

# Functional Gene Group Analysis Reveals a Role of Synaptic Heterotrimeric G Proteins in Cognitive Ability

Dina Ruano,<sup>3</sup> Gonçalo R. Abecasis,<sup>5</sup> Beate Glaser,<sup>4</sup> Esther S. Lips,<sup>1</sup> L. Niels Cornelisse,<sup>1</sup> Arthur P.H. de Jong,<sup>1</sup> David M. Evans,<sup>4</sup> George Davey Smith,<sup>4</sup> Nicolas J. Timpson,<sup>4</sup> August B. Smit,<sup>2</sup> Peter Heutink,<sup>3</sup> Matthijs Verhage,<sup>1</sup> and Danielle Posthuma<sup>1,3,\*</sup>

Although cognitive ability is a highly heritable complex trait, only a few genes have been identified, explaining relatively low proportions of the observed trait variation. This implies that hundreds of genes of small effect may be of importance for cognitive ability. We applied an innovative method in which we tested for the effect of groups of genes defined according to cellular function (functional gene group analysis). Using an initial sample of 627 subjects, this functional gene group analysis detected that synaptic heterotrimeric guanine nucleotide binding proteins (G proteins) play an important role in cognitive ability ( $P_{EMP} = 1.9 \times 10^{-4}$ ). The association with heterotrimeric G proteins was validated in an independent population sample of 1507 subjects. Heterotrimeric G proteins are central relay factors between the activation of plasma membrane receptors by extracellular ligands and the cellular responses that these induce, and they can be considered a point of convergence, or a “signaling bottleneck.” Although alterations in synaptic signaling processes may not be the exclusive explanation for the association of heterotrimeric G proteins with cognitive ability, such alterations may prominently affect the properties of neuronal networks in the brain in such a manner that impaired cognitive ability and lower intelligence are observed. The reported association of synaptic heterotrimeric G proteins with cognitive ability clearly points to a new direction in the study of the genetic basis of cognitive ability.

## Introduction

Cognitive ability is a highly heritable trait, with heritability estimates ranging from 40% in young childhood to 80% in late adulthood.<sup>1–3</sup> Disturbance in cognitive functioning ultimately causes lower test intelligence and is related to various psychiatric conditions, including schizophrenia, mental retardation, and autism. Identifying the actual genetic (and environmental) factors influencing cognitive ability may therefore elucidate the etiological basis of individual differences in both normal and abnormal cognitive functioning.<sup>4</sup> Typically, each of the reported DNA variants associated with cognitive ability explain less than 1%–2% of the variation, and common variants in only a few genes (cholinergic receptor, muscarinic 2, *CHRM2* [MIM 118493], catechol-O-methyltransferase, *COMT* [MIM 116790], and brain-derived neurotrophic factor, *BDNF* [MIM 113505]) have repeatedly shown associations.<sup>5</sup> Despite the high heritability of cognitive ability and the major efforts made to understand it during the past decades, we are currently left with the generally accepted conclusion that “cognitive ability is most likely influenced by many genes of small effect, that possibly interact,”<sup>6,7</sup> without knowing the identity of these genes.

In the past decade, the scale of genotyping and genetic association studies has increased rapidly, from single-locus analysis to genome-wide association studies (GWAS) that

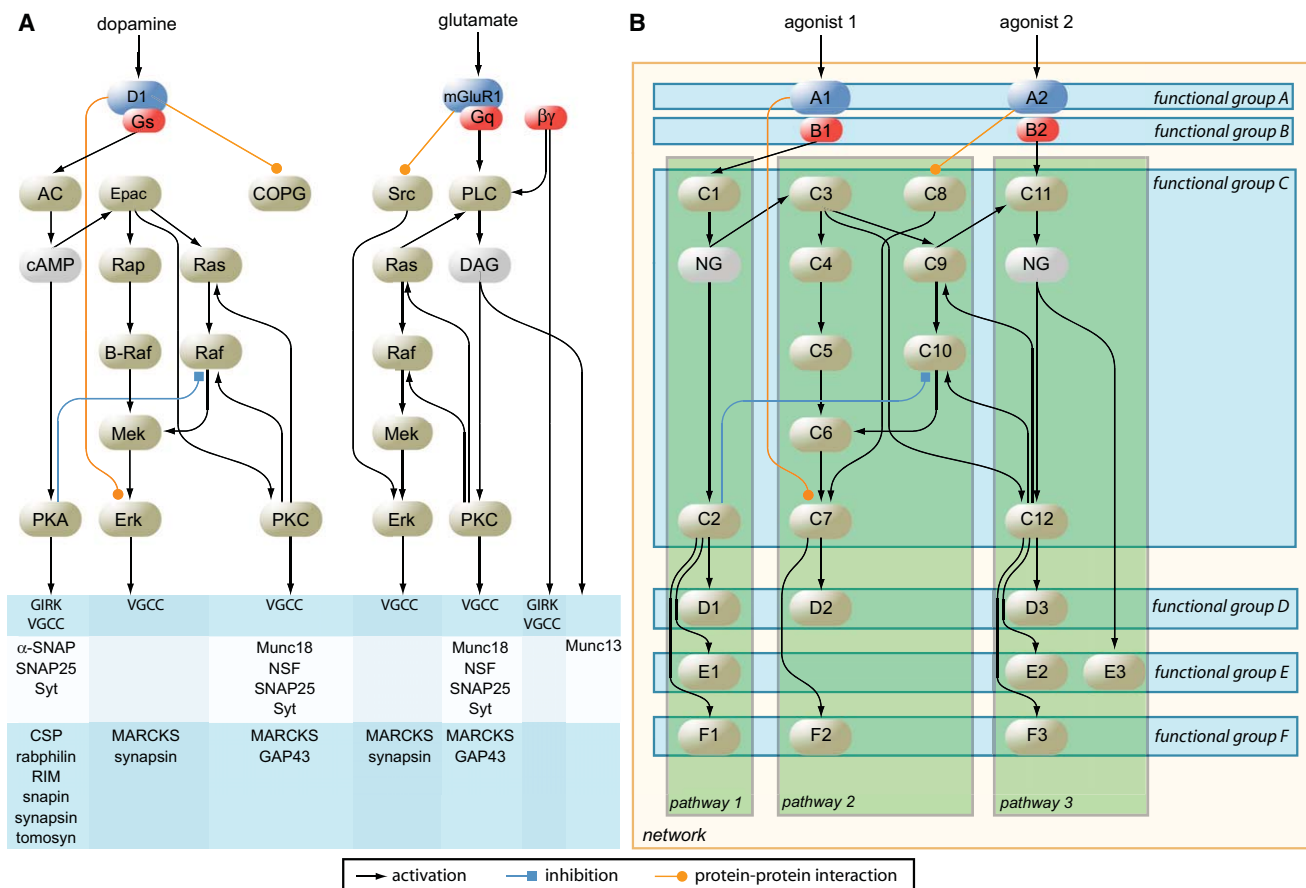
allow the screening of 500,000–1,000,000 SNPs, covering 65%–95% of the human genome, depending on the array of choice. This has proven a successful method of identifying common genes of relatively large effect<sup>8–13</sup> but has been less effective when rare variants of large effect are of importance or when identifying genes of small effect for complex traits, such as attention-deficit hyperactivity disorder (ADHD [MIM 143465]) or schizophrenia (MIM 181500).<sup>14,15</sup>

Collective testing of genes involved in biological pathways has emerged as an alternative strategy for testing the combined effects of genetic variants with small effect size.<sup>16,17</sup> Such pathways are usually defined as a set of proteins that participate in cascades of intracellular reactions, often triggered by extracellular ligands, involving enzyme-catalyzed posttranslational modifications of proteins and/or changes in their subcellular distribution. This ultimately leads to changes in cellular responses, such as altered gene expression or adaptation of cell morphology. Indeed, testing the combined effect of multiple genetic variants in such pathways has been shown to be more powerful than testing single-gene effects.<sup>17</sup> Many proteins, however, are known to act *across* pathways, suggesting that they are not exclusively linked to one pathway. For instance, different neuromodulators, such as dopamine or 5HT, activate receptors in the plasma membrane, which are functionally and often structurally similar to the receptors in other

<sup>1</sup>Department of Functional Genomics, <sup>2</sup>Department of Molecular & Cellular Neurobiology, Center for Neurogenomics and Cognitive Research, <sup>3</sup>Department of Medical Genomics, VU Medical Center, Neuroscience Campus Amsterdam VU University, 1081 HV Amsterdam, The Netherlands; <sup>4</sup>MRC Centre for Causal Analyses in Translational Epidemiology, Department of Social Medicine, University of Bristol, Bristol BS8 2BN, UK; <sup>5</sup>Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

\*Correspondence: danielle.posthuma@cncr.vu.nl

DOI 10.1016/j.ajhg.2009.12.006. ©2010 by The American Society of Human Genetics. All rights reserved.



**Figure 1. Schematic Representation of Vertical Pathways versus Horizontal Functional Groups**

(A) Genes that are involved in two different biological pathways (actual dopamine and glutamate pathways) are illustrated. It can be clearly seen that genes are not exclusively active in one pathway, but tend to be part of more than one pathway.

