

A Genome-Wide Linkage Scan for Distinct Subsets of Schizophrenia Characterized by Age at Onset and Neurocognitive Deficits

Yin-Ju Lien^{1,2,3}, Po-Chang Hsiao², Chih-Min Liu⁴, Stephen V. Faraone⁵, Ming T. Tsuang^{6,7}, Hai-Gwo Hwu^{1,4}, Wei J. Chen^{1,2,3,4*}

1 College of Public Health, Research Center for Genes and Institute of Epidemiology and Preventive Medicine, Environment and Human Health, National Taiwan University, Taipei, Taiwan, **2** Genetic Epidemiology Core Laboratory, Center of Genomic Medicine, National Taiwan University, Taipei, Taiwan, **3** Department of Public Health, College of Public Health, National Taiwan University, Taipei, Taiwan, **4** Department of Psychiatry, College of Medicine and National Taiwan University Hospital, National Taiwan University, Taipei, Taiwan, **5** Departments of Psychiatry and of Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, New York, United States of America, **6** Department of Psychiatry and the Center for Behavioral Genomics, University of California San Diego, San Diego, California, United States of America, **7** Harvard Departments of Epidemiology and Psychiatry, Harvard Institute of Psychiatric Epidemiology and Genetics, Boston, Massachusetts, United States of America

Abstract

Background: As schizophrenia is genetically and phenotypically heterogeneous, targeting genetically informative phenotypes may help identify greater linkage signals. The aim of the study is to evaluate the genetic linkage evidence for schizophrenia in subsets of families with earlier age at onset or greater neurocognitive deficits.

Methods: Patients with schizophrenia ($n = 1,207$) and their first-degree relatives ($n = 1,035$) from 557 families with schizophrenia were recruited from six data collection field research centers throughout Taiwan. Subjects completed a face-to-face semi-structured interview, the Continuous Performance Test (CPT), the Wisconsin Card Sorting Test, and were genotyped with 386 microsatellite markers across the genome.

Results: A maximum nonparametric logarithm of odds (LOD) score of 4.17 at 2q22.1 was found in 295 families ranked by increasing age at onset, which had significant increases in the maximum LOD score compared with those obtained in initial linkage analyses using all available families. Based on this subset, a further subsetting by false alarm rate on the undegraded and degraded CPT obtained further increase in the nested subset-based LOD on 2q22.1, with a score of 7.36 in 228 families and 7.71 in 243 families, respectively.

Conclusion: We found possible evidence of linkage on chromosome 2q22.1 in families of schizophrenia patients with more CPT false alarm rates nested within the families with younger age at onset. These results highlight the importance of incorporating genetically informative phenotypes in unraveling the complex genetics of schizophrenia.

Citation: Lien Y-J, Hsiao P-C, Liu C-M, Faraone SV, Tsuang MT, et al. (2011) A Genome-Wide Linkage Scan for Distinct Subsets of Schizophrenia Characterized by Age at Onset and Neurocognitive Deficits. PLoS ONE 6(8): e24103. doi:10.1371/journal.pone.0024103

Editor: Christian Schönbach, Kyushu Institute of Technology, Japan

Received: May 16, 2011; **Accepted:** July 30, 2011; **Published:** August 29, 2011

Copyright: © 2011 Lien et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by grants from the National Health Research Institutes, Taiwan (DOH 88-HR-825; NHRI-GT-EX89P825P; NHRI-EX90-8825PP; NHRI-EX91, 92, 93-9113PP; NHRI-EX95, 96, 97, 98-9511PP) and the National Science Council, Taiwan (NSC83-0412-B-002-310, NSC84-2331-B-002-187), National Institute of Mental Health (NIMH) grant 1R01-MH-59624-01, and NTU grant 97HP0023. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: wjchen@ntu.edu.tw

Introduction

Family and twin studies have indicated that schizophrenia has a strong genetic basis [1]. However, previous genetic linkage and association studies have not conclusively detected susceptibility loci and genetic variants for schizophrenia [2,3]. Heterogeneity in the etiology of schizophrenia may partly account for the conflicting results. Further, the current diagnostic schemes may not capture the most heritable features of schizophrenia. To overcome these obstacles, using genetically informative phenotypes has been advocated since this may reduce heterogeneity or represent the more direct effects of underlying susceptibility genes [4].

Conceptually, a genetically informative phenotype can be a component phenotype, i.e., a symptom or dimension which is a component of the disorder itself found only in affected individuals, or an intermediate phenotype, i.e., a quantitative phenotype which is found in both affected individuals and their nonpsychotic relatives.

Cumulative evidence suggests that age at onset is one potential component phenotype for schizophrenia. Compared with schizophrenia patients with late onset, those with early age at onset have been associated with more severe cognitive impairments and poorer outcomes [5]. In addition, patients with an early onset were more likely to have relatives affected with schizophrenia [6], and a

greater correlation in age at onset was found in affected monozygotic twin pairs than in dizygotic twin pairs [7]. Meanwhile, neurocognitive deficits such as sustained attention deficit and executive dysfunction are possible intermediate phenotypes for schizophrenia. Sustained attention deficits as measured on the Continuous Performance Test (CPT) [8] are present, not only in schizophrenia patients, but also in subjects with schizotypal personality disorder and in nonpsychotic relatives of schizophrenia patients [9]. Family studies have found an elevated recurrence risk ratio for CPT deficits among nonpsychotic parents or siblings [10] and a positive association between the severity of the CPT deficits and the familial loading for schizophrenia [11]. The performance deficits in schizophrenia patients are stable in clinical course and not amenable to treatment with neuroleptics [12]. Executive functioning as measured by the Wisconsin Card Sorting Test (WCST) [13] have also been found to be impaired in schizophrenia patients [14] and their first degree relatives [15], though the magnitude of familial aggregation is modest at best [16].

A number of candidate gene studies have explored the relations of certain genetic variants to schizophrenia by subgrouping patients using these genetically informative phenotypes of schizophrenia, including age at onset [17], and differential impairments on the CPT or WCST [18]. However, the subgrouping in these studies was mainly based on a predetermined cut-off, which may not result in the most homogenous subsets. An attractive alternative for exploring susceptibility loci influencing different informative phenotypes of schizophrenia is a discovery-based genome-wide scan such as ordered subset analysis (OSA), a method not requiring a priori specification of the subset [19]. Given that age at onset is only partially associated with the severity of neurocognitive deficits in schizophrenia [20], subsetting families using a combination of these phenotypes may help further reveal a more homogeneous subset of families with greater linkage signals.

This study aimed to evaluate the genetic linkage evidence for empirically derived subtypes of schizophrenia by applying ordered subset linkage analyses in a large sample of families of siblings co-affected with schizophrenia, which consisted mainly of a single ethnic group [21]. We hypothesized that using a homogeneous subtype of schizophrenia on the basis of age at onset or neurocognitive deficits can help obtain increased linkage signals as well as identify additional regions as compared with the initial genome-wide linkage scan of schizophrenia.

Methods

Subjects and clinical evaluation

Participants of this study included patients with schizophrenia and their first-degree relatives recruited from six data collection research center throughout the nation in the Taiwan Schizophrenia Linkage Study between 1998 and 2002. The study design, ascertainment process, and sample characteristics have been described in more detail elsewhere [22]. In brief, the large nationwide sample of schizophrenia families were ascertained on the basis of sib-pairs co-affected with DSM-IV schizophrenia, which included only families of Han Chinese ancestry. The original linkage sample consisted of 1,207 affected individuals (57 parents and 1,150 siblings) and 1,035 unaffected individuals (764 parents and 828 siblings) from 557 schizophrenia families [23]. All subjects provided written informed consent after the study procedures were explained to them. Best-estimate final diagnoses were made by two board-certified research psychiatrists independently on the basis of all clinical information, including the Diagnostic Interview for Genetic Studies (DIGS) [24], the Family Interview for Genetic Studies

(FIGS) [25], hospital records, and the interviewer's notes. The study was approved by both the US Department of Health and Human Services and the National Taiwan University Hospital's Internal Review Board of Human Studies.

