

# A Genome-wide Study of Common SNPs and CNVs in Cognitive Performance in the CANTAB battery.

Anna C Need<sup>1</sup>, Deborah K Attix<sup>2</sup>, Jill M McEvoy<sup>1</sup>, Elizabeth T Cirulli<sup>1</sup>, Kristen L Linney<sup>1</sup>, Priscilla Hunt<sup>3</sup>, Dongliang Ge<sup>1</sup>, Erin L Heinzen<sup>1</sup>, Jessica M Maia<sup>1</sup>, Kevin V Shianna<sup>4</sup>, Michael E Weale<sup>5</sup>, Lynn F Cherkas<sup>6</sup>, Gail Clement<sup>6</sup>, Tim D Spector<sup>6</sup>, Greg Gibson<sup>3</sup>, David B Goldstein<sup>1\*</sup>.

1. Center for Human Genome Variation, Institute for Genome Sciences and Policy, Duke University, 450 Research Drive, Box 91009, Durham, North Carolina 27708, USA.
2. Division of Neurology and Division of Medical Psychology, Duke University Medical Center, Durham, NC 27705, USA
3. Department of Genetics, Gardner Hall, North Carolina State University, Raleigh, North Carolina 27695, USA.
4. Genomic Analysis Facility, Institute for Genome Sciences and Policy, 450 Research Drive, Box 91009, Duke University, Durham, North Carolina 27708, USA.
5. Department of Medical & Molecular Genetics, King's College London, Guy's Hospital Campus, London SE1 9RT, UK
6. Department of Twin Research and Genetic Epidemiology, King's College London, St. Thomas' Hospital Campus, London, UK

\*Corresponding author: David B. Goldstein, PhD, Center Director, Center for Human Genome Variation, Institute for Genome Sciences and Policy, 450 Research Drive, Box 91009, Duke University,

Durham, North Carolina 27708, USA.

Tel: 919 684 0896

Fax: 919 668 6787

E-mail: [d.goldstein@duke.edu](mailto:d.goldstein@duke.edu)

## Abstract

Psychiatric disorders such as schizophrenia are commonly accompanied by cognitive impairments that are treatment resistant and crucial to functional outcome. There has been great interest in studying cognitive measures as endophenotypes for psychiatric disorders, with the hope that their genetic basis will be clearer. To investigate this, we performed a genome-wide association study involving 11 cognitive phenotypes from the CANTAB battery. We showed these measures to be heritable by comparing the correlation in 100 monozygotic and 100 dizygotic twin pairs. The full battery was tested in approximately 750 subjects, and for spatial and verbal recognition memory (SRM and VRM), we investigated a further 500 individuals to search for smaller genetic effects. We were unable to find any genome-wide significant associations with either SNPs or common copy number variants (CNVs). Nor could we formally replicate any polymorphism that has been previously associated with cognition, although we found a weak signal of lower-than-expected p values for variants in a set of 10 candidate genes. We additionally investigated SNPs in genomic loci that have been shown to harbor rare variants that associate with neuropsychiatric disorders, to see if they showed any suggestion of association when considered as a separate set. Only *NRXN1* showed evidence of significant association with cognition. These results suggest that common genetic variation does not strongly influence cognition in healthy subjects and that cognitive measures do not represent a more tractable genetic trait than clinical endpoints such as schizophrenia. We discuss a possible role for rare variation in cognitive genomics.

## **Introduction**

Memory and cognitive problems are characteristic of many common psychiatric and neurologic disorders. Patients with psychiatric disorders such as schizophrenia and bipolar disorder manifest with cognitive impairments that are unresponsive to treatment (1, 2) and it has been shown that the level of cognitive function is a strong predictor of the ultimate outcome of the disease (3, 4). Alzheimer's disease is characterized by severe memory and cognitive decline with no effective treatment. It is therefore essential, both for the patients and for society, to develop treatments that improve cognitive symptoms in these disorders.

It was hoped that the development of whole-genome technologies and their application to common disease would shed light on the neurobiology underlying psychiatric disease, leading to better treatment options. Unfortunately, several genome-wide association studies have now been performed with schizophrenia patients and controls and have shown it to be unlikely that common single nucleotide polymorphisms (SNPs) or copy number variants (CNVs), as represented in current genome-wide SNP platforms, will succeed in explaining much of the variation in disease predisposition (5-10). Findings have been similar for bipolar disorder (11-14). Although very large scale genetic studies have yet to be reported, it now seems that common variation does not account in large part for the strong heritability of these disorders (15, 16).

One suggested explanation for this 'missing heritability' (17) is that clinical endpoints do not reflect biologically cohesive entities, and that patients with diagnoses such as bipolar disorder and schizophrenia may in fact have a whole range of overlapping syndromes which could be better classified and more successfully investigated if they were defined by more specific phenotypic measures, or endophenotypes (18, 19). Such measures would be easier to investigate genetically because they would represent a more direct link between the genetic dysfunction and its specific biological consequences (20-22).

Cognitive function has been shown to be highly heritable (23-26). Additionally, cognitive impairments associated with neuropsychiatric disorders are also present in unaffected relatives at a higher rate than in the general population and are present in patients even when the illness is not active (27). These facts suggest that cognitive measures perfectly fit the description of ‘endophenotypes’ for neuropsychiatric disease (22). It has therefore been suggested that by studying relatively small sample sizes of healthy subjects we will be able to find associations between cognitive phenotypes and common genetic variation.

To date, a single genome-wide association study has examined the role of common genetic variation in memory (using a word recall task) and implicated the genes *KIBRA*(28) and *CAMTA1* (29). We could not, however, replicate the *KIBRA* association in our own studies (30) and the *CAMTA1* association has yet to be independently replicated. In addition, over the past five years, hundreds of papers have reported associations between particular genetic polymorphisms and memory or other cognitive phenotypes in both healthy volunteers and in patients with schizophrenia and other neuropsychiatric disorders (15). Many of these studies have concentrated on a small number of genes in the dopamine and serotonin pathways (e.g. *COMT*(31-34), *DRD4* (35, 36), *HTR2A*(37-39), or the val66met polymorphism in *BDNF* (40, 41). However, despite the huge body of research, findings have been inconclusive. The interpretation of these studies has been hampered by small sample sizes, publication bias (42, 43) and lack of correction for population stratification. Unfortunately, this has not been unusual in the field of candidate gene association studies (44). In the wake of genome-wide association studies, however, clear standards of evidence for both executing and interpreting human genetic association standards have emerged (45, 46). Here we attempt to apply those standards in our interpretation of the association evidence in the same candidate genes that have been widely studied in cognitive genomics, as well as in analyzing whole genome association data for multiple cognitive phenotypes.

We have investigated the role of common genetic variation in human cognition as assessed using tests from the Cambridge Neuropsychological Test Automated Battery (CANTAB). The CANTAB is a set of automated, computerized tests that assess different aspects of cognitive function. The tests are based on

tasks that have been successfully used to investigate the neural and genetic basis of cognitive function in animals, and have been shown to detect cognitive impairments in many neuropsychiatric conditions including Alzheimer's disease and schizophrenia, as well as being sensitive to differences in cognitive performance in healthy subjects (47-51). A recent study showed that subjects at a high genetic risk of schizophrenia were impaired in performing CANTAB spatial working memory tests, regardless of whether or not they were manifesting any psychotic symptoms (52). Other studies have shown impairments in CANTAB test performance at early stages of schizophrenia development (53) and also independently of current symptoms in both schizophrenia and bipolar disorder (54). These studies provide support for the use of CANTAB tests as measurements for schizophrenia endophenotypes.

It is of value to examine the CANTAB test measures individually due to the likelihood that the different test measures reflect different cognitive processes that are underlain by different molecular pathways.

However, it is possible that there are also genetic determinants of a general cognitive ability. We therefore examine both individual cognitive measures (n=10) as well as a principal component measure that reflects performance across all tests.

## Results

The full CANTAB battery (see methods) was assessed in 1000 subjects and the short battery comprising just the SRM and VRM tests was assessed in a further 630 subjects. After genotyping quality control checks, Eigenstrat analysis and exclusions based on drugs or illness, 1295 subjects remained, comprising 789 that completed the full battery and 506 that completed the short battery. Details of the cohort are displayed in Table 1.

To select which CANTAB measures to use for phenotypic analysis, we used data from 99 monozygotic and 99 dizygotic twins that performed the full CANTAB test battery to assess heritability of the different measures (Table 2). The least heritable measures were PAL first trial score, SWM within errors (all stages) and SWM between errors (4 box stage). These measures were therefore not examined as phenotypes for genetic analysis. To reduce redundancy between the measures, we also dropped SWM between errors (6 boxes) as this had a low heritability and was highly correlated with SWM between errors (8 boxes). Although SSP had a low heritability (14%), there was no more heritable alternative measure from this task so we retained this as a phenotypic measure.

In linear regression models (implemented in PLINK (55)) including sex and 18 EIGENSTRAT axes as covariates (as well as additional covariates shown in Table 3), we found that no single polymorphism showed a genome-wide significant association in the discovery cohort: the lowest p value for any phenotype was  $1.1 \times 10^{-7}$ . The most strongly associated SNPs are shown in Table 3 and the top 100 associated SNPs for each phenotype are shown in Supplementary Table S1. Only six SNPs appeared in the top 100 associations for more than one phenotype (not including PC1). These are shown in Supplementary Table S1 and include SNPs in the 3'UTR of *MDM4* and *KATNAL1*, and intronic SNPs in *HTR2A* (*HTR2A* SNPs further discussed below). Associations with PC1 were among the least significant of all of the traits.

