The psychosis susceptibility gene ZNF804A is associated with less impaired cognitive performance in schizophrenia.

James T.R. Walters MRCPsych, B.M.[1], Aiden Corvin MRCPsych., Ph.D.[2,6], Hywel Williams Ph.D.[1], Milan Dragovic Ph.D.[8], Emma M Quinn M.Sc.[2], Róisín Judge B.A.[2], Daniel J. Smith M.D. [1], Nadine Norton Ph.D.[1], Ina Giegling Ph.D.[3], Annette M. Hartmann Ph.D.[3], Hans-Jürgen Möller M.D.[4], Pierandrea Muglia M.D.[5], Valentina Moskvina Ph.D.[1], Therese O'Donoghue M.Sc.[2,6], Bharti Morar Ph.D.[7], Matthew Cooper M.Sc.[7,9], David Chandler Ph.D.[7], Assen Jablensky M.D., Ph.D.[8], Michael Gill MRCPsych., M.D. [2,6], Michael J. Owen FRCPsych, Ph.D.[1], Luba Kaladjieva M.D., Ph.D.[7], Derek W. Morris Ph.D.[2], Michael C. O'Donovan FRCPsych, Ph.D.[1], Dan Rujescu M.D.[3], Gary Donohoe D.Clin.Psych, Ph.D.[2,6]

Affiliations

1 MRC Centre for Neuropsychiatric Genetics and Genomics, Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Heath Park, Cardiff, UK.
2 Neuropsychiatric Genetics Research Group, Dept of Psychiatry & Institute of Molecular Medicine, Trinity College Dublin, Ireland.
3 Molecular and Clinical Neurobiology, Ludwig-Maximilians-University, Munich, Germany.
4 Department of Psychiatry, Ludwig-Maximilians-University, Munich, Germany.
5 Medical Genetics, GlaxoSmithKline R&D, Verona, Italy.
6 Trinity Institute of Neuroscience, Trinity College, Dublin.
7 Centre for Medical Research/Western Australian Institute for Medical Research, The University of Western Australia, Perth, Australia.
8 Centre for Clinical Research in Neuropsychiatry, The University of Western Australia, Perth, Australia.
9 Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, Perth, Australia.

Address for correspondence

Dr. Gary Donohoe.
Neuropsychiatric Genetics Research Group,
Dept of Psychiatry,
Trinity Health Sciences Building,
St. James's Hospital,
Dublin 8.
Telephone
+353 1 896 2467
Fax
+353 1 896 3405
E-mail
Gary.Donohoe@tcd.ie

Word count: 4699
Number of Tables: 3 Number of Figures: 1

Abstract

Background: The Zinc Finger Protein 804A gene (*ZNF804A*) has been implicated in schizophrenia (SZ) susceptibility by several genome-wide association studies (GWAS). *ZNF804A* is brain-expressed, but of unknown function. **Objective:** To investigate whether the identified risk allele at the disease associated single nucleotide polymorphism (SNP) rs1344706 is associated with variation in neuropsychological performance in patients and controls. **Design:** A comparison of both cases and controls grouped according to *ZNF804A* genotype (AA v AC v CC) on selected measures of cognition in two independent samples. **Setting:** Unrelated patients from general adult psychiatric inpatient and outpatient services and unrelated healthy volunteers from the general population were ascertained. **Participants:** Patients with DSM-IV diagnosed schizophrenia and healthy controls from independent samples of Irish (n=297 cases and n=165 controls) and German (n=251 cases and n=1472 controls) nationality. **Method:** A two-stage study. We tested for association between *ZNF804A* rs1344706 and cognitive functions known to be impaired in schizophrenia (IQ, episodic memory, working memory, and attentional control) in an Irish discovery sample. We then tested significant results in a German replication sample. **Result:** An interaction effect between *ZNF804A* genotype and diagnosis was observed for measures of episodic and working memory in the Irish patient sample but not controls. These findings replicated in the same direction in the German sample. Furthermore, in both samples the association between *ZNF804A* and schizophrenia strengthened when patients with lower general cognitive function were excluded. **Discussion:** In a disorder characterized by heterogeneity, a risk variant at ZNF804A appears to delineate a patient subgroup characterized by relatively spared cognitive ability. Further work is required to establish if this represents a discrete molecular pathogenesis that differs from other patient groups and whether this also has consequences for nosology, illness course or treatment.

Introduction

Schizophrenia (SZ) has a lifetime risk of approximately 1% and is a major cause of global disability[1]. Despite its substantial heritability (h^2~80%)[2], identifying the genetic variations responsible for schizophrenia has proved challenging, as with other non-psychiatric complex disorders[3, 4]. A recent genome wide association analysis (GWAS) identified the single nucleotide polymorphism rs1344706 located at gene *ZNF804A* (OMIM: 612282) as achieving genome wide significance for psychosis (9.96×10^{-9} (OR 1.12))[5]. Despite being relatively under-powered to replicate such a modest effect, two of three recently reported large SZ GWAS studies supported association with the same risk allele [6-8].

SNP rs1344706 is located in an intron of *ZNF804A* which maps to chromosome 2q32.1. The human *ZNF804A* gene consists of four exons which transcribe a protein of 1210 amino acids with a predicted molecular weight of 137kDa. The encoded protein is uncharacterized, but analysis of the protein sequence shows a zinc finger domain at the N-terminal end, suggesting that it may bind DNA and have a role in regulating gene expression. The Allen Brain Atlas indicates the mouse orthologue Zfp804A is widely expressed in the brain [9]. Lim et al [10] demonstrated using a yeast 2-hybrid system that *ZNF804A* bound Ataxin-1, which is encoded by *ATXN1* (OMIM:601556). Mutations in this gene cause spinocerebellar-ataxia-1 and it has been implicated, albeit inconclusively, in SZ by several small studies (SZGene: http://www.schizophreniaforum.org/res/sczgene/geneoverview.asp?geneid=354).

In their SZ GWAS study, Stefansson and colleagues [6] identified association with genes involved in brain development and cognition. Neurocognitive deficits are core features of schizophrenia and may better represent underlying pathophysiology than clinical diagnostic categories [11]. Whether or not the measurement of these deficits increases power to detect association with psychiatric risk genes[12], they have the unique advantage of enabling *in vivo* functional investigation of candidate genes at the level of brain behavior in large samples of patients and healthy controls. The utility of such an approach has been demonstrated by

findings with existing candidate genes for schizophrenia (including DISC1, NRG1, DTNBP1, DAOA(G72), RGS4; discussed in [13]). No neuropsychological investigation of ZNF804A has taken place to date, although evidence that ZNF804A is likely to influence brain function derives from a recent study by Esslinger et al. [14]. This study, investigating the same risk allele at rs1344706, identified association with altered connectivity in healthy controls both within and between regions including the dorsolateral prefrontal cortex and hippocampus. No behavioral sequelae of this altered connectivity were associated with this finding, although the sample size (n=115) may have been under powered to detect such differences.

