Published online 27 May 2011 *Nucleic Acids Research, 2011, Vol. 39, Web Server issue* ***W437–W443***
doi:10.1093/nar/gkr391

ICSNPathway: identify candidate causal SNPs and pathways from genome-wide association study by one analytical framework

Kunlin Zhang[1], Suhua Chang[1,2], Sijia Cui[1,2], Liyuan Guo[1], Liuyan Zhang[1,2] and Jing Wang[1,*]

[1]Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China, [2]Graduate University of Chinese Academy of Sciences, Beijing 100049, China

Received February 18, 2011; Revised April 21, 2011; Accepted May 3, 2011

ABSTRACT

Genome-wide association study (GWAS) is widely utilized to identify genes involved in human complex disease or some other trait. One key challenge for GWAS data interpretation is to identify causal SNPs and provide profound evidence on how they affect the trait. Currently, researches are focusing on identification of candidate causal variants from the most significant SNPs of GWAS, while there is lack of support on biological mechanisms as represented by pathways. Although pathway-based analysis (PBA) has been designed to identify disease-related pathways by analyzing the full list of SNPs from GWAS, it does not emphasize on interpreting causal SNPs. To our knowledge, so far there is no web server available to solve the challenge for GWAS data interpretation within one analytical framework. ICSNPathway is developed to identify candidate causal SNPs and their corresponding candidate causal pathways from GWAS by integrating linkage disequilibrium (LD) analysis, functional SNP annotation and PBA. ICSNPathway provides a feasible solution to bridge the gap between GWAS and disease mechanism study by generating hypothesis of SNP → gene → pathway(s). The ICSNPathway server is freely available at http://icsnpathway.psych.ac.cn/.

INTRODUCTION

Genome-wide association study (GWAS) (1) is a routine approach to identify novel genetic susceptibility by utilizing genome-wide SNP (single nucleotide polymorphism) array. There have been more than 800 GWAS applications to date (http://www.genome.gov/gwastudies) (2) and the number keeps increasing. From a large amount of genome-wide variants (~300–1000 K or more), a GWAS investigation generally identifies a few SNPs that are statistically significantly associated with a human complex disease or some trait. As GWAS serves as initializations of future genetic and mechanism study of complex traits, one of the key challenges of GWAS data interpretation is to identify causal SNPs (the SNPs that affect trait) and provide profound evidence and hypothesis on the mechanism through which they affect the trait (3).

Currently, there is some research focusing on inferring candidate causal variants from the most significant SNPs (i.e. SNPs with *P*-value below certain threshold. *P*-value $< 10^{-5}$ is utilized by NHGRI GWAS Catalog (2).) or prioritizing the most significant SNPs by linkage disequilibrium (LD) analysis and functional SNP annotation (4–13). However, these analyses can only annotate SNPs to genes. Since a gene can be involved in a variety of pathways, to further annotate genes to pathways, which represent certain biological mechanisms of the complex disease, would require more evidences such as the combined genetic effect. Although pathway-based analysis (PBA) has been developed to identify disease-related pathways (14–18), it emphasizes on interpreting the full list of GWAS SNPs, instead of the most significant SNPs, by searching a large pathway database [e.g. all pathways in the KEGG database (19)]. So the key intention of PBA is to identify novel pathways associated with traits instead of candidate causal pathways that represent the way in which the candidate causal SNPs affect traits.

A feasible proposal to address the above challenges is to establish one unified analytical framework to combine the analysis of candidate causal SNPs and PBA, to generate hypothesis of SNP → gene → pathway(s) which represents that the candidate causal SNP alters the role of its corresponding gene/protein in the context of the pathway(s) associated with traits. So far, there is no web server available to provide such a composed solution. Here, we propose the ICSNPathway (Identify candidate

*To whom correspondence should be addressed. Tel: +86 10 6485 5841; Fax: +86 10 6485 5841; Email: wangjing@psych.ac.cn

Causal SNPs and Pathways) web server, an analytical framework for comprehensive interpretation of GWAS data by integrating LD analysis, functional SNP annotation and pathway-based analysis. ICSNPathway aims to provide an open platform to facilitate researchers to identify the candidate causal SNPs and candidate causal mechanisms of human diseases or traits and to guide future genetic and mechanism study.

