $L_GL_A$, and $L_AL_A$. After exclusion of identifiable errors, genotyping call rates ranged from 97.6% to 99.2% for the 13 SNPs excluding rs25531



**Figure 1.** SLC6A4 (5HTT) gene locus, polymorphisms, and assays. **(A)** Genomic locus map showing exons 1 to 15 (solid and open boxes; open boxes are exons transcribed only in some splicing variants), the length polymorphic region (5HTTLPR; striped box), the VNTR (STin2, striped box), and SNPs (rs#) with minor allele listed first. *Indicates a nonsynonymous exonic SNP. The exon/intron numbering is based on the most recent Ensembl update (SLC6A4-001; OTTHUMT00000256114, June 2007) (http://www.ensembl.org/Homo_sapiens/exonview?transcript=OTTHUMT00000256114; db = Vega), hence STin2 is positioned in intron 3. **(B)** Sequence of the 5HTTLPR with 16 copies of repeat unit plus flanking unique genomic sequence. The shaded and boxed sequence (of repeats VI–VIII) indicates the 43 bp deleted in the short (S) version. The dash in the sequence cc-ccc indicates the removal of one c residue according to recent resequencing data. Within the deleted sequence is an (a/g) SNP (rs25531). Underlined sequences show primers used to obtain various PCR products; primer names are shown to the right. The sequence PEX-1 (bold italics) is used in a special primer extension reaction to assay rs25531 as explained in the text and Supplement 1. **(C)** A 3.5% agarose gel with fourfold multiplexed samples: DNAs subjected to PCR with either L3/R3 or L3/R4 primers. DNA samples from each amplification were pooled in pairs and loaded on to 24 lanes of the gel (half is displayed) in two loadings separated in time, so that 96 DNAs were genotyped on one gel. The sizes of the Long and Short alleles are shown at the bottom (primers have 10nt 5'-extensions, not shown, which add 20 bp to product size); m1, m2 = marker lanes (50-bp ladder) for each loading. Representative examples of genotypes are LS (het, note the weak L band) Lane-1 Loading-2 both L3R3 and L3R4. LL: Lane-1 Loading-1 both L3R3 and L3R4; SS: Lane-3 Loading-1 L3R3; Lane-9 Loading-1 L3R3. 5HTT, serotonin transporter gene; 5HTTLPR, length polymorphism repeat (LPR) in the promoter region of the serotonin transporter gene; bp, base pair; PCR, polymerase chain reaction; S, short; SNP, single nucleotide polymorphism; STin2, serotonin transporter intron 2; VNTR, variable number of tandem repeats.

**Table 2.** Number of Genotyped Individuals Contributing to the Association Analyses

| Diagnosis (Indenting Denotes Subsetting) | Diagnoses Identified from the SSAGA Interview Administered Between 1992 and 2000 (or 2001–2005) and the CIDI Interview Administered Between 1996 and 1999 | | | Quantitative Personality Traits Measured in Questionnaires Administered Between 1988 and 1990 | |
|---|---|---|---|---|---|
| | SSAGA | CIDI | Combined | Trait | Unrelated Individuals |
| Control | 724 | 327 | 1051[a] | EPQ-Neuroticism | 2651 |
| Anxiety and/or Depression | 691 | 470 | 1161[a] | EPQ-Extraversion | 2643 |
| Anxiety with depression[b] | 81 | 184 | 265 | Harm Avoidance | 2645 |
| Depression | 587 | 366 | 953 | | |
| Recurrent early onset | 83 | 129 | 212 | | |
| Depression without anxiety[b] | 506 | 182 | 688 | | |
| Anxiety | 185 | 288 | 473 | | |
| Generalized anxiety disorder | | 129 | 129 | | |
| Social phobia | 58 | 105 | 163 | | |
| Panic/agoraphobia | 141 | 108 | 249 | | |
| Obsessive-compulsive disorder | | 79 | 79 | | |
| Anxiety without depression[b] | 104 | 104 | 208 | | |

CIDI, Composite International Diagnostic Interview; EPQ, Eysenck Personality Questionnaire; SSAGA, Semi-Structured Assessment for the Genetics of Alcoholism.

[a]A total dataset of 1051 control subjects and 1161 cases, all unrelated.

[b]Mutually exclusive case sets, so numbers sum to anxiety and/or depression cases. Note that these classifications are limited by the diagnoses that could be allocated by the questionnaires.

but were only 96.9% for 5HTTLPR and 95.4% for 5HTTLPR + rs25531; these latter rates are lower than normal for our laboratory but nonetheless much improved over the original assay for 5HTTLPR. Tests of deviation from Hardy-Weinberg equilibrium (HWE) and estimates of linkage disequilibrium were calculated in Haploview 4.0 (www.broad.mit.edu/mpg/haploview) (48) using the SSAGA sample (a population sample) with one individual randomly selected per family ($n = 2341$). Within Haploview, the Tagger (49) option was used to force exclude 5HTTLPR to see if any one, two, or three marker combination could be used to predict 5HTTLPR genotype. Similarly, we used Tagger to predict rs25531 genotype from other genotyped SNPs, those with LL 5HTTLPR genotypes.

