Original Article

IQdb: an intelligence quotient score-associated gene resource for human intelligence

Lei Kong¹, Lu Cheng¹, Li-ya Fan², Min Zhao^{1,*} and Hong Qu^{1,*}

¹Center for Bioinformatics, State Key Laboratory of Protein and Plant Gene Research, College of Life Sciences, Peking University, Beijing 100871, P.R. China and ²Software Service Delivery Excellence, IBM China Research Lab, Beijing, 100193 China

*Corresponding author: Tel/Fax: +86 10 6275 5206; Email: zhaomin.2006@gmail.com Correspondence may also be addressed to Hong Qu. Tel/Fax: +86 10 6275 5206; Email: quh@mail.cbi.pku.edu.cn.

Submitted 10 April 2013; Revised 19 July 2013; Accepted 8 August 2013

Citation details: Kong,L., Cheng,L., Fan,L.Y. et al. IQdb: an intelligence quotient score-associated gene resource for human intelligence. Database, Vol. 2013, Article ID bat063, doi:10.1093/database/bat063

Intelligence quotient (IQ) is the most widely used phenotype to characterize human cognitive abilities. Recent advances in studies on human intelligence have identified many new susceptibility genes. However, the genetic mechanisms involved in IQ score and the relationship between IQ score and the risk of mental disorders have won little attention. To address the genetic complexity of IQ score, we have developed IQdb (http://IQdb.cbi.pku.edu.cn), a publicly available database for exploring IQ-associated human genes. In total, we collected 158 experimental verified genes from literature as a core dataset in IQdb. In addition, 46 genomic regions related to IQ score have been curated from literature. Based on the core dataset and 46 confirmed linked genomic regions, more than 6932 potential IQ-related genes are expanded using data of protein–protein interactions. A systematic gene ranking approach was applied to all the collected and expanded genes to represent the relative importance of all the 7090 genes in IQdb. Our further systematic pathway analysis reveals that IQ-associated genes are significantly enriched in multiple signal events, especially related to cognitive systems. Of the 158 genes in the core dataset, 81 are involved in various psychotic and mental disorders. This comprehensive gene resource illustrates the importance of IQdb to our understanding on human intelligence, and highlights the utility of IQdb for elucidating the functions of IQ-associated genes and the cross-talk mechanisms among cognition-related pathways in some mental disorders for community.

Database URL: http://IQdb.cbi.pku.edu.cn.

Introduction

Human intelligence refers to a set of cognitive abilities, such as thinking, remembering, reading, learning, problem solving and using language. The high genetic heterogeneity of intelligence poses an enormous challenge for understanding molecular mechanisms for cognition. Intelligence quotient (IQ) is the most widely used phenotype for characterizing human intelligence in psychometric studies. It is not surprising that IQ score is consistently associated with a number of mental disorders such as schizophrenia, autism, depression and anxiety (1–3). Although genetic epidemiology of the relationship between IQ score and the risk of related mental disorders becomes increasingly clear with various lines of studies, there are no substantial achievements to contribute to understanding the molecular mechanisms underlying human intelligence and relevant mental disorders.

As a quantitative trait, the heritability behind an observed IQ score is due to complex genetic interactions between multiple genes of small effect sizes (4–6). Genetic association studies have identified many candidate genes for human intelligence; however, many candidates fail to be replicated between studies and populations (4).

[©] The Author(s) 2013. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited. Page 1 of 7 (page number not for citation purposes)

Additionally, current genetic predisposition information is scattered in literature and, to date, there has been no systematic collection and analysis. Hence, there is no detailed investigation on the common molecular mechanisms between IQ score and the risk of related mental disorder. Development of a more comprehensive gene resource is really desired to gain a more complete molecular picture for intelligence and relevant disorders.

In this article, we present the IQdb, an IQ-associated gene database for ongoing development of genes relevant to intelligence and serving as a reference dataset for understanding the mechanisms of human intelligence. The resultant gene list, preferably in IQdb with additional functional and genetic information, including gene association study, family-based linkage study, genome-wide association study and other functional studies, would be a valuable resource for the community. In addition, our systematic pathway and disease enrichment analyses reveal that the IQ-associated genes enriched in multiple signal events are involved with many cancers and mental disorders. To the best of our knowledge, IQdb is the first example of an integrated and comprehensive gene resource that helps to elucidate the relationship between IQ score and genetic risk factors in mental disorders. Our collection could have profound implications for the diagnosis, treatment and prevention of some intelligence-related mental disorders.

Data Annotations

Collection of core dataset, experimental verified candidate genes

As shown in Figure 1, this comprehensive collection of gene and genomic information for IQdb was accomplished by curating from published articles using the following four steps:

- (i) An extensive literature search, particularly concerning family-based linkage studies, population association studies, genome-wide association studies and other functional analyses, was conducted through PubMed (on 10 January 2013) using the following search terms: ["intelligence quotient" (Title/Abstract) OR "IQ" (Title/Abstract)] AND ["genome-wide association study" (Title/Abstract) OR "genome wide association study" (Title/Abstract)] OR ["genome" (Title/Abstract) OR "genetic" (Title/Abstract)] OR ["association" (Title/Abstract)] OR ["association" (Title/Abstract)] OR ["association" (Title/Abstract)].
- (ii) The retrieved 2307 abstracts were highlighted with query keywords and grouped by the function in Entrez system in Related Articles.
- (iii) The 2307 abstracts were read manually to curate the experimental verified candidate genes, single-nucleotide polymorphisms (SNPs) and genomic regions relevant to IQ and other related information such as experimental methods and studied population.

(iv) All the names of experimental verified candidate gene and SNPs were manually mapped to 158 Entrez Gene IDs and 139 SNP IDs. For accuracy, we excluded all negative reports. Finally, we defined the 158 genes as a core dataset with high confidence. In addition, 46 genomic regions were also curated from linkage studies (4). To expand the IQ-associated gene list, we overlapped the genes to these curated 46 genomic regions based on RefSeq gene annotation from UCSC genome browser (7).

Expanding and ranking candidate genes from genomic regions and protein–protein interactions

The molecular basis underlying IQ score is still unclear because of its high genetic heterogeneity. Classical identification of candidate genes in individual studies often focuses on verifying specific genes/variants predisposing to IQ. Therefore, systematic evaluation and summary of relationship between all candidate genes is rare. In this article, we first expanded the IQ-associated genes based on the core dataset using linked genomic regions and protein-protein interactions. Using a multi-dimensional evidence-based candidate gene prioritization approach (8), the relative importance of each expanded gene was estimated based on the supported evidence from literature, genomics regions and functional roles. For instance, 3898 genes locating in the 46 curated genomic regions were expanded. And 3063 genes that interacted with 158 genes in the core dataset were further introduced from the BioGRID (9), HPRD (10) and BIND (11) databases. Finally, 7090 genes, including the genes in the core dataset, were integrated together as a most comprehensive IQ-associated gene list.

