FEBS openbio

Mining disease genes using integrated protein–protein interaction and gene–gene co-regulation information

CrossMark

Jin Li [a,b,c,*,1], Limei Wang [c,d,1], Maozu Guo [a,*], Ruijie Zhang [c], Qiguo Dai [a], Xiaoyan Liu [a], Chunyu Wang [a], Zhixia Teng [a], Ping Xuan [a], Mingming Zhang [c]

[a] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang, China
[b] School of Life Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang, China
[c] College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang, China
[d] School of Basic Medical Sciences, Harbin Medical University, Harbin, Heilongjiang, China

ARTICLE INFO

ABSTRACT

In humans, despite the rapid increase in disease-associated gene discovery, a large proportion of disease-associated genes are still unknown. Many network-based approaches have been used to prioritize disease genes. Many networks, such as the protein–protein interaction (PPI), KEGG, and gene co-expression networks, have been used. Expression quantitative trait loci (eQTLs) have been successfully applied for the determination of genes associated with several diseases. In this study, we constructed an eQTL-based gene–gene co-regulation network (GGCRN) and used it to mine for disease genes. We adopted the random walk with restart (RWR) algorithm to mine for genes associated with Alzheimer disease. Compared to the Human Protein Reference Database (HPRD) PPI network alone, the integrated HPRD PPI and GGCRN networks provided faster convergence and revealed new disease-related genes. Therefore, using the RWR algorithm for integrated PPI and GGCRN is an effective method for disease-associated gene mining.

1. Introduction

In humans, despite the rapid increase in the discovery of disease–associated genes, the molecular basis of many diseases is still known. Even for diseases for which the molecular basis is partially understood, a large proportion of the associated genes are still unknown. The known disease-associated genes have been reported to represent only a very small proportion of the actual number of disease-associated genes [1,2]. Hence, mining for disease genes remains important.

Network-based approaches to human disease have multiple biological and clinical applications [3,4]. Many molecular networks have been constructed experimentally to characterize the physical and/or functional interactions between biomolecules [4,5]. There are many methods for disease gene mining using molecular networks, such as the direct neighborhood [6–13], Shortest path length [13–16], Diffusion kernel [8], random walk with restart [8,9,17], propagation flow [18], and clique backbone [19] methods. The random walk with restart (RWR) method has been reported to have the best performance in terms of precision and recall, while both the random walk and propagation flow methods are superior to the clustering and neighborhood methods [20,21]. The most useful network is the protein–protein interaction (PPI) network [6,8,9]. Some other resources are also used in disease gene mining, such as gene ontology, gene co-expression network, KEGG, structure, and TRANSFAC [7,10,15,16].

Expression quantitative trait loci (eQTLs) analyses of DNA use hundreds of thousands of single-nucleotide polymorphism (SNP) markers that capture human genetic variation [22]. This strategy has been successfully applied to several diseases, such as celiac disease [23], asthma [24] and type 2 diabetes [25]. An eQTL is a locus that regulates a gene expression phenotype [26]. If two genes are regulated by one or more of the same SNPs, they are considered to be co-regulated. Obviously, this co-regulation is only one type of gene interactions. We constructed a gene–gene co-regulation

network (GGCRN) using eQTL data and believe that it will be useful for disease gene mining.

In this study, we developed a GGCRN, integrated it with the PPI network, and used the RWR method to mine for candidate disease genes. Using Alzheimer disease (AD) as an example, we demonstrated that this newly developed GGCRN is an effective resource for disease gene mining.

2. Materials and methods

2.1. Materials

2.1.1. Protein–protein interaction data

The Human Protein Reference Database (HPRD) describes interaction networks in the human proteome [27]. All information in the HPRD has been manually extracted from the literature by expert biologists who read, interpreted and analyzed the published data. For this study, we used HPRD, release 9, which contains 38,989 protein–protein interactions among 9605 proteins.