We then went onto examine the role of common CNVs in performance of the CANTAB tests using a set of 286 SNPs shown to tag common CNVs (ref (56) and provided by pers.comm. from Drs Altshuler and McCarroll). Of the 286 tagging SNPs, 178 were represented on all three genotyping chips employed during this study and passed QC in our dataset. None of the CNV-tagging SNPs showed an association with any phenotype in our dataset that was significant at  $p < 0.05$  after correction for all 178 tagging SNPs and 11 phenotypes (significance threshold =  $p = 0.000026$ ). The most strongly associated CNV-tagging SNP was rs4735895, which tagged a CNV at chr8:584944-585908 within the *ERICH1* gene, and associated with IED extradimensional shift errors at  $p = 0.0008$ . One SNP associated with 6/11 of the phenotypes (IED, SSP, PAL 8 patterns, VRM, PRM and PC1), with p values ranging from 0.04 to 0.002. This SNP, rs7829965, tagged a CNV at chr8:24201375-24207011 that was just upstream from the *ADAM18* gene.

Next, we attempted to replicate previous published associations with memory phenotypes where this was possible in our dataset. We searched for genetic association studies performed on healthy human subjects in whom a genetic variant had been associated with a cognitive task based on short-term spatial working memory, immediate verbal recall, paired associates learning, attentional set shifting or sustained attention. For these we looked to see if the specific previously-associated variant was represented in our study, and also what was the lowest p value in our study in the gene and surrounding 20kb (Table 4).

For the 8 previously associated SNPs that were represented either directly or by proxy in our study, 4 associated with  $p < 0.05$  with one or more of our phenotypes, and a further two showed borderline associations at  $p \leq 0.06$ . However, when the significance threshold is corrected for the 11 phenotypes, none of these associations are significant. Since these SNPs may not tag the causal variants, we also looked for association with all genotyped SNPs in each gene, which revealed some associations that withstood correction for multiple testing for all SNPs tested within the individual genes. The strongest associated SNPs were a cluster of three intronic *HTR2A* SNPs: rs2770298, rs9316235 and rs927544, which associated with PAL errors at 8 patterns (PAL8) at  $p = 0.00003-0.00004$ . The first two of these are in

strong linkage disequilibrium (LD) with each other and are not obviously functional. The third SNP, rs927544, is only in partial LD with the other two associated SNPs, and associates with expression of the *HTR2A* gene at  $p=0.04$  (in immortalized B-lymphocytes, using data from the GENEVAR database(57), annotated by WGAviewer (58) ), but not in healthy brain tissue from the SNPEXPRESS database (<http://people.genome.duke.edu/~dg48/SNPExpress/>)(59)). It is interesting to note that these SNPs are not in strong LD with the previously associated functional SNP rs6314, (which did not associate with any of our phenotypes), nor with the other SNPs recently associated with 5-minute and 24-hour delayed (but not immediate) verbal recall (38).

Two other genes, *NRG1* and *KIBRA* (aka *WWCI*) had associations that remained significant after correction for all tested SNPs within the individual genes. The strongest associated SNP in *KIBRA* was rs17633196, which associated with PAL8 with a p value of 0.0002. This intronic SNP was not associated with the previously associated SNP rs17070145 ( $r^2 = 0.06$  in HapMap CEU) but was associated ( $r^2 = 0.77$ ) with a non-synonymous coding SNP, rs17551608 that associated with PAL8 at  $p=0.0007$ . Given that we have previously failed to replicate effects of *KIBRA* on long-term verbal memory (30), these data apparently associating *KIBRA* with short-term spatial memory should be considered with caution. The strongest associated SNP in *NRG1* was rs1481765, an intronic SNP with no obvious function.

It should be noted that, if we test 869 independent SNPs with 11 different phenotypes, we would be likely to see associations at or below 0.00003 one in four times ( $0.00003 * 9559 = 0.27$ ) by chance. We therefore find no convincing evidence supporting the role of any specific polymorphism in and near the set of candidate genes. We can also address whether there is evidence of a general enrichment of low p values amongst the 869 polymorphisms by evaluating a quantile-quantile plot comparing observed and expected p values (Figure 1). This plot reveals an apparent excess of low p values, suggesting the possibility of multiple variants with modest effects within and near these candidate genes.

Finally, a number of genomic regions have been implicated in neurocognitive processes due to the association of rare structural changes with neurological or psychiatric illness. It is possible that these deletions are indicative of genomic regions that harbor common or rare genetic variants that are important in cognitive processes. We therefore focused specifically on these regions to see if there was evidence for increased association with healthy cognition. We selected 7 large deletion regions: 1q21, 8p11, 15q11, 15q13, 16p11.2, 16p12.4-13.11 and 22q11, and five specific genes: *NRXN1*, *CNTN4* and *APBA2* (table 5). Looking just at the PC1 score, we examined the association p values for each of these regions to see if there was an excess of low p values. We found that only one of the regions: the *NRXN1* gene, showed stronger association values than expected by chance (lowest p = 0.00003, significance threshold after Bonferroni correction for 310 SNPs, p=0.0002). The Q-Q plot indicated that the p values for this region for association with PC1 were generally lower than would be expected by chance (supplementary figure 1). Since there was high linkage disequilibrium (LD) across this region, we repeated the analysis, this time using plink (55) to prune the dataset such that all SNPs were in approximate linkage equilibrium with each other. After this, 57 of the 310 original SNPs remained, and the lowest p value was 0.00050, which remained lower than the Bonferroni corrected p value threshold for 57 SNPs of 0.0009, and the highest associated p values were still much lower than expected under the null hypothesis (supplementary figure 1). Because ten subjects had deletions in the *NRXN1* gene (ranging in size from 5kb to 322kb), we looked to see if these were associated with a particular allele of the strongest associated *NRXN1* SNP (rs4971648), however these deletions were distributed across the rs4971648 genotypes so are unlikely to be responsible for this association. None of the other regions showed evidence of significant association with PC1 (table 5). Some of the Q-Q plots indicated a degree of association much less than would be expected by chance, but after elimination of SNPs in LD, the p value distributions conformed approximately to that expected under the null hypothesis, although some remained somewhat lower than expected (supplementary figure 2).

## Discussion

In summary, we have looked for genetic associations with ten cognitive phenotypes and their first principal component in a large sample size using a genome-wide SNP panel thought to represent the majority of common genetic variation in non-African populations (60). We were not able to identify any common SNPs that associated with any of the cognitive tests after correction for all of the tested SNPs.

We also examined for the first time the role of common CNVs in these cognitive phenotypes, using a set of CNV-tagging SNPs reported by McCarroll *et al.* (56). Again, we could not find any associations that remained significant after correction for multiple testing. However, in this case we cannot claim to have performed a comprehensive analysis, since many known common CNVs cannot be identified with the genotyping platforms used in this study.

While our study is a negative one in the sense that there is no polymorphism that can be clearly connected to cognitive performance, there is an intriguing suggestion in our data of a role for polymorphisms near previously studied candidate genes. We emphasize that this observation does not provide a replication of any previous findings because the variants showing the strongest association in our study are not the same variants previously associated, violating now accepted standards of replication (46). On the other hand, the apparent enrichment of low p values in the candidate gene does warrant consideration. One possible explanation is that there are multiple causal variants in these genes creating variable signals of association. We must view these observations as highly tentative, however, since it is difficult to assign formal statistics to patterns in Q-Q plots.

Because several genomic regions have recently been associated with neurologic and psychiatric disorders such as epilepsy, autism, schizophrenia and mental retardation, we also focused specifically on a set of these regions to see if there was evidence for association with healthy cognition. Perhaps surprisingly, these regions were in general rather less associated with our PC1 cognitive score than would be expected under the null hypothesis. The *NRXNI* gene, however, did show an excess of low p values, and the lowest

p value withstood correction for all the SNPs tested at that locus. This suggests one of two possibilities: either the strongly associated SNPs themselves (all intronic) or a genetic variant they are in LD with directly affect the cognitive scores, or that there are multiple rare genetic variants in this region that by chance occur more frequently with one allele of a common variant than the other. Given the multiple recent reports of rare deletions in *NRXNI* in schizophrenia and autism (6, 61-65), this locus seems to warrant further investigation.

In order to assess power, we first note that since we are dealing with a quantitative trait, the power is a function of both the inter-genotype differences and the allele frequency, and these parameters are confounded into the proportion of population variation that the polymorphism is responsible for. A further caveat is that the variant must be sufficiently common to be well represented either directly or indirectly on the gene chips (that is, they have a minor allele frequency of at least about 5%). We assume that such variants are nearly perfectly tagged, and estimate power by simply considering the proportion of variation for the trait in the total population that a given SNP is responsible for.

We can then evaluate the probability that a test statistic will reach a given significance threshold for such a SNP. Thus, for the tests assessed in approximately 750 people, we had 95% power to detect a common variant that explained 5% or more of the variation in the cognitive trait, 84% power to detect a variant accounting for 4% and 58% power to detect a variant accounting for 3%. For the larger sample sizes (SRM and VRM), we had close to 100% power to detect a variant accounting for 4% or more of the variance, 95% power for a variant explaining 3%, and 67% power to detect a variant accounting for 2% or more of the variance. Although it is possible that we could have failed to detect some genetic variants in this study with strong effect sizes, especially those that are not well tagged, it is highly improbable that we would have failed to detect multiple such associations. We can therefore conclusively rule out the possibility that the high heritability of human memory and related cognitive phenotypes is fully accounted for by a small number of SNPs with very strong effects.

