



Published in final edited form as:

Psychiatr Genet. 2009 December ; 19(6): 292–304.

Association study of *DTNBP1* with schizophrenia in a US sample

Lingjun Zuo^{1,2}, Xingguang Luo^{1,2}, Henry R. Kranzler³, Lingeng Lu⁴, Robert A. Rosenheck^{1,2}, Joyce Cramer^{1,2}, Daniel P. van Kammen⁵, Joseph Erdos^{1,2}, Dennis S. Charney⁶, John Krystal^{1,2}, and Joel Gelernter^{1,2,7,8}

¹ Department of Psychiatry, Yale University School of Medicine, New Haven, CT

² VA Connecticut Healthcare System, West Haven Campus, CT

³ Department of Psychiatry, University of Connecticut School of Medicine, Farmington, CT

⁴ Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT

⁵ CHDI, Inc. New York, NY

⁶ Department of Psychiatry, Mount Sinai School of Medicine, New York, NY, USA

⁷ Department of Genetics, Yale University School of Medicine, New Haven, CT

⁸ Department of Neurobiology, Yale University School of Medicine, New Haven, CT

Abstract

Background—Straub et al. (2002b) located a susceptibility region for schizophrenia at the *DTNBP1* locus. At least 40 studies (including one study in US populations) attempted to replicate this original finding, but the reported findings are highly diverse and at least five pathways by which dysbindin protein might be involved in schizophrenia have been proposed. The present study aimed to test the association in two common US populations by using powerful analytic methods.

Methods—Six markers at *DTNBP1* were genotyped by mass spectroscopy (“MassARRAY” technique) in a sample of 663 subjects, including 346 healthy subjects [298 European-Americans (EAs) and 48 African-Americans (AAs)] and 317 subjects with schizophrenia (235 EAs and 82 AAs). Thirty-eight ancestry-informative markers (AIMs) were genotyped in this sample to infer the ancestry proportions. Diplotype, haplotype, genotype, and allele frequency distributions were compared between cases and controls, controlling for possible population stratification, admixture, and sex-specific effects, and taking interaction effects into account, using a logistic regression analysis (an extended structured association (SA) method).

Results—Conventional case-control comparisons showed that genotypes of the markers P1578 (rs1018381) and P1583 (rs909706) were nominally associated with schizophrenia in EAs and in AAs, respectively. These associations became less or non-significant after controlling for population stratification and admixture effects (using SA or regression analysis), and became non-significant after correction for multiple testing. However, regression analysis demonstrated that the common diplotypes (ACCCTT/GCCGCC or GCCGCC/GCCGCC) and the interaction effects of haplotypes GCCGCC × GCCGCC significantly affected risk for schizophrenia in EAs, effects that were modified by sex. Fine-mapping using δ or J statistics located the specific markers (δ : P1328; J: P1333) closest to the putative risk sites in EAs.

Conclusions—The present study shows that *DTNBPI* is a risk gene for schizophrenia in EAs. Variation at *DTNBPI* may modify risk for schizophrenia in this population.

Keywords

schizophrenia; *DTNBPI*; admixture; structured association (SA) method

Introduction

At least twenty-three complete or nearly complete genome scans for schizophrenia in 27 samples have been published, which have localized risk regions for schizophrenia to numerous different chromosomes (reviewed by Sullivan 2005). Since Straub et al. (1995) and Kendler et al. (1996) initially reported the linkage of markers mapped to chromosome 6p24-21 to schizophrenia spectrum disorders, there have been at least 15 additional linkage studies; of these, at least 7 provided supportive evidence for susceptibility loci on chromosome 6p (Schwab et al. 1995; Levinson et al. 1996; Maziade et al. 1997; Lindholm et al. 1999; Turecki et al. 1997; Straub et al. 2002a; Lewis et al. 2003). These susceptibility loci span a broad region of 25Mb between D6S296 and D6S291, including four possibly distinct subregions: 6p25-24, 6p24, 6p23-22, and 6p21 (reviewed by Straub et al. 2002b). Association studies using linkage disequilibrium mapping methods have served to fine-map the risk alleles within these subregions. Using a family-based association method, Straub et al. (2002b) initially identified the dystrobrevin-binding protein 1 gene, i.e., the dysbindin gene (*DTNBPI*) at 6p22.3, as a susceptibility gene for schizophrenia based on a set of 270 Irish high-density pedigrees. They found several polymorphisms within this gene that associated with schizophrenia (see Table 2). At least 40 family-based or population-based association studies have attempted to replicate this initial finding (summarized in Table 1), although not necessarily in the strict sense of repeating the design and methods of the initial study. One linkage study in an Israeli isolate directly located a risk region for major psychiatric disorders at the *DTNBPI* locus (Kohn et al. 2004). At least twenty-two association studies supported the associations between *DTNBPI* and schizophrenia in different populations, but seven did not. However, the positive findings from these studies were variable: (1) Some putative risk alleles (even at the same marker locus) are minor alleles in some populations [e.g., P1635^G in Irish (Straub et al. 2002b)] but common alleles in other populations [e.g., P1635^A in Bulgarian (Kirov et al. 2004) and in German-Israeli (Schwab et al. 2003)]; some risk haplotypes are rare in some populations [e.g., in Irish (Straub et al. 2002b; van den Oord et al. 2003)] but common in other populations [e.g., in German (Schwab et al. 2003), in Chinese (Tang et al. 2003) and in Japanese (Numakawa et al. 2004)], and some common or rare haplotypes protect against disease (Williams et al. 2004); (2) Some markers or haplotypes, even in the same population (e.g., Irish), are associated with schizophrenia in some studies (e.g., Straub et al. 2002b and Williams et al. 2004), but not in other studies (e.g., Morris et al. 2003); (3) The most significantly-associated risk markers are different across different studies (e.g., P1635 in Straub et al. 2002b, Kirov et al. 2004 and Numakawa et al. 2004; but P1320 in Schwab et al. 2003); (4) The risk or protective haplotypes have different block boundaries in different populations and the numbers of these haplotypes differ among studies.

Most researchers (e.g., Straub et al. 2002b; Funke et al. 2004; Bray et al. 2005) have attributed the diversity of findings to allelic heterogeneity or haplotypic heterogeneity *per se*. However, other potential issues need to be considered, particularly, sampling bias and sampling variance. The sampling in all of these studies was non-random. Serious sampling bias may lead to “surprising” (i.e., unexpected) findings that might be false. Mild sampling bias or sampling variance may lead to inconsistent findings within a single gene, because of various stratification effects from variables such as population, familiarity, age, sex, etc.,

which are not random in the sampling. (1) *Population stratification effects*. Different populations have different evolutionary histories with different numbers of generations. The difference in generation number (i.e., the age of the population) leads to a difference in recombination that leads to LD decay and thus results in different haplotype block sizes between populations. Population-specific gene frequency distributions and haplotype block sizes often lead to population-specificity of associations between genes and diseases. To guard against false positives, and because of unfeasibility of random sampling from all populations, many researchers limit sampling to one specific population (even for family-based studies) or conduct analyses only within single populations. Thus, the findings are also limited to that specific population, and replication is required in other populations if the findings are to be generalized. Although more than 30 studies have attempted to replicate the original linkage findings, the results were highly diverse; only two studies were performed in US populations. Additionally, ten studies were performed in at least seven European subpopulations, with diverse findings (see Tables 1, 2), which suggests that the European-Americans that originate from different geographic regions should be taken as potentially admixed. African-Americans (AAs) and Hispanics have high admixture with EAs (Parra et al, 1998; Hoggart et al, 2003), and thus should also be taken as admixed populations. Within a single population, especially those admixed populations, population-based studies are also vulnerable to admixture effects, but this has not been considered in previous studies (see Table 1). Although the family-based design is thought to be immune to population stratification effect, it may be not immune to other stratification effects. (2) *Familiality stratification effects*. Familial patients represent a specific subgroup, different from sporadic patients. The sporadic/familial distinction might lead to the different findings between the family-based studies and the population-based studies, even in the same population. Familiality could confound the association analysis. (3) *Age stratification effects*. Schizophrenia is an age-dependent phenotype. Age could be a stratification factor confounding the association analysis. (4) *Sex stratification effects*. Sex-specificity of schizophrenia has been reported by several studies (e.g., Franzek and Beckmann, 1992; Hafner et al. 1993; Kitamura et al. 1993; Sham et al. 1994; Leung and Chue, 2000; Aleman et al. 2003). In the present study, we stratified the sample by sex or took sex as a covariate in the regression analysis. (5) *Other known or unknown factors might generate stratification effects that contribute to the diversity of findings*. Because completely random sampling so as to randomly distribute these factors usually is unfeasible, replication is very important. As for previous studies, important factors were not randomized in our case-control sample, but the stratification effects of the main confounding factors, including population and sex, were controlled for in our analysis. Additionally, our sample is reasonably representative of the general population, based on similar allele and haplotype frequency distributions of different markers to those from other independent studies within the same populations, e.g., the *OPRM1* haplotype frequency distribution in our AA controls (Luo et al, 2003) is similar to that in the study by Crowley et al. (2003) (comment in Luo et al, 2005a; other comparisons are not shown here), thus, our sample is apparently reasonable for a replication study.

