# Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder

Williams, H.J.; Norton, N.; Dwyer, S.; Moskvina, V.; Nikolov, I.; Carroll, L.; Georgieva, L.; Williams, N.M.; Morris, D.W.; Quinn, E.M.

Link to publication in University of Groningen/UMCG research database

npg

# ORIGINAL ARTICLE

# Fine mapping of *ZNF804A* and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder

HJ Williams[1], N Norton[1], S Dwyer[1], V Moskvina[1], I Nikolov[1], L Carroll[1], L Georgieva[1], NM Williams[1], DW Morris[2], EM Quinn[2], I Giegling[3], M Ikeda[1,4], J Wood[5,6], T Lencz[7,8,9], C Hultman[10], P Lichtenstein[10], D Thiselton[11,12], BS Maher[11,12,13], Molecular Genetics of Schizophrenia Collaboration (MGS)[14], International Schizophrenia Consortium (ISC)[14], SGENE-plus[14], GROUP[14], AK Malhotra[7,8,9], B Riley[11,12,13], KS Kendler[11,12,13], M Gill[2], P Sullivan[15], P Sklar[16,17,18,19], S Purcell[16,17,18,19], VL Nimgaonkar[5,6], G Kirov[1], P Holmans[1], A Corvin[2], D Rujescu[3], N Craddock[1], MJ Owen[1] and MC O'Donovan[1]

[1]MRC Centre for Neuropsychiatric Genetics and Genomics, Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff, UK; [2]Neuropsychiatric Genetics Research Group, School of Medicine, Trinity College Dublin, Dublin, Ireland; [3]Division of Molecular and Clinical Neurobiology, Ludwig-Maximilians-University, Munich, Germany; [4]Department of Psychiatry, Fujita Health University School of Medicine, Aichi, Japan; [5]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; [6]Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA; [7]Center for Translational Psychiatry, The Feinstein Institute for Medical Research, North Shore—Long Island Jewish Health System, Manhasset, NY, USA; [8]Division of Research, The Zucker Hillside Hospital, Glen Oaks, NY, USA; [9]Department of Psychiatry and Behavioral Sciences, Albert Einstein College of Medicine, Bronx, NY, USA; [10]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; [11]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA; [12]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA; [13]Department of Human Genetics, Virginia Commonwealth University, Richmond, VA, USA; [15]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; [16]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; [17]Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; [18]Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA and [19]Broad Institute of Harvard and MIT, Cambridge, MA, USA

**A recent genome-wide association study (GWAS) reported evidence for association between rs1344706 within *ZNF804A* (encoding zinc-finger protein 804A) and schizophrenia ($P = 1.61 \times 10^{-7}$), and stronger evidence when the phenotype was broadened to include bipolar disorder ($P = 9.96 \times 10^{-9}$). In this study we provide additional evidence for association through meta-analysis of a larger data set (schizophrenia/schizoaffective disorder $N = 18\,945$, schizophrenia plus bipolar disorder $N = 21\,274$ and controls $N = 38\,675$). We also sought to better localize the association signal using a combination of *de novo* polymorphism discovery in exons, pooled *de novo* polymorphism discovery spanning the genomic sequence of the locus and high-density linkage disequilibrium (LD) mapping. The meta-analysis provided evidence for association between rs1344706 that surpasses widely accepted benchmarks of significance by several orders of magnitude for both schizophrenia ($P = 2.5 \times 10^{-11}$, odds ratio (OR) 1.10, 95% confidence interval 1.07–1.14) and schizophrenia and bipolar disorder combined ($P = 4.1 \times 10^{-13}$, OR 1.11, 95% confidence interval 1.07–1.14). After *de novo* polymorphism discovery and detailed association analysis, rs1344706 remained the most strongly associated marker in the gene. The allelic association at the *ZNF804A* locus is now one of the most compelling in schizophrenia to date, and supports the accumulating data suggesting overlapping genetic risk between schizophrenia and bipolar disorder.**
*Molecular Psychiatry* (2011) **16,** 429–441; doi:10.1038/mp.2010.36; published online 6 April 2010

## Introduction

Genetic epidemiology reveals that genes account for >80% of the population variance in risk[1,2] of schizophrenia. Numerous genetic associations to schizophrenia have been reported, but which of these, if any, are true associations remains widely disputed.[3]

The advent of genome-wide association study (GWAS) technology has recently proven to be successful in allowing the identification of strongly supported genetic associations for many other common phenotypes, and this is now true for schizophrenia.[4–7]

Recently, we undertook a GWAS of 479 UK schizophrenia cases and 2937 controls,[4] with follow-up of the strongest findings in approximately 17 000 subjects. In the combined data set of approximately 20 000 subjects, the single-nucleotide polymorphism (SNP) with the strongest evidence for association ($P = 1.61 \times 10^{-7}$) to schizophrenia was rs1344706 within *ZNF804A* (encoding zinc-finger protein 804A). Although this falls short of a widely accepted threshold for genome-wide significant association of $P < 7.2 \times 10^{-8}$,[8] we did obtain evidence that surpasses this ($P = 9.96 \times 10^{-9}$), when the phenotype was broadened to include patients with bipolar disorder, a phenotype for which there is considerable overlap in clinical features and increasing epidemiological and molecular genetic evidence for shared genetic risk with schizophrenia.[9] Subsequently, independent associations between schizophrenia and the same allele of rs1344706 have been reported by the International Schizophrenia Consortium,[5] The Irish Case/Control Study of Schizophrenia[10] and the SGENE-plus consortium.[11] In a total of over 5000 patients with psychiatric disorders, the latter[11] also reported three copy number variants (CNVs) at the *ZNF804A* locus, a deletion in an individual with schizophrenia, a duplication in an individual with bipolar disorder and a deletion in an individual with anxiety disorder. This contrasted with no CNVs at the locus in almost 40 000 controls ($P = 0.0016$). From additional reported data sets, they also noted duplication CNVs in three people with autism, and an additional three CNVs out of a total of approximately 12 000 controls, two of whom were $< 18$ years and were therefore not past the characteristic age at onset for schizophrenia. Overall, the CNV data suggest that rare structural variants at the *ZNF804A* locus may also be involved in the risk of a range of psychiatric disorders, although support for this is not unequivocal.

As our GWAS[4] was based upon fewer than 1/20th of all common SNPs in the genome, it seemed unlikely that the best associated SNP in our study was the true functional variant. In the present study, we tried to localize the association signal using a combination of *de novo* polymorphism discovery targeted at exons, pooled *de novo* polymorphism discovery spanning the genomic sequence of the locus and high-density linkage disequilibrium (LD) mapping. rs1344706, the SNP originally highlighted by the original GWAS study, remained the most strongly associated SNP at the locus. The evidence for association for rs1344706 was further tested by a meta-analysis of approximately 60 000 subjects (schizophrenia/schizoaffective disorder $N = 18\,945$, schizophrenia plus bipolar disorder $N = 21\,274$ and controls $N = 38\,675$) derived from the original publication and additional data sets. The overall statistical support in schizophrenia surpassed genome-wide significance ($P = 2.5 \times 10^{-11}$) by more than three orders of magnitude, as did the support in schizophrenia and bipolar disorder combined ($P = 4.1 \times 10^{-13}$). Moreover, the evidence for association in replication samples (that is, after exclusion of the discovery GWAS) was also genome-wide significant, making rs1344706 in *ZNF804A* the most robustly supported common allele for schizophrenia reported to date. The strength of the findings also strongly support accumulating data that genetic risk for schizophrenia and bipolar disorder[12] overlap.

## Materials and methods

### *Targeted* de novo *polymorphism discovery in exons*
Coding sequences were screened for polymorphisms in 163 UK individuals with schizophrenia and 135 UK blood donor controls. Genomic sequences corresponding to all exons were extracted from the UCSC Genome Browser (March 2006), based upon Refseq transcript NM_194250. Amplicons with a maximum length of 500 bp ($n = 13$) were designed to span coding exons plus a minimum of 72 bases of flanking sequence. Where multiple amplicons were required to cover an exon, these overlapped by a minimum of 81 bp. High-resolution melting analysis was performed for 11 amplicons using a LightScanner (Idaho Technologies, Salt Lake City, Utah, USA) according to the manufacturer's instructions. Potential DNA variants identified by this approach were characterized by sequencing the relevant PCR product using BigDye chemistry on a 3100 capillary sequencer (Applied Biosystems, Foster City, CA, USA). Two amplicons failed to optimize in the presence of the high-resolution melting analysis detection reagent and were screened for variants by sequencing. Identified SNPs were genotyped in the CEU HapMap trios to assess their LD relationships and to detect Mendelian inconsistencies (of which there were none).