2.1.2. EQTL data

We used human brain tissue data for this disease gene mining study. The data were obtained from a series of 193 neuropathologically normal human brain samples using the Affymetrix GeneChip Human Mapping 500 K Array Set and Illumina HumanRefseq-8 Expression BeadChip platforms [28]. The eQTLs were determined by Matrix eQTL [29]. In this study, the cis-eQTL definition was a SNP within the gene body +1 Mb up/down stream of the gene body. We calculated cis-eQTLs and trans-eQTLs and performed FDR adjustment (q value < 0.1) separately; then we combined the cis-eQTLs and trans-eQTLs. Finally, we obtained 25,866 significant SNP-gene association pairs of 3709 genes. The results can be downloaded from the seeQTL database [30].

2.1.3. AD-related genes

Online Mendelian Inheritance in Man (OMIM) is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated daily. AD is classified as a neurodegenerative disorder, and it is associated with plaques and tangles in the brain [31]. We obtained 29 AD-related terms from OMIM [32]. After removing the terms with no approved gene symbol, we obtained 15 AD-related genes. Of these 15 genes, 14, 4 and 14 genes were present in the HPRD PPI, the GGCRN, and the HPRD PPI and GGCRN integrated network (Union network). We used the 14 genes that were present in the Union network for the subsequent analyses. See Supplemental Table 1.

2.2. Methods

2.2.1. Gene–gene co-regulation network construction

The human brain data included 25,866 significant SNP-gene association pairs of 3709 genes. For each gene, we first extracted the SNPs that regulate it, and we called these significant related SNPs. If a SNP regulated two genes, we called it a common SNP of the two genes. We considered two genes to be co-regulated if a specific proportion of SNPs regulated both genes. Mathematically, for any 2 genes (G_i and G_j), there are n_1 and n_2 significant related SNPs, respectively. The gene–gene co-regulation coefficient is defined as

$$\text{coreco}(G_i, G_j) = \frac{\#(\text{SNPs in } G_i \cap \text{SNPs in } G_j)}{\#(\text{SNPs in } G_i \cup \text{SNPs in } G_j)},$$

Where $\#(A)$ is the element number in set A. In other words, $\#(\text{SNPs in } G_i \cap \text{SNPs in } G_j)$ is the number of common SNPs that regulate both gene G_i and G_j; and $\#(\text{SNPs in } G_i \cup \text{SNPs in } G_j)$ is

the number of SNPs that regulate gene G_i or G_j. For example, if G_i and G_j have 100 and 80 significant related SNPs, respectively, and 30 of them are common SNPs for G_i and G_j, then the co-regulation coefficient is $30/(100 + 80 - 30) = 0.2$. After calculating all co-regulation coefficients for all gene pairs, a reasonable threshold value for filtering the significant co-regulated gene pairs had to be established; for this purpose, we used the clustering coefficient difference maximization method [33]. The main function of this method is the determination of the difference in the maximum clustering coefficient difference between a real network and a random network if the real network is highly credible. Finally, we obtained the GGCRN using the significant co-regulated gene pairs.

2.2.2. Random walk with restart algorithm

In this paper, we focused on the genetic data resources rather than on statistical methods. Therefore, we adopted a classic and efficient method. The random walk algorithm (RW) for graphs is defined as an iterative walker's random transition from its current node to a neighboring node, and this is initiated at a given source node [34,35]. The random walk with restart algorithm (RWR) [36] is a variant of the random walk that allows for the restart of the walk at every time step at source node s with probability r. Formally, the RWR is defined as:

$$p^{t+1} = (1 - r)Wp^t + rp^0$$

where W is the column-normalized adjacency matrix of the graph and p^t is a vector in which the ith element holds the probability of being at node i at time step t. A special case is the initial probability vector, p^0, which is the probability of being at source node, s. In our application, p^0 was constructed such that equal probabilities were assigned to the known disease genes, with the sum of the probabilities equal to 1. Genes were ranked according to the values in the steady-state probability vector p^N. This was obtained by performing the iteration until the change between p^t and p^{t+1} fell below 10^{-6}. The results of RWR are affected by the restart probability, r. We perform a numerical experiment to select the proper r value.

The flow chart is illustrated in Fig. 1.

