Follow-up of a Major Psychosis Linkage Site in 13q13-q14 Reveals Significant Association in Both Case-Control and Family Samples

Supplemental Information

Supplementary Methods

Diagnoses and Phenotype Description

A lifetime best-estimate DSM-IV diagnosis was made for the unrelated schizophrenia (SZ) cases and the kindred members using personal interview, information from relatives and extensive medical records. We applied a stringent diagnosis procedure outlined in previous reports (1, 2). In brief, all available information across lifetime from different sources (all medical records, family informant interviews, personal structured interview) was reviewed blindly by four research diagnosticians. The board of diagnosticians also specified the presence or absence of psychotic features in bipolar disorder (BP) patients according to DSM-IV.

Family history of mental disorders of the SZ unrelated cases was also obtained from the same sources. In the kindred sample, we used as phenotype a narrow SZ definition restricted to SZ and a broad definition comprising SZ narrow plus schizophreniform disorder and schizotypal personality. The BP narrow phenotype was restricted to BP I and the broad definition included BP I, BP II, and recurrent major depression. We also defined a narrow and broad "common locus" (CL) phenotype. The narrow CL phenotype included BP narrow, SZ narrow and schizoaffective disorder (SAD). The broad CL definition included the broad definitions of BP and SZ, in addition to SAD. The number of affected subjects for each phenotype definition is reported in Table 1. The 467 genotyped kindred members satisfying the criteria stated in the Methods section were treated as unaffected subjects in the association analyses.

Genotyping

In the case-control sample, single nucleotide polymorphisms (SNPs) were analyzed using the minisequencing approach of the Illumina genotyping platform with a customized array including the HumanHap300 BeadChip and 57,000 additional SNPs, for a total of 375,174 SNPs. All subjects had a genotype called for at least 97 percent of the SNPs. Our region of interest derived from the linkage evidence in our kindred sample contained 2,150 SNPs. For replication in the kindred sample, SNPs were analyzed using an in house minisequencing approach (3) adapted for the LiCor sequencers, where genotypes are called automatically using the software SAGA (LICOR). A melting temperature procedure with a cold oligonucleotide probe specific to one of the two nucleotides of the SNP was also used with High Resolution Melting kit and a real time PCR (480 LightCycler), both from Roche. Mendelian inheritance was checked using the computer software PedCheck (4), and 10% blind replicates were included for genotyping quality control.

Copy Number Variant (CNV) Detection in the Case-Control Sample

CNVs were inferred using the hidden Markov model implemented in PennCNV (5), which uses the log R ratio and B allele frequency produced by the Illumina BeadStudio software to infer hidden states corresponding to copy number. Population frequencies of the B allele were estimated from our sample and other model parameters were set to the values estimated by Wang *et al.* (5).

Genotyping Quality Criteria in the Case-Control Sample

Only SNPs with a minimum call rate of 98 percent and minor allele frequency above one percent in the combined case-control sample were retained. We discarded SNPs with Hardy-Weinberg equilibrium chi-square test *p*-value less than 2.5×10^{-5}, corresponding to a 0.05 significance level divided by the number of SNPs in the region. This left 2,081 SNPs to be included in the analysis. Only 2 discordant genotypes were observed at these 2,081 SNPs among 30 subjects genotyped in duplicate, for a concordance rate of 99.997 percent.

Assessment of Population Substructure

A principal component (PC) analysis of the genotypes of the 339,228 autosomal SNPs was performed with Eigensoft version 3.0 (6) (genepath.med.harvard.edu/~reich/Software.htm) to investigate population structure in our case-control sample. Ancestry differences between the case and control samples were tested by comparing the two groups on the first ten PCs using a Tracy-Widom statistic implemented in Eigensoft (6). The software was also used to estimate the inflation factor λ based on the genomic control method (7).