### *Identification of SNPs associated with* ZNF804A *expression (eQTLs)*
To identify SNPs that might be expression quantitative trait loci (eQTLs), we extracted genotype and *ZNF804A* lymphoblastoid expression data from the CEU HapMap samples deposited in GeneVar (http://www.sanger.ac.uk/humgen/genevar/). To analyse these data, we used the method described by the group that developed the database.[13]

### *Association studies*

*Functional and tagSNP (fatSNP) study.* A two-stage fatSNP study was designed to extract a high proportion of the known genetic variation spanning *ZNF804A* (chr2:184958600–185722680, National Center for Biotechnology Information (NCBI) b36), with putative functional variants given particular priority. Details of all samples used in these two stages of the study are given in our previous manuscript.[4]

*fatSNP stage 1.* SNPs: We aimed to tag at $r^2 = 1$ all SNPs with minor allele frequency (MAF) of $\geqslant 0.01$ in the HapMap CEU samples plus SNPs identified by *de novo* polymorphism discovery of exons. This panel of SNPs was derived using Tagger (Haploview, version 4.1[14]) with forced inclusion of all synonymous and non-synonymous variants identified by *de novo* discovery and from public databases, and all SNPs associated with expression of *ZNF804A* mRNA in the GeneVar database. We also force included SNPs previously genotyped in the previous GWAS.[4]

Samples: In fatSNP stage 1 (fatSNP1), the resultant set of SNPs (excluding those for which data were already available from the GWAS study) were genotyped in a subset of the GWAS samples from the earlier study.[4] This fatSNP1 sample comprised 479 UK cases with schizophrenia and the 1958 birth cohort control sample ($N = 1445$) used by the Wellcome Trust Case/Control Consortium (WTCCC)[15] that we had also used in our GWAS study. The results of fatSNP1 were used to select SNPs for follow-up in additional samples using criteria detailed below.

*fatSNP stage 2.* SNPs: For follow-up in fatSNP stage 2 (fatSNP2), we identified all markers that were associated with $P \leqslant 1 \times 10^{-4}$ in fatSNP1, or that had surpassed that threshold in the GWAS study. We additionally followed up non-synonymous SNPs that were associated with $P \leqslant 0.10$ in those analyses. To prune the marker set, we applied a highly conservative definition of non-redundancy, removing one of any pair of markers that were highly correlated ($r^2 \geqslant 0.97$) in the fatSNP1 sample.

Samples: For fatSNP phase 2, we genotyped the UK National Blood Service controls ($N = 1428$) used by the WTCCC,[15] which we had also used in our GWAS study, and additional UK schizophrenia samples ($N = 163$) for whom GWAS array data were not available, which we had used as part of the follow-up sample in the previous publication.[4] The combined set of UK cases ($N = 642$) and blood donor/1958 birth cohort controls ($N = 2873$ available for in-house genotyping, $N = 2937$ with Affymetrix, Santa Clara, CA, USA; 500K data) are subsequently referred to for brevity as the Cardiff Full sample. fatSNP2 also included three additional case–control series from Bulgaria, Ireland (referred to as the Dublin sample) and Germany (referred to as the Munich sample), the combination of which is referred to, for brevity (and consistency with the previous paper), as REP1 and comprises an additional 1664 cases and 3541 controls. All cases met DSM-IV (Diagnostic and Statistical Manual of Mental Disorders-fourth edition) criteria for schizophrenia; full details of these samples are given in our previous study.[4] The results of the REP1 sample were combined with those of the Cardiff Full sample to derive an overall fatSNP2 *P*-value as described below.

*fatSNP genotyping and analysis*
Most SNP assays for the fatSNP studies were genotyped in the Cardiff laboratory using the Sequenom iPlexGold system (San Diego, CA, USA). Additional SNPs were genotyped in Cardiff using TaqMan (on demand) probes (Applied Biosystems; rs17584522) or Amplifluor AssayArchitect (Millipore, Billerica, MA, USA; rs12476147) using an Analyst AD fluorometer (Molecular Devices, Silicon Valley, CA, USA). In addition, rs17584522 was genotyped by the Dublin group with TaqMan (on demand) probes using a 7900HT Sequence Detection System (Applied Biosystems). For SNPs for which we could not design a reliable assay using either method, we genotyped the HapMap CEU sample using SNaPshot chemistry (Applied Biosystems) and a 3100 capillary sequencer (Applied Biosystems). For rs5836928, a 3 bp insertion deletion, we genotyped the HapMap CEU sample using fluorescently tagged primers with genotypes discriminated by size using a 3100 capillary sequencer (Applied Biosystems).

For quality control estimates we determined call rates for each SNP in all samples. Included in the Dublin and Munich samples were duplicated DNA samples. Within the UK, Bulgarian and Dublin sample, we included samples from the HapMap CEU population to determine genotyping congruence against the HapMap data. In the sample genotyped by the Dublin group for marker rs17584522, there were both duplicate samples and HapMap samples with which to check genotyping quality control.

fatSNP statistical analyses were as in the previous study.[4] We used a trend test for within-sample analysis and to combine the data across samples, a Cochran–Mantel–Haenszel test conditioning by site as implemented in PLINK version 1.05.[16]

Imputation was performed on the fatSNP1 sample with both PLINK[16] using the proxy-drop command and Beagle version 3.0.[17]

*Genomic resequencing*
After completion of fatSNP2, we aimed to identify SNPs within the wider genomic regions not targeted by the exon-focussed *de novo* discovery phase. This was performed using 60 cases drawn randomly from the Cardiff Full sample. From these, we generated six pooled samples each containing DNA from 10 samples. PCR primers were designed to span 381 988 bp of the genomic sequence in and around *ZNF804A* (Figure 1). Long-range PCRs were generated using the Roche Expand Long Template PCR system (Basel, Switzerland). The PCR products were gel purified after electrophoresis, and sent to Illumina (San Diego, CA, USA) for sequencing using Solexa technology. The resultant single-end reads were assembled and analysed for putative SNPs using both DNAStar (http://www.dnastar.com) and MAQ (http://maq.sourceforge.net). We used the default parameters for DNAStar analysis. The MAQ output was filtered using previously published thresholds.[18]

*Illumina genotyping*
Putative SNPs identified after genomic resequencing were annotated according to the Illumina OPA (Oligo Pool All) design criteria, and the files were submitted
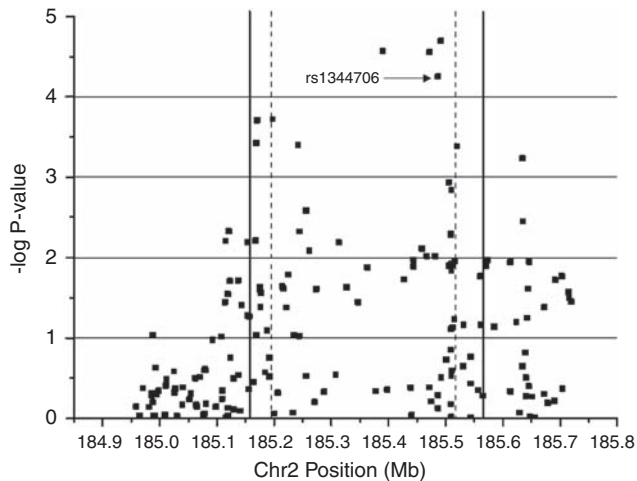
432



**Figure 1** fatSNP1 results for *ZNF804A*. Armitage trend *P*-values ($-\log_{10}$) for 176 SNPs genotyped during fatSNP phase 1. Dashed lines represent the boundaries of *ZNF804A* transcript (NM_194250). Solid lines represent the region targeted by genomic resequencing. rs1344706 was the most significant SNP in the earlier GWAS study (O'Donovan *et al*.[4]). Chromosome position based on NCBI build36.

for assay design (http://illumina.com). Designed SNPs were genotyped in the CEU trios by the Cardiff laboratory using the GoldenGate Genotyping assay for VeraCode and analysed using a BeadXpress instrument using VeraScan 1.1 (Ilumina). BeadStudio V3.2 (Illumina) was used to make genotype calls and compile the data output.

