HAL
open science

Meta-analysis indicates that common variants at the DISC1 locus are not associated with schizophrenia

Jonathan Flint, Iain Mathieson, Marcus Munafò

To cite this version:

Jonathan Flint, Iain Mathieson, Marcus Munafò. Meta-analysis indicates that common variants at the DISC1 locus are not associated with schizophrenia. Molecular Psychiatry, 2011, 10.1038/mp.2011.41 . hal-00631259

HAL Id: hal-00631259

https://hal.science/hal-00631259v1

Submitted on 12 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Meta-analysis indicates that common variants at the *DISC1* locus are not associated with schizophrenia

Iain Mathieson[1], Marcus R. Munafò[2], Jonathan Flint[1]

1. Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, United Kingdom
2. School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, United Kingdom.

Correspondence to: Jonathan Flint

Email jf@well.ox.ac.uk

Phone 01865 287512

Abstract

Several polymorphisms in the Disrupted-in-Schizophrenia-1 (DISC1) gene are reported to be associated with schizophrenia. However to date there has been little effort to evaluate the evidence for association systematically. We carried out an imputation-driven meta-analysis, the most comprehensive to date, using data collected from 10 candidate gene studies and 3 genome wide association studies containing a total of 11,626 cases and 15,237 controls. We tested 1,241 SNPs in total, and estimated that our power to detect an effect from a variant with minor allele frequency > 5% was 99% for an odds ratio of 1.5 and 51% for an odds ratio of 1.1. We find no evidence that common variants at the DISC1 locus are associated with schizophrenia.

Keywords: DISC1, schizophrenia, meta-analysis, association, case-control, GWAS

Introduction

Schizophrenia is a debilitating psychiatric illness with a lifetime prevalence of around 0.4% [1] and a heritability estimated to be between 80 and 85% [2, 3]. One schizophrenia susceptibility gene that has attracted much attention is *Disrupted-in-Schizophrenia -1* (*DISC1*). *DISC1* is one of three genes (with *DISC2* and *TSNAX*) lying at the site of a balanced t(1;11)(q42.1;q14.3) translocation which has a strong association (LOD=7.1) with schizophrenia and other mental illness in a large Scottish pedigree [4] [5].

Further evidence that *DISC1* genetic variants predispose to schizophrenia in particular comes from four sources. First, the expression of the protein product DISC1 is reduced in the Scottish pedigree in which the translocation segregates. Second, DISC1 interacts physically with PDE4B. Another Scottish family with schizophrenia has a translocation that reduces the expression of PDE4B [6]. Third, association between schizophrenia and variants at the *DISC1* locus has been reported [7-20]. Fourth, sequencing part of the *DISC1* gene in 288 cases identified six rare coding variants not found in 10,000 controls [21].

These findings, along with evidence of DISC1's involvement in other mental illnesses including bipolar disorder [19], depression [11] and autism [22] have spurred research into understanding what DISC1 does in the brain. Studies of DISC1's interactome have revealed that it is involved in synaptic development pathways [23] and interacts with proteins important for neural development such as NDEL1 [24-26]. These results seem promising for DISC1's role in schizophrenia in light of the common hypothesis that the origin of the condition is neurodevelopmental.

However, there has been no comprehensive meta-analysis of schizophrenia association study findings for *DISC1*, despite the fact that the association evidence is not totally consistent. Association has been reported over almost the entire gene (Figure 1 [7-20]), but the largest study to date failed to detect any association [27], and genome-wide association studies (GWAS) have not found a clear signal at *DISC1*.

Meta-analysis of *DISC1* findings is difficult to carry out because the gene is large (414 Kb) and includes many variants (2,603 SNPs in dbSNP 131). There is no consensus about which markers have the strongest association, no markers are tested in all studies, and most are tested in only a small number of studies. Therefore we decided to try to combine data from studies that use different markers by imputation. This method estimates results from a marker that has not been genotyped by exploiting linkage disequilibrium (LD) with markers that have been genotyped.

Until recently there was insufficient sequence information at the DISC1 locus to make imputation feasible; however release of data from phase 3 of the International HapMap project [28] and from the first 1000 Genomes pilot [29] provide the most comprehensive catalogue of common human genetic variation available today, including rare variants. We set out to use these publicly available data to impute genotypes in DISC1 association studies and to combine results from all available sources, including both GWAS and candidate gene studies, thus overcoming the problems described above.

Methods

Selection of studies for inclusion

We identified candidate gene studies and GWAS for schizophrenia or schizoaffective disorder in unrelated case-control samples drawn from populations of predominantly European ancestry with at least 100 cases and controls. Studies reporting data on both male and female cases and controls were included. We did not include studies where data from multiple individuals was pooled and analyzed together. We excluded studies where the samples overlapped significantly with another identified study.

Systematic literature search

We considered all studies listed for DISC1 on the SZGene database [30]. We also searched PubMed with the search terms “DISC1” AND “Schizophrenia” AND “Association” and also for the terms “Schizophrenia” AND “GWAS”. Once these had been collected, their bibliographies were then searched for additional references. The abstracts of studies thus identified were examined with respect to the inclusion criteria above. Studies failing to meet the criteria were discarded, and the full text of remaining studies was examined to determine whether or not they should be included.

Data extraction

For each study, the following data were extracted: 1) Author(s) and year of publication, 2) Sample size, 3) Data (individual genotypes, genotype counts or minor allele frequencies in order of preference). Where the required data were not available in the published study or from web-based supplementary information, we contacted the authors of the paper by e-mail to request access to the data. For GWAS we requested data for the region extending 100Kb upstream of *DISC1* and downstream of *TSNAX* (Chr1:231,572,000-232,277,000 in the hg19 genome assembly), which covers all reported associations.

Quality control

In addition to any quality control performed by the authors of the studies, we excluded any SNPs which failed the following tests: 1) Minor allele frequency (MAF) in cases or controls < 0.01. 2) Where alleles were given in the study data, we checked them against dbSNP alleles [31] and excluded the SNP if the reported alleles differed from the dbSNP alleles and could not be explained by a strand flip. We also excluded the SNP if no alleles were specified in the study data and the MAF was > 0.45; otherwise we coded the SNP by the minor allele. 3) For studies where we had genotype counts available, we excluded SNPs that had a Hardy-Weinberg equilibrium test P-value of < 0.001.

Imputation

One of the problems encountered in previous attempts at meta-analysis is the lack of overlap between studies in terms of tested SNPs. We addressed this problem by imputing genotypes for studies where we had full genotype data. Imputation was carried out using IMPUTE v2 [32, 33] with reference panels of CEU haplotypes from HapMap 3 and the 1000 Genomes pilot 1. The CEU haplotypes are taken from samples from the Centre d'Etude du Polymorphisme Humain collected from individuals living in Utah, USA, with ancestry from northern and western Europe. The haplotype data were obtained from the IMPUTE website (http://mathgen.stats.ox.ac.uk/impute/impute.html). For studies where we did not have full genotype data (we had genotype counts or minor allele frequencies), we imputed minor allele frequencies using the linear predictor implemented in the BLIMP package [34] and the CEU haplotypes from 1000 Genomes pilot 1. We attempted to impute for all studies which genotyped at more than ten SNPs.