To calculate the relativities of all 7090 genes, a benchmark dataset including 19 IQ-associated genes with positive evidence was compiled from a classical review (4) (Supplementary File 1). Then, we followed a gene prioritization approach (12) to generate a candidate weight matrix pool including $d^{N} = 4^{3}$ weight vectors, where N represents the number of evidence, including literature, linkage regions and interactions, and d = N + 1 represents possible different weights, from 1 to 4 in the weight vectors. A combined score for each gene was then calculated by summing up the products of the scores and the corresponding weights from the three evidence (8). All the 7090 candidate genes, including 19 benchmark genes, were sorted by their combined scores. We selected the optimal weight matrix [4, 1, 1] that gave the 95% benchmark genes the highest rank among the top 5% of all candidate genes. Based on the matrix, we evaluated the relevance of the 7090 introduced genes with IQ score, which was useful for users to get potential genes for further screening.

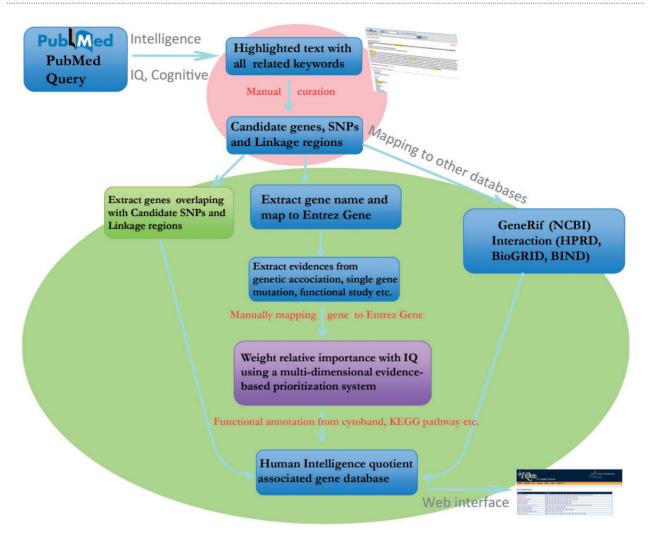


Figure 1. Pipeline for collection, expansion and annotation of IQ-associated genes.

Biological function annotations

Extensive functional information has been retrieved and integrated for better understanding the function of the IQassociated genes, such as cross-links to NCBI Entrez gene (13), OMIM (13), UniProt (14), Ensembl (15) and Gene Ontology (16). Comprehensive mRNA expression profiling data are also collected from BioGPS (17), Allen Brain Atlas (18) and RNA-Seq (19–24). Several popular pathway databases are used to get comprehensive pathway-related information, including BioCyc (25), KEGG Pathway (26), PID Curated (27), PANTHER (28), PID Reactome (29, 30), rate-limiting enzyme database (31), pathway localization database (32) and transporter substrate database (33). Other possible association diseases are also integrated from GAD (gene association database) (34), KEGG Disease (35), FunDO (36, 37), NHGRI (38) and OMIM (13). In addition, the original IQ-related literature references in the NCBI PubMed database are linked to each gene in the core dataset. We shall routinely update the core dataset based on newly published literature and other diseaserelated databases. An automatic annotation pipeline is implemented to integrate functional information from Gene annotation (39), Gene Ontology annotation, HPRD/BIND/ BioGRID interaction annotation, KEGG LIGAND/BioCarta signaling event annotation (40, 41) and OMIM annotation using Perl Script and Swiss knife module (10, 42–44). The result shows the automatic pipeline allows IQ-associated gene annotation to be easily updated when new versions of external databases are available.

Data Analysis

Pathway enrichment analysis on the core dataset

So far, gene-set pathway enrichment analyses have been probably the most practical and successful data mining approaches to explore underlying molecular causes of complex diseases (45). In this article, the hypergeometric test was used to calculate whether a given set of object pairs had a different frequency of annotation pairs than that would be expected by chance, and it gave the sample sizes involved and the expected frequency of such pairs. Using all known genes in the human genome as background, we have identified the statistically significant enriched pathways and diseases for core dataset genes in IQdb.

Enriched functional pathways for the 158 genes in the core dataset are mainly related to neuronal function such as cocaine addiction, long-term potentiation, dopamine degradation and neurotransmitter release cycle (Table 1). In addition to neuron-related pathways and neurotransmitter biosynthesis and degradation, the genes were also highly enriched in developmental biology. These results highlight that multiple neurotransmitter-related signaling events are related to cognitive process. As the majority of molecules in these signaling pathways play fundamental roles in response to environment signals, regulating neuronal development and synaptic function, integration of these different signals together is the key step to process information. In summary, the level of complexity of signaling systems involved in cognitive systems stems from the functions of components as fundamental cellular roles.

Enrichment diseases for the 158 IQ-related genes in the core dataset

As a fundamental role of cognition, it is not surprising that the genes are consistently associated with a number of complex diseases. Although it is difficult to measure how much the IQ score may have contributed to certain diseases based on gene content, it might give a clue that helps to generate hypotheses to examine the potential role of IQ score as a risk factor in relevant disease. A guick disease analysis has revealed that the 158 genes in the core dataset are related to a broad spectrum of human diseases such as various cancers and mental disorders (Table 2). In total, 81 genes are related to psychotic and mental disorders. The mental disorders mainly include schizophrenia, autism, depression, bipolar, obsessive-compulsive disorder and Parkinson's disease. Plenty of previous reports suggest that early-onset and adult-onset schizophrenia are associated with intellectual deficits (46, 47). However, the underlying common molecular mechanism between schizophrenia and IQ scores is still unknown. In IQdb, 37 genes related to schizophrenia are highly enriched in neurotransmitter metabolism pathways, including 'Adrenaline and noradrenaline biosynthesis', 'Dopamine clearance from the synaptic cleft' and 'Arginine and proline metabolism'. These pathways suggest that the early-onset and adultonset schizophrenia might be related to some compound metabolisms such as dopamine metabolism. Most interestingly, several IQ-related genes are associated with several mental disorders. For instance, SLC6A4 is associated with autistic disorder, schizophrenia, obsessive compulsive

Table 1. The statistically significant enriched pathways ofIQ-associated genes in the core dataset from different pathway databases

Pathway	Source	Corrected <i>P</i> -value*
Neuronal system	Reactome	4.28E-04
Cocaine addiction	KEGG PATHWAY	3.95E-03
Long-term potentiation	KEGG PATHWAY	9.04E-03
Dopamine degradation	BioCyc	1.88E-02
Developmental biology	Reactome	2.51E-02
Noradrenaline and adrenaline degradation	ВіоСус	2.51E-02
Adrenaline and noradrenaline biosynthesis	PANTHER	2.76E-02
Arginine and proline metabolism	KEGG PATHWAY	3.79E-02
Serotonin neurotransmitter release cycle	PID Reactome	4.18E-02
Dopamine neurotransmitter release cycle	PID Reactome	4.18E-02
Neurotransmitter release cycle	PID Reactome	4.18E-02

*The corrected *P*-value was calculated by Fisher exact test followed by Benjamini–Hochberg multiple testing correction using the Ingenuity Pathway Tool.