The diversity of findings may also result from design variance and variation in methods among studies. (1) The replication studies used family-based and population-based association designs, which differ in power, e.g., several studies claimed that case-control studies can be more powerful than family-based studies in identifying disease genes, both for qualitative traits (Risch et al. 1996; 1998; 2000) and for quantitative traits (Van den Oord. 1999). (2) Multiple genes with minor effects might contribute to the risk for complex diseases. Detection of these minor effects is sensitive to study power. Differing power may be due to different designs, different sample sizes, different marker sets, and different analytic methods, which may lead to different results. For example, most studies do not exactly “repeat” the initial design and methods, but involve further exploratory analyses and

aim to generally replicate the findings in the sense of identifying some relationship between markers or haplotypes at the locus, and the phenotype.

In summary, the present study aimed to replicate the study by Straub et al. (2002b), but adopted design features intended to overcome some limitations present in some other replication studies. These include (1) controlling for population stratification, admixture effects, and sex stratification effects; (2) preserving the haplotypes with unknown phases in the analysis; (3) taking marker-marker epistasis into account; (4) waiving the requirement of the HWE assumption on haplotype reconstruction; (5) avoiding multiple tests due to involving multiple populations and multiple markers; and (6) fine-mapping the risk sites.

Materials and Methods

1. Subjects

Six hundred sixty-three subjects were included in the study: 346 healthy controls [298 European-Americans (EAs) and 48 African-Americans (AAs)] and 317 subjects with schizophrenia (235 EAs and 82 AAs). Four hundred twenty-seven subjects were male and 208 were female. Males constituted 98.2% of the cases and 41.3% of the controls. Cases and controls showed a roughly matched age distribution. The population groups for individual subjects were classified by ancestry proportions rather than self-report (see below). The diagnosis of schizophrenia was according to DSM-III-R criteria (American Psychiatric Association, 1987) as determined by the Structured Clinical Interview for DSM-III-R (SCID) (Spitzer et al, 1992). The control subjects were screened using the SCID, the Computerized Diagnostic Interview Schedule for DSM-III-R (Blouin et al, 1988), or the Schedule for Affective Disorders and Schizophrenia (Spitzer and Endicott 1975) to exclude major Axis I disorders, including substance dependence, psychotic disorders (including schizophrenia or schizophrenia-like disorders), mood disorders, and major anxiety disorders.

Subjects were recruited at the VA Connecticut Healthcare System, West Haven Campus, the University of Connecticut Health Center, or 14 other Veterans Affairs medical centers (described in Rosenheck et al, 1997). The study was approved by the Institutional Review Boards (IRB) at Yale University School of Medicine, University of Connecticut Health Center, VA Connecticut Healthcare Center, and in some cases additional IRBs at sample collection sites. All subjects signed informed consent, with the exception of a subsample collected at Highland Drive VA (Pittsburgh), which was determined by the Yale IRB to be exempt from review because the research involved use of existing anonymous samples.

2. Marker inclusion

Six markers within *DTNBPI* were genotyped in the present study, including two markers (P1583: rs909706 and P1578: rs1018381) at intron 1, one marker (P1320: rs760761) at intron 3, one marker (P1655: rs2619539) at intron 5, one marker (P1333: rs742105) at intron 7, and one marker (P1328: rs742106) at intron 9 (see Table 3). These markers were selected from the original 12 markers in the study by Straub et al. (2002b), because they could be genotyped by multiplex PCR in the MassARRAY system, and we could validate their allele frequencies in a small sample prior to the high-throughput genotyping. All six markers have also been examined in many other studies and most of them were found to be associated with schizophrenia (see Table 2). The six markers span a total of 136Kb, with an average intermarker distance of 22kb. Most of them are tagSNPs in the HapMap database (www.hapmap.org), and cover most of the information content of *DTNBPI*.

Thirty-eight ancestry-informative markers (AIMs) unlinked to *DTNBPI*, including 37 STRs and one Duffy antigen gene (*FY*) marker (rs2814778) that is highly ancestry-informative, were also genotyped, to examine the population structure of our sample. These markers were

employed in the studies by Stein et al. (2004), Kaufman et al (2004), and Luo et al. (2005b;c); their characteristics were described in the study by Yang et al. (2005), and the genotyping methods have been described in these studies.

3. Genotyping

Genomic DNA was extracted from peripheral blood by standard methods. The six SNPs were genotyped by Matrix-Assisted Laser Desorption Ionization - Time of Flight (MALDI-TOF) Mass Spectrometry via the Sequenom MassARRAY system (SEQUENOM, Inc., San Diego, CA, USA) in three 2-plex PCR, using six pairs of primers. These multiplexes and primers were designed using the MassARRAY™ Assay Design Software and all primers were extended by a 5′ “cap” sequence “ACGTTGGATG” to increase the molecular weight of these primers to > 9000 daltons, so that any residual PCR primer would not interfere with the SNP genotyping software, that is, the PCR primers would not be in the mass range of 5000–9000 Daltons that is used in the genotyping process. PCR was performed in a final volume of 5 µl for each system, which included 2.5–5.0 ng genomic DNA, 200 nM each PCR primer for uniplex reactions or 50 nM each PCR primer for multiplex reactions, 200 µM each dNTP, 1 × HotStar buffer, HotStar Taq polymerase (Qiagen, Inc., Valencia, CA, USA), and 2.5 mM MgCl₂. A strict validation experiment was performed prior to high-throughput genotyping: (1) PCR conditions for multiplex PCR were optimized based on the manufacturer’s recommendation until the genotypes completely agreed with those by the uniplex PCR; (2) One large CEPH family pedigree including 27 individuals, from whom DNAs were available through Coriell (<http://locus.umdj.edu/nigms/ceph/ceph.html>), were genotyped by the optimized PCR to assure that the genotypes completely agreed with the Mendelian rule; (3) The success rate for each multiplex PCR was higher than 90%. Both positive controls (CEPH DNA sample) and negative controls (water and blank) were included in the high-throughput genotyping.

4. Statistical analysis

(1) LD analysis, Hardy-Weinberg Equilibrium (HWE) test, and case-control comparisons for allele and genotype frequency distributions—Pairwise LD between any two *DTNBPI* markers was analyzed separately by population, i.e., EAs and AAs. The value of the standardized disequilibrium coefficient, D' , for each LD pair was calculated and the statistical significance for D' was tested. HWE of the genotype frequency distribution for each marker was tested within different populations, and separately in cases and controls.

The allele and genotype frequencies of the *DTNBPI* markers in different phenotype groups are shown in Table 4. Associations between either the alleles or the genotypes and the phenotypes were analyzed by comparing the allele and genotype frequency distributions between cases and controls (within EAs and AAs, respectively) with exact tests. All of the above tests were performed via PowerMarker software (Liu & Muse, 2004).

(2) Structured association (SA) analysis—EAs and AAs can be taken as admixed populations with different degrees of admixture (Parra et al. 1998; Hoggart et al. 2003; Shriver et al. 2003; Collins-Schramm et al. 2004). The extent of admixture (i.e., ancestry proportions) can be estimated using the program STRUCTURE (Pritchard et al, 2000a) to analyze the 38 AIMs (Yang et al, 2005; Luo et al, 2005c). The case-control design is vulnerable to admixture effects, but the admixture effects on case-control association analysis can be controlled for using the program STRAT (Pritchard et al, 2000b), which adjusts for ancestry proportions to yield a so-called structured association (SA) analysis. The SA method is limited to genotypewise and allelwise analyses. Therefore, the ancestry proportions were also entered into the regression models described below for an extended

analysis, which included diplotypewise, haplotypewise, genotypewise, and allelewise analyses and tested for the population-specificity of associations.

In the present study, in order to increase statistical power by expanding the sample, EAs and AAs were combined as a single admixed sample for association study. Then, EAs and AAs were analyzed separately to identify the sources of the observed associations.

(3) Haplotype and diplotype probability estimation and case-control comparisons for haplotype and diplotype frequency distributions—The program PHASE was used to reconstruct haplotypes and to estimate the probabilities of all likely pairs of haplotypes (i.e., diplotypes) for every individual in this study. This program was developed by Stephens et al. (2001; 2003), based on a Bayesian approach and the Partition Ligation algorithm. These algorithms may be more accurate in reconstructing haplotypes than the Expectation-Maximum (EM) algorithm, especially when the HWE does not hold among some markers, as is the case for our data (see Table 4) (Stephens et al. 2001; Stephens and Donnelly 2003; Niu et al. 2002). In spite of its advantages, PHASE still has its limitations, and thus the below regression analysis on the diplotype and haplotype probabilities estimated by PHASE has to be considered as being exploratory. The haplotypes were reconstructed within two separate subgroups, that is, the genetically-inferred EAs (European ancestry proportion > 0.5) and the genetically-inferred AAs (African ancestry proportion > 0.5).

