Molecular Psychiatry (2010) 15, 319–325

www.nature.com/mp

ORIGINAL ARTICLE

Population-based linkage analysis of schizophrenia and bipolar case–control cohorts identifies a potential susceptibility locus on 19q13

C Francks[1], F Tozzi[1], A Farmer[2], JB Vincent[3], D Rujescu[4], D St Clair[5] and P Muglia[1]

[1]*Division of Medical Genetics, GlaxoSmithKline, Verona, Italy;* [2]*MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, London, UK;* [3]*Centre for Addiction and Mental Health, University of Toronto, Toronto, ON, Canada;* [4]*Division of Molecular and Clinical Neurobiology, Department of Psychiatry, Ludwig-Maximilians University, Munich, Germany and* [5]*Department of Mental Health, University of Aberdeen, Aberdeen, UK*

Population-based linkage analysis is a new method for analysing genomewide single nucleotide polymorphism (*SNP*) genotype data in case–control samples, which does not assume a common disease, common variant model. The genome is scanned for extended segments that show increased identity-by-descent sharing within case–case pairs, relative to case–control or control–control pairs. The method is robust to allelic heterogeneity and is suited to mapping genes which contain multiple, rare susceptibility variants of relatively high penetrance. We analysed genomewide *SNP* datasets for two schizophrenia case–control cohorts, collected in Aberdeen (461 cases, 459 controls) and Munich (429 cases, 428 controls). Population-based linkage testing must be performed within homogeneous samples and it was therefore necessary to analyse the cohorts separately. Each cohort was first subjected to several procedures to improve genetic homogeneity, including identity-by-state outlier detection and multidimensional scaling analysis. When testing only cases who reported a positive family history of major psychiatric disease, consistent with a model of strongly penetrant susceptibility alleles, we saw a distinct peak on chromosome 19q in both cohorts that appeared in meta-analysis ($P = 0.000016$) to surpass the traditional level for genomewide significance for complex trait linkage. The linkage signal was also present in a third case–control sample for familial bipolar disorder, such that meta-analysing all three datasets together yielded a linkage $P = 0.0000026$. A model of rare but highly penetrant disease alleles may be more applicable to some instances of major psychiatric diseases than the common disease common variant model, and we therefore suggest that other genome scan datasets are analysed with this new, complementary method.
Molecular Psychiatry (2010) **15,** 319–325; doi:10.1038/mp.2008.100; published online 16 September 2008

Keywords: schizophrenia; segmental sharing; case–control; genome scan

Introduction

Schizophrenia and bipolar disorder are multifactorial and heterogeneous psychiatric illnesses with strong genetic components, and they may partly share aetiologies.[1] Although genomewide association (GWA) scans have recently delivered a wealth of new insights into the genetic causes of many common diseases,[2] progress for psychiatric diseases is slower, with no validated findings having yet emerged from this approach. Pooling of multiple GWA studies may deliver robust findings. However, in parallel to those efforts, we decided to interrogate three existing case–control GWA datasets for schizophrenia and bipolar disorder under an alternative model to the common disease, common variant model, which is assumed by standard association testing.[2] We decided to analyse our datasets under a model of many, rare susceptibility alleles of relatively strong effect on disease risk, each of which is envisaged to explain only a small patient subpopulation, but would have a large determining effect on disease when present.[3] In support of this model it is interesting to note recent discoveries in autism[4,5] and schizophrenia[6] of potentially important, but rare, structural variations, sometimes *de novo*, but sometimes inherited, that are envisaged to have high penetrance for these psychiatric diseases.

Purcell *et al.*[7] recently described a population-based linkage method that is suitable for application to case–control cohorts for which genomewide single nucleotide polymorphism (*SNP*) genotype data are available, and that is designed to detect relatively rare, major genetic disease forms. The principle is to scan the genome for relatively long, extended