In this study we investigate the influence on cognition of rs1344706, the variant showing strongest evidence of association in the GWAS by O'Donovan et al.[5]. Neuropsychological assessment was designed to measure cognitive functions known to be impaired in SZ – IQ, working memory, episodic memory and attention. The neuropsychological tests examined in the different samples were selected a priori to be as equivalent as possible. To enable a clear investigation of the effects of *ZNF804A* variation on cognition we focussed on a single statistically strong risk variant, used a large discovery sample, and employed a study design which allowed replication of findings in an independent dataset.

Methods

Sample characteristics

Irish patient and control samples (Discovery samples): This sample consisted of 297 cases and 165 controls. 71 of the case participants were genotyped as part of the prior GWAS study [5], the remaining 226 case participants and all control samples were independent of that study. Cases consisted of clinically stable patients with a DSM-IV diagnosis of SZ recruited from five sites across Ireland. Inclusion criteria required that participants were aged 18 to 65 years, had no history of co-morbid psychiatric disorder, substance abuse in the preceding six months, or prior head injury with loss of consciousness or a history of seizures. Diagnosis was confirmed by trained psychiatrists using the Structured Clinical Interview for

DSM-IV Axis 1 Diagnoses (SCID; [15]). Additional diagnostic details and clinical sample characteristics were ascertained at time of interview including symptom severity (SAPS/SANS;[16]) and medication dosage.

The healthy control sample was recruited on the basis of responses to local media advertisements. Control participants were only included if they were aged between 18 and 65 and satisfied, based on clinical interview, the criteria of having no history of major mental health problems, intellectual disability or acquired brain injury, and no history of substance misuse in the preceding six months based on self report. Control participants were also excluded from the study if they reported having a 1st degree relative with a history of psychosis. All patients and control assessments were conducted in accordance with the relevant ethics committees' approval from each participating site. All patients and controls were of Irish ancestry (i.e. four grandparents born in Ireland) and all provided written informed consent.

German patient and control samples (Replication sample): The German sample consisted of 251 clinically stable patients with a DSM-IV diagnosis of SZ and up to 1472 healthy controls, all of whom were genotyped as part of the previous study[5]. Patients were ascertained from mental health services in the Munich area and all participants provided written informed consent. Inclusion criteria were a diagnosis of SZ (over 6 month symptom duration) and age 18-65. Exclusion criteria included a history of head injury or neurological diseases. Detailed medical and psychiatric histories were collected, including a clinical interview using the Structured Clinical Interview for DSM-IV (SCID[15]), to evaluate lifetime Axis I and II diagnoses. Four physicians and one psychologist rated the SCID interviews and all measurements were double-rated by a senior researcher. Participants were also rated for symptoms using the Positive and Negative Symptom Scale[17]. In cases, 68% were of strict German descent (all 4 grandparents born in Germany) and the other 32% were German Caucasian. Of the 251 schizophrenia participants 241 completed the full WAIS-R assessment and 235 completed a comprehensive neuropsychological battery.

Healthy control participants of German descent (all 4 grandparents German) were randomly selected from the general population of Munich, Germany, and contacted by mail. Control patients for this study were only included if aged between 18 and 65. To exclude subjects with central neurological diseases and psychotic disorders or subjects who had first-degree relatives with psychotic disorders, several screenings were conducted before the volunteers were enrolled in the study. First, subjects who responded were initially screened by phone for the absence of neuropsychiatric disorders. Second, detailed medical and psychiatric histories were assessed for both participants and their first-degree relatives by using a semi-structured interview. Third, if no exclusion criteria were fulfilled, they were invited to a comprehensive interview including the SCID[15] to validate the absence of any lifetime psychotic disorder. Additionally, the Family History Assessment Module[18] was conducted to exclude psychotic disorders among their first-degree relatives. A neurological examination was also conducted to exclude subjects with current CNS impairment. In the case of volunteers older than 60 years, the Mini Mental Status Test[19] was performed to exclude subjects with possible cognitive impairment. Of the 1472 control participants, 1470 completed the full WAIS-R assessment and 367 completed tests from a further extensive neuropsychological battery.

Cognitive assessment

This study was designed so that identical or near identical tests of the cognitive domains of general cognition (IQ), episodic memory, working memory, and attention were used for both the Irish discovery samples and the German samples. The number of individual tests within each domain of cognition was limited to minimize multiple testing effects. The Irish discovery sample was used to test for genotypic associations with these tests. Where significant ($p \leq 0.05$) associations were detected, these phenotypes were taken forward to the German sample for replication.

General cognitive functioning (IQ) was measured in the Irish sample using selected subtests (Vocabulary, Similarities, Block Design and Matrix Reasoning) from the Wechsler Adult Intelligence Scale, 3rd edition (WAIS-III; [20]), yielding a full scale, verbal and performance IQ.

For the German sample, IQ was indexed by the German version of the Wechsler Adult Intelligence Scale, revised edition [21] using all available 11 verbal/performance subtests (Vocabulary, comprehension, information, digit span, arithmetic, similarities, block design, picture completion, picture arrangement, object assembly, digit symbol 231 coding). Episodic memory was assessed in the Irish samples using the Logical Memory immediate and delayed from the Wechsler Memory Scale, 3rd edition (WMS-III; [22]) and in the German sample using the Logical memory immediate and delayed from the German version of the Wechsler Memory Scale- Revised [23, 24].

Verbal and spatial working memory were assessed in the Irish samples using the Wechsler Letter Number Sequencing task (WMS-III; [22]) and the spatial working memory task from the Cambridge Automated Neuropsychological Test Battery (CANTAB SWM; [25]). In the German samples, working memory was measured using the Digit Span from the WAIS-R [21] and Spatial Span score from the WMS-R [23, 24]. Attentional control was assessed in the Irish samples using the Continuous performance task, identical pair's version (CPT-IP; [26]) and in the German sample using the Continuous performance task 3-7 version[27]. We have described the memory and attention tasks used here in detail elsewhere (Donohoe et al, in press, [28]).