2.2.3. Direct neighborhood algorithm

We also compared the RWR method with the direct neighborhood (DN) method. In the DN method, the interaction partners in the network were determined for each known disease gene. The more linkages that exist between a gene and known disease genes for a particular disease, the greater the possibility that it is related
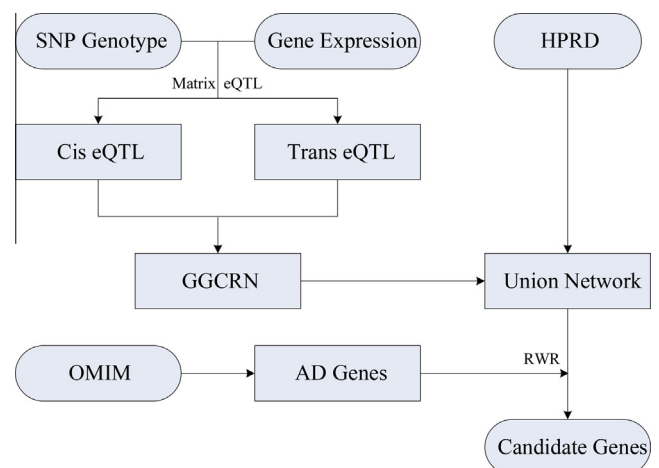

Fig. 1. The flow chart of RWR method using the Union network.

to that disease. In this study, genes with more than 1 linkage with known disease genes were considered candidate disease genes.

3. Results and discussion

3.1. Gene–gene co-regulation network

From the brain data, we identified 181,906 co-regulated gene pairs of 2830 genes with nonzero co-regulation coefficients. We set the threshold to 0 to 1, a distance of 0.01. We calculated the trends in the clustering coefficients, and found the first extreme point of clustering coefficient difference (the real clustering coefficient minus the background clustering coefficient, Fig. 2). As a result, we identified 25,937 gene pairs of 1444 genes under a threshold value 0.63.

3.2. Comparison between HPRD and brain co-regulation network

The comparison of the HPRD PPI and GGCRN network revealed only 4 common gene-gene pairs between 8 genes (Table 1). This indicates that the gene–gene regulation patterns in the GGCRN were different from the protein–protein interactions in the HPRD PPI network. Therefore, the GGCRN can provide new information about gene pairs.

In the following analysis, we used 3 networks, the HPRD PPI network, the GGCRN and the Union network.

3.3. Disease gene mining

In the RWR method, to obtain a reasonable restart probability, r, we performed a numerical experiment. First, we set r to 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9. When r was 0, the RWR could not reach steady state, so we added a stop condition of walking no more than 100,000 times. To mine for disease risk genes, we selected the genes with a steady-state probability greater than the initial known disease genes and considered them candidate genes. The results of this analysis are described in Table 2. We found that when r was 0, we obtained too many candidate genes and that when r was more than 0.1, we obtained too few candidate genes. Then, we set r to 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045 and 0.05. We did not know exactly how many disease genes existed; thus we defined a standard for mining risk genes of mining less than 2 times of the known initial genes. For AD, we examined 14 initial genes in the HPRD PPI and Union networks and 4 initial

Table 1
Comparisons of the HPRD PPI and GGCRN networks.

Networks	Number of genes	Number of gene pairs	Number of AD genes
HPRD PPI	9605	39,023	14
GGCRN	1444	25,937	4
Intersection of HPRD and GGCRN	8	4	0
Union of HPRD and GGCRN	10,209	64,956	14

Table 2
Comparisons of the results of disease gene mining in 3 networks.

r	HPRD PPI	GGCRN	Union network
0	5805	1440	7448
0.01	124	41	304
0.015	58	41	27
0.02	27	2	9
0.025	25	2	4
0.03	7	2	2
0.035	4	2	1
0.04	1	2	1
0.045	1	2	1
0.05	1	2	1
0.1	1	2	1
0.2	0	2	0
0.3	0	2	0
0.4–0.9	0	0	0