OVERVIEW OF THE ICSNPATHWAY APPROACH

The ICSNPathway web server implements a two-stage analysis. The first stage is to pre-select candidate causal SNPs by LD analysis and functional SNP annotation based on the most significant SNPs of GWAS. The second stage is to annotate the biological mechanisms for the pre-selected candidate causal SNPs by using PBA. There are two key basic concepts and one key algorithm applied in ICSNPathway.

One concept is LD analysis, which searches the SNPs in LD with the most significant SNPs of GWAS to ensure to capture more possible candidate causal SNPs based on the extended data set which includes HapMap data (20). The other concept is functional SNPs. ICSNPathway pre-selects candidate causal SNPs based on functional SNPs, which are important for understanding the underlying genetics of human health. Functional SNPs are defined as SNPs that may alter protein, gene expression or the role of protein in context of pathway. The functional SNPs include deleterious and non-deleterious non-synonymous SNPs, SNPs leading to gain or lost of stop codon, SNPs resulting in frame shift, SNPs in essential splice site (the first two bp and last two bp of an intron) and SNPs in regulatory region [including DNase I hypersensitive sites which marks open chromatin, histone modification sites, CCCTC-binding factor (CTCF) sites which characterize insulator/enhancer elements, and transcription factor binding sites (TFBSs)] (21).

The ICSNPathway server implements a PBA algorithm, as named *i*-GSEA (improved-gene set enrichment analysis), on the full list of GWAS SNP *P*-values to detect pathways associated with traits (18). Briefly, (i) each SNP is mapped to its nearest gene according to the SNP and gene localization in Ensembl 61 database (http://www.ensembl.org/biomart/martview) (21), and the maximum $t = -\log(P\text{-value})$ of the SNPs mapped to a gene is assigned to represent the gene. Then all the genes are ranked by decreasing their representation value t. (ii) For each pathway S, ES (enrichment score, i.e. a Kolmogor-Smirnov like running-sum statistics with weight (a)) which measures the tendency that genes of a pathway are located at the top of the ranked gene list, is calculated. (iii) ES is converted to $SPES$ (significant proportion based enrichment score) by multiplying it to m_1/m_2, where m_1 is the proportion of significant genes (defined as genes mapped with at least one of the top 5% most significant SNPs of all SNPs in GWAS) for pathways S and m_2 is the proportion of significant genes for all the genes in the GWAS. (iv) SNP label permutation and normalization are employed to generate the distribution of $SPES$ and to correct gene variation (the bias due to different genes with different number of mapped SNPs) and pathway variation (the bias due to different pathways consisting of different number of genes). (v) Based on all the distributions of $SPESs$ generated by permutation, nominal *P*-value is calculated and false discovery rate (FDR) is computed for multiple testing correction (22).

DESCRIPTION OF THE ICSNPATHWAY WEB SERVER

System development, system overview and data resources

ICSNPathway is written in Java and JSP based on Struts framework and implemented on an Apache web server. AJAX is used for the interface development. The web server is freely available and registration is not required. Besides the web-based browser, users can run ICSNPathway through command lines by using the template program provided in the web site. The overview of ICSNPathway web server is shown in Figure 1.

There are several data resources included in our web server to support the analysis. One is the LD data from HapMap (phases I + II + III, rel #27, downloaded from http://hapmap.ncbi.nlm.nih.gov/downloads/ld_data/2009-04_rel27/) (20) for LD analysis of the most significant SNPs. Another is the SNP function annotation database, which was built based on Ensembl 61 database (http://www.ensembl.org/biomart/martview) (21) and by integrating predictions of deleterious non-synonymous SNPs from PolyPhen-2 (5), SIFT (6) and SNPs3D (7). The third is the pathway database consisting of pathways from KEGG (http://www.genome.jp/kegg/pathway.html) (19), BioCarta (http://www.biocarta.com/genes/index.asp), GO (gene ontology, http://www.geneontology.org/) (23) (level 4 GO terms of biological process domain and molecular function domain) and MSigDB (http://www.broadinstitute.org/gsea/msigdb/index.jsp) v3.0 (24) (curated GO terms of biological process domain and molecular function domain).