*Quantitation of relative allelic expression*
To isolate *cis*-effects from *trans*-effects on gene expression in human brain, we measured the relative expression of each parental copy of *ZNF804A* in post-mortem brain mRNA taken from individuals heterozygous for expressed SNP rs4667001, a marker that is in strong LD ($D' = 1$) with rs1344706 in that brain series ($D' = 0.95$ in our association sample). We have described the methodology and the samples (derived from cerebral cortex of 149 unrelated anonymous individuals (86 males, 63 females; mean age 58 years, s.d. = 19)) in detail elsewhere.[19] As this method uses the level of mRNA expressed from one parental chromosome to control for that expressed by the other, the assay is controlled for (and therefore requires no adjustment for) common confounders that affect more standard measurements of mRNA in native human tissue (for example, drug exposure, agonal state, post-mortem delay, pH, age, and so on).

*Meta-analysis of additional data sets*

*Schizophrenia.* To further test the evidence for association, we combined the data for rs1344706 from our earlier publication[4] with data from four published GWAS studies, one unpublished Swedish GWAS data set (termed SW3, from P Sullivan/C Hultman, unpublished data), data from additional

samples we genotyped in house and data from a *ZNF804A* candidate gene study from a group who had contacted us before genotyping their sample[10] and whose data are therefore unbiased with regard to the result.

The numbers of cases and controls for each of the samples that we included in the meta-analysis are given in Supplementary Table 5. Details of the samples from the published GWAS studies of the International Schizophrenia Consortium (ISC),[5] the Molecular Genetics of Schizophrenia (MGS) consortium,[7] the SGENE-plus consortium[6] and of Lencz *et al*.[20] are given in the primary GWAS publications. The SW3 sample had been genotyped using an Affymetrix 6 chip at the Broad Institute of MIT and was part of the same on-going patient collections as samples SW1 and SW2 from the ISC data set.[21] Details of the samples used in the candidate gene study of *ZNF804A* based upon the Irish Case/Control Study of Schizophrenia sample are also available in a primary manuscript.[10] It should be noted that those samples do not overlap with any of the Irish samples provided by the Dublin group.

Further to their GWAS study, the SGENE-plus and GROUP consortia obtained genotypes on additional samples from Europe and China (2865 schizophrenia cases and 4493 controls) using a Centaurus assay (Nanogen, San Diego, CA, USA). Details of these are given in Steinberg *et al*.[11] but briefly, these comprise schizophrenia case/control samples from China (460/466), Denmark/Aarhus (236/500), Denmark/Copenhagen (513/1338), Germany/Bonn (275/510), Germany/Munich (178/ 320), Hungary (264/ 223), Norway (201/357), Russia (483/ 487) and Sweden (255/292).

In addition, we obtained genotype data on a sample (488 cases and 540 controls) collected by the Pittsburgh group, details of which are given in Talkowski *et al*.[22] The group from Dublin (Corvin and colleagues) provided genotypes based upon a Taqman assay for an extra set of 352 Irish cases and 178 Irish controls who had not been included in fatSNP2. The diagnostic and ascertainment practices were as for the part of their sample described in O'Donovan *et al*.[4]

Subjects who overlapped between studies were removed as follows. In our previous paper, we reported data from the MGS European American sample. However, as the GWAS sample from that group was larger than what we had used, we include in this study the MGS GWAS data rather than our earlier data. The Bulgarian and Irish samples included in fatSNP2 substantially overlap with samples used by the ISC; hence, the ISC data for those populations were excluded. A data set from Aberdeen was included in the GWAS of both the ISC and SGENE-plus. The SGENE-plus data were excluded. It should be additionally noted that the samples from Bonn, Munich, and China that were included in the SGENE-plus study[11] did not overlap with the samples we used in our previous paper.[4] We did not include any data from the GWAS of Need *et al*.[23] as all those

data are subsumed in other data sets included here. We did not include data from the CATIE study[24] as the controls from that study substantially overlap with those used by the MGS study.[7]

The use of unadjusted genotype counts is not appropriate for some samples (for example, SGENE-plus because of relatedness among the Icelandic Groups, and MGS because of extensive structure in the data set); therefore, we used the inverse variance method of meta-analysis. For each study, variance was calculated based on the $2 \times 2$ contingency table counts, or the 95% confidence interval values provided by SGENE-plus, the MGS and the unpublished Swedish GWAS. Heterogeneity between studies was assessed using Cochran's Q statistic.

*Bipolar disorder.* Given that we had previously reported stronger evidence for association findings with the inclusion of bipolar samples, to the above meta-analysis, we added the UK bipolar data set of the WTCCC.[15] Data for additional bipolar samples from Iceland ($n = 404$) and Norway ($n = 205$) were also provided by the SGENE-plus/GROUP consortia. To avoid the problem of shared controls, as before,[4] the WTCCC bipolar data were combined with the UK schizophrenia cases before testing against the UK controls. Similarly, the Icelandic and Norwegian cases were combined with the respective schizophrenia cases from these countries.

## Results

### *Targeted* de novo *polymorphism detection*
We detected no synonymous or non-synonymous variants in exons 1–3. In exon 4, we detected 21 variants, 14 non-synonymous and 7 synonymous (Supplementary Table 1).

### *eQTL results*
In the GeneVar CEU database, we found three SNPs in high LD ($r^2 = 1$) associated with expression of *ZNF804A* mRNA ($P = 0.006$). An additional 16 SNPs, including the best GWAS SNP in *ZNF804A* (rs1344706), were associated with expression at $P < 0.05$ (Supplementary Table 2). The disease-associated allele of rs1344706 was associated with higher expression.

### *fatSNP study*

*fatSNP1.* Within the target region, 887 SNPs were listed in HapMap (Rel 23a/phase II Mar08 SNP database b126) of which 508 had MAF of $\geqslant 0.01$. To these, we added the genotype data for 21 SNPs from our exonic *de novo* polymorphism detection. From these, we identified 209 non-redundant ($r^2 = 1$ in CEU) markers. Of the 209 markers, we were able to obtain genotypes for 176 in fatSNP phase 1. Overall, the genotyped markers provide coverage of 91% of target alleles with MAF of $\geqslant 0.01$ at $r^2 = 1$ and 96% coverage at $r^2 \geqslant 0.9$. For alleles with MAF of $\geqslant 0.05$, the respective figures are 93 and 97%. Association

results are summarized in Figure 1 and in Supplementary Table 3. A total of 12 markers (in addition to rs1344706) met our criteria for follow-up (Table 1; note that rs12613195 was inc1uded as it attained the threshold in the full GWAS sample. Also note that rs6726421 was taken forward as a perfect proxy for the non-synonymous SNPs described in this table). Notably, in this subsample of our GWAS (fatSNP1), three markers yielded slightly stronger evidence for association than our original GWAS marker (Table 1). Of these, two (rs1583048 and rs3931790, $r^2 = 1$) were the putative eQTLs with strongest evidence for association with expression in GeneVar (rs7593816, an equally strong eQTL that is perfectly correlated with those SNPs that failed genotyping), with the schizophrenia-associated allele being associated with higher gene expression (Supplementary Table 2). The other SNP, rs17584522, is intronic.

*fatSNP2.* Of the 12 SNPs (Table 1) meeting the association criteria for follow-up, we selected 8 SNPs based upon $r^2 < 0.97$ (Table 2). In the Cardiff Full sample, only intronic marker rs17584522 was associated with schizophrenia at a level of significance within one order of magnitude of that of rs1344706, the initial 'hit-SNP' from our earlier paper (Table 2). Table 2 shows that all SNPs tested in fatSNP phase 2 are weakly and moderately correlated with rs1344706 ($r^2_{max} = 0.43$) but are in moderately strong LD ($D'_{min} = 0.7$). To test whether any of these associations are independent of rs1344706, we performed forward stepwise logistic regression analysis in the Cardiff Full sample, including all fatSNP phase 2 SNPs. Only one SNP (rs12613195) was nominally significant ($P = 0.021$) after allowing for the effects of rs1344706 (data not shown). Given the multiple testing of SNPs, we conclude that there is no convincing evidence for an association signal independent of rs1344706. Moreover, haplotype analysis based upon fatSNP phase 2 markers did not produce results more significant than that of single marker analysis (data not shown). A combined analysis of the fatSNP2 samples revealed rs1344706 as the most significantly associated SNP ($P = 8.31 \times 10^{-6}$). Single locus data for each of the individual samples are given in Supplementary Table 4.