**Association Analysis**

Sum scores of 12 item responses in each EPQ domain resulted in quantitative scores for neuroticism and extraversion. Harm avoidance comprised sum scores of 18 items from the TPQ. Scores for MZ twins were averaged. Sum scores were transformed using the averaged angular transformation (50). Scores used for analysis were standardized residuals (standardized separately for each sex) after regression of the transformed scores to remove effects of sex and age. This standardization was conducted using the population sample from the full study sample ($n > 20,000$), not just those genotyped (see [40]). Association analysis for quantitative traits used one individual per family, selecting the individual that deviated most from the population mean, and were conducted in PLINK (pngu.mgh.harvard.edu/~purcell/plink/) (51); to account for potential bias in the ascertainment of samples for analysis, permutation $p$ values are reported.

Cases were selected as unrelated individuals with a DSM-IV (35) lifetime diagnosis of major depression or an anxiety disorder. Early age of onset for depression was defined as those with self-reported age of onset of less than 30 years. Control subjects were selected as unrelated individuals from families in which no siblings who completed the questionnaires (including those not supplying a DNA sample) received a diagnosis of either major

depression or anxiety disorder from any of the surveys. Details of selection of individuals for analysis within families are provided in Supplement 1. Semi-Structured Assessment for the Genetics of Alcoholism control subjects were only selected if their neuroticism scores were less than the population average. Cases and control subjects were divided into two independent samples according to the interview instrument. Families who participated in both studies were allocated to the CIDI case-control sample, unless the only anxiety or depression diagnoses for the family were allocated from the SSAGA instrument (2.7% of those measured on both instruments). Thus, the CIDI and SSAGA case-control datasets (Table 2) are independent samples from the same population. An additional 133 cases were identified by the SSAGA re-interview. Association analysis was conducted for cases with major depression disorder (depression), any anxiety disorder (anxiety), or anxiety and/or depression (if they qualified for a diagnosis of any anxiety disorder and/or major depression). Analyses were also conducted for the case subset groups thought to represent genetically more homogeneous groups: anxiety with depression (if they qualified for both diagnoses) and recurrent, early age of onset depression (Table 2). Following up on literature reports discussed above, association analysis of all markers was also conducted for panic/agoraphobia and obsessive-compulsive disorder (OCD). However, for the other specific anxiety disorders, follow-up analyses were only conducted for SNPs showing association in the primary analyses. Logistic regression association analysis was conducted using PLINK (51) under a model that assumes additivity of allelic effects on the log risk scale. Power calculations and details of the permutation test used to determine the empiric significance of the pattern of association observed across datasets and phenotypes are provided in Supplement 1.

**Results**

The numbers of genotyped individuals for the two independent samples used in the association analysis are shown in Table 2. The ascertainment of the CIDI sample means that both the

**Table 3.** Description of the Diagnosis Classes in Terms of the Standardized Personality Measures, Age, and Gender in Independent SSAGA and CIDI Subsets

Personality Questionnaires Administered Between 1988 and 1989

| | EPQ-Neuroticism | | | EPQ-Extraversion | | | Harm Avoidance | | | Age | | Gender |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | *n* | Mean | SD | *n* | Mean | SD | *n* | Mean | SD | Age Mean | Age SD | % Female |
| SSAGA Interview Administered Between 1992 and 2000 (or 2001–2005) | | | | | | | | | | | | |
| Control | 724 | −.82 | .53 | 711 | .24 | .97 | 714 | −.55 | .85 | 41.9 | 12.4 | 64.7 |
| Anxiety and/or depression | 618 | .31 | 1.00 | 617 | .02 | 1.05 | 613 | .17 | 1.01 | 38.9 | 10.1 | 70.2 |
| Anxiety with depression | 73 | .83 | 1.08 | 73 | .02 | 1.11 | 73 | .54 | 1.11 | 39.0 | 8.4 | 78.1 |
| Anxiety | 164 | .63 | 1.13 | 164 | −.11 | 1.07 | 163 | .47 | 1.10 | 40.3 | 10.8 | 75.0 |
| Depression | 527 | .28 | .98 | 526 | .07 | 1.04 | 523 | .13 | .99 | 38.4 | 9.6 | 69.8 |
| CIDI Interview Administered Between 1996 and 1999 | | | | | | | | | | | | |
| Control | 321 | −1.45 | .44 | 318 | .23 | .94 | 321 | −.83 | .76 | 34.9 | 11.9 | 61.4 |
| Anxiety and/or depression | 449 | .82 | 1.34 | 440 | −.09 | 1.07 | 445 | .55 | 1.17 | 35.1 | 10.5 | 65.5 |
| Anxiety with depression | 178 | 1.12 | 1.13 | 174 | −.31 | 1.03 | 176 | .87 | 1.02 | 35.7 | 10.8 | 70.2 |
| Anxiety | 274 | 1.01 | 1.20 | 268 | −.22 | 1.05 | 261 | .76 | 1.06 | 35.7 | 10.9 | 64.6 |
| Depression | 353 | .82 | 1.35 | 346 | −.10 | 1.07 | 350 | .55 | 1.19 | 34.9 | 10.3 | 68.6 |

CIDI, Composite International Diagnostic Interview; EPQ, Eysenck Personality Questionnaire; SSAGA, Semi-Structured Assessment for the Genetics of Alcoholism.

cases and control subjects are likely to be more extreme that those of the SSAGA sample, and this is reflected in the summary statistics based on EPQ neuroticism and harm avoidance (Table 3). collected several years earlier.

