Molecular Psychiatry (2004) 9, 698–704

www.nature.com/mp

ORIGINAL RESEARCH ARTICLE

Identification of a novel neuregulin 1 at-risk haplotype in Han schizophrenia Chinese patients, but no association with the Icelandic/Scottish risk haplotype

T Li[1,2], H Stefansson[3], E Gudfinnsson[3], G Cai[2], X Liu[2], RM Murray[1], V Steinthorsdottir[3], D Januel[1], VG Gudnadottir[3], H Petursson[4], A Ingason[3], JR Gulcher[3], K Stefansson[3] and DA Collier[1]

[1]*Division of Psychological Medicine, Institute of Psychiatry, London, UK;* [2]*West China Hospital, Sichuan University, PR China;* [3]*Decode Genetics, Reykjavík, Iceland;* [4]*Division of Psychiatry, Landspitali University Hospital, Reykjavík, Iceland*

To determine if neuregulin 1 (NRG1) is associated with schizophrenia in Asian populations, we investigated a Han Chinese population using both a family trio design and a case–control design. A total of 25 microsatellite markers and single nucleotide polymorphisms (SNPs) were genotyped spanning the 1.1 Mb NRG1 gene including markers of a seven-marker haplotype at the 5′end of the gene found to be in excess in Icelandic and Scottish schizophrenia patients. The alleles of the individual markers forming the seven marker at-risk haplotype are not likely to be causative as they are not in excess in patients in the Chinese population studied here. However using unrelated patients, we find a novel haplotype ($HAP_{China\ 1}$), immediately upstream of the Icelandic haplotype, in excess in patients (11.9% in patients *vs* 4.2% in controls; $P = 0.0000065$, risk ratio (rr) 3.1), which was not significant when parental controls were used. Another haplotype ($HAP_{China\ 2}$) overlapping the Icelandic risk haplotype was found in excess in the Chinese (8.5% of patients *vs* 4.0% of unrelated controls; $P = 0.003$, rr 2.2) and was also significant using parental controls only ($P = 0.0047$, rr 2.1). A four-marker haplotype at the 3′ end of the NRG1 gene, $HAP_{China\ 3}$, was found at a frequency of 23.8% in patients and 13.7% in nontransmitted parental haplotypes ($P = 0.000042$, rr = 2.0) but was not significant in the case–control comparison. We conclude that different haplotypes within the boundaries of the NRG1 gene may be associated with schizophrenia in the Han Chinese.
Molecular Psychiatry (2004) **9,** 698–704. doi:10.1038/sj.mp.4001485
Published online 9 March 2004

Keywords: 8p; linkage; glial growth factor; psychosis; genetics; transmission; family

Schizophrenia is a debilitating brain disease that affects up to 1% of the population worldwide. There is no cure, and treatments, which are mainly based on antagonism of dopamine, serotonin and other neurotransmitter receptors in the brain, are only partially effective. Although the aetiology of schizophrenia is not well understood, there is longstanding evidence of a genetic component, and recent twin studies place the heritability of schizophrenia at more than 80%.[1]

Despite this high heritability, numerous candidate gene association studies have failed to identify susceptibility genes clearly. There are a variety of reasons for this failure, including probable locus and aetiological heterogeneity, characterised by the probable existence of numerous genes of small effect, the involvement of several biological pathways, lack of a pathophysiology to help with the selection of candidate genes, and methodological difficulties related to study design, statistical power and diagnostic validity and reliability.[2] Linkage analysis has been more successful, with good evidence for genetic loci identified on chromosomes 8p, 22q, 13q and several other locations in the genome.[3,4] These linkage findings have led to systematic genetic mapping efforts and positional candidate gene analysis of these loci.[5]