(B) Genes are grouped according to similar cellular function, and naturally each gene is exclusively assigned to one functional group.

pathways. A higher degree of convergence exists for downstream signaling steps, such as guanine nucleotide binding proteins (G proteins), effector enzymes, second messengers, kinases, and classes of substrates. It is conceivable that genetic variation that influences complex traits accumulates at foci of convergence that act *across* different pathways—but have similar cellular function—and that can influence multiple biological pathways in a similar manner. We refer to grouping genes according to cellular function as “horizontal grouping,” as opposed to the classical grouping of genes in pathways, which we refer to as “vertical grouping” (Figure 1).

Horizontal grouping may be especially powerful in *synaptic* protein networks. Many different pathways have been described, all of which modulate synaptic function. Importantly, these pathways have a high degree of convergence, which has a common consequence; i.e., many modulate synaptic strength and change the coupling between neurons. Therefore, it is conceivable that genetic variation in (many) different pathways leads to similar consequences in synaptic function and hence to a common phenotypic effect. Collective testing of genes *across* pathways addresses this possibility.

Previous gene-finding studies for cognitive ability have focused on targeted candidate-gene testing or whole-genome linkage or associations studies. Here, we focus on evaluating the combined effect of multiple genes in both vertical pathways and horizontal functional gene groups. Results from the functional gene group approach clearly suggest the involvement of synaptic heterotrimeric G proteins in cognitive ability.

## Subjects and Methods

### Study Subjects

Subjects were part of the international collaborative research effort on the genetics of ADHD (the International Multi-Center ADHD Gene project [IMAGE]), and data of these subjects were obtained from the Database of Genotypes and Phenotypes (dbGaP). The IMAGE project was one of the six projects selected by the National Institutes of Health (NIH) in 2006 to be genotyped in the first phase of the Genetic Association Information Network (GAIN).

Characterization of the samples used in the IMAGE project has been described elsewhere in detail.<sup>18–20</sup> In brief, 947 nuclear families of European descent (2844 individuals) from eight countries (Belgium, England, Germany, Holland, Ireland, Israel, Spain, and

Switzerland) were included in the analysis. In this sample, standardized intelligence quotient (IQ) scores were available for 627 unrelated subjects (87.7% males). The age at the time of data collection for these 627 subjects ranged from 5 to 19 yrs, with a mean age of 11.0 (standard deviation [SD] = 2.7). Their IQ was assessed with the Wechsler Intelligence Scale for Children (WISC-III-R),<sup>21</sup> in accordance with procedures described by Sattler.<sup>22</sup> No information on medication status or treatment was included in the analysis.

### Genotyping and Quality Control

Genotyping was performed by Perlegen Sciences with the use of their genotyping platform, which comprises 600,000 tagging SNPs designed to be in high linkage disequilibrium with untyped SNPs for the three HapMap populations. Genotype data were cleaned by the National Center for Biotechnology Information (NCBI) with the use of the GAIN QA/QC software package before being uploaded to the dbGaP database. Details of the quality control (QC) process have been reported elsewhere.<sup>23</sup> A total of 438,783 SNPs survived QC and were available in the downloaded data set.

### Definition of Functional Gene Groups and Biological Pathways

As opposed to defining groups of genes according to the neuromodulator involved, we grouped genes horizontally; that is, across the classically defined pathways. Functional gene grouping was based on cellular function and relies on previous protein identification and data mining for synaptic genes and gene function. Because no selective purification method exists for the purification and analysis of presynaptic proteins, the criterion for proteins from the presynapse was based on manually curated data mining and confirmed by comparison to proteomic data from whole-synapse (pre plus post) analyses and previously characterized presynaptic subcomplexes in solubilized preparations.

Determination of genes expressed in the presynaptic terminal was based on the following set of criteria:

- Null mutation produces a secretion phenotype in synapses.
- Activation of the gene product (e.g., receptor) or blockade thereof modulates secretion in synapses.
- Immuno-electronmicroscopy detects the gene product in the terminal.
- The gene product is enriched in affinity-purified synaptic vesicle fraction.
- Overexpression produces a secretion phenotype in synapses.
- Immunocytochemistry colocalizes the gene product specifically with the presynaptic marker.
- The gene product is enriched in purified brain synaptosomes.
- Null mutation produces a secretion phenotype in secretory cells and the gene is expressed in the human brain.

Inclusion of postsynaptic proteins was based on the presence of proteins in proteomics analyses of the synapse, including preparations of synaptic membrane fractions<sup>24</sup> as well as the postsynaptic density.<sup>25,26</sup>

Synaptic genes are subdivided into 17 functional groups of genes on the basis of shared function into a biological process (i.e., horizontal grouping) and manually curated published data. These groups are as follows: cell adhesion and transsynaptic

signaling molecules; cell metabolism (synaptic metabolic enzymes and their cofactors, excluding mitochondrial proteins); endocytosis (proteins involved in endocytosis); excitability (voltage-gated ion channels); exocytosis (proteins involved in regulated secretion); G protein relay (G protein subunits); GPCR signaling (G protein-coupled receptors); intracellular signal transduction (enzymes downstream of G protein or tyrosine kinase [TK] signaling); intracellular trafficking (vesicle adaptors, sorting proteins, motor proteins); ion balance/ transport (ion and solute carriers and exchangers); ligand-gated ion-channel signaling; neurotransmitter metabolism (metabolizing enzymes); peptide/neurotrophin signaling (neuropeptide, trophic factors, hormones); protein clustering (scaffolding proteins); RNA and protein synthesis, folding, and breakdown; structural plasticity (cytoskeletal proteins and their regulators); and TK signaling (tyrosine receptor kinases).

One group was formed of remaining genes that are known to be expressed in the synapse but do not share any cellular function with other genes (this gene group was called “unknown”).

In addition, four canonical pathways (vertical grouping) for synaptic modulation were defined: the dopaminergic, glutamatergic, serotonergic, and cannabinoid pathways. These pathway definitions are in line with previously published definitions and can be obtained, for example, from regular textbooks or online biological-pathway definitions. All locus IDs in each functional gene group and canonical pathway, as well as the main references used to determine the functional groups, are available online in Tables S1 and S2.

### Data Processing

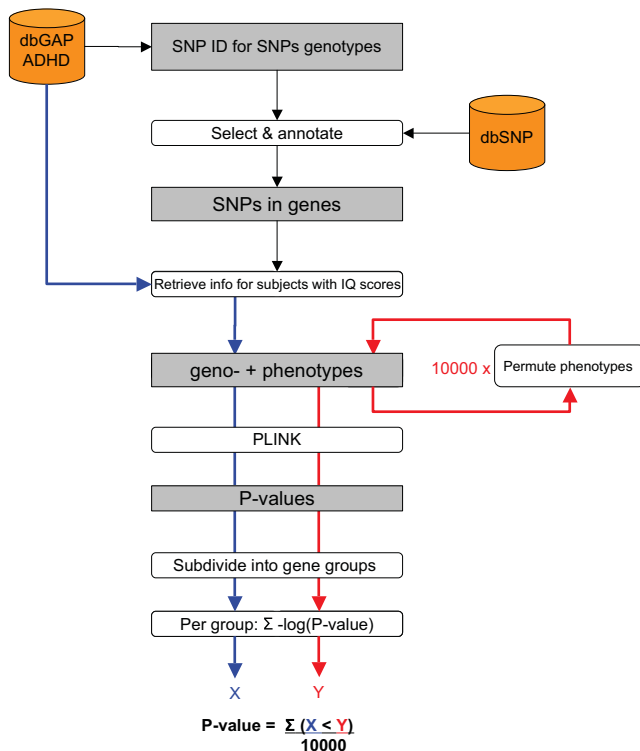
A schematic overview of the sequence of data-processing steps applied in this study is provided in Figure 2.

#### SNP Annotation

All SNPs that survived QC in the GAIN-IMAGE project were annotated according to the criteria of NCBI's SNP database (dbSNP). All SNPs that occur within or in the vicinity of genes were identified. A SNP was assigned to a gene on the basis of positions of the gene's reference sequence (RefSeq), with the addition of 2 kb to the 5' end and 500 bp to the 3' end of the gene. This region has been shown to include promoters (between -300 and -50 bp from the transcriptional start site [TSS]), as well as putative negative elements (between -1000 and -500 upstream of the TSS) and canonical hexanucleotide signals (20-50 bp from the 3' end of the pre-RNA).<sup>27</sup> SNPs mapped to multiple genes were assigned to each of these genes.