genes in the co-regulation network. Using the co-regulation network, we obtained 41 candidate genes with r set at 0.015 and 2 candidate genes with r set at 0.02. As there were only 4 initial genes, this was not a good result, and this is because the co-regulation network is small, only containing 1444 genes. Therefore, it is unsuitable for candidate gene mining using only the GGCRN. Then, we compared the results using the HPRD PPI and Union network. With r set at 0.01, the numbers of candidate genes obtained were still insufficient. While, with an r value larger than 0.015, we obtained more candidate genes using the HPRD PPI network than the Union network with the same restart probability. In other words, to obtain the same number of candidate genes using these two networks, a higher restart probability is needed for the HPRD PPI network than for the Union network. This indicates that more convergent results can be obtained using the Union network than the HPRD PPI network alone. Therefore, the GGCRN constructed by eQTL data is a useful resource for mining disease genes.
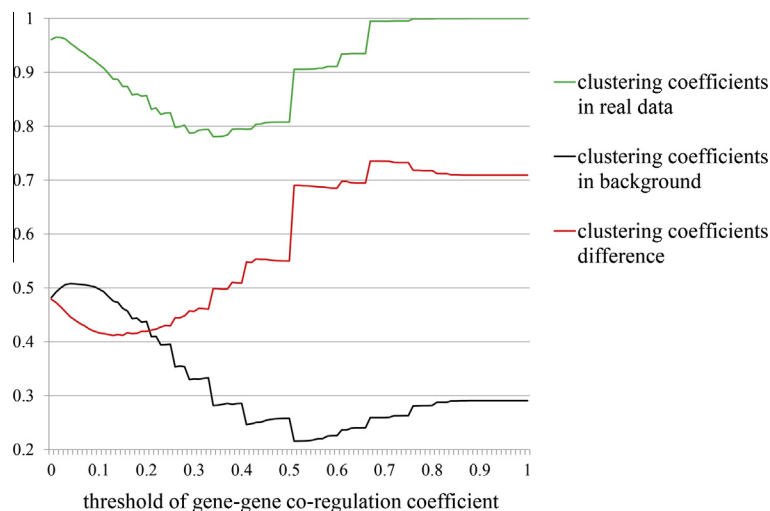

Fig. 2. The trends of the clustering coefficients.

In the DN method, genes with more than 1 linkage with known disease genes were considered candidate disease genes. In the co-regulation network, there were only 4 initial known AD genes, and no gene had more than 1 linkage with known disease genes. In the HPRD PPI and Union networks, there were 14 initial known AD genes, and there were 38 and 39 genes with more than 1 linkage with known disease genes, respectively. These results show that the co-regulation network is not suitable for gene mining by the DN method, because it is too small. Furthermore, the results of gene mining with DN method using the combination of the co-regulation network and the HPRD PPI network were almost the same as those for the HPRD PPI network alone. The detailed results of this analysis are shown in Supplemental Table 2.

For gene mining with the RWR method using the Union network we obtained 27 candidate genes with r set at 0.015, and for gene mining with the DN method using the Union network we obtained 39 genes with more than 1 linkage with known disease genes. Only 1 gene was common between these two gene sets. Even when we included all genes that neighbored known AD genes with the DN method, only 11 genes were common between these two gene sets. These results show that the RWR method tends to identify more indirect neighborhood genes, and this is consistent with the gene function pattern of biological networks: genes that interact indirectly may perform the same functions as those that interact directly. Additionally, it was previously reported that the RWR method has the best performance in terms of precision and recall compared to other neighborhood methods [20,21].

3.4. Gene annotation

To test the usefulness of the candidate genes mined by RWR, we compared them to those published in the literature. We first verified the 4 most likely AD risk genes (CP, EP300, TP53, and YWHAG, with a restart probability 0.035 in the HPRD PPI and 0.025 in the Union network). The protein encoded by CP is a metalloprotein that binds most of the copper in plasma and is involved in the per-oxidation of Fe(II) transferrin to Fe(III) transferrin. Mutations in this gene cause aceruloplasminemia, which is associated with neu-rologic abnormalities. Some studies confirmed that ceruloplasmin is increased in patients with AD [37,38]. EP300 encodes the aden-ovirus E1A-associated cellular p300 transcriptional co-activator protein, which mediates cAMP-gene regulation by binding specifi-cally to phosphorylated CREB protein and has been shown to be related to AD [39,40]. TP53 (P53) encodes a tumor suppressor pro-tein that contains transcriptional activation, DNA binding, and oligomerization domains. TP53 is involved in neurodegenerative syndromesis, and is up-regulated in AD [41–43]. YWHAG has been shown to interact with RAF1 and protein kinase C and is involved in various signal transduction pathways. A gene expression analy-sis showed that YWHAG is down-regulated in AD [44]. These results indicate that these 4 genes are related to AD with high confidence.