Size and power calculations

To calculate the size and power of the meta-analysis, we simulated case-control data from haplotypes drawn from the 1000 Genomes pilot 1. To calculate the P-value required to obtain a given significance level, we sampled data from the 1000 Genomes haplotypes, with the same characteristics as the studies in the sample

(numbers of cases, controls and testing the same SNPs), calculated effect sizes and P-values as described below and obtained the empirical significance levels for each SNP. We repeated the simulation and determined the required nominal point-wise significance level required to achieve a gene-wide significance level of 0.05.

To calculate the power of the meta-analysis, we again sampled from the 1000 Genomes data with the same characteristics as the meta-analysis samples, but in each simulation, we chose a SNP at random and weighted the sampling probabilities in cases to ensure that the chosen SNP had a specified OR in cases relative to controls. We performed the analysis as described below and recorded whether the significance level determined above was reached by any SNP. The proportion of tests with any SNP reaching significance gave us an estimate of the power of the meta-analysis for a given effect size. For size calculations we performed 10,000 simulations, and for power calculations we performed 1000 per datapoint.

Statistical analysis

Log-odds ratios (equivalent to logistic regression coefficients) and their standard deviations were calculated for each SNP in each study and then combined across studies using fixed and random effects models [35]. We tested whether any of the pooled P-values were more significant than the thresholds calculated above. We also tested whether the distribution of P-values differed significantly from that

expected under the null hypothesis. For each SNP, we investigated whether there was any evidence for heterogeneity using the I^2 statistic [36].

For the imputed genotypes, we calculated the test statistics using SNPTEST [32], which takes into account the genotype uncertainty generated by the imputation. For imputed allele frequencies, we calculated the proportion of the genotypic variance explained by the imputation, and filtered out any SNP where this was below 80%. We then calculated test statistics as for the original data. These were then combined with the test statistics from genotyped SNPs as above. We considered separately: 1) the genotyped data, 2) the genotyped data plus the imputed data, but considered only loci that were genotyped in at least one study, and 3) the genotyped data plus all the imputed data. We removed each study from the meta-analysis in turn to determine whether any single study had a disproportionate effect on the analysis. We also calculated Wakefield approximate Bayes factors [37] from estimates of the standardized effect sizes and their variances, assuming a prior for the effect size that was normally distributed with mean 0 and variance 0.015. This prior is equivalent to the belief that for each SNP, there is a 95% probability that its odds ratio lies between 0.79 and 1.28.

Results

We obtained data for eleven candidate gene studies and four GWAS containing a total of 11,626 cases and 15,237 controls (Table 1). We could not obtain data from one case-control study and one GWAS matching our criteria because there was no response from the authors. Three studies were excluded because they overlapped

substantially with other studies (Table 2). The total number of separate SNPs genotyped in any study was 401.

Imputation dramatically increased the number of studies testing each SNP. Before imputation, 100 SNPs were tested in a single study only. After imputation, there were only 13 such SNPs. The mean number of studies testing each SNP increased from 2.6 to 4.3, if only sites genotyped in a single study were considered, and to 3.4 if all imputed sites were used. Ultimately, we obtained data on a total of 1,241 separate SNPs. Many of the SNPs in the full imputed dataset were tested in only two studies and were heavily dependent on the International Schizophrenia Consortium data [38].

For testing the 401 genotyped loci we used a point-wise significance level of 1.92×10^{-4} corresponding to a gene-wide significance level of 0.05. We generated this threshold by simulation, using the exact SNPs and sample sizes available in the included studies. It is equivalent to a multiple testing correction for 260 (95% CI 250-269) independent tests. For the 1,241 SNPs in the imputed data, the correction is equivalent to 444 independent tests (95% CI 424-467). With these cutoffs, we estimated that for common variants (MAF>5%) the power to detect a true association at OR=1.5 with the imputed data restricted to genotyped loci was 99%. For OR=1.1 the power was 51%.

For rare variants, power is much lower. For variants with MAF between 1% and 5%, the corresponding figures are 92%, and 6%. For variants with MAF below 1%, which we did not try to test, the power is very low, even for high odds ratios. These figures are the power to detect an association at any locus given the odds ratio at the causal locus, which is not necessarily the one detected. Figure 2 shows

power as a function of minor allele frequency for different odds ratios at the causal allele.

Using these significance thresholds, we detected no significant association with any tested SNP (figures 3 and 4). The most associated SNP was rs1765805, 160 Kb upstream of *DISC1* ($p=6\times10^{-4}$, genotyped in 1860 cases and 2389 controls), but this was heavily dependent on a single study [38] and not supported by nearby SNPs. The next highest associations were found in the regions of rs1331046 ($p=1.4\times10^{-3}$, imputed in 3610 cases and 3938 controls), downstream of the gene, and rs17817356 ($p=7\times10^{-3}$, genotyped in 4405 cases and 7427 controls, imputed in 2291 cases and 2293 controls), between exons 6 and 7, and about 30Kb upstream from DISC2. Like rs1765805, the association at rs1331046 was driven by the ISC data, but the association at rs17817356 was supported by the ISC and other datasets as well. In the random effects analysis, the associations at rs1765805 and rs17817356 remained at the same significance level but the association at rs1331046 was substantially reduced ($p>0.03$).

Excluding the ISC study significantly altered our results, and was the only study whose omission did so. As noted above, the associations at rs1765085 and rs1331046 were supported only by the ISC. However, this may be because few other studies tested these SNPs. Both SNPs are outside the gene, in a region where we had lower coverage, since most candidate gene studies have concentrated on SNPs inside the gene.

To assess heterogeneity, we calculated the I^2 statistic for all SNPs. This was zero for the majority (70%) of loci and distributed between 0 and 80% for the rest. Tests of heterogeneity are generally underpowered when, as here, there are few

studies [39, 40], so we cannot make conclusive statements about the presence or absence of heterogeneity. However given that most genetic associations display between-study heterogeneity [41] and that fixed and random effects models give the same results when there is no heterogeneity, it seems prudent to use the more conservative random effects model to estimate effect size or to prioritize markers for further analysis.

Using the effects estimated by the random effects model, the approximate Bayes factors reveal a similar pattern to the p-values (figure 5), and indicate that the region around rs17817356 has the most evidence in favor of association and that there is no consistent evidence for association elsewhere in the gene.

