Protein Cell 2012, 3(7): 526–534
DOI 10.1007/s13238-012-2931-7

Protein & Cell

RESEARCH ARTICLE

Prioritization of candidate genes for attention deficit hyperactivity disorder by computational analysis of multiple data sources

Suhua Chang[1,2], Weina Zhang[1,2], Lei Gao[1,2], Jing Wang[1] ✉

[1] Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
[2] Graduate University of the Chinese Academy of Sciences, Beijing 100039, China
✉ Correspondence: wangjing@psych.ac.cn
Received April 5, 2012 Accepted May 15, 2012

ABSTRACT

Attention deficit hyperactivity disorder (ADHD) is a common, highly heritable psychiatric disorder characterized by hyperactivity, inattention and increased impulsivity. In recent years, a large number of genetic studies for ADHD have been published and related genetic data has been accumulated dramatically. To provide researchers a comprehensive ADHD genetic resource, we previously developed the first genetic database for ADHD (ADHDgene). The abundant genetic data provides novel candidates for further study. Meanwhile, it also brings new challenge for selecting promising candidate genes for replication and verification research. In this study, we surveyed the computational tools for candidate gene prioritization and selected five tools, which integrate multiple data sources for gene prioritization, to prioritize ADHD candidate genes in ADHDgene. The prioritization analysis resulted in 16 prioritized candidate genes, which are mainly involved in several major neurotransmitter systems or in nervous system development pathways. Among these genes, nervous system development related genes, especially *SNAP25*, *STX1A* and the gene-gene interactions related with each of them deserve further investigations. Our results may provide new insight for further verification study and facilitate the exploration of pathogenesis mechanism of ADHD.

KEYWORDS gene prioritization, attention deficit hyperactivity disorder, candidate genes, multiple data sources

INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is a common psychiatric disorder among children and adolescents with a prevalence of about 3%–7% (Faraone and Doyle, 2001). The core clinical features of ADHD are lack of concentration, motor hyperactivity and severe impulsiveness, and it is often accompanied with other psychiatric disorders (Biederman et al., 1991), which affect the education and social relationship of children. Twin and adoption studies of ADHD suggest that genetic influences contribute substantially to its etiology, with heritability estimates ranging from 75% to 91% (Faraone et al., 2005). Thus, unraveling the genetic basis of ADHD is of fundamental importance in uncovering disease mechanism and potentially leads to novel diagnostic procedures and pharmacological treatment. Up to now, large numbers of association studies (Gizer et al., 2009), linkage studies (Zhou et al., 2008) and meta-analyses (Neale et al., 2010) have been conducted to study the genetic mechanism of ADHD. To address the genetic complexity of ADHD and the heterogeneity of studies, we previously established the first genetic database for ADHD, named ADHDgene, to provide researchers a comprehensive and well-organized collection of genetic data of ADHD (Zhang et al., 2012). Finally, ADHDgene provided 3589 ADHD candidate genes, in which, 213 genes were literature-origin and 3376 were from extended functional analysis, including linkage disequilibrium (LD) analysis, pathway-based analysis (PBA) of ADHD genome-wide association study (GWAS) data (Zhang et al., 2010, 2011a) and gene mapping.

The abundant genetic data provides plentiful novel can-

didates for ADHD genetic study, but inconsistence and low replication rate exist for many genetic factors. Further candidate gene association studies are needed in larger samples for replication. However, geneticists are always confused with how to choose promising genes for replication from enormous amounts of candidate genes. Meanwhile, among these genetic factors, which are the causal variants and how these causal variants contribute to the pathogenesis of disease are still incompletely understood. Experimental research for molecular function of specific genes would be helpful to understand the etiology and pathophysiology of ADHD. However, it is infeasible for molecular biologists to validate each candidate. They would need to choose more reasonable genes before experiments. For this purpose, selection of the most promising candidates from large numbers of genetic factors will greatly reduce the time and cost. Therefore, the bioinformatics community has introduced the concept of gene prioritization to take advantage of both the progress made in computational biology and the large amount of genomic data publicly available.

A basic assumption used by most of the strategies for gene prioritization is the "guilt-by-association" principle (Tranchevent et al., 2010), which means the genes that are closer to disease genes in a network or other kinds of genomic data will be more likely to be associated with the same disease. Based on this principle, the inputs of these tools normally include two types of data. One is the prior knowledge about the genetic disorder of interest, which can be either a set of genes known to play a role in the disease or a set of keywords that describe the disease. The other is the candidate search space. The output normally is either a ranking list of the candidate genes or a selection of the most promising candidates. To evaluate how the candidate genes are "close" to the training genes, a set of characteristics that are most likely to fit the underlying disease genes (training genes) need to be defined from enormous amounts of information and then are used to score candidates. So, both high coverage and high quality data sources are important to make accurate predictions. A single data type might not be powerful enough to predict the disease-causing genes accurately while the use of several complementary data sources allow much more accurate predictions (Tranchevent et al., 2010). Until now, most of the approaches make use of a wide range of data sources covering distinct views of the genes, including functional annotations (e.g. Gene Ontology (GO) (Ashburner et al., 2000)), expression (e.g. Atlas gene expression from Ensembl (Flicek et al., 2011)), regulatory information (e.g. TOUCAN (Aerts et al., 2005)), text (co-citation and functional mining), protein-protein interaction networks (e.g. HPRD (Keshava Prasad et al., 2009)), pathways (e.g. KEGG (Kanehisa et al., 2010)), sequence similarity (BLAST (Altschul et al., 1990)), phenotype (e.g. OMIM (McKusick, 2007)), protein domain conservation (e.g. InterPro (Mulder and Apweiler, 2008)) and chemical components. Besides high coverage and high quality data sources, another important issue for gene prioritization is the computational method to define the similarity between training genes and candidate genes. For different data sources, the similarity measures might be different (Kohler et al., 2008; Chen et al., 2009b) and finally, the scores from different data sources need to be integrated into one measure (Aerts et al., 2006; Chen et al., 2011a). Because every tool uses separate benchmark, it is difficult to compare the performance of different tools. By now, there have been several application studies of gene prioritization tools in diseases (e.g. on type 2 diabetes (Tiffin et al., 2006; Elbers et al., 2007; Teber et al., 2009)). Several gene prioritization tools were used in parallel in these studies to strengthen the results based on the hypothesis that genes selected by the most independent methods are at least likely to be false positives of the approach used (Elbers et al., 2007).