(4) Regression analysis—A backward stepwise logistic regression analysis was used to test associations between gene and disease. We modeled the analysis with the following equation: $\ln[p/(1-p)] = \beta_0 + \sum \beta_i X_i + \sum \beta_{ij} X_i X_j$, where p is probability of disease; $X_i X_j$ is the interaction between X_i and X_j ; β is regression coefficient; β_i can be interpreted as the magnitude of main effect of X_i , when all other predictor items are equal to 0; $\sum \beta = \beta_i + \beta_{ij}$ can be interpreted as the magnitude of total effect of X_i , when $X_j = 1$ and all other X s are equal to 0; other X s can be interpreted similarly to this. Four kinds of regression models were employed in the present study: X_i includes African ancestry proportions predicted by the program STRUCTURE, sex of individuals, and diplotype probabilities (model 1), haplotype probabilities (model 2), genotypes (model 3), or alleles (model 4). In models 1 and 2, only diplotypes or haplotypes with frequencies > 0.01 (see Table 5) were included; the interaction effects between haplotypes were also considered (diplotypes and haplotypes *per se* have incorporated the interaction information between SNPs). In models 3 and 4, only two genotypes and one allele from each SNP were included, respectively; and the two-way interaction effects between alleles or between genotypes from different SNPs were included as well. In all four models, the interaction effects between sex and diplotypes, haplotypes, genotypes, or alleles were also considered.

Regression analysis using the haplotype probabilities as predictors (model 2) is called haplotype trend regression (HTR). The probabilities, instead of the categories, of haplotypes being included in HTR makes HTR more powerful, because the probabilities preserve more information than does the direct use of categorical variables. The rationale of HTR was first described by Zaykin et al. (2002) and HTR has been widely applied. Regression analysis using the diplotype probabilities as predictors (model 1) is called diplotype trend regression (DTR). DTR has been successfully applied in many previous studies (e.g., Luo et al., 2005b; c; 2006) and its advantages have been demonstrated. DTR increases effective sample size by combining different populations in a single model, avoids multiple testing that would accrue due to the inclusion of multiple populations and markers, controls for population stratification and admixture effects and the potential confounding by sex, allows uncertainty for haplotype inference, obviates the HWE assumption, and takes marker–marker interactions into account.

At a single locus, two alleles could be incorporated into a genotype. Similarly, at the multiple loci, the haplotype information content could be incorporated into the diplotype. The information content of alleles and genotypes from multiple loci could be incorporated into multi-locus haplotypes and diplotypes, respectively. Therefore, the above four regression models are not independent of each other; they actually are equivalent to a single regression model that does not require for correction for multiple testing. Among the four regression models, the diplotype trend regression model (model 1) is most powerful.

Within each regression analysis, multiple predictor variables are tested. These kinds of multiple testing are corrected by the degree of freedom. Thus, p-values derived from the regression analysis do not require for further correction for multiple testing and the significant level (α) is set at 0.05.

(5) Fine-mapping the risk locus—Many measures for LD in case-control samples, e.g., the population attributable risk δ (Levin and Bertell, 1978; Devlin and Risch 1995), have been advanced as tools to fine-map risk loci. Many measures for HWD in case-only samples have also been advanced to fine-map the risk loci, including F , F' , J and J' (Feder et al. 1996; Jiang et al. 2001). These statistics were used for fine-mapping the risk locus in the present study. Because there are no methods available to test the statistical significance for δ or J statistics, we used diplotype trend regression analysis or haplotype trend regression analysis (at the “whole gene” level) to test the statistical significance of gene-disease association first, and then we used a δ or J statistic to fine-map the risk site (at a “single-point” level) within this gene. The marker with the highest δ or J value is thought to be closest to the putative disease locus.

Results

1. *DTNBP1* markers were in several haplotype blocks in patterns that differed by population. Genotype frequency distributions of some markers were in HWD in cases or in controls. Genotypes and/or alleles of two markers were nominally associated with schizophrenia in EAs and AAs, respectively. Diplotype GCCGCC/GCCGCC was nominally associated with schizophrenia in males and/or females in EAs or EAs+AAs.

Pairwise LD analysis showed that P1333 and P1655 were in one haplotype block both in EAs ($D'=0.995$) and AAs ($D'=1.000$); in EAs, P1320 and P1578 belonged to one haplotype block ($D'=0.993$) and in AAs, P1578 and P1583 were in one haplotype block ($D'=1.000$). There were no significant differences in LD between cases and controls (data not shown).

In EAs, P1320 ($p=0.002$) and P1578 ($p=0.031$) were in HWD in controls; in AAs, P1333 ($p=0.011$) and P1655 ($p=0.013$) were in HWD in controls and P1328 ($p=0.001$) was in HWD in cases. After correction for multiple testing using SNPSpD (Nyholt, 2004; Luo et al. 2005b), where $\alpha=0.01$, only P1320 in EA controls and P1328 in AA cases remained in significant HWD (see Table 4).

Case-control comparisons showed that the genotypes of P1578 ($p=0.015$) and P1583 ($p=0.052$) were nominally associated with schizophrenia in EAs and in AAs, respectively. After controlling for population stratification and admixture effect using the SA method, genotypes of P1578 were suggestively associated with schizophrenia in the combined sample ($p=0.070$, in EAs+AAs); decomposing the association by ethnicity, genotypes ($p=0.045$) and alleles ($p=0.047$) of P1578 yielded a significant association with phenotype in EAs, and genotypes of P1583 were suggestively associated with phenotype in AAs ($p=0.081$). After correction

for multiple testing using SNPSpD, where $\alpha=0.01$, none of these remained significantly associated with schizophrenia (see Table 4).

Case-control comparisons showed that the frequencies of diplotypes GCCGCC/GCCGCC were nominally associated with schizophrenia in EAs or in EAs+AAs. In males, its frequencies were nominally lower in cases than controls (In EAs: $f=0.123$ vs. 0.203 ; $p=0.076$; in EAs+AAs: $f=0.125$ vs. 0.204 ; $p=0.053$). In females, its frequencies were nominally higher in cases than controls (In EAs: $f=0.647$ vs. 0.132 ; $p=0.049$; in EAs+AAs: $f=0.485$ vs. 0.137 ; $p=0.101$).

2. Two ancestries, i.e., European and African, were detected in our sample.

One hundred percent of self-reported EAs were “genetic” EAs (European ancestry proportion >0.5) and 100% of self-reported AAs were “genetic” AAs (African ancestry proportion >0.5). Within the 533 EA subjects, the degree of admixture was 1.9% (the total estimated weight of African ancestry proportions divided by $N: 9.89 \div 533$); the degree of admixture was 2.9% for EA cases and 1.0% for EA controls. Within the 130 AA subjects, the degree of admixture detected was 4.7% (the total estimated weight of European ancestry proportions divided by $N: 6.1 \div 130$); the degree of admixture was 5.8% for AA cases and 2.8% for AA controls.

3. Regression analysis demonstrated that diplotypes and haplotypes at the *DTNBPI* locus were associated with schizophrenia (Table 6).

Only the independent variables that have contributions to the risk for disease and whose contributions were statistically significant were retained in the final logistic regression equations (see Table 6). Four kinds of regression models including the diplotypewise, haplotypewise, genotypewise, and/or allelwise analyses showed that men were more common among cases compared with controls in the combined sample (i.e., EAs+AAs), in EAs, and in AAs ($\beta_{\text{female}} < 0$, and $\Sigma\beta = \beta_{\text{female}} + \beta_s$ for interaction effects of female < 0); African ancestry was more common in patients with schizophrenia compared with controls in the combined sample ($\beta_{\text{ancestry}} > 0$).

In females (i.e., when female=1), (i) the diplotype ACCCTT/GCCGCC significantly increased risk for schizophrenia in the combined sample ($\Sigma\beta = (\beta \text{ for ACCCTT/GCCGCC}) + (\beta \text{ for female} \times \text{ACCCTT/GCCGCC}) = 2.408 - 0.609 > 0$; $p < 0.05$); (ii) the diplotype GCCGCC/GCCGCC significantly increased risk for schizophrenia in EAs ($\beta = 2.499 > 0$; $p < 0.05$); (iii) the interaction of haplotypes GCCGCC \times GCCGCC significantly increased risk for schizophrenia both in the combined sample ($\Sigma\beta = (\beta \text{ for GCCGCC} \times \text{GCCGCC}) + (\beta \text{ for female} \times \text{GCCGCC} \times \text{GCCGCC}) = 2.646 - 0.679 = 1.967 > 0$; $p < 0.05$) and in EAs ($\beta = 3.016 > 0$; $p < 0.05$); the magnitude of this interaction effect (GCCGCC \times GCCGCC) in EAs ($\beta = 3.016$) did not increase when combining EAs and AAs ($\Sigma\beta = 1.967$). In males (i.e., when female=0), the diplotype ACCCTT/GCCGCC ($\beta = -0.609$; $p < 0.05$) and the interaction of haplotypes GCCGCC \times GCCGCC ($\beta = -0.679$; $p < 0.05$) significantly decreased risk for schizophrenia in the combined sample.

