BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 9 2014, pages 1287–1289
doi:10.1093/bioinformatics/btt743

Genome analysis

Advance Access publication January 11, 2014

PANOGA: a web server for identification of SNP-targeted pathways from genome-wide association study data

Burcu Bakir-Gungor[1,2,*], Ece Egemen[3] and Osman Ugur Sezerman[4]

[1]Department of Genetics and Bioinformatics, Faculty of Arts and Sciences, Bahcesehir University, 34353, Besiktas, Istanbul, Turkey, [2]Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Abdullah Gul University, 38039, Kayseri, Turkey, [3]Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Sabanci University, 34956, Tuzla, Istanbul, Turkey and [4]Department of Biological Sciences and Bioengineering, Faculty of Engineering and Natural Sciences, Sabanci University, 34956, Tuzla, Istanbul, Turkey

Associate Editor: John Hancock

ABSTRACT

Summary: Genome-wide association studies (GWAS) have revolutionized the search for the variants underlying human complex diseases. However, in a typical GWAS, only a minority of the single-nucleotide polymorphisms (SNPs) with the strongest evidence of association is explained. One possible reason of complex diseases is the alterations in the activity of several biological pathways. Here we present a web server called Pathway and Network-Oriented GWAS Analysis to devise functionally important pathways through the identification of SNP-targeted genes within these pathways. The strength of our methodology stems from its multidimensional perspective, where we combine evidence from the following five resources: (i) genetic association information obtained through GWAS, (ii) SNP functional information, (iii) protein–protein interaction network, (iv) linkage disequilibrium and (v) biochemical pathways.

Availability: PANOGA web server is freely available at: http://panoga.sabanciuniv.edu/. The source code is available to academic users 'as is' on request.

Contact: burcub@gatech.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

Received on July 19, 2013; revised on December 16, 2013; accepted on December 18, 2013

1 INTRODUCTION

Genome-wide association studies (GWAS) aim to discover a subset of single-nucleotide polymorphisms (SNPs) that are associated with the onset and progression of complex disease phenotypes at a genome-wide scale. These studies have produced extraordinary successes in the past few years, with numerous new genetic variants identified with levels of significance that were once thought impossible to reach (Visscher *et al.*, 2012). Standard GWAS approaches focus on single SNP analysis, and hence, it is difficult to incorporate weakly associated SNPs (Ball, 2011). For multifactorial diseases, this incorporation is especially important because no particular variant on a particular gene may have a strong effect, but the combination of the multiple variants with small effects on the polygenic disease explains the overall susceptibility to disease (Cirulli and Goldstein, 2010). In this regard, pathway-based approaches to GWAS, in which prior biological knowledge is integrated into association studies and sets of genes are analyzed simultaneously for association with a disease, have been proposed recently (Peng *et al.*, 2010).

Because the analysis of single variants within isolated genes is not informative enough to explain the underlying disease mechanisms, another recent trend to further mine GWAS data is to integrate protein–protein interaction (PPI) networks (Schadt and Bjorkegren, 2012). Following a systems biology approach, these methods evaluate the pairwise or collective effects of genetic variants, where the functional consequence of each susceptibility allele may interfere with the interaction between pairs of proteins. A few studies available to date on the identification of the molecular interaction networks associated with the GWAS genes indicated that this approach can help to discover formerly unknown mechanisms underlying human diseases at the molecular level (Bebek *et al.*, 2012).

To better understand the biological processes underlying complex diseases, we also considered the functional effect of a typed SNP in GWAS. Although the DNA polymorphisms that change protein function can have significant consequences, other types of SNPs, such as synonymous SNPs, do not have such serious effects in disease development mechanism. Functionally important SNPs, such as those that change amino acids, splicing sites; those that lead to gain or loss of stop codon; those that result in frame shift; and those that are found in regulatory region, are priority targets in disease studies and large-scale genotyping projects (Calabrese *et al.*, 2009). Hence, we decided that SNP functional knowledge is valuable information to strengthen our pathway and network-oriented GWAS analysis (PANOGA) method.

To support the steps from SNP functionalization into PPI network analysis to functional enrichment in a single program, we developed a web server for post-GWAS analysis called PANOGA. With its multifactorial basis, PANOGA has a good potential to decipher the combination of biological processes underlying disease. PANOGA has been tested on several complex diseases, including rheumatoid arthritis (Bakir-Gungor and Sezerman, 2011), intracranial aneurysm (Bakir-Gungor and Sezerman, 2013), epilepsy (Bakir-Gungor *et al.*, 2013) and Behcet's disease (Bakir-Gungor *et al.*, 2012), and proved to be

*To whom correspondence should be addressed.