#### Statistical Analysis Method for Collective Testing of the Association of Gene Groups or Pathways

First, SNP-by-SNP analysis was carried out via linear regression. The  $\Sigma\text{-log}_{10}(P)$  was calculated with the use of all p values obtained by single SNP analysis of selected SNPs in a group of genes, as defined above. The  $\Sigma\text{-log}_{10}(P)$  is, however, not directly interpretable, because SNPs with the lowest p values will have the largest contribution to the calculated sum. As a result, this procedure can detect association of gene groups with a relatively low number of SNPs with low p values, even if some SNPs in the group have conventionally nonsignificant p values. In addition, the number of SNPs in a group will affect the test statistic  $\Sigma\text{-log}_{10}(P)$ , in the sense that a group containing more SNPs will have a higher  $\Sigma\text{-log}_{10}(P)$  as compared to a group with fewer SNPs, assuming similar p values per SNP. Also, groups that contain SNPs with low p values and multiple SNPs in high linkage disequilibrium (LD) with those SNPs will have a higher  $\Sigma\text{-log}_{10}(P)$  than will groups



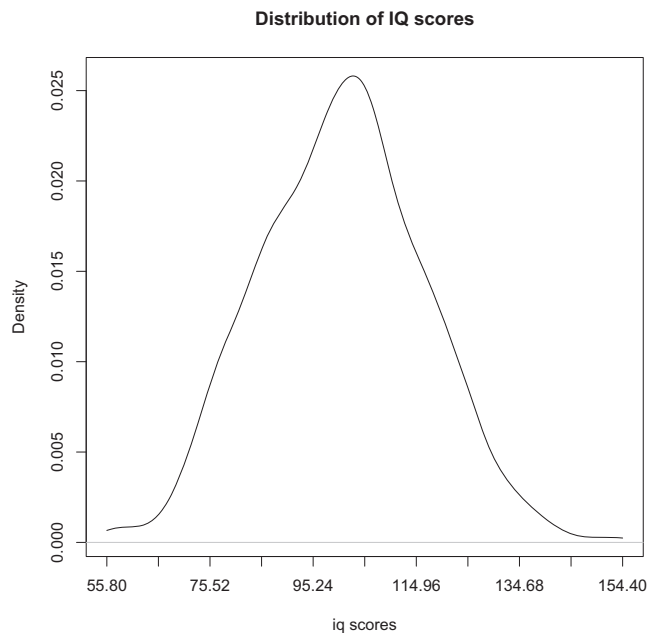
**Figure 2. Data-Processing Steps**

From the dbGAP database, IDs were retrieved for SNPs genotyped in the GAIN-IMAGE study that included IQ phenotypes. From the dbSNP database, those SNPs that occur within and in the vicinity of genes were selected and annotated with the corresponding gene IDs. IQ scores were retrieved from the dbGAP database, as well as genotypes for the selected SNPs. p values were calculated for each SNP with PLINK. The SNPs and their p values were subdivided into functional groups. Per group, the negative log of all p values was summed to obtain value X. The last three steps were repeated after permutation of the IQ scores between subjects, giving a value Y for each permutation and each group. Per group, the number of times that Y was bigger than X was divided by the number of permutations to obtain the group's p value. Grey boxes represent data. Rounded boxes represent data-processing steps. Cylinders represent databases.

that contain SNPs of such low p values but without multiple SNPs in LD with those SNPs, given that p values are a function of LD between SNPs. Sample size will also affect the  $\Sigma\text{-log}_{10}(P)$ , because larger sample sizes are more likely to yield lower p values and thus higher  $\Sigma\text{-log}_{10}(P)$ .

So that unbiased interpretation of the test statistic was allowed, permutations ( $n = 10,000$ , or  $100,000$  when necessary) were carried out, conditional upon these factors, by permutation of phenotypes (i.e., the IQ scores) over genotypes. With this permutation procedure, only the relation between any genetic variant and the phenotype is disconnected, whereas the genomic haplotypic structure is kept intact. In addition, each selected group of genes will include exactly the same number of SNPs and genes, as well as the same haplotypic structure, and is based on the same sample size as that of the original data set.

For each group, we then determined by permutation how likely the observed  $\Sigma\text{-log}_{10}(P)$  of that group was, given the number of SNPs and genes, the LD structure, and the sample size. A conservative Bonferroni correction was used to correct for multiple testing of different gene groups. Because we tested 23 different gene



**Figure 3. Distribution of the IQ Scores in the 627 Individuals Analyzed in the GAIN-IMAGE Sample**

groups (all synaptic genes + 17 functional gene groups + one miscellaneous group and four biological pathways), the significance level was set at  $0.05/23 = 0.0022$ . Given this significance level, the discovery sample of 627 subjects had sufficient power for detection of effect sizes of at least 2.4% of the total variation.

### Software

All SNP-by-SNP analyses were conducted in PLINK.<sup>28</sup> SNP selection, permutation of the data sets, and calculations of the combined effect were implemented in scripts written in the R software package.

### Results

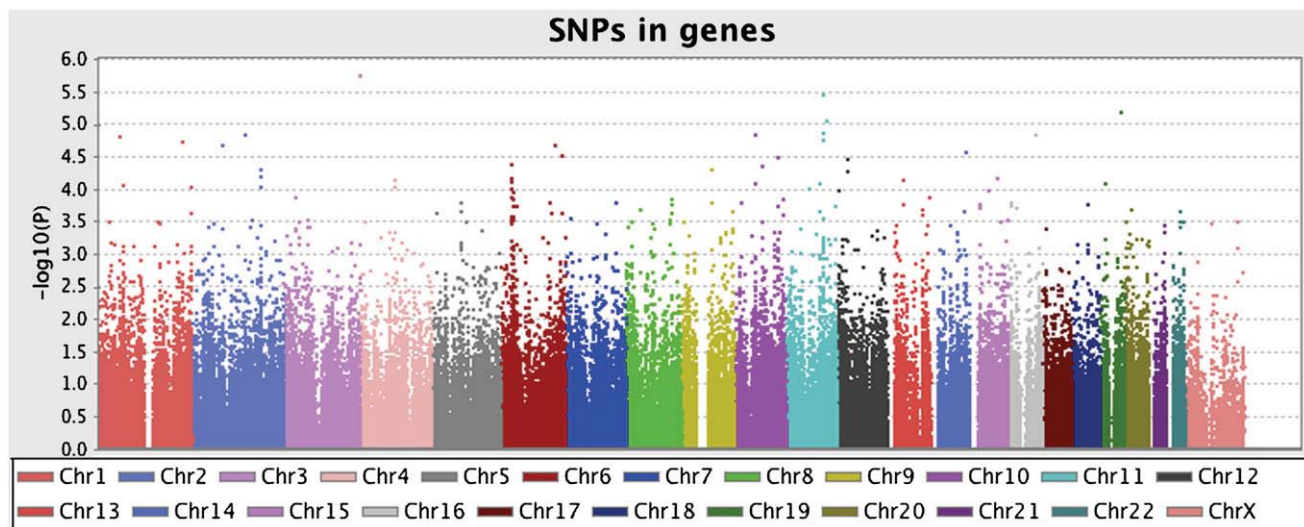
#### IQ Scores Normally Distributed in the GAIN-ADHD Sample

Phenotypic and genotypic data were obtained from dbGAP. IQ scores from 627 individuals were quantified with the Wechsler Intelligence Scales for Children (WISC)<sup>29</sup> and had a mean of 100.7 (SD = 15.7) and a median of 101.6. Skewness was calculated to be 0.063, and the excess kurtosis relative to a normal distribution was  $-0.057$  (Figure 3). Although this sample was originally ascertained for ADHD, this suggests that IQ scores are normally distributed in this sample.

#### Genome-wide SNP-by-SNP Analysis Does Not Detect Robust Associations

After QC, a total number of 438,783 SNPs were available for genome-wide association analysis. Of these, 179,725 SNPs mapped to 16,674 genes. The median number of SNPs per gene was 4.0, and the mean was 11.1, ranging from 1 to 1271 SNPs. First, we tested whether variation in single genes showed robust association with cognitive





**Figure 4. Manhattan Plot Showing Results of SNP-by-SNP Association with Cognitive Ability of the 179,725 SNPs Expressed in Genes** Of these SNPs, 10,237 (5.7%) had p values below 0.05. At lower thresholds, the numbers of p values are as follows: 2230 lower than  $1 \times 10^{-2}$ , 272 lower than  $1 \times 10^{-3}$ , 38 lower than  $1 \times 10^{-4}$ , 4 lower than  $1 \times 10^{-5}$ , and the lowest p value was  $1.8 \times 10^{-6}$ .

ability by applying a standard (SNP-by-SNP) genome-wide association analysis with the use of regression analysis implemented in PLINK<sup>28</sup> and including only SNPs mapped to genes. Results of the SNP-by-SNP association analyses for all SNPs mapped to genes are shown in Figure 4.

With the genome-wide threshold being  $1 \times 10^{-8}$ , none of the single SNPs reached significance. Obviously, this demonstrates mainly a lack of power and underscores the need for larger sample sizes or more sophisticated use of available information. With the current sample size of 627 subjects and a genome-wide significance level of  $1 \times 10^{-8}$ , there was sufficient power for the detection of SNPs explaining at least 6.7% of the variance (Genetic Power Calculator<sup>30</sup>). For the detection of SNPs of small effect size (e.g., 2%) with reasonable power (0.80), a sample of 2138 subjects would have been needed.