Association Analysis in the Case-Control Sample

Allelic association with genotyped SNPs was tested using Fisher exact tests in the 2 x 2 table of alleles x case-control status, as well as Cochran-Armitage trend tests in the 3 x 2 table of genotype x case-control status. Fisher exact tests were also performed with genotype frequencies under the dominant and recessive models. We tested association with untyped variants in the region extending 200 kb on either side of the association signal detected using genotyped SNPs in an attempt to better characterize the association. We applied two complementary approaches:

imputation of genotypes at untyped SNPs and global tests of haplotypes (frequency > 1%) over short windows of SNPs (8). Genotype imputation was performed using genotype data from the 1000 Genomes Project (June 2011 data release, www.1000genomes.org) using IMPUTE version 2 (9) (mathgen.stats.ox.ac.uk/impute/impute.html). Association with the imputed SNP genotypes was then tested using a score test derived from the missing data likelihood under a logistic model implemented in SNPTEST version 2.2 (mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html). Association to both genotyped and imputed SNPs was also tested conditionally on genotypes and on allele counts of the genotyped SNPs showing the strongest association. We used the score test under a logistic model implemented in SNPTEST. Haplotype association tests were performed on sliding windows of three and five consecutive SNPs. Association to any haplotype formed by the SNPs in each window and to individual haplotypes was tested using score tests accounting for missing phase information and missing genotypes at a subset of markers using the haplo.score R function (10). Odds ratios (ORs) attached to haplotypes were estimated under the generalized linear model with missing phase information of Lake *et al.* (11) using the haplo.cc R function.

Association Analysis in the Family Sample

Allelic log-ORs were estimated under a logistic model estimated using generalized estimating equations (GEEs) with an independence working correlation structure between the subjects in the same family. Allelic association was tested using a Wald statistic for the log-OR where the variance was estimated using an empirical variance estimate robust to intra-familial correlation (12). The same approach was then applied to the combined case-control and family samples. The generalized disequilibrium test (13), a score test robust to population stratification,

was also applied to confirm the GEE Wald test results. For haplotypic association, the most likely haplotype pair for each subject given the genotype data on the entire kindred was inferred by maximum likelihood using Superlink (14) (cbl-fog.cs.technion.ac.il/superlink/) and haplotypes were recoded as alleles of a multi-allelic marker for the analysis. All statistical tests were two-sided.

Association Analysis in the Combined Case-Control and Family Samples

The logistic regression model included a term for the allele count of a SNP and a term for the sample of origin (case-control vs. family) to adjust for differences in allele frequency between the samples. The model was estimated using GEE as for the analysis in the family sample, with the subjects from the case-control sample treated as one-member families.

Evaluation of Linkage Disequilibrium (LD)

The squared correlation and the Lewontin's D' coefficient among pairs of genotyped SNPs were estimated in the control sample using standard algorithms implemented in the function LD of the R package *genetics*.

Table S1. Association of single nucleotide polymorphism alleles in the kindred sample with the narrow definition of schizophrenia and non-affected adult relatives older than 39 years.

	Alleles		MAF[a]			
Marker	**Minor**	**Major**	**Cases**	**NAARs[b]**	**OR[c] (95% CI)**	***P*[d]**
All families (119 cases, 378 NAARs)						
rs2120753	G	A	0.429	0.486	0.80 (0.58, 1.10)	0.17
rs1156026	T	C	0.538	0.456	1.43 (0.98, 2.08)	0.064
Schizophrenia families (64 cases, 101 NAARs)						
rs2120753	G	A	0.516	0.632	0.59 (0.38, 0.92)	0.019
rs1156026	T	C	0.523	0.381	2.04 (1.22, 3.39)	0.0063

[a] Minor allele frequency.
[b] Non-affected adult relatives older than 39 years.
[c] Generalized estimating equations odds ratio estimate and its 95% confidence interval calculated using the empirical variance estimate.
[d] *P*-value of Wald test calculated using the empirical variance estimate.

Table S2. Association of single nucleotide polymorphism alleles considering either all cases or early-onset cases only for the narrow definition of schizophrenia, bipolar disorder and the common locus phenotype in the kindred sample with non-affected adult relatives older than 39 years.

			MAF[a]						
	Alleles		Cases		NAARs[b]	All Cases vs NAARs		Early-Onset Cases vs NAARs	
Marker	**Minor**	**Major**	**All**	**Early-onset[e]**	**(*n* = 378)**	**OR[c] (95% CI)**	***P*[d]**	**OR[c] (95% CI)**	***P*[d]**
Schizophrenia (119 cases, 58 early-onset cases)									
rs2120753	G	A	0.429	0.377	0.486	0.80 (0.58, 1.10)	0.17	0.65 (0.42, 1.00)	0.049
rs1156026	T	C	0.538	0.638	0.456	1.43 (0.98, 2.08)	0.064	2.25 (1.40, 3.61)	7.5e-4
Bipolar disorder (117 cases, 40 early-onset cases)									
rs2120753	G	A	0.461	0.382	0.486	0.91 (0.64, 1.29)	0.604	0.67 (0.39, 1.15)	0.15
rs1156026	T	C	0.449	0.512	0.456	0.97 (0.69, 1.37)	0.855	1.26 (0.69, 2.29)	0.45
Common locus phenotype (273 cases, 118 early-onset cases)									
rs2120753	G	A	0.458	0.396	0.486	0.9 (0.72, 1.13)	0.361	0.71 (0.53, 0.95)	0.023
rs1156026	T	C	0.487	0.589	0.456	1.14 (0.88, 1.47)	0.326	1.74 (1.24, 2.44)	0.0014