The DIGS includes a psychosis section that inquires about the age at onset of the first psychotic episode that could not be attributed to medical illness, medications, or substance abuse. When the information with regard to onset of the first psychotic episode on the interview was unclear, research psychiatrists would review the medical history and determine the age at onset.

Neurocognitive assessment

Probands and their relatives completed both the CPT and the WCST. The degraded and undegraded version of CPT '-1-9' procedure has been described in detail elsewhere [26]. Briefly, numbers from 0 to 9 were randomly presented for 50 msec each, at a rate of one per second. Each subject undertook two CPT '-1-9' sessions: the undegraded 1-9 tasks and 25% degraded 1-9 tasks. Sensitivity (d'), derived from the hit rate (probability of response to target trials) and false alarm rate (probability of response to nontarget trials), reflects an individual's ability to discriminate target stimuli from nontarget ones. In addition, the reaction time (i.e., the mean time to respond correctly) for each session was also used for the analyses.

For the WCST, we used a computerized version [27]. Subjects were required to match response cards to the four stimulus cards along one of three dimensions (color, form, or number) by pressing one of the 1 to 4 number keys on the computer keyboard. Subjects were not informed of the correct sorting principle, nor were they told when the principle would shift during the test, but they were given feedback ("Right" or "Wrong") on the screen after each trial. The testing continued until all 128 cards were sorted. Eight performance indices as described in the WCST manual [28] were used for subsequent analyses: Total Errors, Nonperseverative Errors, Perseverative Errors, Perseverative Responses, Categories Achieved, Trials to Complete First Category, Conceptual Level Response, and Failure to Maintain Set.

The adjusted z scores of the CPT indices were derived by means of standardizing the raw scores with adjustments for sex, age, and education against a community sample of 345 individuals [26], while adjusted z scores of the WCST with adjustments for sex, age, and education derived for individual against another group of 440 healthy controls [29]. Because some adjusted z scores were extremely small (< -5) or large (> 5), these scores were winsorized as -5 or 5 , respectively.

Genotyping

The original genotyping was conducted by the Center for Inherited Disease Research, with 386 markers spaced at an average of 9-cM intervals, following the center's standard genotyping procedures (<http://www.cidr.jhmi.edu>). A series of procedures for quality control was adopted for the genotyping, such as duplicate assays with a discordancy rate of 0.06%, and checking for non-Mendelian inheritance and pedigree inconsistencies with an overall rate of 0.39% in genotype and family errors [21]. Erroneous genotypes were removed accordingly. Marker distances were generated by using the sex-averaged Marshfield genetic map. The data included for this study were collected through NIMH's genetic initiative for schizophrenia and details about how to access these data can be found at the following website: <http://zork.wustl.edu/nimh>.

Statistical analysis

Ordered subset linkage analyses. Based on the original sample of genome-wide linkage analysis for schizophrenia, a series

of OSAs [19] were conducted using nonparametric logarithm of odds (LOD) scores by means of Kong and Cox linear model in affected pedigree members [30] derived from software Merlin [31] as input. The method ranks each family by a family-level value of a disease-related covariate of interest and identifies the contiguous subset of families that maximize the evidence for linkage.

In this study, a series of family's covariates was explored, including the average values of the age at onset, CPT scores, and WCST scores from affected siblings. The whole procedures are depicted in the upper part of **Figure 1**. Of the original 557 families, 556 families had information on age at onset, 509 families on undegraded CPT scores, 499 families on degraded CPT scores, and 444 families on WCST scores, respectively. At the beginning of OSA, all the families were ranked according to the magnitude of a chosen family covariate, with those having the same covariate given the same rank. The genome-wide linkage analysis was then conducted in a series of subsets incrementally, starting with the rank that had the greatest family covariate. The procedure was repeated until all families were included in the final subset.

Since earlier age at onset as well as greater CPT-related or WCST-related deficits were associated with greater familial loading for schizophrenia, the following indices were ranked using ascending order: age at onset, four CPT indices (i.e., hit rate and *d'* on both undegraded and degraded task), and three WCST indices (i.e., Categories Achieved, Conceptual Level Response, and Failure to Maintain Set). Meanwhile, the remaining indices were ranked using the descending order: four CPT indices (i.e., false alarm rate and reaction time on both undegraded and degraded task) and five WCST indices (i.e., Total Errors, Non-perseverative Errors, Perseverative Errors, Perseverative Responses, Trials to Complete First Category).

Chromosome-wide significance. The statistical significance of the difference between the greatest LOD scores in a subset of families identified by the OSA and the LOD scores at the same position in the full set of families was evaluated via 1,000 permutations. Under the null hypothesis that the ranking of the covariate is independent of the family's LOD scores on the target

chromosome, the families were randomly permuted with respect to the covariate ranking and a chromosome-wide *p* value for each chromosome was yielded. A *P* value of ≤ 0.05 was considered significant, suggesting that a particular subset of individuals in the sample is more strongly linked to the specified chromosomal region than the whole sample. The procedures are summarized in the left part of **Figure 2**.

Nested ordered subset linkage analysis. Because OSA-derived subsets of families might have tied covariates, we attempted to introduce another subphenotype to further dissect the families (right part of **Figure 1**). In an OSA-derived subset of families characterized by a covariate that resulted in the greatest LOD scores, the families were further re-ranked by means of another subphenotype. The nested OSA procedure was repeated to seek a nested subset of families that resulted in a further increase in the linkage signal on the particular chromosome than the original subset. A *P* value of <0.05 in permutation was considered significant.

Genome-wide significance. After the permutation on the composition of a subset of families, which provides chromosome-wide significance level, we further evaluated its genome-wide significance by simulating the genotype data of the families and then estimating the empirical significance level with correction for multiple phenotypes, as illustrated in the right part of **Figure 2**. First, the genome-wide significance level of a subset-based maximum LOD score was derived from 1,000 simulations in our genotype data generated by the gene-dropping algorithm in Merlin. Simulated data were based on our original family structure, marker informativeness, spacing, and missing status, with phenotypic measurement and affection status being preserved. An *ad hoc* correction procedure was performed to account for the multiple informative phenotypes used in this study. For the OSA part, there were 17 covariates examined (including age at onset and 16 indices on the CPT and the WCST), and thus the threshold of a corrected genome-wide significance level was set as 0.0029 (i.e., $0.05/17$). For the nested OSA part, there were 33 covariates examined (17 covariates in the OSA and 16 indices on

