

EXPLORING CANDIDATE GENES FOR EPILEPSY BY COMPUTATIONAL DISEASE-GENE IDENTIFICATION STRATEGY

Sha Y*, Liu Q, Wang Y, Dong C, Song L

***Corresponding Author:** Ying Sha, Department of Neurology, First Hospital of Jilin University, Jilin Province, People's Republic of China; Tel.: +86-1387-878-9232; Fax: +86-1387-778-7078; E-mail: shaying2010@126.com

ABSTRACT

Epilepsy is a complex disease with a strong genetic component. So far, studies have focused on experimental validation or genome-wide linkage scans for epilepsy susceptibility genes in multiple populations. We have used four bioinformatic tools (SNPs3D, PROSPECTR and SUSPECTS, GenWanderer, PosMed) to analyze 16 susceptibility loci selected from a literature search. Pathways and regulatory network analyses were performed using the Ingenuity Pathways Analysis (IPA) software. We identified a subset of 48 candidate epilepsy susceptibility genes. Five significant canonical pathways, in four typical networks, were identified: GABA receptor signaling, interleukin-6 (IL-6) signaling, G-protein coupled receptor signaling, type 2 diabetes mellitus signaling and airway inflammation in asthma. We concluded that online analytical tools provide a powerful way to reveal candidate genes which can greatly reduce experimental time. Our study contributes to further experimental tests for epilepsy susceptibility genes.

Keywords: Epilepsy, Susceptibility genes, Prioritization, Pathways and networks

Department of Neurology, First Hospital of Jilin University, Jilin Province, People's Republic of China

INTRODUCTION

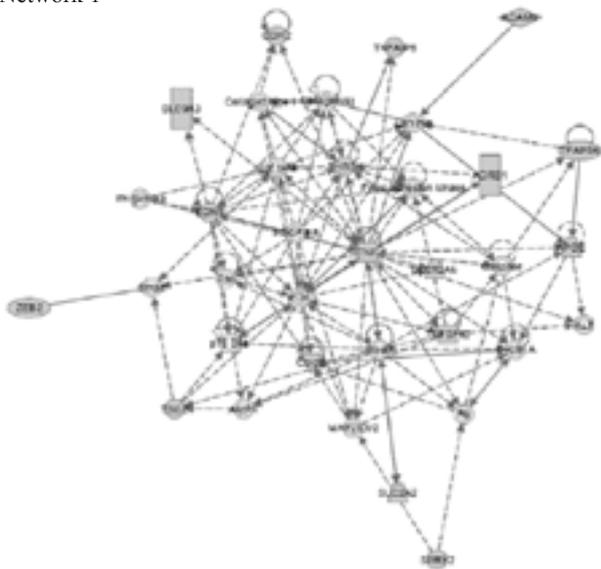
Complex diseases such as epilepsy are influenced by multiple genes with small individual effects, gene interactions and the environment [1]. Epilepsy is a complex disease with a strong genetic component. So far, studies have focused on experimental validation or genome-wide linkage scans for epilepsy susceptibility genes in several populations [5-8].

Promising bioinformatic tools are available for disease-gene identification, which use genomic information extracted from public online databases. These tools have been successfully used to prioritize candidate genes for type 2 diabetes mellitus and for obesity [2-4]. We have applied four bioinformatic tools to prioritize candidate epilepsy susceptibility genes susceptible to epilepsy and subjected the selected genes to pathway analysis and network construction.