### *Quality control estimates*
In each fatSNP2 sample, SNPs had call rates in cases and controls of $\geqslant 97\%$ with the exception of rs12476147 in the UK sample (92% in cases), rs12613195 in the Munich sample (96% in cases) and both rs1344706 (93% in cases) and rs17584522 in the Dublin sample (95 and 93% in cases and controls, respectively; Supplementary Table 4). In the UK samples, the more complete imputed genotypes for rs12476147 gave similar results to the array SNP. No marker in any of the populations studied had a Hardy-Weinberg equilibrium *P*-value $< 0.05$. Of $\sim 1300$

**Table 1** Results of fatSNP phase 1

| | SNP | | | UK fatSNP1 case = 479 con = 1445 | | | |
|---|---|---|---|---|---|---|---|
| *SNP ID* | *Position*[a] | *Class* | *Source*[b] | *Case*[c] | *Con*[c] | *OR* | *ARM* |
| rs12613195[d] | 185197465 | Intronic | GWAS | 0.27 | 0.34 | 0.73 | $2.19 \times 10^{-4}$ |
| rs17584522 | 185389974 | Intronic | fatSNP | 0.18 | 0.25 | 0.68 | $2.69 \times 10^{-5}$ |
| rs3931790[e] | 185471620 | Intronic | eQTL | 0.15 | 0.21 | 0.66 | $2.77 \times 10^{-5}$ |
| rs1344706 | 185486672 | Intronic | GWAS | 0.33 | 0.41 | 0.72 | $4.82 \times 10^{-5}$ |
| rs1583048[e] | 185491385 | Intronic | eQTL | 0.15 | 0.21 | 0.65 | $2.00 \times 10^{-5}$ |
| rs6726421[f] | 185505086 | Intronic | fatSNP | 0.45 | 0.41 | 1.22 | 0.011 |
| rs12476147 | 185509149 | Non-synonymous | *de novo* | 0.41 | 0.37 | 1.22 | 0.012 |
| ss177930705 | 185509348 | Non-synonymous | *de novo* | 0.04 | 0.03 | 1.42 | 0.077 |
| rs35676856[g] | 185509803 | Non-synonymous | *de novo* | 0.09 | 0.06 | 1.47 | 0.0052 |
| rs61739290[g] | 185509841 | Non-synonymous | *de novo* | 0.09 | 0.07 | 1.46 | 0.0050 |
| rs4667001[f] | 185509991 | Non-synonymous | *de novo* | 0.45 | 0.41 | 1.21 | 0.015 |
| rs1366842[f] | 185510487 | Non-synonymous | *de novo* | 0.45 | 0.41 | 1.22 | 0.012 |
| rs3731834 | 185511608 | Non-synonymous | *de novo* | 0.15 | 0.17 | 0.83 | 0.074 |

Abbreviations: ARM, Armitage trend test *P*-value; con, controls; eQTL, expression quantitative trait locus; fatSNP, functional and tagSNP; GWAS, genome-wide association study; OR, odds ratio; SNP, single-nucleotide polymorphism.
[a]Position is based on NCBI build36.
[b]Source is the phase of study the SNP is derived from.
[c]Case and Con show minor allele frequency.
[d]*P*-value in larger GWAS sample (O'Donovan *et al.*[4]) = $2.18 \times 10^{-5}$.
[e]$r^2 > 0.97$ in CEPH sample, GWAS sample and Bulgarian REP1 sample.
[f]$r^2 > 0.99$ in CEPH sample, GWAS sample and Bulgarian REP1 sample.
[g]$r^2 > 0.99$ in CEPH sample, GWAS sample and Bulgarian REP1 sample.

**Table 2** Results of fatSNP phase 2 and schizophrenia meta-analysis

| | | | | Cardiff Full | | REP1 | | fatSNP2 | | SZ meta | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Cases n = 642 controls n = 2873 | | Cases n = 1664 controls n = 3541 | | Cases n = 2306 controls n = 6414 | | Cases n = 18 945 controls n = 38 675 | |
| *SNP ID* | *Class* | *Allele* | *$r^2$ (D')* | *OR* | *ARM* | *OR* | *CMH*[a] | *OR* | *CMH* | *OR* | *P* |
| rs12613195 | GWAS | C | 0.33 (0.75) | 1.35 | $1.20 \times 10^{-5}$ | 1.09 | 0.038 | 1.17 | $6.03 \times 10^{-5}$ | | |
| rs17584522 | fatSNP | C | 0.24 (0.7) | 1.48 | $6.21 \times 10^{-7}$ | 1.08 | 0.080 | 1.20 | $5.34 \times 10^{-5}$ | | |
| rs1344706 | GWAS | T | NA | 1.38 | $7.08 \times 10^{-7}$ | 1.09 | 0.026 | 1.18 | $8.31 \times 10^{-6}$ | 1.10 | $2.54 \times 10^{-11}$ |
| rs1583048 | eQTL | A | 0.36 (1) | 1.46 | $7.40 \times 10^{-6}$ | 1.08 | 0.10 | 1.18 | $2.85 \times 10^{-4}$ | | |
| rs6726421[b] | Tag non-synonymous | C | 0.43 (0.95) | 1.21 | 0.0022 | 1.12 | 0.0050 | 1.15 | $1.06 \times 10^{-4}$ | | |
| rs12476147[c] | Non-synonymous | A | 0.43 (0.94) | 1.23 | 0.0055 | 1.11 | 0.011 | 1.14 | $6.54 \times 10^{-4}$ | | |
| ss177930705 | Non-synonymous | C | 0.01 (1) | 1.13 | 0.49 | 0.99 | 0.92 | 1.04 | 0.74 | | |
| rs61739290 | Non-synonymous | G | 0.07 (1) | 1.31 | 0.017 | 1.03 | 0.36 | 1.12 | 0.099 | | |
| rs3731834 | Non-synonymous | C | 0.19 (1) | 1.26 | 0.0080 | 1.02 | 0.36 | 1.09 | 0.072 | | |

Abbreviations: ARM, Armitage trend test *P*-value; CMH, Cochran-Mantel-Haenszel *P*-value; eQTL, expression quantitative trait locus; GWAS, genome-wide association study; NA, not applicable; OR, odds ratio; SNP, single-nucleotide polymorphism.
Cardiff Full = data from O'Donovan *et al.*;[4] REP1 = three replication samples (Bulgarian, Dublin 1 and Munich) combined; fatSNP2 = Cardiff Full plus REP1 samples combined; Sz Meta = meta analysis of all available schizophrenia samples.
Allele = associated allele in Cardiff Full; $r^2(D')$ = LD metrics pairwise with rs1344706; OR = odds ratio.
[a]One-tailed tests when ORs are in the same direction as Cardiff Full.
[b]Intronic SNP assayed as a proxy for non-synonymous SNP rs4667001.
[c]UK extra Schizophrenia data unavailable.

genotype pairs (HapMap or duplicates) for the 8 independent markers genotyped in our follow-up study, we observed only 1 genotyping discrepancy. For rs17584522, which was genotyped by the Dublin group, the comparison of 72 duplicate samples and 83 HapMap samples revealed no discrepancies.

*Meta-analysis*
Meta-analysis of rs1344706 in all available schizophrenia data sets provided strong evidence for association with schizophrenia ($P = 2.54 \times 10^{-11}$) with an estimated odds ratio (OR) of 1.10 (95% confidence interval 1.07–14) and no evidence for heterogeneity across studies (Cochran's Q statistic, $P = 0.35$, 22 d.f.). A forest plot of the individual studies is given in Figure 2 with further details in Supplementary Table 5. A sensitivity analysis (Supplementary Table 6) revealed that attaining genome-wide significant evidence for association was not dependent on the inclusion of any one sample, although not unexpectedly, the least significant evidence ($P = 1.19 \times 10^{-8}$) for association was observed when the discovery sample was removed, consistent with the expected inflation in the estimated effect size in that sample, although it is notable that a genome-wide significant
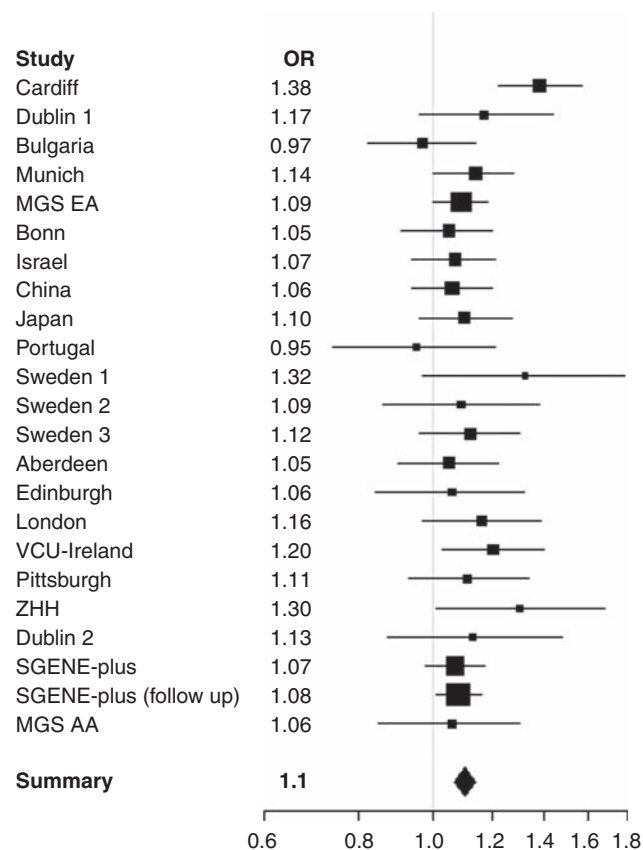


**Figure 2** Forest plot of rs1344706 schizophrenia meta-analysis. OR, is odds ratio for each individual study. Summary is combined OR for all studies. Squares are drawn proportional to the weight of the sample and lines represent 95% confidence intervals.

threshold was surpassed even after removal of this discovery sample.