### Linkage Disequilibrium Around 5HTTLPR

None of the markers showed evidence for Hardy-Weinberg disequilibrium. Figure 2 shows the linkage disequilibrium as both |D′| and $r^2$ measures; although the |D′| values are high, the corresponding $r^2$ values are lower, reflecting differences in allele frequencies (52). The best prediction of 5HTTLPR is provided by rs4251417 and rs2020934 (Figure 2B), where the CA haplotype is coupled with the S allele and the CG haplotype is coupled with the L allele of 5HTTLPR ($r^2$ = .72). No useful prediction could be made of 5HTTLPR-rs25531 or of rs25531 within LL individuals. Genotype data are available from the corresponding author on request.

### Association Analysis

Single nucleotide polymorphisms rs6354 and rs2020936 (markers 4 and 6) show association ($p < .05$) with all the depression and anxiety phenotypes presented in Figure 1, with

the T and A alleles of these SNPs, respectively, being more frequent in cases than control subjects [.82 versus .72, odds ratio (OR) = 1.19 (1.03–1.38)]. These SNPs are in high LD ($r^2$ = .98) and so represent a single association signal. Although the associations for the two independent datasets are only significant for some analyses, the direction of the association is the same, so the association in the combined dataset is significant. The association with depression is marginally significant, OR = 1.17 (1.01–1.26). However, the association with the two subsets of cases thought to represent genetically more homogenous groups (24,25) both show higher association odds ratios: recurrent, early-onset depression (Figures 3E and 4), OR = 1.55 (1.16–2.06), and comorbid group of anxiety with depression (Figures 3D and 4), OR = 1.31 (1.02–1.69). Permutation analysis showed that this pattern of association occurred only 62 times in 10,000 permutations across all markers under the null hypothesis, resulting in an empiric $p$ value that accounts for multiple testing of .0062. Odds ratios of association with rs6354 for the different anxiety disorders are presented in Figure 4. These results must be interpreted recognizing the constraints on possible diagnostic classes from the questionnaires. Necessarily, these results should
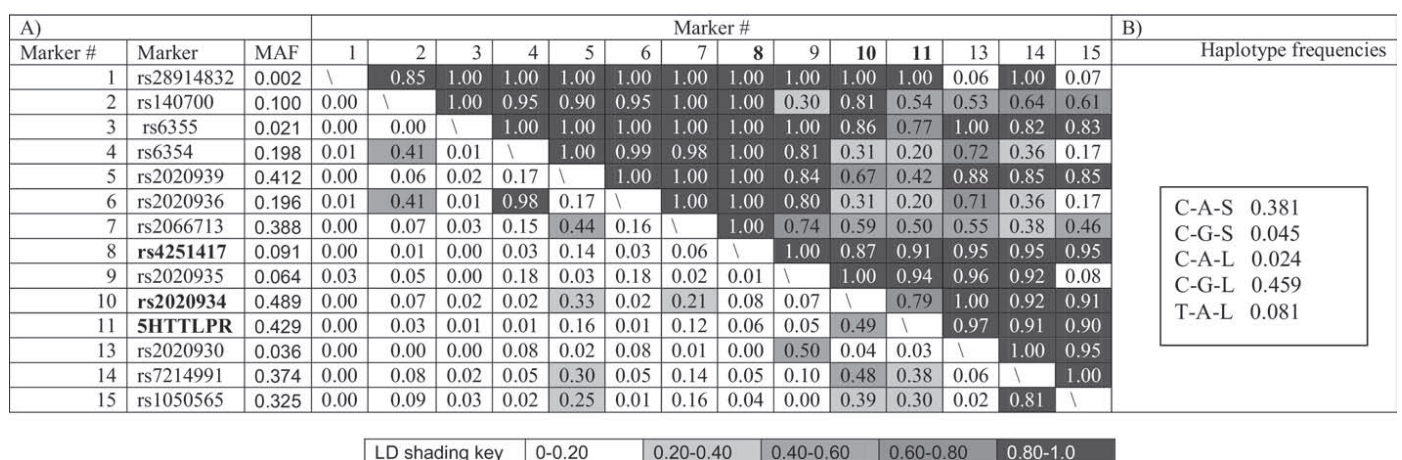
**Figure 2. (A)** Allele frequencies and linkage disequilibrium of genotyped markers, |D′| above the diagonal (\) and $r^2$ below the diagonal. **(B)** Haplotype frequencies of rs4251417-rs2020934-5HTTLPR. 5HTTLPR, length polymorphism repeat (LPR) in the promoter region of the serotonin transporter gene.
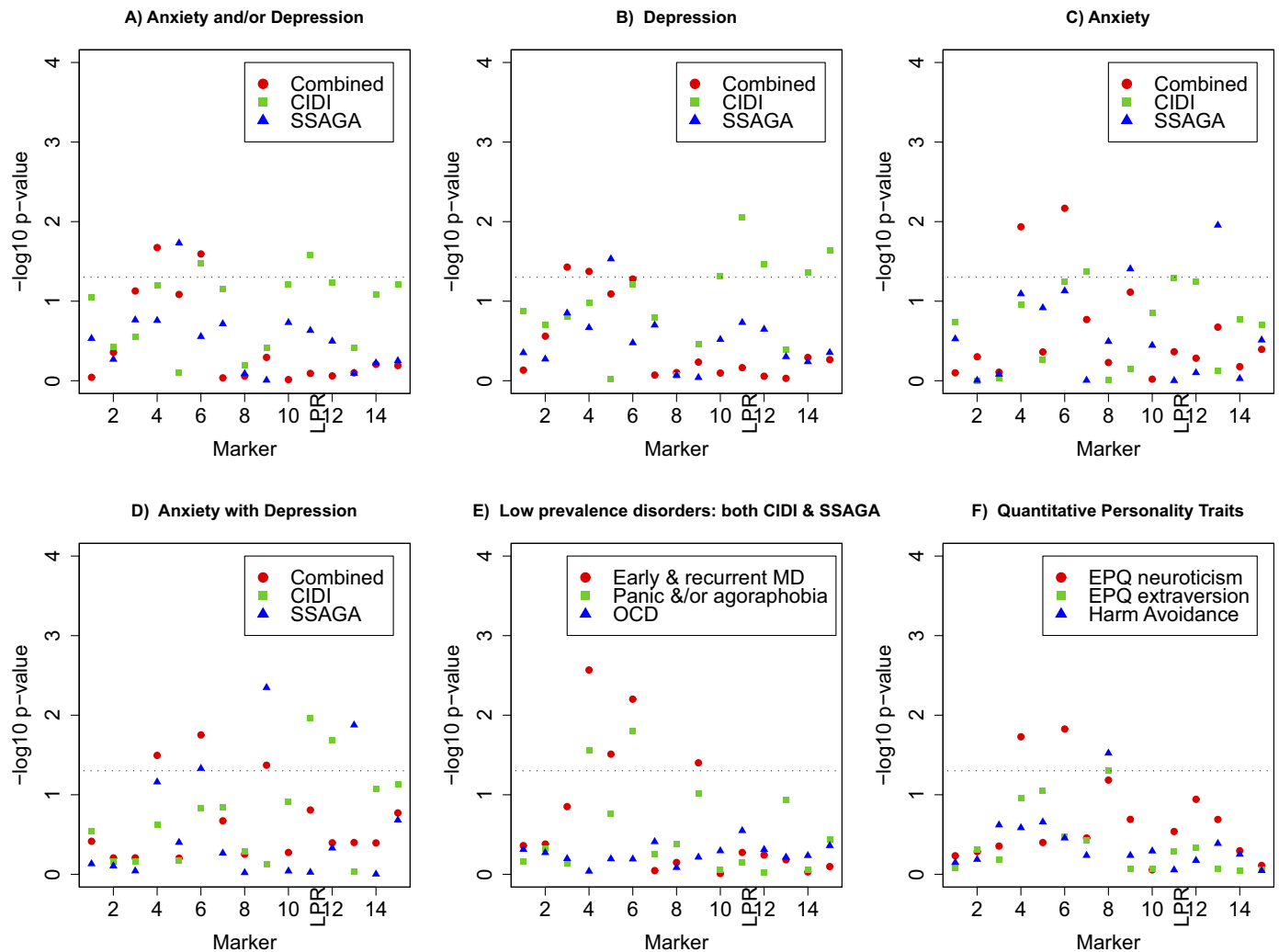
**Figure 3.** Association analysis results. Above the dotted horizontal line represents $p < .05$. Marker numbers as in Figure 2. For each SNP, when the circle representing the combined dataset analysis is at the top (highest $-\log_{10}$ [$p$ value]) of the three symbols, then both the SSAGA and CIDI samples show a trend in association in the same direction. In contrast, when the combined analysis result has the lowest $-\log_{10}$ ($p$ value) of the three analyses, the associations in the SSAGA and CIDI samples are with different alleles. CIDI, Composite International Diagnostic Interview; SNP, single nucleotide poymorphism; SSAGA, Semi-Structured Assessment for the Genetics of Alcoholism.

be considered as hypothesis generating for future studies. The associations for rs6354 and rs2020936 with neuroticism were .019 and .015, respectively (Figure 3F), but there is no evidence for association between either harm avoidance or extraversion, which had phenotypic correlations of .66 and −.25, respectively, with neuroticism. Association ($p < .05$) is seen for rs4251417 and harm avoidance. Haplotype analysis of combinations of SNPs revealed no associations more significant than those of the individual genotyped markers.