Input data

The required input data of ICSNPathway is the full list of GWAS SNP *P*-values (rs-IDs along with corresponding association test *P*-values separated by tab and without head line). To define the most significant SNPs, users may input the *P*-value threshold or use the default parameter. Another optional input is the customized pathway data for PBA. If users make hypothesis that some pathways, which may not be included in our web server, might be associated with certain trait, they can upload the pathway data onto the server. The uploaded pathway data can either be combined with the existing pathway database in our server or be used as a separate pathway database.

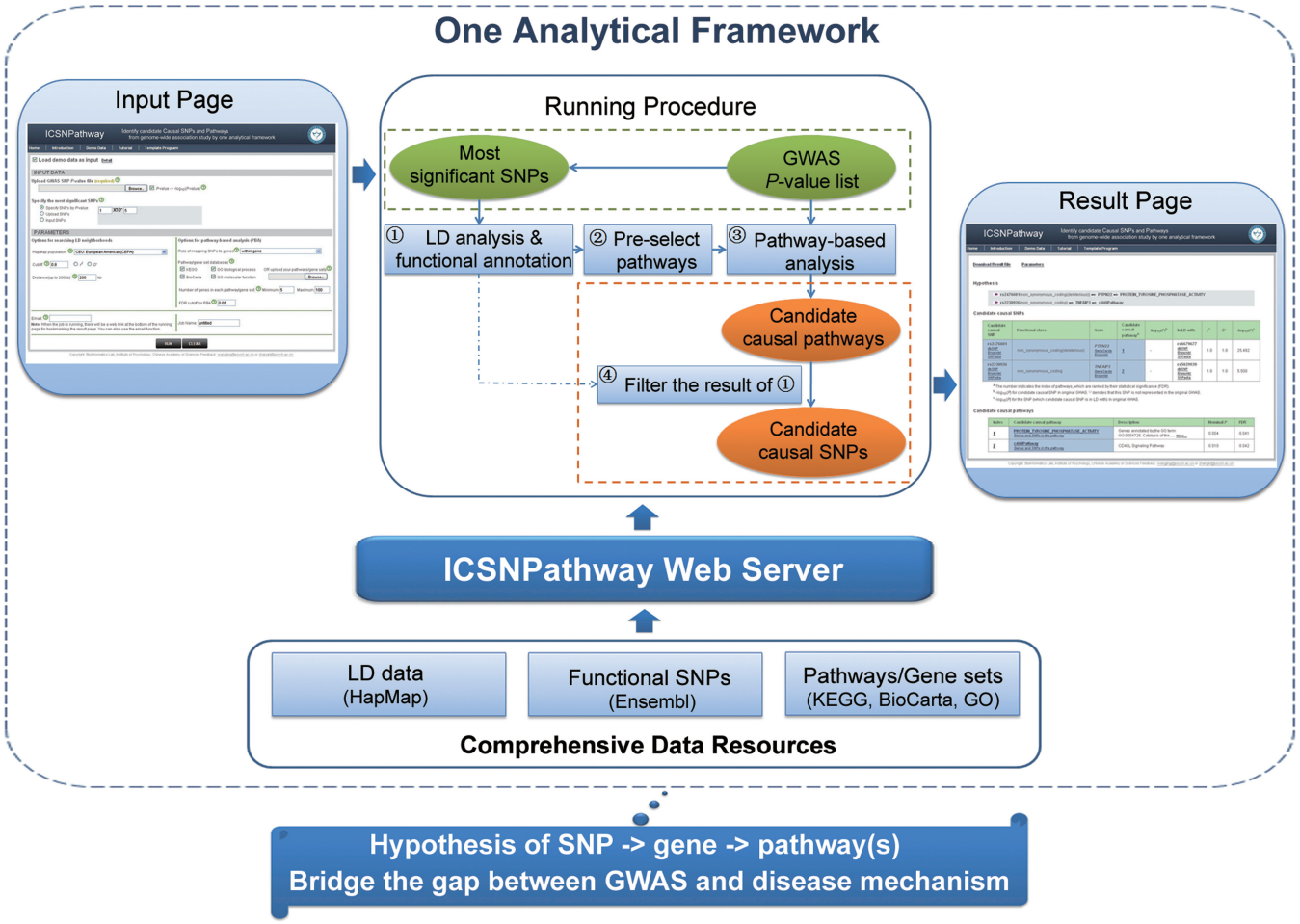

Figure 1. System overview of the ICSNPathway web server.