Correspondence: Dr C Francks, Division of Medical Genetics, GlaxoSmithKline, Via Fleming 4, Verona 37135, Italy.
E-mail: clyde.2.francks@gsk.com
Received 13 May 2008; revised 29 July 2008; accepted 7 August 2008; published online 16 September 2008

segments (typically over 1 Mb) that are shared identity-by-descent (IBD) by case–case pairs relatively more often than by case–control or control–control pairs. The method makes use of the fact that, even in case–control cohorts of apparently 'unrelated' individuals, there will still be a useful degree of residual relatedness between many pairs of subjects in the cohort. Assuming a homogeneous, random-mating population, and using genomewide *SNP* genotype data, it is possible to use the observed genotypic identity-by-state (IBS) between pairs of distantly related subjects to estimate their overall extent of IBD sharing, and their IBD sharing at any particular genomic location.[7] In practise one specifies a minimum number of adjacent *SNPs* in low linkage disequilibrium (LD), and then scans across the genome to search for such extended segments shared IBD (we used a minimum of 75 adjacent *SNPs*, which corresponded on average to genomic intervals of 2 Mb). The significance of any increase in case–case sharing at a given genomic location is estimated empirically by permutation, and the method is robust to allelic heterogeneity at causative genes, because all case–case pairs that share IBD across a given region can contribute to the linkage signal, even if different pairs carry independent mutations in the same gene. As all pairs of subjects in the analysis are only distantly related (we used a maximum IBD sharing of 5% across the genome to consider any pair of subjects in the analysis), the prior probability of a pair sharing at any given genomic location is low. This is a source of power in this approach, because when a pair shares a segment IBD this is far more unlikely to occur by chance than in standard linkage analysis in pedigrees.

Here we present an application of population-based linkage analysis to three case–control cohorts for schizophrenia or bipolar disorder. Case–control association analysis of these cohorts will be described elsewhere, but briefly, no *SNP* association has been detected that survives genomewide multiple testing, either individually within each of the datasets, or in meta-analysis of them, and we are currently collaborating to include these datasets in larger association studies. Here we wished to investigate an alternative, multiple rare variant hypothesis, that is not well tested in the usual association paradigm.

Materials and methods

Schizophrenia case–control cohorts

One cohort comprised 429 schizophrenia patients (age 39.2 ± 10.4 years, range 19–70) and 428 healthy controls (age 48.8 ± 14.7 years, range 22–75), all self-identifying as of German or central European ancestry and collected in Munich. The second cohort comprised 461 schizophrenia patients and 459 controls, all self-identifying as of Scottish or north European ancestry, collected in Aberdeen, Scotland. Patients for both cohorts were ascertained following a matching clinical protocol. To be enrolled as a case, participants must have had both a Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV) and an International Classification of Diseases 10 (ICD-10) diagnosis of schizophrenia. In the Munich and Aberdeen cohorts, respectively, subtypes were observed in the following proportions: paranoid 77.6 and 86.2%, disorganised 15.6 and 7.5%, catatonic 2.2 and 2.1% and undifferentiated 4.6 and 4.2%. Detailed medical and psychiatric histories were collected, including a clinical interview using the Structured Clinical Interview for DSM-IV (SCID), to evaluate lifetime axis I and II diagnoses. In both cohorts four physicians and one psychologist rated the SCID interviews and all measurements were double rated by a senior researcher. Cohen's κ of 0.80 indicated good interrater reliability. Positive or negative family history of disease was assessed by self-report of occurrence in first-degree relatives. Exclusion criteria included a history of head injury or neurological diseases. All case participants were outpatients or stable inpatients. Further details of the Munich cohort and protocol are available in Van den Oord *et al*.[8] All cases and controls gave informed consent. The study was approved by the respective local ethical committees.

Healthy volunteers were randomly selected from the general population both for the Munich and Aberdeen cohorts (ascertained by mail for Munich, and by general practitioners for Aberdeen). In the Munich cohort several screenings were conducted before the volunteers were enrolled in the study to exclude subjects with central neurological diseases and psychotic disorders or subjects who had first-degree relatives with psychotic disorders. First, subjects who responded were screened by phone. Second, detailed medical and psychiatric histories were assessed for the volunteers and their first-degree relatives using systematic forms. Third, if no exclusion criteria were fulfilled, they were invited to a comprehensive interview including the SCID to validate the absence of psychotic disorders. Finally, a neurological examination was conducted to exclude subjects with current CNS impairment. In the case that the volunteers were older than 60 years, the Mini Mental Status Test[9] was performed to exclude subjects with possible cognitive impairment. The enrolment procedure was similar for the Aberdeen controls, although a formal SCID was not undertaken.

Bipolar cohort

This study included 1960 bipolar case/control subjects from a multicentre study including Caucasian subjects from three different sites: 377 cases and 270 controls were recruited at the Centre for Addiction and Mental Health in Toronto, Canada; 493 cases and 477 controls were recruited at the Institute of Psychiatry in London, UK; 95 cases and 186 controls were recruited at the University of Dundee, UK.