Genotyping

The SNP rs1344706 was genotyped in the German sample and a proportion of the Irish sample using the Sequenom iPLEX Gold system (further details are provided in [5]). The call rate for the iPLEX genotyping was >95% in the Irish sample and >99% in the German sample. Both case/control samples were in Hardy-Weinberg Equilibrium (HWE; p>0.05). The remainder of the Irish sample (n=529) was genotyped using a Taqman® SNP Genotyping Assay on a 7900HT Sequence Detection System (Applied Biosystems). The call rate for the Taqman genotyping was >95% and both case and control samples were in HWE (p>0.05). Along with the Irish samples, a number of HapMap CEU DNA samples (n=90: www.hapmap.org) and duplicates (n=60) were genotyped for rs1344706 for quality control purposes. All genotypes were found to be concordant with either the available online HapMap data or each other for this SNP.

Statistical analyses

To inform appropriate adjustments in the primary cognitive analyses, association between *ZNF804A* rs1344706 and demographic variables was investigated using one-way ANOVAs. In the case of symptom severity, a principal components analysis (PCA) was undertaken separately in each of the three samples based on SAPS/SANS scores in the Irish sample and PANSS scores in the German sample (both analyses previously described in Donohoe et al., in press); differences associated with genotype were analyzed using MANCOVA in which age and gender were included as covariates.

Association between *ZNF804A* rs1344706 and the phenotypes of general cognitive function, episodic memory, working memory, and attentional control was tested using a general factorial design in SPSS 14 [29]. In the original GWAS study there was no difference in genotypic versus allelic model; as there was no evidence on which to test a specific dominant or recessive model, and as sample sizes allowed, our analysis was based on a comparison of all three genotype groups. *ZNF804A* genotype (AA versus AC versus CC) and diagnosis (cases versus controls) were entered as fixed effects. In a series of ANCOVAs, scores for each neuropsychological subtest were entered as the dependent variables, with age and gender included as covariates as appropriate. Significant interaction effects were further explored by examining simple effects in cases and controls. Those tests showing significant results in the Irish sample were then taken forward to the same analyses in the German samples.

Results

ZNF804 & Demographics and Clinical variables:

Demographic and clinical characteristics by rs1344706 genotype for the 2 samples appear in **Table 1**. In the Irish sample no difference was observed in age, years in education, or gender between the genotype groups for either the full sample, or when patients and controls were considered separately. In terms of clinical symptom severity, no difference was observed between genotype groups for PCA derived negative or disorganized symptoms. For positive

symptom severity a trend level difference between groups was observed (F=2.31; p=.06); mean values for the three groups suggested that the homozygous carriers of the A risk allele presented somewhat higher positive scores than the other genotype groups. This trend was observed in the absence of any association between genotype and medication dose.

For the German sample no significant differences were apparent in age, gender distribution and education by genotype in cases or controls (**Table 1**). In terms of clinical severity, a significant difference was observed for the PCA derived depressive factor (F= 4.61, p=0.01). Pairwise comparisons revealed that CC genotype scored higher on this depressive factor than the AC (t=2.19, p=0.03) and AA (t=3.06, p=0.002) genotype groups. Genotype was not associated with any other symptom factor score; neither were any differences in medication dosage by genotype observed.

Insert Table 1 here

Cognitive analysis of ZNF804

Mean scores for each of the 4 cognitive domains of IQ, working memory, episodic memory, and attention by ZNF804 genotype group for cases and controls in the Irish discovery sample are presented in **Table 2.** As expected, patients performed significantly below controls on all cognitive tests administered in both the Irish and German samples (all p-values<0.0001).

Irish discovery sample For the Irish samples, *ZNF804* genotype was not associated with differences in IQ, either as a main effect or as an interaction effect with case/control status. By contrast, a significant interaction between *ZNF804A* genotype and case/control status revealed association with variance on both working memory tasks in patients but not in controls (verbal working memory: F=4.2; p=0.02; spatial working memory: F=5.04; p=-.007; see Table 2). Effect size estimates (partial η^2) indicated that in cases genotype explains 2.8% of verbal working memory and 4.4% of spatial working memory. Tukey *post hoc* analyses within the patient group revealed that the homozygous risk AA genotype performed significantly *better* than the homozygous CC genotype group (Verbal working memory:

H=1.53; p=.046; Spatial working memory: H=11.38; p=.045). Finally, although a small trend for a main effect of genotype on spatial working memory was apparent, this appeared to be primarily driven by cases; separate analysis of controls revealed no such trend.

A significant association between *ZNF804A* genotype and verbal episodic memory was also observed in patients and not in controls for both immediate and, to a lesser extent, delayed verbal memory. Effect size estimates indicated that in cases genotype explains 2.7% of immediate logical memory scores and 1.1% of delayed verbal memory scores (**Table 2**). For immediate verbal episodic memory, Tukey *post hoc* analysis revealed that this difference was driven by both homozygous risk carriers (H=1.86; p=0.012) and heterozygous risk carriers (H=1.83; p=0.015) again performing significantly *better* than homozygous non-risk carriers. For delayed verbal episodic memory the affect of genotype was smaller and post hoc analysis did not reveal specific between group differences.

Finally, scores for attentional control (the CPT task) was available for patients and not for controls. Based on ANCOVA again with age and gender used as covariates, we found no effect of *ZNF804A* genotype on CPT either in terms of the two letter or three letter conditions (**Table 2**).

Insert Table 2 here

German replication sample

The significant results from the Irish sample were taken forward to the replication samples and are presented in **Table 3**. The results from the Irish sample replicated in the German sample cases for all the cognitive tests – verbal and spatial working memory and episodic memory. Furthermore the allelic direction of effect was the same in that those with AA genotype performed better than AC and CC genotypes on all significant tests. For both spatial and episodic memory Tukey post hoc analysis confirmed that this was mainly driven by a significant difference between AA and CC genotypes (Spatial working memory: H=2.39, p=0.047; immediate episodic memory H=2.81, p=0.015, delayed episodic memory H=3.12,

p=0.009). In the case of verbal working memory inter-genotype group differences were observed between the homozygous risk carriers and the heterozygous risk carriers only (H=2.67, p=0.022). Effect size estimates, again based on partial η^2, indicated that in cases genotype explains 3.3% of variance in spatial working memory, 3.1% of verbal working memory and 3% of both immediate and delayed episodic memory.