Running procedure

There are mainly four steps included in the ICSNPathway analytical framework. (i) Search for SNPs in LD with (r^2 or $D' >$ user-defined threshold, default: $r^2 > 0.8$) and located in the flanking region (with distance up to 200 kb up- and downstream) of the most significant SNPs of GWAS; then perform functional annotation on these SNPs (including the most significant SNPs) by querying the internal SNP function annotation database. (ii) Map the functional SNPs to their corresponding genes and then select the pathways containing any one of the mapped genes from the default or customized pathway database as search space for PBA. (iii) Perform PBA using algorithm described above on the full list of GWAS SNP *P*-values. (iv) Identify candidate causal SNPs and corresponding candidate causal pathways to generate study hypothesis of SNP → gene → pathway(s) for disease mechanism. The candidate causal pathways are defined as pathways identified by PBA and the candidate causal SNPs are defined as functional SNPs both in LD with the most significant SNPs and in the candidate causal pathways. With default parameters, usually it will take less than three minutes for a single run which includes ~150 most significant SNPs and ~455 K GWAS SNP *P*-values. The GWAS data uploaded by users will be deleted automatically as soon as the above 4-step analysis is finished.

The main default parameters used in ICSNPathway include: (i) threshold to specify the most significant SNPs: *P*-value $< 10^{-5}$, (ii) HapMap population: CEU, (iii) LD cutoff: $r^2 > 0.8$, (iv)distance for searching LD neighborhoods: 200 kb, (v) rule of mapping SNPs to genes: within gene, (vi) pathway/gene set database: KEGG, BioCarta, GO biological process and GO molecular function, (vii) number of genes in each pathway/gene set: minimum 5 and maximum 100 and (viii) FDR (False Discovery Rate) cutoff for multiple testing correction for PBA: 0.05.

Output and analyzing a GWAS investigation for rheumatoid arthritis

The output includes lists of candidate causal SNPs and corresponding candidate causal pathways with detailed information for each SNP and pathway. The information for each candidate causal SNP includes rs-ID, functional class, corresponding gene, the candidate causal pathway(s) it represents in, its $-\log_{10}$(*P*-value) in original GWAS, the most significant SNP that it is in LD with, r^2, D', and $-\log_{10}$(*P*-value) of the most significant SNP in original

Table 1. Candidate causal SNPs of RA

Candidate causal SNP	Functional class	Gene	Candidate causal pathway[a]	$-\log_{10}(P)$[b]	In LD with	r^2	D'	$-\log_{10}(P)$[c]
rs2476601	non-synonymous (deleterious)	PTPN22	1	–	rs6679677	1.0	1.0	25.5
rs2230926	non-synonymous	TNFAIP3	2	–	rs5029939	1.0	1.0	5.5

[a]The number indicates the index of pathways, which are ranked by their statistical significance (FDR).
[b]$-\log_{10}(P)$ for candidate causal SNP in original GWAS. '-' denotes that this SNP is not represented in the original GWAS.
[c]$-\log_{10}(P)$ for the SNP (which the candidate causal SNP is in LD with) in original GWAS.

Table 2. Candidate causal pathways of RA

Index	Candidate pathway	Nominal *P*	FDR
1	protein tyrosine phosphatase activity (GO:0004725)	0.004	0.041
2	CD40L Signaling Pathway (cd40Pathway)	0.010	0.042

GWAS. The information for each candidate causal pathway includes detailed description, genes, nominal *P*-value and FDR (both are for PBA) to help evaluate the statistical significance of the candidate causal pathway. Hypothesis derived from the above analysis is expressed as [SNP (functional class) → gene → pathway(s)].