As before,[4] combined analysis of the schizophrenia and bipolar samples from all studies reported here provided even stronger evidence ($P = 4.1 \times 10^{-13}$) for association (Table 3).

For a subset of the non-fatSNP2 samples we were able to obtain data for additional fatSNP2 markers (or high-quality proxies). These data (plus for comparison the data for rs1344706 in the same subset) are presented in Supplementary Table 7. The significance of association at rs1344706 in the meta-analysis of these particular samples was more than two orders of magnitude better than those of the next best marker (rs17584522).

*Complete genomic sequencing*
We designed 55 PCR amplimeres with 1 kb overlaps (average amplimere length 7908 bases, range 6967–8362). The region spanned ~382.3 kb (chr2: 185163292–185545280, NCBI b36) (Figure 1). We were not successful in amplifying five of the amplimeres, representing a total of ~8% of the target sequence (29 341 bases).

We identified 825 putative SNPs. Of these, 245 were either present in the CEU HapMap or had been genotyped by us in that sample, leaving 580 potential variants whose LD relationships were unknown. We were able to design assays for 34% of these SNPs ($n = 198$). The remaining 66% putative SNPs comprised mainly sites in repetitive elements (49%), with the rest (17%) in unique sequence. We attempted to confirm 19 of those SNPs that were in repetitive elements by more routine sequencing methods (Big-Dye chemistry, Applied Biosystems). Of these, 16 (85%) were not confirmed suggesting that the vast majority of the sites in highly repetitive elements were sequencing artefacts.

In the CEU sample, of the 198 SNP assays in our Illumina panel, 18 did not yield readable genotypes and an additional 8 assays had a call rate of < 0.95. In total, we were able to design assays (or obtain data from HapMap) for a high proportion (~71%) of the likely genuine SNPs (Supplementary Table 8). Of 172 SNPs for which we obtained new data, only 22 (13%) with MAF of $\geqslant 0.01$ were not previously tagged at $r^2 = 1$ and only 16 (9%) at $r^2 = 0.9$. This low yield of untagged SNPs suggests that further sequencing endeavours (or efforts to genotype the SNPs for which we could not design Illumina assays) are unlikely to extract much additional genetic information from the region. Information about the 172 confirmed SNPs is provided in Supplementary Table 9.

On the basis of a combination of HapMap and our in-house polymorphism discovery, the target region contains a total of 651 confirmed SNPs with MAF of $\geqslant 0.01$. Of these, 87% are tagged at $r^2 = 1$ by the markers we had already genotyped, whereas 94% are tagged at $r^2 = 0.9$. Of 13% of markers ($n = 80$) that were not tagged at $r^2 = 1$, we were able to impute 74% ($n = 59$) at the recommended 'good practice'[16] value of

436

**Table 3** Results of genotyping bipolar disorder samples for rs1344706

| | UK bipolar[a] | | UK psychosis[b] | | PSY meta[c] | |
|---|---|---|---|---|---|---|
| | Cases n = 1735 controls n = 2873 | | Cases = 2377 controls n = 2873 | | Cases n = 21 274 controls n = 38 675 | |
| SNP | OR | ARM | OR | CMH | OR | P[d] |
| rs1344706 | 1.17 | $5.26 \times 10^{-4}$ | 1.22 | $9.74 \times 10^{-7}$ | 1.11 | $4.10-10^{-13}$ |

Abbreviations: ARM, Armitage trend test *P*-value; CMH, Cochran–Mantel–Haenszel *P*-value; OR, odds ratio; SNP, single-nucleotide polymorphism.
[a]UK biploar is UK WTCCC bipolar sample (controls are those used for Cardiff Full).
[b]UK psychosis is Cardiff Full and UK bipolar samples combined.
[c]PSY meta is Sz meta samples plus additional data from all available bipolar disorder samples (see text for details).
[d]*P* is inverse variance meta-analysis *P*-value.

INFO (information content metric) $\geqslant 0.8$ using PLINK and 60% (*n* = 48) with an $r^2$ score of $\geqslant 0.9$ using Beagle, in the fatSNP1 sample. None was associated with schizophrenia at a threshold of $P < 1 \times 10^{-4}$ and none was highly correlated with rs1344706 (Supplementary Table 10). Including these successfully imputed SNPs as if they were directly genotyped, we estimate that our coverage of all known SNPs in the region with MAF of $\geqslant 0.01$ is increased to 97% ($r^2 = 1$) and to 99% for $r^2 = 0.9$. Importantly, of the markers (21 out of a total of 651 confirmed markers) that we could not impute, none was even moderately highly correlated with rs1344706 (max $r^2 = 0.13$), making it very unlikely that they could account for the strong association signal observed at that locus.

To estimate the effectiveness of common variant discovery in the pooled sequencing analysis, we determined the number of known HapMaP SNPs with a frequency of $> 0.10$, and of $> 0.20$ that were detected by pooled genomic sequencing in our sample. At these thresholds, we detected, respectively, 85 and 87% of known CEU HapMaP SNPs. These data suggest that the efficiency of mutation discovery was good for alleles with MAF of $> 0.10$. Moreover, as discussed above, given the coverage we already had for the novel SNPs we detected, it is unlikely that detection of additional SNPs would have provided much additional genetic information with respect to common alleles.

*Allelic expression*
To identify *cis*-acting eQTL effects on expression of *ZNF804A*, we assayed mRNA samples from 34 individuals heterozygous for non-synonymous SNP rs4667001 (Figure 3). The G allele at this locus was associated with a 1.13-fold increase (s.d. 0.08) in *ZNF804A* expression (*t*-test $P = 2.59 \times 10^{-7}$, unequal variance). rs4667001 has D' = 1 with rs1344706 in this sample. The higher expression G allele is always in phase with risk allele rs1344706T, with this haplotype representing $\sim 75\%$ of all rs1344706T risk alleles. The underexpressed A allele at rs4667001 resides on a mixture of haplotypes of which the majority ($> 70\%$) carry the non-risk allele rs1344706G. Thus, consistent with our analysis of the GeneVar data, the risk allele at rs1344706 is generally carried by a higher *ZNF804A* expression haplotype. If rs1344706 is *per se* the eQTL responsible for association between that SNP and *ZNF804A* expression, we expect only those subjects who are heterozygous for it to show unequal allelic expression (as homozygotes for rs1344706 would carry two functionally equivalent eQTL alleles). Our observed data were not compatible with this, with there being no difference in the degree of differential expression in rs1344706 homozygotes compared with heterozygotes (*t*-test $P = 0.84$, Figure 3). This suggests that although rs1344706 is associated with expression, it is not responsible for it, or at least that it is not the only *cis*-acting eQTL. Similarly, analyses of the most strongly associated eQTLs from GeneVar revealed that they are also associated with expression, but the comparison of homozygotes and heterozygotes for each putative eQTL suggest none are likely the causal eQTLs (Supplementary Figure 1).

## Discussion

Following our original GWAS,[4] the association between schizophrenia and rs1344706, which lies in intron 2 of *ZNF804A,* has been independently replicated in three studies.[5,10,11] As our study was based upon $< 5\%$ of all common SNPs in the genome, we assumed that rs1344706 was unlikely to be the susceptibility variant. Thus, we undertook a fine-mapping study, the aim of which was to identify the variant that was directly responsible for association. In spite of extensive investigation, we were unable to detect a more strongly associated variant. However, meta-analysis based upon $\sim 60\,000$ subjects provided very strong evidence for association between rs1344706 and schizophrenia ($P = 2.54 \times 10^{-11}$) and also a combined schizophrenia bipolar phenotype ($P = 4.1 \times 10^{-13}$).