#### Analysis of Synaptic Functional Gene Groups Identifies Association with Cognitive Ability

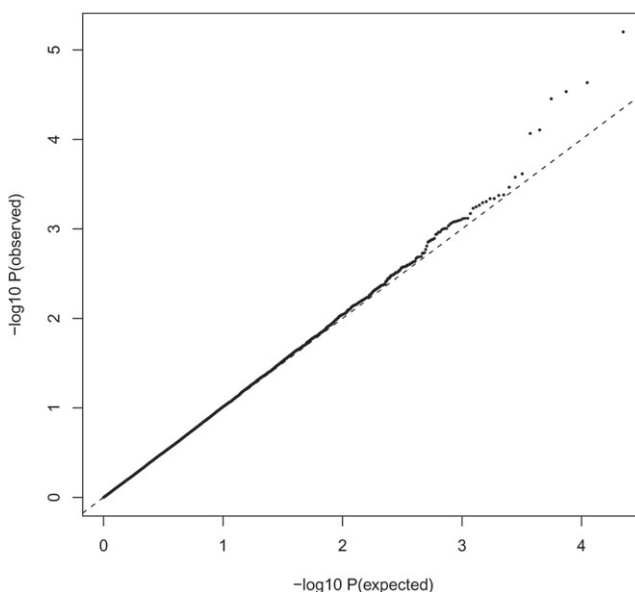
Before testing detailed functional groups, we first tested whether genes expressed in synapses are associated with cognitive ability. This group contains 1024 genes, of which 900 (22,324 SNPs) were included on the Perlegen chip. Our results suggest that SNPs associated with cognitive ability are not randomly distributed across the genome, but tend to cluster in genes that are known to be expressed in synapses ( $p = 0.001$ ) (Figure 5).

The 900 genes that are known to be expressed in the synapse and were included in the Perlegen chip were assigned to one of 17 functional gene groups on the basis of cellular function. Because some of the synaptic genes did not fit into any of the functional synaptic gene networks defined, we also tested this group of remaining synaptic genes as a negative control (the “unknown” group). The 49 genes in this group are expressed in

the synaptic terminal, but their function is currently unknown.

Table 1 lists the results of the joint association analysis of SNPs within each of the functional groups, and Figure 6 and Figure 7 provide the quantile-quantile (Q-Q) plots for each of the tested functional gene groups.

As expected, the “unknown” group did not show any evidence for an association with cognitive ability (empirical  $P_{EMP} = 0.880$ ). The group of synaptic heterotrimeric



**Figure 5. Results of Association Analysis of All Synaptic Genes with Cognitive Ability**

Distribution of expected (under the null hypothesis) versus observed p values for SNPs in genes expressed in synapses. The dashed diagonal line represents the line obtained if the observed distribution did not deviate from the expected distribution. All p values are corrected for  $\lambda$ .

**Table 1. Results of the Association Analysis of Functional Gene Groups and Biological Pathways with Cognitive Ability**

Gene Group	N Genes	N SNPs	$\Sigma\text{-log}_{10}(P)$	$P_{EMP}$
All synaptic genes	900	22325	10146	<b>0.001</b>
<b>Synaptic Functional Gene Group</b>				
G protein relay	25	359	227	<b>0.00019</b>
Neurotransmitter metabolism	26	405	210	0.024
Endocytosis	24	291	153	0.058
Tyrosine kinase signaling	7	512	256	0.064
Cell metabolism	48	368	179	0.081
Excitability	48	1697	793	0.083
Cell adhesion and transsynaptic signaling molecules	74	5209	2359	0.109
Ion balance/transport	37	436	209	0.149
Structural plasticity	81	1474	676	0.160
GPCR signaling	37	1174	540	0.191
Intracellular signal transduction	137	3570	1591	0.236
Ligand gated ion channel signaling	36	1139	517	0.242
RNA and protein synthesis, folding and breakdown	58	597	266	0.341
Protein clustering	44	1322	582	0.395
Intracellular trafficking	66	667	293	0.406
Exocytosis	79	1593	681	0.603
Peptide/neurotrophin signaling	24	645	265	0.698
Unknown	49	867	347	0.880
<b>Biological Synaptic Signaling Pathways</b>				
Glutamate	60	1968	865	0.3883
Dopamine	69	1584	687	0.5006
Serotonin	102	3146	1348	0.6211
Cannabinoid	81	2568	1069	0.8309

Bold values indicate significance after correction for multiple testing.  $P_{EMP}$  denotes empirical p value, based on 10,000 (or 100,000 when necessary) permutations of the data.

The groups have been ordered by their p values. For each selected group, the number of genes and the number of SNPs are listed. In the last two columns, the sum of the negative  $\log_{10}$  of the p value in the original data and the empirical p value are listed for each group.

G proteins, however, yielded an overall test statistic for association with cognitive ability that was better than that expected for a group that size on the basis of chance alone ( $P_{EMP} = 1.9 \times 10^{-4}$ ), given the preset threshold of 0.0022. This suggests that the combined effect of genes within this group plays a role in explaining variation in cognitive ability. In the case of a single SNP, such a p value would correspond to an overall calculated effect size of at

least 3.3% of the variation in cognitive ability.<sup>30</sup> However, it should be noted that this calculated effect size (based on p value, sample size, and significance level) is based on single-SNP effects and is difficult to interpret when many (nonindependent) SNPs are involved.

None of the genes in the group of heterotrimeric G proteins have been associated with cognitive ability previously. This functional gene group consists of the following genes:

G protein alpha 11 (*GNAI1* [MIM 139313]), G protein alpha 12 (*GNAI2* [MIM 604394]), G protein alpha 13 (*GNAI3* [MIM 604406]), G protein alpha 14 (*GNAI4* [MIM 604397]), G protein alpha 15 (*GNAI5* [MIM 139314]), G protein, alpha inhibiting activity polypeptide 1 (*GNAI1* [MIM 139310]), G protein, alpha inhibiting activity polypeptide 2 (*GNAI2* [MIM 139360]), G protein, alpha inhibiting activity polypeptide 3 (*GNAI3* [MIM 139370]), G protein, alpha activating activity polypeptide, olfactory type (*GNAL* [MIM 139312]), G protein, alpha activating activity polypeptide O (*GNAO1* [MIM 139311]), G protein, q polypeptide (*GNAQ* [MIM 600998]), GNAS complex locus (*GNAS* [MIM 139320]), G protein, alpha transducing activity polypeptide 1 (*GNAT1* [MIM 139330]), G protein, alpha z polypeptide (*GNAZ* [MIM 139160]), G protein, beta polypeptide 1 (*GNB1* [MIM 139380]), G protein, beta polypeptide 2 (*GNB2* [MIM 139390]), G protein, beta polypeptide 3 (*GNB3* [MIM 139130]), G protein, beta polypeptide 4 (*GNB4* [MIM 610863]), G protein, beta polypeptide 5 (*GNB5* [MIM 604447]), G protein, gamma 2 (*GNG2* [MIM 606981]), G protein, gamma 3 (*GNG3* [MIM 608941]), G protein, gamma 4 (*GNG4* [MIM 604388]), G protein, gamma 5 (*GNG5* [MIM 600874]), G protein, gamma 7 (*GNG7* [MIM 604430]), G protein, gamma 10 (*GNG10* [MIM 604389]), G protein, gamma 11 (*GNG11* [MIM 604390]), G protein, gamma 12 (*GNG12*).

Two genes in this group were not included in our analyses because there were no SNPs available in the Perlegen chip: *GNAT1* and *GNG3*. Notably, none of the single SNPs within this network would have been detected in the context of a genome-wide SNP-by-SNP analysis, with the lowest p value being  $4.9 \times 10^{-5}$ . In fact, only four SNPs within this functional gene group reached a p value below  $10^{-3}$ , and only 12.8% of SNPs had a p value below 0.05. Three of these four SNPs with a p value below  $10^{-3}$  are in the *GNAQ* gene, and the 46 SNPs with a p value lower than 0.05 are distributed throughout 11 different genes: *GNAI4*, *GNAQ*, *GNG2*, *GNAS*, *GNG4*, *GNG11*, *GNB5*, *GNAI1*, *GNAL*, *GNAO1*, *GNB3*, with frequencies of 13, 9, 8, 5, 3, 3, 1, 1, 1, 1, and 1, respectively. This suggests that the effect of the functional group cannot be explained by the effect of a few individual SNPs or genes but must be ascribed to the combined effect of multiple genes in the functional gene group. This is also evident in the Q-Q plot of the heterotrimeric G proteins in Figure 6 (upper left panel), which shows that the distribution of the observed p values for the functional group of genes encoding heterotrimeric G proteins are consistently lower than