There are three commonly tested non-synonymous SNPs in the gene, rs3738401, rs821616 and rs6675281. The variants rs821616 and rs6675281 have been reported to have functional effects, affecting centrosomal PCM1 localization [42, 43], expression of splice variants [44], hippocampal function [45] and interaction with NDEL1 which is important for neurite outgrowth [46]. However we detected no significant associations at any of these SNPs (p= 0.11, 0.51 and 0.93 in the raw data and 0.49, 0.60 and 0.28 in the imputed).

Given the potential contribution of rare variants in *DISC1* to schizophrenia risk, we searched additional sources of data to identify further functional variants not represented in dbSNP. In the 1000 Genomes pilot data, we found one additional non-synonymous coding SNP. We also looked in the results from a recent exome sequencing study of 200 Danish individuals [47] and identified five novel nonsynonymous SNPs. Of these six variants, two had been discovered in cases and controls in Song *et al*. [21] and 4 were novel (Table 3).

Discussion

We have carried out a comprehensive analysis of candidate schizophrenia susceptibility variants in the DISC1 gene. We found no significant association with any variant. We calculated approximate Bayes factors for each SNP to identify the regions where there was most evidence of association and found that the best evidence of association was in a region between exons 6 and 7. Our meta-analysis had to deal with the problem that of 401 variants reported in the literature, 120 markers were tested in a single study, and a further 110 in just two studies. The lack of overlap in the markers tested in different studies poses a challenge to meta-analysis that we overcame by imputing genotypes. By combining results from genome-wide association and candidate gene studies together with linkage disequilibrium data from the HapMap and 1000 Genomes projects we were able to test the involvement of 1,241 SNPs in a combined sample of 11,626 cases and 15,237 controls.

Our results raise four questions. First, what range of effects have we tested? Second, to what extent have we excluded the involvement of common variants throughout the gene? Third, can we draw any conclusions about the role of rare variants at the DISC1 locus? Finally, why have we been unable to replicate the many previously reported associations?

Power calculations help to answer the first question (figure 2). Power depends critically on the sample size, which is not straightforward to estimate. While the total number of cases and controls in our sample was 11,626 and 15,237, not every SNP

was tested in every individual. In the raw data, the mean numbers of cases and controls testing any single SNP were 2,781 and 3,782. We cannot directly calculate the sample sizes for the imputed data since we used imputed genotype probabilities in our tests, but comparison of the variance of test statistics suggest that the mean improvement was equivalent to increasing the sample size by 49% to a mean of 4,144 cases and 5,653 controls. These figures are heavily skewed in the sense that some variants have much larger sample sizes. Imputation also allowed us to test variants that had not been genotyped in any study.

Our power calculations are only approximate since we sampled from a fairly limited set of haplotypes. However, they can give us an indication of what effects we have tested. For variants with MAF > 5%, we had 99% power to detect an effect with OR=1.5 and 51% power to detect an effect with OR=1.1. However, it is possible that effects will be too small to detect, even for high frequency variants. Robust association for common variants with OR of < 1.1 have been reported in a study of height [48] and it is not inconceivable that similarly small effects operate in schizophrenia [49]. It has been argued that such variants can explain a significant proportion of the heritability for common diseases and traits [38, 49, 50].

Results from the HapMap and 1000 Genomes projects provide an answer to the second question. Genotyped SNPs included 32% of all SNPs with MAF > 5% in the 1000 Genomes, and imputed SNPs included 89%. For SNPs with MAF between 1% and 5%, the corresponding figures are 10% and 41%. For genotyped loci, 95% of the gene is within 5.0Kb of a tested SNP, 50% is within 1.4Kb and the maximum distance without a tested SNP is 13.7Kb. For the imputed data, the figures are

2.6Kb, 0.6Kb and 7.2Kb. These small regions make it unlikely that any common SNP could exist that we have not, at least indirectly, tested.

Third, the effects may arise from alleles too rare to be tested here. It is not surprising that common variants do not have a large effect on susceptibility to such a debilitating disease, but we might expect rare variants to have larger effects. While we tested variants down to a MAF of 1%, we had very low power at this level, and we could not even test all the known variants of this frequency. Further stages of the 1000 Genomes project will allow most variants at this frequency to be tested. On the other hand, association between DISC1 and the disease might derive from extremely rare, essentially private, mutations of large effect [51, 52]. One study reported the identification of coding mutations in schizophrenia cases that were not detected in 10,000 control subjects [21]. However this study did not search in healthy controls for rare variants that were not found in cases. In two studies sampling 260 individuals in total, we found four novel nonsynonymous mutations, and rediscovered two that had been identified in the previous study. This suggests that coding mutations in DISC1 may be more common in healthy individuals than previously assumed. However it should also be noted that control subjects were not actively screened for the absence of SZ, a family history of SZ, or related mental illnesses.

Finally, many previous studies have reported association. There are a number of reasons why we might have failed to replicate these results; publication bias, winner's curse, allelic heterogeneity and population, sex, or subgroup specific heterogeneity. We have not attempted to investigate these here, but it is worth noting that the presence of heterogeneity of effects would particularly tend to confound large GWAS and meta-analyses such as this one.

Our meta-analysis does not address the question of *DISC1*'s involvement in other mental illness, such as bipolar disorder, major depressive disorder or autism spectrum disorder**,** nor does it answer all the questions about the association of *DISC1* with schizophrenia. Although we found no significant association with common variants, many studies have reported sex specific, and interaction effects [9, 11, 12, 18, 19], which we did not test. Consideration of the power required to detect such effects suggests that if they do exist, significantly larger studies will be required to detect them convincingly. For example, if we were to test all pairwise SNP-SNP interactions for all imputed SNPs in this study, we would be increasing the multiple testing factor by around 100,000. We would also be reducing the sample size for each test, particularly for combinations of rare variants making it difficult, if not impossible, to reach statistical significance. Hypothesis driven tests derived from functional studies and model systems may prove more productive than data mining in this case.

Whatever the true effect of the factors described above, it seems likely that we will need to consider new study designs, and new statistical techniques in order to further investigate the effect of *DISC1* variation on schizophrenia and other psychiatric illnesses.

Acknowledgments

Iain Mathieson and Jonathan Flint are supported by the Wellcome Trust. The authors would like to thank all the researchers who provided data for this study.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. *Plos Med* 2005; **2:** 413-433.

2. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait - Evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* 2003; **60:** 1187-1192.

3. 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.

4. Jacobs PA, Brunton M, Frackiewicz A, Newton M, Cook PJL, Robson EB. Studies on a family with three cytogenetic markers. *Ann Hum Genet* 1970; **33:** 325-336.

5. St Clair D, Blackwood D, Muir W, Carothers A, Walker M, Spowart G *et al.* Association within a family of a balanced autosomal translocation with major mental illness. *Lancet* 1990; **336:** 13-16.

6. Millar JK, Pickard BS, Mackie S, James R, Christie S, Buchanan SR *et al.* DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. *Science* 2005; **310:** 1187-1191.

7. Callicott JH, Straub RE, Pezawas L, Egan MF, Mattay VS, Hariri AR *et al.* Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia. *Proc Natl Acad Sci USA* 2005; **102:** 8627-8632.