Analysis of whether ZNF804A's association with SZ is moderated by cognition

Given the counter-intuitive evidence that carriers of the *ZNF804A* risk genotype presented less impaired cognitive performance than non carriers, we tested the post hoc hypothesis that *ZNF804A* was delineating a subgroup of patients characterized by relatively intact cognitive performance. To do so we used chi square statistics to calculate the association between *ZNF804A* and SZ in sub-samples with higher cognitive ability (indexed by IQ as a more general index of cognitive ability than memory function). Higher cognitive ability was based on IQ scores of equal to or greater than 70, equal to or greater than 80, equal to or greater than 90, equal to or greater than 100, equal to or greater than 110, and equal to or greater than 120 respectively. In both Irish and German samples, as the phenotype was narrowed to individuals with IQ in the average range, the association between ZNF804A and SZ became more statistically significant (see **Tables 4&5** and **Figure 1**). For example, in the Irish samples, as the minimum IQ for inclusion approaches 90 (the range for normal IQ) the odds ratio of the allelic association between ZNF804A and SZ increases from 1.37 (when all samples are included) to 2.29 (CI 1.41-3.72). In the German sample the same trend of increasing association between ZNF804A and SZ also emerges, becoming statistically significant for patients with high average IQ (IQ>110; odds ratio 1.63; CI 1.12-2.38).

Discussion

With regards to its credibility as a risk variant, the SNP rs1344706 at *ZNF804A* is the first variant to have achieved genome wide level significance for psychosis, with the association replicating in multiple independent samples. Despite the modest effect size, two independent GWAS studies have provided support for association with the same risk allele[7,8]. Although this finding may have a very small effect on disease risk, it is potentially important to our

understanding of the genetic mechanisms and pathways that contribute to susceptibility. Little is known about the function of *ZNF804A* and this study sought to elucidate the phenotypic effects of the identified risk allele on indices of neuropsychological function. Our study's design sought to overcome weaknesses of earlier gene-cognition studies (e.g. small sample size and high multiple testing burden) by investigating a single GWAS significant variant and then seeking to replicate significant findings in a large independent dataset using comparable cognitive tests. In our study, carrying the risk allele at SNP rs1344706 was associated with variation in cognitive performance in patients. Specifically, SZ carriers of the AA risk genotype performed relatively *better* on measures of episodic memory and working memory based on two independent Irish and German samples that had both contributed to the original GWAS finding.

While the association with cognition observed may appear counter-intuitive, it is important to note that the risk allele at *ZNF804A* is not so much associated with *better* cognitive performance in the present study as with *less impaired* cognitive performance. We interpret these data to mean that *ZNF804A* is delineating an illness susceptibility pathway that is independent of a deleterious effect on cognition, and hence is characterized by relatively spared cognitive ability compared to other patient subgroups whose pathway into illness is being mediated by a greater burden of more cognitively deleterious gene variants[30]. Support for this view derives from the following: first, the association is only seen for cases; if *ZNF804A* was having a positive impact on cognitive performance this should also be apparent in controls, but this is not the case; second, in the association analysis between *ZNF804A* and SZ, when cases with lower cognitive ability are excluded, the association signal strengthens, indicating that the relationship between *ZNF804A* and psychosis is particularly apparent in those with relatively intact cognitive function. This finding that *ZNF804A* may encode for a cognitively spared psychosis subtype complements earlier evidence in the original GWAS study that inclusion of bipolar patients – a patient group typically associated with less severe cognitive impairments[31] – also led to a strengthening of its association with psychosis[5]. The fact that the association is found in these two independently collected

samples of individuals with schizophrenia also counteracts the argument that ZNF804A is in fact a gene for bipolar disorder.

By elucidating *ZNF804A*'s role in delineating a schizophrenia subtype characterized by relatively less cognitive impairment, our study has direct relevance to one of the main criticisms of the endophenotype approach: the unproven 'assumption' that impaired cognition lies on the pathway from gene to disease phenotype[32]. In this analysis we have demonstrated the value of cognition independent of any assumption that impaired cognition lies on the pathway between *ZNF804A* as a risk gene and disease phenotype, in that this cognitive approach has been helpful in characterizing *what kind* of schizophrenia is associated with this variant. This finding is not unique - recent studies of both *PPP1R1B*, encoding DARPP-32, and *CHI3L1* found that the schizophrenia-associated risk alleles at both gene loci were associated with relatively spared performance[30,33]. Given the heterogeneity of the schizophrenia syndrome and the fact it is possible to be schizophrenic and cognitively intact[34, 35] it is perhaps to be expected that not all identified susceptibility genes will have detrimental effects on broad cognitive abilities.

In contravening the typical expectations for intermediate phenotypes (of not being associated with poorer function), an obvious concern is that the observed associations might spuriously result from one or more confounding demographic, clinical, or cognitive variables in the patient group. We were able to examine this possibility using a wide variety of clinical and general cognitive indicators, including age, medication, gender, education, and clinical symptom severity. No differences between genotype groups on any of these variables were observed. One factor which might have confounded our results was the above average IQ of the Irish healthy controls. However, IQ in the German healthy control sample was in the average range and *post hoc* analysis of this variable failed to reveal significant differences associated with genotype. Furthermore, in the one other study based on exactly the same samples reported here, the effect of genotype on cognitive ability was apparent in both patient and control samples in both the Irish and German datasets (Donohoe et al., in press). The

impact of *ZNF804A* variation being apparent only in patients in both Irish and German datasets is therefore unlikely to be explainable purely in term of this confounder.

As the primary association between *ZNF804A* and cognition was with better memory function in patients we speculated about whether this association generalized to other memory tests. We therefore genotyped ZNF rs1344706 in an Australian sample for whom data were available on a verbal list learning task – the Rey Auditory verbal learning test (SZ patients n= 385; controls n=211; full sample details provided in Supplemental Material (SM1)). Data on working memory was only available in 40 cases and so association with this function could not be tested. Only a trend level association was observed for immediate episodic memory ($F_{3,593}$=2.4; p=.09; see SM table 2), indicating that the association between *ZNF804A* and the story recall task used in the Irish and German samples (WMS logical memory task[22, 24]) did not generalize in the same way to performance on a verbal list learning task. Understanding such inter-test differences is important for interpreting cognitive genetics studies. As well as behavioral and psychometric differences, differences in the heritability of these measures are unknown, and any such differences are likely to influence reproducibility of results in non-identical tests.