[a] Minor allele frequency.
[b] Non-affected adult relatives older than 39 years.
[c] Generalized estimating equations odds ratio estimate and its 95% confidence interval calculated using the empirical variance estimate.
[d] P-value of Wald test calculated using the empirical variance estimate.
[e] Onset before age 26.

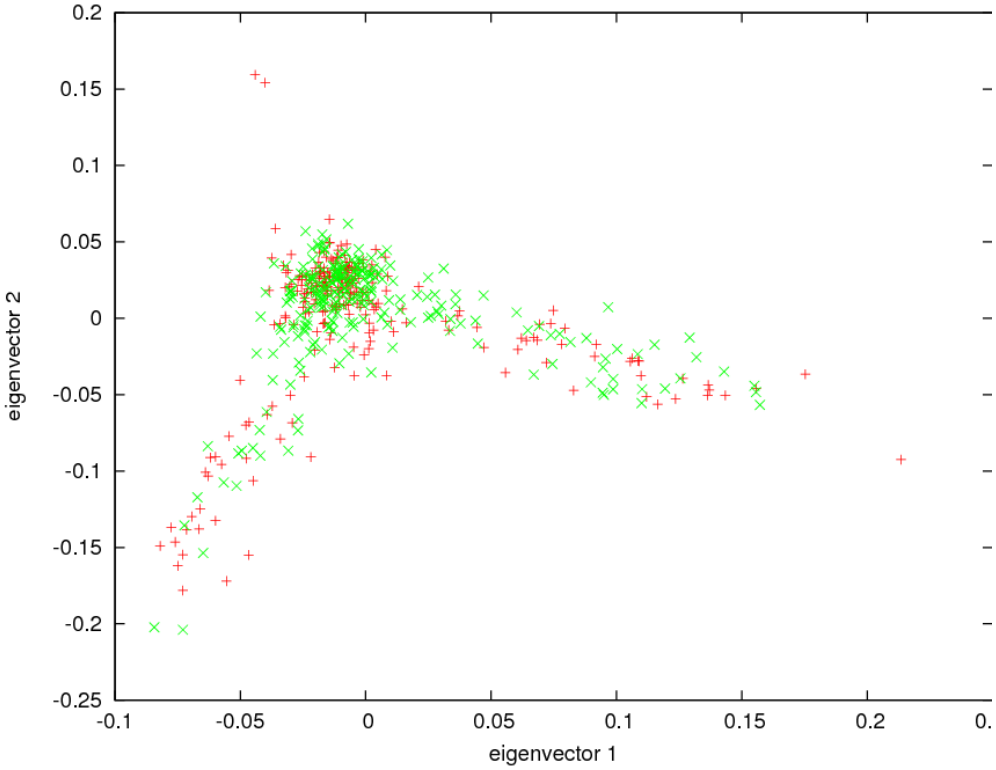

Figure S1. Projection of the cases (red +) and controls (green x) on the first two principal components defined from autosomal single nucleotide polymorphism genotypes.