8. Cannon TD, Hennah W, van Erp TG, Thompson PM, Lonnqvist J, Huttunen M *et al.* Association of DISC1/TRAX haplotypes with schizophrenia, reduced prefrontal gray matter, and impaired short- and long-term memory. *Arch Gen Psychiatry* 2005; **62:** 1205-1213.

9. Chen QY, Chen Q, Feng GY, Lindpaintner K, Wang LJ, Chen ZX *et al.* Case-control association study of Disrupted-in-Schizophrenia-1 (DISC1) gene and schizophrenia in the Chinese population. *J Psychiat Res* 2007; **41:** 428-434.

10. Ekelund J, Hennah W, Hiekkalinna T, Parker A, Meyer J, Lönnqvist J *et al.* Replication of 1q42 linkage in Finnish schizophrenia pedigrees. *Mol Psychiatry* 2004; **9:** 1037-1041.

11. Hashimoto R, Numakawa T, Ohnishi T, Kumamaru E, Yagasaki Y, Ishimoto T *et al.* Impact of the DISC1 Ser704Cys polymorphism on risk for major depression, brain morphology and ERK signaling. *Hum Mol Genet* 2006; **15:** 3024-3033.

12. Hennah W, Thomson P, McQuillin A, Bass N, Loukola A, Anjorin A *et al.* DISC1 association, heterogeneity and interplay in schizophrenia and bipolar disorder. *Mol Psychiatry* 2009; **14:** 865-873.

13. Hennah W, Varilo T, Kestilä M, Paunio T, Arajärvi R, Haukka J *et al.* Haplotype transmission analysis provides evidence of association for DISC1 to schizophrenia and suggests sex-dependent effects. *Hum Mol Genet* 2003; **12:** 3151-3159.

14. Hodgkinson CA, Goldman D, Jaeger J, Persaud S, Kane JM, Lipsky RH *et al.* Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. *Am J Hum Genet* 2004; **75:** 862-872.

15. Qu M, Tang F, Yue W, Ruan Y, Lu T, Liu Z *et al.* Positive association of the Disrupted-in-Schizophrenia-1 gene (DISC1) with schizophrenia in the Chinese Han population. *Am J Med Genet B* 2007; **144B:** 266-270.

16. Rastogi A, Zai C, Likhodi O, Kennedy JL, Wong AH. Genetic association and post-mortem brain mRNA analysis of DISC1 and related genes in schizophrenia. *Schizophr Res* 2009; **114:** 39-49.

17. Saetre P, Agartz I, De Franciscis A, Lundmark P, Djurovic S, K√§hler A *et al.* Association between a disrupted-in-schizophrenia 1 (DISC1) single nucleotide polymorphism and schizophrenia in a combined Scandinavian case-control sample. *Schizophr Res* 2008; **106:** 237-241.

18. Schumacher J, Laje G, Abou Jamra R, Becker T, M√ºhleisen TW, Vasilescu C *et al.* The DISC locus and schizophrenia: evidence from an association study in a central European sample and from a meta-analysis across different European populations. *Hum Mol Genet* 2009; **18:** 2719-2727.

19. Thomson PA, Wray NR, Millar JK, Evans KL, Hellard SL, Condie A *et al.* Association between the TRAX/DISC locus and both bipolar disorder and schizophrenia in the Scottish population. *Mol Psychiatry* 2005; **10:** 657-668, 616.

20. Zhang F, Sarginson J, Crombie C, Walker N, St Clair D, Shaw D. Genetic association between schizophrenia and the DISC1 gene in the Scottish population. *Am J Med Genet B* 2006; **141B:** 155-159.

21. Song W, Li W, Feng J, Heston LL, Scaringe WA, Sommer SS. Identification of high risk DISC1 structural variants with a 2% attributable risk for schizophrenia. *Biochem Bioph Res Co* 2008; **367:** 700-706.

22. Kilpinen H, Ylisaukko-Oja T, Hennah W, Palo OM, Varilo T, Vanhala R *et al.* Association of DISC1 with autism and Asperger syndrome. *Mol Psychiatry* 2008; **13:** 187-196.

23. Camargo LM, Collura V, Rain JC, Mizuguchi K, Hermjakob H, Kerrien S *et al.* Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. *Mol Psychiatry* 2007; **12:** 74-86.

24. Brandon NJ, Handford EJ, Schurov I, Rain JC, Pelling M, Duran-Jimeniz B *et al.* Disrupted in Schizophrenia 1 and Nudel form a neurodevelopmentally regulated protein complex: implications for schizophrenia and other major neurological disorders. *Mol Cell Neurosci* 2004; **25:** 42-55.

25. Morris JA, Kandpal G, Ma L, Austin CP. DISC1 (Disrupted-In-Schizophrenia 1) is a centrosome-associated protein that interacts with MAP1A, MIPT3, ATF4/5 and NUDEL: regulation and loss of interaction with mutation. *Hum Mol Genet* 2003; **12:** 1591-1608.

26. Ozeki Y, Tomoda T, Kleiderlein J, Kamiya A, Bord L, Fujii K *et al.* Disrupted-in-Schizophrenia-1 (DISC-1): mutant truncation prevents binding to NudE-like (NUDEL) and inhibits neurite outgrowth. *Proc Natl Acad Sci USA* 2003; **100:** 289-294.

27. Sanders AR, Duan J, Levinson DF, Shi J, He D, Hou C *et al.* No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. *Am J Psychiatry* 2008; **165:** 497-506.

28. Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; **467:** 52-58.

29. Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA *et al.* A map of human genome variation from population-scale sequencing. *Nature* 2010; **467:** 1061-1073.

30. Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ *et al.* Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* 2008; **40:** 827-834.

31. Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. (dbSNP Build ID: 131). http://www.ncbi.nlm.nih.gov/SNP/, Accessed 27th June 2010.

32. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39:** 906-913.

33. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *Plos Genet* 2009; **5:** e1000529.

34. Wen X, Stephens M. Using linear predictors to impute allele frequencies from summary or pooled genotype data. *Ann Appl Stat* 2010; **In press**.

35. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7:** 177-188.

36. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21:** 1539-1558.

37. Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol* 2009; **33:** 79-86.

38. International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460:** 748-752.

39. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods* 2006; **11:** 193-206.

40. Kavvoura FK, Ioannidis JPA. Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. *Hum Genet* 2008; **123:** 1-14.

41. Ioannidis JPA, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet* 2001; **29:** 306-309.

42. Eastwood SL, Hodgkinson CA, Harrison PJ. DISC-1 Leu607Phe alleles differentially affect centrosomal PCM1 localization and neurotransmitter release. *Mol Psychiatry* 2009; **14:** 556-557.

43. Eastwood SL, Walker M, Hyde TM, Kleinman JE, Harrison PJ. The DISC1 Ser704Cys substitution affects centrosomal localization of its binding partner PCM1 in glia in human brain. *Hum Mol Genet* 2010; **19:** 2487-2496.

