Molecular Psychiatry (2004) 9, 208–212
© 2004 Nature Publishing Group All rights reserved 1359-4184/04 $25.00
www.nature.com/mp

ORIGINAL RESEARCH ARTICLE

Confirmation and refinement of an 'at-risk' haplotype for schizophrenia suggests the EST cluster, *Hs*.97362, as a potential susceptibility gene at the Neuregulin-1 locus

AP Corvin*[,1], DW Morris*[,1], K McGhee[1], S Schwaiger[1], P Scully[2], J Quinn[2], D Meagher[2], D St Clair[3], JL Waddington[2,4] and M Gill[1]

[1]*Neuropsychiatric Genetics Group, Department of Psychiatry, Institute of Molecular Medicine, Trinity College, Dublin 2, Ireland;* [2]*Stanley Research Unit, St Davnet's Hospital, Monaghan, Ireland;* [3]*Aberdeen Royal Infirmary and University of Aberdeen Medical School, Aberdeen, Scotland, UK;* [4]*Department of Clinical Pharmacology, Royal College of Surgeons, 123 St Stephen's Green, Dublin 2, Ireland*

Two recent association studies have implicated the neuregulin-1 gene (*NRG1*) at chromosome 8p21–22 as a susceptibility gene for schizophrenia. Stefansson *et al* identified three 'at-risk' haplotypes (HapA, B and C) which spanned the *NRG1* locus and shared a common core haplotype. Subsequently, they demonstrated evidence that the core haplotype was associated with schizophrenia in an independent Scottish sample. To confirm and refine this haplotype we investigated the *NRG1* locus in an independent Irish case–control sample. We did not find the core haplotype to be associated in our sample. However, we identified a refined 2-marker haplotype ($HapB_{IRE}$) that shared common alleles with one of the Icelandic 'at-risk' haplotypes and is in significant excess in the Irish cases (19.4%) *vs* controls (12.3%) ($P=0.013$). This refined 'at-risk' haplotype is also in significant excess in the Scottish case sample (17.0% *vs* 13.5%; $P=0.036$). Interestingly, this refined 'at-risk' haplotype is positioned close to an EST cluster of unknown function (*Hs*.97362) within intron 1 of *NRG1*.
Molecular Psychiatry (2004) **9,** 208–212. doi:10.1038/sj.mp.4001412

Keywords: schizophrenia; neuregulin-1; risk haplotype; association; linkage disequilibrium

Schizophrenia (OMIM 181500; http://www.ncbi.nlm.gov/Omim/) is a chronic psychiatric disorder affecting 0.5–1% of the general population. Expression of the schizophrenia phenotype is likely to involve interaction of several susceptibility genes with each other and with environmental risk. Identifying susceptibility genes and molecular pathways contributing to schizophrenia may be a powerful method of understanding schizophrenia pathobiology and developing new treatment strategies. Recently, Stefansson *et al*[1] reported *neuregulin-1* (*NRG1*) as a candidate gene associated with schizophrenia in an Icelandic population. Since then the authors have confirmed this finding in an independent Scottish case–control sample.[2]

Stefansson *et al*[1] identified *NRG1* as a positional and functional candidate gene for schizophrenia. They established association with *NRG1* by fine mapping a chromosome 8p21–22 locus that had shown suggestive evidence of linkage in a schizophrenia genome scan of 33 Icelandic families. Previously, four schizophrenia genome scans,[3–6] including a sample from the Irish population,[3] had reported suggestive evidence of a chromosome 8p susceptibility locus. As has been the case with other complex genetic disorders, this linkage finding has not been universally replicated, but a recent meta-analysis of 20 schizophrenia genome scans has provided significant statistical support for chromosome 8p as a schizophrenia susceptibility locus.[7]

As the aetiology of schizophrenia is obscure, *a priori* evidence for *NRG1* as a functional candidate gene is weak. By necessity, current hypotheses of schizophrenia pathobiology are broad and the prevailing view is that schizophrenia is a neurodevelopmental disorder which impacts on CNS maturation. This may be mediated through abnormal cortical wiring or dysregulation of synaptic function with secondary downstream effects on neurotransmitter systems.[8] *NRG1* isoforms may play a role at each stage in this process through influencing gliogenesis and neuronal migration during CNS development or by inhibiting long-term potentiation (a form of synaptic plasticity;).[9] *NRG1* also regulates the development and function of cortical GABAerbic interneurones that have been implicated in schizophrenia.[10] However, many other CNS expressed genes may have similar effects and be equally plausible candidates.