 Table 2. The top 10 enriched diseases of IQ-associated genes

 in the core dataset with experimental supports

Disease	Source	Corrected <i>P</i> -value*
Behavior disease	FunDO	8.71E-09
Psychotic disorder	FunDO	1.42E-08
Autistic disorder	FunDO	2.98E-07
Cognitive function	GAD	9.38E-06
Schizophrenia	GAD	5.66E-05
Obsessive compulsive disorder	GAD	4.26E-04
Noonan syndrome	KEGG DISEASE	6.58E-04
Other congenital disorders	KEGG DISEASE	1.14E-03
Bipolar disorder	FunDO	3.95E-03
Congenital disorders of development	KEGG DISEASE	4.74E-03

*The corrected *P*-value was calculated by Fisher exact test followed by Benjamini–Hochberg multiple testing correction using the Ingenuity Pathway Tool.

disorder, bipolar disorder, personality disorders, affective disorder, attention deficit hyperactivity disorder, suicide, Alzheimer's disease and depression. Thus, the relationships between common IQ-associated genes and diseases are promising for future biological experiments or replication efforts to discover the underlying common pathways. In

Original Article

Α	Quick Search		G				в		arch By	Gene ID						
Basic information	Basic inform	nation						Ger	ne ID:			sul	bmit (example:	6197)	
Expression	Gene ID	3759						Sec	arch By	Pathway	& Dise	ase				
Pathway	Dataset	core dataset						Det	hway:						Constantio	Terrentedes
Reference	Gene Symbol	KCNJ2														Transmission
	Alias	HHBIRK1 HHIRK	1 IRK1 KIR2	.1 LQT7	SQT3			Dise	ease:			sub	mit (e	example: s	chizophr	enia)
Nucleotide Sequence	Chromosome	17						En	arch By	Charmon						
Protein Sequence	Cytoband Description	17q24.3 potassium inward rectifier K+ chan	ily-rectifying cl inel KIR2.1 in	hannel, sub ward rectif	family J, me ier potassiu	mber 2;IRK	-1 cardiac i 2 potassium	nwar		Chromos	ome		submit	(examp	ple: X)	
	Gene Type	protein-coding							oband			su	ubmit	(example:	Xp22.2)	
	Tissue	N/A						_				(inter-				
	Specific							Sec	arch By	Evidence	Public	ation				
	Weight Score Evidence	8 "ATS subjects un	warrally had m	ore school	difficulties	than their e	iblings despi	ite ri 5.44		- D. Luis					1	
	Xref	ENSG00000123700		ore school	uniculues	ulan ulen :	aburigs, despi								ubmit	
	Xref [hsa:3759						Sea	irch in Rela	ted Senter	nce:			2	submit	
	Xref S	4504835						5	arch By	Evonacci	on Law	a l				
	Xref 📿	P63252						000	aren by	capiessi	on Leve					
								-			≤ UH	RLowcov	≤ -			and
	Expression	Level									≤ bra	inLowcov	<u>s</u> -			and
	UHRLowcov	brainLowe	cov adij	pose b	rain bre	ast col	on heart	liver	lymphi	lode	skelMu	cle	testes	MCF7	1	
	0.04	1.04	1.0	2	.02 0.7	3 0.3	7 4.22	0.21	0.74		2.86		0.33	0.0		
				oellum3	cerebel	laure d			rebellum6	BT4	74HME	HME	MB43	5 T47	D	
	cerebellum1 0.14	0.28	0.21	Jettum's	0.23		cerebellum5 0.32	0.2		0.02		0.11	0.0	0.0		
c		0.28	0.21	Related F	0.23			0.2	21							
с			0.21 Top 10 F	Related F	0.23 Pathways			0.1 Ge	21 me IDs	0.02		0.11	0.0	0.0		2265 6616
с		0.28	0.21 Top 10 Pathway Synaptic	Related F	0.23 Pathways			0.2 Ge 375	21 nne IDs 59 , 1139 , 4	0.02	, 2904 , 0	0.11	0.0	0.0	141 , 6197	, 3265 , 6616
	0.14	0.28	0.21 Top 10 F Pathway Synaptic Prostate	Related F Transmissio cancer	0.23 Pathways			0.2 Ge 375 665	21 nne IDs 59 , 1139 , 4 54 , 673 , 32	0.02	, 2904 , 6 3845 , 58	0.11	0.0	0.0	141 , 6197 87 , 2950	, 3265 , 6616
Choose program to use and dat	0.14	0.28 D	0.21 Top 10 F Pathway Synaptic Prostate Signaling	Related F Transmissio cancer by Insulin	0.23 Pathways			Ge 375 665 936	21 me IDs 59 , 1139 , 4 54 , 673 , 32 65 , 6654 , 7	0.02 4128 , 1312 265 , 5604 , 7249 , 7248	, 2904 , (3845 , 58 , 5604 , 3	0.11	0.0 7 , 22999 , 2033 , 4 , 5781 ,	0.0	141 , 6197 87 , 2950	
Choose program to use and dat Program blastp - Database	0.14 abase to search: IQdb_protein	0.28 D	0.21 Top 10 I Pathway Synaptic Prostate Signaling Long-ter	Related F Transmissio cancer by Insulin i m potentia	0.23 Pathways on receptor tion			0.2 Ge 375 665 936 290	21 sne IDs 59, 1139, 4 54, 673, 32 65, 6654, 3 04, 673, 32	0.02 4128 , 1312 265 , 5604 , 7249 , 7248 265 , 5604 ,	, 2904 , 0 3845 , 50 , 5604 , 3 3845 , 61	0.11 6804 , 113 194 , 2263 1845 , 589	0.0 7 , 22999 , 2033 , 4 , 5781 , , 2033 ,	0.0 , 673 , 11 2260 , 138 , 2263 , 32 1387	141 , 6197 87 , 2950	
Choose program to use and dat Program blastp - Database	0.14 abase to search: IQdb_protein	0.28 D	0.21 Top 10 I Pathway Synaptic Prostate Signaling Long-ter Renal ce	Related F Transmissio cancer by Insulin in m potentia II carcinom	0.23 Pathways on receptor tion a			0.2 Ge 375 665 290 665	ne IDs 59 , 1139 , 4 54 , 673 , 32 65 , 6654 , 7 04 , 673 , 32 54 , 673 , 32	0.02 4128 , 1312 265 , 5604 , 7249 , 7248 265 , 5604 , 265 , 5604 ,	, 2904 , 0 3845 , 50 , 5604 , 3 3845 , 60 3845 , 50	0.11 8804 , 113 194 , 2263 1845 , 589 197 , 5894 194 , 5781	0.0 7 , 22999 , 2033 , 4 , 5781 , , 2033 , , 2033 ,	0.0	141 , 6197 87 , 2950 265 , 2260	
Choose program to use and dat Program blastp • Database	0.14 abase to search: IQdb_protein	0.28 D	0.21 Top 10 I Pathway Synaptic Prostate Signaling Long-ter Renal ce Neuroac	Related F Transmissio cancer by Insulin n m potentia II carcinom tive ligand-	0.23 Pathways on receptor tion	teraction		0.2 Ge 375 665 936 290 665 181	ne IDs 59 , 1139 , 4 54 , 673 , 32 65 , 6654 , 7 04 , 673 , 32 54 , 673 , 32	0.02 4128 , 1312 265 , 5604 , 7249 , 7248 265 , 5604 , 265 , 5604 , 1139 , 7442	, 2904 , 0 3845 , 56 3845 , 61 3845 , 51 3845 , 51	0.11 8804 , 113 194 , 2263 1845 , 589 197 , 5894 194 , 5781 193 , 5021	0.0 7 , 22999 , 2033 , 4 , 5781 , , 2033 , , 2033 , , 1137 ,	0.0	141 , 6197 87 , 2950 265 , 2260	