An important context for interpreting these data is provided by the recent imaging study of the same genetic variant reported by Esslinger et al.[14]. In a sample of healthy controls, they found that the *ZNF804A* risk genotype was associated with altered connectivity in dorsolateral prefrontal cortex (DLPFC), the hippocampus, and the amygdala. Altered connectivity within and between these brain regions has been associated with SZ; its association with *ZNF804A* provided the first evidence of the gene's functional involvement in brain activity. The fact that the two aspects of cognition implicated in the present study – episodic and working memory – are the aspects of cognition sub-served by the brain regions identified by Esslinger et al.[14] – the DLPFC and hippocampus – again implicate *ZNF804A* in biological processes relevant to these regions. However, there are two caveats to this interpretation. First, Esslinger et al.[14] observed an association between ZNF804A and altered brain connectivity in healthy controls, but no association between the risk variant and neuropsychological performance in this

population. In our study we similarly found no association between *ZNF804A* variation and neuropsychological performance in healthy controls, but did find an association in patients. Second, Esslinger et al[14] demonstrated enhanced functional connectivity between the DLPFC and hippocampus. Although hypothesized to have a deleterious effect on cognition, the true functional significance of this finding for cognition and psychosis is unknown as patients were not included in that study. Given our results, these findings raise the interesting question as to the effect of this variant is on brain connectivity and function in patients. Undertaking this imaging-based analysis in patient groups will therefore be an important next step in linking these brain imaging and neuropsychological findings.

Conclusion

These findings have potential relevance for the nosology of schizophrenia and related psychotic disorders, particularly given that the genetic evidence for traditional clinical dichotomies is a subject of considerable discussion [36, 37]. The present study supports the approach of sub-grouping schizophrenia patients to better understand molecular and biological processes, as has been done for other complex genetic diseases (e.g. breast cancer[38]). Our findings suggests that those with less compromised cognitive functioning may show a different pattern of association or may even be a genetically distinct group worthy of further study in genetic association studies. In light of the stronger association between *ZNF804A* in the combined schizophrenia/bipolar sample in the original GWAS study[5], our data suggest that *ZNF804A* is indexing a psychosis pathway defined by cognitive rather than diagnostic characteristics. If confirmed, defining the molecular etiology involved in this group may have important diagnostic, prognostic and therapeutic implications.

Acknowledgments

The authors sincerely thank all patients who contributed to this study and all staff who facilitated their involvement. Recruitment of the patients from Munich, Germany was partially supported by GlaxoSmithKline (GSK). We are grateful to the Genetics Research Centre GmbH, an initiative by GSK and LMU. Genotyping of the German sample was funded by grants from the MRC and Wellcome Trust. Recruitment and genotyping of the Irish sample was supported by the Wellcome Trust and Science Foundation Ireland (SFI). Dr. Donohoe was supported by an SFI research frontiers grant and a NARSAD Young Investigator Award. Dr. Walters was supported by an MRC/Welsh Assembly clinical research training fellowship. The Australian study was supported by grants 404046 and 513874 from the Australian national health and medical research council (to AJ and LK). All authors acknowledge that they have no conflict of interest in relation to the data contained in this manuscript.

References

1. Prince M, Patel V, Saxena S, et al. Global mental health 1 - No health without mental health. *Lancet.* 2007;370(9590):859-877.
2. Cardno AG, Gottesman II. Twin studies of schizophrenia: From bow-and-arrow concordances to star wars mx and functional genomics. *American Journal of Medical Genetics.* 2000;97(1):12-17.
3. Sanders AR, Duan J, Levinson DF, et al. No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: Implications for psychiatric genetics. *American Journal of Psychiatry.* 2008;165(4):497-506.
4. Todd JA. Statistical false positive or true disease pathway? *Nature Genetics.* 2006;38(7):731-733.
5. O'Donovan MC, Craddock N, Norton N, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nature Genetics.* 2008;40(9):1053-1055.
6. Stefansson H, Ophoff RA, Steinberg S, et al. Common variants conferring risk of schizophrenia. *Nature.* 2009;advanced online publication.
7. International Schizophrenia Corsortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 2009;advanced online publication.
8. Shi J, Levinson DF, Duan J, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature.* 2009;advanced online publication.
9. Atlas AB, http://mouse.brain-map.org/welcome.do;jsessionid=16FD10AF0D38032F4A2FAB2E58B93118. Accessed 01/03/09.
10. Lim J, Hao T, Shaw C, et al. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell.* 2006;125(4):801-814.
11. Gur RE, Calkins ME, Gur RC, et al. The Consortium on the Genetics of Schizophrenia: neurocognitive endophenotypes. *Schizophrenia Bulletin.* 2007;33(1):49-68.
12. Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *American Journal of Psychiatry.* 2003;160(4):636-645.
13. O'Donovan M, Craddock N, Owen M. Schizophrenia: Complex genetics, not fairy tales: A commentary on The emperors of the schizophrenia polygene have no clothes by Crow (2008). *Psychological Medicine.* Dec 2008;38(12):1697-1699.
14. Esslinger C, Walter H, Kirsch P, et al. Neural Mechanisms of a Genome-Wide Supported Psychosis Variant. *Science.* 2009;324(5927):605-.
15. First MB, Spitzer RL, Gibbon M, Williams JB. *Structured Clinical Interview for DSM-IV Axis I Disorders - Patient Edition*. New York: Biometrics Research Department, New York State Psychiatric Institute; 1995.
16. Andreasen NC. *Scale for the Assessment of Negative Symptoms/Scale for the Assessment of Positive Symptoms (Manual)*. Iowa City: University of Iowa Press; 1984.
17. Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (Panss) for Schizophrenia. *Schizophrenia Bulletin.* 1987;13(2):261-276.
18. Rice JP, Reich T, Bucholz KK, et al. Comparison of direct interview and family history diagnoses of alcohol dependence. *Alcohol Clin Exp Res.* 1995;19:1018-1023.
19. Folstein MF, Folstein SE, McHugh PR. *Mini-Mental-Status-Test. German Version*. Beltz: Weinheim; 1990.
20. Wechsler D. *Wechsler Adult Intelligence Scale, Third edition (WAIS-III)*. New York: The Psychological Corporation; 1997.
21. Tewes U. *HAWIE-R Hamburg-Wechsler Intelligenztest fu¨r Erwachsene*. Goettingen: Hogrefe; 1991.
22. Wechsler D. *Wechsler Memory Scale, Third edition (WMS-III)*. New York: The Psychological Corporation; 1998.

23. Härting C, Markowitsch HJ, Neufeld H, Calabrese P, Deisinger K, Kessler J. *Wechsler Memory Scale- Revised; German Version: Wechsler Gedächtnistest; Revidierte Fassung*. Bern: Huber; 2000.

24. Wechsler D. *Wechsler Memory Scale - Revised.* San Antonio: The Psychological Corporation; 1987.

25. Cognition C. *Cambridge Neuropsychological Test Automated Battery, expedio version (CANTABexpedio).* Cambridge: Cambridge Cognition Ltd; 2003.