In this study, in order to pick out promising candidate genes for further verification study from the large numbers of candidates provided in ADHDgene, we combined five web interface available tools, which all integrate multiple data sources and use different methods for similarity calculation, to prioritize ADHD candidate genes in ADHDgene. Our results may provide new insight for further replication study and facilitate the exploration of pathogenesis mechanism of ADHD.

RESULTS

Prioritized genes

After integrating the prioritization results from five tools, 16 candidate genes were detected as promising ADHD genes by at least 3 out of 5 methods (a summary of these candidate genes is shown in Table 1). There are 4 of the 16 genes supported by four tools and 12 of the 16 genes supported by three tools. Based on the hypothesis about the genetic influences of major neurotransmitter systems on ADHD (Sharp et al., 2009), 12 of the 16 genes are involved in dopaminergic, serotonergic, noradrenergic or nicotinic neurotransmitter systems. Among these genes, *SLC6A4*, *DBH*, *HTR1B* and *HTR2A* have been examined in many studies, but there are some genes (e.g. *HTR3A*, *ADRB2* and *CHRNA7*) which have not been examined much by now and may be promising candidate for further study.

Besides the genes involved in the major neurotransmitter systems, the most interesting predictive results are the genes about nervous system development pathways, including *SYT1*, *STX1A*, *VAMP2* and *SNAP23*. Although these genes were not investigated too much in previous studies, there is even no genetic study on *SNAP23* which is mapped by PBA pathways in ADHDgene, they were still pinpointed by prioritization tools and *SYT1* was even selected by four tools together. Fig. 1 illustrates the training gene sub-network produced by ToppNet (Chen et al., 2009b) which was created by

Table 1 List of 16 genes selected by the five gene prioritization tools

Pathways/ systems	Gene HUGO ID	Gene name	Supported by tools	No. of studies (significant/ trend/non-significant/)[a]
Dopaminergic neurotransmitter system	DBH	Dopamine beta-hydroxylase (dopamine beta-monooxygenase)	DIR, Endeavour, ToppGene	16 (9/0/7)
	DRD2	Dopamine receptor D2	Endeavour, ToppGene, ToppNet, TargetMine	9 (3/0/6)
	DRD3	Dopamine receptor D3	DIR, Endeavour, ToppGene, ToppNet	8 (1/0/7)
Serotonergic neurotransmitter system	SLC6A4	Solute carrier family 6 (neurotransmitter transporter, serotonin), member 4	Endeavour, ToppGene, TargetMine	24 (13/0/11)
	HTR1B	5-hydroxytryptamine (serotonin) receptor 1B	Endeavour, ToppGene, TargetMine	12 (5/0/7)
	HTR2A	5-hydroxytryptamine (serotonin) receptor 2A	Endeavour, ToppGene, TargetMine	11 (5/0/6)
	TPH1	Tryptophan hydroxylase 1	DIR, Endeavour, ToppGene, TargetMine	6 (2/0/4)
	MAOB	Monoamine oxidase B	DIR, ToppGene, TargetMine	6 (2/0/4)
	TH	Tyrosine hydroxylase	DIR, Endeavour, ToppGene	6 (1/0/5)
	HTR3A	5-hydroxytryptamine (serotonin) receptor 3A	Endeavour, ToppGene, TargetMine	1 (0/0/1)
Noradrenergic neurotransmitter system	ADRB2	Adrenergic, beta-2-, receptor, surface	Endeavour, ToppNet, TargetMine	1 (1/0/0)
Nicotinic neurotransmitter system	CHRNA7	Cholinergic receptor, nicotinic, alpha 7	Endeavour, ToppGene, TargetMine	1 (1/0/0)
Nervous system development pathways	SNAP23	Synaptosomal-associated protein, 23 kDa	Endeavour, ToppGene, ToppNet	0 (0/0/0)
	SYT1	Synaptotagmin I	DIR, Endeavour, ToppGene, ToppNet	3 (1/0/2)
	STX1A	Syntaxin 1A (brain)	Endeavour, ToppGene, ToppNet	2 (1/0/1)
	VAMP2	Vesicle-associated membrane protein 2 (synaptobrevin 2)	Endeavour, ToppGene, ToppNet	2 (0/0/2)

[a] The numbers are from ADHD gene.