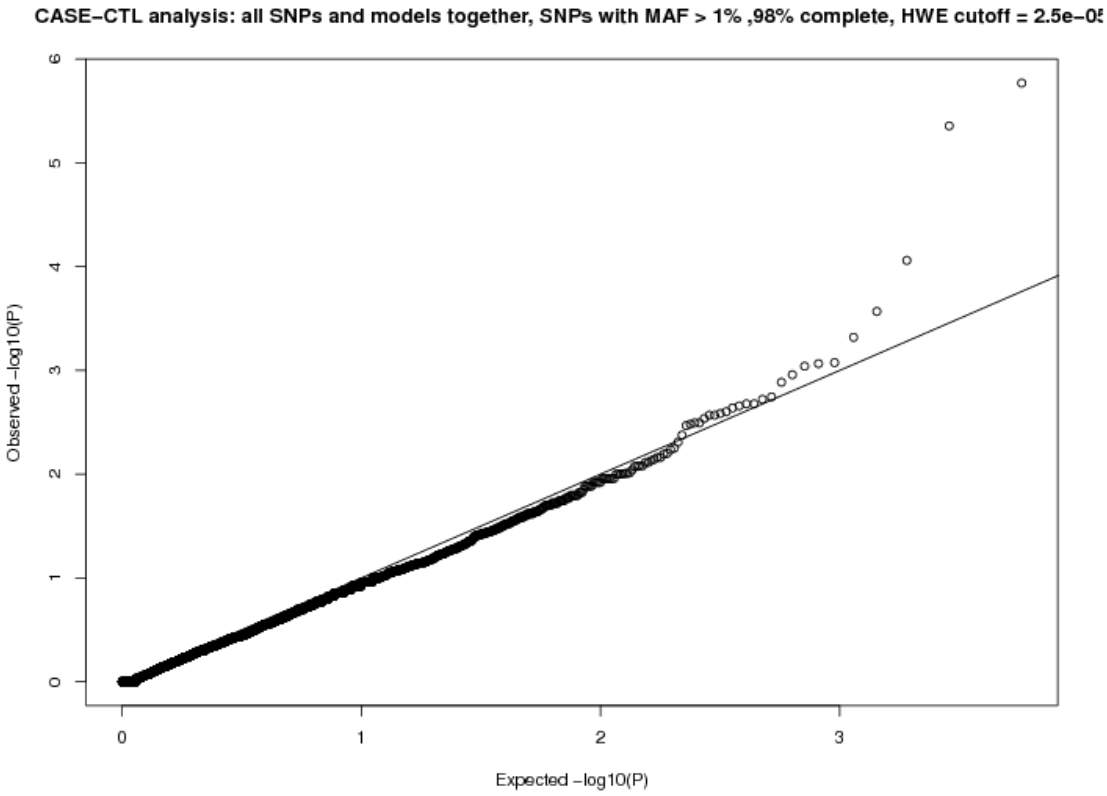

Figure S2. Observed vs. expected distribution of $-\log_{10}$ *p*-values in the case-control analysis of the single nucleotide polymorphisms (SNPs) in the candidate region. The results of the analyses under the allelic, dominant and recessive models are pooled together. MAF, minor allele frequency; HWE, Hardy-Weinberg equilibrium.