In a recent publication by Stefansson *et al*,[6] linkage to chromosome 8p followed by haplotype mapping with microsatellites and single nucleotide polymorphisms (SNPs) identified at-risk neuregulin 1 (NRG1) haplotypes for schizophrenia in patients from Iceland. Three microsatellite at-risk haplotypes, each individually in excess in patients, were found to have the same SNP core haplotype upstream of the first 5′ exon of NRG1. This finding has now been replicated in two large case–control studies from Scotland[7] and the UK.[8] The haplotype frequencies in Scottish and Icelandic patients and in Scottish and Icelandic controls, respectively, are very similar. A second replication study in a UK population found a small excess of the same NRG1 risk haplotype, particularly

Correspondence: Dr D Collier, Department of Psychological Medicine, The Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF, UK.
E-mail: d.collier@iop.kcl.ac.uk
Received 05 August 2003; revised 04 November 2003; accepted 26 November 2003

when patients with a family history were analysed,[8] and analysis of a Han Chinese population using one SNP from Stefansson *et al*,[6] and two others selected at random was also positive.[9] Thus, evidence is accumulating that NRG1 may be a genuine risk gene for schizophrenia. However, association is restricted to haplotypes, and the underlying functional variant(s) that give rise to disease risk have not yet been identified.

One means to approach the identification of risk alleles is to examine diverse geographic groups to search for common patterns of association. Under the hypothesis that the NRG1-associated haplotype is a surrogate marker for a susceptibility polymorphism then, depending on the age and nature of the underlying risk alleles, a different associated haplotype would be expected for different ethnic groups, even though the risk allele(s) may be the same. If there are a large number of variants with substantial allelic heterogeneity at disease-causing loci, then there may be many risk haplotypes each having weak association with disease. Analysis of diverse ethnic groups may help to answer these questions, not just through simple replication but also through confirmation of the region of the gene in which schizophrenia susceptibility lies. In this study, we have investigated a Han Chinese population using both a family trio design and a case–control design, to determine if the haplotype found in Iceland and Scotland is present in the Chinese, and if not, whether alternative risk haplotypes can be identified in the gene.

Materials and methods

Patient sample

Schizophrenia patients, their families and control subjects were recruited from Sichuan Province, SW China. The family trio design used 184 trios (parents and one affected offspring) and 138 sibling pairs (parents and at least two affected siblings). The case–control design used 298 unrelated patients from the families above and 336 normal Han Chinese controls. All patients were interviewed by an experienced psychiatrist using the SCID. Diagnosis was made according to DSM-III-R or DSM-IV criteria (American Psychiatric Association). Information was also collected from examination of medical records and all other available sources. Genomic DNA was extracted from peripheral blood according to standard phenol–chloroform methods. This study was approved by the South London and Maudsley Trust ethical committee. Informed consent was obtained from all patients and control individuals and DNA samples were anonymised.

Haplotype estimation

We estimated haplotype frequencies for unrelated patients and unrelated controls. Two sets of control haplotypes were created: one for unrelated controls and the other for nontransmitted parental alleles for patients. To handle missing genotypes and the lack of familial information to derive the phase, our own implementation of a likelihood approach, using the expectation-maximization (EM) algorithm[10] as a computational tool, was applied to estimate the haplotype frequencies. Under the null hypothesis, the affected individuals and controls are assumed to have identical frequencies of all haplotypes. Under the alternative hypothesis, the candidate at-risk haplotype is allowed to have a higher frequency in affected individuals than controls, while the ratios of the frequencies of all other haplotypes are assumed to be the same in both groups. Likelihoods are maximized separately under both hypotheses and corresponding one-degree-of-freedom likelihood ratio statistics are used to evaluate statistical significance. While our own computer program was developed to fit our chosen models, and to handle missing genotypes and haplotypes with many markers efficiently, our use of the EM-algorithm is very similar to methods used by others.[11–13] Although applied in a slightly different setting, the one-degree-of-freedom model we use is essentially that used by Clayton and Jones.[14,15] The method used here is identical to that used in Stefansson *et al*.[6]