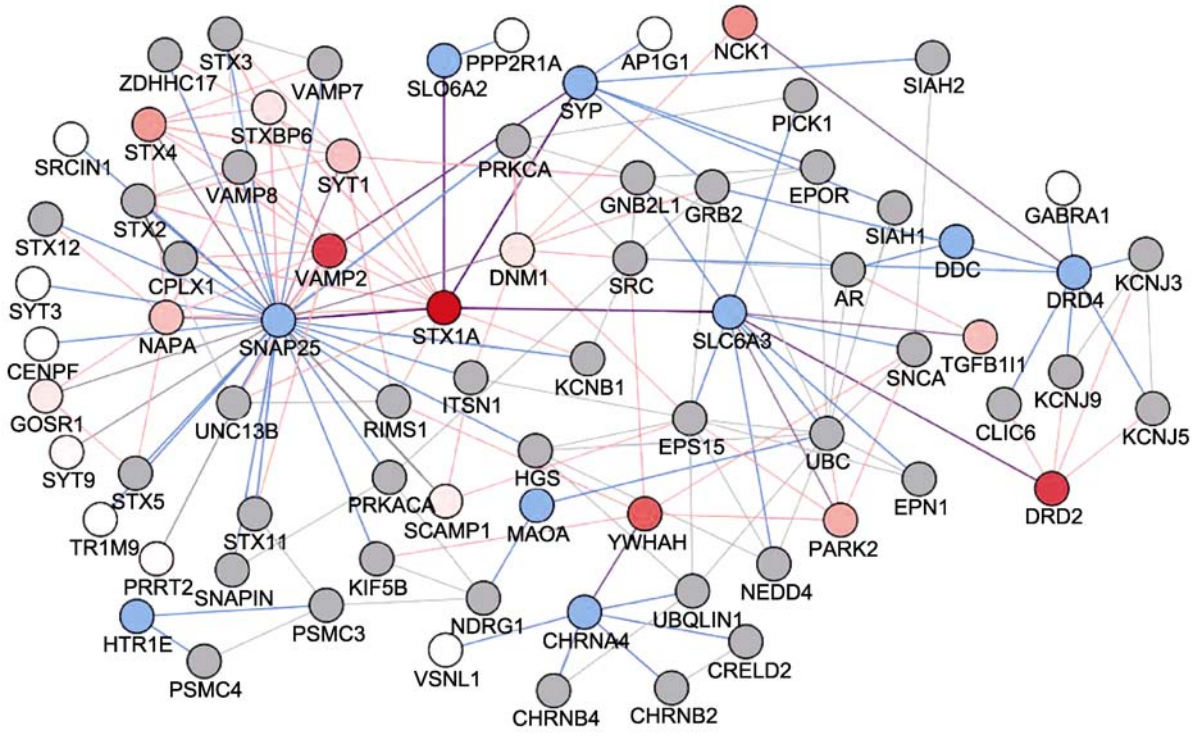

Figure 1. Training gene sub-network produced by ToppNet. Cytoscape (Smoot et al., 2011) is used to rearrange the layout of the network. Node legend: blue circle: training gene, red circle: test gene (the darker denotes the node connect with more training genes), gray circle: k-Order connected (connected with other nodes besides the direct training gene), white circle: k-Order non-connected (only connected with one training gene, no other connected genes). Edge legend: purple line: connection between training gene and test gene, blue line: connection between training gene and non-test genes, pink line: connection between test genes, gray: others.

using training genes and their immediate interactants. The network shows *STX1A* is strongly connected with four training genes (*SNAP25*, *SLC6A2*, *SLC6A3* and *SYP*), and *SNAP25* interacts with several important genes, including *SYT1*, *VAMP2* and *STX1A* in our final results. *SNAP25* codes for a protein involved in axonal growth and synaptic plasticity, as well as in the docking and fusion of synaptic vesicles in presynaptic neurons necessary for the regulation of neurotransmitter release (Sollner et al., 1993). *STX1A* encodes a member of the superfamily of syntaxins, which are nervous system-specific proteins implicated in the docking of synaptic vesicles with the presynaptic plasma membrane. Syntaxin and *SNAP25*, together with vesicle-associated membrane protein termed synaptobrevin/VAMP form a complex which serves as a binding site for pre- and post-synaptic exocytosis (Kennedy and Ehlers, 2011). Synaptobrevin/VAMP and syntaxin are believed to be involved in vesicular transport in most cells, while *SNAP25* is present almost exclusively in the brain (Liu et al., 2011). These results showed these nervous system development related genes and gene-gene interactions among them form an important network, which might contribute to the development of ADHD in a combination pattern and deserve more attention in the future studies.

Result evaluation

To evaluate the influence of the training genes on the final ranking result, we used another training data set (see detailed gene list in MATERIALS AND METHODS part) to execute the same gene prioritization procedure as our primary training data set. The comparison of these two ranking results showed 6 genes of the 16 candidate genes from our primary analysis were replicated in the second ranking results including *DRD2*, *DRD3*, *SYT1*, *ADRB2*, *STX1A* and *TH*, 4 genes of the 16 candidate genes are the training genes in the second ranking procedure including *DBH*, *HTR1B*, *HTR2A* and *SLC6A4*. Meanwhile, the most interesting results *STX1A* and *SYT1* were both pinpointed by four tools in the second ranking results.

The choice of data sources for gene prioritization tools might affect the final ranking result. In our primary analysis, we used all data sources provided by tools. We have also tried to remove several computational data sources to compare the results. When we removed data sources like 'GO', 'BLAST' and 'Interaction-String' in Endeavour and features 'GO' and 'computational' in ToppGene, we got 15 prioritized genes. Comparison with the original 16 prioritized genes showed that only gene *HTR3A*, *CHRNA7*, *SNAP23* and *ADRB2* were missed and *HTR5A*, *HTR1A* and *TGFB1I1*were gained, which shows the computational data sources will only bring minor changes on the results and will not affect the most interesting results such as *STX1A* and *SYT1*.