26. Cornblatt BA, Risch NJ, Faris G, Friedman D, Erlenmeyerkimling L. The Continuous Performance-Test, Identical Pairs Version (Cpt-Ip) .1. New Findings About Sustained Attention in Normal-Families. *Psychiatry Research.* 1988;26(2):223-238.

27. Nuechterlein K, Asarnow R. *3-7 Continuous Performance Test*. Los Angeles: University of California; 2004.

28. Peters K, Wiltshire S, Henders AK, et al. Comprehensive Analysis of Tagging Sequence Variants in DTNBP1 Shows No Association With Schizophrenia or With Its Composite Neurocognitive Endophenotypes. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics.* 2008;147B(7):1159-1166.

29. SPSS. *Statistical package for the social sciences (SPSS) version 14.0*. Chicago, Illinois: SPSS Inc; 2005.

30. Yang MS, Morris DW, Donohoe G, et al. Chitinase-3-like 1 (CHI3L1) gene and schizophrenia: genetic association and a potential functional mechanism. *Biological Psychiatry.* 2008;64(2):98-103.

31. Schretlen DJ, Cascella NG, Meyer SM, et al. Neuropsychological Functioning in Bipolar Disorder and Schizophrenia. *Biological Psychiatry.* 2007;62(2):179-186.

32. Walters JT, Owen MJ. Endophenotypes in psychiatric genetics. *Mol Psychiatry.* 2007;12(10):886-890.

33. Meyer-Lindenberg A, Straub RE, Lipska BK, et al. Genetic evidence implicating DARPP-32 in human frontostriatal structure, function, and cognition. *Journal of Clinical Investigation.* 2007;117(3):672-682.

34. Palmer BW, Heaton RK, Paulsen JS, et al. Is It Possible to Be Schizophrenic Yet Neuropsychologically Normal? *Neuropsychology.* 1997;11(3):437-446.

35. Goldstein G, Shemansky WJ, Allen DN. Cognitive function in schizoaffective disorder and clinical subtypes of schizophrenia. *Archives of Clinical Neuropsychology.* 2005;20(2):153-159.

36. Owen MJ, Craddock N. Diagnosis of functional psychoses: time to face the future. *Lancet.* 2009;373(9659):190-191.

37. 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(3):482-490.

38. Garcia-Closas M, Hall P, Nevanlinna H, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *Plos Genetics.* 2008;4(4)(e1000054).

39. Wing JK, Babor T, Brugha T, et al. Scan - Schedules for Clinical-Assessment in Neuropsychiatry. *Archives of General Psychiatry.* 1990;47(6):589-593.

40. Hallmayer JF, Kalaydjieva L, Badcock J, et al. Genetic evidence for a distinct subtype of schizophrenia characterized by pervasive cognitive deficit. *American Journal of Human Genetics.* 2005;77(3):468-476.

41. Jablensky A. Subtyping schizophrenia: implications for genetic research. *Molecular Psychiatry.* 2006;11(9):815-836.

42. Rey A. Psychological examination of traumatic encephalopathy.Archieves de Psychologic, 1941, 28, 286-340. sections translated by J. Corwin, & F.W. Bylsma. *The Clinical Neuropsychologist.* 1993:4-9.

List of tables and figures:

Irish Sample	**Cases**					**Controls**				
	AA (n=134)	***AC (n=127)***	***CC (n=36)***	***F value/χ^2***	***p***	***AA (n=61)***	***AC (n=77)***	***CC (n=27)***	***F value/χ^2***	***p***
Age (s.d.)	42.2 (12.4)	40.8 (11.3)	40.2 (11.7)	0.64	0.530	36.8 (12.3)	38.0 (13.0)	37.1(12.7)	0.152	0.860
Gender (% female)	34.3	23.6	25	3.90	0.140	45	51.9	57.7	1.33	0.520
Education (Years)	13.1 (2.4)	13.3 (2.6)	12.7 (2.3)	0.78	0.460	16.0 (2.2)	15.7 (2.1)	15.8 (2.8)	0.364	0.700
Medication mg (s.d.)	442.6 (401.0)	444.3 (356.3)	444.1 (379.3)	0.01	0.999					
German Sample	**Cases**					**Controls**				
	AA (n=85)	***AC (n=125)***	***CC (n=41)***	***F value/χ^2***	***p***	***AA (n=512)***	***AC (n=707)***	***CC (n=253)***	***F value/χ^2***	***p***
Age (s.d.)	38.8 (10.8)	36.0 (10.1)	36.8 (11.1)	1.93	0.147	44.6 (14.2)	44.5 (14.2)	44.4 (14.5)	0.022	0.978
Gender (% female)	44.7	31.2	39.0	4.039	0.133	54.5	55.7	58.5	1.103	0.576
Education (% completing high school)	76.5	60.0	65.4	6.354	0.174	80.1	80.9	77.9	3.158	0.532
Medication mg (s.d.)	666 (738)	696 (696)	821 (794)	0.587	0.557					

Table 1: Sample characteristics according to ZNF804A genotype of cases and controls for both Irish and German samples

Cognitive function	Test or Subscale	sample	n	Mean (SD) AA	Mean (SD) AC	Mean (SD) CC	$F_{Case\ v\ Controls}$	p	$F_{Main\ effect}$	p	$F_{Interaction\ effect}$	p	$F_{simple\ effect}$	p
IQ	Abbreviated Full Scale IQ	cases	288	90.5 (19.0)	88.5 (15.8)	85.0 (14.0)	361.0	<.0001	0.78	0.46	1.96	0.140		
		controls	164	122 (14.4)	119.6 (14.4)	125.2 (12.7)								
Working Memory	LN sequence	cases	276	7.7 (.29)	7.4 (.31)	6.0 (.57)	295.2	<.0001	0.11	0.90	**4.20**	**0.020**	**4.781**	**0.009**
		controls	163	13.1 (3.7)	13.2 (3.2)	14.3 (2.8)							1.382	0.254
	CANTAB SWM	cases	287	-0.81 (0.1)	-1.1 (0.1)	-1.8 (0.2)	136.4	<.0001	2.6	0.08	**5.04**	**0.007**	**7.159**	**0.001**
		controls	153	.25 (.90)	.26 (.78)	.37 (.59)							0.230	0.795
Episodic Memory	Logical Memory Immediate	cases	283	6.44 (.28)	6.42 (.29)	4.59 (.54)	361.3	<.0001	0.35	0.70	**6.53**	**0.002**	**4.586**	**0.011**
		controls	161	12.5 (2.6)	12.1 (2.7)	13.7 (2.8)							**3.585**	**0.030**
	Logical Memory Delayed	cases	283	7.24 (.27)	7.15 (.28)	6.18 (.51)	344.9	<.0001	0.12	0.89	**3.26**	**0.040**	1.552	0.214
		controls	160	12.5 (2.7)	12.1 (2.7)	13.9 (2.5)							2.119	0.124
Attentional control	CPT_IP (3 letters)	cases	192	1.9 (1.1)	2.0 (1.0)	1.9 (0.9)	-	-	0.145	0.87	-	-	-	-
		controls	none	-	-	-	-	-	-	-	-	-	-	-