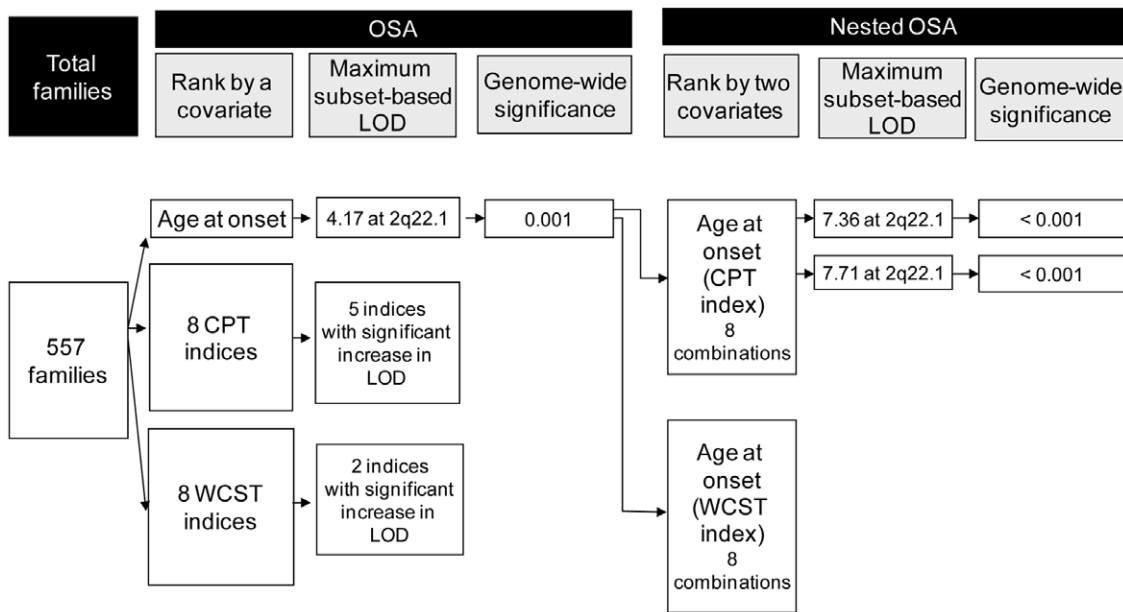


Figure 1. Flow chart of the ordered subset analysis (OSA) and nested OSA of genetic linkage in 557 families of siblings co-affected with schizophrenia. CPT, Continuous Performance Test; WCST, Wisconsin Card Sorting Test; LOD, logarithm of odds. doi:10.1371/journal.pone.0024103.g001

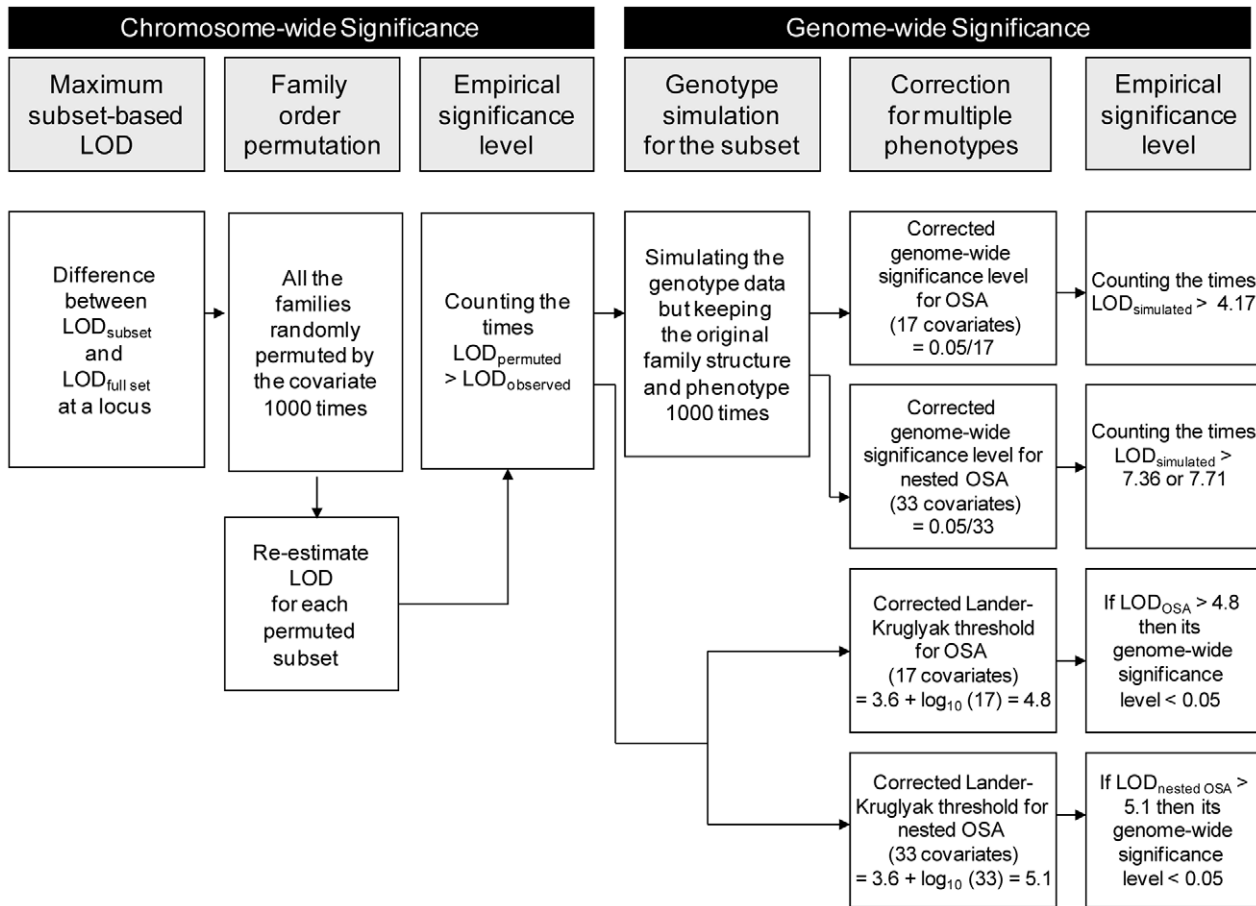


Figure 2. Flow chart of the estimation of chromosome-wide significance (left panel) and genome-wide significance (right panel) for a maximum subset-based logarithm of odds (LOD) in the ordered subset analysis (OSA) and nested OSA of genetic linkage in 557 families of siblings co-affected with schizophrenia.

doi:10.1371/journal.pone.0024103.g002

the CPT and the WCST), and the threshold of a corrected genome-wide significance level was set as 0.0015 (i.e., $0.05/33$).

Second, according to the Lander and Kruglyak threshold [32], which was also derived from simulations, a LOD score of greater than 3.6 indicates attaining genome-wide significance. To maintain the overall genome-wide significance level after considering multiple informative phenotypes used in this study, an *ad hoc* correction procedure was also performed, which was calculated as $LOD_{(corrected)} = 3.6 + \log_{10}(\text{numbers of tests})$ [33]. Accordingly, the threshold of a LOD score reaching genome-wide significance was raised to 4.8 for the OSA part, and to 5.1 for the nested OSA part.

Family covariates comparison. The family covariate values between the subset of families that resulted in the greatest linkage signals and the remaining families were then compared. A linear mixed effect model with family treated as a random effect was applied using the Proc MIXED of the SAS software version 9.1 (SAS Institute, Cary, NC, USA) to adjust for within-family correlation.

Results

The distributions of individual covariates, which were averaged over the affected siblings of each family, are listed in **Table 1**. On the basis of the adjusted z scores, the affected siblings had poorer performance across all of the indices that were based on the

correctness of the response on both the CPT and WCST in comparison with the normative data. Meanwhile, the affected siblings had longer reaction time on both the undegraded and degraded CPT than the community norm. Furthermore, the measures of age at onset and neurocognitive performances employed in this study demonstrated significant familial resemblance, with an affected sib-pair correlation of 0.29 for age at onset ($P < 0.0001$), 0.14 to 0.31 for CPT indices (P ranging from 0.004 to 0.0001), and 0.12 to 0.24 for WCST indices (P ranging from 0.04 to 0.0001).