*De novo* polymorphism discovery based upon mutation scanning of all *ZNF804A* coding exons did
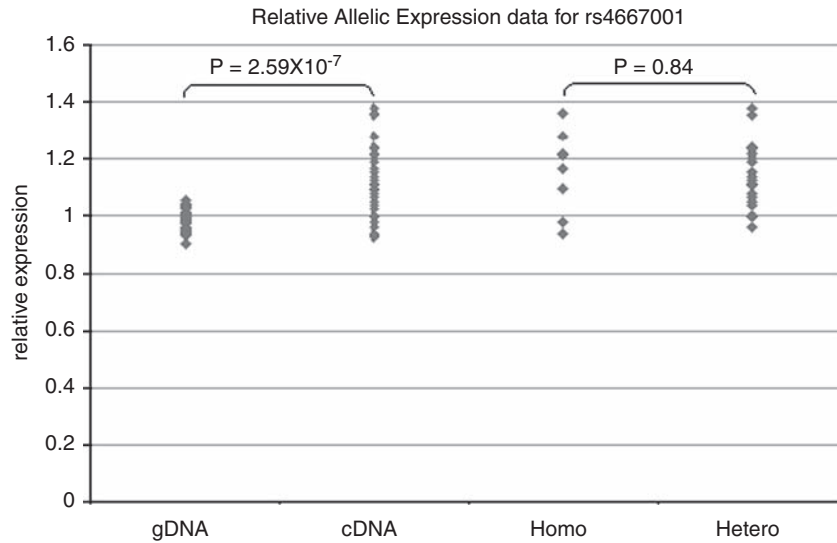
**Figure 3** Relative allelic expression data for rs4667001. *P* is results of *t*-test of mean relative expression. gDNA is relative allelic expression ratio for genomic DNA samples heterozygous for rs4667001. cDNA is relative allelic expression ratio for the corresponding complimentary DNA samples heterozygous for rs4667001. Homo is relative allelic expression ratio for heterozygous rs4667001 cDNA samples stratified for homozygosity at rs1344706. Hetero is relative allelic expression ratio for heterozygous rs4667001 cDNA samples stratified for heterozygosity at rs1344706.

not reveal evidence for the existence of a common non-synonymous variant (MAF of $\geqslant 0.01$) that explains the original signal in our sample (Table 1), in the fatSNP2 meta-analysis (Table 2) or in the larger meta-analyses of those samples for which data were available for these SNPs (Supplementary Table 7). The second approach that we applied was to test SNPs that, in the GeneVar database, were associated with *ZNF804A* expression. In fatSNP phase 1, two of these markers were more significantly associated with schizophrenia than rs1344706 (Table 1). As those two markers were in perfect LD, we followed up only one, rs1583048. As was the case for the non-synonymous variants, the evidence for association in fatSNP2 (Table 2) and in the larger meta-analyses of those samples for which data were available for these SNPs (Supplementary Table 7) was much weaker than for rs1344706. Therefore, it is likely that the signal observed at rs1583048 derives from LD ($D' = 1$) with rs1344706. That the signals are not independent was also supported by the fact that rs1583048 was not required in the regression model.

Detailed tagging analysis of the *ZNF804A* locus based upon HapMap SNPs did not identify any variants that were more strongly associated in the Cardiff Full sample or in fatSNP2. Addressing the possibility that there exists a common variant that might be more strongly associated than rs1344706, but that is not present in the HapMap, we undertook sequencing across most of the genomic region. Although this uncovered a number of additional variants not present in the HapMap, these additional SNPs were well covered by the existing genotyped SNPs. Despite the fact that the vast majority (99%) of

all the known variation across *ZNF804A* could be imputed and/or tagged at least at $r^2 > 0.9$, no additional markers were more strongly associated than rs1344706. Moreover, few markers were even moderately correlated with rs1344706. In Supplementary Table 11, we list all markers with $r^2 > 0.2$ in relation to rs1344706 and their imputed *P*-values. If rs1344706 was only weakly or moderately correlated with a true susceptibility variant, we would expect the association signal at that second SNP to be considerably stronger than that observed at rs1344706. The absence of such a signal suggests that none of these moderately correlated SNPs are likely to be responsible for the signal detected at rs1344706. On this basis, we conclude that rs1344706 is the most likely susceptibility variant.

Several caveats should be mentioned. These are: (1) we identified a number of putative SNPs that we were unable to genotype, (2) $\sim 8\%$ of the target sequence was refractory to sequencing and (3) our genomic sequencing was not 100% sensitive as we identified a high proportion of, but not all, known variation in the region. Thus, we cannot conclude with certainty that the true functional variant did not elude us, but given that it was not captured in the 651 SNPs (*de novo* plus HapMap) for which we had good coverage, and that very little additional genetic information was extracted by the novel SNPs that we had detected through sequencing the genomic region, this does not seem likely.

At the outset, we assumed it unlikely that a GWAS study would identify the true pathogenic variant. However, a recent study[5] supports the existence of very large numbers of common alleles of weak effect.

Power considerations suggest few of these can be expected to achieve high levels of significance even in samples substantially larger than those we used in our discovery GWAS. However, although the power to identify any *one specific* risk allele is low, the power to identify *one of many* alleles is enhanced if there are very many of these to be detected. One factor dictating this will be the degree to which true risk variants are tagged by array SNPs, with weak effects requiring very high LD for both detection and reliable directional replication.[25] It follows that those risk alleles that by chance are included on arrays are most likely to be detected and replicated. Thus, in the context of a highly polygenic disorder with weak genetic effects and a underpowered discovery sample, these true associations that are first detected by GWAS studies are, as we have observed here, likely to either correspond to susceptibility alleles, or be in perfect LD with them.

In terms of functional mechanisms, we observed that the risk allele of rs1344706 is associated with higher *ZNF804A* expression in our analysis of the GeneVar data. That finding is compatible with the analysis of Riley *et al.*[10] who reported that the risk allele at rs1344706 is associated with higher *ZNF804A* expression in human brain. However, although our analysis of allele-specific expression confirms that rs1344706 is generally carried on a higher expression haplotype, it does not seem to be the eQTL responsible for higher expression, suggesting that the relationship between that SNP and expression is not relevant to disease. Alternatively, there may be additional eQTLs (and susceptibility alleles) at the locus that we have failed to uncover.

In the absence of evidence for a non-synonymous SNP that explains the association or for direct effects of rs1344706 on expression, if rs1344706 is the true causal variant, its influence on gene function remains to be elucidated. It remains possible that it exerts effects through expression, but similar to many eQTLs, these are cell specific,[26] tissue-region specific or specific to certain developmental phases. Further expression studies in samples not available to the researchers will be required to test these hypotheses. However, the observations of deletion and partial duplication CNVs at the *ZNF804A* in a schizophrenia and bipolar proband, respectively,[11] similarly suggests that simple upregulation of *ZNF804A* may not be the mechanism relevant to risk to schizophrenia and other major psychiatric disorders.

The establishment of *ZNF804A* as a risk factor for schizophrenia and bipolar disorder is one of several successes arising from recent large-scale genetic studies of major psychiatric disorders. These have implicated common alleles for psychiatric disorders at much higher levels of confidence than previous genetic approaches, among which the most robust current findings in schizophrenia are *ZNF804A*, Neurogranin (*NRGN*), Trancription Factor 4 (*TCF4*) and a locus spanning several megabases of chromosome 6 in the region of the major histocompatibility

complex.[4–7] In bipolar disorder, *calcium channel, voltage-dependent, L type, alpha 1C subunit* (*CACNA1C*) and *ankyrin 3, node of Ranvier* (*ANK3*) have been strongly supported as susceptibility genes.[27] At least two of the specific loci, *ZNF804A* and *CACNA1C*, influence risk for both disorders[28] a finding that supports the hypothesis that schizophrenia and BD are not aetiologically distinct. Similar to *ZNF804A*, in general, the common risk alleles identified by GWAS have small effect sizes (OR < 1.25), although the associated allele at *ANK3* may have a somewhat larger, but still weak effect (OR ~ 1.45).[27] It is, however, clear that many more common risk alleles remain to be identified, a substantial component (at least 30%) of the variance in risk of schizophrenia being attributable to risk alleles of very small effect, and many of these also influence risk of bipolar disorder.[5] In the case of schizophrenia, several rare susceptibility CNVs have also been detected; in contrast to common alleles, these have fairly large effect sizes (OR > 3) on disease risk. However, similar to the common alleles, these CNVs are not specific to individual disorders as defined by widely used classification systems; the same CNVs associated with schizophrenia additionally influence risk of other neurodevelopmental disorders such as autism, epilepsy and mental retardation.[12] Although very little of the genetic risk of either schizophrenia or bipolar disorder is currently explained, there are grounds for optimism that larger studies will reveal more about the origins of these disorders. Moreover, the existing findings already challenge current concepts of disease classification[9] and point to some pathophysiological mechanisms, for example, the involvement of calcium channels in bipolar disorder (*CACNA1C*).