Table 2. Irish sample cognitive analysis by ZNF804A genotype. Bold typeface indicates significant results; Shaded area indicates tests with significant results that were taken forward for replication

GermanSample				**AA**	**AC**	**CC**								
Cognitive function	**Test or Subscale**	**sample**	**n**	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**	**$F_{Case\ v\ Controls}$**	**p**	**$F_{Main\ effect}$**	**p**	**$F_{Interaction\ effect}$**	**p**	**$F_{Simple\ effect}$**	**p**
Working Memory	Digit Span	cases	237	14.5 (4.0)	13.2 (3.3)	13.6 (3.5)	32.70	p<.0001	**3.71**	**0.025**	**4.00**	**0.018**	**3.73**	**0.025**
		controls	1836	14.4 (4.0)	14.4 (3.8)	14.3 (4.0)							0.037	0.964
	WMS Spatial WM	cases	243	15.8 (3.1)	15.0 (3.2)	14.3 (3.0)	115.20	p<.0001	**3.25**	**0.040**	**3.24**	**0.040**	**4.18**	**0.017**
		controls	374	17.3 (3.6)	17.2 (3.3)	17.3 (2.5)							0.159	0.853
Episodic Memory	Log Mem - I	cases	239	25.4 (8.4)	24.4 (8.0)	21.5 (6.3)	156.50	p<.0001	1.95	0.144	**3.32**	**0.037**	**3.20**	**0.043**
		controls	376	30.7 (6.7)	31.0 (6.2)	32.0 (5.0)							0.398	0.672
	Log Mem - D	cases	239	28.3 (10.4)	26.8 (9.7)	23.1 (9.1)	155.80	p<.0001	2.42	0.090	**4.69**	**0.010**	**3.60**	**0.029**
		controls	376	34.2 (7.9)	34.8 (6.4)	35.8 (6.2)							0.718	0.49

Table 3: Cognitive analysis by ZNF804A genotype in German replication sample.

Cases and controls in analysis (IQ cutoff)	N (Controls/ Cases)	Odds Ratio	95% CI OR	Allelic		Genotypic	
				X2	p	X2	p
All	Controls 165 Cases 297	1.37	0.95-1.98	2.88	0.089	4.69	0.096
>70	Controls 164 Cases 256	1.49	1.02-2.18	4.29	0.038	4.79	0.091
>80	Controls 162 Cases 186	1.74	1.14-2.65	6.76	0.009	4.43	0.109
>90	Controls 136 Cases 161	2.29	1.41-3.72	11.47	0.001	8.99	0.011
>100	Controls 154 Cases 68	4.59	2.81-9.71	18.39	$1.8x10^{-5}$	18.24	0.0001
>110	Controls 40 Cases 126	6.10	2.11-17.54	13.70	0.0002	9.10	0.011
>120	Controls 88 Cases 14	9.70	1.26-76.9	6.90	.009	6.92	0.031

Table 4: Association between ZNF and SZ (both allelic and by genotype) in the Irish samples excluding cases according to general cognitive ability.

Cases and controls in analysis (IQ cutoff)	N (Controls/ Cases)	Odds Ratio	95% CI OR	Allelic		Genotypic	
				X2	p	X2	p
None	Controls 1462 Cases 251	1.00	0.82-1.21	0	0.989	0.283	0.868
>70	Controls 1460 Cases 225	0.96	0.79-1.18	0.124	0.725	0.196	0.907
>80	Controls 1457 Cases 207	1.02	0.83-1.26	0.048	0.826	0.119	0.942
>90	Controls 1394 Cases 169	1.05	0.83-1.32	0.162	0.688	0.164	0.921
>100	Controls 1236 Cases 126	1.16	0.89-1.52	1.2	0.273	1.44	0.486
>110	Controls 913 Cases 70	1.63	1.12-2.38	6.75	0.009	6.83	0.033
>120	Controls 563 Cases 37	2.20	1.28-3.80	8.48	0.004	8.68	0.013

Table 5: Association between ZNF and SZ (both allelic and by genotype) in the German samples according to general cognitive ability.