The 16 prioritized genes together with 13 training genes (29 genes in total) could be regarded as reliable ADHD candidate genes. Recently, Kang et al. (2011) published a spatio-temporal transcriptome of the human brain. We used their dataset to check the expression profiling of the 29 ADHD genes. The result shows five genes are region-enriched differentially expressed (Fig. 2), in which *DRD2*, *DRD3* and *HTR1B* are enriched in striatum, but *DRD2* and *DRD3* are differentially expressed in earlier development periods than *HTR1B*; *CHRNA4* and *TPH2* are enriched in mediodorsal nucleus of the thalamus from late fetal to late adulthood or adolescence. It has been reported that methylphenidate-elicited dopamine increases in ventral striatum are associated with long-term symptom improvement in adults with attention deficit hyperactivity disorder (Volkow et al., 2012). Smaller regional volume on the lateral thalamic surface was found in ADHD patient with hyperactivity and larger regional volumes on the medial thalamic surface was found in ADHD patient with inattention. These suggest the involvement of thalamic subcircuits is different in the pathogenesis of different ADHD symptoms (Ivanov et al., 2010) and regulation of gene expression in these regions may contribute to ADHD.

DISCUSSION

Prioritized genes in major neurotransmitter systems

The main drug for ADHD treatment methylphenidate acts by

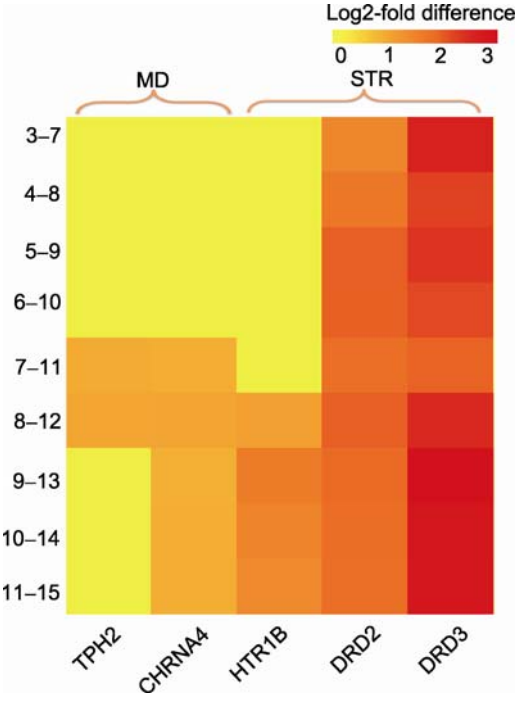

Figure 2. Heat map matrix to show the differential expression of five region-enriched genes at different periods of 1–15. MD: mediodorsal nucleus of the thalamus, STR: striatum. 1–15 denote periods of human development and adulthood as defined in (Kang et al., 2011).

blocking the dopamine transporter, which together with the association of ADHD with those executive neuropsychological functions and frontostriatal dopaminergic pathways make the dopaminergic system become the most intensively analyzed neurotransmitter system in ADHD (Swanson et al., 2007). *DRD2*, *DRD3* and *DBH* in the 16 candidate genes as well as *DRD4* and *SLC6A3* in the training genes are all dopaminergic system related genes. Their associations with ADHD have been investigated in many studies (Gizer et al., 2009), especially, *DRD4* and *SLC5A3* are the top two genes with the most studies in ADHDgene (Zhang et al., 2012). The selections of *DRD2* and *DRD3* by prioritization tools are mainly because of their similar function with *DRD4* although the studies about them are less and the results from different studies are inconsistent. Dopamine beta hydroxylase (DBH) converts dopamine to norepinephrine, and also represents an interesting candidate gene for ADHD. A meta-analysis (Gizer et al., 2009) shows that *Taq*I polymorphism in *DBH* shows a trend association with ADHD. The meta-analysis results also indicate there is significant heterogeneity of effect sizes across studies for *DRD2* and *DBH*; further studies are needed to determine their effect during the development of ADHD.

Serotonin dysregulation has been related to impulsive behavior in children, and thus genes involved in serotonergic system have been hypothesized to play a causal role in ADHD (Oades, 2008). Solute carrier family 6 (neurotransmitter transporter, serotonin) member 4 (*SLC6A4*) is the most tested serotonergic system related gene in ADHDgene (Zhang et al., 2012). Several 5-hydroxytryptamine (serotonin) receptors (*HTR1E*, *HTR1B*, *HTR2A* and *HTR3A*) were also good candidates associated with ADHD. *HTR3A* has only one negative result by now in ADHDgene (Ribases et al., 2009), but it is selected by these tools because of its function similarity with other serotonin receptors. Tryptophan hydroxylase is the rate-limiting enzyme in the production of serotonin. Because recent studies showed *TPH2* is responsible for tryptophan hydroxylase expression in the brain (Walther et al., 2003), the association studies of ADHD have focused on *TPH2* rather than *TPH1* (Gizer et al., 2009; Sharp et al., 2009). The monoamine oxidase genes encode enzymes involved in the metabolism of dopamine, serotonin, and norepinephrine. Two monoamine oxidase genes, *MAOA* and *MAOB*, are interesting candidates for association with ADHD. By now, more studies were conducted for *MAOA* than *MAOB* because of some specific supports for *MAOA* (Gizer et al., 2009), but the prioritization tools also detected *MAOB* from the training gene *MAOA*.

Adrenergic neurotransmitters are hypothesized to influence attentional processes and certain aspects of executive control (Arnsten, 2006). The most frequently investigated genes of the noradrenergic system are those encoding the noradrenaline transporter (*SLC6A2*) and the adrenergic receptors (including *ADRA2A*, *ADRA2C*, *ADRA1A*, *ADRA2B*, *ADRB1* and *ADRB2*). In our predicted results, *ADRB2* was selected while *SLC6A2* as training gene. The association between *ADRB2* and ADHD was first tested in a sample of 776 DSM-IV ADHD combined type cases (Brookes et al., 2006), which is the main source of our training genes. One interesting thing is that this gene is excluded from our training genes because the number of its significant result is less than two, but finally it is selected by prediction tools into our candidate genes. So, further attention on this gene may have some unexpected result.