For the association analysis in the CIDI sample, we observed association ($p < .05$) for the 5HTTLPR (L allele), most significantly with the depression cases. We interpret this as a chance result, as the association is not replicated in the SSAGA sample for which any trend in association is with the S allele. A detailed investigation is provided in Supplement 1. We find no evidence for association of 5HTTLPR (with or without rs25531) with panic/agoraphobia, OCD, neuroticism, extraversion, or harm avoidance (Figure 3E and 3F). As a consequence, our results provide no evidence, either way, for the usefulness of genotyping of SNP rs25531 to subclassify the L allele. A full report of

allele and genotype frequencies and $p$ values is available online (Supplement 2).

## Discussion

Preferential amplification of one allele over another is recognized as a systematic problem in PCR assays for variable repeat polymorphisms (22). In pilot studies using the standard 5HTTLPR assay (44), we found overamplification of the S allele, as also reported by others (22,23). Yonan *et al.* (23) found that the assay was sensitive to the concentration of magnesium chloride ($MgCl_2$) and Hardy-Weinberg equilibrium frequencies of genotypes could only be achieved at lower concentrations of $MgCl_2$ and found that a disease association with 5HTTLPR disappeared when the revised assay was used. Although genotyping errors could lead to false detection of association, consistent association with one allele detected across studies would require a mechanism by which a suboptimal assay systematically biases genotypes of cases versus control subjects. Such a mechanism is hard to contemplate, but ascertainment, age, and storage of DNA
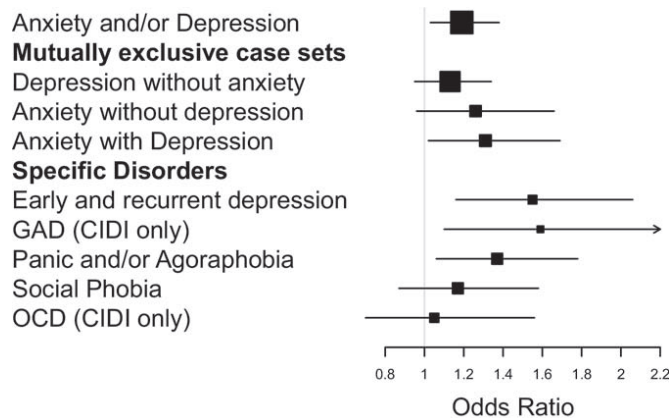
**Figure 4.** Forrest plot of odds ratio and their 95% confidence intervals for different case groups for rs6354 showing increased frequency of allele T in cases compared to control subjects (frequency in control subjects .78). The case groups are combined from both the SSAGA and CIDI questionnaires and reflect the diagnostic classes achievable from them.

samples can often differ between case and control sample sets. Our concern with the 5HTTLPR assay cannot conclusively undermine the meta-analysis results of a large number of studies in which the odds ratio for S versus L alleles was estimated to be 1.12 from 24 studies (11), as it is possible that optimal genotyping conditions in each study had been achieved. However, it is noteworthy that many of the contributing studies had too few samples to detect anything but extreme deviation from HWE. Furthermore, potential problems with the 5HTTLPR assay are often overlooked in discussions of results of 5HTTLPR associations (e.g., [2]).

We have developed a new assay for 5HTTLPR based on two pairs of PCR primers that bind directly to the repeat sequences and that can be multiplexed into a single assay. Further, fourfold multiplexing was achieved by delayed double loading of sample replicates onto the same gel. In operation, our protocol took only 1 hour to genotype 384 different DNA samples, providing replication of two assays for each sample. Genotyping of 13 SNPs within a 38-kb region flanking 5HTTLPR has enabled us to examine linkage disequilibrium landscape of the region and identify a two-SNP haplotype proxy for it; the CA haplotype of rs4251417 and rs2020934 is coupled with the short allele of 5HTTLPR ($r^2 = .72$) (Figure 2). Ideally, a higher $r^2$ is desirable for SNP proxies (usually an $r^2$ threshold of .8 is set for selection of tagging SNPs), but with large sample sizes, genotyping this two-SNP proxy could usefully replace the time-consuming 5HTTLPR assay in association studies, i.e., a sample size increased by a factor of 1/.72 is required when genotyping the two-SNP proxy to achieve the same power as direct genotyping of 5HTTLPR (53). Others (45,46,54) have genotyped SNPs in the 38-kb region surrounding 5HTTLPR as part of association studies for autism or bipolar disorder. However, none of these studies genotyped SNP rs4251417 whose minor T allele is useful in splitting the majority of L alleles that couple with the A allele of rs2020934 (Figure 2B). Despite the large number of association studies for depression and anxiety-related traits that have genotyped 5HTTLPR, none (to our knowledge) has genotyped more than a couple of additional SNPs in the region. Unfortunately, rs2020934 has not been genotyped as part of the HapMap project and has not been included on any of the genome-wide SNP platforms. SNP rs4251417 is included on the Illumina 610K and 1M chips, but on its own, it is not a useful proxy for 5HTTLPR ($r^2 = .06$).