Genotyping

SNPs: SNP8NRG221533, SNP8NRG241930 and SNP8NRG243177 were scored using a fluorescent-based method.[16] SNPNRG112232 is not significantly polymorphic in the Chinese and is thus not included in the analysis. The following microsatellite markers spanning the NRG1 gene were genotyped as described by Gretarsdottir *et al*:[17] D8S1770, D8S1769, 29H12-1, D8S1711, 29H12-121L21, 478B14-642, 478B14-848, 487-2, 420M9-1395, 420M9-1, D8S1810, 473C15-439, 72H22-1, 82H10-79B8, 244L21-8557, 225C17-4, 317J8-2123, 317J8-1, 317J8-2, 317J8-4858, S8S4792, S8S1765. Amplimers can be found on deCODE Genetics Web site and detailed sequence information in GenBank (AF491780) and TPA (Third party annotation) (BK000383).

LD mapping

In order to study the LD with microsatellites, an extension to the definitions of D' [18] and its statistical signicance for biallelic markers were utilized. The extended D' is averaged over all the possible allele combinations of the two markers D', weighted by the marginal allele probabilities.[19] The corresponding P-value is defined as the minimum P-value for the pair of markers over the same combinations, provided the joint probability is higher than 0.05. Plotting for all marker combinations, D' in the upper-left corner and the P-value in the lower-right corner suggest the LD structure of the region (see Figure 2).

Results

Genotypes were successfully obtained for 25 polymorphisms: 22 microsatellites and three SNPs, across

the NRG1 gene in family trios (184 affected schizophrenic patients and their parents) and sibling pairs (138 schizophrenic families with at least two affected siblings and their parents). The positions of markers genotyped in the exon 1 region of the sample are shown in Figure 1.

Genotypes from 298 unrelated schizophrenia patients from the above families could be used in the association analysis. Two sets of controls were used in this study: a set of unrelated Han Chinese controls ($N=336$) and a set of nontransmitted parental haplotypes ($N=552$ chromosomes). We did a case–control study first and then used the nontransmitted parental haplotypes as control haplotypes, enabling us to compare directly the frequency of the nontransmitted haplotypes to their frequencies in the case–control study.

Despite the fact that almost all of the individual markers found in the extended Icelandic risk haplotypes called HapA, HapB, HapC1 and HapC2[6] are polymorphic in the Chinese population studied here, we were not able to detect the extended haplotypes in the Chinese population. However, the frequency of a seven-marker at-risk haplotype, 290 kb, at the 5′ end of the NRG1 gene where the extended haplotypes coincide (same alleles for the seven markers) was estimated. As described in Stefansson *et al*,[6] not all of the seven markers have to be genotyped to identify the haplotype due to linkage disequilibrium (LD) within the 290 kb region. Five out of the seven markers were genotyped and the frequency of the seven-marker haplotype estimated 95 described in Materials and methods. The *P*-values are two-sided, without adjustment for multiple comparisons, and are based on the EM algorithm. Single point *P*-values and risk ratio (rr) are given in Table 1 for the five genotyped markers of the seven-marker haplotype as well as for the five-marker haplotype, a good surrogate for the seven-marker haplotype. Neither the haplotype nor alleles for individual markers associating with the haplotype were found in excess in this Chinese population (Table 1).

Even though the seven-marker at-risk haplotype is not in excess in the Chinese patients, other candidate at-risk haplotypes may exist. We analysed the microsatellite haplotypes within the boundaries identified by HapA, HapB and HapC, that is, including the first 5′ exon of NRG1 and upstream sequence using both unrelated controls haplotypes and nontransmitted haplotypes from parents of patients. We searched for haplotypes overlapping the NRG1 locus region using the two different control groups and identified three interesting haplotypes called $HAP_{China\ 1}$, $HAP_{China\ 2}$ and $HAP_{China\ 3}$. The *P*-values given are two-sided and uncorrected for multiple testing.

Using unrelated patients we find a haplotype ($HAP_{China\ 1}$), immediately upstream of the Icelandic at-risk haplotype, in excess in patients (Figure 1). The frequency of the $HAP_{China\ 1}$ is 11.9% in patients and 4.2% in controls (Table 2). This gives a *P*-value of 0.0000065 and a relative risk of 3.1. When we use the nontransmitted parental haplotypes as controls, the control frequency changes from 4.2 to 9.2% and the uncorrected *P*-value increases from 0.0000065 to 0.12.