Table 2 summarizes the results of OSA that had a significant increase (empirical $P \leq 0.05$) in a subset-based maximum LOD score from the initial LOD scores of the whole sample. The results of the OSA identified distinct chromosomal regions implicated for schizophrenia with different informative phenotypes: 1) 2q22.1 for younger age at onset; 2) 7q32.2, 8q13.1, and 9q34 for greater undegraded CPT-related deficits; 3) 7q32.3 for greater degraded CPT-related deficits; and 4) 8q13.1 and 8q21.1 for greater WCST-related deficits. For the five loci with significant increases in subset-based LOD scores, the percentage of linked families in each subset ranged from 10% to 53% of the whole sample. When the families within the subset were compared with the remaining families for the corresponding subphenotype, there were indeed significant differences between the two groups in the covariates, with earlier age at onset or more neurocognitive deficits in the subset of families. Of those covariate/locus combinations with a

Table 1. Distribution of age at onset, CPT scores, and WCST scores in affected siblings.

Variables	Raw scores			Adjusted z scores ^a		
	Mean (SD)	Q1 ^b	Q3 ^c	Mean (SD)	Q1 ^b	Q3 ^c
Age at onset (N = 1,085)	22.6 (6.3)	18.0	26.0	---	---	---
CPT-'1-9' indices						
Undegraded (N = 917)						
Hit rate	0.7 (0.3)	0.5	0.9	-1.9 (2.0)*	-3.9	-0.2
False alarm rate	0.1 (0.1)	0.0	0.1	1.5 (2.0)*	3.0	-0.1
d'	2.6 (1.5)	1.6	3.7	-2.3 (1.8)*	-3.9	-0.9
Reaction time (centi-sec)	45.1 (11.2)	37.0	51.0	1.0 (1.4)*	0.1	1.8
Degraded (N = 881)						
Hit rate	0.4 (0.4)	0.0	0.8	-2.5 (1.8)*	-4.2	-0.8
False alarm rate	0.1 (0.1)	0.0	0.1	1.2 (1.8)*	-0.2	2.2
d'	1.5 (1.6)	0.0	2.8	-2.6 (1.5)*	-3.9	-1.4
Reaction time (centi-sec)	53.6 (12.8)	45.0	66.0	1.0 (1.6)*	0.0	2.5
WCST indices (N = 718)						
Total Errors	73.4 (24.3)	55.0	95.0	1.4 (1.2)*	0.6	2.3
Non-perseverative Errors	29.2 (25.2)	11.0	38.0	0.5 (1.9)*	-1.0	1.2
Perseverative Errors	44.3 (28.4)	19.0	72.0	1.4 (1.8)*	-0.2	3.1
Perseverative Response	54.7 (39.3)	20.0	93.0	1.4 (1.9)*	-0.3	3.2
Categories Achieved	1.9 (2.5)	0.0	3.0	-1.0 (0.8)*	-1.6	-0.7
Conceptual Level Response	26.1 (24.1)	4.7	43.0	-1.3 (1.1)*	-2.2	-0.7
Failure to Maintain Set	0.8 (1.3)	0.0	1.0	-0.5 (1.0)*	-1.1	-0.2
Trials to Complete First Category	74.3 (54.2)	13.0	129.0	1.0 (1.5)*	-0.6	2.4

^aThe adjusted z scores were derived by means of standardizing the raw scores with adjustments for sex, age, and education against our norm data.

^bCut-off at the lowest 25% of data.

^cCut-off at the lowest 75% of data.

* $P < 0.0001$ for the t-test examining if the adjusted z scores significantly different than 0.

doi:10.1371/journal.pone.0024103.t001

significantly increased LOD score, only “age at onset/2q22.1” reached genome-wide significance based on simulations after an *ad hoc* correction for multiple informative phenotypes, whereas no locus reached the Lander and Kruglyak genome-wide significance threshold after an *ad hoc* correction for multiple informative phenotypes.

Among the loci identified in **Table 2**, the greatest LOD scores of 4.17 at D2S442 (147 cM) was found in 53% of the families ($n = 295$, mean age at onset = 18.70) ranked by age at onset. Since many of the families had the same mean values of age at onset, we were interested in whether a further subsetting by one CPT or WCST index could result in further increase in the LOD score on the same chromosomal region. Before the subsetting, we examined the relations of age at onset to either a CPT or WCST index and found no significant correlations in any of them (data not shown). As displayed in **Table 3**, among these CPT-related and WCST-related deficits, only undegraded and degraded CPT false alarm rate reached significant increase in the nested OSA-based LOD on 2q22.1, with a score of 7.36 (empirical $P = 0.004$) in 81% of 282 families (mean adjusted z scores of false alarm rate = 1.84, mean age at onset = 18.40) and 7.71 (empirical $P = 0.001$) in 87% of 281 families (mean adjusted z scores of false alarm rate = 1.49, mean age at onset = 18.40), respectively. Of note, both the LOD score of 7.36 (nominal $P < 0.00001$) and of 7.71 (nominal $P < 0.00001$) reached genome-wide significance, regardless of which method was used in the significance evaluation, i.e., simulated genotype data or the adjusted Lander and Kruglyak threshold after

accounting for multiple phenotypes. More severe CPT or WCST deficits were indeed found in the families of the nested subset than those in the remaining families, as summarized in **Figure 1**.

The distributions of the genome-wide LOD scores of the nested OSAs that were based on age at onset and CPT false alarm rates over 22 autosomal chromosomes are illustrated in **Figure 3**. The corresponding chromosome-wide LOD distributions on chromosome 2 from the results of OSA and nested OSA are shown in **Figure 4**. Compared with the initial linkage analysis for schizophrenia using all available families, OSA using age at earlier onset as covariate did strengthen and narrow the linkage peak. Moreover, the strengthening pattern was further augmented when additionally incorporating undegraded or degraded CPT false alarm rate as another family covariate in the nested OSA.

Discussion

This study found that subsets of families of patients with schizophrenia characterized by younger age at onset or neuro-cognitive deficits exhibited increased linkage signals and reached genome-wide significance on 2q22.1, which was nearby a suggestive locus (a LOD score of 1.16 on 2q14.1) in the initial genome-wide scan for the whole sample [21]. Furthermore, a subset with greater CPT deficit nested within the families with younger age at onset increased the linkage signal on 2q22.1 to a LOD score as high as 7.71. The results have provided intriguing clues for gene search for a complex disease as schizophrenia.

Table 2. Ordered-subset analyses for schizophrenia by age at onset, CPT, or WCST (only results with a significant change in LOD scores are shown here).

Covariates	Cytogenetic location	Markers	Change in LOD		Maximum subset-based		Mean family Covariates value (SE)	
			(permutation p value) ^a	Initial LOD	LOD (genome-wide empirical p value) ^b	No. Families used/total (%) ^c	Within subset	Remaining families
Age at onset	2q22.1	D2S442	3.3 (0.007)	0.87	4.17 (0.001)	295/556 (53.1)	18.70 (0.16)**	26.86 (0.26)
CPT-'1-9' indices								
<i>Undegraded</i>								
Hit rate	8q13.1	D8S1136	2.25 (0.02)	0.95	3.20 (0.01)	160/509 (31.4)	-4.06 (0.08)**	-0.97 (0.06)
False alarm rate	7q32.3	D7S1804	1.80 (0.01)	1.25	3.05 (0.01)	95/509 (18.7)	1.87 (0.13)*	1.37 (0.07)
d'	8q13.1	D8S1136	0.95 (0.05)	1.83	2.78 (0.02)	158/509 (31.0)	-4.20 (0.06)**	-1.54 (0.07)
Reaction time	9q34	D9S2157	0.43 (0.01)	3.00	3.43 (0.02)	49/509 (9.6)	3.46 (1.00)**	0.81 (1.18)
<i>Degraded</i>								
Reaction time	7q32.3	D7S1804	1.35 (0.004)	2.30	3.65 (0.01)	161/499 (32.3)	2.46 (0.86)**	0.34 (1.39)
WCST indices								
Non-perseverative Errors	8q21.1	D8S1119	3.31 (0.004)	0.01	3.32 (0.01)	86/444 (19.4)	3.33 (0.13)**	-0.15 (0.06)
Categories Achieved	8q13.1	D8S1136	0.51 (0.05)	1.79	2.30 (0.07)	151/444 (34.0)	-1.63 (0.01)**	-0.70 (0.04)