We have conducted an association study of 5HTTLPR, including the rs25531 polymorphism found on the L allele, as well as an additional 13 SNPs in a 38-kb region surrounding the 5HTTLPR, using a large population cohort rich in psychiatric and personality phenotypic information. Our cases and control subjects represent two independent study samples. The ascertainment for the sample completing the CIDI questionnaire is likely to have identified more extreme cases and control subjects, and this is quantified through personality measures collected several years before administration of the psychiatric interviews (Table 3). Although we might expect to observe more extreme associations with the CIDI cohort, we would expect the direction of the association to be the same in the two cohorts if a true association exists. We found no consistent evidence of association with 5HTTLPR or with functional genotype classes of 5HTTLPR with SNP rs25531. However, we did find consistent evidence of association with SNPs rs6354 and rs2020936 in both independent samples. These two SNPs are in high LD ($r^2 = .95$) and so represent a single association signal; the LD with 5HTTLPR for these SNPs is $r^2 = .01$ and $|D'| = .20$. The association was upheld across multiple (correlated) phenotypic measures with larger differences in allele frequencies occurring with more extreme, homogenous phenotypes (although because of the smaller sample sizes, association $p$ values were not always smaller). For example, association $p$ values with SNP rs6354 for depression and/or anxiety, comorbid anxiety with depression, and recurrent, early-onset depression are .021, .032, and .0027, respectively, reflecting control versus case frequencies of the T allele of .78 versus .81, .83, and .85, and generating allelic OR (95% confidence interval [CI]) of 1.19 (1.03–1.38), 1.31 (1.02–1.69), and 1.55 (1.16–2.06), respectively. The higher OR for subsets of cases that are likely to be genetically more homogeneous provides support that this association is not a false positive. Indeed, permutation testing confirmed this, generating a significance of $p = .0062$, which includes correction for testing of multiple markers. The association was supported through analysis of the quantitative trait neuroticism with measures available on more than 2600 individuals collected several years before the diagnostic psychiatric interviews, $p = .019$ with rs6354, reflecting a standardized difference between homozygotes of $d = .2$, with a higher score associated with their T and A alleles, respectively. There was no evidence for association with extraversion or harm avoidance with these SNPs, and within the anxiety disorders, no evidence was found for association with OCD, although the number of cases was low.

Single nucleotide polymorphisms rs6354 and rs2020936 are positioned in the 15.5-kb region that separates the 5HTTLPR and STin2 (they are 1.3 kb and 2.2 kb from STin2, respectively). These SNPs are not in LD with 5HTTLPR (Figure 2). Using European (CEU) genotypes downloaded from the HapMap (47) website and haplotype information (55), we deduce that rs6354 and rs2020936 are in the same haplotype block as STin2, but we are unable to deduce the likely coupling of alleles between these polymorphisms (Supplement 1). As for 5HTTLPR, mixed association results have been reported for STin2. A meta-analysis of eight association studies of STin2 and major depression (11) reported an OR of 1.33 for the rare allele 9 versus 12, but this failed to reach significance because of small sample size, low frequency of the risk allele, and heterogeneity between studies. Their result for allele 10 versus 12 was not significant. However, a recent association study of panic and social anxiety disorder reported significant association with rs140701 ($|D'| = 1$, $r^2 \approx .2$ with rs6354) and no association with 5HTTLPR (56).

Furthermore, in a study of allelic expression of variants within SLC6A4 and its promoter, the most significant correlation was found with two SNPs in intron 1 (rs16965628 and rs2020933, $r^2$ between them of .79), which are not in LD with 5HTTLPR. Taken together, these studies provide evidence for a role for SLC6A4, but not necessarily 5HTTLPR, in the etiology of depression and anxiety disorder, although whether these results reflect single or multiple causal variants is unclear.

