Pilot Study on Schizophrenia in Sardinia

Jurg Ott^{a, b} Fabio Macciardi^{c, d} Yuanyuan Shen^a Mauro G. Carta^e
Andrea Murru^e Riccardo Triunfo^{f, g} Renato Robledo^{e, f} Antonietta Rinaldi^{e, f}
Licinio Contu^{e, f} Marcello Siniscalco^{b, f}

^aBeijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China; ^bRockefeller University, Laboratory of Statistical Genetics, New York, N.Y., and ^cDepartment of Psychiatry and Human Behavior, University of California at Irvine, Irvine, Calif., USA; ^dDepartment of Science and Biomedical Technologies, University of Milan, Milan, ^eFaculty of Medicine, University of Cagliari, and ^fSardinian Center for Genome Diversity, Cagliari, and ^gCRS4, Center for Advanced Studies, Research and Development in Sardinia, Cagliari, Italy

Key Words

Schizophrenia · Genome-wide association study · Genetic association

Abstract

Objective: Based on a small sample of cases with schizophrenia and control individuals from an isolated population, a genome-wide association study was undertaken to find variants conferring susceptibility to this disease. **Methods:** Standard association tests were employed, followed by newer multilocus association methods (genotype patterns). **Results:** Individually, no variant produced a significant result. However, the best two variants (rs1360382 on chromosome 9 and rs1303 on chromosome 14) showed significantly different genotype pattern distributions between patients and control individuals. The risk genotype pattern AA-TT is highly predictive of schizophrenia, with estimated sensitivity and specificity of 1 and 0.96, respectively. **Conclusions:** These findings support the hypothesis that schizophrenia is partly due to multiple genetic variants, each with a relatively small effect.

Copyright © 2010 S. Karger AG, Basel

Introduction

In recent years, a large number of genetic variants potentially contributing to schizophrenia have been proposed, and some of them have been replicated and emerged as significant from meta-analyses [1–5]. Micro-deletions and duplications have also been implicated in the vulnerability to schizophrenia [6–8].

For genome-wide genetic case-control association studies, the value of more homogeneous, isolated populations has been pointed out [9, 10] and debated [11]. In particular, the island of Sardinia, located west of the Italian mainland, is unusual in that it has been genetically isolated for many centuries. In spite of their derivation from a common founder group of the late Neolithic age, the marked genetic diversity of contemporary individual Sardinian isolates was substantiated in the elegant studies of Cavalli-Sforza et al. [12]. With the exception of a few major towns, whose present large populations are the result of the urbanization process of the last 50 years, the majority of Sardinians are still subdivided into about 300 isolated villages of various dimensions, each stemming from the same ancestral founder group of the late Neolithic age, yet different from each other through the well-known triggers of genomic diversity such as founder effect, assortative mating, adaptive selection, and chance (genetic

drift). These circumstances, enriched by the availability of precious genealogical data from governmental archives since 1860 and from church records dating back several centuries, offer a unique potential to search for meaningful associations between genomic and/or environmental factors and the expression of complex traits and diseases. For example, in a study of centenarians [13], analysis of Y-chromosomal markers showed that ‘the Sardinian population had two main founder populations that have evolved in isolation for at least 5,000 years. These findings set the stage for future studies on longevity and other complex traits in Sardinia’. One of the advantages of this situation is that background linkage disequilibrium tends to be high so that a smaller number of markers will be as efficient as a much larger number in large outbred populations. As an example, the town of Talana was established a thousand years ago by an extremely small number of original settlers – 75% of the town’s current population descends from 8 founder fathers and 10 founder mothers – and currently comprises 1,200 inhabitants. Most of the marriages occur between people of the same community and 35% of the marriages are consanguineous [14].

Over many years, some of us have been collecting data on schizophrenics and control individuals in several of these isolated townships [15]. As this was a pilot study, we took a small sample from one specific village, Sestu, to carry out a genome-wide screen for schizophrenia susceptibility variants.