^aFamilies were randomly permuted for 1000 times with respect to the covariate ranking and a chromosome-wide p value for each chromosome was yielded.
^bSignificance level derived from simulations; a genome-wide empirical P-value <0.0029 [i.e., 0.05/(17 covariates)] is denoted in boldface as reaching genome-wide significance.
^cThe number of total families varied due to missing information on the covariate; subsets consisting of ≥ 15% of total families were reported here.
 *P<0.001, **P<0.0001, for the mixed effect model comparing the families covariate values between the subset of families and the remaining one.
 doi:10.1371/journal.pone.0024103.t002

Table 3. Nested ordered-subset analysis in the subgroup with early-onset schizophrenia by CPT or WCST.

Covariates	Cytogenetic location	Markers	Change in LOD		Maximum subset-based		Mean family covariate value (SE)	
			(permutation p value) ^a	Initial LOD	LOD (genome-wide empirical p value) ^b	No. Families used/total (%)	Within nested subset	Remaining families
CPT-'1-9' indices								
Undegraded hit rate	2q22.1	D2S442	0.23 (0.54)	4.23	4.46	275/282 (97.5)	---	---
Undegraded false alarm rate	2q22.1	D2S442	3.13 (0.004)	4.23	7.36 (<0.001)	228/282 (80.9)	1.84 (0.09)*	-0.04 (0.03)
Undegraded d'	2q22.1	D2S442	0.85 (0.08)	4.23	5.08	206/282 (73.0)	---	---
Undegraded reaction time	2q22.1	D2S442	0.48 (0.38)	4.23	4.71	268/282 (95.0)	---	---
Degraded hit rate	2q22.1	D2S442	0.70 (0.16)	4.23	4.93	242/281 (86.1)	---	---
Degraded false alarm rate	2q22.1	D2S442	3.48 (0.001)	4.23	7.71 (<0.001)	243/281 (86.5)	1.49 (0.09)*	-0.26 (0.04)
Degraded d'	2q22.1	D2S442	1.06 (0.09)	4.23	5.29	251/281 (89.3)	---	---
Degraded reaction time	2q22.1	D2S442	1.13 (0.07)	4.23	5.36	262/281 (93.2)	---	---
WCST indices								
Total Errors	2q14.1	D2S410	0.53 (0.08)	3.84	4.37	67/263 (25.5)	---	---
Non-perseverative Errors	2q22.1	D2S442	0.68 (0.16)	3.64	4.32	247/263 (93.9)	---	---
Perseverative Errors	2q22.1	D2S442	0.00 (1.00)	3.64	4.17	263/263 (100.0)	---	---
Perseverative Response	2q22.1	D2S442	0.00 (1.00)	3.84	3.84	263/263 (100.0)	---	---
Categories Achieved	2q14.1	D2S410	0.56 (0.34)	3.18	3.74	55/263 (20.9)	---	---
Conceptual Level Response	2q14.1	D2S410	0.41 (0.40)	3.18	3.59	72/263 (27.4)	---	---
Failure to Maintain Set	2q22.1	D2S442	0.69 (0.47)	3.64	3.87	255/263 (97.0)	---	---
Trials to Complete First Category	2q22.1	D2S442	0.76 (0.15)	3.64	4.40	53/263 (20.2)	---	---

^aFamilies were randomly permuted for 1000 times with respect to the covariate ranking and a chromosome-wide p value for each chromosome was yielded.
^bSignificance level derived from simulations; a genome-wide empirical p-value <0.0015 [i.e., 0.05/33 covariates] is denoted in boldface as reaching genome-wide significance.
 *P<0.0001 for the mixed effect model comparing the families covariate values between the nested subset of families and the remaining ones.
 doi:10.1371/journal.pone.0024103.t003

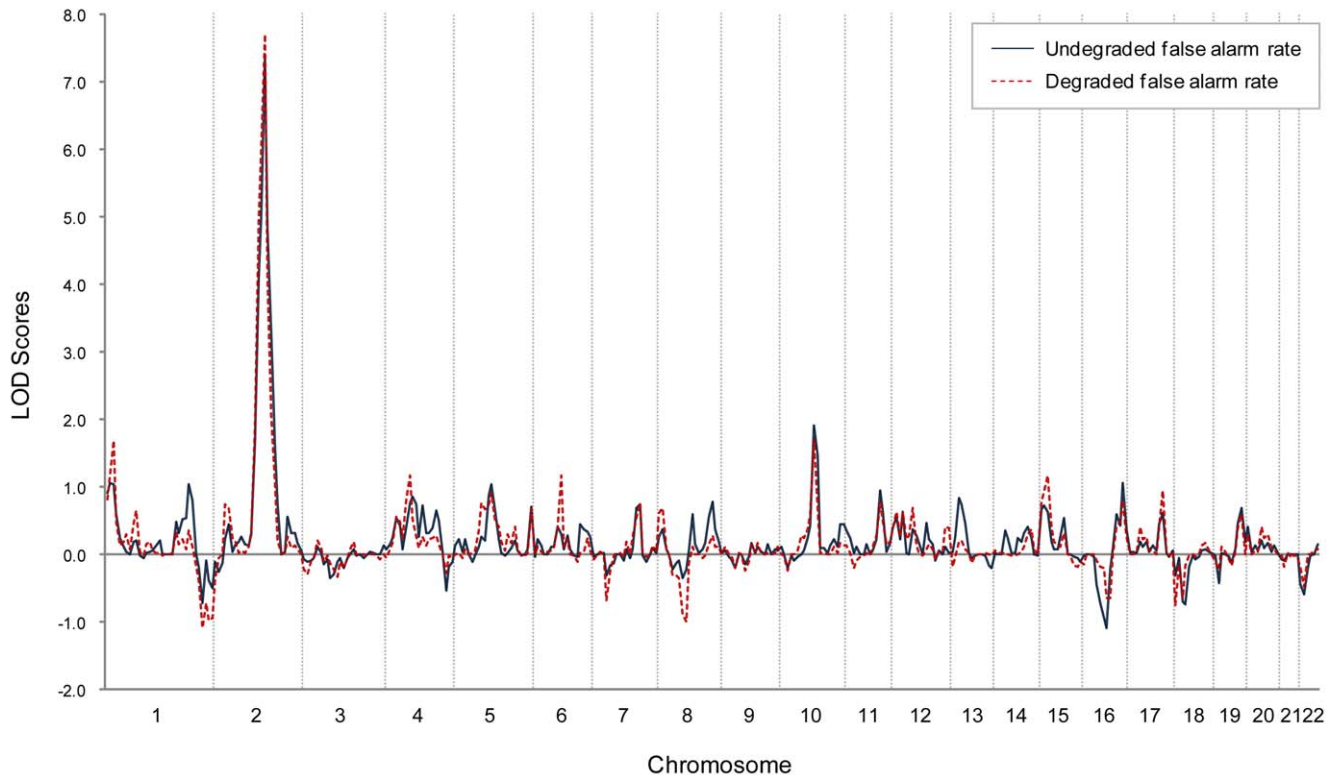


Figure 3. Genome-wide nonparametric LOD scores of two indices attaining genome-wide significance in the nested OSA linkage analyses over 22 autosomal chromosomes.

doi:10.1371/journal.pone.0024103.g003

Our nested OSA approach demonstrates an exploratory way to incorporate various schizophrenia-related characteristics in deciphering the genetic linkage signal to schizophrenia for families with tied covariates (i.e., the same mean values of age at onset). Our results revealed that in the onset-based subset of families having maximal linkage signals, families with tied covariates might in fact have linkage scores in the opposite directions. Accordingly, a further subsetting by one CPT or WCST index might partition the families into more homogeneous subsets and then result in more increase in the LOD scores. Unlike other statistical methods for detecting linkage by dissecting heterogeneity, which require a prior stratification of the data [34] and do not consider quantitative trait information [35], the OSA approach takes advantage of disease-related phenotype information and has been successfully applied in other complex diseases [36,37].