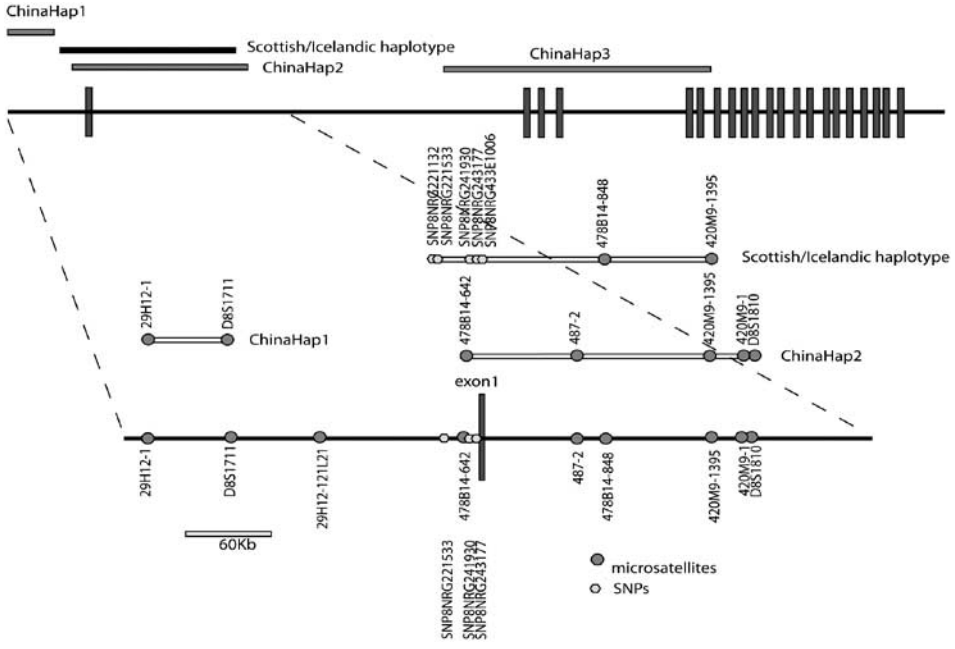

Figure 1 Marker and haplotype positions in the NRG1 gene. In the top panel, the 21 currently known exons for NRG1 are shown together with the positions of the previously published Scottish/Icelandic haplotype (dark bar) and the three Chinese haplotypes described in the current study (light bars). In the lower section, the exon 1 region containing $HAP_{China\ 1}$ and $HAP_{China\ 2}$ is shown, together with the markers genotyped in the present study. SNPs are indicated by filled hexagons and microsatellite markers by filled circles.

Table 1 Allele frequency for five markers from a seven-marker haplotype found in excess in schizophrenia patients in Iceland, Scotland and UK

	Percent frequency, significance and risk ratio (rr) for unrelated patients and controls			
Markers	*Patients*	*Control*	P	*rr*
SNP8NRG221533	55.2 (285)	59.1 (250)	0.22	0.85
SNP8NRG241930	95.7 (290)	94.7 (259)	0.48	1.2
SNP8NRG243177	58.9 (292)	63.2 (256)	0.16	0.83
Three-marker SNP haplotype[a]	53.4 (292)	55.9 (267)	0.42	0.90
Five-marker haplotype[b]	6.8 (292)	9.9 (276)	0.10	0.67

Estimates for three-marker and five-marker haplotypes are also given. The five-marker haplotype is a good surrogate haplotype for the seven-marker haplotype.
[a]The three-marker haplotype consists of SNP markers only. Marker order and names are as follows: SNP8NRG221533, SNP8NRG241930, SNP8NRG243177. Alleles for the SNP haplotype are shown for individual SNPs in the table.
[b]The five-marker haplotype consists of the following markers in this order: SNP8NRG221533, SNP8NRG241930, SNP8NRG243177, 478B14-848 and 420M9-1395. Markers 478B14-848 and 420M9-1395 are microsatellite markers. The other three markers are SNP markers. For patients, the number of genotyped individuals is given within parentheses. Control frequency is here estimated from nontransmitted parental haplotypes and the number in parentheses refers to the number of patients for whom nontransmitted parental haplotypes could be estimated.