There are a number of factors that may serve to reduce the power of this study design to detect associations with certain types of genetic variant. Since the majority of this cohort had some form of higher education, and were aged below 40, we have reduced ability to detect associations that are specific to older or less well educated populations. It is also possible that EIGENSTRAT has not completely corrected for population stratification and this may reduce the power to find truly associated variants that differ strongly in frequency between the two main race/ ethnicities in the study (Asian and European).

We also note that there are many types of learning and memory that we have not assessed in this study, including delayed or remote memory - which could theoretically be under much simpler genetic control. We cannot generalize these findings to other cognitive phenotypes. However, we have examined multiple cognitive domains in this study, and, in a separate study, have also failed to find genome-wide association with traditional neuropsychological test measures (Cirulli et al, unpublished data). This evidence, together with the similar findings emerging from multiple neuropsychiatric traits, suggest that neurocognition, in health and disease, is not going to be a simple phenotype to genetically decipher.

There are a number of possible explanations for the lack of positive findings in this study. It is possible that the traits being measured are too noisy to gauge a reliable phenotype. The phenotypes used in this study, as for other genetic studies of cognition, were collected during a single test session, and many non-genetic variables are known to affect cognitive performance from day to day, including fatigue, hunger, motivation, affective distress, illness. However, we performed heritability assessments on the data gathered in this study by comparing dizygotic and monozygotic twin pairs, and found that most of the measures had a very substantial heritable component by this measure. Some of the heritabilities (especially IED and PRM) may be over- inflated here due to a low correlation in dizygotic twins, however the heritability of memory as measured by CANTAB tests is supported elsewhere using other methods (23). This suggests that these measures are suitable phenotypes for a genetic study. It is possible that methodological weaknesses inherent in the monozygotic-dizygotic twin comparisons have led to systematic overestimation of the heritabilities of cognitive traits. For instance, monozygotic twins, due to

their very similar appearances, may be treated more similarly than dizygotic twins in educational environments. Other types of heritability study e.g. comparison of adopted children to biological parents (66), or of twins reared apart (24) give estimates of heritability of around 0.3 to 0.5 for cognitive tests similar to those used here. However no measure of heritability is without flaws and it is possible that the genetic component of cognition has been systematically overestimated.

Three further explanations for these findings remain consistent with a role for common variation in normal cognition: firstly, that CNVs, or other types of common genetic variant (e.g. microsatellites) that were not well represented by tagging SNPs in this study, could account for the heritability of these traits. However, due to the very large SNP panel used, and the degree to which these SNPs represent the total amount of common variation in the genome (56, 60), it is unlikely that most of the genetic contribution to cognition happens to be both common and unrepresented.

Secondly, it could be that common genetic variation underlies cognitive traits but that the variants interact with each other to such an extent that they do not produce detectable main effects in these sample sizes. Tackling this possibility is going to be difficult, given the number of possibilities for even simple two-way interactions. One is faced with a high risk of either type I error (if one fails to adequately correct for multiple testing) or type II error (since the interaction effects would have to be huge to achieve formal significance in consideration of the number of tests). With currently assembled datasets it seems implausible to perform non-targeted screens for interactive genetic variants with effects on neurocognition. Interactive effects between predetermined candidate variants could be assessed but presently there are no good candidate polymorphisms that consistently replicate between different datasets. Thirdly, the genetic variation could be completely attributed to hundreds of common variants each with a tiny effect size, however there are good reasons for doubting that is the case for cognitive disorders, and arguably for normal cognitive function (6,58).

As has been suggested for other neuropsychiatric traits (67, 68), it is possible that epigenetic modifications affect cognition in healthy subjects. This kind of genetic contribution would be undetectable with these methods and remains to be explored.

Perhaps the most likely explanation for these findings is that the genetic variation underlying neurocognitive traits is too rare to be detectable with current genotyping platforms. There is strong recent evidence for this in other neurocognitive traits such as schizophrenia, autism and epilepsy, in which large, rare, copy number variants have been associated with disorders that have not shown obvious associations with common variation in genome-wide screens (6, 69-72). This hypothesis has considerable implications for the future study of cognitive traits.

Whether the majority of associated variants will be only marginally or substantially less common than those we have examined in genome-wide association studies we do not yet know. If the variants are only marginally below the detection threshold for current genome-wide association studies (minor allele frequency around 3-5%), then the new catalogue of variants down to a frequency of 1% generated by the 1000 genomes project (73) will provide a useful set of variants for further study of cognitive phenotypes. However, very large sample sizes would be needed, particularly given the measurement error inherent in cognitive testing .

If, on the other hand, the most important variants are less than 1%, this framework will not work, and complete genomic resequencing at a high coverage will be necessary to detect memory-influencing genetic variants. However, it is not obvious how to apply this methodology to a trait such as normal variation in cognition. For traits that clearly have extreme phenotypes, discovery of very rare associated genetic variants may be possible by searching for enrichment of causal variants in subjects with the extreme phenotypes, as compared to controls. However, the relevant extreme phenotypes for cognition are not obvious. Consideration might be given to looking for association with neurophysiologic phenotypes, such as ion channel activity in cultured neurons, or synaptic vesicle size (74); or to detailed

phenotyping of cognitive tasks that may more directly reflect such specific activity, such as visual attention (75). Another possibility would be to use subjects with exceptional memory (76), or very high IQ. Alternatively, if the endophenotype hypothesis is correct, the same genes that control cognition in healthy subjects are those that have gone awry in schizophrenia. If this is the case, a promising direction would be to search for rare genetic variants that are enriched in schizophrenia patients (e.g. *NRXNI*) and then look at the effects of rare variants in those same genes in healthy people.

Finally, one clear and surprising finding that has emerged from the recent associations of rare CNVs with neuropsychiatric disease is that the same rare variants are associated with multiple neuropsychiatric conditions (such as autism, mental retardation and schizophrenia), rather than being confined to particular disease classifications, and are also present in apparently unaffected people (6, 72, 77-79). Since all these neuropsychiatric conditions can be associated with some form of cognitive impairment, there is some possibility that these rare genetic variants are acting indirectly to contribute to these disorders by causing cognitive changes. Differences in the ultimate phenotype (which, if any, disease state) may then be dependent on interaction with other genetic or environmental influences. If this is the case, detailed cognitive assessments of patients and unaffected relatives carrying such variants may reveal specific cognitive deficits associated with the rare variants.

In conclusion, our findings indicate that cognitive endophenotypes will not be the simple solution to the problem of complexity in schizophrenia genetics, and that, like schizophrenia itself, the heritability of cognitive traits cannot be accounted for by common variants with strong effect. We suggest that these findings may be attributed to a stronger role for rare variation in cognition than previously expected and suggest some possible future directions for cognitive genomics.

## Materials and Methods

**Study Subjects and test administration.** US participants were recruited by means of IRB-approved advertisements displayed around the Duke University and North Carolina State University campuses, and emailed to student and staff list servers. Subjects that performed the full battery were scheduled for a test session during which they filled out consent forms and questionnaires, gave 20ml of blood and performed the cognitive tests. The twins were recruited during a scheduled interview as part of their participation in an ongoing twin study at the Department of Twin Research & Genetic Epidemiology at King's College, London, UK. Their DNA was already banked but they filled out the same questionnaire data and underwent the same cognitive tests as the Duke and NC State participants. Both twins were tested during the same interview session on the same day. The CANTAB battery comprised the following tests, in this order: paired associates learning (PAL), spatial working memory (SWM), verbal recall (VRM) intra-extradimensional set shifting (IED), rapid visual processing (RVP), pattern recognition memory (PRM), spatial span (SSP) and spatial recognition memory (SRM) (see supplementary methods for further details of the test battery). The battery took about 1 hour to complete and was administered to all participants in a private room under supervision of a trained administrator who read the instructions from a script. The test scores were automatically assigned and stored in the CANTAB computers for later extraction. Participants that took part in the short battery were recruited by means of posters and display boards in high traffic areas of the campuses. They filled out consent forms and questionnaires (supplementary methods), gave a saliva sample and underwent testing in a private room at the time of recruitment, without appointment. Testing took 5-10 minutes and was performed by trained personnel as for the full battery. A total of 326 of the full battery participants, as well as all 674 of the short battery participants underwent a modified test battery, designed to remove some practice stages and eliminate ceiling effects. Details of the modifications are described in the supplementary methods.

**Heritability check:** The following test scores were extracted from the twin data (all tested with CANTAB clinical mode): PAL total errors 6 patterns, PAL total errors 8 patterns, PAL first trial memory,

SWM between for 4, 6 and 8 box stages, SWM total within for 4, 6 and 8 box stages, SWM strategy, IED EDS errors, RVP probability of a hit, PRM percent correct, SSP span length, SRM percent correct and VRM free recall. The heritability of each measure was then estimated as twice the difference in correlation between MZ and DZ twins:  $h^2=2(r(MZ)-r(DZ))$ .

**Principal Components Analysis.** The following measures were included in a principal components analysis: PAL Errors 6 patterns, PAL Errors 8 patterns, SWM Between errors (8 boxes), SWM Strategy, IED EDS Errors, RVP Probability of hit, PRM % Correct, SRM % Correct, SSP Span length, VRM Immediate recall. All subjects with no missing data for any of the measures were included, regardless of whether they did the original or the modified test battery (n=701). Only one of each twin pair was included. Three components had an Eigenvalue above 1: the first, with an Eigenvalue of 3.63, accounted for 36% of the variation in the data and all tests contributed in the same direction to this axis, suggesting that this axis may reflect general cognitive ability. When PC1 was compared between monozygotic and dizygotic twins, the heritability was found to be 0.88. The score on the principal component axis 1 was therefore also included as a cognitive phenotype.