Subjects and Methods

Probands were selected from the clinical records of the public psychiatric unit that provides psychiatric care to the village of Sestu (in Italy, psychiatric care is free of charge and dispensed by the National Health System). All patients with a diagnosis of schizophrenia and schizoaffective disorders who fulfilled the criteria of residence were evaluated by two clinicians (with at least three years of experience) by means of the Structured Clinical Interview for DSM-IV Disorders (SCID-IV) of Axis I. SCID-IV is a semi-structured interview aimed at formulating the diagnoses covered by Axis I of DSM-IV. Only patients who fulfilled the DSM-IV criteria for schizophrenia were included in the study. In total, 37 individuals (14 schizophrenics and 23 controls) were recruited for this pilot study (table 1).

The selected individuals had to be distantly related, that is, not having ancestors in common at least for the past four generations. This principle has previously been shown to be extremely fruitful; in fact, it allowed the localization of a (albeit mendelian) disorder on the basis of only three distantly related patients [16].

Genotyping was carried out with Affymetrix 400K SNP chips using standard protocols. Of the original number of 356,075 SNPs, quality control removed the following numbers of SNPs: 39,843 had no chromosomal assignment or were monomorphic; 34,592

Table 1. Characteristics of study individuals

	Number of individuals		Mean age, years		
	females	males	females	males	both
Cases	3	11	48.3	41.6	43.1
Controls	13	10	46.2	51.8	48.6

had call rates <99% (we adopted this stringent criterion because low call rates have been shown to be associated with genotyping errors [17]), and 26,587 had a minor allele frequency of <0.10 (we chose this threshold as our sample size was very small). These steps left a total of 255,053 SNPs for further analysis. We also tested for deviations from Hardy-Weinberg equilibrium (HWE) and adopted a threshold. We chose a critical limit for the χ^2 (1 d.f.) test of 30, which approximately corresponds to a Bonferroni-corrected significance level of 0.01 (255,053 SNPs). This stringent criterion was chosen to reject only those SNPs with very strong deviations from HWE. None of the remaining SNPs violated this condition.

Statistical Analysis

For each SNP, we applied two test statistics as follows: (1) χ^2 (2 d.f.) in the 2×3 table containing numbers of genotypes for cases/controls (two rows) and three SNP genotypes (columns) (genotype test), and (2) χ^2 (1 d.f.) in the 2×2 table of alleles (allele test). Nominal significance levels (p values) were calculated numerically [18] for each SNP and test statistic. For a given test statistic, empirical levels of false discovery rates (FDR) were assigned to p values according to the Benjamini-Hochberg method [19, 20].

To evaluate the combined contribution of more than one SNP, we constructed diplotypes (genotype patterns), that is, sets of genotypes at different SNPs, wherever these SNPs are in the genome. The aim was to explore whether diplotype frequencies are different in case and control individuals [21]. As the number of individuals was quite limited, we only carried out these tests for two and three most significant SNPs found in single-locus analysis. While SNPs may be detected by any single-locus association test, we applied the genotype test as originally proposed [21], because it is sensitive to various types of association [22]. To allow for the fact that SNPs were selected because they showed some disease association by themselves, we evaluated the significance of our findings in a permutation framework that included selecting SNPs under the null hypothesis of no association [21]. Significance levels were determined in 10,000 randomization samples.

Results

Even though our sample was very small, we carried out a cluster analysis with the PLINK program [23]. It showed only a single cluster, so we concluded that there was at least no major population admixture.

The genotype test did not reveal any remarkable results; the smallest FDR value was 0.30 (results not shown).