Table 2 Frequency, significance and rr of the microsatellite haplotypes associated with schizophrenia in the Han Chinese population

		Frequency (%) and risk for unrelated patients and unrelated controls				*Frequency (%) and risk for unrelated patients and nontransmitted parental alleles as controls*			
Markers	*Allele*	*Patients*	*Control*	P	*rr*	*Patients*	*Control*	P	*rr*
$HAP_{China\ 1}$		11.9(231)	4.2(271)	0.0000065	3.1	11.7	9.2(205)	0.12	1.3
29H12-1	6	14.6(267)	8.3(307)	0.00038	1.9	14.6	13.2(231)	0.26	1.1
D8S1711	−2	46.8(263)	42.9(297)	0.10	1.2	46.8	42.5(221)	0.10	1.2
$HAP_{China\ 2}$		8.5(298)	4.0(336)	0.0030	2.2	7.6	3.8(269)	0.0047	2.1
478B14-642	4	33.4(292)	28.6(325)	0.035	1.3	33.2	27.8(253)	0.025	1.3
487-2	20	65.8(279)	65.1(311)	0.41	1.0	65.9	63.4(240)	0.020	1.1
420M9-1395	−2	67.9(238)	66.8(312)	0.36	1.0	65.9	63.4(240)	0.020	1.1
D8S1810	16	58.0(257)	54.7(296)	0.14	1.1	58.2	52.2(198)	0.037	1.3
$HAP_{China\ 3}$		26.9(297)	25.3(333)	0.29	1.1	23.8	13.7(269)	0.000042	2.0
317J8-2123	4	77.6(277)	78.0(318)	0.44	1.0	77.9	72.6(232)	0.025	1.3
317J8-1	0	62.3(280)	64.5(306)	0.22	0.91	62.5	62.5(234)	0.50	1.0
317J8-2	2	65.4(279)	65.8(301)	0.45	1.0	65.4	61.7(240)	0.11	1.2
317J8-4858	0	58.5(266)	52.9(323)	0.03	1.3	58.5	48.5(242)	0.00095	1.5

Parentheses show the number of patients with each haplotype. Both case–control and family-based analyses are shown. $Hap_{China\ 1}$ was chosen from the case–control study as the most significant haplotype associating with the disease. $Hap_{China\ 3}$ was the most significant haplotype when nontransmitted parental haplotypes were used as control haplotypes. $Hap_{China\ 2}$ is found in excess both when controls from the case–control study were used and when nontransmitted parental alleles were used as controls.

Apart from chance differences, population stratification is one likely explanation for this difference between case–control and family-based analysis.

Another haplotype overlapping the seven-marker at-risk haplotype found in Iceland is found in excess in the Chinese population studied here. This haplotype ($HAP_{China\ 2}$; Figure 1) is found in 8.5% of patients and 4.0% of sporadic controls. $HAP_{China\ 2}$ is in excess whether the nontransmitted parental at-risk haplotypes are used as controls or the population controls are used ($P=0.003$, rr 2.2 and $P=0.0047$, rr 2.1, respectively). The difference in frequency for $HAP_{China\ 2}$ between patients and controls cannot, therefore, be explained by population stratification since nontransmitted haplotypes give similar results as the unrelated controls.