Another interesting neurotransmitter system with suggested association with ADHD is the nicotinic neurotransmitter system. One of the evidences is that nicotine administration has been shown to improve attention and working memory deficits in adults diagnosed with ADHD (Levin, 2002). The nicotinic acetylcholine receptor alpha 4 subunit gene (*CHRNA4*) received the most attention thus far and it is also in our training data set. Another member of nicotinic acetylcholine receptors superfamily *CHRNA7*, homolog of *CHRNA4*, is in our prioritized results. Although it was only examined in one study and the result was non-significant, it may be also an interesting candidate because of its similarity with *CHRNA4*.

Prioritized genes in nervous system development pathways

The most studied nervous system development related gene is *SNAP25* (synaptosomal-associated protein, 25 kDa). It has been suggested as a genetic susceptibility factor in ADHD in several studies (Gizer et al., 2009). *STX1A* (syntaxin 1A (brain)) has been suggested to be a genetic susceptibility factor for autism (Nakamura et al., 2011). However, until now, the genetic studies about *SYT1*, *STX1A*, *VAMP2* and *SNAP23* in ADHD are still scarce. *SYT1* got evidences from four prioritization tools to be promising candidate for ADHD, but was only studied in three researches with one significant result in Chinese Han population (Guan et al., 2009). *STX1A* also only got one significant result in a United Kingdom population (Brookes et al., 2005), while, there is no significant result for *VAMP2* thus far. *SNAP23*, a homolog of *SNAP25*, is structurally and functionally similar to *SNAP25* and binds tightly to multiple syntaxins and synaptobrevins/VAMPs (Frank et al., 2011). It is also selected by three tools as one of our final candidate genes although by now there is no genetic study about it in ADHD. To sum up, nervous system development pathway related genes, especially *STX1A*, *SYT1*, *SNAP25* and their related interactions, deserve more investigation to help the exploration of the pathogenesis mechanism of ADHD.

CONCLUSION

The first genetic database of ADHD we developed at previous study (Zhang et al., 2012) provides abundant and novel

candidates for ADHD genetic study, and meanwhile, brings new challenge for the selection of promising candidates for verification study. In this study, we used five gene prioritization tools based on multiple data sources and different methods for similarity calculation to prioritize all candidate genes in ADHDgene. Relatively credible genes from IMAGE team with one extra criterion were chosen as training genes and genes predicted by at least three tools were selected as promising candidate genes. Our analysis results provide 16 promising candidate genes for future replication and verification studies. Besides the genes involved in major neurotransmitter systems, the genes related with nervous systems development pathways may deserve more investigation in future studies, especially the gene-gene interactions related with *STX1A* and *SNAP25*. The results may provide new insights for the pathogenesis mechanism research of ADHD.

MATERIALS AND METHODS

Training data and test data

For the gene prioritization tools based on "guilt-by-association" principle, the training data set is important to get credible result. By now, there are no widely accepted ADHD-related genes as the core genes used in (Sun et al., 2009) for schizophrenia. The International Multicenter ADHD Genetics project (IMAGE) is an international collaborative study that aims to identify genes that increase the risk for ADHD using QTL linkage and association strategies. IMAGE investigators had selected 51 pre-specified ADHD autosomal candidate genes and used this gene set for candidate association study in a sample of 776 DSM-IV ADHD combined type cases ascertained for IMAGE (Brookes et al., 2006). Finally, they found nominal significance with one or more SNPs in 18 genes. These genes were also analyzed in one ADHD GWAS study (Lasky-Su et al., 2008) and a meta-analysis of GWASs of ADHD (Neale et al., 2010) to examine the distribution of the association *P*-values of the SNPs in these genes. Among these 18 genes, 13 are replicated in at least two studies as significant results according to ADHDgene. Thus, we choose the 13 genes as training data set, they are as the following: *CHRNA4*, *DDC*, *DRD4*, *FADS2*, *HTR1E*, *MAOA*, *PNMT*, *SLC6A2*, *SLC6A3*, *SLC9A9*, *SNAP25*, *SYP* and *TPH2*. To evaluate the influence of the training data on the final ranking result, another training data set was used for result comparison, which includes genes in ADHDgene with at least ten study results and five significant results: *DRD4*, *SLC6A3*, *SLC6A4*, *COMT*, *DRD5*, *SNAP25*, *MAOA*, *DBH*, *BDNF*, *SLC6A2*, *HTR1B*, *HTR2A* and *TPH2*. The test data set included 3576 genes in ADHDgene except those genes in training data set.

Selection of gene prioritization tools

Because our training data and test data are both candidate gene lists, the tools using keywords or expression dataset as training data and the tools using region or genome as candidate genes were excluded. According to the inputs/outputs statistics in "Gene Prioritization Portal" website (Tranchevent et al., 2010), only seven tools satisfy the above criterion: DIR (Chen et al., 2011a), Endeavour (Aerts et al., 2006; Tranchevent et al., 2008), GeneWanderer (Kohler et al., 2008), Prioritizer (Franke et al., 2006), TargetMine (Chen et al., 2011b), ToppGene and ToppNet (Chen et al., 2009b). When we further checked the websites of these tools, we found GeneWanderer only accepts genomic region as candidates, not gene list; Prioritizer has been moved from a living web application to a program to download and the candidates for the local version software can also only be genomic regions. In addition, the data sets in GeneWanderer and Prioritizer have not been updated for relatively long time. So finally, we chose the other five tools to prioritize ADHD candidate genes.