As an example, we investigated a GWAS of rheumatoid arthritis (RA) (25). The result of this GWAS includes ~455 K GWAS SNP *P*-values and 154 SNPs with *P*-value $< 10^{-5}$. Utilizing the ~455 K GWAS SNP *P*-values as input and the 154 SNPs as the most significant SNPs, with default parameters, ICSNPathway identified two candidate causal SNPs (rs2476601 and rs2230926) and two candidate causal pathways ('protein tyrosine phosphatase activity' and 'CD40L signaling pathway') (Tables 1 and 2). SNP rs2476601 is in LD with rs6679677 ($r^2 = 1.0$), which is with genome-wide significance in the original GWAS (*P*-value = 3.2×10^{-26}). SNP rs2230926 is in LD with rs5029939 ($r^2 = 1.0$), which does not reach genome-wide significance in the original GWAS (*P*-value = 3.2×10^{-6}). Although rs2476601 and rs2230926 were not presented in the original GWAS for RA, both of them were proved to be associated with RA in additional GWASs for RA (26–29).

The two candidate causal SNPs and two candidate causal pathways indicate two hypotheses of biological mechanisms. One is [rs2476601 (non-synonymous, deleterious) → *PTPN22* → protein tyrosine phosphatase activity] and the other is [rs2230926 (non-synonymous) → *TNFAIP3* → CD40L signaling pathway]. rs2476601(C/T) locates within the coding region of gene *PTPN22*, which encodes lymphoid tyrosine phosphatase (LYP). LYP acts as a key negative regulator of T cell receptor (TCR) signaling. It can phosphorylate and be phosphorylated by protein tyrosine kinase Csk, another negative regulator of TCR signaling. The risk allele T of rs2476601 leads to a R620W substitution within the protein. The substitution affects the interaction between LYP and Csk and reduces phosphorylation of LYP on tyrosine. The change of protein tyrosine phosphatase activity affects TCR signaling and abnormal TCR signaling has been considered as a major risk factor for autoimmunity, such as RA (30). For rs2230926 (T/G), it is a coding SNP in gene *TNFAIP3*, which encodes protein A20, which is a participant of CD40L signaling pathway. A20 has anti-inflammatory activity by inhibiting TNF-induced NF-kB activity in CD40L signaling pathway. The allele G of rs2230926 results in a F127C substitution, which will reduce the effectiveness of A20 to inhibit NF-kB activity and may affect the susceptibility of RA (31). The above two hypotheses were derived by GWAS data interpretation using ICSNPathway and are both well supported by experimental evidences.

DISCUSSION

In order to solve the challenge for GWAS data interpretation, our web server is developed to identify candidate causal SNPs and corresponding candidate causal pathways in one analytical framework. ICSNPathway will help researchers to derive mechanism hypothesis of SNP → gene → pathway(s) for complex disease study. It is well-known that complex disease is caused by multiple genetic factors interacting with environmental factors and complex molecular network and cellular pathways usually play important roles in susceptibility of complex diseases (32). As pathways represent the combined genetic effect, the candidate causal SNPs, which are supported by the pathways associated with traits, are supposed to have much higher confidence to be true than those that are lacking in support of such pathways. Meanwhile, ICSNPathway considers not only the strong association signal of most significant SNPs, but also the combined effect of modest SNPs, which ensures a comprehensive analysis.

It should be noted that ICSNPathway is not intended to be used to predict true causal SNPs and pathways since for complex diseases, due to the limited understanding of their genetic basis, currently there is no concrete evidence to be used to establish the predictive properties (e.g. assessments of false-positive rates) for ICSNPathway. So the outputs of ICSNPathway are candidate causal SNPs and pathways, plus the mechanism hypotheses of SNP → gene → pathway(s) based on them. An important application of the ICSNPathway results is to allow investigators to

Table 3. Summary and comparison of some web tools for GWAS

Web tool	Input	Output		
		proxy SNP	functional SNP annotation	pathway associated with trait
SNAP (4)	a list of SNPs	Yes	No	No
PolyPhen-2 (5)	single non-synonymous SNP	No	Yes	No
SIFT (6)	a list of non-synonymous SNPs	No	Yes	No
SNPs3D (7)	single non-synonymous SNP	No	Yes	No
PANTHER (8)	single non-synonymous SNP	No	Yes	No
FASTSNP (9)	a list of SNPs	No	Yes	No
F-SNP (10)	single SNP	No	Yes	No
CandiSNPer (11)	single SNP	Yes	Yes	No
SPOT (12)	a list of SNPs, with or without *P*-values	Yes	Yes	No
GenomePipe of SNPinfo (13)	a list of GWAS SNP *P*-values	Yes	Yes	No
GeSBAP (16)	full list of GWAS SNP *P*-values or genotype data	No	No	Yes
GSA-SNP (17)[a]	full list of GWAS SNP *P*-values	No	No	Yes
i-GSEA4GWAS (18)	full list of GWAS SNP *P*-values	No	No	Yes
ICSNPathway	full list of GWAS SNP *P*-values	Yes	Yes	Yes