An important feature of this study is that we used age at onset and several facets of the CPT and WCST as covariates to reduce phenotypic and genetic heterogeneity of schizophrenia. These are important features as they clearly impact the outcome and severity of the illness. In addition, cognitive deficit is a core symptom for schizophrenia [38] and the endophenotypic nature of this deficit, more so for CPT [10] than for WCST performance [29], may help index underlying genetic liability or vulnerability of the disease. Particularly, greatest increases in LOD scores were obtained for the families of schizophrenia patients with more CPT false alarm rates nested within the families with younger age at onset. Our findings were consistent with previous studies indicating that some genetic variants were associated with both younger age at onset and more severe cognitive deficits of schizophrenia [39]. Therefore, the combination of clinical and cognitive features used in this study may

aid the identification of susceptibility and modifying genes for schizophrenia.

Unlike previous linkage studies for psychiatric disorders that used selective indices of neurocognitive tasks, either as a composite profile of neurocognitive performance [40,41] or part of multivariate traits [42], this study used all available indices that exhibited significant familial resemblance. Our approach might avoid missing potentially important endophenotypes, such as CPT reaction time [43] or CPT hit rate [44]. Although the psychometric construction of the indices on the tests might be inter-correlated, the cognitive deficits underlying individual indices might be attributed to quite different psychopathologies. For example, previous studies indicated that individual CPT or WCST indices are differentially associated with different symptom dimensions for schizophrenia [43,45,46] or schizotypy [26,47]. In addition, the brain dysfunctions underlying WCST perseverative errors might be different from those underlying nonperseverative errors [9,48,49,50].

However, our approach might suffer from the issue of multiple testing. To fully address this, several strategies were used in this study to help interpret the significance of the OSA results. First, we used permutation tests to estimate the statistical significance of the change in LOD score from the subsets, which controlled for inflation in the false-positive rate induced by examining multiple family subsets for a given covariate. Second, two simulation-based strategies were applied to estimate the genome-wide significance level of subset-based LOD scores, with one based on the simulated genotype data of this sample and the other one based on adjusting the Lander-Kruglyak threshold that was also obtained by simulations. Such simulations can help control for the inflation in the false-positive rate over multiple chromosomes and markers

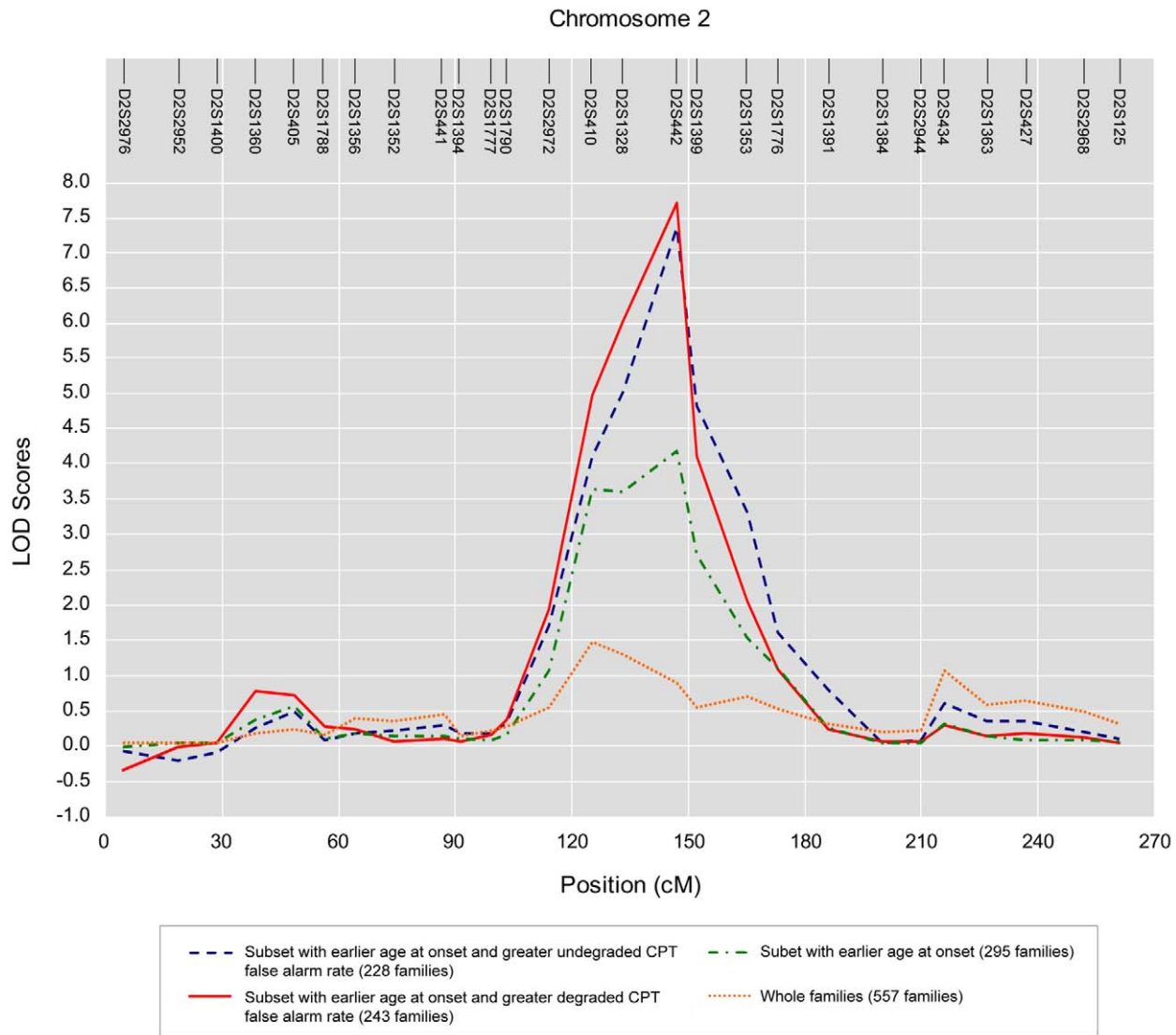


Figure 4. The result of OSA linkage analysis in the subset of families defined by a variety of covariates, including earlier age at onset, greater undegraded CPT false alarm rate nested within earlier age at onset, greater degraded CPT false alarm rate nested within earlier age at onset, versus that of the original linkage analysis of the whole sample on chromosome 2.
doi:10.1371/journal.pone.0024103.g004

since genome-wide significance was estimated by the proportion of times that a particular LOD score was reached anywhere in the simulated genome. Finally, we used a Bonferroni correction to adjust for multiple trait-related covariates after empirical genome-wide P values were derived from simulations. The current results of particular subset-based NPL- Z scores attaining genome-wide significance might be very conservative because a Bonferroni-type correction of the P value was applied for the correlated indices among the CPT or WCST. It should be pointed out that it is very difficult to assess the statistical significance of our subset-based linkage signals due to the complexity inherent in the many possible sources of multiple testing. Accordingly, the results reported here might be better treated as exploratory and independent replication is warranted to reassure the robustness of the findings.