To compare between the results of gene mining using the HPRD PPI and Union networks, we analyzed the 27 genes detected in the HPRD PPI with r set at 0.02 and the 27 genes detected in the Union network with r set at 0.015. The detailed results are shown in Supplemental Table 3. Twenty-two genes were common between the two networks, and they were in front of the 27 genes detected in the HPRD PPI. We then analyzed the 10 genes that were uncom-mon between the two networks, and we analyzed the ranks of these genes in the results. The 5 genes that only appeared in the results of the HPRD PPI network ranked from 28 to 37 in the results of the Union network, which was just below the level of signifi-cance (rank 1–27). However, the 5 genes that only appeared in the results of the Union network ranked from 66 to 547 in the results of the HPRD PPI network. Thus, the 5 genes that only

Table 3
The 10 candidate genes uncommon between the HPRD PPI and Union networks.

Genes	Rank in the results of HPRD	Rank in the results of Union network
CTNNB1	23	28
RB1	24	36
TRAF2	25	30
AKT1	26	34
PIK3R1	27	37
RAF1	66	16
MCM2	423	20
RPS6KA1	476	24
MDM2	126	26
MAP3K3	547	27

appeared in the results of the Union network are new discoveries based on this gene–gene co-regulation information (Table 3).

We verified these 5 genes by comparison with those published in the literature. RAF1 is the cellular homolog of the viral raf gene (v-raf). It is one of the physiological activators of the ERK pathway, which plays key roles in several steps of tumorigenesis, including cancer cell proliferation, migration, and invasion [45]. Raf-1 activa-tion effectively mediates Ras-dependent signals in AD [46]. The protein encoded by MCM2 is one of the highly conserved mini-chromosome maintenance proteins (MCM) that are involved in the initiation of eukaryotic genome replication. Phosphorylated MCM2 (pMCM2) is markedly associated with neurofibrillary tan-gles, neuropil threads, and dystrophic neurites in AD [47]. RPS6KA1 encodes a member of the RSK family of serine/threonine kinases. This kinase contains 2 nonidentical kinase catalytic domains and phosphorylates various substrates, including mem-bers of the mitogen-activated kinase (MAPK) signaling pathway. Deviation from the strict control of MAPK signaling pathways has been implicated in the development of many human diseases, including AD[45]. MDM2 encodes a nuclear-localized E3 ubiquitin ligase. The encoded protein can promote tumor formation by tar-geting tumor suppressor proteins, such as p53, for proteasomal degradation. This gene is transcriptionally regulated by p53 [48], which is involved in neurodegenerative syndromes, and is up-regu-lated in AD [41–43]. MAP3K3 (mitogen-activated protein kinase kinase kinase 3) is a 626-amino acid polypeptide. This protein directly regulates the stress-activated protein kinase (SAPK) and extracellular signal-regulated protein kinase (ERK) pathways. MAP3K3 phosphorylates RCAN1, and RCAN1-1L is overexpressed in neurons in AD patients [49]. Furthermore, two polymorphisms of the RCAN1 gene are associated with AD [50]. These results indi-cated that these 5 genes are related to AD.

Compared with using the HPRD PPI network alone, more dis-ease genes can be identified by integrating the HPRD PPI and GGCRN networks. The results of this study indicate that eQTL-based gene–gene regulation information is a useful resource for disease gene mining.

4. Conclusion

There are certain relationships between genes that are associ-ated with the same SNPs. The eQTL-based GGCRN developed here provides extra gene–gene interaction information than the HPRD PPI network. We integrated the GGCRN and HPRD PPI networks and used the RWR method and DN method to mine for disease gene candidates. The RWR method identified more indirect neigh-borhood candidate genes than the DN method. Applying the AD data, compared to using the HPRD PPI network, using the inte-grated network obtained better results. With the same restart probability, the integrated network provided faster convergence and identified some new disease-related genes.