Individuals were eligible for inclusion in this study only if all the following criteria applied: (1) they were at least 18 years of age at the time of entering the study, (2) they were of Caucasian background and (3)

they gave voluntary written consent to participate in the study. Furthermore, each case was assessed when euthymic and had to have been diagnosed (lifetime) with the DSM-IV/ICD-10 bipolar I or bipolar II disorder. In addition to the general inclusion criteria listed above, each control individual had to show they were devoid of any bipolar disorder related psychiatric illness as determined by a self-reported questionnaire. Affected individuals were excluded if they: (1) had ever experienced mood incongruent psychotic symptoms, (2) had a lifetime intravenous drug use or dependency diagnosis, (3) had bipolar disorder solely because of alcohol/substance abuse or secondary to medical illness/medication. For each affected individual, detailed clinical information was collected through a psychiatric interview administered by an experienced research assistant who received official training to administer Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interviews. All cases were interviewed using SCAN. Items of psychopathology in the SCAN interview were rated for presence and severity according to the worst depressive and the worst manic episodes identified by the individual. Demographics, country of origin, personal and family medical history, alcohol and smoking habits, medications taken during the past 6 months, and vital signs (for example weight, height and heart rate) were assessed in all participants. Positive or negative family history of disease was assessed by self-report of occurrence in first-degree relatives.

The three recruiting hospitals obtained approval by their respective Research Ethical Boards to conduct this study, and all individuals gave their written informed consent to participate in the study and for the use of their DNA.

Genotyping of schizophrenia cohorts

The Munich cohort was genotyped using the Illumina HumanHap300 chip with a total of 317 503 *SNPs*, and the Aberdeen cohort was genotyped using the Illumina HumanHap550 chip with a total of 555 352 *SNPs*. A series of quality control (QC) checks and tests of cryptic relatedness were carried out (similar to the bipolar sample, see Supplementary Methods), ultimately excluding a total of 15 and 28 participants in Munich and Aberdeen, respectively. A 'one percent rule' was applied, that discarded from analysis any *SNP* that had more than 1% of samples that could not be reliably scored. After employing this rule the average success rate of genotyping was 98.4% and the concordance rate for duplicate genotyping was 99.997%.

Genotyping of bipolar cohort

Samples were genotyped in blinded fashion using the Illumina Infinium HumanHap 550 genotyping platform at Illumina, and according to published Illumina protocol.

The details of our extensive QC procedure are given in Supplementary Information on-line. In total, after all of these procedures, 1779 subjects were available for genetic analysis, 883 cases and 896 controls.

Improving genetic homogeneity

In association analysis of these cohorts we have corrected for underlying population structure by using principal component-based methods.[10] For population-based linkage analysis this quantitative approach to handling stratification is not possible currently, yet the analysis depends crucially on having a genetically homogeneous sample in which cases and controls do not differ systematically, and in which all subjects can reasonably be imagined to be drawn from a single, random-mating population.[7] We knew from previous analysis (unpublished) that the Munich and Aberdeen schizophrenia cohorts differ markedly from one another in principal components analysis, and therefore we analysed each of the schizophrenia cohorts, and the bipolar cohort, separately for linkage, with the aim of combining them by downstream meta-analysis. Nonetheless, we found it necessary to perform additional cleanup steps within PLINK[7] version 1.00 to improve within-cohort genetic homogeneity for the purposes of this analysis.

Within each cohort separately we first calculated IBS values for all pairs of subjects, and transformed these to cohort-specific standardised *Z*-scores, then eliminated outlier subjects whose genomewide IBS sharing was < -3 s.ds. in relation to any of their nearest five neighbours. We then performed multi-dimensional scaling analysis and plotted subjects out on the first two dimensions (Supplementary Figure S1). Briefly, the Aberdeen cohort appeared as a single homogeneous distribution. The Munich cohort appeared largely homogeneous but 60 subjects did not fit into the main cluster and we removed them from further analysis. The bipolar cohort was more variable, and we eliminated 333 subjects (19% of the cohort) who did not fit tightly within the main cluster. We then carried out tests for systematic differences in genomewide IBS sharing in cases and controls, but there was no significant evidence for differences of within- and between-group genetic similarities. We derived inbreeding coefficients for all subjects based on genomewide rates of homozygosity but did not find it necessary to eliminate any subjects for this.