The most striking linkage signal found in this study was obtained from a nested OSA on chromosome 2q22.1 (D2S4422 at 147.4 cM). Intriguingly, a rank-based meta-analysis showed evidence of linkage in this region [51]. A recent genome-wide

association study also found several SNPs on 2q22.1 to be associated with schizophrenia [52]. What this study adds is that this locus was likely to be involved with not only younger age at onset but also greater CPT deficit in false alarm rate in schizophrenia. The region contains potential susceptibility genes for schizophrenia, like HNMT (Histamine N-methyltransferase) and NR4A2 (alias NURR1, nuclear receptor subfamily 4, group A, member 2, an orphan nuclear receptor and putative transcription factor for the dopamine transporter). Although previous studies found that neither HNMT [53] nor NR4A 2 [54,55,56] were associated with schizophrenia, the negative findings might be attributed to the lack of sub-grouping by particular quantitative traits in these studies. Further fine-mapping studies using SNP arrays or deep resequencing methods among patients with early age at onset and CPT deficits may help identify the common genetic variants or rare coding sequence variants that are involved in the genetic susceptibility to this subtype of schizophrenia.

This study is limited in several aspects. First, our samples consisted of adults with Han Chinese ancestry only, and hence the results may not extrapolate to populations of other ethnicities where allele frequencies may vary. Second, despite our use of different strategies to deal with multiple testing, our linkage findings may not rule out the possibilities of false positive or false negative because genome-wide statistical significance of OSA-derived subset results remains a difficult issue. Third, the age at onset of schizophrenia in this study was determined mainly on the initiation of first psychotic symptoms. Although the age at first functional impairment may be closer to the true onset of the disease, there is an overall agreement that the onset of the first psychotic episode is a better marker for the onset of schizophrenia because it has a higher specificity than negative or nonspecific symptoms [57].

In summary, we obtained possible evidence of linkage on chromosome 2q22.1 in families of schizophrenia patients with more CPT false alarm rates nested within the families with younger age at onset. Our results suggest the presence of etiologic

heterogeneity in schizophrenia and highlight the importance of unraveling the complex genetics of schizophrenia by incorporating genetically informative phenotypes.

Acknowledgments

The authors thank other participants in the Taiwan Schizophrenia Linkage Study who helped in the recruitment and evaluation of schizophrenia patients, including Drs. Ming-Hsien Hsieh, Tzung-Jeng Hwang, Shi K. Liu, Ching-Jui Chang, Hung-Jung Chang, Hai Ho, Ping-Ju Chang, Shi-Chin Guo, Hsien-Yuan Lane, Su-Kuan Lin, Fu-Chuan Wei, and Joseph J. and Cheng, and Miss Shih-Wei Lin and Drs. Hui-Chun Tsuang and Sheng-Hsiang Lin for their help in compiling relevant data.

Author Contributions

Conceived and designed the experiments: SVF MTT HGH WJC. Performed the experiments: YJL CML HGH WJC. Analyzed the data: YJL PCH. Wrote the paper: YJL WJC.

References

1. Tsuang M (2000) Schizophrenia: genes and environment. *Biol Psychiatry* 47: 210–220.
2. Crow TJ (2007) How and why genetic linkage has not solved the problem of psychosis: Review and hypothesis. *Am J Psychiatry* 164: 13–21.
3. Sullivan PF (2008) Schizophrenia genetics: the search for a hard lead. *Cur Opin Psychiatry* 21: 157–160.
4. Szatmari P, Maziade M, Zwaigenbaum L, Merette C, Roy MA, et al. (2007) Informative phenotypes for genetic studies of psychiatric disorders. *Am J Med Genet B Neuropsychiatr Genet* 144B: 581–588.
5. Hollis C (2000) Adult outcomes of child- and adolescent-onset schizophrenia: Diagnostic stability and predictive validity. *Am J Psychiatry* 157: 1652–1659.
6. Kendler KS, Maclean CJ (1990) Estimating familial effects on age at onset and liability to schizophrenia. I. Results of a large sample family study. *Genet Epidemiol* 7: 409–417.
7. Kendler KS, Tsuang MT, Hays P (1987) Age at onset in schizophrenia: A familial perspective. *Arch Gen Psychiatry* 44: 881–890.
8. Rosvold HE, Mirsky AF, Sarason I, Bransome ED, Jr., Beck LH (1956) A Continuous Performance Test of brain damage. *J Consult Psychol* 20: 343–350.
9. Chen WJ, Faraone SV (2000) Sustained attention deficits as markers of genetic susceptibility to schizophrenia. *Am J Med Genet* 97: 52–57.
10. Chen WJ, Chang C-H, Liu SK, Hwang TJ, Hwu H-G, et al. (2004) Sustained attention deficits in nonpsychotic relatives of schizophrenic patients: a recurrence risk ratio analysis. *Biol Psychiatry* 55: 995–1000.
11. Tsuang HC, Lin SH, Liu SK, Hsieh MH, Hwang TJ, et al. (2006) More severe sustained attention deficits in nonpsychotic siblings of multiplex schizophrenia families than in those of simplex ones. *Schizophr Res* 87: 172–180.
12. Liu SK, Chen WJ, Chang C-J, Lin H-N (2000) Effects of atypical neuroleptics on sustained attention deficits in schizophrenia: a trial of risperidone versus haloperidol. *Neuropsychopharmacol* 22: 311–319.
13. Robinson AL, Heaton RK, Lehman RAW, Stulson DW (1980) The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal-lobe lesions. *J Consult Clin Psychol* 48: 605–614.
14. Goldberg TE, Weinberger DR, Berman KF, Pliskin NH, Podd MH (1987) Further evidence for dementia of the prefrontal type in schizophrenia? A controlled study of teaching the Wisconsin Card Sorting Test. *Arch Gen Psychiatry* 44: 1008–1014.
15. Szoke A, Schurhoff F, Mathieu F, Meary A, Ionescu S, et al. (2005) Tests of executive functions in first-degree relatives of schizophrenic patients: a meta-analysis. *Psychol Med* 35: 771–782.
16. Laurent A, Biloa-Tang M, Bougerol T, Duly D, Anchisi AM, et al. (2000) Executive/attentional performance and measures of schizotypy in patients with schizophrenia and in their nonpsychotic first-degree relatives. *Schizophr Res* 46: 269–283.
17. Renou J, De Luca V, Zai CC, Bulgin N, Remington G, et al. (2007) Multiple variants of the DRD3, but not BDNF gene, influence age-at-onset of schizophrenia. *Mol Psychiatry* 12: 1058–1060.
18. Liu YL, Fann CSJ, Liu CM, Chang CC, Yang WC, et al. (2007) More evidence supports the association of PPP3CC with schizophrenia. *Mol Psychiatry* 12: 966–974.
19. Hauser ER, Watanabe RM, Duren WL, Bass MP, Langefeld CD, et al. (2004) Ordered subset analysis in genetic linkage mapping of complex traits. *Genet Epidemiol* 27: 53–63.
20. Tuulio-Henriksson A, Partonen T, Suvisaari J, Haukka J, Lonnqvist J (2004) Age at onset and cognitive functioning in schizophrenia. *Br J Psychiatry* 185: 215–219.
21. Faraone SV, Hwu H-G, Liu C-M, Chen WJ, Tsuang M-M, et al. (2006) Genome scan of Han Chinese schizophrenia families from Taiwan: confirmation of linkage to 10q22.3. *Am J Psychiatry* 163: 1760–1766.
22. Hwu H-G, Faraone SV, Liu C-M, Chen WJ, Liu S-K, et al. (2005) Taiwan Schizophrenia Linkage Study: The field study. *Am J Med Genet B Neuropsychiatr Genet* 134B: 30–36.
23. Faraone SV, Lasky-Su J, Glatt SJ, Eerdekewegh PV, Tsuang MT (2006) Early onset bipolar disorder: possible linkage to chromosome 9q34. *Bipolar Disord* 8: 144–151.
24. Nurnberger JI, Jr., Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, et al. (1994) Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch Gen Psychiatry* 51: 849–859.
25. NIMH Genetics Initiative (1992) Family Interview for Genetic Studies. RockvilleMD: National Institute of Mental Health.
26. Chen WJ, Hsiao CK, Hsiao L-L, Hwu H-G (1998) Performance of the Continuous Performance Test among community samples. *Schizophr Bull* 24: 163–174.
27. Lin CCH, Chen WJ, Yang HJ, Hsiao CK, Tien AY (2000) Performance on the Wisconsin Card Sorting Test among adolescents in Taiwan: Norms, factorial structure, and relation to schizotypy. *J Clin Exp Neuropsychol* 22: 69–79.
28. Heaton RK, Chelune GJ, Talley JL, Kay GG, Curtiss G (1993) Wisconsin Card Sorting Test manual: Revised and expanded. Odessa, FL: Psychological Assessment Resources.
29. Lin SH, Liu CM, Hwang TJ, Hsieh MH, Hsiao PC, et al. (2011) Performance on the Wisconsin Card Sorting Test in families of schizophrenia patients with different familial loadings. *Schizophr Bull*, Advance Access published March 9, 2011 (doi:10.1093/schbul/sbr2014).
30. Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 61: 1179–1188.
31. Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30: 97–101.
32. Lander E, Kruglyak L (1995) Genetic dissection of complex traits - guidelines for interpreting and reporting linkage results. *Nat Genet* 11: 241–247.
33. Kidd KK, Ott J (1984) Power and sample-size in linkage studies. *Cytogenet Cell Genet* 37: 510–511.
34. Risch N (1988) A new statistical test for linkage heterogeneity. *Am J Hum Genet* 42: 353–364.
35. Ott J (1986) Linkage probability and its approximate confidence-interval under possible heterogeneity. *Genet Epidemiol Supplement* 1: 251–257.
36. Scott WK, Hauser ER, Schmechel DE, Welsh-Bohmer KA, Small GW, et al. (2003) Ordered-subsets linkage analysis detects novel Alzheimer disease loci on chromosomes 2q34 and 15q22. *Am J Hum Genet* 73: 1041–1051.
37. Shao YJ, Cuccaro ML, Hauser ER, Rairford KL, Menold MM, et al. (2003) Fine mapping of autistic disorder to chromosome 15q11-q13 by use of phenotypic subtypes. *Am J Hum Genet* 72: 539–548.
38. Joyce EM, Roiser JP (2007) Cognitive heterogeneity in schizophrenia. *Cur Opin Psychiatry* 20: 268–272.
39. Joyce E (2005) Origins of cognitive dysfunction in schizophrenia: clues from age at onset. *Br J Psychiatry* 186: 93–95.
40. Hallmayer JF, Jablensky A, Michie P, Woodbury M, Salmon B, et al. (2003) Linkage analysis of candidate regions using a composite neurocognitive phenotype correlated with schizophrenia. *Mol Psychiatry* 8: 511–523.
41. Hallmayer JF, Kalaydjieva L, Badcock J, Dragovic M, Howell S, et al. (2005) Genetic evidence for a distinct subtype of schizophrenia characterized by pervasive cognitive deficit. *Am J Hum Genet* 77: 468–476.