FUNCTIONAL GENE GROUPS

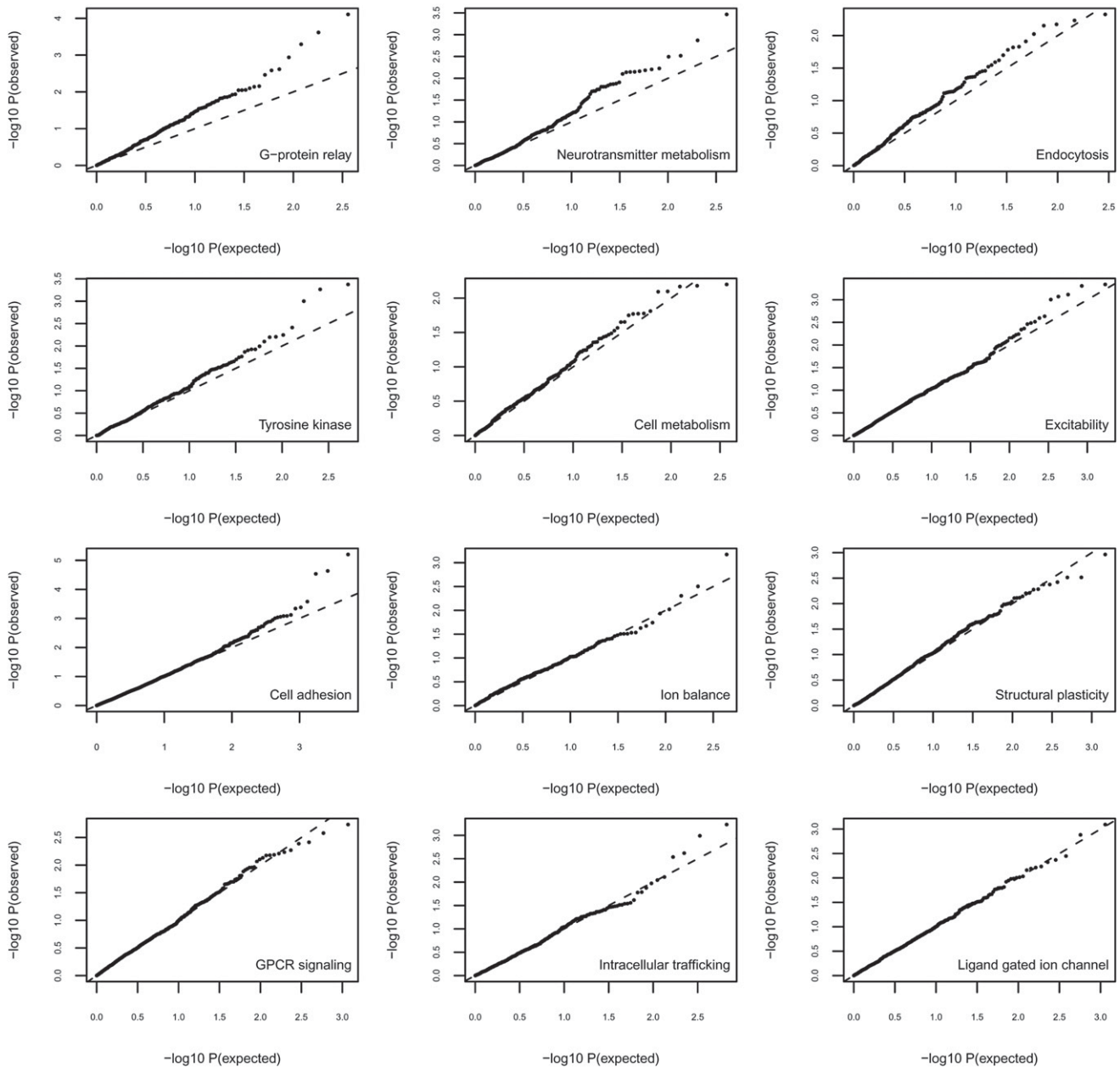


Figure 6. Results of Association Analysis Corrected for  $\lambda$  of the Twelve Most Significant Functional Gene Groups

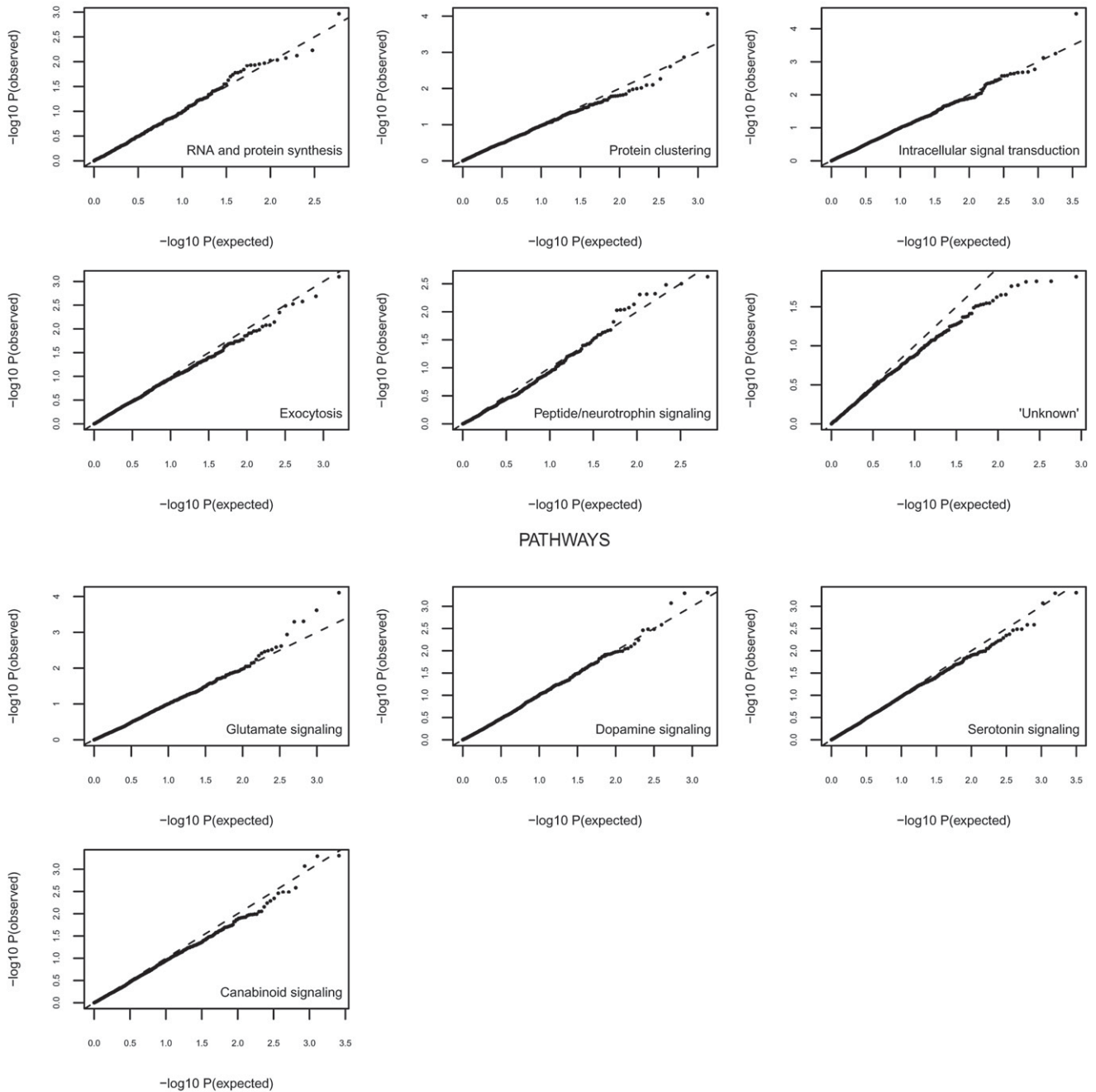
expected under the null hypothesis of uniform distribution, with no single  $p$  value standing out.

**Biological Pathway Analysis Shows No Association with Cognitive Ability**

Because the total collection of synaptic genes was associated with cognitive ability, we tested four biological pathways that are known to involve synaptic functioning as one of their roles in cognitive ability. These pathways involve the dopaminergic, glutamate, serotonergic, and cannabinoid pathways and follow the rationale of vertical grouping. Of these pathways, the serotonin pathway

showed the largest number of genes overlapping with the synaptic genes. The dopaminergic, glutamatergic, serotonergic, and cannabinoid pathways yielded the following empirically derived  $p$  values: dopaminergic,  $P_{EMP} = 0.5006$ ; glutamatergic,  $P_{EMP} = 0.3883$ ; serotonergic,  $P_{EMP} = 0.6211$ ; cannabinoid,  $P_{EMP} = 0.8309$  (see also Table 1 and Figure 7), suggesting that collective testing of genes in synaptically relevant biological pathways is less successful in identifying genetic variation underlying cognitive ability than collectively testing genes that are grouped according to function in a biological process (horizontal grouping).

FUNCTIONAL GENE GROUPS (CONT.)