[a]GSA-SNP is a stand-alone tool.

test 'a priori' hypothesis concerning pathways by using candidate causal SNPs as the practical starting point. For such applications, although there is no concrete evidence of prediction, some investigators may still wish to test certain pathway-related hypotheses, particularly if there is some 'a priori' connection between the pathway and the disease of interest.

At the time of writing, there is no web-based tool that performs the same function as ICSNPathway, namely to identify candidate causal SNPs and their corresponding candidate causal pathways from GWAS by integrating linkage disequilibrium (LD) analysis, functional SNP annotation and PBA. The summary and comparison of current GWAS web tools are shown in Table 3, which can be classified into two types. One type of web tool is for identification of candidate causal SNPs or prioritizing SNPs, which includes SNAP (4), PolyPhen-2 (5), SIFT (6), SNPs3D (7), PANTHER (8), FASTSNP (9), F-SNP (10), CandiSNPer (11), SPOT (12) and SNPinfo (13). The SNAP web server employs LD analysis to identify the proxy SNPs of the input SNPs by using HapMap (20). The web tools of PolyPhen-2, SIFT, SNPs3D and PANTHER focus on non-synonymous SNPs to predict their (deleterious) impact on protein. FASTSNP and F-SNP annotate functional information (deleterious non-synonymous, splice site, etc) to the input SNPs. While the web servers like CandiSNPer, SPOT and SNPinfo implement both LD analysis and functional SNP annotation to annotate or prioritize the input SNP(s). The other type of GWAS web tools is for identification of pathways associated with disease, which includes GeSBAP (16), GSA-SNP (17) and *i*-GSEA4GWAS (18). These three tools focus on interpretation of the full list GWAS data by applying three different PBA approaches of segmentation test, a collection of three methods (Z-statistic method, restandardized GSA and GSEA), and *i*-GSEA respectively to identify disease-associated pathways. The first type of web tools provides the output of SNP(s) along with function annotation (except SNAP which does not provide function annotation) to provide the hypothesis of candidate causal SNP → gene, without taking into account the information of disease-associated pathways. While the second type of web tools identifies pathways associated with trait to generate the hypotheses only in pathway level, without analysis on significant SNPs and genes. ICSNPathway will bridge the gap between these two types of web tools by implementing both the analysis of candidate causal SNPs and PBA. The ICSNPathway web server will help improve GWAS data interpretation from variants to biological mechanisms to well guide future biological mechanism studies.

We set two strict default parameters in ICSNPathway for pathway-based analysis. One is for parameter 'rule of mapping SNPs to genes' (default: within gene, which means that only the P-values of the SNPs located within genes are utilized in PBA) and the other is for parameter 'FDR cutoff for multiple testing correction for PBA' (default: 0.05). These default settings ensure that the result of PBA is based on the association signals inside genes and with statistical significance. Users may also adjust the 'rule of mapping SNPs to genes' to '500 kb upstream and downstream of gene' and/or relax the 'FDR cutoff for multiple testing correction for PBA' to 0.25, for example. In this way users may get more candidate causal SNPs and pathways to try the possibility to have more novel findings, but the confidence of the results will be reduced and this will also increase the background noise for the true results.