Choose program to use and dat Program blastp • Database	0.14 abase to search: IQdb_protein	0.28 D	0.21 Top 10 f Pathway Synaptic Prostate Signaling Long-ter Renal ce Neuroac Arginine	Related F Transmissio cancer by Insulin I m potentia II carcinom tive ligand- and proline	0.23 Pathways on receptor tion a receptor in	teraction n		0.2 Ge 375 665 934 290 665 181 412	21 me IDs 55, 1139, 4 54, 673, 32 55, 6654, 7 04, 673, 32 54, 673, 32 14, 1815, 1	0.02 4128 , 1312 265 , 5604 , 7249 , 7248 265 , 5604 , 1139 , 7442 5625 , 1610	, 2904 , 6 3845 , 58 , 5604 , 3 3845 , 61 3845 , 58 , 886 , 77 , 4846 , 2	0.11 	0.0 7 , 22999 , 2033 , 4 , 5781 , , 2033 , , 2033 , , 1137 , 2	0.0 , 673 , 111 2260 , 138 , 2263 , 32 1387 1387 1387 1813 , 290	141 , 6197 37 , 2950 265 , 2260 04 , 1141 ,	
C Choose program to use and dat Program blastp • Database Enter sequence below in FAST/	0.14 abase to search: IQdb_protein	0.28 D	0.21 Top 10 I Pathway Synaptic Prostate Signaling Long-ter Renal ce Neuroac Arginine Neurotro	Related F Transmissic cancer by Insulin m potentia Il carcinom tive ligand- and proline ophin signal	0.23 Pathways on receptor tion a receptor in e metaboliss ing pathway	teraction n		0.2 Ge 375 665 2936 296 665 181 412 490	ne IDs 59, 1139, 4 54, 673, 32 65, 6654, 7 04, 673, 32 14, 1815, 7 29, 4128, 5	0.02 4128 , 1312 265 , 5604 , 7249 , 7248 265 , 5604 , 265 , 5604 , 1139 , 7442 5625 , 1610 354 , 673 ,	, 2904 , 6 3845 , 56 3845 , 604 , 3 3845 , 60 3845 , 56 , 886 , 77 , 4846 , 1 5604 , 384	0.11 6804 , 113 194 , 2263 1845 , 589 197 , 5894 194 , 5781 153 , 5021 1628 , 484 15 , 6197 ,	0.0 7 , 22999 , 2033 , 4 , 5781 , , 2033 , , 2033 , , 1137 , 2	0.0 , 673 , 111 2260 , 138 , 2263 , 32 1387 1387 1387 1813 , 290	141 , 6197 37 , 2950 265 , 2260 04 , 1141 ,	
Choose program to use and dat Program blastp - Database	0.14 abase to search: IQdb_protein	0.28 D	0.21 Top 10 I Pathway Synaptic Prostate Signaling Long-tree Renal ce Neuroac Arginine Neurotro Adrenalis	Related F Transmissic cancer by Insulin in m potentia Il carcinom tive ligand-i and profine ophin signal- ne and nora	0.23 Pathways on receptor tion a receptor in e metaboliss ing pathway adrenaline b	teraction n		0.2 Ge 375 665 936 290 665 181 412 412 490 412	me IDs 59, 1139, 4 55, 673, 3 55, 6654, 5 04, 673, 3 54, 673, 3 14, 1815, 2 29, 4128, 5 08, 627, 6 08, 627, 6	0.02 4128 , 1312 265 , 5604 , 7249 , 7248 265 , 5604 , 265 , 5604 , 1139 , 7442 5625 , 1610 554 , 673 , 1312 , 1621	, 2904 , 6 3845 , 58 3845 , 60 3845 , 60 3845 , 60 3845 , 50 , 886 , 77 , 4846 , 2 5604 , 384 , 6616 , 6	0.11 	0.0 7,22999 ,2033, 4,5781, ,2033, ,2033, ,1137, 2 5894,5	0.0 , 673 , 11 2260 , 138 , 2263 , 32 1387 1387 1813 , 290 781 , 3265	141 , 6197 37 , 2950 265 , 2260 04 , 1141 ,	3356 , 5745
Choose program to use and dat Program blastp • Database	0.14 abase to search: IQdb_protein	0.28 D	0.21 Top 10 I Pathway Synaptic Prostate Signaling Long-tree Renal ce Neuroac Arginine Neurotro Adrenalis	Related F Transmissic cancer by Insulin m potentia Il carcinom tive ligand- and proline ophin signal	0.23 Pathways on receptor tion a receptor in e metaboliss ing pathway adrenaline b	teraction n		0.2 Ge 375 665 936 290 665 181 412 412 490 412	me IDs 59, 1139, 4 55, 673, 3 55, 6654, 5 04, 673, 3 54, 673, 3 14, 1815, 2 29, 4128, 5 08, 627, 6 08, 627, 6	0.02 4128 , 1312 265 , 5604 , 7249 , 7248 265 , 5604 , 265 , 5604 , 1139 , 7442 5625 , 1610 554 , 673 , 1312 , 1621	, 2904 , 6 3845 , 58 3845 , 60 3845 , 60 3845 , 60 3845 , 50 , 886 , 77 , 4846 , 2 5604 , 384 , 6616 , 6	0.11 	0.0 7,22999 ,2033, 4,5781, ,2033, ,2033, ,1137, 2 5894,5	0.0 , 673 , 11 2260 , 138 , 2263 , 32 1387 1387 1813 , 290 781 , 3265	141 , 6197 37 , 2950 265 , 2260 04 , 1141 ,	
Choose program to use and dat Program blastp - Database Enter sequence below in FASTA	0.14 abase to search: IQdb_protein	0.28 D	0.21 Top 10 I Pathway Synaptic Prostate Signaling Long-tree Renal ce Neuroac Arginine Neurotro Adrenalis	Related F Transmissic cancer by Insulin i m potential Il carcinom tive ligand- and proline ophin signal ne and nora mating path	0.23 Pathways on receptor tion a receptor in pathway adrenatine to way	teraction n		0.2 Ge 375 665 936 290 665 181 412 412 490 412	me IDs 59, 1139, 4 55, 673, 3 55, 6654, 5 04, 673, 3 54, 673, 3 14, 1815, 2 29, 4128, 5 08, 627, 6 08, 627, 6	0.02 4128 , 1312 265 , 5604 , 7249 , 7248 265 , 5604 , 265 , 5604 , 1139 , 7442 5625 , 1610 554 , 673 , 1312 , 1621	, 2904 , 6 3845 , 58 3845 , 60 3845 , 60 3845 , 60 3845 , 50 , 886 , 77 , 4846 , 2 5604 , 384 , 6616 , 6	0.11 	0.0 7,22999 ,2033, 4,5781, ,2033, ,2033, ,1137, 2 5894,5	0.0 , 673 , 11 2260 , 138 , 2263 , 32 1387 1387 1813 , 290 781 , 3265	141 , 6197 37 , 2950 265 , 2260 04 , 1141 ,	3356 , 5745
Choose program to use and dat Program blastp Database Enter sequence below in FASTA Or load it from disk	0.14 abase to search: IQdb_protein format	0.28 D	0.21 Top 10 I Pathway Synaptic Signaling Long-ter Neuroac Arginine Neurotr Adrenalii MAPK sig	Related F Transmissic cancer by Insulin i m potential Il carcinom tive ligand- and proline ophin signal ne and nora mating path	0.23 Pathways on receptor tion a receptor in pathway adrenatine to way	teraction n		0.2 Ge 375 665 936 290 665 181 412 412 490 412	me IDs 59, 1139, 4 55, 673, 3 55, 6654, 5 04, 673, 3 54, 673, 3 14, 1815, 2 29, 4128, 5 08, 627, 6 08, 627, 6	0.02 4128 , 1312 265 , 5604 , 7249 , 7248 265 , 5604 , 265 , 5604 , 1139 , 7442 5625 , 1610 554 , 673 , 1312 , 1621	, 2904 , 6 3845 , 58 3845 , 60 3845 , 60 3845 , 60 3845 , 50 , 886 , 77 , 4846 , 2 5604 , 384 , 6616 , 6	0.11 	0.0 7,22999 ,2033, 4,5781, ,2033, ,2033, ,1137, 2 5894,5	0.0 , 673 , 11 2260 , 138 , 2263 , 32 1387 1387 1813 , 290 781 , 3265	141 , 6197 37 , 2950 265 , 2260 04 , 1141 ,	3356 , 5745