Regression analysis did not detect associations in either EAs or AAs between any genotype or allele and schizophrenia. In AAs, no gene effects were found in the diplotypewise or the haplotypewise analyses either.

4. Fine-mapping the risk locus using δ , the putative risk locus was closest to P1328 in EAs. Fine-mapping the risk locus using J, the putative risk locus was closest to P1333 in EAs. (δ and J values are shown in the legend of Table 4).

Discussion

The present study demonstrated that the diplotypes and haplotypes at *DTNBPI* locus affected risk for schizophrenia in EAs. We conclude that *DTNBPI* is a risk gene for schizophrenia, and it may harbor a risk locus for the disorder. The present study has also provided additional map information regarding the probable location of functional variants within the locus.

Conventional case-control comparisons on allele and genotype frequency distributions showed that two polymorphisms (P1578: $p=0.015$ and P1583: $p=0.052$) were nominally associated with schizophrenia in EAs and in AAs, respectively. The associations became less significant after controlling for population stratification and admixture effects using the SA method and were no longer significant after the correction for multiple tests. These findings suggest that the methods were not powerful enough to identify the gene as a susceptibility gene, possibly due to a small effect size. Moreover, these two methods have other limitations. For example, the conventional case-control comparison method is vulnerable to population stratification and admixture effects; the SA method cannot handle unphased haplotype data; both methods are limited by multiple testing; potential confounders such as sex and age cannot be controlled for by either method; and neither method is capable of considering marker-marker interaction effects. These limitations reduce the statistical power, accuracy, and robustness of both methods, so that the results are considered to be exploratory.

Regression analysis overcomes these limitations and thus increases the statistical power and leads to more accurate and robust findings. Cases and controls, and EAs and AAs, were combined in one regression model to increase sample size; different markers were entered in one regression model to avoid multiple tests; ancestry proportions were entered as a covariate in the regression model to control for population stratification and admixture effects on association analysis; data on sex were entered in the regression model as a covariate to take into account the sex-specificity of the prevalence of schizophrenia and correct for asymmetric sampling of cases and controls, thereby controlling for its stratification effects and potential confounding effects on the association analysis; the phased and unphased diplotype and haplotype data, which are thought to contain more information than single markers in many cases, were included in the analysis; finally, marker-marker interaction effects and marker-covariate interaction effects were considered to avoid erroneously interpreting the main effect of each marker in the presence of a significant interaction.

P1328 was found to be in HWD in AA cases but in HWE in AA controls, which may be an indication of association between P1328 and schizophrenia (Feder et al. 1996; Nielsen et al. 1999; Jiang et al. 2001; Hoh et al. 2001; Lee 2003; Hao et al. 2004; Wittke-Thompson et al. 2005; Luo et al. 2005b). P1320 and P1578 were in HWD in EA controls, and P1333 and P1655 were in HWD in AA controls, which most likely resulted from sampling bias, or unrecognized copy number variation in this genomic region [such variation in other regions has been related to schizophrenia risk (Walsh et al, 2008)]; it is unlikely to have resulted from genotyping errors [the genotyping missing rates for these four markers in those groups were 2.01%, 2.68%, 2.08% and 0%, respectively]. The presence of HWD led us to use a Bayesian approach and the Partition Ligation algorithm instead of the Expectation-Maximum (EM) algorithm to reconstruct diplotypes and haplotypes. When the diplotype and haplotype data were analyzed in our regression models, the regression method was independent of the HWE assumption. Additionally, the predicted diplotype and haplotype probabilities that can be analyzed by regression methods are continuous variables, which usually are more informative than diplotypes or haplotypes (i.e., categorical variables).

In view of the advantages of the regression method and the limitations of conventional association analysis methods (including the HWD test, case-control comparison and the SA method), we believe that the results from the regression analysis are more accurate and robust, and the different results may reflect the greater accuracy and robustness of the regression method. Using regression analysis, we found that: (1) Males predominate in cases both in EAs and AAs in our sample. Although the incidence of schizophrenia differs by sex (McGrath et al, 2004), the imbalance on sex in our sample is mostly due to a sampling bias (males constituted 98.2% of the cases and 41.3% of the controls). Thus, sex data were taken as a key confounder for gene-disease association analysis and the interaction effects between sex and gene were also considered. (2) African ancestry was more common in cases (82/317 (25.9%) were AA) than controls (48/346 (13.9%) were AA). However, when we analyzed the data separately for EAs and AAs, we noted that African ancestry was also more common in EA cases than controls (2.9% vs. 1.0%; $p=0.001$), and European ancestry was more common in AA cases than controls (5.8% vs. 2.8%; $p=0.052$), suggesting that the degree of admixture *per se* was higher in patients with schizophrenia than in controls. This asymmetry in the degree of admixture between cases and controls is probably also attributable to sampling bias. It is difficult to avoid biased sampling in relation to admixture, since it is not feasible to measure the degree of admixture clinically in order to match cases and controls during the sampling process. Instead, a genetic experiment makes it possible to measure the extent of admixture, so that its potential confounding effects on association analysis can be controlled for. An alternative explanation for the association between the degree of admixture and schizophrenia is that admixture *per se* may increase risk for schizophrenia. Further studies are warranted to test this hypothesis. (3) In the combined sample or in EAs, the most common diplotype and haplotype were ACCCTT/GCCGCC ($f=0.161$) and GCCGCC ($f=0.370$), respectively; the second most common diplotype is GCCGCC/GCCGCC ($f=0.153$). In AAs, the most common diplotype and haplotype were GCCGCC/GCCGCC ($f=0.157$) and GCCGCC ($f=0.291$). Generally, for the majority of individuals in a population, the common haplotypes and diplotypes protect against a disease that is present at low frequency in the population, as observed in the present study. However, these gene effects can be modified by sex. For example, in the combined sample, the common diplotype ACCCTT/GCCGCC and the interaction of the common haplotypes GCCGCC \times GCCGCC protected against ($\beta < 0$) schizophrenia only in males, which constituted the majority of our sample. This is basically consistent with the results from straightforward case-control comparison on diplotype frequency distributions, i.e., the frequencies of diplotype GCCGCC/GCCGCC were lower in male cases than male controls both in EAs and in combined sample. In females, the common diplotype and haplotype increased risk for schizophrenia ($\beta > 0$), both in EAs and in the combined sample. This is also basically consistent with the results from straightforward case-control comparison on diplotype frequency distributions, i.e., the frequencies of diplotype GCCGCC/GCCGCC were higher in female cases than female controls both in EAs and in combined sample. (The findings in females might be chance findings given that only 2% of cases were women). The associations of diplotypes and haplotypes with schizophrenia suggest that *DTNBP1* may harbor a disease locus for schizophrenia, which is in LD with these risk or protective diplotypes or haplotypes. The magnitude of the interaction effect of GCCGCC \times GCCGCC in EAs ($\beta=3.016$) did not increase when combined with AAs ($\Sigma\beta=1.967$), which suggests that the gene effects were significant mainly in EAs and that the addition of AA subjects did not increase information. Fine-mapping using δ or J located the specific markers (δ : P1328; J: P1333) closest to the putative risk sites in EAs. These fine-mapping methods have limitations; for example, δ is subject to the assumption of HWE and J ignores the information from controls, which may explain the different localization provided by the methods.

All of the risk markers were located in introns spanning *DTNBPI*. They may affect risk for schizophrenia in three possible ways. First, these markers may be in LD with nearby functional variation. However, despite intensive resequencing efforts (e.g., Liao and Chen, 2004; Williams et al, 2004), no *DTNBPI* coding variants have yet been identified. Second, *DTNBPI* has at least 12 different known mRNA transcripts resulting from alternative splicing (Williams et al. 2004), and these markers may be involved in the post-transcriptional alternative RNA splicing process, so that genetic variation in pre-mRNA may yield distinct mature mRNAs, which can be translated to distinct dysbindin proteins with differing, or even opposing activities. Third, these intronic variants and their haplotypes *per se* might directly affect the expression of dysbindin protein in the brain and thus directly affect the susceptibility to schizophrenia. Recently, Bray et al. (2003, 2005) detected strong allele-specific and haplotype-specific expression of *DTNBPI* in the brain, and several other studies have reported significant reduction of *DTNBPI* expression in the brains of patients with schizophrenia (Weickert et al, 2004; Numakawa et al, 2004; McClintock et al, 2003; Talbot et al, 2004), raising the possibility that *cis*-acting variation may contribute to the role of *DTNBPI* in the etiology of schizophrenia. Because the present study does not repeat the design and methods of the initial study, but involve novel approaches, replication of our findings is warranted in the future.