Population-based linkage analysis

We used PLINK7 version 1.00 to carry out population-based linkage analysis. Analysis was performed separately in each case–control cohort. We pruned our genomewide *SNP* datasets to roughly 110 000 *SNPs* that are in relatively low LD (see Supplementary Methods), and we used an approximation of 1 Mb = 1 cM to define a genomewide genetic map for these *SNPs*. We set a maximum genotyping failure rate per *SNP* of 1%, and a minimum minor allele frequency of 5%, to only use the most informative *SNPs* in which we have the highest confidence of genotype quality. We did not consider subject pairings whose genomewide IBD was > 5 or $< 0.5\%$,

as more closely related pairs are likely to dominate the analysis, and less closely related pairs are unlikely to share extended segments IBD but add considerably to run time. We used 200 000 permutations to estimate empirical significance of linkage at each location in the genome. We used a minimum of 75 adjacent *SNPs* to identify shared segments, which for 110 000 *SNPs* distributed genomewide corresponds to an average interval of 2 Mb, assuming a 3 Gb genome. Linkage results from different cohorts were then merged by meta-analysis by use of Fisher's combination of *P*-values.

Results

Genomewide population-based linkage analysis of the entire Munich and Aberdeen schizophrenia case–control cohorts produced no signals that overlapped significantly in meta-analysis, nor overlapped noticeably with the bipolar cohort (not shown). However, our main focus was to work with major gene, familial forms of these disorders, in accordance with the model of inherited rare variants of strong penetrance, for which one expects a marked elevation of disease frequency in first-degree relatives of affected cases. We therefore selected for further analysis only those cases who reported a positive family history of major psychiatric disease similar to their own. In the schizophrenia cohorts, roughly 20% of cases met this definition, and in the bipolar, roughly 40% of cases. By making this selection we aimed to reduce the occurrence of sporadic, environmental or polygenic forms of the disease.

When analysing only the cases with a positive family history, we found a peak on chromosome 19q13 in meta-analysis of the two schizophrenia datasets, $P = 0.000016$, that was 1.5 orders of magnitude more significant than the rest of the signals in the genome. Although the exact genomewide correction for this analytic approach has yet to be determined, a *P* of 0.000016 surpasses the traditional threshold for calling genomewide significance for linkage in complex trait analysis.[11] (The significance threshold in linkage testing is considerably higher than for GWA testing because the genomic resolution is lower, and the effective amount of independent testing is correspondingly lower.) The segments shared IBD by case–case pairs at the 19q13 locus ranged from 1.5 Mb, spanning 75 consecutive low-LD *SNPs*, to 6.6 Mb, spanning 244 consecutive low-LD *SNPs*. There were 10 case–case pairs who shared segments IBD across the region, made up from 15 individual cases, and the data indicated that the shared segments were derived from a maximum of 6 separate founder chromosomes, which may carry different ancestral mutations in an underlying gene.

The bipolar dataset provided further support for the 19q13 linkage (Supplementary Figure S2), such that in meta-analysis of all three datasets together the peak *P*-value decreased by roughly a further order of magnitude to $P = 0.0000026$ (Figure 1). In the bipolar dataset there were 32 case–case pairs derived from 53 cases who shared segments IBD across the most strongly linked region. The overlap of the 19q13 linkages in the three cohorts is shown in Figure 2, illustrated by the cohort-specific, empirical, permutation-derived *P*-values. Meta-analysis of all three cohorts (Figure 1) yielded a maximally linked interval, comparable to the '1-LOD' region obtained in standard linkage analysis, of 2.3 Mb on 19q13, from 46.5 to 48.8 Mb on National Centre for Biotechnology Information (NCBI) build 36, which is a genomic region very dense in genes (54 genes in the region). The overlapping linkages on 19q13 may be consistent with an underlying genetic predisposition on 19q13 to an affective psychotic disorder that is sometimes classified as either bipolar or schizophrenia.