Choose program to use and dat Program blastp Database Enter sequence below in FASTA Or load it from disk Set subsequence: From	0.14 abase to search: p IQdb_protein , format	0.28 D	0.21 Top 10 I Pathway Synaptic Prostate Signaling Long-ture Renal ce Neuroac Arginine Neurotro Adrenalin MAPK sig	Related F Transmissio cancer by Insulin m potential Il carcinom tive ligand- and proline sphin signal ee and norar maling path Related I	0.23 Pathways on receptor tion a metabolisr ing pathway defenatione way Diseases Gene IDS 2778 , 627	teraction n / biosynthesis	4, 5649, 6533	0.7 Gec 3733 66556 93665 2900 66556 1811 412 490 412 490	21 me IDs 59, 1139, 4 54, 673, 33 55, 6654, 5 04, 673, 33 54, 673, 33 14, 1815, 7 29, 4128, 7 08, 627, 23 08, 627, 23 6531, 2718	0.02 4128 , 1312 265 , 5604 , 7249 , 7248 265 , 5604 , 1139 , 7442 5625 , 1610 554 , 673 , 1 312 , 1621 316 , 6654 ,	, 2904, (3845, 55 3845, 664, 2 3845, 6 3845, 5 8866, 7 7 4846, 1 5604, 384 6615, (673, 566	0.11 804 , 113 94 , 2263 845 , 589 97 , 5894 94 , 5781 553 , 5021 628 , 4844 64 , 4763 , 4 7 , 4763 , 5	0.0 7,22999 ,2033, 4,5781 ,2033, ,1137, 2 5894,5 5894,5	0.0 , 673 , 11 2260 , 138 , 2263 , 32 1387 1387 1387 1387 1813 , 290 781 , 3265 894 , 3845	141, 6197 37, 2950 265, 2260 04, 1141, 5	3356 , 5745 2263 , 3265 ,
Choose program to use and dat Program blastp Database Enter sequence below in FASTA	0.14 abase to search: p IQdb_protein , format	0.28 D	0.21 Top 10 4 Pathway Synaptic Prostate Signaling Long-ter Renal ce Neuroac Arginine Neuroac Neuroac Neuroac Arginine Neuroac	Related F Transmissic cancer by Insulin on m potential Il carcinom tive tigand- and profine sphin signal- and profine sphin signal- ne and noru- rating path Related I renia	0.23 Pathways on receptor tion a metabolism ing pathway Diseases Gene IDs Gene IDs 2778 , 627 6616 , 599	teraction n / ioosynthesis ', 886 , 290- 9 , 1621 , 45	0.32 4 , 5649 , 6533 524 , 267012 ,	0.7 Cee 3773 6665 9333 2992 6655 1812 49	21 me IDs 59, 1139, 4 54, 673, 33 55, 6654, 5 04, 673, 33 54, 673, 33 14, 1815, 7 28, 472, 8 08, 627, 4 29, 4128, 7 08, 627, 23 6531, 2718 00752	0.02 4128 , 1312 265 , 5604 , 7249 , 7248 265 , 5604 , 265 , 5604 , 265 , 5604 , 139 , 7442 5625 , 1610 554 , 673 , 1 312 , 1621 316 , 6654 , 35 , 2944 , 1	, 2904, 4 3845, 51 3845, 61 3845, 61 3845, 51 3845, 51 , 4846, 1 5604, 384 , 6616, 0 673, 560	0.11 804,1113 94,2263 845,589 97,5894 94,5781 53,5021 55,51 14,4763, 5,4128,	0.0	0.0 , 673 , 11 12260 , 138 3, 2263 , 32 1387 1387 781 , 3265 781 , 3265 894 , 3845	141, 6197 7, 2950 265, 2260 24, 1141, 5 5	3356 , 5745 , 2263 , 3265 , 1752 , 5621 , .
Choose program to use and dat Program blastp Database Enter sequence below in FASTA Or load it from disk Set subsequence: From	0.14 abase to search: p IQdb_protein , format	0.28 D	0.21 Top 10 I Pathway Synptic Prosta Signaling Long-ter Renal ce Neuroar Arginine Neurotr Adrenali MAPK sig Top 10 I Disease schizoph Autistic	Related F Transmissio cancer by Insulin m potential Il carcinom tive ligand- and proline ophin signal- ne and norr spaling path Related I renia	0.23 Pathways Pathway	teraction n / siosynthesis , 886 , 290. 9 , 1621 , 43 035 , 1139 ,	0.32 4, 5649, 6533 524, 267012, 4128, 627, '	0.7 Cec 377 666 933 299 665 665 665 665 665 665 645 645	21 me IDs 59, 1139, - 54, 673, 33 54, 673, 33 54, 673, 33 54, 673, 33 14, 1815, - 29, 4128, - 08, 627, 62 29, 4128, - 08, 627, 23 6531, 2718 10752 1, 1137, 80	0.02 4128 , 1312 265 , 5604 , 27249 , 7248 265 , 5604 , 265 , 5604 , 265 , 5604 , 265 , 5604 , 265 , 1610 554 , 673 , 1312 , 1621 316 , 6654 , 55 , 2944 , 1 258 , 1141	, 2904, (3845, 5; 5604, ; 3845, 5; 3845, 5; 3845, 5; 886, 7; 4846, 7; 4846	0.11 8804 , 1113 194 , 2263 845 , 589 97 , 5894 194 , 5781 1628 , 484 194 , 5781 1628 , 482 5 , 4128 , 5 , 4128 , 248 , 6432	0.0 7,22999 ,2033, ,2033, ,2033, ,2033, ,1137, 2 5894,5 6197,5 6197,5	0.0 , 673 , 11 2260 , 138 , 2263 , 33 1387 1387 1813 , 290 781 , 3265 894 , 3845	141, 6197 7, 2950 265, 2260 34, 1141, 5 5 5, 2259, . , , 875, 9 , , , 875, 9	3356 , 5745 2263 , 3265 , 1752 , 5621 ,
Choose program to use and dat Program blastp Database Enter sequence below in FASTA Or load it from disk Set subsequence: From	0.14 abase to search: p IQdb_protein , format	0.28 D	0.21 Top 10 I Pathway Synaptic Prostate Signaling Long-ter Renal ce Neuroac Arginine Neurotr Adrenali MAPK sig Top 10 I Disease schizoph Autistic Psychoti	Related F Transmissio cancer by Insulin n m potential Il carcinom tive ligand- and proline ophin signal disorder and c disorder c disorder	0.23 Pathways Pathway	teraction n / siosynthesis , 886 , 290. 9 , 1621 , 43 035 , 1139 ,	0.32 4 , 5649 , 6533 524 , 267012 ,	0.7 Cec 377 666 933 299 665 665 665 665 665 665 645 645	21 me IDs 59, 1139, - 54, 673, 33 54, 673, 33 54, 673, 33 54, 673, 33 14, 1815, - 29, 4128, - 08, 627, 62 29, 4128, - 08, 627, 23 6531, 2718 10752 1, 1137, 80	0.02 4128 , 1312 265 , 5604 , 27249 , 7248 265 , 5604 , 265 , 5604 , 265 , 5604 , 265 , 5604 , 265 , 1610 554 , 673 , 1312 , 1621 316 , 6654 , 55 , 2944 , 1 258 , 1141	, 2904, (3845, 5; 5604, ; 3845, 5; 3845, 5; 3845, 5; 886, 7; 4846, 7; 4846	0.11 8804 , 1113 194 , 2263 845 , 589 97 , 5894 194 , 5781 1628 , 484 194 , 5781 1628 , 482 5 , 4128 , 5 , 4128 , 248 , 6432	0.0 7,22999 ,2033, ,2033, ,2033, ,2033, ,1137, 2 5894,5 6197,5 6197,5	0.0 , 673 , 11 2260 , 138 , 2263 , 33 1387 1387 1813 , 290 781 , 3265 894 , 3845	141, 6197 7, 2950 265, 2260 34, 1141, 5 5 5, 2259, . , , 875, 9 , , , 875, 9	3356 , 5745 , 2263 , 3265 , 1752 , 5621 , .