44. Nakata K, Lipska BK, Hyde TM, Ye T, Newburn EN, Morita Y *et al.* DISC1 splice variants are upregulated in schizophrenia and associated with risk polymorphisms. *Proc Natl Acad Sci USA* 2009; **106:** 15873-15878.

45. Callicott JH, Straub RE, Pezawas L, Egan MF, Mattay VS, Hariri AR *et al.* Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia. *Proc Natl Acad Sci USA* 2005; **102:** 8627-8632.

46. Kamiya A, Tomoda T, Chang J, Takaki M, Zhan CX, Morita M *et al.* DISC1-NDEL1/NUDEL protein interaction, an essential component for neurite outgrowth, is modulated by genetic variations of DISC1. *Hum Mol Genet* 2006; **15:** 3313-3323.

47. Li Y, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H *et al.* Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet* 2010; **42:** 969-972.

48. Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S *et al.* Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet* 2008; **40:** 584-591.

49. Wray NR, Visscher PM. Narrowing the Boundaries of the Genetic Architecture of Schizophrenia. *Schizophrenia Bull* 2010; **36:** 14-23.

50. Yang JA, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010; **42:** 565-U131.

51. McClellan JM, Susser E, King MC. Schizophrenia: a common disease caused by multiple rare alleles. *Brit J Psychiatry* 2007; **190:** 194-199.

52. Mitchell KJ, Porteous DJ. Rethinking the Genetic Architecture of Schizophrenia. *Schizophr Res* 2010; **117:** 222-222.

53. Devon RS, Anderson S, Teague PW, Burgess P, Kipari TM, Semple CA *et al.* Identification of polymorphisms within Disrupted in Schizophrenia 1 and Disrupted in Schizophrenia 2, and an investigation of their association with schizophrenia and bipolar affective disorder. *Psychiatr Genet* 2001; **11:** 71-78.

54. Wood LS, Pickering EH, Dechairo BM. Significant support for DAO as a schizophrenia susceptibility locus: examination of five genes putatively associated with schizophrenia. *Biol Psychiatry* 2007; **61:** 1195-1199.

55. O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V *et al.* Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet* 2008; **40:** 1053-1055.

56. Betcheva ET, Mushiroda T, Takahashi A, Kubo M, Karachanak SK, Zaharieva IT *et al.* Case-control association study of 59 candidate genes reveals the DRD2 SNP rs6277 (C957T) as the only susceptibility factor for schizophrenia in the Bulgarian population. *J Hum Genet* 2009; **54:** 98-107.

57. Need AC, Ge D, Weale ME, Maia J, Feng S, Heinzen EL *et al.* A genome-wide investigation of SNPs and CNVs in schizophrenia. *Plos Genet* 2009; **5:** e1000373.

58. Athanasiu L, Mattingsdal M, K√§hler AK, Brown A, Gustafsson O, Agartz I *et al.* Gene variants associated with schizophrenia in a Norwegian genome-wide study are replicated in a large European cohort. *J Psychiat Res* 2010.

59. Lencz T, Morgan TV, Athanasiou M, Dain B, Reed CR, Kane JM *et al.* Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry* 2007; **12:** 572-580.

60. Nicodemus KK, Kolachana BS, Vakkalanka R, Straub RE, Giegling I, Egan MF *et al.* Evidence for statistical epistasis between catechol-O-methyltransferase (COMT) and polymorphisms in RGS4, G72 (DAOA), GRM3, and DISC1: influence on risk of schizophrenia. *Hum Genet* 2007; **120:** 889-906.

61. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D *et al.* Common variants conferring risk of schizophrenia. *Nature* 2009; **460:** 744-747.

Tables

Captions;

Table 1: Details of studies analysed. [C] Candidate gene(s). [G] GWAS. [M] Minor allele frequencies. [S] Summary genotypes [F] Full genotypes. Gen genotyped. Imp Imputed.

Table 2: Studies which met our inclusion criteria but which were not included. "No response from author" refers to email response to the corresponding author. Specifically, [i] we received an initial response but no response to two further emails, [ii] no response after three initial emails.

Table 3: Nonsynonymous SNPs not found in dbSNP identified in 260 individuals. Two have been found in a previous study and four are novel.

Table 1

SNPs Tested

Study	Year	Gen	Imp	Cases	Controls
Devon *et al.* [CM 53]	2001	3	0	656	344
Hodgkinson *et al.* [CF 14]	2004	38	62	242	213

Wood *et al.* [CF 54]	2007	74	328	451	291
Song *et al.* [CS 21]	2008	3	0	288	288
Saetre *et al.* [CS 17]	2008	15	52	837	1473
Sanders *et al.* [CF 27]	2008	114	521	1870	2002
O'Donovan *et al.* [GM 55]	2008	115	180	478	2937
Betcheva *et al.* [CM 56]	2009	14	46	255	556
Hennah *et al.* [CS 12]	2009	62	136	1263	1275
Rastogi *et al.* [CS 16]	2009	9	0	210	210
Schumacher *et al.* [CM 18]	2009	108	210	782	839
Need *et al.* [GS 57]	2009	72	138	673	871
ISC *et al.* [GF 38]	2009	224	832	3322	3587
Athanasiu *et al.* [GF 58]	2010	250	955	298	351

Table 2

Study	Year	Reason for Exclusion
Thomson *et al.* [19]	2005	Samples reported in Hennah *et al.* [12]
Zhang *et al.* [20]	2006	Samples partially reported in Hennah *et al.* [12]
Lencz *et al.* [59]	2007	Samples reported in Hodgkinson *et al.* [14]
Nicodemus *et al.* [60]	2007	No response from author [i]
Stefansson *et al.* [61]	2009	No response from author [ii]

Table 3

Position (hg18)	NT Substitution	AA Substitution	Source	Song et al.?
229,896,375	C/T	A83V	[47]	Yes
229,896,606	G/T	W160L	[47]	No

229,896,987	C/T	P287L	[47]	No
229,897,063	T/G	C312W	[47]	No
229,969,594	T/C	I452F	[47]	No
230,969,594	A/T	E751Q	[29]	Yes

Figures

Figure 1: Reported DISC1 associations including both family and association studies from population of all ethnicities and indicating regions reported as being significantly associated with schizophrenia [7-20]. These includes family based linkage studies and studies in non-European populations which were not included in the present study.