When we look at other microsatellite haplotypes in the large NRG1 gene downstream of the region harbouring the at-risk haplotype, they show less significant uncorrected *P*-values than haplotype $HAP_{China\ 1}$ when unrelated controls are used. When nontransmitted haplotypes are used, however, a

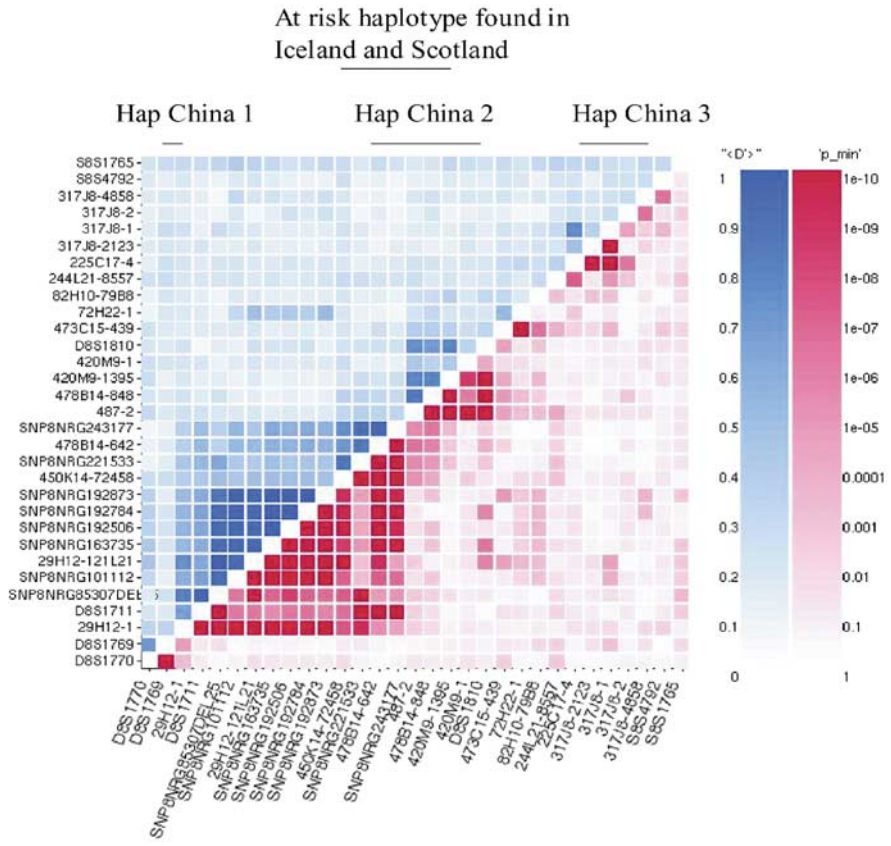

Figure 2 LD relationships between microsatellite markers in the NRG1 gene in the Chinese population under study. Two measures of LD are shown: D' values on the upper-left side of the plot and *P*-values on the lower-right side. Scales for the LD strength are provided for both measures to the right. Note the existence of a haplotype block at the 5′ end of the NRG1 gene, which is flanked by $HAP_{China\ 1}$ and $HAP_{China\ 2}$. The relative positions of the Chinese associated haplotypes and the Icelandic/Scottish haplotypes are illustrated above the figure. The strongest block of LD in $HAP_{China\ 2}$ was between the three microsatellites D8S1810, 487-2 and 420M9-1395 ($D' > 0.9$).

four-marker haplotype, $HAP_{China\ 3}$ was found in 23.8% frequency in patients and 13.7% in nontransmitted parental haplotypes ($P = 0.000042$, $rr = 2.0$). Most of the NRG1 exons are located within this haplotype at the 3′ end of the gene (Figure 1). The extent of LD across the region is illustrated in Figure 2. The five markers in the $HAP_{China\ 2}$ haplotype all show significant LD with each other, with the strongest LD relationships between the three microsatellites: D8S1810, 487-2 and 420M9-1395. LD between the 5′ marker 478B14-642 and this group of three microsatellites was much lower ($D' < 0.2$) although still significant. Note the existence of a haplotype block at the 5′ end of the NRG1 gene, which is flanked by $HAP_{China\ 1}$ and $HAP_{China\ 2}$ (Figure 2).