1. Hariri AR, Holmes A (2006): Genetics of emotional regulation: The role of the serotonin transporter in neural function. *Trends Cogn Sci* 10:182–191.
2. Canli T, Lesch KP (2007): Long story short: The serotonin transporter in emotion regulation and social cognition. *Nat Neurosci* 10:1103–1109.
3. Papakostas GI, Thase ME, Fava M, Nelson JC, Shelton RC (2007): Are antidepressant drugs that combine serotonergic and noradrenergic mechanisms of action more effective than the selective serotonin reuptake inhibitors in treating major depressive disorder? A meta-analysis of studies of newer agents. *Biol Psychiatry* 62:1217–1227.
4. Hu XZ, Oroszi G, Chun J, Smith TL, Goldman D, Schuckit MA (2005): An expanded evaluation of the relationship of four alleles to the level of response to alcohol and the alcoholism risk. *Alcohol Clin Exp Res* 29:8–16.
5. Hu XZ, Lipsky RH, Zhu G, Akhtar LA, Taubman J, Greenberg BD, *et al.* (2006): Serotonin transporter promoter gain-of-function genotypes are linked to obsessive-compulsive disorder. *Am J Hum Genet* 78:815–826.
6. Kraft JB, Slager SL, McGrath PJ, Hamilton SP (2005): Sequence analysis of the serotonin transporter and associations with antidepressant response. *Biol Psychiatry* 58:374–381.
7. Nakamura M, Ueno S, Sano A, Tanabe H (2000): The human serotonin transporter gene linked polymorphism (5-HTTLPR) shows ten novel allelic variants. *Mol Psychiatry* 5:32–38.
8. Wendland JR, Martin BJ, Kruse MR, Lesch KP, Murphy DL (2006): Simultaneous genotyping of four functional loci of human SLC6A4, with a reappraisal of 5-HTTLPR and rs25531. *Mol Psychiatry* 11:224–226.
9. Rasmussen HB, Werge TM (2007): Novel procedure for genotyping of the human serotonin transporter gene-linked polymorphic region (5-HTTLPR)—a region with a high level of allele diversity. *Psychiatr Genet* 17:287–291.
10. Lesch KP, Bengel K, Heils A, Sabol SZ, Greenberg BD, Petri S, *et al.* (1996): Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* 274:1527–1531.
11. Lopez-Leon S, Janssens A, Ladd A, Del-Favero J, Claes SJ, Oostra BA, *et al.* (2008): Meta-analyses of genetic studies on major depressive disorder. *Mol Psychiatry* 13:772–785.
12. Schinka JA (2005): Measurement scale does moderate the association between the serotonin transporter gene and trait anxiety: Comments on Munafo, *et al. Mol Psychiatry* 10:892–893.
13. Willis-Owen SAG, Turri MG, Munafo MR, Surtees PG, Wainwright NWJ, Brixey RD, *et al.* (2005): The serotonin transporter length polymorphism, neuroticism, and depression: A comprehensive assessment of association. *Biol Psychiatry* 58:451–456.
14. Blaya C, Salum GA, Lima MS, Leistner-Segal S, Manfro GG (2007): Lack of association between the serotonin transporter promoter polymorphism (5-HTTLPR) and panic disorder: A systematic review and meta-analysis. *Behav Brain Funct* 3:41.
15. Lin PY (2007): Meta-analysis of the association of serotonin transporter gene polymorphism with obsessive-compulsive disorder. *Prog Neuropsychopharmacol Biol Psychiatry* 31:683–689.
16. Fanous A, Gardner CO, Prescott CA, Cancro R, Kendler KS (2002): Neuroticism, major depression and gender: A population-based twin study. *Psychol Med* 32:719–728.
17. Kendler KS, Gatz M, Gardner CO, Pedersen NL (2006): Personality and major depression: A Swedish longitudinal, population-based twin study. *Arch Gen Psychiatry* 63:1113–1120.
18. Jardine R, Martin NG, Henderson AS (1984): Genetic covariation between neuroticism and the symptoms of anxiety and depression. *Genet Epidemiol* 1:89–107.
19. Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, *et al.* (2003): Influence of life stress on depression: Moderation by a polymorphism in the 5HTT gene. *Science* 301:386–389.
20. Ioannidis JPA, Trikalinos TA, Khoury MJ (2006): Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *Am J Epidemiol* 164:609–614.
21. Kavvoura FK, Ioannidis JPA (2008): Methods for meta-analysis in genetic association studies: A review of their potential and pitfalls. *Hum Genet* 123:1–14.
22. Kaiser R, Tremblay PB, Roots I, Brockmuller J (2002): Validity of PCR with emphasis on variable number of tandem repeat analysis. *Clin Biochem* 35:49–56.
23. Yonan AL, Palmer AA, Gilliam TC (2006): Hardy-Weinberg disequilibrium identified genotyping error of the serotonin transporter (SLC6A4) promoter polymorphism. *Psychiatr Genet* 16:31–34.
24. Holmans P, Weissman MM, Zubenko GS, Scheftner WA, Crowe RR, DePaulo JR, *et al.* (2007): Genetics of recurrent early-onset major depression (GenRED): Final genome scan report. *Am J Psychiatry* 164:248–258.
25. Hettema JM (2008): What is the genetic relationship between anxiety and depression? *Am J Med Genet C Semin Med Genet* 148C:140–146.
26. Heath AC, Cloninger CR, Martin NG (1994): Testing a model for the genetic structure of personality: A comparison of the personality systems of Cloninger and Eysenck. *J Pers Soc Psychol* 66:762–775.
27. Lake RIE, Eaves LJ, Maes HHM, Heath AC, Martin NG (2000): Further evidence against the environmental transmission of individual differences in neuroticism from a collaborative study of 45,850 twins and relatives on two continents. *Behav Genet* 30:223–233.
28. Eysenck SBG, Eysenck HJ, Barrett P (1985): A revised version of the Psychoticism scale. *Pers Individ Dif* 6:21–30.
29. Cloninger CR, Przybeck TR, Svrakic DM (1991): The Tridimensional Personality Questionnaire: U.S. normative data. *Psychol Rep* 69:1047–1057.
30. Bierut LJ, Heath AC, Bucholz KK, Dinwiddie SH, Madden PA, Statham DJ, *et al.* (1999): Major depressive disorder in a community-based twin sample: Are there different genetic and environmental contributions for men and women? *Arch Gen Psychiatry* 56:557–563.
31. Bucholz KK, Cloninger CR, Dinwiddie DH, Hesselbrock VM, Nurnberger JL, Reich T, *et al.* (1994): A new, semi-structured psychiatric interview for use in genetic linkage studies: A report of the reliability of the SSAGA. *J Stud Alcohol* 55:149–158.
32. American Psychiatric Association (1987): *Diagnostic and Statistical Manual of Mental Disorders, 3rd ed rev.* Washington, DC: American Psychiatric Press.
33. Macgregor S, Lind P, Bucholz K, Hansell N, Madden P, Richter M, *et al.* (2009): Associations of ADH and ALDH2 gene variation with self report alcohol reactions, consumption and dependence: An integrated analysis. *Hum Mol Genet* 18:580–593.
34. World Health Organization (1997): *Composite International Diagnostic Interview, Version 2.1.* Geneva: World Health Organization.
35. American Psychiatric Association (1994): *Diagnostic and Statistical Manual of Mental Disorders, 4th ed.* Washington, DC: American Psychiatric Publishing.
36. Risch N, Zhang H (1995): Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* 268:1584–1589.
37. Purcell S, Cherny SS, Hewitt JK, Sham PC (2001): Optimal sibship selection for genotyping in quantitative trait locus linkage analysis. *Hum Hered* 52:1–13.