Choose program to use and dat Program blastp Database Enter sequence below in FASTA Or load it from disk Set subsequence: From	0.14 abase to search: p IQdb_protein , format	0.28 D	0.21 Top 10 I Pathway Synaptic Prostate Signaling Long-ter Renal ce Neuroac Arginine Neurotra Adrenali MAPK sig Top 10 I Disease schizoph Autistic Psychoti Bardet-B syndrom	Related F Transmissic cancer by Insulin and polential It carcinom tive ligand- and proline ophin signal ne and norra maling path Related I renia disorder c disorder c disorder e	0.23 Pathways	teraction n / oiosynthesis 9,1621,48 035,1139, 1,627,295	0.32 4, 5649, 6533 524, 267012, 4128, 627, '	0.7 Gec 3775 6666 6665 6675 6775 67	21 ne IDs 59, 1139, 4 54, 673, 33 65, 6654, 7 04, 673, 33 54, 673, 33 54, 673, 33 54, 673, 33 54, 673, 33 54, 673, 33 29, 4128, 1 29, 4128, 1 08, 627, 64 29, 4128, 1 08, 627, 23 6531, 2718 0052 1, 1137, 80 1610, 6311	0.02 4128 , 1312 265 , 5604 , 7249 , 7248 265 , 5604 , 265 , 5604 , 265 , 5604 , 265 , 5604 , 3112 , 1621 313 , 7442 3625 , 1610 354 , 673 , 1 312 , 1621 316 , 6654 , 316 , 6654 , 316 , 6654 , 317 , 2944 , 1 258 , 1141 , 6532 , 26	, 2904 , (3845 , 56 3845 , 56 3845 , 56 3845 , 51 3845 , 51 3845 , 51 3846 , 7 5604 , 38 866 , 7 4846 , 7 5604 , 38 5604 , 38 5604 , 38 5604 , 38 5604 , 38 5604 , 5 5604 , 5	0.11 804,113 94,2263 845,589 97,594 94,5781 94,5781 94,5781 94,5781 94,4763, 531 44,4763, 55,4128, 44,4763, 1139	0.0 7,22999 ,2033, 4,5781, ,2033, ,2033, ,2033, ,2033, ,2033, ,2033, ,2035, ,2137, 5894,5 6197,5	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	141, 6197 7, 2950 265, 2260 34, 1141, 5 5 5, 2259, . , , 875, 9 , , , 875, 9	3356 , 5745 2263 , 3265 , 1752 , 5621 ,
Choose program to use and dat Program blastp Database Enter sequence below in FASTA Or load it from disk Set subsequence: From	0.14 abase to search: p IQdb_protein , format	0.28 D	0.21 Top 10 I Pathway Synaptic Prostate Signaling Long-treat Renal ce Neuroac Arginine Neurotro Adrenalin MAPK sig Top 10 I Disease schizoph Autistic Psycholi Bardet-B	Related F Transmissic cancer by Insulin and poly Insulin and poly Insulin and poly Insulin and and proline ophin signal ne and norra maling path Related I renia disorder c disorder c disorder e	0.23 Pathways	teraction n / oiosynthesis 9,1621,48 035,1139, 1,627,295	0.32 4, 5649, 653; 524, 267012, 4128, 627, 1 50, 348, 113;	0.7 Gec 3775 6666 6665 6675 6775 67	21 ne IDs 59, 1139, 4 54, 673, 33 65, 6654, 7 04, 673, 33 54, 673, 33 54, 673, 33 54, 673, 33 54, 673, 33 54, 673, 33 29, 4128, 1 29, 4128, 1 08, 627, 64 29, 4128, 1 08, 627, 23 6531, 2718 0052 1, 1137, 80 1610, 6311	0.02 4128 , 1312 265 , 5604 , 7249 , 7248 265 , 5604 , 265 , 5604 , 265 , 5604 , 265 , 5604 , 3112 , 1621 313 , 7442 3625 , 1610 354 , 673 , 1 312 , 1621 316 , 6654 , 316 , 6654 , 316 , 6654 , 317 , 2944 , 1 258 , 1141 , 6532 , 26	, 2904 , (3845 , 56 3845 , 56 3845 , 56 3845 , 51 3845 , 51 3845 , 51 3846 , 7 5604 , 38 866 , 7 4846 , 7 5604 , 38 5604 , 38 5604 , 38 5604 , 38 5604 , 38 5604 , 5 5604 , 5	0.11 804,113 94,2263 845,589 97,594 94,5781 94,5781 94,5781 94,5781 94,4763, 531 44,4763, 55,4128, 44,4763, 1139	0.0 7,22999 ,2033, 4,5781, ,2033, ,2033, ,2033, ,2033, ,2033, ,2033, ,2035, ,2137, 5894,5 6197,5	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	141, 6197 7, 2950 265, 2260 34, 1141, 5 5, 2259, . , 875, 9 2020, 2711	3356 , 5745 2263 , 3265 , 1752 , 5621 ,
Choose program to use and dat Program blastp Database Enter sequence below in FASTA Or load it from disk Set subsequence: From	0.14 abase to search: p IQdb_protein , format	0.28 D	0.21 Top 10 I Pathway Synaptic Prostate Signaling Long-ter Renal ce Neuroac Arginine Neurotra Adrenali MAPK sig Top 10 I Disease schizoph Autistic Psychoti Bardet-B syndrom	Related F Transmissic cancer by Insulin and poly Insulin and poly Insulin and poly Insulin and and proline ophin signal ne and norra maling path Related I renia disorder c disorder c disorder e	0.23 Pathways	teraction n / oiosynthesis 9,1621,48 035,1139, 1,627,295	4, 5649, 653; 24, 267012, 4128, 627, - 30, 348, 1133 582, 583, 8	0.7 Gec 3775 6666 6665 6675 6775 67	21 ne IDs 59, 1139, 4 54, 673, 33 65, 6654, 7 04, 673, 33 54, 673, 33 54, 673, 33 54, 673, 33 54, 673, 33 54, 673, 33 29, 4128, 1 29, 4128, 1 08, 627, 64 29, 4128, 1 08, 627, 23 6531, 2718 0052 1, 1137, 80 1610, 6311	0.02 4128 , 1312 265 , 5604 , 7249 , 7248 265 , 5604 , 265 , 5604 , 265 , 5604 , 265 , 5604 , 3112 , 1621 313 , 7442 3625 , 1610 354 , 673 , 1 312 , 1621 316 , 6654 , 316 , 6654 , 316 , 6654 , 317 , 2944 , 1 258 , 1141 , 6532 , 26	, 2904 , (3845 , 56 3845 , 56 3845 , 56 3845 , 51 3845 , 51 3845 , 51 3846 , 7 5604 , 38 866 , 7 4846 , 7 5604 , 38 5604 , 38 5604 , 38 5604 , 38 5604 , 38 5604 , 5 5604 , 5	0.11 804,113 94,2263 845,589 97,594 94,5781 94,5781 94,5781 94,5781 94,5781 94,4763, 53,5021 44,4763, 55,4128,6433 31,1139	0.0 7,22999 ,2033, 4,5781, ,2033, ,2033, ,2033, ,2033, ,2033, ,2033, ,2035, ,2137, 5894,5 6197,5	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	141, 6197 7, 2950 265, 2260 34, 1141, 5 5, 2259, . , 875, 9 2020, 2711	3356 , 5745 2263 , 3265 , 1752 , 5621 ,