**Genotyping and Quality Control.** All further stages included only one of each twin pair. The samples were genotyped on the Illumina Infinium HumanHap550 version1 (555352 SNPs) or version 3 (561466 SNPs) BeadChips or the Infinium HD 610-quad BeadChips (598966 SNPs, invariate CNV probes were automatically excluded from the SNP association files). The Infinium BeadStudio raw data analysis was carried out as described in Fellay et al (80). For further details, see supplementary methods. After QC procedures, 1576 subjects remained. Only SNPs that were present on each of the three genotyping platforms (Illumina 550v1, 550v3 and 610) were included in the genetic association testing (n=535752). Using the 1295 QC-passed subjects, we then used the PLINK "--missing" function to examine missing data. We found that 47,183 SNPs had missing data in 1% or more of subjects. This is because clustering and calling was done within the different genotyping batches and SNP failures differed between the genotyping runs. These 47,183 SNPs were excluded from analysis, and after this all individuals had a genotyping rate of 99% or greater. We then removed a further 12,598 SNPs with a MAF<0.004. This

criterion ensured that at least 6 individuals of the rare genotype are present in the dataset of 768 subjects that had completed the full battery, to control for error in the estimation of asymptotic  $P$ -values (as alleles with MAF this low or lower have no chance of approaching significance). The final SNP dataset then totaled 475,971, with a total genotyping rate of 0.999518.

**Eigenstrat :** To correct for population structure, we used the EIGENSTRAT approach of Price *et al* (81), which derives the principal components of the correlations among the genotyped SNPs. The scores on the significant principal component axes can then be included in the genetic association test to control for the possibility of false positive associations due to population stratification. For further details, see supplementary methods.

**Exclusions based on drugs, illness and injury.** We collected data on neurological and psychiatric history for 827/1347 participants (details of questionnaire in supplementary methods), and based on this we excluded two subjects that had suffered a stroke and one who had had a brain tumor or brain surgery. In addition, a pharmacist examined all the drugs and drug combinations that the subject reported and recommended the exclusion of 49 subjects, of whom 21 were taking amphetamine based drugs, 11 were taking drugs that may cause excessive drowsiness and 17 were removed based on their particular types and combinations of antiepileptics, antipsychotics, antidepressants and antianxiety drugs. This left 1295 individuals in the analysis.

**Genetic association testing.** In order to determine which factors to include as covariates in the genetic analysis we performed linear regressions to determine the effects of different covariates on the different cognitive phenotypes. Only subjects with genetic data were included. The covariates analyzed were: age, sex, whether English was learned as second language, site of testing, whether the original or modified battery was used, whether the full battery or the short battery (comprising just SRM and VRM) was taken, and education. With the exception of sex, which was included in all analyses, the covariates were only included if they showed a significant effect on the particular cognitive phenotype. Due to differences in educational systems between the US and the UK this was classified broadly into four categories: not completed college, current undergraduate student, obtained an undergraduate degree, current graduate

student OR obtained a postgraduate degree. Because some twins were missing education data, for tests where education was a significant predictor of score these subjects were excluded. We tested for SNP associations with the 11 memory phenotypes with a linear regression model using the PLINK software (55) and including 18 EIGENSTRAT axes as covariates, as well as the additional covariates shown in table 3. An additive genetic model was assumed. The results were then annotated using the WGAViewer software (82).

**Funding**

This work was supported by start-up monies awarded to David Goldstein by the Duke University Institute for Genome Sciences & Policy. The TwinsUK cohort receives funding from the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F2- (ENGAGE project grant agreement HEALTH-F4-2007-201413 and the FP-5 GenomEUtwin Project QLG2-CT-2002-01254). The study also receives support from the Dept of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. TDS is an NIHR senior Investigator. The project also received support from a *Biotechnology and Biological Sciences Research Council* (BBSRC) project grant (G20234).

**Acknowledgments**

This study was funded by Duke IGSP start-up monies awarded to David Goldstein. We are grateful to all the participants who took part in this study. We also thank Nicole Walley, Jeff Dawson, Helen Onabanjo, Paola Nicoletti, Ana-Patricia Wagoner, Joshua Elmore, Liisa Bevan and Janice Hunkin for test administration, and Ian Cartland and Stuart Hacker from Cambridge Cognition who helped us design and implement the modified versions of the tests.

**Conflict of Interest Statement**

The authors declare no conflict of interest.

## References

1. Summers, M., Papadopoulou, K., Bruno, S., Cipolotti, L. and Ron, M.A. (2006) Bipolar I and bipolar II disorder: cognition and emotion processing. *Psychol. Med.*, **36**, 1799-1809.
2. Reichenberg, A. and Harvey, P.D. (2007) Neuropsychological impairments in schizophrenia: Integration of performance-based and brain imaging findings. *Psychol. Bull.*, **133**, 833-858.
3. Green, M.F., Kern, R.S. and Heaton, R.K. (2004) Longitudinal studies of cognition and functional outcome in schizophrenia: implications for MATRICS. *Schizophr. Res.*, **72**, 41.
4. Kolakowska, T., Williams, A.O., Ardern, M., Reveley, M.A., Jambor, K., Gelder, M.G. and Mandelbrote, B.M. (1985) Schizophrenia with good and poor outcome. I: Early clinical features, response to neuroleptics and signs of organic dysfunction. *Br. J. Psychiatry*, **146**, 229-239.
5. Lencz, T., Morgan, T.V., Athanasiou, M., Dain, B., Reed, C.R., Kane, J.M., Kucherlapati, R. and Malhotra, A.K. (2007) Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol. Psychiatry*, **12**, 572-580.
6. Need, A.C., Ge, D., Weale, M.E., Maia, J., Feng, S., Heinzen, E.L., Shianna, K.V., Yoon, W., Kasperaviciute, D., Gennarelli, M. *et al.* (2009) A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet*, **5**, e1000373.
7. Sullivan, P.F., Lin, D., Tzeng, J.Y., van den Oord, E., Perkins, D., Stroup, T.S., Wagner, M., Lee, S., Wright, F.A., Zou, F. *et al.* (2008) Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol. Psychiatry*, **13**, 570-584.
8. O'Donovan, M.C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V., Nikolov, I., Hamshere, M., Carroll, L., Georgieva, L. *et al.* (2008) Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat. Genet.*, **advanced online publication**.
9. Shifman, S., Johannesson, M., Bronstein, M., Chen, S.X., Collier, D.A., Craddock, N.J., Kendler, K.S., Li, T., O'Donovan, M., O'Neill, F.A. *et al.* (2008) Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet*, **4**, e28.
10. Kirov, G., Zaharieva, I., Georgieva, L., Moskvina, V., Nikolov, I., Cichon, S., Hillmer, A., Toncheva, D., Owen, M.J. and O'Donovan, M.C. (2008) A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Mol. Psychiatry*.
11. Ferreira, M.A., O'Donovan, M.C., Meng, Y.A., Jones, I.R., Ruderfer, D.M., Jones, L., Fan, J., Kirov, G., Perlis, R.H., Green, E.K. *et al.* (2008) Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat. Genet.*, **40**, 1056-1058.
12. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661-678.
13. Sklar, P., Smoller, J.W., Fan, J., Ferreira, M.A., Perlis, R.H., Chambert, K., Nimgaonkar, V.L., McQueen, M.B., Faraone, S.V., Kirby, A. *et al.* (2008) Whole-genome association study of bipolar disorder. *Mol. Psychiatry*, **13**, 558-569.