Table 2. Results of allelic association tests for the seven best SNPs

Rank	SNP	ch	position Mb	χ^2	FDR-BH	pBon	Risk allele	f _A	f _U	Non-risk allele	OR
1	rs1360382	9	23.4	22.3486	0.302	0.440	A	0.857	0.478	G	5.93
2	rs1303	14	93.9	21.9352	0.302	0.513	T	0.964	0.587	G	13.00
3	rs11718902	3	46.7	21.2555	0.302	0.641	A	0.429	0.065	G	9.42
4	rs3793355	8	105.5	20.9388	0.302	0.702	T	0.464	0.087	C	8.23
5	rs1026421	7	50.3	19.2814	0.438	0.944	A	0.893	0.457	G	8.64
6	rs12722932	1	81.7	19.2814	0.438	0.944	T	0.821	0.457	C	5.07
7	rs2837985	21	41.5	19.1619	0.438	0.953	A	1	0.630	G	33.81

ch = Chromosome; FDR-BH = empirical false discovery rate, Benjamini-Hochberg method; pBon = Bonferroni-corrected significance level; f_A, f_U = allele frequency of risk allele in cases and controls, respectively; OR = odds ratio computed after addition of 0.5 to each cell.

Table 3. Genotype pattern (diplotype) frequencies for SNPs rs1360382 and rs1303 in case and control individuals

Pattern ^a	Cases	Controls	Sum	OR ^b	95% CI
AA-TT	13	0	13	423.000	16.1–11,128.7
AG-TG	0	6	6	0.093	0.005–1.79
Rare ^c	1	17	18	0.027	0.003–0.254
Sum	14	23	37		

^a The first genotype refers to rs1360382, the second genotype to rs1303. ^b OR = Odds ratio computed after addition of 0.5 to each cell, given genotype pattern versus all other patterns. ^c Collection of 6 patterns, each with a frequency of <15%.

The allele test resulted in somewhat better findings, although p values were far from significant (table 2). The four SNPs with the smallest p values all had FDR levels of 0.302 (the next higher FDR value was 0.438). The smallest p value, Bonferroni-corrected for multiple testing, was p = 0.440. All SNPs shown in table 2 were in excellent agreement with HWE. Uncorrected significance levels in the test of HWE ranged from 0.22 to 1 in controls and were all equal to 1 in cases.

Next, we tested whether the two or three SNPs with the smallest p values in table 2 formed genotype patterns with different frequencies in case and control individuals; a pattern was considered common when its frequency in cases and controls combined exceeded 15%; all other patterns were pooled into a 'rare' class. While patterns for three SNPs did not reveal any remarkable findings (results not shown), the best two SNPs, rs1360382 on chromosome 9 and rs1303 on chromosome 14, exhibited pat-

tern frequency differences that are significant with p = 0.0162, corrected for multiple testing. Results are shown in table 3. The total number of possible patterns is $3 \times 3 = 9$, of which eight were observed in the data, and two of these were frequent. In our data, the genotype pattern AA-TT (rs1360382- rs1303) occurs only in cases, while the pattern AG-TG occurs only in controls. Individually, these two SNPs are characterized by their alleles and allele frequencies, as shown in table 2. Each acts in a recessive-like manner. Clearly, the AA-TT genotype pattern is highly predictive of schizophrenia. In the comparison between this risk pattern and the other patterns, based on our small dataset, the estimated sensitivity is equal to 1 with a 95% confidence interval of 0.79–1, and specificity is estimated to be 0.96 (0.79–1.00).

Recently, a large association study of schizophrenia was published [24]. In their online supplementary table 1, these authors listed the results for the top 50 disease-associated SNPs based on 2,663 cases and 13,498 controls. Of these 50 SNPs, 36 were also evaluated in the current study. To test whether some or all of these 36 SNPs are associated with schizophrenia in our study, we carried out the following analyses. In the absence of disease association (null hypothesis), the p values for these SNPs are expected to follow a uniform distribution. However, a histogram with 9 classes, each containing an expected number of 4 SNPs ($4 \times 9 = 36$), clearly shows a preponderance of small p values (fig. 1). χ^2 comparing observed and expected numbers of p values is equal to 20.5 (8 d.f., p = 0.0086). Thus, despite the small sample size, our data clearly contain much information on disease-associated SNPs, more so for multiple SNPs combined than for effects of individual variants.