42. Doyle AE, Ferreira MA, Sklar PB, Lasky-Su J, Petty C, et al. (2008) Multivariate genome-wide linkage scan of neurocognitive traits and ADHD symptoms: suggestive linkage to 3q13. *Am J Med Genet B Neuropsychiatr Genet* 147B: 1399–1411.
43. Wang Q, Chan R, Sun JH, Yao J, Deng W, et al. (2007) Reaction time of the Continuous Performance Test is an endophenotypic marker for schizophrenia: A study of first-episode neuroleptic-naïve schizophrenia, their non-psychotic first-degree relatives and healthy population controls. *Schizophr Res* 89: 293–298.
44. Lien YJ, Liu CM, Faraone SV, Tsuang MT, Hwu HG, et al. (2010) A genome-wide quantitative trait loci scan of neurocognitive performances in schizophrenia families. *Genes Brain Behav* 9: 695–702.
45. Liu SK, Hwu HG, Chen WJ (1997) Clinical symptom dimensions and deficits on the Continuous Performance Test in schizophrenia. *Schizophr Res* 25: 211–219.
46. Nieuwenstein MR, Aleman A, de Haan EHF (2001) Relationship between symptom dimensions and neurocognitive functioning in schizophrenia: a meta-analysis of WCST and CPT studies. *J Psychiatr Res* 35: 119–125.
47. Chen WJ, Hsiao CK, Lin CCH (1997) Schizotypy in community samples: The three-factor structure and correlation with sustained attention. *Journal of Abnormal Psychology* 106: 649–654.
48. Barcelo F (1999) Electrophysiological evidence of two different types of error in the Wisconsin Card Sorting Test. *Neuroreport* 10: 1299–1303.
49. Gooding DC, Tallent KA (2002) Spatial working memory performance in patients with schizoaffective psychosis versus schizophrenia: a tale of two disorders? *Schizophr Res* 53: 209–218.
50. Lacerda AT, Dalgalarondo P, Caetano D, Haas GL, Camargo EE, et al. (2003) Neuropsychological performance and regional cerebral blood flow in obsessive-compulsive disorder. *Prog Neuropsychopharmacol Biol Psychiatry* 27: 657–665.
51. Ng MYM, Levinson DF, Faraone SV, Suarez BK, DeLisi LE, et al. (2009) Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol Psychiatry* 14: 774–785.
52. Sullivan PF, Lin D, Tzeng JY, van den Oord E, Perkins D, et al. (2008) Genome-wide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry* 13: 570–584.
53. Yan L, Szumlanski CL, Rice SR, Sobell JL, Lachman HM, et al. (2000) Histamine N-methyltransferase functional polymorphism: Lack of association with schizophrenia. *Am J Med Genet* 96: 404–406.
54. Buervenich S, Carmine A, Arvidsson M, Xiang FQ, Zhang ZP, et al. (2000) NURR1 mutations in cases of schizophrenia and manic-depressive disorder. *Am J Med Genet* 96: 808–813.
55. Chen YH, Tsai MT, Shaw CK, Chen CH (2001) Mutation analysis of the human NR4A2 gene, an essential gene for midbrain dopaminergic neurogenesis, in schizophrenic patients. *Am J Med Genet* 105: 753–757.
56. Talkowski ME, Kirov G, Bamne M, Georgieva L, Torres G, et al. (2008) A network of dopaminergic gene variations implicated as risk factors for schizophrenia. *Hum Mol Genet* 17: 747–758.
57. Gourion D, Goldberger C, Leroy S, Bourdel MC, Olie JP, et al. (2005) Age at onset of schizophrenia: Interaction between brain-derived neurotrophic factor and dopamine D3 receptor gene variants. *Neuroreport* 16: 1407–1410.