Correspondence: Dr AP Corvin, Neuropsychiatric Genetics Group, Department of Psychiatry, Institute of Molecular Medicine, Trinity Centre for Health Sciences, St. James Hospital, Dublin 8, Ireland. E-mail: acorvin@tcd.ie

*Contributed equally to this work.

Received 16 May 2003; revised 11 July 2003; accepted 12 July 2003

Stefansson *et al*[1] identified a core haplotype ~191 kb in length and this finding was replicated in an independent sample from the Scottish population[2]. There are historical connections between these populations and the Irish population; the founder population of Iceland was of Norse and Gaelic ancestry.[11] A study in the Irish population was the first to report linkage to the 8p chromosomal susceptibility locus which contains *NRG1*. Thus, an independent case–control sample from the larger, related and more out-bred Irish population might have the capacity both to test the finding of association in this region, and refine the location of the putative susceptibility variants.

The sample consisted of 243 cases and 222 controls from the Republic of Ireland. Ethics Committee approval for the study was obtained from all participating hospitals and centres. All cases gave written informed consent and were interviewed by a psychiatrist or psychiatric nurse trained to use the Structured Clinical Interview for DSM (SCID). Diagnosis was based on DSM-IV criteria using all available information (interview, family or staff report and chart review). All cases were over 18 years of age, of Irish origin and had been screened to exclude substance-induced psychotic disorder or psychosis due to a general medical condition. The patients were diagnosed with DSM-IV defined schizophrenia (n=196) or schizoaffective disorder (n=47). The control sample, drawn from Irish blood donors, was not specifically screened for psychiatric illness but individuals were not taking regular prescribed medications.

A second European sample of 609 unrelated patients, diagnosed with DSM-IIIR defined schizophrenia or schizo-affective disorder was available to investigate the genetic findings detailed in this paper. This sample, ascertained in Scotland, was ethnically matched with 618 controls drawn from the same population. Informed consent was obtained from all patients and control individuals. Further detail on ascertainment and diagnosis of this Scottish sample has been reported previously by Stefansson *et al.*[2]

In total, 17 microsatellite markers and three SNPs were genotyped in our sample. These markers spanned a region of ~911 kb, which covers the putative susceptibility locus identified by Stefansson *et al.*[1] The names and relative positions of all 20 markers are detailed in Table 1. All samples were genotyped for the above markers by deCODE Genetics using protocols previously outlined.[1] A strict sample handling protocol was employed and deCODE remained blind to the affection status of the samples until all genotypic data were returned to Dublin and statistical analysis was complete.

Table 1 Results (simulated *P*-values) from association analyses of single markers and of 2-, 3- and 4-marker haplotypes constructed from adjacent markers

#	*Marker name*	*Intermarker distances (kb)*	*Single markers*	*2-Marker haplotypes*	*3-Marker haplotypes*	*4-Marker haplotypes*
M1	D8S1770	—	0.369	0.250[b]	0.043[c]	0.201[d]
M2	29H12-1	128.9[e]	0.349	0.059	0.029	0.067
M3	29H12-7320	5.4	0.116	0.368	0.606	0.914
M4	D8S1711	49.8	0.419	0.638	0.472	0.498
M5	29H12-121L21	65.1	0.278	0.163	0.572	0.884
M6	450K14-72458	79.0	**0.039**	0.552	0.570	0.233
M7	*SNP8NRG221533*[f]	5.7	0.724	0.419	0.432	0.387
M8	478B14-642	18.8	0.177	0.225	0.217	0.179
M9	*SNP8NRG241930*[f]	1.4	0.546	0.109	0.821	0.791
M10	*SNP8NRG243177*[f]	1.2	0.244	0.699	0.652	0.643
M11	487-2	74.9	0.753	0.696	0.643	0.505
M12	*478B14-848*	18.4	0.749	0.307	0.157	0.166
M13	*420M9-1395*	76.6	0.106	0.183	**0.037**	**0.006**
M14	420M9-1	22.4	0.274	0.174	0.082	0.220
M15	D8S1810	3.9	0.133	**0.045**	0.118	**0.004**
M16	420M9-116I12	33.8	0.521	0.710	0.169	0.236
M17	473C15-533	108.6	0.380	0.126	0.205	0.573
M18	473C15-439	66.2	**0.036**	0.341	0.370	—
M19	72H22-1	36.2	0.058	0.347	—	—
M20	72H22-36	114.2	0.919	—	—	—