The premise that ICSNPathway is applicable for a GWAS investigation is that both LD neighborhood(s) of the most significant SNPs from HapMap LD data ($r^2 > 0.8$) and functional information of the LD neighborhood(s) from Ensembl database are available. We evaluated whether our approach could be widely used for the available GWAS investigations. By the end of

March 2011, taking the GWASs for Caucasian population as an example, there have been 3391 SNPs (P-value $< 10^{-5}$) identified by 556 GWAS investigations for 361 traits (http://www.genome.gov/gwastudies). Of these studies, there are a total of 477 SNPs (most significant SNPs plus their LD neighborhoods with functional annotation) in 265 GWASs (~48%) for 185 traits (~51%) that can be analyzed by ICSNPathway. Thus, ICSNPathway is applicable to a high proportion of available GWASs. The most significant SNPs, which represent most significant association signals detected by GWAS, are utilized in the initial step of the ICSNPathway analysis for searching candidate causal SNPs. However there are usually a limited number of such SNPs for a single GWAS, leading to the possible limitation while searching for candidate causal SNPs. To solve this, a possible way is to loose the threshold of P-value while specifying the most significant SNPs or extend the most significant SNPs to SNPs identified by other GWASs for the same trait or any other SNPs that are considered to be associated with trait according to users' knowledge. So users may use the customized most significant SNPs as input (ICSNPathway provides this option while the input of the full list of GWAS SNP P-values is mandatory). Thus, ICSNPathway is extendable to help researchers to test and find proofs for their own hypothesis.

ICSNPathway will be regularly updated to ensure the most up-to-date data resources. In the future, ICSNPathway will be extended to fulfill more functions with more user-friendly options. For example, the function of LD neighborhood searching will be implemented by calculating LD using HapMap genotype data, so that users can search for LD neighborhoods of the most significant SNPs with a more flexible range. In summary, ICSNPathway represents a feasible solution for identifying both candidate causal SNPs and their corresponding candidate causal pathways to bridge the gap between GWAS and biological mechanism study of complex disease.

ACKNOWLEDGEMENTS

We thank all our colleagues and friends in the Chinese Academy of Sciences and from different Universities both inside and outside China who helped us test the web server and provided the valuable suggestions. We thank the anonymous reviewers for their valuable comments and suggestions to help us improve both the web server and the manuscript.

FUNDING

Knowledge Innovation Program of the Chinese Academy of Sciences (KSCX2-EW-J-8); Project for Young Scientists Fund, Institute of Psychology, Chinese Academy of Sciences (O9CX115011). Funding for open access charge: KSCX2-EW-J-8.

Conflict of interest statement. None declared.

REFERENCES

1. McCarthy,M.I., Abecasis,G.R., Cardon,L.R., Goldstein,D.B., Little,J., Ioannidis,J.P. and Hirschhorn,J.N. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.*, **9**, 356–369.
2. Hindorff,L.A., Sethupathy,P., Junkins,H.A., Ramos,E.M., Mehta,J.P., Collins,F.S. and Manolio,T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA*, **106**, 9362–9367.
3. McCarthy,M.I. and Hirschhorn,J.N. (2008) Genome-wide association studies: potential next steps on a genetic journey. *Hum. Mol. Genet.*, **17**, R156–R165.
4. Johnson,A.D., Handsaker,R.E., Pulit,S.L., Nizzari,M.M., O'Donnell,C.J. and de Bakker,P.I. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*, **24**, 2938–2939.
5. Adzhubei,I.A., Schmidt,S., Peshkin,L., Ramensky,V.E., Gerasimova,A., Bork,P., Kondrashov,A.S. and Sunyaev,S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
6. Kumar,P., Henikoff,S. and Ng,P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.*, **4**, 1073–1081.
7. Yue,P., Melamud,E. and Moult,J. (2006) SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinformatics*, **7**, 166.
8. Mi,H., Dong,Q., Muruganujan,A., Gaudet,P., Lewis,S. and Thomas,P.D. (2010) PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.*, **38**, D204–D210.
9. Yuan,H.Y., Chiou,J.J., Tseng,W.H., Liu,C.H., Liu,C.K., Lin,Y.J., Wang,H.H., Yao,A., Chen,Y.T. and Hsu,C.N. (2006) FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res.*, **34**, W635–W641.
10. Lee,P.H. and Shatkay,H. (2008) F-SNP: computationally predicted functional SNPs for disease association studies. *Nucleic Acids Res.*, **36**, D820–D824.
11. Schmitt,A.O., Assmus,J., Bortfeldt,R.H. and Brockmann,G.A. (2010) CandiSNPer: a web tool for the identification of candidate SNPs for causal variants. *Bioinformatics*, **26**, 969–970.
12. Saccone,S.F., Bolze,R., Thomas,P., Quan,J., Mehta,G., Deelman,E., Tischfield,J.A. and Rice,J.P. (2010) SPOT: a web-based tool for using biological databases to prioritize SNPs after a genome-wide association study. *Nucleic Acids Res.*, **38**, W201–W209.
13. Xu,Z. and Taylor,J.A. (2009) SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res.*, **37**, W600–W605.
14. Wang,K., Li,M. and Hakonarson,H. (2010) Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet.*, **11**, 843–854.
15. Cantor,R.M., Lange,K. and Sinsheimer,J.S. (2010) Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *Am. J. Hum. Genet.*, **86**, 6–22.
16. Medina,I., Montaner,D., Bonifaci,N., Pujana,M.A., Carbonell,J., Tarraga,J., Al-Shahrour,F. and Dopazo,J. (2009) Gene set-based analysis of polymorphisms: finding pathways or biological processes associated to traits in genome-wide association studies. *Nucleic Acids Res.*, **37**, W340–W344.
17. Nam,D., Kim,J., Kim,S.Y. and Kim,S. (2010) GSA-SNP: a general approach for gene set analysis of polymorphisms. *Nucleic Acids Res.*, **38**, W749–W754.
18. 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–W95.
19. 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–D360.