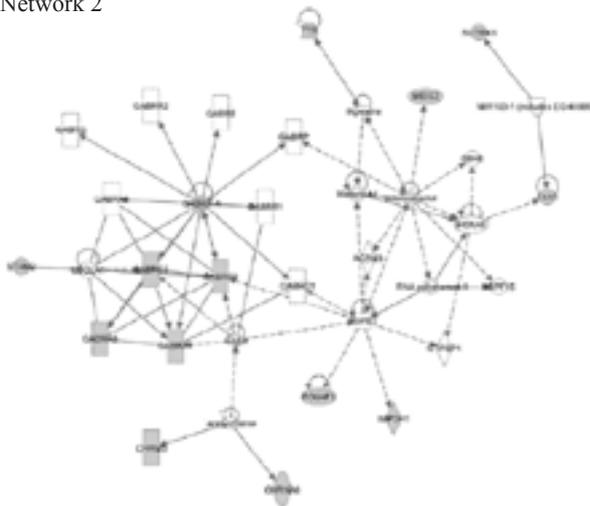
MATERIALS AND METHODS

To date, genome-wide linkage scans across multiple populations have been performed to identify underlying quantitative trait loci. We defined as a significant susceptibility locus with a logarithm of odds ratio score of 3.0 or more, as a suggestive locus with a score between 2.2 and 3.0, and as a nominal locus with a score between 1 and 2.2 [3]. We considered

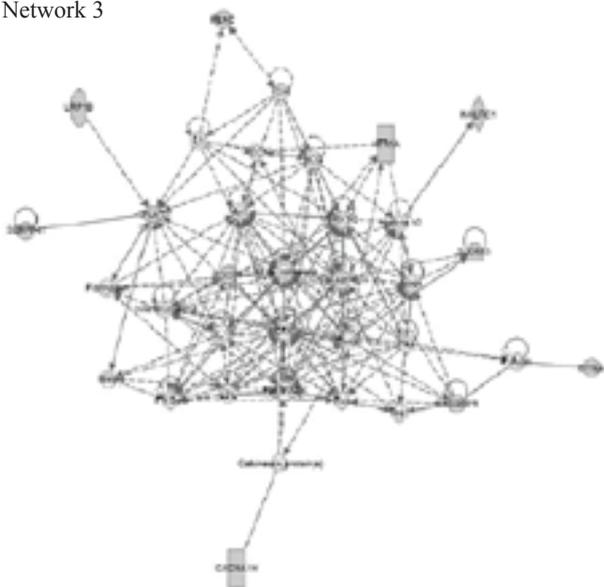
Network 1



Network 2



Network 3



Network 4

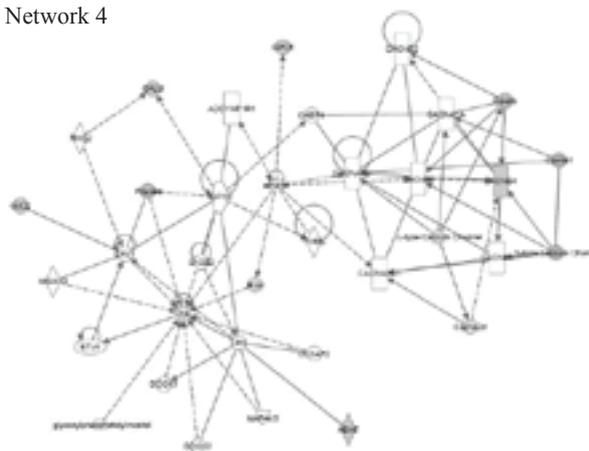


Figure 1. Canonical pathways, the shaped genes are the potential candidate genes identified by the online tools, and others are those associated with the selected candidate genes based on pathway analysis. Network 1. Network 2. Network 3. Network 4.

only the following suggestive and significant loci: 7q32, 16p13, 5p15, 18q12, 8p11-12, 6p22-p11, 6p24, 15q14, 2q22-23, 5q34, 3q26, 10q25-q26, 13q31, 19q13, 5q12-q14, and 10q21-q22 [5-8].