Therefore, disease gene mining using an RWR approach with an integrated network (HPRD PPI and GGCRN) is an effective method. This eQTL-based GGCRN also has some disadvantages such as its coverage, the genes in this co-regulation network are not sufficient.

Author contribution statement

J.L. and M.G. conceived the study. J.L., L.W. and Q.D. did most of the experiments. R.Z., X.L., C.W., Z.T., P.X. and M.Z. analyzed the data. All authors reviewed the manuscript.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fob.2015.03.011.

References

[1] Oldenburg, R.A., Meijers-Heijboer, H., Cornelisse, C.J. and Devilee, P. (2007) Genetic susceptibility for breast cancer: how many more genes to be found? Crit. Rev. Oncol. Hematol. 63, 125–149.

[2] Frayling, T.M. (2007) Genome-wide association studies provide new insights into type 2 diabetes aetiology. Nat. Rev. Genet. 8, 657–662.

[3] Wang, X., Gulbahce, N. and Yu, H. (2011) Network-based methods for human disease gene prediction. Briefings Funct. Genomics 10, 280–293.

[4] Barabasi, A.L., Gulbahce, N. and Loscalzo, J. (2011) Network medicine: a network-based approach to human disease. Nat. Rev. Genet. 12, 56–68.

[5] Vidal, M., Cusick, M.E. and Barabasi, A.L. (2011) Interactome networks and human disease. Cell 144, 986–998.

[6] Oti, M., Snel, B., Huynen, M.A. and Brunner, H.G. (2006) Predicting disease genes using protein–protein interactions. J. Med. Genet. 43, 691–698.

[7] Aerts, S., Lambrechts, D., Maity, S., Van Loo, P., Coessens, B., De Smet, F., Tranchevent, L.C., De Moor, B., Marynen, P., Hassan, B., Carmeliet, P. and Moreau, Y. (2006) Gene prioritization through genomic data fusion. Nat. Biotechnol. 24, 537–544.

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

[9] Navlakha, S. and Kingsford, C. (2010) The power of protein interaction networks for associating genes with diseases. Bioinformatics 26, 1057–1063.

[10] Linghu, B., Snitkin, E.S., Hu, Z., Xia, Y. and Delisi, C. (2009) Genome-wide prioritization of disease genes and identification of disease–disease associations from an integrated human functional linkage network. Genome Biol. 10, R91.

[11] Lage, K., Karlberg, E.O., Storling, Z.M., Olason, P.I., Pedersen, A.G., Rigina, O., Hinsby, A.M., Tumer, Z., Pociot, F., Tommerup, N., Moreau, Y. and Brunak, S. (2007) A human phenome–interactome network of protein complexes implicated in genetic disorders. Nat. Biotechnol. 25, 309–316.

[12] Care, M.A., Bradford, J.R., Needham, C.J., Bulpitt, A.J. and Westhead, D.R. (2009) Combining the interactome and deleterious SNP predictions to improve disease gene identification. Hum. Mutat. 30, 485–492.

[13] Wu, X., Jiang, R., Zhang, M.Q. and Li, S. (2008) Network-based global inference of human disease genes. Mol. Syst. Biol. 4, 189.

[14] Radivojac, P., Peng, K., Clark, W.T., Peters, B.J., Mohan, A., Boyle, S.M. and Mooney, S.D. (2008) An integrated approach to inferring gene–disease associations in humans. Proteins 72, 1030–1037.

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

[16] Karni, S., Soreq, H. and Sharan, R. (2009) A network-based method for predicting disease-causing genes. J. Comput. Biol. 16, 181–189.

[17] Li, Y. and Patra, J.C. (2010) Genome-wide inferring gene–phenotype relationship by walking on the heterogeneous network. Bioinformatics 26, 1219–1224.

[18] Vanunu, O., Magger, O., Ruppin, E., Shlomi, T. and Sharan, R. (2010) Associating genes and protein complexes with disease via network propagation. PLoS Comput. Biol. 6, e1000641.

[19] Yang, L., Zhao, X. and Tang, X. (2014) Predicting disease-related proteins based on clique backbone in protein–protein interaction network. Int. J. Biol. Sci. 10, 677–688.