How dysbindin protein affects risk for schizophrenia is still under investigation. Because dysbindin protein, binding with β -dystrobrevin, is likely a component of the brain dystrophin protein complex (DPC) (Benson et al. 2001), Straub et al. (2002b) speculated that dysbindin protein's involvement in the development of schizophrenia may be mediated by DPC via three possible pathways (reviewed by Straub et al. (2002b)). In addition, Talbot et al. (2004) identified a new presynaptic signaling pathway that is not mediated by DPC. Specifically, the dysbindin protein is also located in presynaptic glutamatergic neurons, independent of DPC. Presynaptic dysbindin reductions are frequent in schizophrenia and are related to glutamatergic alterations in intrinsic hippocampal formation connections. Such changes may contribute to the cognitive deficits common in schizophrenia. Finally, there has also been recent speculation that the mechanism involves a phosphatidylinositol 3-kinase - Akt (PI3-kinase-Akt) signaling pathway (Weickert et al. 2004; Numakawa et al. 2004; McClintock et al. 2003; Talbot et al. 2004; Emamian et al. 2004; Numakawa et al. 2004). Further research is needed to specify the mechanism(s) by which *DTNBPI* contributes to the risk of schizophrenia.

Acknowledgments

This work was supported in part NIH grants R01-DA12849, R01-DA12690, K24-DA15105, R01-AA11330, P50-AA12870, K08-AA13732, K24-AA13736, R01-AA016015, K02-MH01387, and M01-RR06192 (University of Connecticut General Clinical Research Center), by funds from the U.S. Department of Veterans Affairs (the VA Medical Research Program, and the VA Connecticut Massachusetts Mental Illness Research, Education and Clinical Center [MIRECC], and the VA Research Enhancement Award Program [REAP] research center), and Alcoholic Beverage Medical Research Foundation (ABMRF) grant award R06932 (X Luo). Dr. Karl Hager provided valuable assistance with generating genotypes. The helpful comments of the two anonymous reviewers are highly appreciated.

References

- Aleman A, Kahn RS, Selten JP. Sex differences in the risk of schizophrenia: evidence from meta-analysis. *Arch Gen Psychiatry*. 2003; 60:565–571. [PubMed: 12796219]
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders, third edition, revised. Washington, DC: American Psychiatric Press; 1987.
- Benson MA, Newey SE, Martin-Rendon E, Hawkes R, Blake DJ. Dysbindin, a novel coiled-coil-containing protein that interacts with the dystrobrevins in muscle and brain. *J Biol Chem*. 2001; 276:24232–24241. [PubMed: 11316798]

- Blouin AG, Perez EL, Blouin JH. Computerized administration of the diagnostic interview schedule. *Psychiatry Res.* 1988; 23:335–344. [PubMed: 3387505]
- Bray NJ, Buckland PR, Owen MJ, O'Donovan MC. Cis-acting variation in the expression of a high proportion of genes in human brain. *Hum Genet.* 2003; 113:149–153. [PubMed: 12728311]
- Bray NJ, Preece A, Williams NM, Moskvina V, Buckland PR, Owen MJ, O'donovan MC. Haplotypes at the dystrobrevin binding protein 1 (DTNBP1) gene locus mediate risk for schizophrenia through reduced DTNBP1 expression. *Hum Mol Genet.* 2005; 14:1947–1954. [PubMed: 15917270]
- Collins-Schramm HE, Chima B, Morii T, Wah K, Figueroa Y, Criswell LA, Hanson RL, Knowler WC, Silva G, Belmont JW, Seldin MF. Mexican American ancestry-informative markers: examination of population structure and marker characteristics in European Americans, Mexican Americans, Amerindians and Asians. *Hum Genet.* 2004; 114:263–271. [PubMed: 14628215]
- Crowley JJ, Oslin DW, Patkar AA, Gottheil E, DeMaria PA Jr, O'Brien CP, Berrettini WH, Grice DE. A genetic association study of the mu opioid receptor and severe opioid dependence. *Psychiatr Genet.* 2003; 13:169–173. [PubMed: 12960749]
- Devlin B, Risch N. A Comparison of Linkage Disequilibrium Measures for Fine-Scale Mapping. *Genomics.* 1995; 29:311–322. [PubMed: 8666377]
- Emamian ES, Hall D, Birnbaum MJ, Karayiorgou M, Gogos JA. Convergent evidence for impaired AKT1-GSK3beta signaling in schizophrenia. *Nat Genet.* 2004; 36:131–137. [PubMed: 14745448]
- Feder JN, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy DA, Basava A, Dormishian F, Domingo R Jr, Ellis MC, Fullan A, Hinton LM, Jones NL, Kimmel BE, Kronmal GS, Lauer P, Lee VK, Loeb DB, Mapa FA, McClelland E, Meyer NC, Mintier GA, Moeller N, Moore T, Morikang E, Wolff RK, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet.* 1996; 13:399–408. [PubMed: 8696333]
- Franzek E, Beckmann H. Sex differences and distinct subgroups in schizophrenia. A study of 54 chronic hospitalized schizophrenics. *Psychopathology.* 1992; 25:90–99. [PubMed: 1502296]
- Funke B, Finn CT, Plocik AM, Lake S, DeRosse P, Kane JM, Kucherlapati R, Malhotra AK. Association of the DTNBP1 locus with schizophrenia in a U.S. population. *Am J Hum Genet.* 2004; 75:891–898. [PubMed: 15362017]
- Hafner H, Maurer K, Loffler W, Riecher-Rossler A. The influence of age and sex on the onset and early course of schizophrenia. *Br J Psychiatry.* 1993; 162:80–86. [PubMed: 8425144]
- Hao K, Xu X, Laird N, Wang X, Xu X. Power estimation of multiple SNP association test of case-control study and application. *Genet Epidemiol.* 2004; 26:22–30. [PubMed: 14691954]
- Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, McKeigue PM. Control of confounding of genetic associations in stratified populations. *Am J Hum Genet.* 2003; 72:1492–1504. [PubMed: 12817591]
- Hoh J, Wille A, Ott J. Trimming, weighting, and grouping SNPs in human case-control association studies. *Genome Res.* 2001; 11:2115–2119. [PubMed: 11731502]
- Jiang R, Dong J, Wang D, Sun FZ. Fine-scale mapping using Hardy-Weinberg disequilibrium. *Ann Hum Genet.* 2001; 65:207–219. [PubMed: 11427179]
- Kaufman J, Yang BZ, Douglas-Palumberi H, Houshyar S, Lipschitz D, Krystal JH, Gelernter J. Social supports and serotonin transporter gene moderate depression in maltreated children. *Proc Natl Acad Sci USA.* 2004; 101:17316–17321. [PubMed: 15563601]
- Kendler KS, O'Neill FA, Burke J, Murphy B, Duke F, Straub RE, et al. Irish study on high-density schizophrenia families: field methods and power to detect linkage. *Am J Med Genet.* 1996; 67:179–190. [PubMed: 8723045]
- Kitamura T, Fujihara S, Yuzuriha T, Nakagawa Y. Sex differences in schizophrenia: a demographic, symptomatic, life history and genetic study. *Jpn J Psychiatry Neurol.* 1993; 47:819–824. [PubMed: 8201792]
- Kirov G, Ivanov D, Williams NM, Preece A, Nikolov I, Milev R, Koleva S, Dimitrova A, Toncheva D, O'Donovan MC, Owen MJ. Strong evidence for association between the dystrobrevin binding protein 1 gene (DTNBP1) and schizophrenia in 488 parent-offspring trios from Bulgaria. *Biol Psychiatry.* 2004; 55:971–975. [PubMed: 15121479]
- Kohn Y, Danilovich E, Filon D, Oppenheim A, Karni O, Kanyas K, Turetsky N, Korner M, Lerer B. Linkage Disequilibrium in the DTNBP1 (Dysbindin) Gene Region and on Chromosome 1p36