To identify epilepsy-related candidate genes, we used the online tools: SNPs3D [9], PROSPECTR and SUSPECTS (PandS) [10], PosMed [11], and GeneWanderer [12]. We used the following known susceptible genes on GeneWanderer and PandS as a training data set: CHRNA4, CHRNA2, CHRNA2, CHRNA2, KCNQ2, KCNQ3, SCN1A, SCN2A, SCN1B,

GABRA1, GABRG2, CLCN2, LGI1, EFHC1, CHRNA3, and CACNB4 [5-7]. For performing all the online tools, one necessary setting is to provide the start and end locations in the base pair of the corresponding loci.

SNPs3D pinpoints all possible genes regardless of specific loci. The other tools produce rankings, and we used only the top 10 genes from each of the tools. We considered a gene to be an interesting candidate if it was indicated by two or more of the tools.

Identified candidate genes were imported into the Ingenuity Pathways Analysis (IPA) 5.0 to generate putative signaling networks based on pathway interactions extracted from the literature [1]. A network was generated for the input genes using direct and indirect relationships. The networks were ranked by scores that measured the probability that the genes were included in the network by chance alone. Networks with scores >3.0 have a 99.9% confidence of not being generated by chance. Overlapping networks were merged to produce the largest possible network. Correctness of relationships was checked manually on the basis of categorized literature provided by the system. Canonical molecular and cellular pathways associated with the identified genes were elucidated with a statistical significance value (p value of <0.05).

RESULTS AND DISCUSSION

Forty-eight genes were selected as potential epilepsy susceptibility genes (Table 1). Interesting genes indicated by all the tools are known to be associated with epilepsy, such as APOE, GABRA1 and GABRG2. GABRA1 and GABRG2 are subunits of the ligand-gated chloride channel receptors of GABA, the major inhibitory neurotransmitter in the mammalian brain. For the APOE gene, many studies have found that its polymorphism is strongly correlated with different kinds of epilepsy. The ApoE- ϵ allele interacts with longstanding seizures to affect verbal and nonverbal memory performance in patients with medically intractable temporal lobe epilepsy [13]. This allele also increases the risk of postictal confusion [14]. Salzmann *et al.* [15] confirmed that the role of APOE in the temporal lobe epilepsy. In our study, all these widely referred genes were discovered by most of the bioinformatic tools.

Canonical Pathways. Four canonical pathways were selected by the identified candidate genes (Table 2) with significant confidence: GABA receptor signaling, interleukin-6 (IL-6) signaling, G-protein coupled receptor signaling, type 2 diabetes mellitus signaling and airway inflammation in asthma. Here, 35 of the 48 identified genes were used in the pathway analysis of the IPA system. These genes generated four significant networks (Figure 1): Network 1 with a score of 38 and 17 focus genes,

which is related to carbohydrate metabolism, small molecule biochemistry and cellular movement; Network 2 with a score of 27 and 13 focus genes, which is related to psychological disorders, genetic disorder and neurological disease according to the description in the IPA system; Network 3 with a score of 22 and 11 focus genes, related to cardiovascular disease, neurological disease and drug metabolism; and Network 4 with a score of eight and five focus genes, related to cell signaling, molecular transport, vitamin and mineral metabolism.

Comparison of Online Analytical Tools.

There was wide variation between the tools regarding which genes were prioritized, and their rank orders. PosMed and PandS were most similar in their prioritization.

A combination of two or all of the tools was superior for ranking positional candidates [16]. Some methods have a significant identity of candidate genes, which may affect the accuracy of gene prioritization. PanS, GenWanderer and PosMed use the same input information and show a high degree of similarity on their outputs. The selection of a candidate gene by several methods using the same input data may be less valuable than the selection of a candidate by several methods using disparate data sources [1]. In this study, we excluded candidate genes that were solely identified by two of the online tools, *i.e.*, GenWanderer and PosMed. We also restricted the selection to only the top 10 genes as candidate list so as to provide more convergent results.

We concluded that bioinformatic tools were helpful in the hunt for complex disease genes. Our list of most likely candidate genes and the associated pathways should assist further experimental design and analysis, and therefore, our understanding of the pathogenesis of epilepsy.