Basic information about these gene prioritization tools is summarized in Table 2. All of these tools are based on multiple data sources, but the data sources and the computational methods for similarity measure might be different among tools. DIR (Chen et al., 2011a) uses an expandable framework for gene prioritization that can integrate multiple heterogeneous data sources by taking advantage of a unified graphic representation. Endeavour (Aerts et al., 2006) is the first gene prioritization tool using multiple data sources. It was maintained and updated continually (Tranchevent et al., 2008) and has been used in several applications (Aerts et al., 2009; Thienpont et al., 2010). TargetMine (Chen et al., 2011b) is a powerful database for gene enrichment analysis, and the "unsupervised" protocol for prioritization is one of its applications, in which, input candidate genes are enriched to several biological terms (e.g. KEGG pathways, GO terms) and prioritized genes are collected from the significantly enriched biological terms. ToppGene and ToppNet are in one suite of tools for gene functional enrichment and prioritization (Chen et al., 2009b). ToppGene is for candidate gene prioritization based on functional similarity to training gene list (Chen et al., 2007). ToppNet is for ranking candidate genes based on topological features in protein-protein interaction network (Chen et al., 2009a). Both ToppGene and ToppNet were used in our analyses and are listed separately in Table 2 as two different tools. The inputs of DIR, Endeavour, ToppGene and ToppNet are the same, including a list of training genes and a list of candidate genes. The input of TargetMine is only a list of candidate genes, no training data is required. Except the input gene lists, default parameters and all data sources for these tools were used. The outputs of DIR, Endeavour, ToppGene and ToppNet are all ranked candidate genes with scores. Whereas the outputs of ToppNet also include a graphical network and Cytoscape-compatible input file to demonstrate the training gene sub-network. The outputs of TargetMine are candidate gene enriched biological terms, including but not limited to KEGG pathways, GO terms and OMIN phenotypes. Prioritized genes were collected from the top biological terms.

Identification of candidate genes

Because TargetMine pinpoints candidate gene enriched biological terms, the top seven biological terms for KEGG pathways, GO terms and disease ontology terms respectively were collected (according to the original paper, the threshold of top seven terms was by and large most suited to ensuring maximum coverage and minimum overprediction (Chen et al., 2011b)). The intersections among genes from these three types of biological terms were regarded as the results from TargetMine. For the results from DIR, Endeavour, ToppGene and ToppNet which all produce rankings, the top 30 genes from each method were included for comparison. A gene was considered to be

Table 2 Basic information of the gene prioritization tools

Tool/website	Data sources used	Method description
DIR (Chen et al., 2011a) http://cbc.case.edu/dir/Default.aspx	PPI (HyNet, Reactome, BIND, MINT, HPRD), expression, pathway (KEGG), OMIM	Gene-gene relationships and gene-disease relationships are defined based on the overall topology of each network using a diffusion kernel measure. These relationship measures are in turn normalized to derive an overall measure across all networks, which is utilized to rank all candidate genes.
Endeavour (Tranchevent et al., 2008) http://homes.esat.kuleuven.be/~bioiuser/endeavour/tool/endeavourweb.php	Annotation (Ensembl EST, GO, InterPro, KEGG, UniProtKB/Swiss-Prot), BLAST, cis-regulatory information, expression, interaction (BIND, BioGRID, HPRD, IntNetDB, IntAct, MINT, STRING), motif, precalculated data, text	One ranking score is calculated for each data source and global ranking is obtained by fusion of the rankings per data source.
TargetMine (Chen et al., 2011b) http://targetmine.nibio.go.jp:8080/targetmine/begin.do	Entrez Gene, UniProtKB, InterPro, KEGG, BioGRID, GO, UniProtKB GOA, PDBs SIFTS, PPIview, PIP, SCOP, KEGG Orthology, OregAnno, AMADEUS, The ENZYME database, DrugBank, OMIN, Disease Ontology	"Unsupervised" protocol for prioritization, no training data is needed. Prioritized genes are collected from the significantly enriched biological associations.
ToppGene (Chen et al., 2009b) http://toppgene.cchmc.org/prioritization.jsp	GO: molecular function, GO: biological process, mouse phenotype, pathways, protein interactions, protein domains, transcription factor-binding sites, miRNA-target genes, disease-gene associations, drug-gene interactions, and Gene Expression	A fuzzy-based similarity measure was used to compute the similarity between any two genes based on semantic annotations. Similarity scores from different data sources are combined into an overall score using statistical meta-analysis.
ToppNet (Chen et al., 2009b) http://toppgene.cchmc.org/	Protein-protein interaction network (BIND, BioGRID, HPRD)	Prioritize or rank candidate genes based on topological features in protein–protein interaction network using PageRank with Priors, HITS with Priors and K-step Markov algorithms.

interesting as a candidate gene if it was indicated by at least three or more of these tools.

ACKNOWLEDGEMENTS

This work was supported by Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences.

ABBREVIATIONS

ADHD, attention deficit hyperactivity disorder; DBH, dopamine beta hydroxylase; GO, Gene Ontology; GWAS, genome-wide association study; LD, linkage disequilibrium; PBA, pathway-based analysis

REFERENCES

Aerts, S., Lambrechts, D., Maity, S., Van Loo, P., Coessens, B., De Smet, F., Tranchevent, L.-C., De Moor, B., Marynen, P., Hassan, B., et al. (2006). Gene prioritization through genomic data fusion. Nat Biotech 24, 537–544.

Aerts, S., Van Loo, P., Thijs, G., Mayer, H., de Martin, R., Moreau, Y., and De Moor, B. (2005). TOUCAN 2: the all-inclusive open source workbench for regulatory sequence analysis. Nucleic Acids Res 33, W393–396.