14. Baum, A.E., Akula, N., Cabanero, M., Cardona, I., Corona, W., Klemens, B., Schulze, T.G., Cichon, S., Rietschel, M., Nothen, M.M. *et al.* (2008) A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol. Psychiatry*, **13**, 197-207.
15. Burmeister, M., McInnis, M.G. and Zollner, S. (2008) Psychiatric genetics: progress amid controversy. *Nat Rev Genet*, **9**, 527-540.
16. Crow, T.J. (2008) The emperors of the schizophrenia polygene have no clothes. *Psychol. Med.*, **38**, 1681-1685.
17. Maher, B. (2008) Personal genomes: The case of the missing heritability. *Nature*, **456**, 18-21.
18. Lin, P.I. and Mitchell, B.D. (2008) Approaches for unraveling the joint genetic determinants of schizophrenia and bipolar disorder. *Schizophr. Bull.*, **34**, 791-797.
19. Szoke, A., Meary, A., Trandafir, A., Bellivier, F., Roy, I., Schurhoff, F. and Leboyer, M. (2008) Executive deficits in psychotic and bipolar disorders - implications for our understanding of schizoaffective disorder. *Eur Psychiatry*, **23**, 20-25.
20. Cannon, T.D. and Keller, M.C. (2006) Endophenotypes in the genetic analyses of mental disorders. *Annu Rev Clin Psychol*, **2**, 267-290.
21. Braff, D.L., Freedman, R., Schork, N.J. and Gottesman, II (2007) Deconstructing schizophrenia: an overview of the use of endophenotypes in order to understand a complex disorder. *Schizophr. Bull.*, **33**, 21-32.
22. Gottesman, II and Gould, T.D. (2003) The endophenotype concept in psychiatry: etymology and strategic intentions. *Am. J. Psychiatry*, **160**, 636-645.
23. Singer, J.J., MacGregor, A.J., Cherkas, L.F. and Spector, T.D. (2006) Genetic influences on cognitive function using The Cambridge Neuropsychological Test Automated Battery. *Intelligence*, **34**, 421-428.
24. Plomin, R., Pedersen, N.L., Lichtenstein, P. and McClearn, G.E. (1994) Variability and stability in cognitive abilities are largely genetic later in life. *Behav. Genet.*, **24**, 207-215.
25. Finkel, D., Pedersen, N.L., McGue, M. and McClearn, G.E. (1995) Heritability of cognitive abilities in adult twins: comparison of Minnesota and Swedish data. *Behav. Genet.*, **25**, 421-431.
26. Benyamin, B., Wilson, V., Whalley, L.J., Visscher, P.M. and Deary, I.J. (2005) Large, consistent estimates of the heritability of cognitive ability in two entire populations of 11-year-old twins from Scottish mental surveys of 1932 and 1947. *Behav. Genet.*, **35**, 525-534.
27. Trandafir, A., Meary, A., Schurhoff, F., Leboyer, M. and Szoke, A. (2006) Memory tests in first-degree adult relatives of schizophrenic patients: a meta-analysis. *Schizophr. Res.*, **81**, 217-226.
28. Papassotiropoulos, A., Stephan, D.A., Huentelman, M.J., Hoerndli, F.J., Craig, D.W., Pearson, J.V., Huynh, K.D., Brunner, F., Corneveaux, J., Osborne, D. *et al.* (2006) Common Kibra alleles are associated with human memory performance. *Science*, **314**, 475-478.
29. Huentelman, M.J., Papassotiropoulos, A., Craig, D.W., Hoerndli, F.J., Pearson, J.V., Huynh, K.D., Corneveaux, J., Hanggi, J., Mondadori, C.R., Buchmann, A. *et al.* (2007) Calmodulin-binding transcription activator 1 (CAMTA1) alleles predispose human episodic memory performance. *Hum. Mol. Genet.*, **16**, 1469-1477.
30. Need, A.C., Attix, D.K., McEvoy, J.M., Cirulli, E.T., Linney, K.N., Wagoner, A.P., Gumbs, C.E., Giegling, I., Moller, H.J., Francks, C. *et al.* (2008) Failure to replicate

- effect of Kibra on human memory in two large cohorts of European origin. *Am J Med Genet B Neuropsychiatr Genet*, **147B**, 667-668.
31. Sheldrick, A.J., Krug, A., Markov, V., Leube, D., Michel, T.M., Zerres, K., Eggermann, T. and Kircher, T. (2008) Effect of COMT val158met genotype on cognition and personality. *Eur Psychiatry*, **23**, 385-389.
  32. Egan, M.F., Goldberg, T.E., Kolachana, B.S., Callicott, J.H., Mazzanti, C.M., Straub, R.E., Goldman, D. and Weinberger, D.R. (2001) Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 6917-6922.
  33. de Frias, C.M., Annerbrink, K., Westberg, L., Eriksson, E., Adolfsson, R. and Nilsson, L.G. (2004) COMT gene polymorphism is associated with declarative memory in adulthood and old age. *Behav. Genet.*, **34**, 533-539.
  34. Goldberg, T.E., Egan, M.F., Gscheidle, T., Coppola, R., Weickert, T., Kolachana, B.S., Goldman, D. and Weinberger, D.R. (2003) Executive subprocesses in working memory: relationship to catechol-O-methyltransferase Val158Met genotype and schizophrenia. *Arch. Gen. Psychiatry*, **60**, 889-896.
  35. Jonsson, E.G., Bah, J., Melke, J., Abou Jamra, R., Schumacher, J., Westberg, L., Ivo, R., Cichon, S., Propping, P., Nothen, M.M. *et al.* (2004) Monoamine related functional gene variants and relationships to monoamine metabolite concentrations in CSF of healthy volunteers. *BMC Psychiatry*, **4**, 4.
  36. Rybakowski, J.K., Borkowska, A., Czerski, P.M., Kapelski, P., Dmitrzak-Weglarczyk, M. and Hauser, J. (2005) An association study of dopamine receptors polymorphisms and the Wisconsin Card Sorting Test in schizophrenia. *J. Neural Transm.*
  37. Reynolds, C.A., Jansson, M., Gatz, M. and Pedersen, N.L. (2006) Longitudinal change in memory performance associated with HTR2A polymorphism. *Neurobiol. Aging*, **27**, 150-154.
  38. Sigmund, J.C., Vogler, C., Huynh, K.D., de Quervain, D.J. and Papassotiropoulos, A. (2008) Fine-mapping at the HTR2A locus reveals multiple episodic memory-related variants. *Biol. Psychol.*, **79**, 239-242.
  39. de Quervain, D.J., Henke, K., Aerni, A., Coluccia, D., Wollmer, M.A., Hock, C., Nitsch, R.M. and Papassotiropoulos, A. (2003) A functional genetic variation of the 5-HT<sub>2a</sub> receptor affects human memory. *Nat. Neurosci.*, **6**, 1141-1142.
  40. Hariri, A.R., Goldberg, T.E., Mattay, V.S., Kolachana, B.S., Callicott, J.H., Egan, M.F. and Weinberger, D.R. (2003) Brain-derived neurotrophic factor val66met polymorphism affects human memory-related hippocampal activity and predicts memory performance. *J. Neurosci.*, **23**, 6690-6694.
  41. Egan, M.F., Kojima, M., Callicott, J.H., Goldberg, T.E., Kolachana, B.S., Bertolino, A., Zaitsev, E., Gold, B., Goldman, D., Dean, M. *et al.* (2003) The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell*, **112**, 257-269.
  42. Barnett, J.H., Scoriels, L. and Munafo, M.R. (2008) Meta-analysis of the cognitive effects of the catechol-O-methyltransferase gene Val158/108Met polymorphism. *Biol. Psychiatry*, **64**, 137-144.
  43. Munafo, M.R., Stothart, G. and Flint, J. (2009) Bias in genetic association studies and impact factor. *Mol. Psychiatry*, **14**, 119-120.
  44. Hirschhorn, J.N., Lohmueller, K., Byrne, E. and Hirschhorn, K. (2002) A comprehensive review of genetic association studies. *Genet Med*, **4**, 45-61.

45. McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P. and Hirschhorn, J.N. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*, **9**, 356-369.
46. Chanoock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D., Bailey-Wilson, J.E. *et al.* (2007) Replicating genotype-phenotype associations. *Nature*, **447**, 655-660.
47. Fray, P.J. and Robbins, T.W. (1996) CANTAB battery: proposed utility in neurotoxicology. *Neurotoxicol. Teratol.*, **18**, 499-504.
48. Sahakian, B.J. and Owen, A.M. (1992) Computerized assessment in neuropsychiatry using CANTAB: discussion paper. *J. R. Soc. Med.*, **85**, 399-402.
49. Levoux, M.N., Potvin, S., Sepehry, A.A., Sablier, J., Mendrek, A. and Stip, E. (2007) Computerized assessment of cognition in schizophrenia: promises and pitfalls of CANTAB. *Eur Psychiatry*, **22**, 104-115.
50. Robbins, T.W., James, M., Owen, A.M., Sahakian, B.J., Lawrence, A.D., McInnes, L. and Rabbitt, P.M. (1998) A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: implications for theories of executive functioning and cognitive aging. Cambridge Neuropsychological Test Automated Battery. *J. Int. Neuropsychol. Soc.*, **4**, 474-490.
51. Robbins, T.W., James, M., Owen, A.M., Sahakian, B.J., McInnes, L. and Rabbitt, P. (1994) Cambridge Neuropsychological Test Automated Battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers. *Dementia*, **5**, 266-281.
52. O'Connor, M., Harris, J.M., McIntosh, A.M., Owens, D.G., Lawrie, S.M. and Johnstone, E.C. (2009) Specific cognitive deficits in a group at genetic high risk of schizophrenia. *Psychol. Med.*, 1-7.
53. Bartok, E., Berecz, R., Glaub, T. and Degrell, I. (2005) Cognitive functions in prepsychotic patients. *Prog. Neuropsychopharmacol. Biol. Psychiatry*, **29**, 621-625.
54. McKirdy, J., Sussmann, J.E., Hall, J., Lawrie, S.M., Johnstone, E.C. and McIntosh, A.M. (2008) Set shifting and reversal learning in patients with bipolar disorder or schizophrenia. *Psychol. Med.*, 1-5.
55. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559-575.
56. McCarroll, S.A., Kuruville, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., de Bakker, P.I., Maller, J.B., Kirby, A. *et al.* (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.*
57. Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C. *et al.* (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, **315**, 848-853.
58. Ge, D., Zhang, K., Need, A.C., Martin, O., Fellay, J., Urban, T.J., Telenti, A. and Goldstein, D.B. (2008) WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res.*, **18**, 640-643.