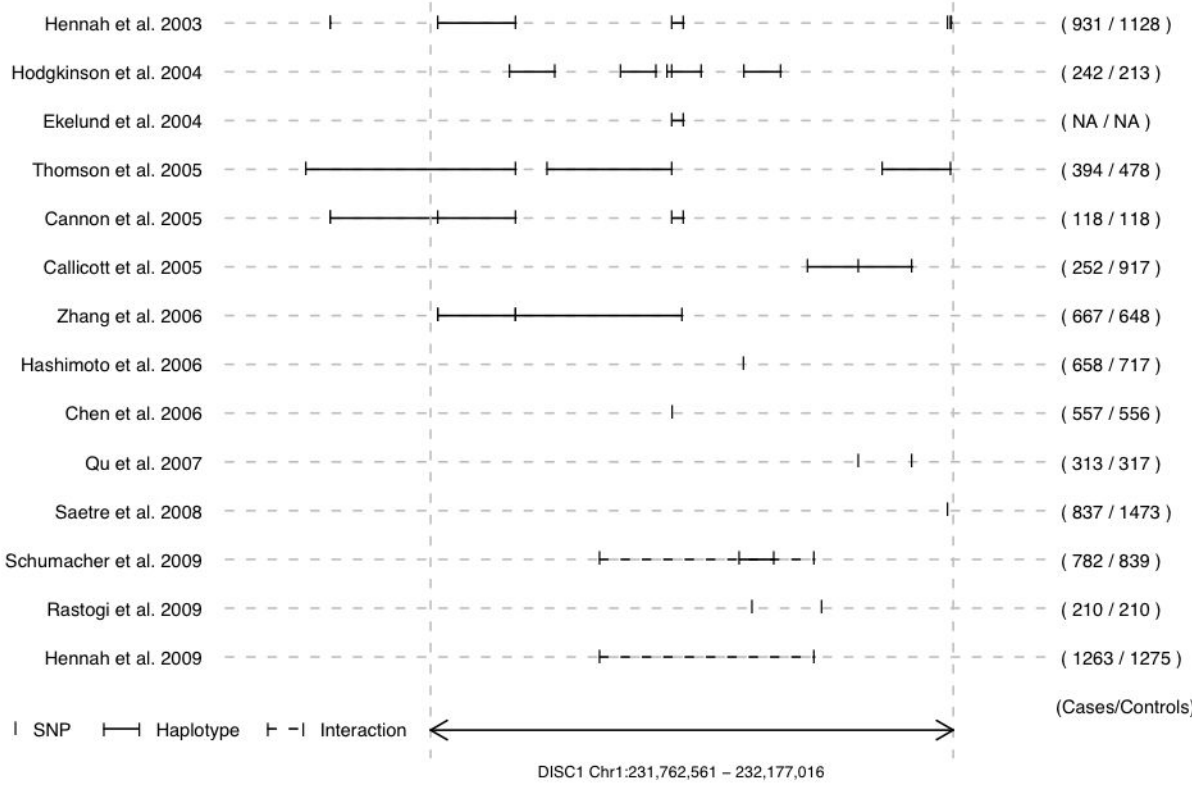
The numbers to the right of the graph indicate numbers of cases/controls in the study.

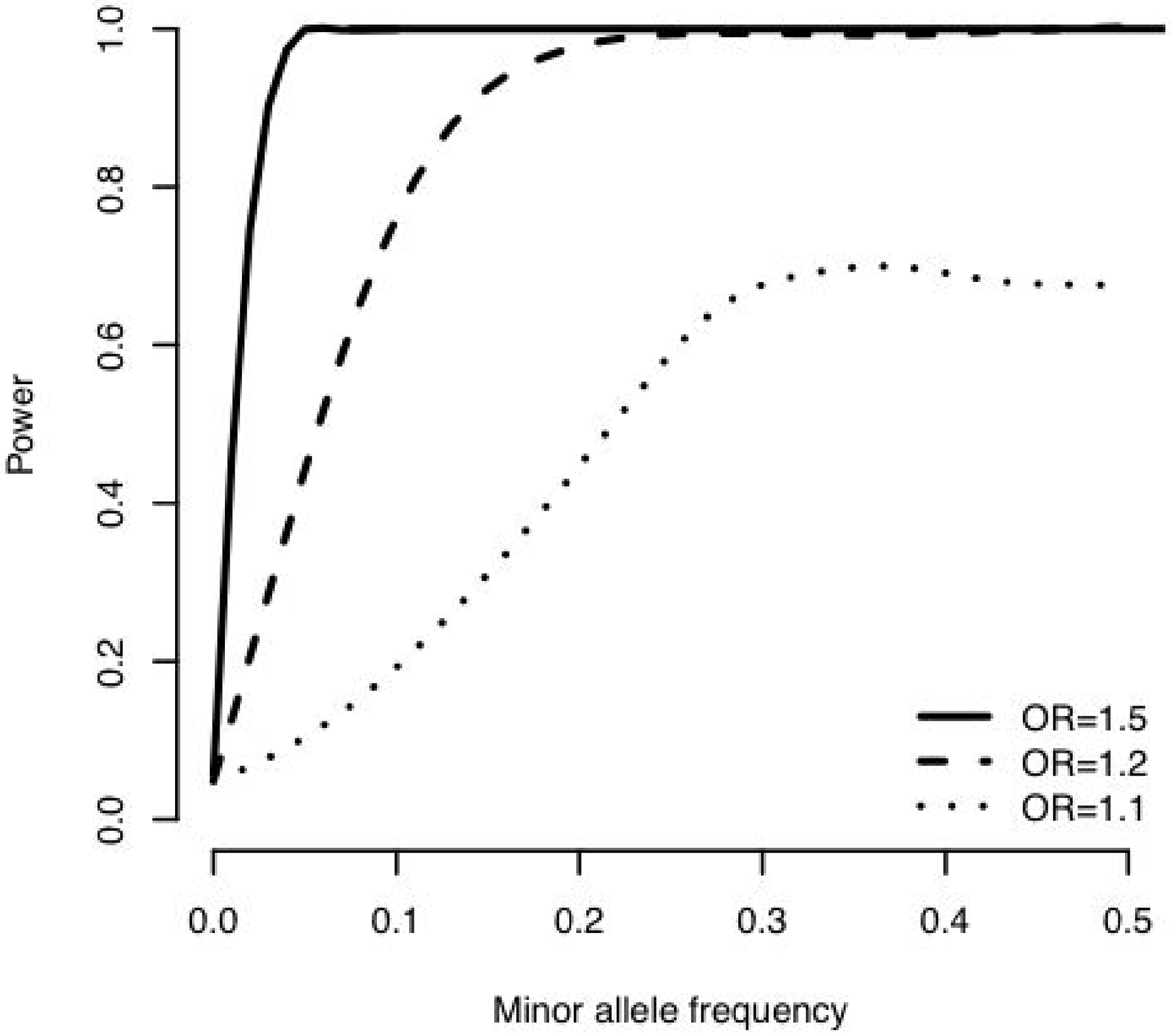
Figure 2: Estimated power as a function of minor allele frequency at the causal allele, showing results for different odds ratios. The power was simulated using variants and haplotypes from the 1000 genomes data and the curves have been fitted as smooth splines of degree 6.