[20] Brown, K.R. and Jurisica, I. (2005) Online predicted human interaction database. Bioinformatics 21, 2076–2082.

[21] Jain, E., Bairoch, A., Duvaud, S., Phan, I., Redaschi, N., Suzek, B.E., Martin, M.J., McGarvey, P. and Gasteiger, E. (2009) Infrastructure for the life sciences: design and implementation of the UniProt website. BMC Bioinf. 10, 136.

[22] Franke, L. and Jansen, R.C. (2009) EQTL analysis in humans. Methods Mol. Biol. 573, 311–328.

[23] Hunt, K.A., Zhernakova, A., Turner, G., Heap, G.A., Franke, L., Bruinenberg, M., Romanos, J., Dinesen, L.C., Ryan, A.W., Panesar, D., Gwilliam, R., Takeuchi, F., McLaren, W.M., Holmes, G.K., Howdle, P.D., Walters, J.R., Sanders, D.S., Playford, R.J., Trynka, G., Mulder, C.J., Mearin, M.L., Verbeek, W.H., Trimble, V., Stevens, F.M., O'Morain, C., Kennedy, N.P., Kelleher, D., Pennington, D.J., Strachan, D.P., McArdle, W.L., Mein, C.A., Wapenaar, M.C., Deloukas, P., McGinnis, R., McManus, R., Wijmenga, C. and van Heel, D.A. (2008) Newly identified genetic risk variants for celiac disease related to the immune response. Nat. Genet. 40, 395–402.

[24] Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufe, A., Rietschel, E., Heinzmann, A., Simma, B., Frischer, T., Willis-Owen, S.A., Wong, K.C., Illig, T., Vogelberg, C., Weiland, S.K., von Mutius, E., Abecasis, G.R., Farrall, M., Gut, I.G., Lathrop, G.M. and Cookson, W.O. (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature 448, 470–473.

[25] Das, S.K. and Sharma, N.K. (2014) Expression quantitative trait analyses to identify causal genetic variants for type 2 diabetes susceptibility. World J. Diabetes 5, 97–114.

[26] Nica, A.C. and Dermitzakis, E.T. (2013) Expression quantitative trait loci: present and future. Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci. 368, 20120362.

[27] Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., Balakrishnan, L., Marimuthu, A., Banerjee, S., Somanathan, D.S., Sebastian, A., Rani, S., Ray, S., Harrys Kishore, C.J., Kanth, S., Ahmed, M., Kashyap, M.K., Mohmood, R., Ramachandra, Y.L., Krishna, V., Rahiman, B.A., Mohan, S., Ranganathan, P., Ramabadran, S., Chaerkady, R. and Pandey, A. (2009) Human Protein Reference Database – 2009 update. Nucleic Acids Res. 37, D767–D772.

[28] Myers, A.J., Gibbs, J.R., Webster, J.A., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., Leung, D., Bryden, L., Nath, P., Zismann, V.L., Joshipura, K., Huentelman, M.J., Hu-Lince, D., Coon, K.D., Craig, D.W., Pearson, J.V., Holmans, P., Heward, C.B., Reiman, E.M., Stephan, D. and Hardy, J. (2007) A survey of genetic human cortical gene expression. Nat. Genet. 39, 1494–1499.

[29] Shabalin, A.A. (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 28, 1353–1358.

[30] Xia, K., Shabalin, A.A., Huang, S., Madar, V., Zhou, Y.H., Wang, W., Zou, F., Sun, W., Sullivan, P.F. and Wright, F.A. (2012) SeeQTL: a searchable database for human eQTLs. Bioinformatics 28, 451–452.

[31] Tiraboschi, P., Hansen, L.A., Thal, L.J. and Corey-Bloom, J. (2004) The importance of neuritic plaques and tangles to the development and evolution of AD. Neurology 62, 1984–1989.

[32] Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A. and McKusick, V.A. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 33, D514–D517.

[33] Elo, L.L., Jarvenpaa, H., Oresic, M., Lahesmaa, R. and Aittokallio, T. (2007) Systematic construction of gene coexpression networks with applications to human T helper cell differentiation process. Bioinformatics 23, 2096–2103.