[a]Individual allele frequencies for all markers are not displayed. These are available on request.[b]Simulated *P*-value for 2-marker haplotype analysis of markers M1 and M2. Data are presented in the same format for each adjacent pair of markers down this column.[c]Simulated *P*-value for 3-marker haplotype analysis of markers M1, M2 and M3. Data are presented in the same format for each set of three adjacent markers down this column.[d]Simulated *P*-value for 4-marker haplotype analysis of markers M1, M2, M3 and M4. Data are presented in the same format for each set of four adjacent markers down this column.[e]Represents the distance between M1 and M2. Data are presented in the format for each adjacent pair of markers down this column.[f]SNP markers; *P*-values for single marker analysis of SNPs are not simulated.Markers in italics are those that define the Icelandic core haplotype.

SNPs were tested for association with the phenotype by using a 2×2 contingency table to calculate a χ^2 statistic. Microsatellite markers were tested for association using the T2 χ^2 statistic of CLUMP and significance was assessed using Monte Carlo simulations.[12] Intermarker linkage disequilibrium (LD) was measured using D';[13] the D' values were calculated using GOLD (http://www.well.ox.ac.uk/asthma/GOLD/index.html). Haplotype association analysis was performed using FASTEHPLUS.[14] This program employs the gene-counting algorithm, a form of the EM algorithm, to calculate haplotype frequencies. FASTEHPLUS is a global test of association for each combination of markers chosen. *P*-values are calculated by simulation. A separate program, GENECOUNTING, was used to output individual haplotype frequencies from cases and from controls[15] for the purpose of testing specific haplotypes. All markers were tested for Hardy–Weinberg equilibrium (HWE) in both case and control samples. This analysis was performed using FSTAT (http://www.unil.ch/izea/softwares/fstat.html) and confirmed that all markers were in HWE (simulated $P > 0.01$). These data and individual allele frequencies at each marker locus are available from our website (see web tables 1–21; http://www.tcd.ie/psychiatry/neuropsychiatry/nrg1.pdf).

We first tested for evidence of association with the core 'at-risk' haplotype as defined by Stefansson *et al*.[1] This haplotype is represented by alleles 1, 0 and 0 from markers SNP8NRG221533 (M7), 478B14-848 (M12) and 420M9-1395 (M13), respectively. In our study, the frequency of the *1 0 0* haplotype was 9.6% in cases and 8.4% in controls (odds ratio (OR)=1.18 (95% confidence interval (95% CI) 0.67–2.09), P=0.534). D' values for all marker combinations were calculated for cases and controls. Both case and control data sets produce a similar pattern of LD across the region. Two 'blocks' of markers (450K14-72458 (M6) to SNP8NRG243177 (M10); 487-2 (M11) to 420M9-116I12 (M16)) exhibit strong intermarker LD, separated by an area of low LD or higher recombination. This is graphically represented for our case sample in Figure 2a. The Icelandic core haplotype extends across these two blocks. Haplotype structure may differ between populations; hence, we investigated whether a different pattern of association was present in the Irish sample.

Therefore, we conducted systematic association analyses using all 20 markers genotyped across this region. The results of these single marker association analyses are detailed in Table 1. Two markers show evidence of association, 450K14-72458 (M6; P=0.039) and 473C15-439 (M18; P=0.036). Next, association analyses were performed using haplotypes constructed from adjacent markers in blocks of 2, 3 or 4 markers in a sliding-window fashion across the region. The results (see Table 1) suggest evidence of association in two sections of the region under investigation. The strongest evidence came from 2-, 3- and 4-marker haplotypes in the region of markers 420M9-1395 (M13) to 473C15-439 (M18). These markers lie within the second haplotype 'block' and overlap with the 3′ end of the Icelandic core haplotype.