Aerts, S., Vilain, S., Hu, S., Tranchevent, L.C., Barriot, R., Yan, J., Moreau, Y., Hassan, B.A., and Quan, X.J. (2009). Integrating computational biology and forward genetics in Drosophila. PLos Genet 5, e1000351.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol 215, 403–410.

Arnsten, A.F. (2006). Fundamentals of attention-deficit/hyperactivity disorder: circuits and pathways. J Clin Psychiatry 67 Suppl 8, 7–12.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25, 25–29.

Biederman, J., Newcorn, J., and Sprich, S. (1991). Comorbidity of attention deficit hyperactivity disorder with conduct, depressive, anxiety, and other disorders. Am J Psychiatry 148, 564–577.

Brookes, K., Xu, X., Chen, W., Zhou, K., Neale, B., Lowe, N., Anney, R., Franke, B., Gill, M., Ebstein, R., 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.

Brookes, K.J., Knight, J., Xu, X., and Asherson, P. (2005). DNA pooling analysis of ADHD and genes regulating vesicle release of neurotransmitters. Am J Med Genet B Neuropsychiatr Genet 139B, 33–37.

Chen, J., Aronow, B.J., and Jegga, A.G. (2009a). Disease candidate

gene identification and prioritization using protein interaction networks. BMC Bioinformatics 10, 73.

Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009b). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res 37, W305-W311.

Chen, J., Xu, H., Aronow, B.J., and Jegga, A.G. (2007). Improved human disease candidate gene prioritization using mouse phenotype. BMC Bioinformatics 8, 392.

Chen, Y., Wang, W., Zhou, Y., Shields, R., Chanda, S.K., Elston, R.C., and Li, J. (2011a). In silico gene prioritization by integrating multiple data sources. PLoS ONE 6, e21137.

Chen, Y.A., Tripathi, L.P., and Mizuguchi, K. (2011b). TargetMine, an integrated data warehouse for candidate gene prioritisation and target discovery. PLoS ONE 6, e17844.

Elbers, C.C., Onland-Moret, N.C., Franke, L., Niehoff, A.G., van der Schouw, Y.T., and Wijmenga, C. (2007). A strategy to search for common obesity and type 2 diabetes genes. Trends in endocrinology and metabolism: TEM 18, 19–26.

Faraone, S.V., and Doyle, A.E. (2001). The nature and heritability of attention-deficit/hyperactivity disorder. Child Adolesc Psychiatr Clin N Am 10, 299–316, viii–ix.

Faraone, S.V., Perlis, R.H., Doyle, A.E., Smoller, J.W., Goralnick, J.J., Holmgren, M.A., and Sklar, P. (2005). Molecular genetics of attention-deficit/hyperactivity disorder. Biol Psychiatry 57, 1313–1323.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., et al. (2011). Ensembl 2011. Nucleic Acids Res 39, D800–806.

Frank, S.P., Thon, K.P., Bischoff, S.C., and Lorentz, A. (2011). SNAP-23 and syntaxin-3 are required for chemokine release by mature human mast cells. Mol Immunol 49, 353–358.

Franke, L., Bakel, H.v., Fokkens, L., de Jong, E.D., Egmont-Petersen, M., and Wijmenga, C. (2006). Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. Am J Hum Genet 78, 1011–1025.

Gizer, I.R., Ficks, C., and Waldman, I.D. (2009). Candidate gene studies of ADHD: a meta-analytic review. Hum Genet 126, 51–90.

Guan, L., Wang, B., Chen, Y., Yang, L., Li, J., Qian, Q., Wang, Z., Faraone, S.V., and Wang, Y. (2009). A high-density single-nucleotide polymorphism screen of 23 candidate genes in attention deficit hyperactivity disorder: suggesting multiple susceptibility genes among Chinese Han population. Mol Psychiatry 14, 546–554.

Ivanov, I., Bansal, R., Hao, X., Zhu, H., Kellendonk, C., Miller, L., Sanchez-Pena, J., Miller, A.M., Chakravarty, M.M., Klahr, K., et al. (2010). Morphological abnormalities of the thalamus in youths with attention deficit hyperactivity disorder. Am J Psychiatry 167, 397–408.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 38, D355–360.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., et al. (2011). Spatio-temporal transcriptome of the human brain. Nature 478, 483–489.

Kennedy, M.J., and Ehlers, M.D. (2011). Mechanisms and function of dendritic exocytosis. Neuron 69, 856–875.

Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., et al. (2009). Human protein reference database--2009 update. Nucleic Acids Res 37, D767–772.

Kohler, S., Bauer, S., Horn, D., and Robinson, P.N. (2008). Walking the interactome for prioritization of candidate disease genes. Am J Hum Genet 82, 949–958.

Lasky-Su, J., Neale, B.M., Franke, B., Anney, R.J., Zhou, K., Maller, J.B., Vasquez, A.A., Chen, W., Asherson, P., Buitelaar, J., et al. (2008). Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. Am J Med Genet B Neuropsychiatr Genet 147B, 1345–1354.

Levin, E.D. (2002). Nicotinic receptor subtypes and cognitive function. J Neurobiol 53, 633–640.

Liu, S.T., Sharon-Friling, R., Ivanova, P., Milne, S.B., Myers, D.S., Rabinowitz, J.D., Brown, H.A., and Shenk, T. (2011). Synaptic vesicle-like lipidome of human cytomegalovirus virions reveals a role for SNARE machinery in virion egress. Proc Natl Acad Sci U S A 108, 12869–12874.

McKusick, V.A. (2007). Mendelian Inheritance in Man and its online version, OMIM. Am J Hum Genet 80, 588–604.