As for the pathophysiological implications of the present finding, ZNF804A is presently a protein of unknown function. The amino acid sequence contains a C2H2-type domain characteristic of the classical zinc-finger (ZnF) family of proteins. These were originally identified as DNA binding molecules with a role in transcription, but proteins with this classic zinc-finger domain are now known to interact with many other types of molecules, including RNA and protein, and in doing so, have many roles in cellular function.[29] Until ZNF804A is functionally characterized, it is not possible to propose specific cellular processes that link the current genetic finding to disease risk.

At the whole-organism level, two recent studies have associated rs1344706 with variation in function. Esslinger *et al.*[30] reported that the schizophrenia risk allele at this locus was associated with reduced connectivity both within DLPFC (dorsolateral prefrontal cortex) and between the right and left DLPFC as indexed by the extent to which activation of these brain structures was temporally correlated during the N-back task (a probe for executive function). In contrast, connectivity was increased between the DLPFC and the hippocampal formation as well as

between a number of other structures. More recently, Walters and colleagues found that the schizophrenia risk allele at rs1344706 was associated with better episodic and working memory in individuals with schizophrenia, but not controls, a finding that they replicated in a German sample. Given that schizophrenia is associated with reduced cognitive function, association with better function seems counterintuitive. However, given the absence of association with better function in controls, rather than being associated with cognitive ability, *ZNF804A* may be associated with a subtype of schizophrenia in which cognition is relatively spared. This hypothesis is in keeping with association between this locus not just with schizophrenia but also with bipolar disorder.[31] To what extent, if at all, the results from the study of Walters and colleagues reflect the observations of altered connectivity in the earlier study[30] is unclear.

Given that the function of the product of this gene is currently unknown, determining this is now a priority in understanding how this genetic association translates into pathophysiology. The identification of its binding partners, be they DNA, RNA or protein, will offer the opportunity to identify a set of further candidate genes, each of which will benefit from downstream genetic analysis based upon a much higher prior probability than typical candidate genes for schizophrenia.

## Conflict of interest

The authors declare no conflict of interest.

## References

1 Cardno AG, Gottesman II II. Twin studies of schizophrenia: from bow-and arrow concordances to star wars Mx and functional genomics (review). *Am J Med Genet* 2000; **97**: 12–17.

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 Williams HJ, Owen MJ, O'Donovan MC. Schizophrenia genetics: new insights from new approaches. *Br Med Bull* 2009; **91**: 61–74.

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

5 The International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.

6 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.

7 Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I *et al*. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 2009; **460**: 753–757.

8 Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* 2008; **32**: 227–234.

9 Craddock N, O'Donovan MC, Owen MJ. Psychosis genetics: modeling the relationship between schizophrenia, bipolar disorder, and mixed (or 'schizoaffective') psychoses. *Schizophr Bull* 2009; **35**: 482–490.

10 Riley B, Thiselton D, Maher BS, Bigdeli T, Wormley B, McMichael GO *et al*. Replication of association between schizophrenia and ZNF804A in the Irish case–control study of schizophrenia sample. *Mol Psychiatry* 2010; **15**: 29–37.

11 Steinberg S, Mors O, Gustafsson O, Werge T, Mortensen PB, Andreassen OA *et al*. Expanding the range of ZNF804A variants conferring risk of psychosis. *Mol Psychiatry* 2010. e-pub ahead of print 5 January 2010; doi:10.1038/mp.2009.149.

12 O'Donovan MC, Craddock NJ, Owen MJ. Genetics of psychosis; insights from views across the genome. *Hum Genet* 2009; **126**: 3–12.

13 Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C *et al*. Population genomics of human gene expression. *Nat Genet* 2007; **39**: 1217–1224.

14 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.

15 WTCCC. Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* 2007; **447**: 661–678.

16 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.

17 Browning BL, Browning SR. A unified approach to genotype imputation and haplotype phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 2009; **84**: 210–223.

18 Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 2008; **18**: 1851–1858.

19 Bray NJ, Preece A, Williams NM, Moskvina V, Buckland P, Owen MJ *et al*. Haplotypes at the dystrobrevin binding protein 1 (DTNBP1) gene locus mediate risk for schizophrenia through reduced DTNBP1 expression. *Hum Mol Genet* 2005; **14**: 1947–1954.

20 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.

21 The International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008; **455**: 237–241.

22 Talkowski ME, Kirov G, Bamne M, Georgieva L, Torres G, Mansour H *et al*. A network of dopaminergic gene variations implicated as risk factors for schizophrenia. *Hum Mol Genet* 2008; **17**: 747–758.

23 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.

24 Sullivan PF, Lin D, Tzeng JY, van den Oord E, Perkins D, Stroup TS *et al*. Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry* 2008; **13**: 570–584.

440

25 Moskvina V, O'Donovan MC. Detailed analysis of the relative power of direct and indirect association studies and the implications for their interpretation. *Hum Hered* 2007; **64**: 63–73.

26 Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 2009; **325**: 1246–1250.

27 Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L *et al.* Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet* 2008; **40**: 1056–1058.

28 Green EK, Grozeva D, Jones I, Jones L, Kirov G, Caesar S *et al.* The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia. *Mol Psychiatry* 2009. e-pub ahead of print 21 July 2009; doi:10.1038/mp.2009.49.

29 Gamsjaeger R, Liew CK, Loughlin FE, Crossley M, Mackay JP. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem Sci* 2007; **32**: 63–70.

30 Esslinger C, Walter H, Kirsch P, Erk S, Schnell K, Arnold C *et al.* Neural mechanisms of a genome-wide supported psychosis variant. *Science* 2009; **324**: 605.

31 Walters JTR, Corvin A, Owen MJ, Williams H, Dragovic M, Quinn EM *et al.* The psychosis susceptibility gene ZNF804A is associated with less impaired cognitive performance in schizophrenia. *Arch Gen Psychiatry* (in press).

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

## Appendix

The following authors are included under:

*Molecular Genetics of Schizophrenia Collaboration*
PV Gejman (Evanston Northwestern Healthcare and Northwestern University, IL, USA), AR Sanders (Evanston Northwestern Healthcare and Northwestern University, IL, USA), J Duan (Evanston Northwestern Healthcare and Northwestern University, IL, USA), DF Levinson (Stanford University, CA, USA), NG Buccola (Louisiana State University Health Sciences Center, LA, USA), BJ Mowry (Queensland Centre for Mental Health Research, and Queensland Institute for Medical Research, Queensland, Australia), R Freedman (University of Colorado Denver, Colorado, USA), F Amin (Atlanta Veterans Affairs Medical Center and Emory University, Atlanta, USA), DW Black (University of Iowa Carver College of Medicine, IA, USA), JM Silverman (Mount Sinai School of Medicine, New York, USA), WJ Byerley (University of California at San Francisco, California, USA), CR Cloninger (Washington University, Missouri, USA).