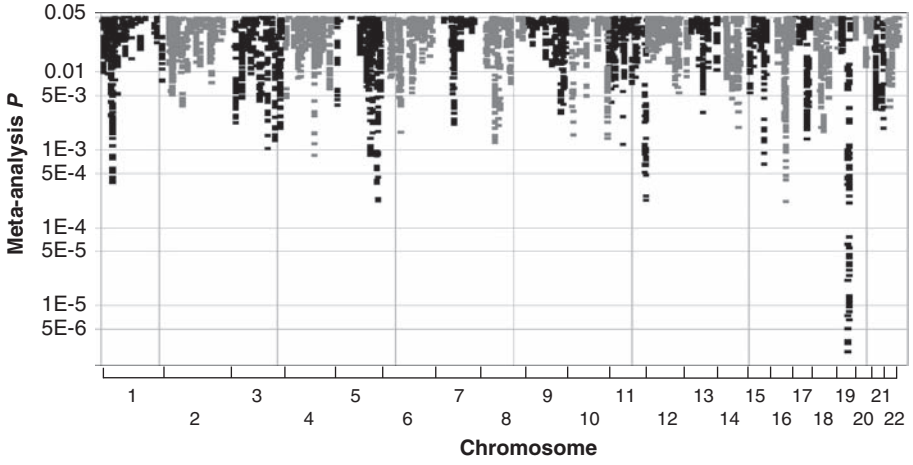

Figure 1 Meta-analysis of genomewide population-based linkage signals in three case–control cohorts for psychiatric disease; Munich SCZ, Aberdeen SCZ, bipolar. The pointwise significance of linkage is given on the *Y* axis. The chromosomes are arranged p- to q-arm along the *X* axis, in numerical order from left to right, and alternating grey/black. The 19q13 signal is two orders of magnitude more significant than the rest of the genomic background.

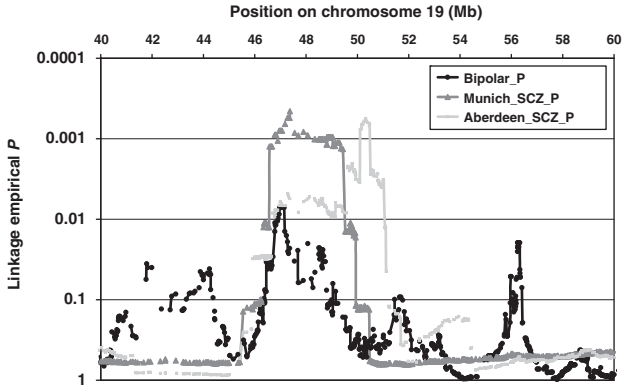

Figure 2 The 19q13 linkage in the three separate cohorts, showing the positional overlap. Empirical, permutation-based *P*-values are shown (derived separately in each cohort from 200 000 permutations).

We checked the deCODE genetic map[12] for chromosome 19 and observed that the sex-averaged recombination rate is very close to 1 across the 19q13 region of linkage (Supplementary Figure S3), and the general level of LD in the region does not appear unusual in the Hapmap,[13] so we think the region meets the assumptions of the analysis and that there are no obvious biases that might have produced a spurious effect at this location. We checked the Haplotter website[14] for evidence of recent positive selection in the region based on analysis of Hapmap data, but no genes in the maximally linked region were assigned a significant *P*-value for evidence of positive selection, although there were some extreme scores of Fay and Wu's *H* and Tajima's *D*, which are both based on detecting shifts in the frequency spectrum of polymorphisms.[14] However, these measures are best calculated from resequencing data, rather than Hapmap data, which are based on a preselection of *SNPs* and do not give an unbiased assessment of the true frequency spectrum of polymorphisms, so at the moment the evidence for selection at this locus remains equivocal.

We have listed the full genomewide linkage *P*-values from our analysis of familial cases in the three cohorts, plus the meta-analysis of them, in Supplementary Table S1 on-line, mapped onto NCBI genomic build 36. The table also includes the locations of known genes so that they can be assessed relative to the peaks and troughs of linkage.

Discussion

Population-based linkage analysis offers an alternative to standard association approaches for analysing genomewide *SNP* genotype datasets in case–control cohorts. Instead of assuming a common disease, common variant model, population-based linkage analysis is designed to detect rare effects of relatively strong penetrance, and relatively recent origin in the population, and the approach is robust to allelic heterogeneity at susceptibility loci. This model may be more applicable to some forms of major psychiatric diseases than the standard association paradigm,[4–6] and our discovery of a potential new effect on 19q13 illustrates the applicability of the method in this field. We encourage other investigators to interrogate their own GWA datasets in this way, as a complementary strategy to everexpanding metaassociation analysis.