Figure 2. Web interface of IQdb. (A) The basic information in each IQ-associated gene page. (B) Query interface for text search. (C) BLAST search interface for comparing query against all sequences in IQdb. (D) Browser interface for genes in top 10 enriched pathways, top 10 enriched diseases and shared cytoband.

14q19

11015

14q20

14a23

14q21

11a25

14q32

14011

14q16

1q41

14q17

7036

14q18

8p12

summary, IQdb is valuable in discovery of potential candidate genes, pathways and potential cross-talks between mental disorder and intelligence using comprehensive annotation and user-friendly interface. As a first effort to systematically collect and extend candidate IQ-associated genes, IQdb is also useful to better clarify the molecular mechanisms related to human intelligence.

Interface Development of Database

7q31

14a24

7q32 7q33 7q34 7q35 7q36

All data and information in IQdb are stored in a free, fast and reliable open-source relational database MySQL on a Linux server. Web-based interface to the database is implemented in object-oriented Java, which is a platform-

11q22

independent language and easy to deploy and update. All the Web applications run under a Tomcat+Apache Web server environment. Based on the JavaServer Pages (JSP) technology, dynamical Web pages for each gene in the database are generated. For genes with different evidence, the comprehensive annotation and links are provided (Figure 2A). Gene expression in various tissues and brain regions is represented in tabular format (Figure 2A). In addition, the original literature to support their association with IQ scores is also complied for the 158 genes in the core dataset. For other expanded genes, literature is compiled from the NCBI GeneRIF database (48), which may be useful for users to judge their potential roles with IQ or other cognitive processes.

IQdb allows users to do text guery (Figure 2B), or to run BLAST search against the sequences in IQdb (Figure 2C). To provide a powerful text-based query, six different userfriendly input forms are provided for Entrez Gene ID, pathway and disease annotation, genomic region, literature content and gene expression range in 22 tissues or brains regions. Moreover, a quick full-text search for GeneID, gene symbol or gene alias and publication is on the top right of each page, which is efficient for users to access any data in the database, especially literature-based annotations. In addition, users can browse the data in IQdb in a variety of ways, including significantly enriched pathway, related disease, reported linkage region and chromosome number (Figure 2D). Finally, for any advanced study, IQdb provides all downloadable genetic and population information in a plain text for all the collected 139 SNPs related to IQ.

Conclusions

IQdb is constructed as a free database and analysis server to enable users to rapidly search and retrieve summarized IQassociated genes. Enrichment pathway analyses reveal that multiple signal events related to IQ-associated genes are involved in cognitive systems. Central questions should focus on integration of various signaling pathways to process information. In addition, comprehensive disease enrichment analyses interlink IQ-associated genes with many relevant cancers and mental disorders. IQdb is freely available at http://iqdb.cbi.pku.edu.cn.

Supplementary Data

Supplementary data are available at Database Online.

Funding

This work was supported by the National High-tech 863 Program of China [grant numbers 2006AA02A312, 2008BAI64B01], the National Natural Science Foundation of China [grant number 31171270] and the National Science and Technology Infrastructure Program [grant number 2009FY120100].

Conflict of interest. None declared.