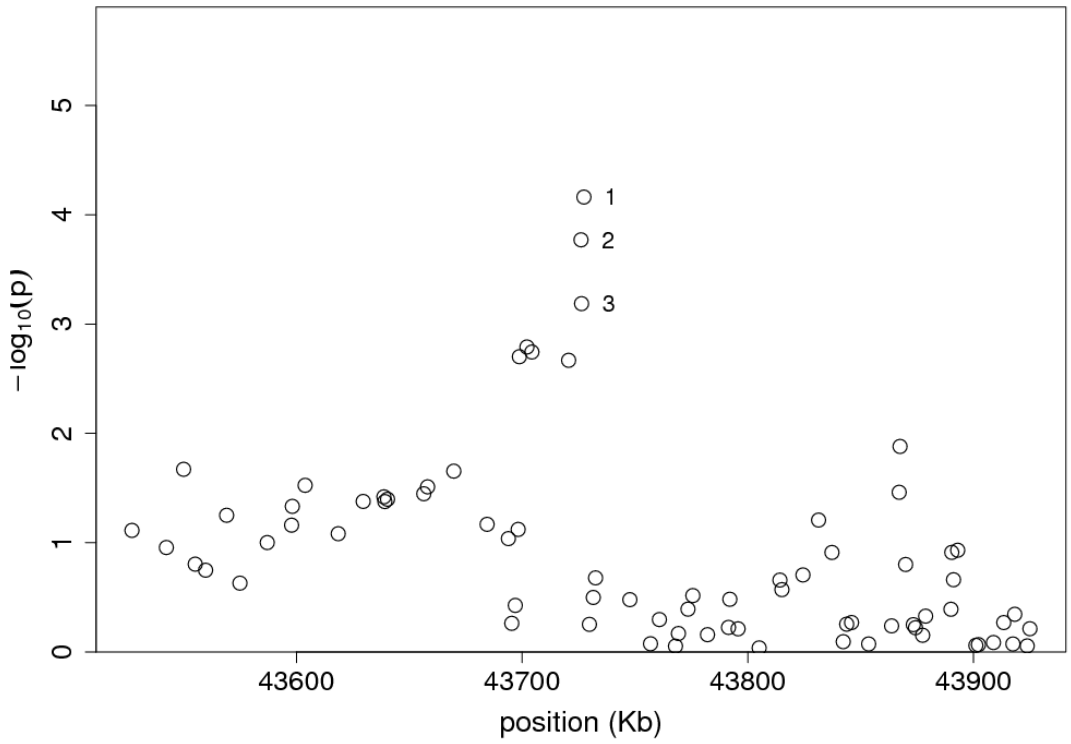

Figure S3: Global association tests with haplotypes consisting of three single nucleotide polymorphisms within 200 kb on either side of rs1156026. 1: rs1156026 - rs2657099 - rs1008913 triplet, 2: rs2120753 - rs2657100 - rs1156026 triplet, 3: rs2657100 - rs1156026 - rs2657099 triplet.

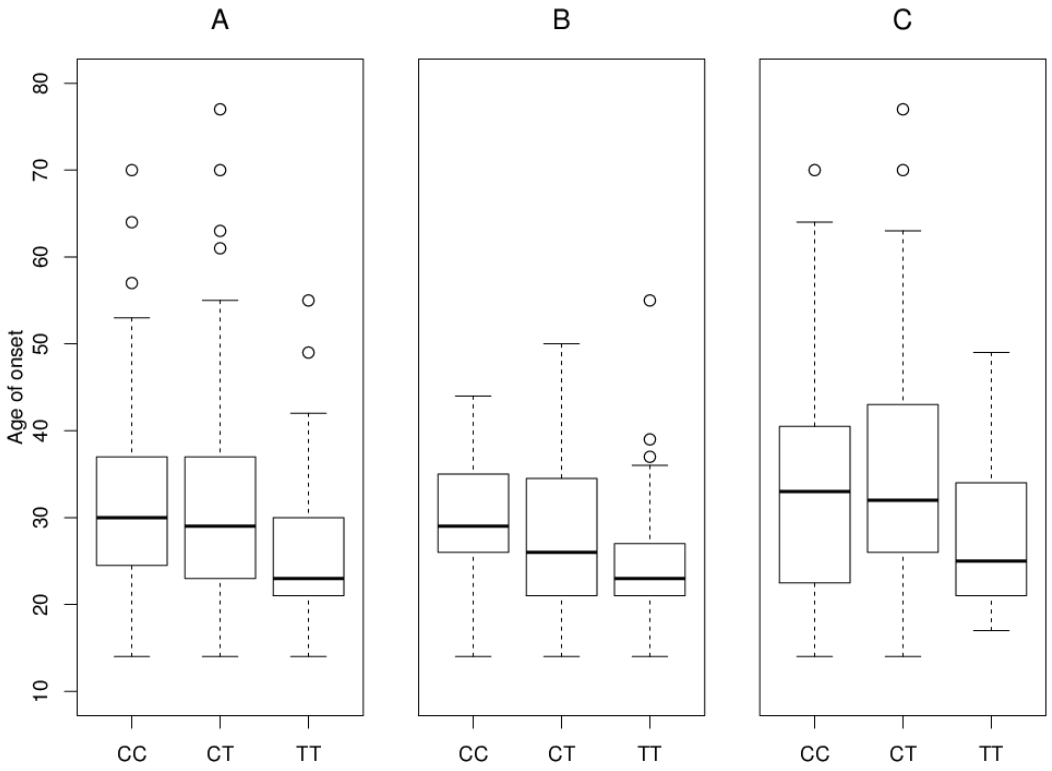

Figure S4: Distribution of the age of onset of (**A**) the common locus, (**B**) the schizophrenia and (**C**) the bipolar disorder phenotypes for the three genotypes of rs1156026. The distribution is shown for the narrow version of each phenotype.