In addition, we have to say that using different online analytical tools might generate different results, and we selected the tools that were successfully applied to previous studies. We only selected four tools from a number of available resources, all of which easily perform the analyses, fast to obtain results and the corresponding results are reliable. In fact, there are some more tools widely applied in previous studies. However, due to a variety of reasons, we did not adopt them into our study. For example, GeneSeeker [17] is not currently operating and we cannot obtain a complete result;

Table 1. Final 48 candidate epilepsy susceptibility genes identified by two or more online tools

Genes	SNPs3D	PosMed	Gene Wanderer	PandS
ABAT	•	•		•
ADAM9		•	•	•
ADRB1	•	•	•	
APOE	•	•	•	
BCHE		•		
CACNA1H	•		•	
CACNB4	•	•	•	•
CALU		•	•	•
CDC2	•		•	
CDH2		•	•	•
CHRM5		•	•	•
CHRNA6	•	•		
DTNA		•	•	•
EDN1		•	•	•
FLNC		•	•	•
FXR1		•		
GABRA1	•	•	•	•
GABRA6	•	•	•	•
GABRB2	•	•	•	•
GABRG2	•	•	•	•
GJD2	•	•		
GPC5		•		
HTR1A	•	•		
IMPDH1		•	•	•
LRP1B		•	•	•
MEIS2		•	•	•
NHLRC1	•	•		
PLAT		•	•	•
POU4F1		•		
RASGRP1		•	•	•
RPL3L	•		•	
SERPINI1	•			•
SLC12A6	•	•	•	•
SLC2A2		•		
SLC6A3	•	•		
SLC9A3		•	•	•

Table 1 / continue

SLITRK1		•		
SOCS1	•		•	
SPRY2		•		
STAM2		•		
TFAP2A		•	•	•
THBS1		•	•	•
TNF	•	•	•	
TNFAIP6		•	•	•
TNFSF10	•	•	•	
TSC2	•	•		
TTR	•	•	•	•
ZEB2	•	•	•	

Table 1/ end

Table 2. Canonical cellular signaling pathways and genes identified by Ingenuity Pathways Analysis software

Canonical Pathways	p Value	Log (p)	Genes
GABA receptor	3.60E-07	6.443697	ABAT, GABRA1, GABRA6, GABRB2, GABRG2
IL-6	2.81E-03	2.551294	SOCS1, TNF, TNFAIP6
G-protein coupled receptor	4.18E-03	2.378824	ADRB1, CHRM5, HTR1A, RASGRP1
Type 2 diabetes mellitus	5.71E-03	2.243364	SLC2A2, SOCS1, TNF
Airway inflammation in asthma	1.29E-02	1.88941	TNF

Endeavor is very slow during the prioritization process; Prioritizer [18] needs to download a large size of full software and associated data, while the output text for the G2D tool [19] is not easily edited for comparison.

REFERENCES

- Huang QY, Li GHY, Cheung WMW, Song YQ. Prediction of osteoporosis candidate genes by computational disease-gene identification strategy. *J Hum Genet.* 2008; 53(7): 644-655.
- Teber ET, Liu JY, Ballouz S, Fatkin D, Wouters MA. Comparison of automated candidate gene prediction systems using genes implicated in type 2 diabetes by genome-wide association studies. *BMC Bioinformatics.* 2009; 10(1): S69.
- Elbers CC, Onland-Moret NC, Franke L, Niehoff AG, van der Schouw YT, Wijmenga C. A strategy to search for common obesity and type 2 diabetes genes. *Trends Endocrin Met.* 2006; 18(1): 19-26.
- Tiffin N, Adie E, Turner F, Brunner H, van Driel M, Oti M, Lopez-Bigas N, Ouzounis C, Perez-Iratxeta C, Andrade-Navarro M, Adeyemo A, Patti M, Semple CA, Hide W. Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes. *Nucleic Acids Res.* 2006; 34(10): 3067-3081.
- Durner M, Keddache MA, Tomasini L, Shinnar S, Resor SR, Cohen J, Harden C, Moshe SL, Rosenbaum D, Kang H, Ballaban-Gil K, Hertz S, Labar DR, Luciano D, Wallace S, Yohai D, Klotz I, Dicker E, Greenberg DA. Genome scan of idiopathic generalized epilepsy: evidence for major susceptibility gene and modifying genes influencing the seizure type. *Ann Neurol.* 2001; 3(3): 328-335.
- Pinto D, Westland B, de Haan GJ, Rudolf G, da Silva BM, Hirsch E, Lindhout D, Trenité DG, Koeleman BP. Genome-wide linkage scan of epilepsy-related photoparoxysmal electroencephalographic response: evidence for linkage on chromosomes 7q32 and 16p13. *Hum Mol Genet.* 2005; 14(1): 171-178.