Fig. 1. Histogram of p values for 36 SNPs selected because they were previously shown to be associated with schizophrenia.

Discussion

Clearly, the number of cases and controls used in this pilot study is extremely small. However, in a similarly small study previously carried out as a proof of principle, we successfully detected the well-known close association between high myopia and the genetic marker D18S63 [25]. That study employed 15 high myopia patients and 35 controls carefully selected in Carloforte (Italy) who had no ancestors in common for the last three to four generations. The current study used an analogously collected sample of schizophrenics and control individuals. Thus, the reported results are worth being followed up in a larger study.

References

- Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ, Tanzi RE, Bertram L: Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* 2008;40:827–834.
- Cichon S, Craddock N, Daly M, Faraone SV, Gejman PV, Kelsoe J, Lehner T, Levinson DF, Moran A, Sklar P, Sullivan PF: Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry* 2009;166:540–556.
- Rujescu D, Ingason A, Cichon S, et al: Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet* 2009; 18:988–996.
- Tosic M, Ott J, Barral S, Bovet P, Deppen P, Gheorghita F, Matthey ML, Parnas J, Preisig M, Saraga M, Solida A, Timm S, Wang AG, Werge T, Cuénod M, Do KQ: Schizophrenia and oxidative stress: glutamate cystine ligase modifier as a susceptibility gene. *Am J Hum Genet* 2006;79:586–592.
- Vazza G, Bertolin C, Scudellaro E, Vettori A, Boaretto F, Rampinelli S, De Sanctis G, Perini G, Peruzzi P, Mostacciolo ML: Genome-wide scan supports the existence of a susceptibility locus for schizophrenia and bipolar disorder on chromosome 15q26. *Mol Psychiatry* 2007;12:87–93.
- Stefansson H, Rujescu D, Cichon S, et al: Large recurrent microdeletions associated with schizophrenia. *Nature* 2008;455:232–236.

Of the two SNPs whose genotype patterns are implicated in the vulnerability to schizophrenia (table 3), rs1360382 is unremarkable (not in a gene), but rs1303 is located in the coding region of the SERPINA1 gene on chromosome 14q32.1. This gene is listed in UniGene as serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1, and, according to OMIM, variants of it are responsible for the autosomal recessive trait alpha-1-antitrypsin deficiency. We have found one report suggesting an association between alleles in this gene and whether schizophrenics had a family history or not, although in the same report, there were no significant differences between schizophrenics and controls [26]. While none of the recently published large studies on schizophrenia [24, 27] reported associations of the two SNPs implicated in our study, it is conceivable that mutations associated with schizophrenia are different in small isolated populations and large outbred populations.

Acknowledgments

This work was supported by NSFC grant (project number 30730057) from the Chinese Government. The authors also express their thanks to (1) the *Prefettura di Cagliari* for having permitted the consultation of the *Stato Civile*, which recorded the birth dates of the patients and control individuals; (2) Prof. Bernardo Carpinello, director of the public health unit that provides psychiatric care to the Sardinian village of Sestu, for kindly allowing the ascertainment of 14 schizophrenic patients; (3) the *Fondazione Marino Golinelli*, Bologna, Italy, for its generous support of our work through the Biotechne of Cagliari, and (4) the *Istituto di Tecnologie Biomediche*, CNR, Segrate, Milano, Italy, for the generous hospitality offered during our visits in conjunction with the present research. This work was also supported in part by NIH grants AG026916 and HL084410.