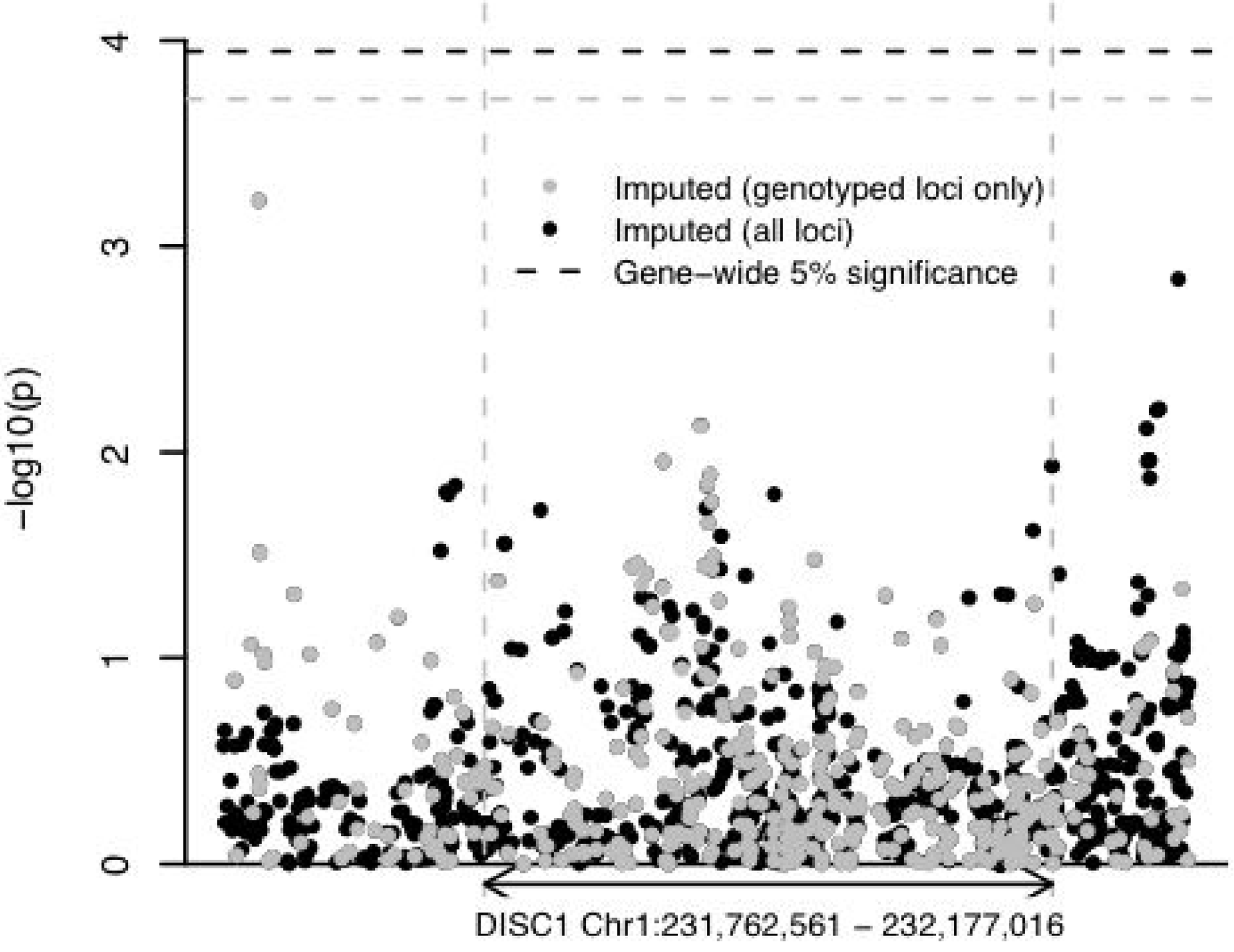
Figure 3: Pooled P-values for association. Each point represents a single SNP. The horizontal axis indicates chromosomal position, and the vertical axis indicates negative $\log_{10}$ of the fixed effect pooled P-value for that SNP. The black points show the imputed SNPs, and the grey points show the subset of the imputed SNPs which were genotyped in at least one study. The horizontal dashed lines show the corrected significance level which is equivalent to a 5% gene-wide significance level for each dataset.

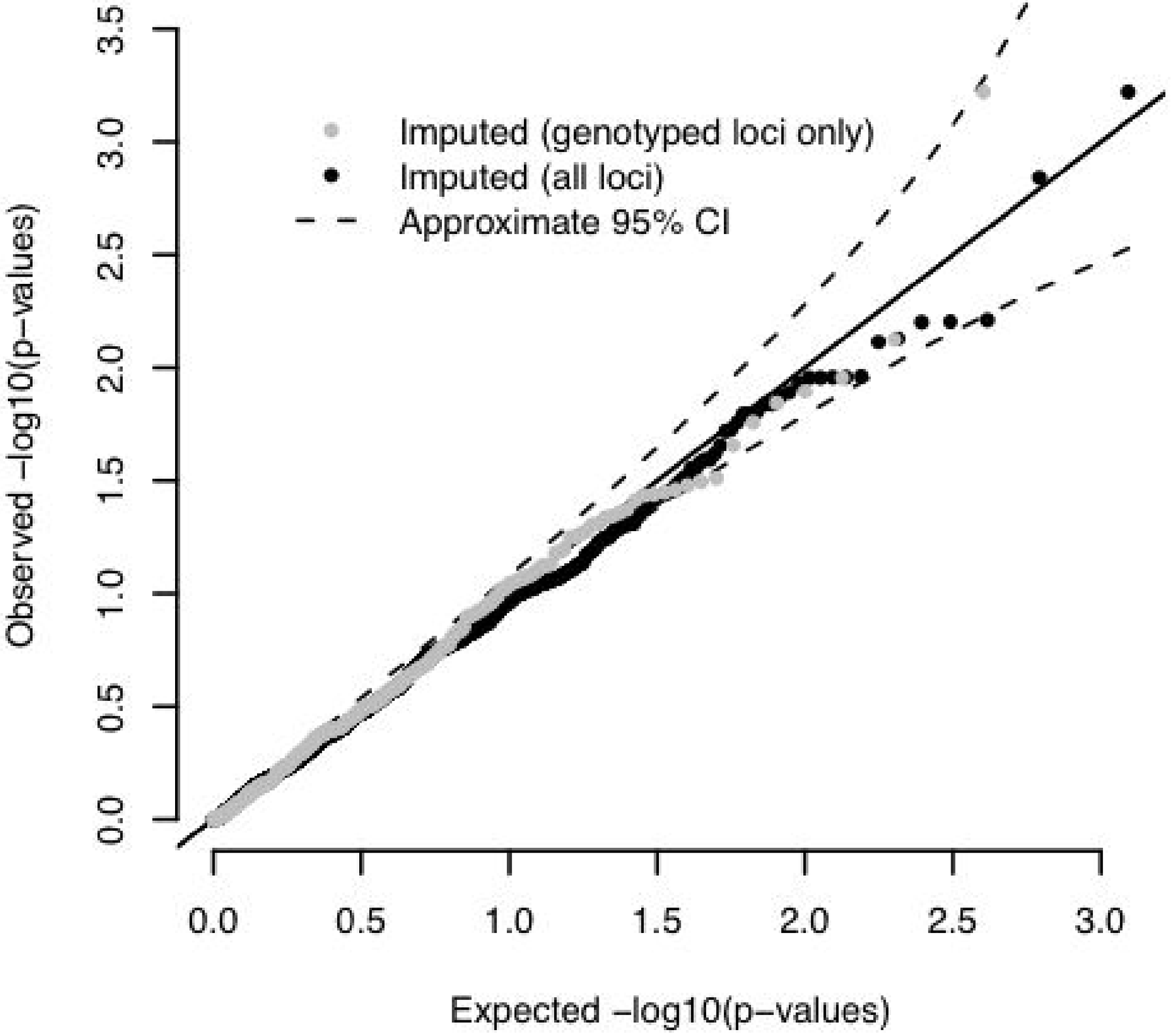
Figure 4: Quantile-quantile plot of P-values. QQ plot of the p-values shown in figure 2. The solid line shows the expected values under the null hypothesis of no association and the dashed lines show the upper and lower 95% quantiles.