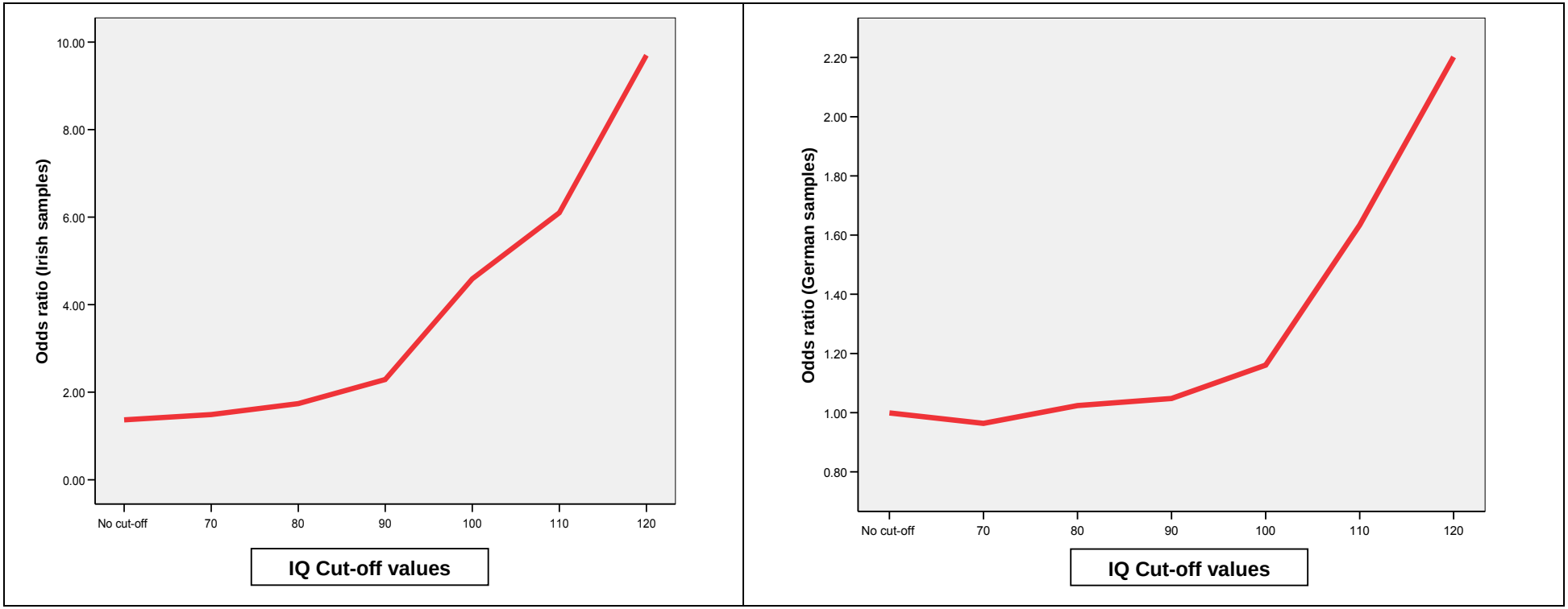

Figure 1. Changes in associated odds ratios for ZNF804A and SZ according on IQ.

Supplementary Materials – Post hoc analysis of episodic memory in an Australian case control sample

Sample characteristics

Australian patient and control samples: This sample, recruited from Perth, Western Australia, consisted of 394 cases and 214 controls of European descent (of whom ~75% were of Anglo-Irish ancestry). Cases comprised clinically stable patients recruited among consecutive admissions to a psychiatric hospital and community-based mental health services in Perth, Western Australia. Clinical diagnosis of affected subjects was based on interviews using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN), version 2.0 [39], a review of case records, and a structured developmental history obtained from a key family member (usually the mother). Research diagnoses were established by consensus between two senior clinicians, who reviewed independently the entire diagnostic information, including the videotape of the SCAN interview, and assigned DSM-IV lifetime diagnoses. Symptom severity was measured using the diagnostic module of the Diagnostic Interview for Psychosis (Castle et al., 2006). Control subjects were recruited by random sampling from local telephone directories (42%), or among Red Cross blood donors (58%), with screening for psychopathology used to exclude individuals with previous diagnosis of psychotic illness in themselves or in any of their first-degree relatives. All participants (including controls) were administered a battery of tests assessing several domains of neurocognitive function, as described in detail [40, 41]. Written informed consent was obtained from all participating subjects. The study was approved by the Human Research Ethics Committee of The University of Western Australia and the North Metropolitan Health Area Ethics Committee, Perth, Western Australia.

Cognitive Assessment: In the Australian sample episodic memory was tested using the Rey Auditory-Verbal Learning Test - immediate and delayed conditions[42]. This task is a brief paper and pencil measure of memory span and new learning. The task consists of 15 nouns which are read aloud for 5 consecutive trials, each trial followed by a free recall test. After a 20-minute delay participants are again required to recall as many of the list of words as they

732 can. As with the other two samples and as to be expected cases performed more poorly than
733 controls on both the immediate and delayed conditions (P<0.0001).
734
735 **Genotyping**
736 In the Australian sample SNP rs1344706 was genotyped with a 5'-exonuclease allelic
737 discrimination assay (Taqman SNP genotyping assay) according to the manufacturer's
738 protocol (Applied Biosystems, Foster City, CA). The ZNF804 rs1344706 'A' allele was not
739 associated with increased risk for schizophrenia in this sample (X^2=0.26; p=0.60; OR=0.94;
740 95%CI:0.73-1.20).
741
742 **Results**
743 From Supplementary table 1 it is evident that for the Australian sample no significant
744 differences were observed for age or gender. A small difference associated with years in
745 education was observed; the GT genotype group had slightly higher (p=.048) years in
746 education, although *post hoc* Tukey analysis did not reveal specific between group
747 differences. Comparisons of symptom severity associated with genotype were again based
748 on a PCA of symptom scores derived from the Diagnostic Interview for Psychosis (DIP,
749 described in Castle et al., 2006). No significant differences in symptom severity between
750 genotype groups were detected. Neither were differences in medication dosage (again
751 measured in terms of chlorpromazine equivalents) observed.
752
753 **Cognitive analysis of ZNF804**
754 Results of our analysis of association between ZNF804 genotype and cognition in the
755 Australian sample are presented in **Table S2**. We failed to find evidence of association
756 between ZNF804 genotype and either immediate or delayed verbal memory, although a trend
757 towards a main effect of genotype and immediate verbal episodic memory was observed in
758 cases (p=.09); in terms of effect size (again derived from partial η^2) this explained less that
759 1% of the variance in immediate verbal memory scores.
760
761

Australian Sample	Cases					Controls				Control
	AA (n=153)	***AC (n=180)***	***CC (n=52)***	***F value/χ^2***	***p***	***AA (n=74)***	***AC (n=112)***	***CC (n=25)***	***F value/χ^2***	***p***
Age (s.d.)	34.1 (9.8)	33.8 (9.7)	33.6 (8.8)	0.05	0.95	41.1 (14.0)	38.5 (13.1)	40.0 (12.5)	0.87	0.42
Gender (% female)	22.9	19.4	17.3	0.98	0.61	41.9	33.9	24.0	2.9	0.24
Education	11.0 (1.8)	11.2 (2.0)	11.1 (2.1)	0.48	0.62	12.5 (2.4)	13.3 (2.7)	12.2 (3.1)	3.09	0.05
Medication (s.d.)	806.0 (536.3)	796.7 (493.3)	702.9 (744.0)	0.44	0.64					

Table S1: Australian sample characteristics according to ZNF804A genotype of cases and controls

Australian Sample

				AA	**AC**	**CC**	***Cases v controls***		***Main effect of genotype***		**Interaction effect**	
Episodic Memory	RAVLT_I	cases	385	19.6(0.45)	20.2 (.41)	21.9 (0.76)	168.36	p<.0001	2.40	0.09	0.26	0.77
		controls	211	27.6(.65)	27.8(.52)	28.7(1.1)						
	RAVLT_D	cases	385	5.15 (0.23)	5.95 (0.21)	5.67 (0.39)	135.69	p<.0001	1.80	0.17	0.651	0.522
		controls	211	8.9 (0.23)	9.1 (0.27)	9.1 (0.57)						

Table S2. Cognitive analysis by ZNF804A genotype in Australian sample.