**Figure 7. Results of Association Analysis Corrected for  $\lambda$  of the Remaining Functional Gene Groups and All Tested Biological Pathways**

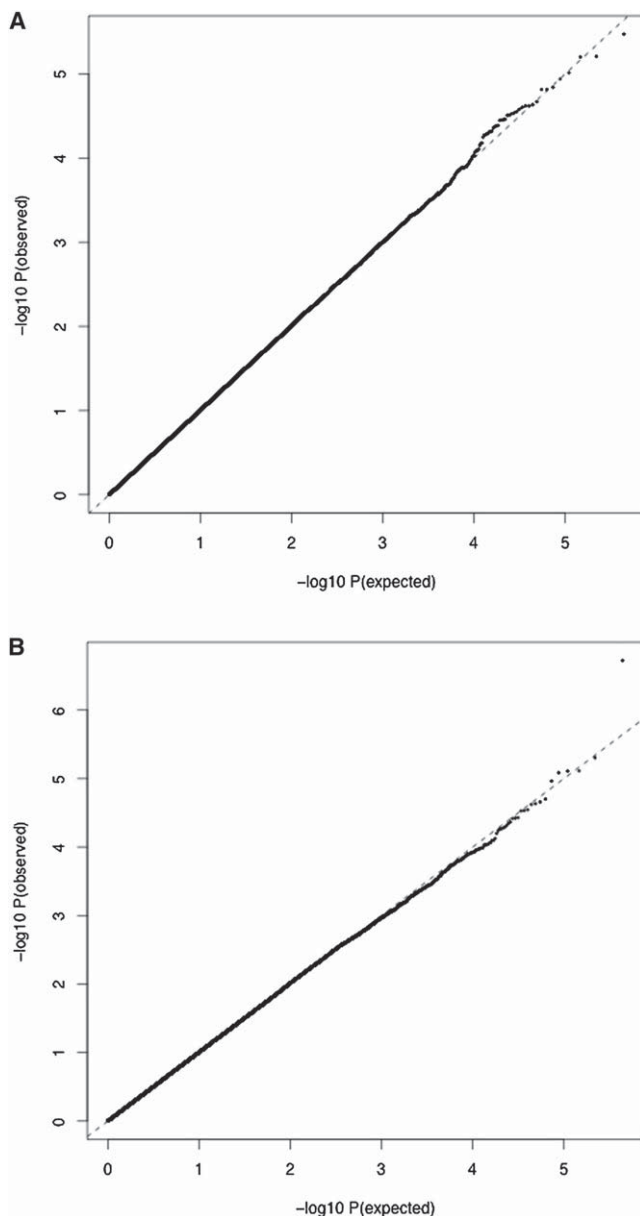
**Correction for Population Stratification**

The possible effects of population stratification on our results were explored via two methods. The Q-Q plot (corrected for  $\lambda$ ) based on all SNPs available for analysis is provided in Figure 8A. Without correction for  $\lambda$ , a slight deviation from the uncorrected expected distribution of p values under the null hypothesis was present, which was quantified in a genomic inflation factor of  $\lambda = 1.05672$ . This deviation may be due to true association or may be indicative of false positives due to population stratifica-

tion. Applying the genomic control correction method, we corrected all test statistics by the genomic inflation factor. Although all synaptic genes were no longer statistically significant as a group, this did not significantly change the results for the heterotrimeric G proteins ( $P_{EMP} = 0.00062$ ).

Second, given that the primary sample is known to include subsamples as a result of data being collected in different sites and countries, we calculated Z scores within each site and conducted all analyses again. The Z score





**Figure 8. Distribution of Association Results for All SNPs that Survived QC**

Distribution of expected (under the null hypothesis) versus observed p values for all SNPs on the Perlegen platform that survived QC in the initial sample. The dashed diagonal represents the line obtained if the observed distribution did not deviate from the expected distribution.

(A) Results of association with standardized IQ scores according to the IQ test manuals (all p values have been corrected for  $\lambda$ ).

(B) Standardized IQ scores additionally corrected for differences across collection sites.

procedure ensures that there are no mean trait differences left across subpopulations and therefore rules out spurious associations due to the known subpopulation structure. The genomic inflation factor using the Z scores was calculated as  $\lambda = 1$  (see Q-Q plot in Figure 8B). Again, the results remained significant ( $P_{EMP} = 0.0015$ ) for heterotrimeric G proteins.

On the basis of these results, we conclude that possible spurious effects due to population stratification cannot account for the detected association of the heterotrimeric G proteins and cognitive ability.

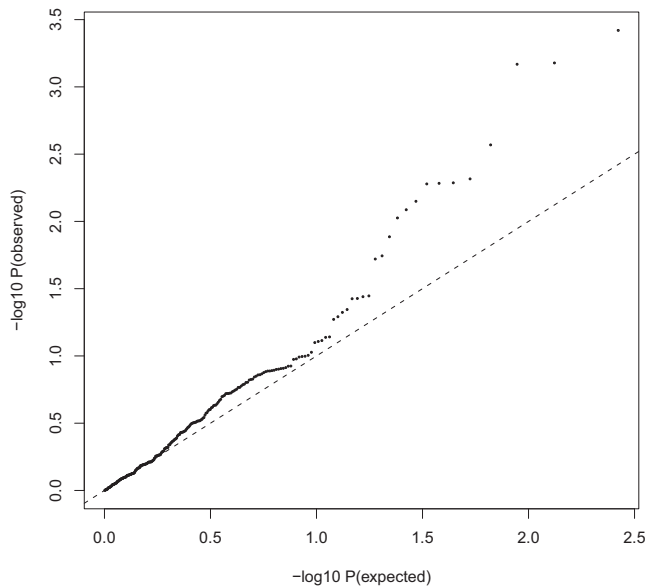
#### Validation of Significant Functional Gene Group in Independent Population-Based Sample

The Avon Longitudinal Study of Parents and Children (ALSPAC) study served as a validation sample. ALSPAC is a UK-based, population-based, prospective birth cohort with extensive data collection on health and development of children and their parents, predominantly those of white European origin, and has been described previously.<sup>31</sup> Ethical approval for the study was obtained from the ALSPAC Law and Ethics Committee and the local research ethics committees. Genotyping on 1543 individuals was initially performed with the Illumina HumanHap 300K BeadChip for 1568 blood-derived DNA samples. After QC, the clean data set comprised 1507 samples (excluding individuals with potential non-European ancestry, more than 5% missing genotype data, sex-inconsistent X-heterozygosity or a genome-wide heterozygosity of more than 36.4% or less than 34.3%).

Because this validation sample is based on a general population sample and does not consist of individuals ascertained on the basis of ADHD, any replicated effect would confirm that the observed association is related to cognitive ability in general and is not specific to individual differences in cognitive ability in an ADHD population. The validation sample ( $n = 1507$ ) had 100% power for detection of an effect size of 3.3% against a significance level of 0.05 (one test conducted).

Cognitive ability was measured in children 8 yrs of age with the Wechsler Intelligence Scale for Children (WISC-IIIK).<sup>29</sup> A short version of the test, consisting of alternate items only (with the exception of the coding task), was conducted by trained psychologists.<sup>32</sup> Verbal (information, similarities, arithmetic, vocabulary, comprehension) and performance (picture completion, coding, picture arrangement, block design, object assembly) subscales were administered, the subtests scaled and scores for total IQ derived.

Of the 27 genes in the heterotrimeric G protein relay group, four genes were not covered in the validation sample. The validation sample included 265 SNPs mapped to 23 genes (versus 359 SNPs mapped to 25 genes in the original sample) in the G protein group. The difference in available SNPs was due to the difference in platforms used in the original and validation samples. The two genes (*GNB2* and *GNG11*) that were present in the original sample but were not covered in the validation sample include the *GNG11* gene, which was one of the most significant genes of the group, with three SNPs showing a p value  $< 0.05$  (rs4262 = 0.009793; rs180236 = 0.01612; rs180241 = 0.02378). Figure 9 shows the Q-Q plot of all tested SNPs in the validation sample and suggests that all or most SNPs are important and that no single SNP drives the observed association of the G protein group.



**Figure 9. Q-Q Plot of All SNPs in the Heterotrimeric G Protein Relay Group in the Validation Sample**

Of the 46 SNPs in the G protein group that had a  $p$  value  $< 0.05$  in the original data set, 27 SNPs had a proxy SNP with an  $r^2 > 0.8$ , of which seven SNPs are identical between the two data sets and another seven SNPs have an  $r^2$  of 1. For reasons of comparison, gene coverage was determined on the basis of LD structure and genomic density and was based on the HapMap CEU LD structure. It was calculated by the sum of the typed SNPs as well as the tagged SNPs divided by the total known common SNPs within a gene (see Table 2).

In line with the recommendation by Holmans et al.,<sup>16</sup> the validation study focused on the functional gene group rather than on a direct SNP-by-SNP replication. This may result in a conservative  $p$  value in the validation sample, given that the SNPs that showed most evidence for association in the initial sample are not directly included in the validation study. However, the main goal is to validate the association with the functional gene group, not with single SNPs. This is also in line with the assumption advocated here and in Holmans et al.,<sup>16</sup> that it is more powerful to treat the functional gene group as the unit of analysis and not the single SNPs.