Discussion

The NRG1 gene was originally identified as a susceptibility gene for schizophrenia by using a combination of linkage and association approaches, which identified a seven-marker risk haplotype consisting of five SNPs and two microsatellites.[6] This haplotype was also found in excess in schizophrenia patients in Scotland[7] and the UK[8] populations sharing recent ancestry with the Norse and Gaelic-derived Icelandic population, and a different haplotype in a Han Chinese population.[9] In the present study, we genotyped five of the seven markers from the at-risk haplotype and estimated the frequency of the five-marker haplotype, a good surrogate for the seven-marker haplotype. The five-marker haplotype (surrogate for the seven-marker haplotype) is not associating with schizophrenia. In fact, the frequency is higher in the control group (Table 1). Unlike the Han Chinese sample, we did not see an excess of the SNP SNP8NRG221533 in our case–control analysis and the allele is not transmitted more often than expected to patients. The reason as to why the association is found to the SNP8NRG221533 allele in Yang *et al*[9] study but not in our study is not clear, and will require further analysis in other samples from China.

The over-riding extended haplotypes, HapA, HapB, HapC1 and HapC2 could not be identified in the Chinese population described here probably as they are spanning several LD blocks. Instead we have identified three different potentially associated

haplotypes, which we call $HAP_{China\ 1}$, $HAP_{China\ 2}$ and $HAP_{China\ 3}$. The first two of these are at the 5′ end of the gene and the third spans most of the NRG1 exons at the 3′ end of the gene. The $HAP_{China\ 1}$ was associated with schizophrenia in the case–control analysis but not the family-based analysis, indicating the possibility of population stratification in the controls, whereas $HAP_{China\ 3}$ was in excess in patients only in the family-based analysis. However, $HAP_{China\ 2}$ was in excess in patients in both analysis, and furthermore is in the same position as the Icelandic/Scottish risk haplotype, albeit with a different set of markers and alleles. The family-based approach may be a conservative one, whereas the use of unrelated controls as the comparison group is sensitive to population stratification, which may lead to false-positive association.

Our failure to identify the same Icelandic/Scottish schizophrenia risk haplotype in the Chinese is perhaps unsurprising since Northern European and Chinese population are separated by tens of thousands of years, and a different relationship between risk haplotypes and underlying pathogenic variants may have evolved during this time. It is possible that the same as-yet-unidentified disease-causing genetic variant(s) exist in both the Chinese and European populations but are associated with different overlying haplotypes. It is also possible that there is locus heterogeneity at the NRG locus, with several risk alleles that differ in occurrence and frequency between geographic populations, or even multiple rare alleles associated with various haplotypes that will hamper attempts at replication.

The causative mutation in NRG has not yet been identified since no individual polymorphisms show strength of association comparable to haplotypes. The apparent absence of changes in the amino-acid structure of the genes[6] implies a noncoding polymorphism with regulatory effects, which could lie anywhere in a considerable region within or surrounding the gene. Most causative alleles will lie on associated haplotypes because of the simplified haplotype block structure of much of the genome.[20] Consequently, it will be difficult to distinguish between a causative polymorphism and association through LD by genetic methods alone. In addition, the extent of conservation of haplotypes and their block structure across ethnic groups is not clear, since factors such as genetic drift, selection, mutation, nonrandom mating combined with the dependence of haplotypes on allele frequency could have profound effects on haplotype structures.[20] The analysis of different ethnic groups, combined with hierarchical modelling of LD[21] may be the most useful approach to the identification of the underlying pathogenic alleles.

Analysis of haplotypes across an extended region of the genome such as the NRG locus involves the analysis of many combinations of markers, and even though the actual number of haplotypes seen is fewer than the number theoretically possible, the statistical effect of performing multiple tests combined with phase uncertainty is a problem.[14,15] However, frequencies for 2–5-marker haplotypes were estimated from the 25 genotyped markers spanning several LD blocks and multiple testing has not been accounted for. Although the analysis of families helps with phase estimation, none of the microsatellite haplotypes described here are significantly associated with the disease after correcting for multiple testing. Furthermore, difference in haplotype frequency between nontransmitted parental haplotypes and population controls is of concern and stratification between our patients and controls can therefore not be excluded. However, we have identified three haplotypes showing the strongest association in this cohort and other research groups using Asian populations can test these. If relative risk for an at-risk haplotype is as low as 2.0–2.5, the association is difficult to detect unambiguously without direct replication in another cohort.