© The Author 2014. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/bioinformatics/article-abstract/30/9/1287/234082 by guest on 02 April 2020

useful. Throughout the time, the method has evolved with our efforts (Bakir-Gungor and Sezerman, 2012). This web server presents the latest version of PANOGA.

2 METHODS

For a given SNP that was found significant in a GWAS, first, the functional scores are fetched from The Functional Single Nucleotide Polymorphism (F-SNP) (Lee and Shatkay, 2008) and SPOT (Saccone *et al.*, 2010) servers. F-SNP obtains the functional consequence of a SNP from multiple tools at splicing, transcriptional, translational and post-translational levels. SPOT score takes into account the effect of a SNP on the gene transcript (e.g. non-sense, frameshift and 5′- and 3′-UTR), the impact of an amino acid substitution on the protein product, evolutionary conserved regions and all possible linkage disequilibrium proxy SNPs with r2 over a predefined threshold in a specific HapMap sample (Saccone *et al.*, 2010). Second, PANOGA combines the functional score and the GWAS *P*-value of a SNP and obtains P_w-values. Among all known SNP/gene transcript designations, the SNP is associated with the gene on which the SNP has the most important functional effect. Third, PANOGA transfers P_w-values of SNP to its associated gene.

During the network-oriented steps of PANOGA, genes with P_w-values are mapped into a human PPI network. Next, PANOGA finds connected sub-networks of the PPI network that has high total significance of genotypic *P*-values of the disease-predisposing SNPs with respect to the controls. To check for the biological relevance of the identified sub-networks, PANOGA performs functional enrichment. Finally, PANOGA integrates the functional enrichments of the generated sub-networks. The overview of PANOGA methodology is presented in Supplementary Figure S1. The Web site is implemented in HTML/PHP; operations are made with Java on the server side.

3 SERVER DESCRIPTION

PANOGA web server accepts GWAS results including reference SNP ID (rs#) and genotypic *P*-value of association for each tested SNP, as shown in Figure 1A. It is important to note that the PANOGA web server does not require individual genotypes, odds ratio, minor allele frequency or confidence intervals computed in a GWAS, which can have ethical considerations.

PANOGA outputs SNP-targeted pathway and gene tables (Supplementary Tables S1–S4), and customized pathway maps, as shown in Figure 1B. SNP-targeted pathway tables of PANOGA present the number of associated SNPs from GWAS, the number of regulatory SNPs among the disease-predisposing SNPs, the number of SNP-targeted genes and the number of sub-networks that this pathway is found to be statistically significant.

The gene table of PANOGA includes the genes that are part of a functionally important pathway and at the same time identified in a sub-network. For each such gene, this table presents gene symbols, Entrez gene ID, number of times found in sub-network, number of associated pathways, list of associated pathways, number of typed SNPs in GWAS, functional information of the typed SNPs in GWAS, SNP regulatory potential and number of regulatory SNPs.

Customized Kyoto Encyclopedia of Genes and Genomes pathway maps of PANOGA, as shown in Figure 1C, enrich the utility of PANOGA results. These pathway maps help the users to visualize affected genes along different routes within the pathway map.