REFERENCES

- 1. Mortensen, E.L., Sorensen, H.J., Jensen, H.H. et al. (2005) IQ and mental disorder in young men. Br. J. Psychiatry, 187, 407–415.
- Koenen, K.C., Moffitt, T.E., Roberts, A.L. *et al.* (2009) Childhood IQ and adult mental disorders: a test of the cognitive reserve hypothesis. *Am. J. Psychiatry*, **166**, 50–57.
- Kalbfleisch,M.L. and Loughan,A.R. (2011) Impact of IQ discrepancy on executive function in high-functioning autism: insight into twice exceptionality. J. Autism Dev. Disord., 42, 390–400.
- Deary, I.J., Johnson, W. and Houlihan, L.M. (2009) Genetic foundations of human intelligence. *Hum. Genet.*, **126**, 215–232.
- 5. Plomin,R. and Haworth,C.M.A. (2009) Genetics of high cognitive abilities. *Behav. Genet.*, **39**, 347–349.
- Plomin,R. and Spinath,F.M. (2004) Intelligence: genetics, genes, and genomics. J. Pers. Soc. Psychol., 86, 112–129.
- Meyer,L.R., Zweig,A.S., Hinrichs,A.S. *et al.* (2013) The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.*, 41, D64–D69.
- Xu,L.M., Li,J.R., Huang,Y. *et al.* (2012) AutismKB: an evidence-based knowledgebase of autism genetics. *Nucleic Acids Res.*, 40, D1016–D1022.
- Stark, C., Breitkreutz, B.J., Chatr-Aryamontri, A. et al. (2011) The BioGRID Interaction Database: 2011 update. Nucleic Acids Res., 39, D698–D704.
- Prasad,T.S., Goel,R., Kandasamy,K. et al. (2008) Human Protein Reference Database—2009 update. Nucleic Acids Res., 37, D767–D772.
- Willis,R.C. and Hogue,C.W. (2006) Searching, viewing, and visualizing data in the Biomolecular Interaction Network Database (BIND). *Curr. Protoc. Bioinformatics*, Chapter 8, Unit 8.9.
- Sun, J., Jia, P., Fanous, A.H. et al. (2009) A multi-dimensional evidence-based candidate gene prioritization approach for complex diseases-schizophrenia as a case. *Bioinformatics*, 25, 2595–6602.
- Sayers, E.W., Barrett, T., Benson, D.A. et al. (2011) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res., 39, D38–D51.
- 14. Magrane, M. and Consortium, U. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)*, **2011**, bar009.
- 15. Flicek, P., Amode, M.R., Barrell, D. *et al.* (2011) Ensembl 2011. *Nucleic Acids Res.*, **39**, D800–D806.
- Gene Ontology Consortium. (2010) The Gene Ontology in 2010: extensions and refinements. Nucleic Acids Res., 38, D331–D335.
- Su,A.I., Wiltshire,T., Batalov,S. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.
- 18. Jones, A.R., Overly, C.C. and Sunkin, S.M. (2009) The Allen Brain Atlas: 5 years and beyond. *Nat. Rev. Neurosci.*, **10**, 821–828.
- Wang,E.T., Sandberg,R., Luo,S. et al. (2008) Alternative isoform regulation in human tissue transcriptomes. Nature, 456, 470–476.
- Wu,J.Q., Habegger,L., Noisa,P. et al. (2010) Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing. Proc. Natl Acad. Sci. USA, 107, 5254–5259.

- 21. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
- 22. Mortazavi,A., Williams,B.A., McCue,K. *et al.* (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **5**, 621–628.
- Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25, 1105–1111.
- Trapnell, C., Williams, B.A., Pertea, G. *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, 28, 511–515.
- Karp,P.D., Ouzounis,C.A., Moore-Kochlacs,C. et al. (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. Nucleic Acids Res., 33, 6083–6089.
- Kanehisa, M., Araki, M., Goto, S. *et al.* (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, 36, D480–D484.
- 27. Schaefer, C.F., Anthony, K., Krupa, S. et al. (2009) PID: the Pathway Interaction Database. *Nucleic Acids Res.*, **37**, D674–D679.
- Thomas,P.D., Campbell,M.J., Kejariwal,A. et al. (2003) PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.*, 13, 2129–2141.
- Croft, D., O'Kelly, G., Wu, G. et al. (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, 39, D691–D697.
- Matthews, L., Gopinath, G., Gillespie, M. et al. (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.*, 37, D619–D622.
- Zhao, M., Chen, X., Gao, G. et al. (2009) RLEdb: a database of ratelimiting enzymes and their regulation in human, rat, mouse, yeast and E. coli. Cell Res., 19, 793–795.
- 32. Zhao, M. and Qu, H. (2010) PathLocdb: a comprehensive database for the subcellular localization of metabolic pathways and its application to multiple localization analysis. *BMC Genomics*, **11** (Suppl. 4), S13.
- 33. Zhao, M., Chen, Y., Qu, D. and Qu, H. (2011) TSdb: a database of transporter substrates linking metabolic pathways and transporter systems on a genome scale via their shared substrates. *Sci. China Life Sci.*, 54, 60–64.
- 34. Becker,K.G., Barnes,K.C., Bright,T.J. and Wang,S.A. (2004) The genetic association database. *Nat. Genet.*, **36**, 431–432.

- Kanehisa, M., Goto, S., Furumichi, M. et al. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res., 38, D355–D360.
- Osborne, J. D., Flatow, J., Holko, M. et al. (2009) Annotating the human genome with Disease Ontology. BMC Genomics, 10 (Suppl. 1), S6.
- 37. Du,P., Feng,G., Flatow,J. et al. (2009) From disease ontology to disease-ontology lite: statistical methods to adapt a generalpurpose ontology for the test of gene-ontology associations. *Bioinformatics*, 25, i63–i68.
- Hindorff,L.A., Sethupathy,P., Junkins,H.A. *et al.* (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA*, **106**, 9362–9367.
- Maglott, D., Ostell, J., Pruitt, K.D. and Tatusova, T. (2011) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, 39, D52–D57.
- Kanehisa, M., Goto, S., Hattori, M. et al. (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res., 34, D354–D357.
- 41. Nishimura,D. (2001) BioCarta. Biotech Softw. Internet Rep., 2, 117–120.
- 42. Hermjakob, H., Fleischmann, W. and Apweiler, R. (1999) Swissknife-'lazy parsing' of SWISS-PROT entries. *Bioinformatics*, **15**, 771–772.
- 43. Stein,L.D. (2004) Using the Reactome database. *Curr. Protoc. Bioinformatics*, Chapter 8, Unit 8.7.
- Feramisco, J.D., Sadreyev, R.I., Murray, M.L. et al. (2009) Phenotypic and genotypic analyses of genetic skin disease through the Online Mendelian Inheritance in Man (OMIM) database. J. Invest. Dermatol., 129, 2628–2636.
- Huang da,W., Sherman,B.T. and Lempicki,R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.*, **37**, 1–13.
- Goldberg,X., Fatjo-Vilas,M., Munoz,M.J. et al. (2011) Increased familiarity of intellectual deficits in early-onset schizophrenia spectrum disorders. World J. Biol. Psychiatry, 13, 493–500.
- Dickson, H., Laurens, K.R., Cullen, A.E. and Hodgins, S. (2011) Metaanalyses of cognitive and motor function in youth aged 16 years and younger who subsequently develop schizophrenia. *Psychol. Med.*, 1–13.
- Lu,Z., Cohen,K.B. and Hunter,L. (2007) GeneRIF quality assurance as summary revision. *Pac. Symp. Biocomput.*, 269–280.