The signal on 19q13 that we identified, when focussing on cases who reported positive family history of disease, was present in both schizophrenia and bipolar disorder cohorts, which indicates that this locus may be pleiotropic to both disorders, or rather that it may define a psychiatric phenotype that straddles the traditional diagnostic boundary between them (although the linkage was stronger in the two schizophrenia datasets (Figure 2)). Identification of underlying mutations on 19q13 will be necessary, and investigation of the implicated gene in larger numbers of patients, to build confidence around this initial observation of pleiotropy. Of note, 19q12-13 has been implicated previously by linkage analysis in multiplex families that contained cases with either bipolar disorder, manic schizoaffective disorder or recurrent unipolar depression, and the linkage was strongest (LOD = 4.55) when applying the broadest, most inclusive psychiatric phenotype in that analysis.[15]

An obvious drawback of the population-based linkage method is its relatively low resolution at the genomic level, as compared to LD-based association mapping. Our region of maximal linkage on 19q13, based on meta-analysis of all three of our datasets, spans roughly 2.3 Mb. It is somewhat unfortunate that the region we have implicated is very dense in genes (54 genes). There are many 2-Mb regions of the genome in which only a small number of genes are located, and the road to discovering functional mutations would likely be much shorter in such regions. In general then, the genomic resolution of the population-based linkage method is not prohibitive when used in samples of this size, even if the particular signal that we have detected in this study is somewhat difficult to follow up, because of a very high gene density on 19q13.

Nonetheless there are several good candidate genes for future resequencing analysis in the linked region. Of note, glycogen synthase kinase 3-α (GSK3A) is situated at the centre of the region. GSK3A is a multifunctional protein serine kinase, homologous to *Drosophila* 'shaggy' (zeste-white3) and implicated in the control of several regulatory proteins including glycogen synthase and transcription factors.[16] It is also involved in the WNT and PI3K signalling pathways.[16] *GSK3A* is one of six genes that showed differential expression, in both blood and brain tissue, in a systematic expression profiling study of schizophrenia patients relative to controls.[17] *GSK3B* activity is inhibited by lithium and has been implicated in schizophrenia and bipolar disorder by a number of different approaches, and remains a target of interest in psychiatry drug discovery.[18,19] Another candidate gene in the region, glutamate

receptor, ionotropic, kainate 5 (*GRIK5*), forms functional heteromeric kainite preferring ionic channels in combination with other *GRIK* subunit genes,[16] and its close homolog *GRIK4* has been implicated in schizophrenia and bipolar disorder by genetic and cytogenetic analysis.[20] *GSK3A* and *GRIK5* are both strong candidates for future mutation analysis in bipolar disorder and schizophrenia.

After 19q13, the next four most significantly linked loci in meta-analysis of the schizophrenia and bipolar datasets (Figure 1) were on 1p34, 5q34, 11q24 and 16q22, although these were not outstanding in a genomewide context. We checked each of these loci in previously published meta-analyses of linkage scans in multiplex families, and found that 5q34 and 11q24 were within regions ranked second and fourth, respectively, in a meta-analysis of 20 genomewide linkage scans for schizophrenia,[21] and 19q13 was implicated in meta-analysis of 18 genomewide linkage scans for bipolar disorder,[22] although in that study only under a narrow diagnostic model. There was no obvious overlap of our top linkages with an earlier, combined schizophrenia/bipolar meta-analysis of genomewide linkage scans in multiplex families.[23] It may be that population-based linkage analysis is more sensitive to rarer familial forms of disease than traditional linkage approaches (particularly in sib pairs) that typically identify signals arising from a high proportion of families in the analysis. However, clearly our results will require further validation using the population-based linkage approach in independent case–control samples, which we recommend to apply while nonetheless enriching for familial forms of disease whenever possible.

Disease susceptibility genes may contain both common and rare variants with phenotypic consequences, and which can arise independently. We therefore investigated our allelic association data in these cohorts across the linked region of 19q13, but found nothing that helps to refine the signal, as no association was present that survives multiple testing for the number of *SNPs* in the region (this also remained the case when we analysed for association only the cases with positive family history). Allelic association and population-based linkage provide parallel, complementary assessments of the dataset, and are not necessarily expected to highlight the same region unless there are both common and rare variant effects at an underlying, causative gene. In the case of the 19q13 effect, our focus will remain on identifying rare, highly penetrant mutations in the subset of patients that drive the linkage here.