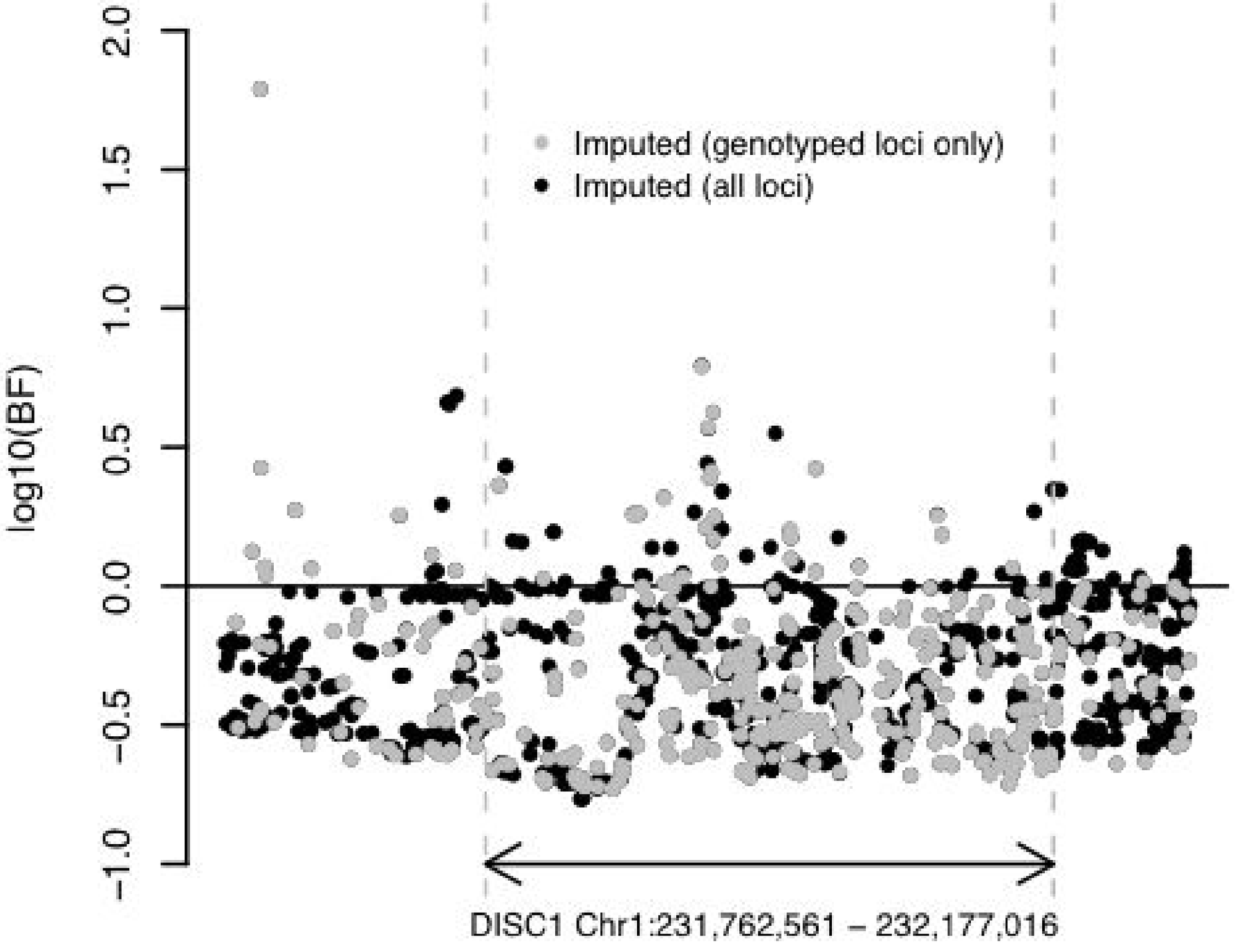
Figure 5: Wakefield approximate Bayes factors for each SNP. Calculated from the random effects pooled P-values and Z-scores assuming a Normal prior on the log-odds ration with mean 0 and variance of 0.015. A Bayes factor > 0 is evidence for association while a Bayes factor <0 is evidence against.


Hennah et al. 2003
(931 / 1128)
Hodgkinson et al. 2004
(242 / 213)
Ekelund et al. 2004
(NA / NA)
Thomson et al. 2005
(394 / 478)
Cannon et al. 2005
(118 / 118)
Callicott et al. 2005
(252 / 917)
Zhang et al. 2006
(667 / 648)
Hashimoto et al. 2006
(658 / 717)
Chen et al. 2006
(557 / 556)
Qu et al. 2007
(313 / 317)
Saetre et al. 2008
(837 / 1473)
Schumacher et al. 2009
(782 / 839)
Rastogi et al. 2009
(210 / 210)
Hennah et al. 2009
(1263 / 1275)
(Cases/Controls)
SNP
Haplotype
Interaction
DISC1 Chr1:231,762,561 – 232,177,016


Power
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Minor allele frequency
OR=1.5
OR=1.2
OR=1.1


Imputed (genotyped loci only)
Imputed (all loci)
Gene−wide 5% significance
−log10(p)
0
1
2
3
4
DISC1 Chr1:231,762,561 − 232,177,016


Imputed (genotyped loci only)
Imputed (all loci)
Approximate 95% CI
Observed −log10(p−values)
Expected −log10(p−values)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5


log10(BF)
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
Imputed (genotyped loci only)
Imputed (all loci)
DISC1 Chr1:231,762,561 – 232,177,016