- Among Psychotic Patients From a Genetic Isolate in Israel: Findings From Identity by Descent Haplotype Sharing Analysis. *Am J Med Genet B (Neuropsychiatric Genetics)*. 2004; 128B:65–70.
- Lee WC. Searching for disease-susceptibility loci by testing for Hardy-Weinberg Disequilibrium in a gene bank of affected individuals. *Am J Epidemiol*. 2003; 158:397–400. [PubMed: 12936892]
- Leung A, Chue P. Sex differences in schizophrenia, a review of the literature. *Acta Psychiatr Scand Suppl*. 2000; 401:3–38. Review. [PubMed: 10887978]
- Levin ML, Bertell R. Re: Simple estimation of population attributable risk from case-control studies. *Am J Epidemiol*. 1978; 108:78–79. [PubMed: 685980]
- Levinson DF, Wildenauer DB, Schwab SG, Albus M, Hallmayer J, Lerer B, et al. Additional support for schizophrenia linkage on chromosomes 6 and 8: a multicenter study. *Am J Med Genet B (Neuropsychiatr Genet)*. 1996; 67B:580–594.
- Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I, Williams NM, Schwab SG, Pulver AE, Faraone SV, et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet*. 2003; 73:34–48. [PubMed: 12802786]
- Liao HM, Chen CH. Mutation analysis of the human dystrobrevin-binding protein 1 gene in schizophrenic patients. *Schizophr Res*. 2004; 71:185–189. [PubMed: 15374586]
- Lindholm E, Ekholm B, Balciuniene J, Johansson G, Castensson A, Koisti M, et al. Linkage analysis of a large Swedish kindred provides further support for a susceptibility locus for schizophrenia on chromosome 6p23. *Am J Med Genet*. 1999; 88:369–377. [PubMed: 10402504]
- Liu K, Muse S. PowerMarker: new genetic data analysis software. Version 3.0. Accessed February 1, 2007 Free program distributed by the author over the internet from <http://www.powermarker.net>
- Luo X, Kranzler HR, Zhao H, Gelernter J. Haplotypes at the OPRM1 locus are associated with susceptibility to substance dependence in European-Americans. *Am J Med Genet B (Neuropsychiatr Genet)*. 2003; 120B:97–108. [PubMed: 12815747]
- Luo X, Kranzler HR, Zuo L, Yang BZ, Lappalainen J, Gelernter J. ADH4 gene variation is associated with alcohol dependence and drug dependence in European Americans: results from family-controlled and population-structured association studies. *Pharmacogenet Genom*. 2005c; 15:755–768.
- Luo X, Kranzler HR, Zuo L, Wang S, Blumberg HP, Gelernter J. CHRM2 gene predisposes to alcohol, drug dependence and affective disorder: results from an extended population structured association study. *Hum Mol Genet*. 2005b; 14:2421–2434. [PubMed: 16000316]
- Luo X, Gelernter J, Zhao H, Kranzler HR. Response to Dr. Kopke's comments on haplotypes at the OPRM1 locus. *Am J Med Genet B (Neuropsychiatr Genet)*. 2005a; 135B:102. [PubMed: 15806579]
- Luo X, Kranzler HR, Zuo L, Wang S, Lappalainen J, Schork NJ, Gelernter J. Diplotype trend regression (DTR) analysis of the ADH gene cluster and ALDH2 gene: Multiple significant associations for alcohol dependence. *Am J Hum Genet*. 2006; 78:973–987. [PubMed: 16685648]
- Maziade M, Bissonnette L, Rouillard E, Martinez M, Turgeon M, Charron L, et al. 6p24–22 region and major psychoses in the Eastern Quebec population. *Le Groupe IREP. Am J Med Genet*. 1997; 74:311–318. [PubMed: 9184316]
- McClintock, W.; Shannon Weickert, C.; Halim, ND.; Lipska, BK.; Hyde, TM.; Herman, MM.; Weinberger, DR.; Kleinman, JE.; Straub, RE. Program No. 317.9. Abstract Viewer/Itinerary Planner. Washington, DC: Society for Neuroscience (Online); 2003. Reduced expression of dysbindin protein in the dorsolateral prefrontal cortex of patients with schizophrenia.
- McGrath J, Saha S, Welham J, El Saadi O, MacCauley C, Chant D. A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Med*. 2004; 2:13. Review. [PubMed: 15115547]
- Morris DW, McGhee KA, Schwaiger S, Scully P, Quinn J, Meagher D, Waddington JL, Gill M, Corvin AP. No evidence for association of the dysbindin gene [DTNBP1] with schizophrenia in an Irish population-based study. *Schizophr Res*. 2003; 60:167–172. [PubMed: 12591580]
- Nielsen DM, Ehm MG, Weir BS. Detecting marker-disease association by testing for Hardy-Weinberg Disequilibrium at a marker locus. *Am J Hum Genet*. 1999; 63:1531–1540. [PubMed: 9867708]
- Niu T, Qin ZD, Xu X, Liu JS. Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *Am J Hum Genet*. 2002; 70:157–169. [PubMed: 11741196]

- Numakawa T, Yagasaki Y, Ishimoto T, Okada T, Suzuki T, Iwata N, Ozaki N, Taguchi T, Tatsumi M, Kamijima K, Straub RE, Weinberger DR, Kunugi H, Hashimoto R. Evidence of novel neuronal functions of dysbindin, a susceptibility gene for schizophrenia. *Hum Mol Genet.* 2004; 13:2699–2708. [PubMed: 15345706]
- Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet.* 2004; 74:765–769. [PubMed: 14997420]
- Owen, MJ.; O'Donovan, MC.; Gottesman, II. Schizophrenia. In: McGuffin, P.; Owen, MJ.; Gottesman, II., editors. *Psychiatric genetics and genomics.* Oxford University Press; New York, New York, USA: 2002. p. 247-266.
- Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD. Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet.* 1998; 63:1839–1851. [PubMed: 9837836]
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000a; 155:945–959. [PubMed: 10835412]
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. *Am J Hum Genet.* 2000b; 67:170–181. [PubMed: 10827107]
- Raybould R, Green EK, MacGregor S, Gordon-Smith K, Heron J, Hyde S, Caesar S, Nikolov I, Williams N, Jones L, O'Donovan MC, Owen MJ, Jones I, Kirov G, Craddock N. Bipolar disorder and polymorphisms in the dysbindin gene (DTNBP1). *Biol Psychiatry.* 2005; 57:696–701. [PubMed: 15820225]
- Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science.* 1996; 273:1516–1517. [PubMed: 8801636]
- Risch N, Teng J. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases. I. DNA pooling. *Genome Res.* 1998; 12:1273–1288. [PubMed: 9872982]
- Risch N. Searching for genetic determinants in the new millennium. *Nature.* 2000; 405:847–856. [PubMed: 10866211]
- Rosenheck R, Cramer J, Xu W, Thomas J, Henderson W, Frisman L, et al. A comparison of clozapine and haloperidol in hospitalized patients with refractory schizophrenia. *N Engl J Med.* 1997; 337:809–815. [PubMed: 9295240]
- Schwab SG, Albus M, Hallmayer J, Borrmann M, Lichtermann D, Ebstein RP, Ackenheil M, Lerer B, Risch N, Maier W, Wildenauer DB. Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. *Nat Genet.* 1995; 11:325–327. [PubMed: 7581458]
- Schwab SG, Knapp M, Mondabon S, Hallmayer J, Borrmann-Hassenbach M, Albus M, Lerer B, Rietschel M, Trixler M, Maier W, Wildenauer DB. Support for association of schizophrenia with genetic variation in the 6p22.3 gene, dysbindin, in sib-pair families with linkage and in an additional sample of triad families. *Am J Hum Genet.* 2003; 72:185–190. [PubMed: 12474144]
- Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, Pfaff C, Jones C, Massac A, Cameron N, Baron A, Jackson T, Argyropoulos G, Jin L, Hoggart CJ, McKeigue PM, Kittles RA. Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet.* 2003; 112:387–399. [PubMed: 12579416]
- Sham PC, Gottesman II, MacLean CJ, Kendler KS. Schizophrenia: sex and familial morbidity. *Psychiatry Res.* 1994; 52:125–134. [PubMed: 7972569]
- Spitzer, RL.; Endicott, J. Schedule for affective disorders and schizophrenia: Lifetime version. New York: New York Biometrics Research Division, New York State Psychiatric Institute; 1975.
- Spitzer RL, Williams JBW, Gibbon M, First MB. The structured clinical interview for DSM-III-R (SCID), I: History, rationale, and description. *Arch Gen Psychiatry.* 1992; 49:624–629. [PubMed: 1637252]
- Stein MB, Schork MJ, Gelernter J. A polymorphism of the β 1-adrenergic receptor is associated with shyness and low extraversion. *Biol Psychiatry.* 2004; 56:217–224. [PubMed: 15312808]
- Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet.* 2003; 73:1162–1169. [PubMed: 14574645]

- Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet.* 2001; 68:978–989. [PubMed: 11254454]
- Straub RE, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, Webb BT, Zhang J, Walsh D, Kendler KS. A potential vulnerability locus for schizophrenia on chromosome 6p24–22: evidence for genetic heterogeneity. *Nat Genet.* 1995; 11:287–293. [PubMed: 7581452]
- Straub RE, MacLean CJ, Ma Y, Webb BT, Myakishev MV, Harris-Kerr C, Wormley B, Sadek H, Kadambi B, O'Neill FA, Walsh D, Kendler KS. Genome-wide scans of three independent sets of 90 Irish multiplex schizophrenia families and follow-up of selected regions in all families provides evidence for multiple susceptibility genes. *Mol Psychiatry.* 2002a; 7:542–559. [PubMed: 12140777]
- Straub RE, Jiang Y, MacLean CJ, Ma Y, Webb BT, Myakishev MV, Harris-Kerr C, Wormley B, Sadek H, Kadambi B, Cesare AJ, Gibberman A, Wang X, O'Neill FA, Walsh D, Kendler KS. Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. *Am J Hum Genet.* 2002b; 71:337–348. [PubMed: 12098102]
- Sullivan PF. The genetics of schizophrenia. *PLoS Med.* 2005; 2(7):e212. [PubMed: 16033310]
- Talbot K, Eidem WL, Tinsley CL, Benson MA, Thompson EW, Smith RJ, Hahn CG, Siegel SJ, Trojanowski JQ, Gur RE, Blake DJ, Arnold SE. Dysbindin-1 is reduced in intrinsic, glutamatergic terminals of the hippocampal formation in schizophrenia. *J Clin Invest.* 2004; 113:1353–1363. [PubMed: 15124027]
- Tang JX, Zhou J, Fan JB, Li XW, Shi YY, Gu NF, Feng GY, Xing YL, Shi JG, He L. Family-based association study of DTNBP1 in 6p22.3 and schizophrenia. *Mol Psychiatry.* 2003; 8:717–718. [PubMed: 12888799]
- Turecki G, Rouleau GA, Joobor R, Mari J, Morgan K. Schizophrenia and chromosome 6p. *Am J Med Genet B (Neuropsychiatr Genet).* 1997; 74B:195–198.
- Van Den Bogaert A, Schumacher J, Schulze TG, Otte AC, Ohlraun S, Kovalenko S, Becker T, Freudenberg J, Jonsson EG, Mattila-Evenden M, Sedvall GC, Czerski PM, Kapelski P, Hauser J, Maier W, Rietschel M, Propping P, Nothen MM, Cichon S. The DTNBP1 (dysbindin) gene contributes to schizophrenia, depending on family history of the disease. *Am J Hum Genet.* 2003; 73:1438–1443. [PubMed: 14618545]
- Van den Oord EJ. A comparison between different designs and tests to detect QTLs in association studies. *Behav Genet.* 1999; 29:245–256.
- Van den Oord EJ, Sullivan PF, Jiang Y, Walsh D, O'Neill FA, Kendler KS, Riley BP. Identification of a high-risk haplotype for the dystrobrevin binding protein 1 (DTNBP1) gene in the Irish study of high-density schizophrenia families. *Mol Psychiatry.* 2003; 8:499–510. [PubMed: 12808430]
- Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science.* 2008; 320:539–543. [PubMed: 18369103]
- Weickert CS, Straub RE, McClintock BW, Matsumoto M, Hashimoto R, Hyde TM, Herman MM, Weinberger DR, Kleinman JE. Human dysbindin (DTNBP1) gene expression in normal brain and in schizophrenic prefrontal cortex and midbrain. *Arch Gen Psychiatry.* 2004; 61:544–555. [PubMed: 15184234]
- Williams NM, Preece A, Spurlock G, Norton N, Williams HJ, Zammi S, O'Donovan MC, Owen MJ. Support for genetic variation in neuregulin 1 and susceptibility to schizophrenia. *Mol Psychiatry.* 2003; 8:485–487. [PubMed: 12808428]
- Williams NM, Preece A, Morris DW, Spurlock G, Bray NJ, Stephens M, Norton N, Williams H, Clement M, Dwyer S, Curran C, Wilkinson J, Moskvina V, Waddington JL, Gill M, Corvin AP, Zammit S, Kirov G, Owen MJ, O'Donovan MC. Identification in 2 independent samples of a novel schizophrenia risk haplotype of the dystrobrevin binding protein gene (DTNBP1). *Arch Gen Psychiatry.* 2004; 61:336–344. [PubMed: 15066891]
- Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational Inferences about Departures from Hardy-Weinberg Equilibrium. *Am J Hum Genet.* 2005; 76:967–986. [PubMed: 15834813]

- Yang BZ, Zhao H, Kranzler HR, Gelernter J. Practical population group assignment with selected informative markers: characteristics and properties of Bayesian clustering via STRUCTURE. *Genet Epidemiol.* 2005; 28:302–312. [PubMed: 15782414]
- Zaykin DV, Westfall PH, Young SS, Karnoub MC, Wagner MJ, Ehm MG. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. *Hum Hered.* 2002; 53:79–91. [PubMed: 12037407]
- Zill P, Baghai TC, Engel R, Zwanzger P, Schule C, Eser D, Behrens S, Rupprecht R, Moller HJ, Ackenheil M, Bondy B. The dysbindin gene in major depression: an association study. *Am J Med Genet B (Neuropsychiatr Genet).* 2004; 129:55–58. [PubMed: 15274041]

Table 1
(Supplemental) Summary of the designs and methods used in previous association studies of *DTNBP1* with schizophrenia

Design	Sample	Size	Ethnic	Phenotype	Genotyping	Program	Reference
Family-based	pedigrees	270	Irish	S	FP-TDI	Simwalk (Simwalk, TRANSMIT, FBAT)	Straub et al., 2002b
	pedigrees	268	Irish (German, Israeli, Arab, Hungary)	S	FP-TDI		van den Oord et al. 2003
	sid-pairs, trios	78,125		S	FP-TDI	MLE	Schwab et al., 2003
	trios	233	Chinese	S	TaqMan (FP-TDI, Amplifluor)	ETDT, TRANSMIT	Tang et al., 2003
	trios	488	Bulgarian	S	SNaPshot	TDTPHASE	Kirov et al., 2004
		219,231 (65FH+)	Irish	S		–	Morris et al. 2003
		418,285	German	S			
		294,113	Polish	S			
		142,272	Swedish	S	Pyro-sequencing		Van Den Bogaert et al., 2003
		32FH+	Swedish	S			
Cases, controls		708,711	Welsh	S	(FP-TDI, SNaPshot or Sequencing)		Williams et al., 2004
		219,231	Irish	S		EHPPLUS	
		293,220	German	MD	SNaPshot	–	Zill et al., 2004
		670,588	Japanese	S	TaqMan	–	Numakawa et al. 2004
			EAs			–	
		524,573	Hispanic	S or SAD			
			EAs		MALDI-ToF	–	Funke et al. 2004
			AAs			–	
		50+94	Chinese	S	Sequencing	–	Liao and Chen, 2004
		726,1407	Welsh	PBD	Amplifluor	UNPHASED	Raybould et al. 2005
Cases, controls	Cases	31	Caucasian	mRNA			
		708,711	Welsh	S	SNaPshot	–	Bray et al. 2005
		30,711	Caucasian	mRNA		–	

FH+, positive family history; S, schizophrenia; MD, major depression; SAD, schizoaffective disorder; PBD, psychotic bipolar disorder; mRNA, DTNBP1 mRNA expression in brain; FP-TDI, fluorescence-polarization template-directed incorporation (FP-TDI) technique; TaqMan, a fluorogenic 5' nuclease assay method; MALDI-ToF, Matrix-Assisted Laser Desorption Ionization - Time of Flight (MALDI-ToF) Mass Spectrometry; Program, the haplotype-reconstruction program; MLE, maximum-likelihood estimate. "–" denotes unknown.

Table 2

(Supplemental) Risk or protective alleles and haplotypes in previous association studies

Population	Markers ¹														Haplotype			Reference																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		18	19	20	21	22	23	24	25	26	27	28	29	30	31	Frequency	Effect	
Irish					G		C	C	T	A	A	A	T			G			G			G	G	C			T					rare	risk	Straub et al., 2002b	
						G	C	C	T	A	A	A																					risk		
(German, Israeli, Arab, Hungary)									A	C	T	A	A	C																			0.060	risk	van den Oord et al. 2003
										A	C	G	C																				risk		
Chinese				A			C	A	C	C	G	C	C										G										most common	risk	Schwab et al., 2003
							C	A	C	C	G	C																					most common	risk	
Bulgarian											G																						0.115	risk	Kirov et al., 2004
Irish																																		risk	Morris et al. 2003
German																																		risk	
Polish																																		risk	
Swedish																																		risk	
Swedish (FH+)																																		risk	Van Den Bogaert et al., 2003
	A		C																														0.178	risk	
Welsh	A																																	risk	Williams et al., 2004
	A																																0.290	protective	
	A																																0.100	risk	
	T																																0.210	risk	
	T																																0.000	protective	
Irish	A																																0.300	protective	Williams et al., 2004
	T																																0.210	risk	
	T																																0.000	protective	
German																																		risk	Zill et al., 2004
Japanese	T																																	risk	Numakawa et al. 2004
	A																																0.026	risk	
EAs																																		risk	
Hispanic																																		risk	Funke et al. 2004
																																		risk	