This led us to investigate whether haplotype associations in our sample shared common allelic structure with the Icelandic 'risk' haplotypes (HapA, B and C; see Figure 1). In such an instance further refinement of 'risk' haplotypes could be possible. To test this hypothesis we specifically examined combinations of markers that had been genotyped in both samples and shared common alleles at individual loci. An intriguing finding from this analysis was for a haplotype constructed from microsatellite

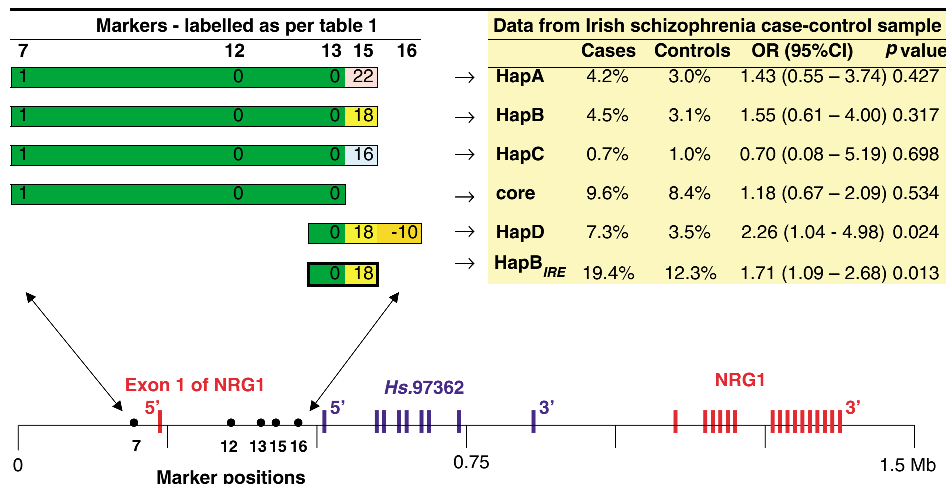

Figure 1 A graphic display of the truncated forms of the 3 'at-risk' haplotypes (HapA, B and C) and core 'at-risk' haplotype identified by Stefansson *et al*.[1] The physical positioning of these haplotypes are indicated by black arrows relative to *NRG1* (exons displayed in red) and the EST cluster *Hs*.97362 (exons displayed in blue). The frequencies of these haplotypes in the case and control samples from the Irish population are detailed. In addition, we present the haplotype frequencies of HapB_{IRE} and HapD that were found to be associated in the Irish sample.

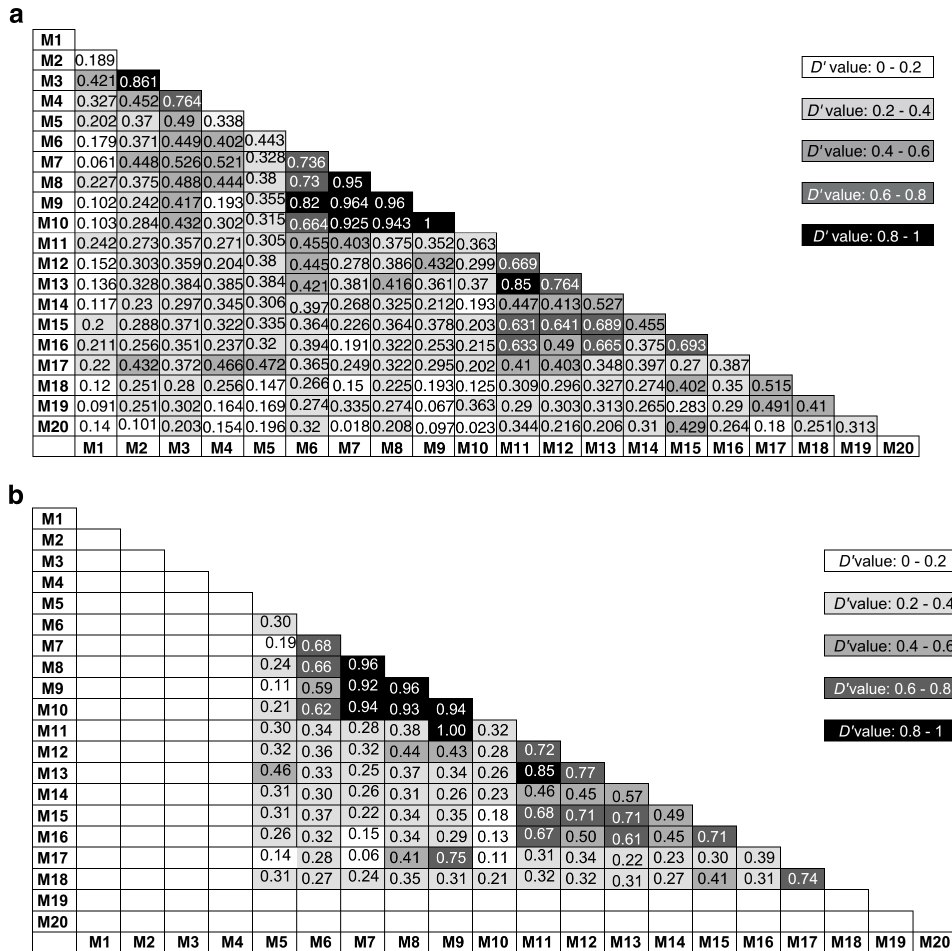

Figure 2 Intermarker LD measured using D' in the Irish schizophrenia case sample (a) and in the Icelandic schizophrenia case sample (b).