7. Delgado-Escueta AV. Advances in genetics of juvenile myoclonic epilepsies. *Epilepsy Curr.* 2007; 7(3): 61-67.
8. Kapoor A, Ratnapriya R, Kuruttukulam G, Anand A. A novel genetic locus for juvenile myoclonic epilepsy at chromosome 5q12-q14. *Hum Genet.* 2007; 121(6): 655-662.
9. Yue P, Melamud E, Moulton J. SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinformatics.* 2006; 7: 166.
10. Adie EA, Adamns RR, Evans KL, Porteous DJ, Pickard BS. SUSPECTS: enabling fast and effective prioritization of positional candidates. *Bioinformatics.* 2006; 22(6): 773-774.
11. Yoshida Y, Makita Y, Heida N, Asano S, Matsushima A, Ishii M, Mochizuki Y, Masuya H, Wakana S, Kobayashi N, Toyoda T. PosMed (Positional Medline): prioritizing genes with an artificial neural network comprising medical documents to accelerate positional cloning. *Nucleic Acids Res.* 2009; 37(Web Server issue): W147-W152.
12. Köhler S, Bauer S, Horn D, Robinson PN. Walking the interactome for prioritization of candidate disease genes. *Am J Hum Genet.* 2008; 82(4): 949-958.
13. Busch RM, Lineweaver TT, Naugle RI, Kim KH, Gong Y, Tillelli CQ, Prayson RA, Bingaman W, Najm IM, Diaz-Arrastia R. ApoE-ε4 is associated with reduced memory in long-standing intractable temporal lobe epilepsy. *Neurology.* 2007; 6(6): 409-414.
14. Chapin JS, Busch RM, Janigro D, Dougherty M, Tillelli CQ, Lineweaver TT, Naugle RI, Diaz-Arrastia R, Najm IM. APOE-ε4 is associated with postictal confusion in patients with medically refractory temporal lobe epilepsy. *Epilepsy Res.* 2008; 81(2-3): 220-224.
15. Salzman A, Perroud N, Crespel A, Lamercy C, Malafosse A. Candidate genes for temporal lobe epilepsy: a replication study. *Neurol Sci.* 2008; 29(6): 397-403.
16. Thornblad TA, Elliott KS, Jowett J, Visscher PM. Prioritization of positional candidate genes using multiple web-based software tools. *Twin Res Hum Genet.* 2007; 10(6): 861-870.
17. van Driel MA, Cuelenaere K, Kemmeren PP, Leunissen JA, Brunner HG, Vriend G. GeneSeeker: extraction and integration of human disease-related information from web-based genetic databases. *Nucleic Acids Res.* 2005; 33(Web Server issue): W758-W761.
18. Prioritizer: <http://pcdoeglas.med.rug.nl/prioritizer/>.
19. Perez-Iratxeta C, Wjst M, Bork P, Andrade MA. G2D: A tool for mining genes associated to disease. *BMC Genet.* 2005; 6: 45.