20. Altshuler,D.M., Gibbs,R.A., Peltonen,L., Dermitzakis,E., Schaffner,S.F., Yu,F., Bonnen,P.E., de Bakker,P.I., Deloukas,P., Gabriel,S.B. *et al.* (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**, 52–58.
21. Flicek,P., Aken,B.L., Ballester,B., Beal,K., Bragin,E., Brent,S., Chen,Y., Clapham,P., Coates,G., Fairley,S. *et al.* (2010) Ensembl's 10th year. *Nucleic Acids Res.*, **38**, D557–D562.
22. Reiner,A., Yekutieli,D. and Benjamini,Y. (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*, **19**, 368–375.
23. 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.
24. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*, **102**, 15545–15550.
25. WTCCC. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.
26. Plenge,R.M., Seielstad,M., Padyukov,L., Lee,A.T., Remmers,E.F., Ding,B., Liew,A., Khalili,H., Chandrasekaran,A., Davies,L.R. *et al.* (2007) TRAF1-C5 as a risk locus for rheumatoid arthritis–a genomewide study. *N. Engl. J. Med.*, **357**, 1199–1209.
27. Gregersen,P.K., Amos,C.I., Lee,A.T., Lu,Y., Remmers,E.F., Kastner,D.L., Seldin,M.F., Criswell,L.A., Plenge,R.M., Holers,V.M. *et al.* (2009) REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat. Genet.*, **41**, 820–823.
28. Stahl,E.A., Raychaudhuri,S., Remmers,E.F., Xie,G., Eyre,S., Thomson,B.P., Li,Y., Kurreeman,F.A., Zhernakova,A., Hinks,A. *et al.* (2010) Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.*, **42**, 508–514.
29. Kochi,Y., Okada,Y., Suzuki,A., Ikari,K., Terao,C., Takahashi,A., Yamazaki,K., Hosono,N., Myouzen,K., Tsunoda,T. *et al.* (2010) A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. *Nat. Genet.*, **42**, 515–519.
30. Fiorillo,E., Orru,V., Stanford,S.M., Liu,Y., Salek,M., Rapini,N., Schenone,A.D., Saccucci,P., Delogu,L.G., Angelini,F. *et al.* (2010) Autoimmune-associated PTPN22 R620W variation reduces phosphorylation of lymphoid phosphatase on an inhibitory tyrosine residue. *J. Biol. Chem.*, **285**, 26506–26518.
31. Musone,S.L., Taylor,K.E., Lu,T.T., Nititham,J., Ferreira,R.C., Ortmann,W., Shifrin,N., Petri,M.A., Kamboh,M.I., Manzi,S. *et al.* (2008) Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. *Nat. Genet.*, **40**, 1062–1064.
32. Schadt,E.E. (2009) Molecular networks as sensors and drivers of common human diseases. *Nature*, **461**, 218–223.