- 7 International Schizophrenia Consortium: Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008;455:237–241.
- 8 Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A, Stray SM, Rippey CF, Roccanova P, Makarov V, Lakshmi B, Findling RL, Sikich L, Stromberg T, Merriman B, Gogtay N, Butler P, Eckstrand K, Noory L, Gochman P, Long R, Chen Z, Davis S, Baker C, Eichler EE, Meltzer PS, Nelson SF, Singleton AB, Lee MK, Rapoport JL, King MC, Sebat J: Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008;320:539–543.
- 9 Khoury MJ, Beaty TH, Cohen BH: *Fundamentals of Genetic Epidemiology*. New York, Oxford University Press, 1993.
- 10 Service S, Sabatti C, Freimer N: Tag SNPs chosen from HapMap perform well in several population isolates. *Genet Epidemiol* 2007;31:189–194.
- 11 Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, Cucca F, Todd JA: The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. *Nat Genet* 2000;25:320–323.
- 12 Cavalli-Sforza LL, Menozzi P, Piazza A: *The History and Geography of Human Genes*. Princeton, N.J., Princeton University Press, 1994.
- 13 Passarino G, Underhill PA, Cavalli-Sforza LL, Semino O, Pes GM, Carru C, Ferrucci L, Bonafè M, Franceschi C, Deiana L, Baggio G, De Benedictis G: Y chromosome binary markers to study the high prevalence of males in Sardinian centenarians and the genetic structure of the Sardinian population. *Hum Hered* 2001;52:136–139.
- 14 Meldolesi A: Italy toasts Sardinia with *ShardNA*. *Nat Biotechnol* 2000;18:1032.
- 15 Siniscalco M, Robledo R, Bender PK, Carcasì C, Contu L, Beck JC: Population genomics in Sardinia: a novel approach to hunt for genomic combinations underlying complex traits and diseases. *Cytogenet Cell Genet* 1999;86:148–152.
- 16 Houwen RH, Baharloo S, Blankenship K, Raeymaekers P, Juyn J, Sandkuijl LA, Freimer NB: Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. *Nat Genet* 1994;8:380–386.
- 17 Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, Pe'er I, Burt NP, Blumenstiel B, DeFelice M, Parkin M, Barry R, Winslow W, Healy C, Graham RR, Neale BM, Izmailova E, Roubenoff R, Parker AN, Glass R, Karlson EW, Maher N, Hafler DA, Lee DM, Seldin MF, Remmers EF, Lee AT, Padyukov L, Alfredsson L, Coby J, Weinblatt ME, Gabriel SB, Purcell S, Klareskog L, Gregersen PK, Shadick NA, Daly MJ, Altshuler D: Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 2007;39:1477–1482.
- 18 Abramowitz M, Stegun IA: *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*, ed 9. New York, Dover Publications, 1972.
- 19 Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc Ser B* 1995;57:289–300.
- 20 Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I: Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 2001;125:279–284.
- 21 Long Q, Zhang Q, Ott J: Detecting disease-associated genotype patterns. *BMC Bioinformatics* 2009;10:S75.
- 22 Zheng G, Freidlin B, Gastwirth JL: Comparison of robust tests for genetic association using case-control studies. *IMS Lecture Notes-Monograph Series* 2006;49:253–265.
- 23 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–575.
- 24 Stefansson H, Ophoff RA, Steinberg S, et al: Common variants conferring risk of schizophrenia. *Nature* 2009;460:744–747.
- 25 Heath S, Robledo R, Beggs W, Feola G, Parodo C, Rinaldi A, Contu L, Dana D, Stambolian D, Siniscalco M: A novel approach to search for identity by descent in small samples of patients and controls from the same mendelian breeding unit: a pilot study on myopia. *Hum Hered* 2001;52:183–190.
- 26 Rudduck C, Franzen G, Lindstrom L: Alpha 1-antitrypsin types in schizophrenia. *Hum Hered* 1985;35:193–195.
- 27 O'Donovan MC, Craddock N, Norton N, et al: Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet* 2008;40:1053–1055.