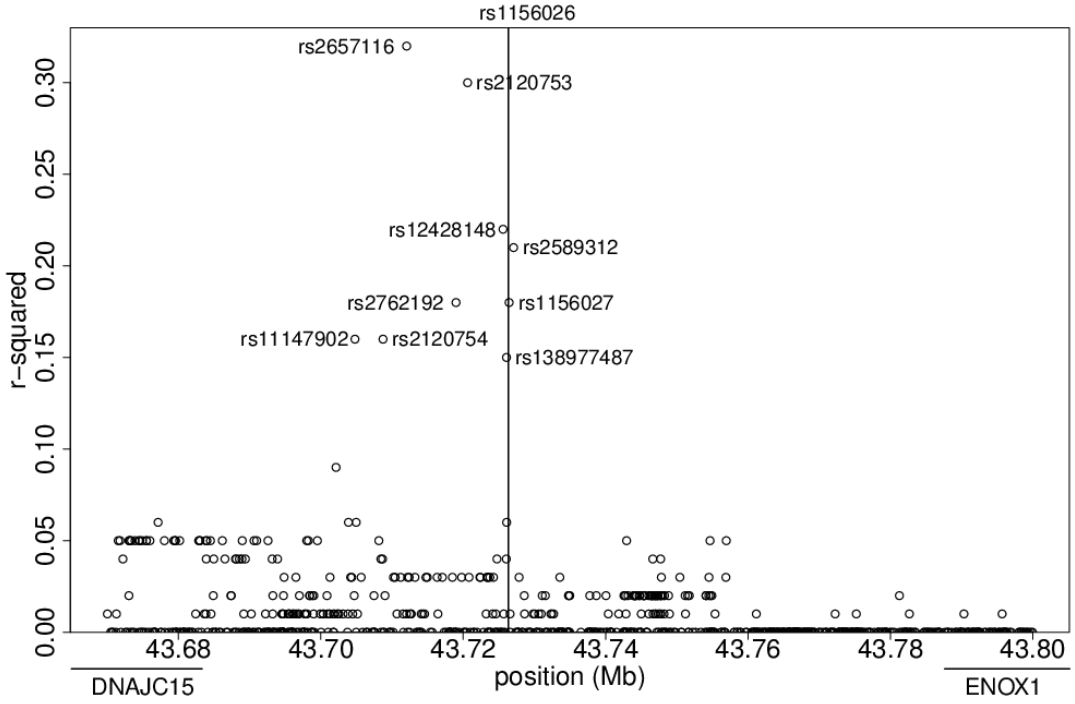

Figure S5: Linkage disequilibrium between rs1156026 and neighboring single nucleotide polymorphisms, measured by the squared correlation in the data from the 1000 Genomes project (Phase I version 3, Aug 2012 update).

Supplemental References

1. Maziade M, Roy MA, Fournier JP, Cliche D, Merette C, Caron C, *et al.* (1992): Reliability of best-estimate diagnosis in genetic linkage studies of major psychoses: results from the Quebec pedigree studies. *Am J Psychiatry* 149:1674-1686.

2. Roy MA, Lanctot G, Merette C, Cliche D, Fournier JP, Boutin P, *et al.* (1997): Clinical and methodological factors related to reliability of the best-estimate diagnostic procedure. *Am J Psychiatry* 154:1726-1733.

3. Sun X, Ding H, Hung K, Guo B (2000): A new MALDI-TOF based mini-sequencing assay for genotyping of SNPS. *Nucleic Acids Research* 28:e68.

4. O'Connell JR, Weeks DE (1998): PedCheck: A Program for Identification of Genotype Incompatibilities in Linkage Analysis. *Am J Hum Genet* 63:259-266.

5. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, *et al.* (2007): PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 17:1665-1674.

6. Patterson N, Price AL, Reich D (2006): Population structure and eigenanalysis. *PLoS Genet* 2:e190.

7. Devlin B, Roeder K, Wasserman L (2001): Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol* 60:155-166.

8. Huang BE, Amos CI, Lin DY (2007): Detecting haplotype effects in genomewide association studies. *Genet Epidemiol* 31:803-812.

9. Howie BN, Donnelly P, Marchini J (2009): A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5:e1000529.

10. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA (2002): Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 70:425-434.

11. Lake SL, Lyon H, Tantisira K, Silverman EK, Weiss ST, Laird NM, Schaid DJ (2003): Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. *Hum Hered* 55:56-65.

12. Zeger SL, Liang KY, Albert PS (1988): Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 44:1049-1060.

13. Chen WM, Manichaikul A, Rich SS (2009): A generalized family-based association test for dichotomous traits. *Am J Hum Genet* 85:364-376.

14. Fishelson M, Dovgolevsky N, Geiger D (2005): Maximum likelihood haplotyping for general pedigrees. *Hum Hered* 59:41-60.