38. Kirk KM, Birley AJ, Statham DJ, Haddon B, Lake RI, Andrews JG, *et al.* (2000): Anxiety and depression in twin and sib pairs extremely discordant and concordant for neuroticism: Prodromus to a linkage study. *Twin Res* 3:299–309.

39. Birley AJ, Gillespie NA, Heath AC, Sullivan PF, Boomsma DI, Martin NG (2006): Heritability and nineteen-year stability of long and short EPQ-R neuroticism scales. *Pers Individ Dif* 40:737–747.

40. Wray NR, Birley AJ, Sullivan PF, Visscher PM, Martin NG (2007): Genetic and phenotypic stability of measures of neuroticism over 22 years. *Twin Res Hum Genet* 10:695–702.

41. Saccone SF, Pergadia ML, Loukola A, Broms U, Montgomery GW, Wang JC, *et al.* (2007): Genetic linkage to chromosome 22q12 for a heavy-smoking quantitative trait in two independent samples. *Am J Hum Genet* 80:856–866.

42. Hansell NK, Agrawal A, Whitfield JB, Morley KI, Zhu G, Lind PA, *et al.* (2008): Long-term stability and heritability of telephone interview measures of alcohol consumption and dependence. *Twin Res Hum Genet* 11:287–305.

43. Miller SA, Dykes DD, Polesky HF (1988): A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215.

44. Heils A, Teufel A, Petri S, Stöber G, Riederer P, Bengel D, *et al.* (1996): Allelic variation of human serotonin transporter gene expression. *J Neurochem* 66:2621–2624.

45. Kim SJ, Cox N, Courchesne R, Lord C, Corsello C, Akshoomoff N, *et al.* (2002): Transmission disequilibrium mapping at the serotonin transporter gene (SLC6A4) region in autistic disorder. *Mol Psychiatry* 7:278–288.

46. Ramoz N, Reichert JG, Corwin TE, Smith CJ, Silverman JM, Hollander E, *et al.* (2006): Lack of evidence for association of the serotonin transporter gene SLC6A4 with autism. *Biol Psychiatry* 60:186–191.

47. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, *et al.* (2007): A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861.

48. Barrett JC, Fry B, Maller J, Daly MJ (2005): Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265.

49. de Bakker PIW, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D (2005): Efficiency and power in genetic association studies. *Nat Genet* 37:1217–1223.

50. Freeman MF, Tukey JW (1950): Transformations related to the angular and the square root. *Ann Math Stat* 21:607–611.

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

52. Wray NR (2005): Allele frequencies and the l2 measure of linkage disequilibrium: Impact on design and interpretation of association studies. *Twin Res Hum Genet* 8:87–94.

53. Pritchard JK, Przeworski M (2001): Linkage disequilibrium in humans: Models and data. *Am J Hum Genet* 69:1–14.

54. Mansour HA, Talkowski ME, Wood J, Pless L, Bamne M, Chowdari KV, *et al.* (2005): Serotonin gene polymorphisms and bipolar I disorder: Focus on the serotonin transporter. *Ann Med* 37:590–602.

55. Zaboli G, Jonsson EG, Gizatullin R, De Franciscis A, Asberg M, Leopardi R (2008): Haplotype analysis confirms association of the serotonin transporter (5-HTT) gene with schizophrenia but not with major depression. *Am J Med Genet B Neuropsychiatr Genet* 147:301–307.

56. Strug LJ, Suresh R, Fyer AJ, Talati A, Adams PB, Li W, *et al.* (2008): Panic disorder is associated with the serotonin transporter gene (SLC6A4) but not the promoter region (5-HTTLPR) [published online ahead of print July 29]. *Mol Psychiatry*.