Analyses were conducted in PLINK<sup>28</sup> similarly to the method applied in the original sample, and 10,000 permutations were used to determine the empirical  $p$  value of the combined effect of all included SNPs. The  $\Sigma\text{-log}_{10}(P)$  of the heterotrimeric G proteins was 136, with an empirical  $p$  value of 0.047, validating our findings in an independent cohort.

## Discussion

Here, we applied an innovative functional gene group approach to cognitive ability. Functional gene groups

were defined on the basis of shared cellular function of genes, as determined by previous protein identification and data mining for synaptic genes and gene function. Initially, we identified the group of synaptic heterotrimeric G proteins to be associated with cognitive ability in a relatively small sample of 627 subjects. We replicated these findings in an independent large cohort of 1507 subjects.

Association of the group of heterotrimeric G proteins could not be attributed to a single gene or a single SNP, because none of the individual  $p$  values came close to genome-wide significance, confirming the importance of focusing on the gene group as the unit of analysis and not on single-SNP effects.

Heterotrimeric G proteins consist of three subunits ( $\alpha$ ,  $\beta$ , and  $\gamma$ ), for which a total of 33 genes are found in the human genome. Twenty-seven of these are ubiquitously expressed in the synapse. Many of these proteins are also expressed in the nonsynaptic areas of neurons and in other cells inside and outside the brain. Heterotrimeric G proteins are central relay factors between the activation of plasma membrane receptors by extracellular ligands and the cellular responses that these induce. Some signaling molecules in the brain can also activate ionic currents upon binding to ionotropic receptors, but for most signaling molecules no parallel or alternative pathways next to G protein receptor-coupled signaling exist to induce cellular responses. Therefore, heterotrimeric G proteins may be considered a point of convergence; a kind of “signaling bottleneck.” Although alterations in synaptic processes may not be the exclusive explanation for the association of heterotrimeric G proteins with cognitive ability, it is plausible that such alterations prominently affect the properties of neuronal networks in the brain in such a manner that impaired cognition and lower intelligence is observed. Synaptic processes are thought to have a central role in the “real time” processing capacities of the brain—for instance, in discrimination tasks, working memory, attention, and decision making<sup>33–35</sup>—as well as in adaptations required for “long-term synaptic modulation” in learning and memory.<sup>36</sup>

It is worth noting that the second most significant (although not below the conservative threshold of significance) functional gene group associated with cognitive ability in this study was the group of “transmitter synthesizing and metabolizing proteins,” which includes the *COMT* and monoamine oxidase A (*MAOA* [MIM 309850]) genes. Previous studies have systematically pointed to a role of metabolic enzymes in cognitive ability. In fact, the *COMT* gene has been associated with many different cognitive traits.<sup>37</sup> These results suggest that further exploration of a role of the group of transmitter synthesizing and metabolizing proteins may be indicated.

## Evaluating the Combined Effect of Multiple Genes in Functional Gene Groups

The functional gene group approach may also prove fruitful for other complex traits or common disorders that are

**Table 2. Gene Coverage Rates of the 27 Genes in the Heterotrimeric G Protein Group in the Initial and Validation Samples**

Gene	LID	HapMap (N SNPs)		Initial Sample (N SNPs)			Validation Sample (N SNPs)		
		Total	Common <sup>a</sup>	Genotyped	Tagged <sup>b</sup>	Cov. Rate	Genotyped	Tagged <sup>b</sup>	Cov. Rate
<i>GNA11</i>	2767	29	24	3	0	0.13	7	9	0.67
<i>GNA12</i>	2768	180	118	21	87	0.92	16	82	0.83
<i>GNA13</i>	10672	57	8	2	5	0.88	2	5	0.88
<i>GNA14</i>	9630	348	217	73	68	0.65	39	82	0.56
<i>GNA15</i>	2769	29	20	7	3	0.50	6	4	0.50
<i>GNAI1</i>	2770	138	71	19	23	0.59	10	24	0.48
<i>GNAI2</i>	2771	14	3	1	0	0.33	1	0	0.33
<i>GNAI3</i>	2773	61	26	4	16	0.77	5	17	0.85
<i>GNAL</i>	2774	176	87	30	37	0.77	20	48	0.78
<i>GNAO1</i>	2775	242	143	29	60	0.62	20	84	0.73
<i>GNAQ</i>	2776	369	208	26	170	0.94	20	110	0.63
<i>GNAS</i>	2778	97	45	14	8	0.49	11	16	0.60
<i>GNAT1</i>	2779	3	1	0	0	0.00	0	0	0.00
<i>GNAZ</i>	2781	83	41	5	32	0.90	7	19	0.63
<i>GNB1</i>	2782	76	27	7	11	0.67	3	8	0.41
<i>GNB2</i>	2783	1	0	1	0	-	0	0	-
<i>GNB3</i>	2784	10	4	4	2	1.50	4	2	1.50
<i>GNB4</i>	59345	62	31	3	11	0.45	5	4	0.29
<i>GNB5</i>	10681	94	41	7	28	0.85	7	25	0.78
<i>GNG10</i>	2790	6	4	2	1	0.75	3	1	1.00
<i>GNG11</i>	2791	11	7	6	1	1.00	0	0	0.00
<i>GNG12</i>	55970	201	117	20	81	0.86	16	63	0.68
<i>GNG2</i>	54331	255	114	31	55	0.75	16	60	0.67
<i>GNG3</i>	2785	0	0	0	0	0.00	0	0	0.00
<i>GNG4</i>	2786	113	70	18	22	0.57	16	34	0.71
<i>GNG5</i>	2787	9	5	1	0	0.20	2	2	0.80
<i>GNG7</i>	2788	148	97	25	25	0.52	28	25	0.55
Total		2812	1529	359	746	0.72 <sup>c</sup>	264	724	0.65 <sup>c</sup>

Abbreviations are as follows: LID, locus ID; Cov. Rate, coverage rate.

<sup>a</sup> Common SNPs, based on MAF > 0.05.

<sup>b</sup> Tagged SNPs, based on pairwise  $r^2 \geq 0.80$ .

<sup>c</sup> Average coverage rate.

potentially influenced by many genes of small effect. One of the main assumptions of current GWAS is that common diseases are caused by at least a few common genetic variants of relatively large effect (the common disease, common variant [CDCV] hypothesis). If this assumption is not met (e.g., many common alleles with small effect, genetic heterogeneity), GWAS will not work, because in different individuals different variants will account for a disease status or trait level.<sup>38</sup> However, if these different variants share a common molecular function, thus sustaining a common biological process, focusing on their combined effect will still be a valuable approach. The func-

tional gene group approach may thus be able to detect genes even when there is large genetic heterogeneity at the SNP level, as long as the many alleles of small effect share a common function in a biological process.

Unlike previous methods, the functional gene group approach does not rely on post hoc formulation of pathways or networks, but instead takes a hypothesis-driven approach by directly testing functional gene groups. This is opposed to the commonly followed strategy in which the most significant SNPs from a genome-wide association analysis are annotated to search for possible biological pathways associated with the trait. Although this may

prove a successful strategy, it will not detect any pathways or functional networks of genes in which most genes are of small and equal effect size, because it focuses initially on SNPs with the largest effects. The applied permutation procedure renders the functional gene group analysis independent of the number of SNPs per gene, of the number of genes per pathway or functional group, and of gene differences in LD structure. Functional gene group analysis may be further preferred because it circumvents the multiple-testing problem and because effect sizes are likely to be increased in comparison to single-SNP effects (because these are now a function of the combined effect, rather than of effects of single genes). This approach may therefore aid in resolving the “case of the missing heritability.”<sup>15</sup> It should be noted, however, that the effect size as estimated in the current study explained 3.3% of the observed variation in cognitive ability, which is considered large in comparison to effects of single genes for cognitive ability but is still modest in terms of estimated heritability of cognitive ability. Additional pathways or functional gene groups are therefore also likely to contribute to variation in cognitive ability.

To our knowledge, this is the first study reporting on a functional role of synaptic heterotrimeric G proteins in cognitive ability, and it thereby directs future research into the genetic basis of cognitive ability toward synaptic signaling processes. At the same time, these results underscore the notion that pathway analysis or group analysis is more informative as the unit of analysis than is single-SNP analysis, because it is directly related to biological function. The functional gene group approach adapted in this study relies on grouping genes according to similar cellular function on the basis of extensive lab experiments (whole-synapse analyses and solubilized preparations) and data mining. This functional gene group approach complements existing methods<sup>16</sup> and may also be useful in identifying etiological factors underlying complex diseases for which classic, genome-wide, SNP-by-SNP analysis has been unsuccessful so far.

### Supplemental Data

Supplemental Data include two tables and can be found with this article online at <http://www.ajhg.org>.