markers 420M9-1395 (M13) and D8S1810 (M15). The *0 18* haplotype from markers M13 and M15 is in significant excess when patients are compared with controls (19.4% *vs* 12.3%; OR 1.71 (95% CI 1.09–2.68), P=0.013). This 2-marker haplotype shares the same alleles at markers 420M9-1395 (M13) and D8S1810 (M15) as HapB in the Icelandic sample (see Figure 1). We term this refined 'at-risk' haplotype HapB_{IRE}.

This study provides additional support for a susceptibility locus for schizophrenia on chromosome 8p21–22 by replicating the finding of association previously reported[1,2] in an independent case–control sample from the Irish population. We have identified a refined associated haplotype (HapB_{IRE}) that overlaps with one of the at-risk haplotypes (HapB) reported by Stefansson *et al.*[1] To confirm this finding we have tested for evidence of association with this haplotype in the Scottish schizophrenia sample described by Stefansson *et al.*[2] We again find an excess of HapB_{IRE} in schizophrenia cases compared to controls (17.0% *vs* 13.5%; OR 1.31 (95% CI 1.01–1.70), P=0.036). However, it must be noted that there is not a significant excess of HapB_{IRE} in the Icelandic cases (20.3%) compared to controls (20.1%) (HStefansson, personal communication).

HapB_{IRE} is derived from two markers separated by a distance of ~99 kb. HapB_{IRE} is within the second of two haplotype 'blocks' that we have identified from Irish intermarker LD data (Figure 2a) and which extends from 487-2 (M11) to 420M9-116I12 (M16). However, it is of interest to note that at this marker resolution a similar pattern of LD is present in the Icelandic sample (Figure 2b). The Icelandic core haplotype extends across the low LD 'gap' between SNP8NRG243177 (M10) and 420M9-116I12 (M11) and includes all of the first haplotype 'block' between markers 450K14-72458 (M6) to SNP8NRG243177 (M10). The absence of association with the Icelandic core haplotype in our sample may be due to one of the two factors. Our study may have lacked power to detect association with the Icelandic core haplotype. A power calculation indicates that we had 65% power to detect an effect of this size at a significance level of 0.05. Alternatively, this may be due to differences in haplotype structure between samples. In a population isolate such as Iceland it is reasonable to expect that a particular haplotype or haplotypes

containing a susceptibility variant might bridge a region of low LD and be represented by an extended haplotype. However, it would be less easy to explain how, in a more out-bred population such as the Irish, the evidence of association in the second haplotype 'block' might be due to a susceptibility variant located in the first haplotype 'block' where no evidence of association exists in our sample. In the Scottish sample the same LD pattern was present and a significant excess of both the Icelandic core haplotype[2] and the $HapB_{IRE}$ haplotype was identified. By contrast, $HapB_{IRE}$ is not in significant excess in Icelandic cases, and this haplotype is more common in the control population. The complexity of haplotype structure is well recognised, as a result definitive interpretation of the three data sets is difficult and further investigation will be required. Our conclusion is that the susceptibility allele (or alleles) are within the second haplotype 'block' currently delineated by markers 487-2 (M11) and 420M9-116I12 (M16).

We attempted to further investigate the extent of the association in the region of $HapB_{IRE}$ using the additional markers typed in our sample. This identified a 3-marker haplotype (HapD) that extended 3' of $HapB_{IRE}$ and included 420M9-116I12 (M16; see Figure 1). This haplotype was in significant excess in our sample (7.3% in cases *vs* 3.5% in controls; OR 2.26 (95% CI 1.04–4.98), P=0.024), but not in the Scottish sample (5.2% in cases *vs* 5.8% in controls). From these data, we suggest that the most likely position of the causative variant(s) is in the region of $HapB_{IRE}$. Further analysis of the likely position of the susceptibility variant or variants would then be dependent on the extent of LD between it and surrounding markers. Were this region to display extensive LD, more exact positioning of a causative variant may require investigation in older populations where LD may have dissipated to a greater degree.