*SGENE-plus*
H Stefansson (deCODE genetics, Reykjavik, Iceland), S Steinberg (deCODE genetics, Reykjavik, Iceland), E Strengman (Universiteitsweg, Utrecht, The Netherlands), T Hansen (Copenhagen University Hospital, Roskilde, Denmark), HB Rasmussen(Copenhagen University Hospital, Roskilde, Denmark), O Gustafsson (University of Oslo, Oslo, Norway), S Djurovic (University of Oslo, Oslo, Norway), I Giegling (Ludwig-Maximilians-University, Munich, Germany), M Nyegaard (Aarhus University, Arhus C, Denmark), OP Pietiläinen (Institute of Molecular Medicine, Helsinki, Finland and Wellcome Trust Sanger Institute, Cambridge UK), A Tuulio-Henriksson (National Public Health Institute, Helsinki, Finland), E Sigurdsson (National University Hospital, Reykjavik, Iceland), H Petursson (National University Hospital, Reykjavik, Iceland), B Glenthøj (Copenhagen University Hospital, Glostrup, Denmark), G Jürgens (Bispebjerg University Hospital, Copenhagen, Denmark), I Melle (University of Oslo, Oslo, Norway), M Rietschel (University of Heidelberg, Mannheim, Germany), AD Børglum (Aarhus University Hospital, Risskov, Denmark), A Ingason (deCODE genetics, Reykjavik, Iceland), U Thorsteinsdottir (deCODE genetics, Reykjavik, Iceland), A Kong (deCODE genetics, Reykjavik, Iceland), P Muglia (GlaxoSmithKline R&D, Verona, Italy), LA Kiemeney (Radboud University, Nijmegen, The Netherlands), B Franke (Radboud University, Nijmegen, The Netherlands), M Ruggeri (University of Verona, Verona, Italy), S Tosato (University of Verona, Verona, Italy), TE Thorgeirsson (deCODE genetics, Reykjavik, Iceland), O Mors (Aarhus University Hospital, Risskov, Denmark), PB Mortensen (Aarhus University, Aarhus, Denmark), I Bitter (Semmelweis University, Budapest, Hungary), EG Jönsson (Karolinska Institutet and Hospital, Stockholm, Sweden), S Cichon (University of Bonn, Bonn, Germany), MM Nöthen (University of Bonn, Bonn, Germany), OA Andreassen (University of Oslo, Oslo, Norway), V Golimbet (Russian Academy of Medical Sciences, Moscow, Russia), T Li (Institute of Psychiatry, London, UK), T Werge (Copenhagen University Hospital, Roskilde, Denmark), RA Ophoff (UCLA, Los Angeles, USA and University Medical Center Utrecht, Utrecht, The Netherlands), D St Clair (University of Aberdeen, Aberdeen, UK), DA Collier (Institute of Psychiatry, London, UK), L Peltonen (Institute of Molecular Medicine, Helsinki, Finland and Wellcome Trust Sanger Institute, Cambridge, UK), D Rujescu (Ludwig-Maximilians-University, Munich, Germany) and K Stefansson (deCODE genetics, Reykjavik, Iceland).

*Genetic Risk and Outcome in Psychosis (GROUP)*
RS Kahn (Rudolf Magnus Institute of Neuroscience, Utrecht, The Netherlands), DH Linszen (Academic Medical Centre University of Amsterdam, Amsterdam, The Netherlands), J van Os (Maastricht University Medical Centre, Maastricht, The Netherlands), D Wiersma (University of Groningen, Groningen, The Netherlands), R Bruggeman (University of Groningen, Groningen, The Netherlands), W Cahn (Rudolf Magnus Institute of Neuroscience, Utrecht, The Netherlands), L de Haan (Academic Medical Centre

University of Amsterdam, Amsterdam, The Netherlands), L Krabbendam (Maastricht University Medical Centre, Maastricht, The Netherlands) and Inez Myin-Germeys (Maastricht University Medical Centre, Maastricht, The Netherlands).

*International Schizophrenia Consortium (ISC)*
Michael C O'Donovan (Cardiff University, Cardiff, UK), George K Kirov (Cardiff University, Cardiff, UK), Nick J Craddock (Cardiff University, Cardiff, UK), Peter A Holmans (Cardiff University, Cardiff, UK), Nigel M Williams (Cardiff University, Cardiff, UK), Lyudmila Georgieva (Cardiff University, Cardiff, UK), Ivan Nikolov (Cardiff University, Cardiff, UK), N Norton (Cardiff University, Cardiff, UK), H Williams (Cardiff University, Cardiff, UK), Draga Toncheva (University Hospital Maichin Dom, Sofia, Bulgaria), Vihra Milanova (Alexander University Hospital, Sofia, Bulgaria), Michael J Owen (Cardiff University, Cardiff, UK), Christina M Hultman (Karolinska Institutet, Stockholm, Sweden and Uppsala University, Uppsala, Sweden), Paul Lichtenstein (Karolinska Institutet, Stockholm, Sweden), Patrick Sullivan (University of North Carolina at Chapel Hill, NC, USA), Derek W Morris (Trinity College Dublin, Dublin, Ireland), Colm T O'Dushlaine (Trinity College Dublin, Dublin, Ireland), Elaine Kenny (Trinity College Dublin, Dublin, Ireland), Emma M Quinn (Trinity College Dublin, Dublin, Ireland), Michael Gill (Trinity College Dublin, Dublin, Ireland), Aiden Corvin (Trinity College Dublin, Dublin, Ireland), Andrew McQuillin (University College London, London, UK), Khalid Choudhury (University College London, London, UK), Susmita Datta (University College London, London, UK), Jonathan Pimm (University College London, London, UK), Srinivasa Thirumalai (West Berkshire NHS Trust, Reading, UK), Vinay Puri (University College London, London, UK), Robert Krasucki (University College London, London, UK), Jacob Lawrence (University College London, London, UK), Digby Quested (University of Oxford, Oxford, UK), Nicholas Bass (University College London, London, UK), Hugh Gurling (University College London, London, UK), Caroline Crombie (University of Aberdeen, Aberdeen, UK), Gillian Fraser (University of Aberdeen, Aberdeen, UK), Soh Leh Kuan (University of Aberdeen, Aberdeen, UK), Nicholas Walker (Ravenscraig Hospital, Greenock, UK), David St Clair (University of Aberdeen, Aberdeen, UK), Douglas HR Blackwood (University of Edinburgh, Edinburgh, UK), Walter J Muir (University of Edinburgh, Edinburgh, UK), Kevin A McGhee (University of Edinburgh, Edinburgh, UK), Ben Pickard (University of Edinburgh, Edinburgh, UK), Pat Malloy (University of Edinburgh, Edinburgh, UK), Alan W Maclean (University of Edinburgh, Edinburgh, UK), Margaret Van Beck (University of Edinburgh, Edinburgh, UK), Naomi R Wray (Queensland Institute of Medical Research, Queensland, Australia), Stuart Macgregor (Queensland Institute of Medical Research, Queensland, Australia), Peter M. Visscher (Queensland Institute of Medical Research, Queensland, Australia), Michele T Pato (University of Southern California, California, USA), Helena Medeiros (University of Southern California, California, USA), Frank Middleton (Upstate Medical University, New York, USA), Celia Carvalho (University of Southern California, California, USA), Christopher Morley (Upstate Medical University, New York, USA), Ayman Fanous (University of Southern California, California, USA and Washington VA Medical Center, Washington, USA and Georgetown University School of Medicine, Washington DC, USA and Virginia Commonwealth University, Virginia, USA), David Conti (University of Southern California, California, USA), James A. Knowles (University of Southern California, California, USA), Carlos Paz Ferreira (Department of Psychiatry, Azores, Portugal), Antonio Macedo (University of Coimbra, Coimbra, Portugal), M Helena Azevedo (University of Coimbra, Coimbra, Portugal), Carlos N Pato (University of Southern California, California, USA); Massachusetts General Hospital Jennifer L Stone (Massachusetts General Hospital, Massachusetts, USA and The Broad Institute of Harvard and MIT, Massachusetts, USA), Douglas M Ruderfer (Massachusetts General Hospital, Massachusetts, USA and The Broad Institute of Harvard and MIT, Massachusetts, USA), Andrew N Kirby (Massachusetts General Hospital, Massachusetts, USA and The Broad Institute of Harvard and MIT, Massachusetts, USA), Manuel AR Ferreira (Massachusetts General Hospital, Massachusetts, USA and The Broad Institute of Harvard and MIT, Massachusetts, USA), Mark J Daly (Massachusetts General Hospital, Massachusetts, USA and The Broad Institute of Harvard and MIT, Massachusetts, USA), Shaun M Purcell (Massachusetts General Hospital, Massachusetts, USA and The Broad Institute of Harvard and MIT, Massachusetts, USA), Jennifer L Stone (Massachusetts General Hospital, Massachusetts, USA and The Broad Institute of Harvard and MIT, Massachusetts, USA), Kimberly Chambert (The Broad Institute of Harvard and MIT, Massachusetts, USA), Douglas M Ruderfer (Massachusetts General Hospital, Massachusetts, USA and The Broad Institute of Harvard and MIT, Massachusetts, USA), Finny Kuruvilla (The Broad Institute of Harvard and MIT, Massachusetts, USA), Stacey B Gabriel (The Broad Institute of Harvard and MIT, Massachusetts, USA), Kristin Ardlie (The Broad Institute of Harvard and MIT, Massachusetts, USA), Jennifer L Moran (The Broad Institute of Harvard and MIT, Massachusetts, USA), Edward M Scolnick (The Broad Institute of Harvard and MIT, Massachusetts, USA), Pamela Sklar (Massachusetts General Hospital, Massachusetts, USA and The Broad Institute of Harvard and MIT, Massachusetts, USA).