[34] Pearson, K. (1905) The problem of the random walk. Nature 72, 1.

[35] Lange, K. and Sobel, E. (1991) A random walk method for computing genetic location scores. Am. J. Hum. Genet. 49, 1320–1334.

[36] Ham, B., Min, D. and Sohn, K. (2013) A generalized random walk with restart and its application in depth up-sampling and interactive segmentation. IEEE Trans. Image Process. 22, 2574–2588.

[37] Connor, J.R., Tucker, P., Johnson, M. and Snyder, B. (1993) Ceruloplasmin levels in the human superior temporal gyrus in aging and Alzheimer's disease. Neurosci. Lett. 159, 88–90.

[38] Squitti, R., Pasqualetti, P., Dal Forno, G., Moffa, F., Cassetta, E., Lupoi, D., Vernieri, F., Rossi, L., Baldassini, M. and Rossini, P.M. (2005) Excess of serum copper not related to ceruloplasmin in Alzheimer disease. Neurology 64, 1040–1046.

[39] Saura, C.A. and Valero, J. (2011) The role of CREB signaling in Alzheimer's disease and other cognitive disorders. Rev. Neurosci. 22, 153–169.

[40] Urdinguio, R.G., Sanchez-Mut, J.V. and Esteller, M. (2009) Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. Lancet Neurol. 8, 1056–1072.

[41] Hooper, C., Meimaridou, E., Tavassoli, M., Melino, G., Lovestone, S. and Killick, R. (2007) P53 is upregulated in Alzheimer's disease and induces tau phosphorylation in HEK293a cells. Neurosci. Lett. 418, 34–37.

[42] Kitamura, Y., Shimohama, S., Kamoshima, W., Matsuoka, Y., Nomura, Y. and Taniguchi, T. (1997) Changes of p53 in the brains of patients with Alzheimer's disease. Biochem. Biophys. Res. Commun. 232, 418–421.

[43] Vousden, K.H. and Lane, D.P. (2007) P53 in health and disease. Nat. Rev. Mol. Cell Biol. 8, 275–283.

[44] Hokama, M., Oka, S., Leon, J., Ninomiya, T., Honda, H., Sasaki, K., Iwaki, T., Ohara, T., Sasaki, T., Laferla, F.M., Kiyohara, Y. and Nakabeppu, Y. (2014) Altered expression of diabetes-related genes in Alzheimer's disease brains: the Hisayama study. Cereb. Cortex 24 (9), 2476–2488, http://dx.doi.org/10.1093/cercor/bht101.

[45] Kim, E.K. and Choi, E.J. (2010) Pathological roles of MAPK signaling pathways in human diseases. Biochim. Biophys. Acta 1802, 396–405.

[46] Mei, M., Su, B., Harrison, K., Chao, M., Siedlak, S.L., Previll, L.A., Jackson, L., Cai, D.X. and Zhu, X. (2006) Distribution, levels and phosphorylation of Raf-1 in Alzheimer's disease. J. Neurochem. 99, 1377–1388.

[47] Bonda, D.J., Evans, T.A., Santocanale, C., Llosa, J.C., Vina, J., Bajic, V., Castellani, R.J., Siedlak, S.L., Perry, G., Smith, M.A. and Lee, H.G. (2009) Evidence for the progression through S-phase in the ectopic cell cycle re-entry of neurons in Alzheimer disease. Aging 1, 382–388.

[48] Picksley, S.M. and Lane, D.P. (1993) The p53-mdm2 autoregulatory feedback loop: a paradigm for the regulation of growth control by p53? Bioessays 15, 689–690.

[49] Harris, C.D., Ermak, G. and Davies, K.J. (2007) RCAN1-1L is overexpressed in neurons of Alzheimer's disease patients. FEBS J. 274, 1715–1724.

[50] Lin, K.G., Tang, M., Guo, Y.B., Han, H.Y. and Lin, Y.H. (2011) Two polymorphisms of RCAN1 gene associated with Alzheimer's disease in the Chinese Han population. East Asian Arch. Psychiatry 21, 79–84.