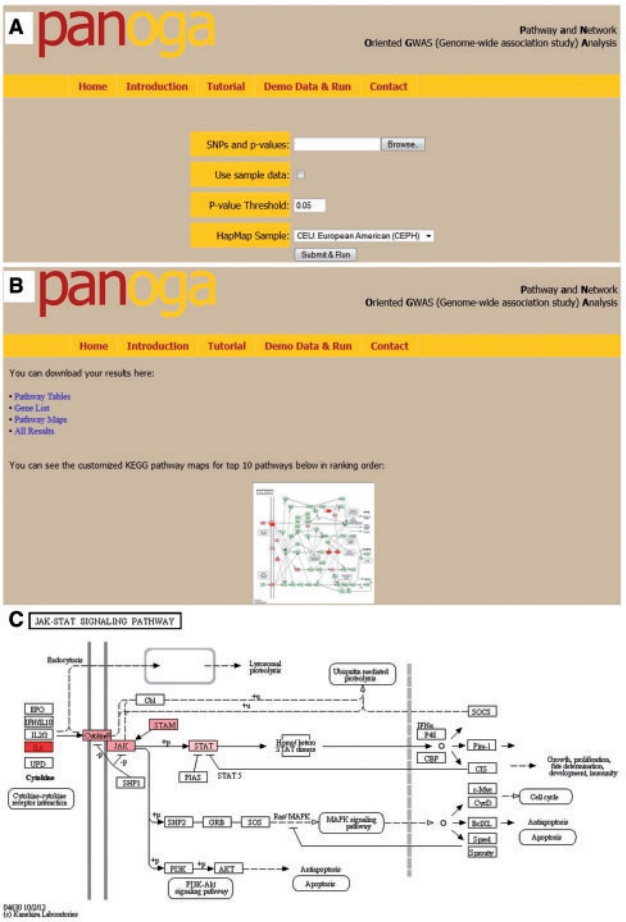

Fig. 1. A snapshot of the web server input (**A**) and results page (**B**). A link from the results page for customized Kyoto Encyclopedia of Genes and Genomes pathway maps opens the zoomed-in version for visual display (**C**). The shade of red color in genes indicates the number of targeted SNPs (typed in the GWAS of disease) per base pair of the gene. Red refers to the highest targeted gene, whereas white refers to a gene product not targeted by the SNPs

4 CONCLUSIONS

PANOGA is a web server with multiple functions that support pathway and network-oriented SNP–phenotype association analysis in a single system without requiring any additional manual processes. PANOGA represents a feasible solution for the identification of SNP-targeted pathways (candidate causal pathways) to bridge the gap between GWAS and biological mechanisms of complex diseases.

Funding: The work of B.B.G. has been supported by the Abdullah Gül University Support Foundation (AGUV).

Conflict of Interest: none declared.

REFERENCES

Bakir-Gungor,B. and Sezerman,O.U. (2013) The identification of pathway markers in intracranial aneurysm using genome-wide association data from two different populations. *Plos One*, **8**, e57022.

Downloaded from https://academic.oup.com/bioinformatics/article-abstract/30/9/1287/234082 by guest on 02 April 2020

Bakir-Gungor,B. *et al.* (2013) Identifying SNP targeted pathways in partial epilepsies with genome-wide association study data. *Epilepsy Res.*, **105**, 92–102.

Bakir-Gungor,B. *et al.* (2012) *The Identification of Pathway Markers in Behçet's Disease Using GWA Data From Two Different Populations, Scientific Meeting of the American College of Rheumatology, November 9–14*. Washington, DC.

Bakir-Gungor,B. and Sezerman,O.U. (2012) Identification of SNP targeted pathways from GWAS data. *Nature Protocol Exchange*, [Epub ahead of print, doi:10.1038/protex.2012.019].

Bakir-Gungor,B. and Sezerman,O.U. (2011) A new methodology to associate SNPs with human diseases according to their pathway related context. *PLos One*, **6**, e26277.

Ball,R.D. (2011) Experimental designs for robust detection of effects in genome-wide case-control studies. *Genetics*, **189**, 1497–1514.

Bebek,G. *et al.* (2012) Network biology methods integrating biological data for translational science. *Brief Bioinform.*, **13**, 446–459.

Calabrese,R. *et al.* (2009) Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum. Mutat.*, **30**, 1237–1244.

Cirulli,E.T. and Goldstein,D.B. (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing, *Nat. Rev. Genet.*, **11**, 415–425.

Lee,P.H. and Shatkay,H. (2008) F-SNP: computationally predicted functional SNPs for disease association studies. *Nucleic Acids Res.*, **36**, D820–D824.

Peng,G. *et al.* (2010) Gene and pathway-based second-wave analysis of genome-wide association studies. *Eur. J. Hum. Genet.*, **18**, 111–117.

Saccone,S.F. *et al.* (2010) SPOT: a web-based tool for using biological databases to prioritize SNPs after a GWA study. *Nucleic Acids Res.*, **38**, W201–W209.

Schadt,E.E. and Bjorkegren,J.L.M. (2012) NEW: network-enabled wisdom in biology, medicine, and health care. *Sci. Transl. Med.*, **4**, 115rv1.

Visscher,P.M. *et al.* (2012) Five years of GWAS discovery. *Am. J. Hum. Genet.*, **90**, 7–24.

Downloaded from https://academic.oup.com/bioinformatics/article-abstract/30/9/1287/234082 by guest on 02 April 2020