### Acknowledgments

The work of D.R. is supported by the Portuguese Foundation for Science and Technology under grant no. SFRH / BPD / 28725 / 2006. D.P. is supported by NWO/VIDI 016-065-318. We would like to thank the Young Academy of the Royal Dutch Academy of Arts and Science and the Simonsfonds for providing funds to D.P. and D.R., respectively, to facilitate collaboration with G.R.A. A.B.S. is supported by the Centre for Medical Systems Biology (CMSB). Statistical analyses were carried out on the Genetic Cluster Computer, which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003). The genotyping of the samples was provided through the Genetic Association Infor-

mation Network (GAIN). The data set(s) used for the analyses described in this manuscript were obtained from the dbGaP database through dbGaP accession no. phs000016.v1.p1. Samples and associated phenotype data for the Whole-Genome Association Study of Attention Deficit Hyperactivity Disorder were provided by S Faraone. We are extremely grateful to all the families who took part in this study, to the midwives for their help in recruiting them, and to the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. The UK Medical Research Council (grant ref. 74882), the Wellcome Trust (grant ref. 076467), and the University of Bristol provided core support for ALSPAC. D.P. wishes to thank Zoltan Bochdanovits and Peter Visscher for reading earlier versions of this paper and/or for valuable discussions on the applied methods. We thank Synaptologics BV (<http://www.synaptologics.com>) for providing the database of functional gene groups.

Received: September 9, 2009

Revised: November 26, 2009

Accepted: December 4, 2009

Published online: January 7, 2010

### Web Resources

The URLs for data presented herein are as follows:

dbGaP, <http://view.ncbi.nlm.nih.gov/dbgap-controlled>

dbSNP, <http://www.ncbi.nlm.nih.gov/SNP>.

GAIN QA and QC software package <http://www.sph.umich.edu/csg/abecasis/GainQC/>

Genetics Cluster Computer, <http://www.geneticcluster.org>

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/>

PLINK software, <http://pngu.mgh.harvard.edu/~purcell/plink/>

R software, <http://www.r-project.org>

### References

1. Bouchard, T.J. Jr., and McGue, M. (1981). Familial studies of intelligence: a review. *Science* 212, 1055–1059.
2. Posthuma, D., Baaré, W.F., Hulshoff Pol, H.E., Kahn, R.S., Boomsma, D.I., and De Geus, E.J. (2003). Genetic correlations between brain volumes and the WAIS-III dimensions of verbal comprehension, working memory, perceptual organization, and processing speed. *Twin Res.* 6, 131–139.
3. Posthuma, D., de Geus, E.J., and Boomsma, D.I. (2001). Perceptual speed and IQ are associated through common genetic factors. *Behav. Genet.* 31, 593–602.
4. Plomin, R., and Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychol. Bull.* 131, 592–617.
5. Deary, I.J., Johnson, W., and Houlihan, L.M. (2009). Genetic foundations of human intelligence. *Hum. Genet.* 126, 215–232.
6. McClearn, G.E., Johansson, B., Berg, S., Pedersen, N.L., Ahern, F., Petril, S.A., and Plomin, R. (1997). Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science* 276, 1560–1563.
7. Plomin, R. (1999). Genetics and general cognitive ability. *Nature* 402(6761, Suppl), C25–C29.
8. Saxena, R., Voight, B.F., Lyssenko, V., Burt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N.,



- Daly, M.J., et al; Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316, 1331–1336.
9. Rioux, J.D., Xavier, R.J., Taylor, K.D., Silverberg, M.S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M.M., Datta, L.W., et al. (2007). Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat. Genet.* 39, 596–604.
  10. Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L.T., Gudbjartsson, D., Helgason, A., Rafnar, T., Bergthorsson, J.T., Agnarsson, B.A., Baker, A., et al. (2007). Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat. Genet.* 39, 631–637.
  11. Dewan, A., Liu, M., Hartman, S., Zhang, S.S., Liu, D.T., Zhao, C., Tam, P.O., Chan, W.M., Lam, D.S., Snyder, M., et al. (2006). HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science* 314, 989–992.
  12. Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., Steinhardt, A.H., Abraham, C., Regueiro, M., Griffiths, A., et al. (2006). A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 314, 1461–1463.
  13. Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W., et al. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316, 889–894.
  14. Altshuler, D., and Daly, M. (2007). Guilt beyond a reasonable doubt. *Nat. Genet.* 39, 813–815.
  15. Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature* 456, 18–21.
  16. Holmans, P., Green, E.K., Pahwa, J.S., Ferreira, M.A., Purcell, S.M., Sklar, P., Owen, M.J., O'Donovan, M.C., Craddock, N., and Wellcome Trust Case-Control Consortium. (2009). Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. *Am. J. Hum. Genet.* 85, 13–24.
  17. Torkamani, A., Topol, E.J., and Schork, N.J. (2008). Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics* 92, 265–272.
  18. Brookes, K., Xu, X., Chen, W., Zhou, K., Neale, B., Lowe, N., Anney, R., Aneey, R., Franke, B., Gill, M., et al. (2006). The analysis of 51 genes in DSM-IV combined type attention deficit hyperactivity disorder: association signals in DRD4, DAT1 and 16 other genes. *Mol. Psychiatry* 11, 934–953.
  19. Kuntsi, J., Neale, B.M., Chen, W., Faraone, S.V., and Asherson, P. (2006). The IMAGE project: methodological issues for the molecular genetic analysis of ADHD. *Behav. Brain Funct.* 2, 27.
  20. Neale, B.M., Sham, P.C., Purcell, S., Banaschewski, T., Buitelaar, J., Franke, B., Sonuga-Barke, E., Ebstein, R., Eisenberg, J., Mulligan, A., et al. (2008). Population differences in the International Multi-Centre ADHD Gene Project. *Genet. Epidemiol.* 32, 98–107.
  21. Wechsler, D. (1997). Wechsler Adult Intelligence Scale (San Antonio: The Psychological Corporation).
  22. Sattler, J.M. (1992). Assessment of Children (San Diego: Jerome M. Sattler Publisher).
  23. Neale, B.M., Lasky-Su, J., Anney, R., Franke, B., Zhou, K., Maller, J.B., Vasquez, A.A., Asherson, P., Chen, W., Banaschewski, T., et al. (2008). Genome-wide association scan of attention deficit hyperactivity disorder. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 147B, 1337–1344.
  24. Li, K.W., Miller, S., Klychnikov, O., Loos, M., Stahl-Zeng, J., Spijker, S., Mayford, M., and Smit, A.B. (2007). Quantitative proteomics and protein network analysis of hippocampal synapses of CaMKIIalpha mutant mice. *J. Proteome Res.* 6, 3127–3133.
  25. Li, K., Hornshaw, M.P., van Minnen, J., Smalla, K.H., Gundelfinger, E.D., and Smit, A.B. (2005). Organelle proteomics of rat synaptic proteins: correlation-profiling by isotope-coded affinity tagging in conjunction with liquid chromatography-tandem mass spectrometry to reveal post-synaptic density specific proteins. *J. Proteome Res.* 4, 725–733.
  26. Emes, R.D., Pocklington, A.J., Anderson, C.N., Bayes, A., Collins, M.O., Vickers, C.A., Croning, M.D., Malik, B.R., Choudhary, J.S., Armstrong, J.D., and Grant, S.G. (2008). Evolutionary expansion and anatomical specialization of synapse proteome complexity. *Nat. Neurosci.* 11, 799–806.
  27. Plumptre, M., and Barnes, M.R. (2007). Predictive functional analysis of polymorphisms: An overview. In *Bioinformatics for Geneticists: A Bioinformatics Primers for the Analysis of Genetic Data*, M.R. Barnes, ed. (Chichester, West Sussex, UK: John Wiley & Sons Ltd), pp. 249–280.
  28. 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., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.
  29. Wechsler, D., Golombok, J., and Rust, J. (1992). WISC-IIIUK: Wechsler Intelligence Scale for Children (Sidcup, UK: The Psychological Corporation).
  30. Purcell, S., Cherny, S.S., and Sham, P.C. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19, 149–150.
  31. Golding, J., Pembrey, M., Jones, R., and ALSPAC Study Team. (2001). ALSPAC—the Avon Longitudinal Study of Parents and Children. I. Study methodology. *Paediatr. Perinat. Epidemiol.* 15, 74–87.
  32. Joinson, C., Heron, J., Butler, R., Von Gontard, A., Butler, U., Emond, A., and Golding, J. (2007). A United Kingdom population-based study of intellectual capacities in children with and without soiling, daytime wetting, and bed-wetting. *Pediatrics* 120, e308–e316.
  33. de Jong, A.P., and Verhage, M. (2009). Presynaptic signal transduction pathways that modulate synaptic transmission. *Curr. Opin. Neurobiol.* 19, 245–253.
  34. Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. *Science* 319, 1543–1546.
  35. Verhage, M., and van Meer, G. (2009). Synaptobrevin, sphingolipids, and secretion: lube ‘n’ go at the synapse. *Neuron* 62, 603–605.
  36. Südhof, T.C., and Malenka, R.C. (2008). Understanding synapses: past, present, and future. *Neuron* 60, 469–476.
  37. Dickinson, D., and Elvevåg, B. (2009). Genes, cognition and brain through a COMT lens. *Neuroscience* 164, 72–87.
  38. Altshuler, D., Daly, M.J., and Lander, E.S. (2008). Genetic mapping in human disease. *Science* 322, 881–888.