59. Heinzen, E.L., Ge, D., Cronin, K.D., Maia, J.M., Shianna, K.V., Gabriel, W.N., Welsh-Bohmer, K.A., Hulette, C.M., Denny, T.N. and Goldstein, D.B. (2008) Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol*, **6**, e1.
60. Barrett, J.C. and Cardon, L.R. (2006) Evaluating coverage of genome-wide association studies. *Nat. Genet.*, **38**, 659-662.
61. Glessner, J.T., Wang, K., Cai, G., Korvatska, O., Kim, C.E., Wood, S., Zhang, H., Estes, A., Brune, C.W., Bradfield, J.P. *et al.* (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*, **459**, 569-573.
62. Kirov, G., Gumus, D., Chen, W., Norton, N., Georgieva, L., Sari, M., O'Donovan, M.C., Erdogan, F., Owen, M.J., Ropers, H.H. *et al.* (2008) Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum. Mol. Genet.*, **17**, 458-465.
63. Vrijenhoek, T., Buizer-Voskamp, J.E., van der Stelt, I., Strengman, E., Sabatti, C., Geurts van Kessel, A., Brunner, H.G., Ophoff, R.A. and Veltman, J.A. (2008) Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am. J. Hum. Genet.*, **83**, 504-510.
64. Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O.P., Barnes, M.R., Toulopoulou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E. *et al.* (2008) Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum. Mol. Genet.*
65. Kim, H.G., Kishikawa, S., Higgins, A.W., Seong, I.S., Donovan, D.J., Shen, Y., Lally, E., Weiss, L.A., Najm, J., Kutsche, K. *et al.* (2008) Disruption of neurexin 1 associated with autism spectrum disorder. *Am. J. Hum. Genet.*, **82**, 199-207.
66. Alarcon, M., Plomin, R., Fulker, D.W., Corley, R. and DeFries, J.C. (1998) Multivariate path analysis of specific cognitive abilities data at 12 years of age in the Colorado Adoption Project. *Behav. Genet.*, **28**, 255-264.
67. Crow, T.J. (2007) How and why genetic linkage has not solved the problem of psychosis: review and hypothesis. *Am. J. Psychiatry*, **164**, 13-21.
68. Abel, T. and Zukin, R.S. (2008) Epigenetic targets of HDAC inhibition in neurodegenerative and psychiatric disorders. *Curr Opin Pharmacol*, **8**, 57-64.
69. Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O.P., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E. *et al.* (2008) Large recurrent microdeletions associated with schizophrenia. *Nature*, **455**, 232-236.
70. Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M.A., Green, T. *et al.* (2008) Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.*, **358**, 667-675.
71. Friedman, J.I., Vrijenhoek, T., Markx, S., Janssen, I.M., van der Vliet, W.A., Faas, B.H., Knoers, N.V., Cahn, W., Kahn, R.S., Edelman, L. *et al.* (2008) CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. *Mol. Psychiatry*, **13**, 261-266.
72. Helbig, I., Mefford, H.C., Sharp, A.J., Guipponi, M., Fichera, M., Franke, A., Muhle, H., de Kovel, C., Baker, C., von Spiczak, S. *et al.* (2009) 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat. Genet.*, **41**, 160-162.
73. Kaiser, J. (2008) DNA sequencing. A plan to capture human diversity in 1000 genomes. *Science*, **319**, 395.

74. Qu, L., Akbergenova, Y., Hu, Y. and Schikorski, T. (2009) Synapse-to-synapse variation in mean synaptic vesicle size and its relationship with synaptic morphology and function. *J. Comp. Neurol.*, **514**, 343-352.
75. Bellgrove, M.A., Chambers, C.D., Johnson, K.A., Daibhis, A., Daly, M., Hawi, Z., Lambert, D., Gill, M. and Robertson, I.H. (2007) Dopaminergic genotype biases spatial attention in healthy children. *Mol. Psychiatry*, **12**, 786-792.
76. Parker, E.S., Cahill, L. and McGaugh, J.L. (2006) A case of unusual autobiographical remembering. *Neurocase*, **12**, 35-49.
77. Mefford, H.C., Sharp, A.J., Baker, C., Itsara, A., Jiang, Z., Buysse, K., Huang, S., Maloney, V.K., Crolla, J.A., Baralle, D. *et al.* (2008) Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N. Engl. J. Med.*, **359**, 1685-1699.
78. Brunetti-Pierri, N., Berg, J.S., Scaglia, F., Belmont, J., Bacino, C.A., Sahoo, T., Lalani, S.R., Graham, B., Lee, B., Shinawi, M. *et al.* (2008) Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat. Genet.*, **40**, 1466-1471.
79. Hannes, F.D., Sharp, A.J., Mefford, H.C., de Ravel, T., Ruivenkamp, C.A., Breuning, M.H., Fryns, J.P., Devriendt, K., Van Buggenhout, G., Vogels, A. *et al.* (2008) Recurrent reciprocal deletions and duplications of 16p13.11: The deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant. *J. Med. Genet.*
80. Fellay, J., Shianna, K.V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A. *et al.* (2007) A whole-genome association study of major determinants for host control of HIV-1. *Science*, **317**, 944-947.
81. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*, **38**, 904-909.
82. Ge, D., Zhang, K., Need, A.C., Martin, O., Fellay, J., Urban, T.J., Telenti, A. and Goldstein, D.B. (2008) WGAViewer: Software for genomic annotation of whole genome association studies. *Genome Res.*
83. Takamiya, K., Mao, L., Hagan, R.L. and Linden, D.J. (2008) The glutamate receptor-interacting protein family of GluR2-binding proteins is required for long-term synaptic depression expression in cerebellar Purkinje cells. *J. Neurosci.*, **28**, 5752-5755.
84. Moessner, R., Marshall, C.R., Sutcliffe, J.S., Skaug, J., Pinto, D., Vincent, J., Zwaigenbaum, L., Fernandez, B., Roberts, W., Szatmari, P. *et al.* (2007) Contribution of SHANK3 mutations to autism spectrum disorder. *Am. J. Hum. Genet.*, **81**, 1289-1297.
85. Durand, C.M., Betancur, C., Boeckers, T.M., Bockmann, J., Chaste, P., Fauchereau, F., Nygren, G., Rastam, M., Gillberg, I.C., Anckarsater, H. *et al.* (2007) Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat. Genet.*, **39**, 25-27.
86. Wong, A.H., Lipska, B.K., Likhodi, O., Boffa, E., Weinberger, D.R., Kennedy, J.L. and Van Tol, H.H. (2005) Cortical gene expression in the neonatal ventral-hippocampal lesion rat model. *Schizophr. Res.*, **77**, 261-270.

87. Tang, Y.P., Shimizu, E., Dube, G.R., Rampon, C., Kerchner, G.A., Zhuo, M., Liu, G. and Tsien, J.Z. (1999) Genetic enhancement of learning and memory in mice. *Nature*, **401**, 63-69.
88. McOmish, C.E., Burrows, E.L., Howard, M. and Hannan, A.J. (2008) PLC-beta1 knockout mice as a model of disrupted cortical development and plasticity: behavioral endophenotypes and dysregulation of RGS4 gene expression. *Hippocampus*, **18**, 824-834.
89. DeLorey, T.M., Handforth, A., Anagnostaras, S.G., Homanics, G.E., Minassian, B.A., Asatourian, A., Fanselow, M.S., Delgado-Escueta, A., Ellison, G.D. and Olsen, R.W. (1998) Mice lacking the beta3 subunit of the GABAA receptor have the epilepsy phenotype and many of the behavioral characteristics of Angelman syndrome. *J. Neurosci.*, **18**, 8505-8514.
90. Zhang, C.S., Bertaso, F., Eulenburg, V., Lerner-Natoli, M., Herin, G.A., Bauer, L., Bockaert, J., Fagni, L., Betz, H. and Scheschonka, A. (2008) Knock-in mice lacking the PDZ-ligand motif of mGluR7a show impaired PKC-dependent autoinhibition of glutamate release, spatial working memory deficits, and increased susceptibility to pentylentetrazol. *J. Neurosci.*, **28**, 8604-8614.
91. Siegl, S., Flor, P.J. and Fendt, M. (2008) Amygdaloid metabotropic glutamate receptor subtype 7 is involved in the acquisition of conditioned fear. *Neuroreport*, **19**, 1147-1150.
92. Goddyn, H., Callaerts-Vegh, Z., Stroobants, S., Dirikx, T., Vansteenwegen, D., Hermans, D., van der Putten, H. and D'Hooge, R. (2008) Deficits in acquisition and extinction of conditioned responses in mGluR7 knockout mice. *Neurobiol. Learn. Mem.*, **90**, 103-111.
93. Fendt, M., Schmid, S., Thakker, D.R., Jacobson, L.H., Yamamoto, R., Mitsukawa, K., Maier, R., Natt, F., Husken, D., Kelly, P.H. *et al.* (2008) mGluR7 facilitates extinction of aversive memories and controls amygdala plasticity. *Mol. Psychiatry*, **13**, 970-979.
94. Callaerts-Vegh, Z., Beckers, T., Ball, S.M., Baeyens, F., Callaerts, P.F., Cryan, J.F., Molnar, E. and D'Hooge, R. (2006) Concomitant deficits in working memory and fear extinction are functionally dissociated from reduced anxiety in metabotropic glutamate receptor 7-deficient mice. *J. Neurosci.*, **26**, 6573-6582.
95. Holscher, C., Schmid, S., Pilz, P.K., Sansig, G., van der Putten, H. and Plappert, C.F. (2005) Lack of the metabotropic glutamate receptor subtype 7 selectively modulates Theta rhythm and working memory. *Learn. Mem.*, **12**, 450-455.
96. Holscher, C., Schmid, S., Pilz, P.K., Sansig, G., van der Putten, H. and Plappert, C.F. (2004) Lack of the metabotropic glutamate receptor subtype 7 selectively impairs short-term working memory but not long-term memory. *Behav. Brain Res.*, **154**, 473-481.
97. Pietilainen, O.P., Paunio, T., Loukola, A., Tuulio-Henriksson, A., Kieseppa, T., Thompson, P., Toga, A.W., van Erp, T.G., Silventoinen, K., Soronen, P. *et al.* (2008) Association of AKT1 with verbal learning, verbal memory, and regional cortical gray matter density in twins. *Am J Med Genet B Neuropsychiatr Genet.*
98. Barnett, J.H., Jones, P.B., Robbins, T.W. and Muller, U. (2007) Effects of the catechol-O-methyltransferase Val158Met polymorphism on executive function: a meta-analysis of the Wisconsin Card Sort Test in schizophrenia and healthy controls. *Mol. Psychiatry*, **12**, 502-509.