Table 3

Marker information

No.	rs number	Alias	Base	Chromosome Position	Map Position
1	rs2619538	P4211 ¹ , A	T/A	15773188	Promoter
2	rs12204704	B	C/T	15773184	Promoter
3	rs2743852	C	C/G	15772743	Promoter
4	rs2619537	D	C/T	15772392	Promoter
5	rs909706	P1583,E	C/T	15768850	intron1
6	rs1997679	P1795	C/T	15766884	intron1
7	rs1474605	P1792	A/G	15766191	intron1
8	rs1018381	P1578	C/T	15765049	intron1
9	rs2619522	P1763	A/C	15761628	intron1
10	rs760761	P1320	C/T	15759111	intron3
11	rs2005976	P1757	G/A	15758781	intron3
<u>12</u>	rs2619528	P1765	G/A	15757808	intron3
13	rs2619550	F'	C/G	15742221	intron4
14	rs1011313	P1325	C/T	15741411	intron4
15	rs2619542	E'	C/T	15737202	intron4
16	rs13198195	F	T/C	15736802	intron4
17	rs13198335	G	A/G	15736727	intron4
18	N/A	H	C/T	15736141	intron4
19	rs3213207	P1635	A/G	15736081	intron4
20	rs12527496	I	C/T	15736060	intron4
21	rs12525702	J	G/A	15735750	intron4
22	rs16876738	K	G/C	15735532	intron5
23	rs2619539	P1655	C/G	15728834	intron5
24	rs3829893	P3170 ¹ , L	C/T	15723616	intron5
25	rs760666	P1287	C/T	15697100	intron7
26	rs12524251	M	C/T	15694111	intron7
27	rs742105	P1333	C/T	15681053	intron7
28	N/A	N	G/-	15658414	intron7

No.	rs number	Alias	Base	Chromosome Position	Map Position
29	N/A	O	A/G	15632897	intron8
30	rs742106	P1328	C/T	15632459	intron9
31	rs1047631	P3230 ^{A,P}	A/G	15631080	3'UTR

The markers in different haplotype blocks are separated by “_” in the first column. Base, major allele/minor allele;

¹Weickert et al. 2004; A-P, markers names by Williams et al. 2004; A', E', and F', markers named by Morris et al. 2003. The six bolded markers were genotyped in the present study. The marker numbers are consistent with Figure 1 and Table 2. N/A, not applicable due to unknown.

Table 4

Genotype and allele frequency distributions

	European-Americans						African-Americans					
	Cases (n=235)			Controls (n=298)			Cases (n=82)			Controls (n=48)		
	n	f		n	f		n	f		n	f	
TT	29	0.148		30	0.102		8	0.136		4	0.087	
TC	84	0.429		135	0.458		11	0.186 ^{H1}		11	0.239	
CC	83	0.423		130	0.441		40	0.678		31	0.674	
T	142	0.362 ^{S1}		195	0.331		27	0.229		19	0.207	
C	250	0.638		395	0.669		91	0.771		73	0.793	
TT	29	0.188		62	0.213		5	0.125		6	0.128	
TC	82	0.532 ^{H1}		140	0.481		16	0.400		10	0.213 ^{H4}	
CC	43	0.279		89	0.306		19	0.475		31	0.660	
T	140	0.455		264	0.454		26	0.325 ^{S2}		22	0.234	
C	168	0.545		318	0.546		54	0.675		72	0.766	
CC	31	0.197		58	0.197		4	0.093		6	0.125	
CG	78	0.497		140	0.475		19	0.442		11	0.229 ^{H5}	
GG	48	0.306		97	0.329		20	0.465		31	0.646	
C	140	0.446		256	0.434		27	0.314		23	0.240	
G	174	0.554		334	0.566		59	0.686		73	0.760	
TT	12	0.069		22	0.075		12	0.255		10	0.213	
TC	55	0.314		80	0.274 ^{H2}		18	0.383 ^{S2}		21	0.447	
CC	108	0.617		190	0.651		17	0.362		16	0.340	
T	79	0.226		124	0.212		42	0.447		41	0.436	
C	271	0.774		460	0.788		52	0.553		53	0.564	
TT	0	0.000		5	0.017		7	0.092		3	0.063	
TC	44	0.201 ^{A1}		38	0.131 ^{H3}		35	0.461		25	0.521	
CC	175	0.799		247	0.852		34	0.447		20	0.417	
T	44	0.100		48	0.083		49	0.322		31	0.323	
C	394	0.900		532	0.917		103	0.678		65	0.677	
AA	27	0.130		36	0.125		2	0.027		3	0.063	

	European-Americans				African-Americans			
	Cases (n=235)		Controls (n=298)		Cases (n=82)		Controls (n=48)	
	n	f	n	f	n	f	n	f
AG	92	0.442	128	0.443	32	0.427 ^{A2}	11	0.229
GG	89	0.428	125	0.433	41	0.547	34	0.708
A	146	0.351	200	0.346	36	0.240	17	0.177
G	270	0.649	378	0.654	114	0.760	79	0.823

n, individual number (for genotypes) or chromosome number (for alleles); f, frequency; $p < 0.05$ for χ^2 Case-control frequency comparison and $H_{Hardy-Weinberg}$ Disequilibrium test;

δ The marker with highest δ among all markers ($\delta_1 = 0.047$ for P1328 in EAs; $\delta_2 = 0.119$ for P1333 in AAs);

J The marker with highest J value among all markers ($J_1 = 0.062$ for P1333 in EAs; $J_2 = 0.182$ for P1320 in AAs).

Table 5

(Supplemental) Haplotype and diplotype frequency distributions

European-Americans		African-Americans	
Haplotype and Diplotype	f	Haplotype and Diplotype	f
GCCGCC	0.370	GCCGCC	0.291
ACCCTT	0.228	GTTGCC	0.286
ACCCTC	0.089	ACCCTC	0.108
GCTCTC	0.086	ACCCTT	0.102
GTTGCC	0.078	GCTGCC	0.049
GCCGCT	0.068	GCCGCT	0.042
GCTCTT	0.026	GCTCTC	0.034
ACCGTC	0.017	GTTGCT	0.033
GCTGCC	0.011	GCTCTT	0.017
ACCCTT/GCCGCC	0.161	GCCGCC/GCCGCC	0.157
GCCGCC/GCCGCC	0.153	GTTGCC/GCCGCC	0.146
ACCCTC/GCCGCC	0.071	GTTGCC/GCTGCC	0.083
GCCGCT/GCCGCC	0.068	ACCCTC/GTTGCC	0.074
GCTCTC/GCCGCC	0.065	GTTGCC/GTTGCC	0.073
ACCCTT/ACCCTT	0.055	ACCCTT/GTTGCC	0.063
ACCCTT/GTTGCC	0.045	ACCCTC/GCCGCC	0.044
ACCCTT/ACCCTC	0.042	ACCCTT/GCCGCC	0.031
ACCCTT/GCTCTC	0.039	ACCCTT/GCCGCT	0.028
GTTGCC/GCCGCC	0.031	GCTCTC/GTTGCC	0.027
ACCCTT/GCCGCT	0.026	ACCCTT/ACCCTC	0.024
GCTCTC/GTTGCC	0.020	ACCCTC/ACCCTC	0.022
ACCGTC/GCCGCC	0.015	GCCGCT/GCCGCC	0.012
GCTCTT/ACCCTT	0.015	GCTCTC/ACCCTC	0.010
ACCCTC/GTTGCC	0.014		
GCTCTT/GCCGCC	0.014		
GCCGCT/ACCCTC	0.012		
GTTGCC/GCTGCC	0.011		
GCTCTC/ACCCTC	0.010		

The haplotypes were constructed by the order from 5' to 3': P1583-P1578-P1320-P1655-P1333-P1328. The disease-associated haplotype and diplotypes are presented in bold.

Table 6
Logistic regression analysis on the relationship between schizophrenia and *DTNBP1*

Models	Variates	Combined sample		European-American		African-American	
		β	p	β	p	β	p
Diploypewise	Female	-5.020	2.1×10^{-12}	-5.322	6.4×10^{-8}	-4.335	3.8×10^{-5}
	Ancestry	0.649	0.031				
	ACCCTT/GCCGCC	-0.609	0.040				
	Female \times ACCCTT/GCCGCC	2.408	0.025				
Haploypewise	Female \times GCCGCC/GCCGCC			2.499	0.043		
	Female	-5.302	7.5×10^{-11}	-5.806	1.3×10^{-6}	-4.335	3.8×10^{-5}
	Ancestry	0.623	0.039				
	GCCGCC \times GCCGCC	-0.679	0.034				
Genotypewise	Female \times GCCGCC \times GCCGCC	2.646	0.030	3.016	0.040		
	Female	-4.416	4.6×10^{-25}	-4.311	2.3×10^{-20}		
	Ancestry	0.788	0.004				
	Female	-4.416	2.0×10^{-48}	-4.311	4.3×10^{-39}		
Allelewise	Ancestry	0.788	5.6×10^{-5}				

Combined sample, all subjects including European-Americans and African-Americans; Ancestry, African ancestry; β , regression coefficient; p, p-value.