Mulder, N.J., and Apweiler, R. (2008). The InterPro database and tools for protein domain analysis. Current protocols in bioinformatics/editoral board, Andreas D. Baxevanis et al. Chapter 2, Unit 27.

Nakamura, K., Iwata, Y., Anitha, A., Miyachi, T., Toyota, T., Yamada, S., Tsujii, M., Tsuchiya, K.J., Iwayama, Y., Yamada, K., et al. (2011). Replication study of Japanese cohorts supports the role of STX1A in autism susceptibility. Prog Neuropsychopharmacol Biol Psychiatry 35, 454–458.

Neale, B.M., Medland, S.E., Ripke, S., Asherson, P., Franke, B., Lesch, K.P., Faraone, S.V., Nguyen, T.T., Schafer, H., Holmans, P., et al. (2010). Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry 49, 884–897.

Oades, R.D. (2008). Dopamine-serotonin interactions in attention-deficit hyperactivity disorder (ADHD). Prog Brain Res 172, 543–565.

Ribases, M., Ramos-Quiroga, J.A., Hervas, A., Bosch, R., Bielsa, A., Gastaminza, X., Artigas, J., Rodriguez-Ben, S., Estivill, X., Casas, M., et al. (2009). Exploration of 19 serotoninergic candidate genes in adults and children with attention-deficit/hyperactivity disorder identifies association for 5HT2A, DDC and MAOB. Mol Psychiatry 14, 71–85.

Sharp, S.I., McQuillin, A., and Gurling, H.M. (2009). Genetics of attention-deficit hyperactivity disorder (ADHD). Neuropharmacology 57, 590–600.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27, 431–432.

Sollner, T., Whiteheart, S.W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P., and Rothman, J.E. (1993). SNAP receptors implicated in vesicle targeting and fusion. Nature 362, 318–324.

Sun, J., Jia, P., Fanous, A.H., Webb, B.T., van den Oord, E.J.C.G.,

Chen, X., Bukszar, J., Kendler, K.S., and Zhao, Z. (2009). A multi-dimensional evidence-based candidate gene prioritization approach for complex diseases-schizophrenia as a case. Bioinformatics 25, 2595–6602.

Swanson, J.M., Kinsbourne, M., Nigg, J., Lanphear, B., Stefanatos, G.A., Volkow, N., Taylor, E., Casey, B.J., Castellanos, F.X., and Wadhwa, P.D. (2007). Etiologic subtypes of attention-deficit/ hyperactivity disorder: brain imaging, molecular genetic and environmental factors and the dopamine hypothesis. Neuropsychol Rev 17, 39–59.

Teber, E.T., Liu, J.Y., Ballouz, S., Fatkin, D., and Wouters, M.A. (2009). Comparison of automated candidate gene prediction systems using genes implicated in type 2 diabetes by genome-wide association studies. BMC Bioinformatics 10 Suppl 1, S69.

Thienpont, B., Zhang, L., Postma, A.V., Breckpot, J., Tranchevent, L.C., Van Loo, P., Mollgard, K., Tommerup, N., Bache, I., Tumer, Z., et al. (2010). Haploinsufficiency of TAB2 causes congenital heart defects in humans. Am J Hum Genet 86, 839–849.

Tiffin, N., Adie, E., Turner, F., Brunner, H.G., van Driel, M.A., Oti, M., Lopez-Bigas, N., Ouzounis, C., Perez-Iratxeta, C., Andrade-Navarro, M.A., et al. (2006). Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes. Nucleic Acids Res 34, 3067–3081.

Tranchevent, L.-C., Barriot, R., Yu, S., Van Vooren, S., Van Loo, P., Coessens, B., De Moor, B., Aerts, S., and Moreau, Y. (2008). Endeavour update: a web resource for gene prioritization in multiple species. Nucleic Acids Res 36, W377–384.

Tranchevent, L.-C., Capdevila, F.B., Nitsch, D., De Moor, B., De Causmaecker, P., and Moreau, Y. (2010). A guide to web tools to prioritize candidate genes. Brief Bioinform 12, 22–32.

Volkow, N.D., Wang, G.J., Tomasi, D., Kollins, S.H., Wigal, T.L., Newcorn, J.H., Telang, F.W., Fowler, J.S., Logan, J., Wong, C.T., et al. (2012). Methylphenidate-elicited dopamine increases in ventral striatum are associated with long-term symptom improvement in adults with attention deficit hyperactivity disorder. J Neurosci 32, 841–849.

Walther, D.J., Peter, J.U., Bashammakh, S., Hortnagl, H., Voits, M., Fink, H., and Bader, M. (2003). Synthesis of serotonin by a second tryptophan hydroxylase isoform. Science 299, 76.

Zhang, K., Chang, S., Cui, S., Guo, L., Zhang, L., and Wang, J. (2011a). ICSNPathway: identify candidate causal SNPs and pathways from genome-wide association study by one analytical framework. Nucleic Acids Res 39, W437–443.

Zhang, K., Cui, S., Chang, S., Zhang, L., and Wang, J. (2010). i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. Nucleic Acids Res 38, W90–95.

Zhang, L., Chang, S., Li, Z., Zhang, K., Du, Y., Ott, J., and Wang, J. (2012). ADHDgene: a genetic database for attention deficit hyperactivity disorder. Nucleic Acids Res 40, D1003–1009.

Zhou, K., Dempfle, A., Arcos-Burgos, M., Bakker, S.C., Banaschewski, T., Biederman, J., Buitelaar, J., Castellanos, F.X., Doyle, A., Ebstein, R.P., et al. (2008). Meta-analysis of genome-wide linkage scans of attention deficit hyperactivity disorder. Am J Med Genet B Neuropsychiatr Genet 147B, 1392–1398.