99. Aerni, A., Traber, R., Hock, C., Roozendaal, B., Schelling, G., Papassotiropoulos, A., Nitsch, R.M., Schnyder, U. and de Quervain, D.J. (2004) Low-dose cortisol for symptoms of posttraumatic stress disorder. *Am. J. Psychiatry*, **161**, 1488-1490.
100. Jansen, A., Krach, S., Krug, A., Markov, V., Eggermann, T., Zerres, K., Stocker, T., Shah, N.J., Nothen, M.M., Treutlein, J. *et al.* (2009) A putative high risk diplotype of the G72 gene is in healthy individuals associated with better performance in working memory functions and altered brain activity in the medial temporal lobe. *Neuroimage*, **45**, 1002-1008.
101. Stefanis, N.C., Trikalinos, T.A., Avramopoulos, D., Smyrnis, N., Evdokimidis, I., Ntzani, E.E., Ioannidis, J.P. and Stefanis, C.N. (2007) Impact of schizophrenia candidate genes on schizotypy and cognitive endophenotypes at the population level. *Biol. Psychiatry*, **62**, 784-792.
102. Stone, J.L. and O'Donovan, M.C. and Gurling, H. and Kirov, G.K. and Blackwood, D.H. and Corvin, A. and Craddock, N.J. and Gill, M. and Hultman, C.M. and Lichtenstein, P. *et al.* (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*.
103. Roohi, J., Montagna, C., Tegay, D.H., Palmer, L.E., DeVincent, C., Pomeroy, J.C., Christian, S.L., Nowak, N. and Hatchwell, E. (2009) Disruption of contactin 4 in three subjects with autism spectrum disorder. *J. Med. Genet.*, **46**, 176-182.
104. Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K.K., Holmans, P., Craddock, N., Owen, M.J. and O'Donovan, M.C. (2009) Support for the involvement of large cnvs in the pathogenesis of schizophrenia. *Hum. Mol. Genet.*
105. Ben-Shachar, S., Lanpher, B., German, J.R., Qasaymeh, M., Potocki, L., Nagamani, S., Franco, L.M., Malphrus, A., Bottenfield, G.W., Spence, J.E. *et al.* (2009) Microdeletion 15q13.3: A locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J. Med. Genet.*
106. Pagnamenta, A.T., Wing, K., Akha, E.S., Knight, S.J., Bolte, S., Schmotzer, G., Duketis, E., Poustka, F., Klauck, S.M., Poustka, A. *et al.* (2009) A 15q13.3 microdeletion segregating with autism. *Eur. J. Hum. Genet.*, **17**, 687-692.
107. Bijlsma, E.K., Gijsbers, A.C., Schuurs-Hoeijmakers, J.H., van Haeringen, A., Franssen van de Putte, D.E., Anderlid, B.M., Lundin, J., Lapunzina, P., Perez Jurado, L.A., Delle Chiaie, B. *et al.* (2009) Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. *Eur J Med Genet*, **52**, 77-87.
108. Mefford, H.C., Cooper, G.M., Zerr, T., Smith, J.D., Baker, C., Shafer, N., Thorland, E.C., Skinner, C., Schwartz, C.E., Nickerson, D.A. *et al.* (2009) A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. *Genome Res*.
109. Ullmann, R., Turner, G., Kirchhoff, M., Chen, W., Tonge, B., Rosenberg, C., Field, M., Vianna-Morgante, A.M., Christie, L., Krepischi-Santos, A.C. *et al.* (2007) Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum. Mutat.*, **28**, 674-682.
110. Hoogendoorn, M.L., Vorstman, J.A., Jalali, G.R., Seltén, J.P., Sinke, R.J., Emanuel, B.S. and Kahn, R.S. (2008) Prevalence of 22q11.2 deletions in 311 Dutch patients with schizophrenia. *Schizophr. Res.*, **98**, 84-88.
111. Bassett, A.S. and Chow, E.W. (2008) Schizophrenia and 22q11.2 deletion syndrome. *Curr Psychiatry Rep*, **10**, 148-157.

112. Karayiorgou, M., Morris, M.A., Morrow, B., Shprintzen, R.J., Goldberg, R., Borrow, J., Gos, A., Nestadt, G., Wolynec, P.S., Lasseter, V.K. *et al.* (1995) Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc. Natl. Acad. Sci. U. S. A.*, **92**, 7612-7616.

**Figure 1.** Quantile-quantile plots of the test statistics for the candidate gene associations with all phenotypes. The shaded area depicts the 95% concentration band

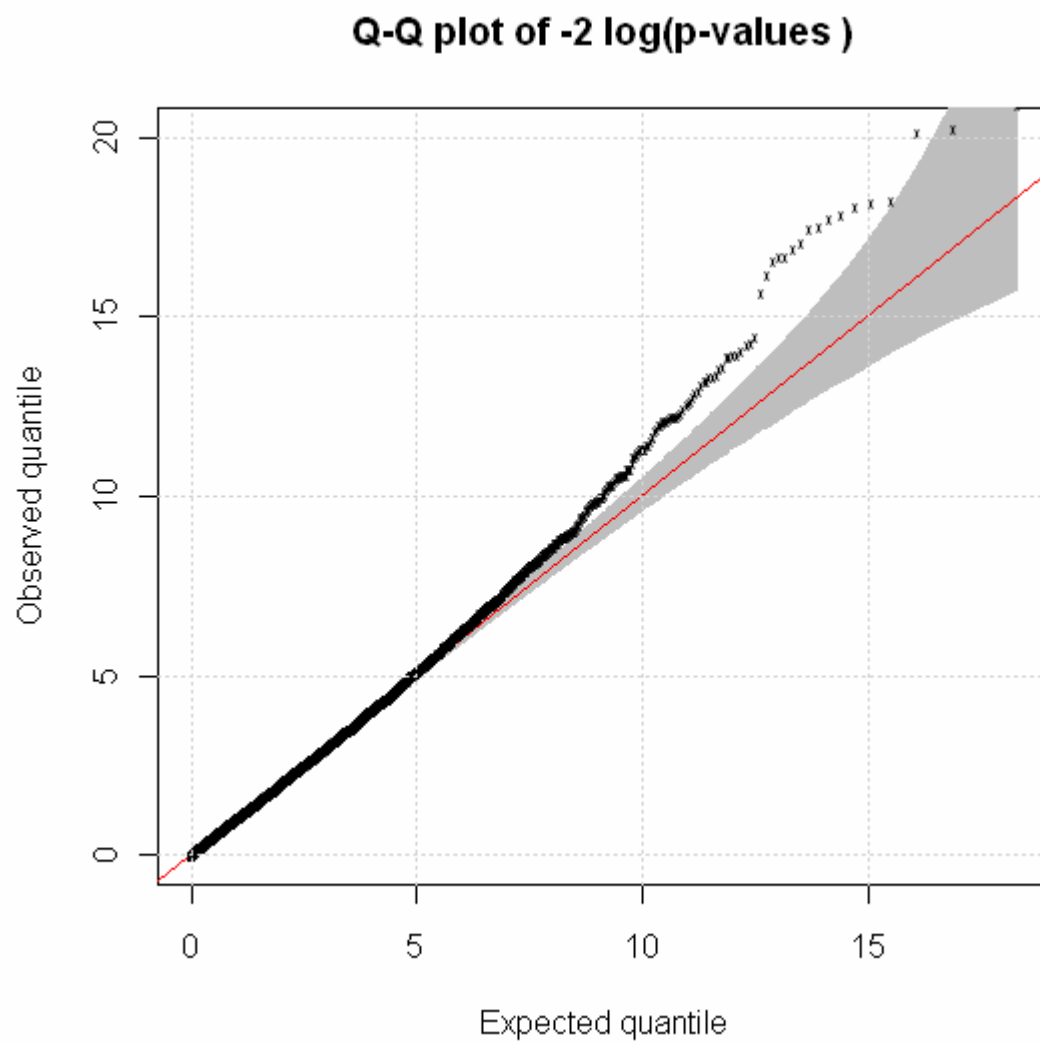


Table 1: Demographics of included participants that took the full CANTAB battery or the short battery comprising just the SRM and VRM tests.