As a proportion of the overall sample of familial cases, roughly 10–15% drove the linkage signal on 19q13 in both the schizophrenia and bipolar cohorts, and similarly for each of the loci on 1p34, 5q34, 11q24 and 16q22. However, this is an overestimate of the disease risk potentially attributable to these loci, as the linkage signals are derived from increased case–case sharing relative to case–control or control–control sharing, and each of these loci has many of these other pair types sharing IBD also. The low population risk attributable to these linkages highlights again that the focus of this approach is on rare but highly familial forms of disorder, which may nonetheless yield critical insights into disease pathogenesis if the underlying genetic causes are identified.

Acknowledgments

Thanks to David Goldstein, Dongliang Ge, Kevin Shianna and Xiangyang Kong for producing and sharing well QC'd genotype data. Shaun Purcell, Silviu-Alin Bacanu and Lon Cardon provided critical thoughts and comments. John Strauss, Sagar Parikh, Jim Kennedy, Peter McGuffin and Ian Reid contributed to the bipolar case–control collection. We thank Lefkos Middleton and Allen Roses for their roles in assembling the case–control cohorts.

References

1 Berrettini W. Bipolar disorder and schizophrenia: not so distant relatives? *World Psychiatry* 2003; **2**: 68–72.
2 McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA *et al*. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; **9**: 356–369.
3 McClellan JM, Susser E, King MC. Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry* 2007; **190**: 194–199.
4 Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T *et al*. Strong association of *de novo* copy number mutations with autism. *Science* 2007; **316**: 445–449.
5 Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J *et al*. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 2008; **82**: 477–488.
6 Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM *et al*. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008; **320**: 539–543.
7 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
8 Van den Oord EJCG, Rujescu D, Robles JR, Giegling I, Birrell C, Bukszar J *et al*. Factor structure and external validity of the PANSS revisited. *Schizophr Res* 2006; **82**: 213–223.
9 Kessler J, Folstein SE, Denzler P. *Mini-Mental-Status-Test: German Version*. Beltz: Weinheim, 1990.
10 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–909.
11 Sawcer S, Jones HB, Judge D, Visser F, Compston A, Goodfellow PN *et al*. Empirical genomewide significance levels established by whole genome simulations. *Genet Epidemiol* 1997; **14**: 223–229.
12 Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B *et al*. A high-resolution recombination map of the human genome. *Nat Genet* 2002; **31**: 241–247.
13 International HapMap Consortium. A second generation human haplotype map of over 3.1 million *SNPs*. *Nature* 2007; **449**: 851–861.
14 Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol* 2006; **4**: e72.
15 Badenhop RF, Moses MJ, Scimone A, Mitchell PB, Ewen-White KR, Rosso A *et al*. A genome screen of 13 bipolar affective disorder pedigrees provides evidence for susceptibility loci on chromosome 3 as well as chromosomes 9, 13 and 19. *Mol Psychiatry* 2002; **7**: 851–859.

16 OMIM—Online Mendelian Inheritance in Man. *OMIM 2008*##http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim.
17 Glatt SJ, Everall IP, Kremen WS, Corbeil J, Sasik R, Khanlou N *et al.* Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. *Proc Natl Acad Sci* 2005; **102**: 15533–15538.
18 Jope RS, Roh MS. Glycogen synthase kinase-3 (GSK3) in psychiatric diseases and therapeutic interventions. *Curr Drug Targets* 2006; **7**: 1421–1434.
19 Adli M, Hollinde DL, Stamm T, Wiethoff K, Tsahuridu M, Kirchheiner J *et al.* Response to lithium augmentation in depression is associated with the glycogen synthase kinase 3-beta -50T/C single nucleotide polymorphism. *Biol Psychiatry* 2007; **62**: 1295–1302.
20 Pickard BS, Malloy MP, Christoforou A, Thomson PA, Evans KL, Morris SW *et al*. Cytogenetic and genetic evidence supports a role for the kainate-type glutamate receptor gene, *GRIK4*, in schizophrenia and bipolar disorder. *Mol Psychiatry* 2006; **11**: 847–857.
21 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.
22 Segurado R, Detera-Wadleigh SD, Levinson DF, Lewis CM, Gill M, Nurnberger Jr JI *et al*. Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: bipolar disorder. *Am J Hum Genet* 2003; **73**: 49–62.
23 Badner JA, Gershon ES. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry* 2002; **7**: 405–411.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)