The refinement of the at-risk haplotype to the region of $HapB_{IRE}$ has potentially important implications for the identity of the putative susceptibility gene (see Figure 1). These markers are located between exon 1 of Glial Growth Factor 2 (*GGF2*), an alternatively spliced isoform of NRG1 and the putative first exon of an EST cluster (*Hs*.97362; http://www.ncbi.nlm.nih.gov/UniGene/). Exon 1 of *GGF2* is not known to be shared by any other isoforms of the NRG1 gene, and is located ~70 kb upstream of 487-2 (M11) and ~2 kb downstream of SNP8NRG243177 (M10). This probably positions this exon within the first haplotype 'block'. Although *GGF2* is widely expressed in the brain during development, *in vivo* function of this isoform has yet to be demonstrated.[9] The next *NRG1* exon is located ~633 kb downstream of 420M9-116I12 (M16). Positioned between 420M9-116I12 (M16) and the second *NRG1* exon is the EST cluster (*Hs*.97362). In addition to the 3 ESTs identified in UniGene, we have identified two further ESTs (BU567364 and BG202570) in the NCBI GenBank database that share homology with this cluster and map to this locus. Together these five ESTs identify nine putative exons, all positioned from ~30 to ~340 kb downstream of 420M9-116I12 (M16). As yet this EST cluster is of unknown function. However, given its close physical proximity to $HapB_{IRE}$, it warrants consideration in the search for the putative susceptibility variant or variants at this locus.

Acknowledgements

The authors gratefully acknowledge the contribution of all participating individuals and institutions in this study. Dr Corvin is a Wellcome Trust Research Fellow in Mental Health. The authors would like to thank the Health Research Board (Ireland), Science Foundation Ireland, Health Education Authority (Ireland) and the Stanley Medical Research Institute (USA) for their support of this research. The authors also gratefully acknowledge deCODE Genetics for their technical assistance.

References

1 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**: 77–892.
2 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.
3 Kendler KS, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F *et al.* Evidence for a schizophrenia vulnerability locus on chromosome 8p in the Irish Study of High-Density Schizophrenia Families. *Am J Psychiatry* 1996; **153**: 1534–1540.
4 Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G *et al.* Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. *Nat Genet* 1998; **20**: 70–73.
5 Kaufmann CA, Suarez B, Malaspina D, Pepple J, Svrakic D, Markel PD *et al.* NIMH Genetics Initiative Millennium Schizophrenia Consortium: linkage analysis of African-American pedigrees. *Am J Med Genet* 1998; **81**: 282–289.
6 Brzustowicz LM, Honer WG, Chow EW, Little D, Hogan J, Hodgkinson K *et al.* Linkage of familial schizophrenia to chromosome 13q32. *Am J Hum Genet* 1999; **65**: 1096–1103.
7 Levinson DF, Lewis CM, Wise LH. Meta-analysis of genome scans for schizophrenia. *Am J Med Genet* 2002; **114**: 700.
8 Weinberger DR. From neuropathology to neurodevelopment. *Lancet* 1995; **346**: 552–557.
9 Falls DL. Neuregulins: functions, forms, and signalling strategies. *Exp Cell Res* 2003; **284**: 14–30.
10 Yau HJ, Wang HF, C Lai, Liu FC. Neural development of the neuregulin receptor ErbB4 in the cerebral cortex and the hippocampus: preferential expression by interneurons tangentially migrating from the ganglionic eminences. *Cereb Cortex* 2003; **13**: 252–264.
11 Helgason A, Hickey E, Goodacre S, Bosnes V, Stefansson K, Ward R *et al.* mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* 2001; **68**: 723–737.
12 Sham PC, Curtis D. Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann Hum Genet* 1995; **59**: 97–105.
13 Lewontin RC. The interaction of selection and linkage I. General considerations: heterotic models. *Genetics* 1964; **49**: 49–67.
14 Zhao JH, Sham PC. Faster haplotype frequency estimation using unrelated subjects. *Hum Hered* 2002; **53**: 36–41.
15 Zhao JH, Lissarrague S, Essioux L, Sham PC. GENECOUNTING: haplotype analysis with missing genotypes. *Bioinformatics* 2002; **18**: 1694–1695.