Acknowledgements

This work was partly funded by NSFC (China), NARSAD, the Psychiatry Research Trust and the Wellcome Trust.

Electronic-database information

Accession Numbers and URLs for data in this article are as follows:

- deCODE genetics http://www.decode. com/nrg1/ markers for SNPs and microsatellite markers in the NRG1 locus sequence;
- GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ for NRG1 (AF491780);
- Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for NRG1 (MIM 142445), SCZD (MIM 181500) and SCZD6 (MIM 603013));
- Third Party Annotation, DDBJ/EMBL/GenBank databases for NRG1 (TPA: BK000383).

References

1 Cardno AG, Gottesman II. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *Am J Med Genet* 2000; **97**: 12–17.
2 Tabor HK, Risch NJ, Myers RM. Opinion: candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* 2002; **3**: 391–397.
3 Badner JA, Gershon ES. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry* 2002; **7**: 405–411.
4 Lewis CM, Levinson DF, Wise LH, Delisi LE, Straub RE, Hovatta I *et al.* Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* 2003; **73**: 34–48.
5 Harrison PJ, Owen MJ. Genes for schizophrenia? Recent findings and their pathophysiological implications. *Lancet* 2003; **361**: 417–419.
6 Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S *et al.* Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet* 2002; **71**: 877–892.

7 Stefansson H, Sarginson J, Kong A, Yates P, Steinthorsdottir V, Gudfinnsson E *et al.* Association of neuregulin 1 with schizophrenia confirmed in a Scottish population. *Am J Hum Genet* 2003; **72**: 83–87.
8 Williams NM, Preece A, Spurlock G, Norton N, Williams HJ, Zammit S *et al.* Support for genetic variation in neuregulin 1 and susceptibility to schizophrenia. *Mol Psychiatry* 2003; **8**: 485–487.
9 Yang JZ, Si TM, Ruan Y, Ling YS, Han YH, Wang XL *et al.* Association study of neuregulin 1 gene with schizophrenia. *Mol Psychiatry* 2003; **8**: 706–709.
10 Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B* 1977; **39**: 1–38.
11 Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995; **12**: 921–927.
12 Hawley M, Kidd K. HAPLO: a program using the EM algorithm to estimate the frequencies of multi-site haplotypes. *J Hered* 1995; **86**: 409–411.
13 Long JC, Williams RC, Urbanek M. An E-M algorithm and testing strategy for multiple-locus haplotypes. *Am J Hum Genet* 1995; **56**: 799–810.
14 Clayton D, Jones H. Transmission/disequilibrium tests for extended marker haplotypes. *Am J Hum Genet* 1999; **65**: 1161–1169.
15 Clayton D. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Am J Hum Genet* 1999; **65**: 1170–1177.
16 Kwok PY. SNP genotyping with fluorescence polarization detection. *Hum Mutat* 2002; **19**: 315–323.
17 Gretarsdottir S, Sveinbjornsdottir S, Jonsson HH, Jakobsson F, Einarsdottir E, Agnarsson U *et al.* Localization of a susceptibility gene for common forms of stroke to 5q12. *Am J Hum Genet* 2002; **70**: 593–603.
18 Lewontin RC. The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics* 1964; **49**: 46–49.
19 Hedrick PW. Gametic disequilibrium measures: proceed with caution. *Genetics* 1987; **117**: 331–341.
20 Cardon LR, Abecasis GR. Using haplotype blocks to map human complex trait loci. *Trends Genet* 2003; **19**: 135–140.
21 Conti, Witt JS. Hierarchical modelling of linkage disequilibrium: genetic structure and spatial relations. *Am J Hum Genet* 2003; **72**: 351–363.