Battery		full battery	short battery	full + short
<b>n</b>		789	506	1295
Sex (% male)		39	53	45
English is not first language (%)		16	20	18
Education	not completed college	11	0	7
	current undergrad	32	72	48
	obtained degree	16	3	11
	postgrad	35	25	31
	missing education	6	0	4
Age mean (sd)		31.2 (14.1)	22.2 (5.5)	27.7 (12.4)
Max age		77	60	77
Min age		18	18	18
Test site (%)	Duke	59	100	75
	London	21	0	13
	NC State	20	0	12
Test version (%)	old	69	0	42
	new	31	100	58
Self-declared race/ethnicity (%)	Caucasian	81	63	74
	East Asian	9	25	16
	South Asian	8	8	8
	Other	2	4	2

**Table 2.** Heritability of CANTAB test measures.

<b>CANTAB Measure</b>	<b>Correlation MZ</b>	<b>Correlation DZ</b>	<b>Heritability (r<sup>2</sup>)</b>
PRM % Correct	0.49	0.04	<b>0.89*</b>
Principal Component 1 (PC1)	0.76	0.32	<b>0.88</b>
VRM Immediate recall	0.44	0.11	<b>0.65*</b>
PAL Errors 6 patterns	0.44	0.16	<b>0.55</b>
SWM Strategy	0.47	0.21	<b>0.53</b>
RVP Probability of hit	0.46	0.26	<b>0.42</b>
PAL Errors 8 patterns	0.46	0.28	<b>0.37</b>
SRM % Correct	0.33	0.17	<b>0.31</b>
SWM Between errors (8 boxes)	0.38	0.24	<b>0.28</b>
IED EDS Errors	0.16	0.03	<b>0.25*</b>
SWM Between errors (6 boxes)	0.28	0.19	<b>0.18</b>
SSP Span length	0.36	0.3	<b>0.14</b>
PAL first trial	0.44	0.39	<b>0.1</b>
SWM Within errors (8 boxes)	0.01	-0.04	<b>0.1</b>
SWM Within errors (6 boxes)	0	0.01	<b>-0.01</b>
SWM Between errors (4 boxes)	0.03	0.1	<b>-0.14</b>
SWM Within errors (4 boxes)	-0.03	0.08	<b>-0.21</b>

MZ: monozygotic twins; DZ: dizygotic twins. PRM: pattern recognition memory, VRM: verbal recognition memory, PAL: paired associates learning, SWM: spatial working memory, RVP: rapid visual processing, SRM: spatial recognition memory IED: intra-extradimensional set shifting, SSP: spatial span. \*Heritability likely to be over-inflated due to very low DZ correlation.

Table 3. Sample sizes, covariates and top hits for the 11 phenotypes.

phenotype	n	covariates <sup>1</sup>	top hit	rs number	closest gene	type of variant	notes
PAL total errors 6 patterns	764	site <sup>2</sup>	1.01E-07	rs5934953	HCCS	81kb upstream	Fifth from top at $p=3 \times 10^{-6}$ is a 3' UTR SNP in <i>GRIP2</i> , a gene involved in synaptic plasticity (83).
PAL total errors 8 patterns	718	education	1.16E-07	rs11856323	COR2B	intronic	
SWM between for 8 box stage	784	ESL, site, battery	2.86E-07	rs2326017	AC103702.3	11kb upstream	Fifteenth from top is an intronic SNP in <i>NRXN1</i> , $p=1.3 \times 10^{-5}$ . Rare <i>NRXN1</i> deletions have been associated with schizophrenia (6, 62, 64) and autism (61, 65).
SWM strategy	721	ESL, education, site <sup>2</sup> , battery	6.30E-06	rs4472969	FXVD2	1.5kb upstream	Fourth from top was an intronic SNP in <i>SHANK3</i> , $p=1.9 \times 10^{-5}$ . Rare genetic variants affecting <i>SHANK3</i> have been associated with autism (61, 84, 85) and possibly epilepsy (61).
IED EDS errors	743	ESL, site	1.69E-06	rs6592284	c11orf73	intronic	Top hit is also close to <i>ME3</i> which shows expression changes in a rat model of schizophrenia (86). Third from top at $p=1.9 \times 10^{-5}$ was an intronic <i>GRIN2B</i> SNP. Overexpression of <i>GRIN2B</i> in mice enhances cognition (87).
RVP probability of a hit	782	ESL, site, battery	2.17E-06	rs6056209	PLCB1	intronic	<i>PLCB1</i> knockout mice are used as schizophrenia models (88).
PRM percent correct	740	education, site <sup>2</sup> , battery	3.23E-07	rs2203512	AC092725.3	88kb downstream	
SSP span length	714	education, site <sup>2</sup> , battery	1.00E-06	rs7729273	AC091951.3	71kb downstream	
SRM percent correct	1227	site <sup>2</sup> , battery, length	5.89E-07	rs2229741	NRIP1	synonymous coding	Second strongest hit is intronic hit in <i>GABRB3</i> , $p=1.9 \times 10^{-6}$ . <i>GABRB3</i> knockout mice have seizures, cognitive deficits and behavioral problems (89).
VRM free recall	1235	education, battery	3.94E-06	rs4258076	SLC24a2	intronic	Fifth from top is intronic hit in <i>GRM7</i> , $1.3 \times 10^{-5}$ . <i>GRM7</i> knockout and transgenic mice show memory

							deficits (90-96)
PC1	707	education, site, battery	2.76E- 06	rs4082514	AC009117. 8	108kb upstream	Sixteenth from top is intronic SNP in <i>NRXN1</i> , rs4971648, $p=3 \times 10^{-5}$ (see above).

<sup>1</sup> Sex and age were included as covariates in all analyses; <sup>2</sup> London only, NC State and Duke scores did not differ; ESL: English learned as a second language, battery: whether the subject performed the original or the modified CANTAB battery, length: whether the subject performed the complete battery or just SRM and VRM.

Table 4. Association results in this study for genes and SNPs that have previously been reported to associate with closely related cognitive phenotypes.

Gene (ref)	Previous association	Previously associated variant	Variant represented in our study	P for previously associated variant in our study	Lowest p in our study (test)	Corrected p, (# tested SNPs in that gene)	Ref
AKT1	Verbal L&M	rs1130214	no proxy		0.013	0.04 (3)	(97)
BDNF	Verbal L&M	rs6265	directly	0.06 (RVP)	0.005	0.07 (15)	(41)
CAMTA1	Verbal L&M	rs4908449	rs1476047*	0.03 (SWM strategy)	0.0009	0.24 (267)	(29)
CLSTN2	Verbal L&M	rs6439886	rs6806737*	0.05 (PAL6)	0.001	0.26 (190)	(28)
COMT	Set shifting	rs4680	directly	0.13 (PAL6)	0.02	0.36 (20)	(98)
CTSD	Spatial L&M	rs17571	directly	0.04 (PRM)	0.03	0.14 (5)	(99)
DAOA	Verbal L&M	rs3918342, rs1421292	rs9519707*, no proxy	0.01 (VRM)	0.004	0.06 (16)	(100)
HTR2A	Verbal L&M	rs6314	directly	0.12 (RVP)	0.00003	<b>0.001 (39)</b>	(39)
KIBRA	Verbal L&M	rs17070145	rs10462974*	0.01 (PAL8)	0.0002	<b>0.01 (69)</b>	(28)
NRG1	Sustained attention and spatial and verbal L&M	rs35753505, rs6994992, SNP8NRG433E1006	No proxy for any		0.0001	<b>0.03 (244)</b>	(101)

L&M: learning and memory; \*perfect proxy:  $r^2$  between previously associated SNP and specified genotyped SNP in HapMap CEU population is 1.

Table 5.

Region	Reason for inclusion	Chr	Implicated region <sup>^</sup>	Region examined*	Size examined *	# SNPs**	lowest p	Bonferroni corrected significance threshold <sup>§</sup>
1q21.1	Deletion associated with schizophrenia, autism, mental retardation, behavioral abnormalities, micro/ macrocephaly and other congenital abnormalities (69, 77, 78, 102),	1	144,106,312-146,293,282	143328536-148205522 <sup>#</sup>	4.88 Mb	241	0.01	0.0002
<b>NRXN1</b>	<b>Deletions including parts of NRXN1 have been associated with schizophrenia (6, 62, 64) and autism (61, 65).</b>	<b>2</b>	<b>49,999,148 - 51,113,178</b>	<b>49875535-51225851</b>	<b>1.35 Mb</b>	<b>310</b>	<b>0.00003</b>	<b>0.0002</b>
CNTN4	Rare deletions and duplications have been implicated in autism (61, 103)	3	2,117,247-3,074,645	2026065-3095571	1.07Mb	432	0.004	0.0001
8p22	Deletion >2Mb found in schizophrenia patient, no deletion of this size seen in controls (6).	8	15,112,352-18,360,000	15090765-18434971	3.34Mb	1062	0.0009	0.00005
15q11.2	Large deletions have been associated with schizophrenia (69, 102, 104)	15	20,306,549-20,777,695	18421386-21339423	2.91Mb	140	0.04	0.0004
APBA2	Deletions have been associated with schizophrenia (6, 62)	15	27,001,135-27,197,807	26017833-27230795	1.21Mb	82	0.04	0.0007
15q13.3	Deletions and duplications have been associated with schizophrenia, autism, epilepsy and other neural phenotypes (69, 72, 102, 105, 106).	15	28,723,577-30,302,218	28079091-30760289	2.68Mb	317	0.008	0.0002
16p11.2	Deletions and duplications have been associated with autism, mental retardation and learning and language disabilities (70, 107).	16	29,500,000-30,100,000	29045893-30453989	1.41Mb	86	0.003	0.0007

16p12.4-13.11	Deletions and duplications have been found in schizophrenia, autism, mental retardation, epilepsy and learning disability(6, 79, 104, 108, 109).	16	15,387,381-18,072,544	14458540-18874875	4.42Mb	526	0.001	0.0001
22q11	22q11 deletion syndrome is associated with a high incidence of psychiatric disorders, particularly schizophrenia (110-112).	22	17,257,788-19,792,353	16974931-20375603	3.4Mb	534	0.0002	0.0001

^ approximate: not all events are identical and endpoints depend to some extent on how the deletion was identified (e.g. using a genome-wide SNP array versus CNV-specific assay); \*implicated region plus approximately 100kb either side; \*\*includes only SNPs included in the association testing (see methods); #Due to sparse coverage in region only 5 SNPs upstream were included; §Corrected for SNPs testing in this region only.