DMBase: An Integrated Genetic Information Resource for Diabetes Mellitus

Sun Young Lee[1], Young-Kyu Park[1], Jaeheup Kim[2] and Young Joo Kim[1,*]

[1]Genomics Resource Center, KRIBB, 125 Gwahag-ro, Yusong-gu, Daejeon 305-806, Republic of Korea
[2]Cogent Biotechnology Inc., Rockville, MD 20850, USA

SYNOPSIS

Diabetes Mellitus (DM), often simply referred to as diabetes, has developed into a major health concern affecting more than 200 million people worldwide with approximately 4 million deaths per year attributed to the presence of the disease. Diabetes mellitus is categorized as Type 1 and Type 2, where Type 1 diabetes represents a lack of insulin production, and Type 2 diabetes is characterized by a relative lack of insulin receptor (i.e., decreased sensitivity to the effect of insulin) and cased by a complex interplay between genetic factors and environmental factors. Up to date, various studies on the pathology and mechanism in terms of genetic experiments have been conducted and approximately hundreds of genes were reported as diabetes mellitus associated genes. At this point, to support studies on the cause and mechanism of diabetes mellitus, an efficient database system to provide genetic variants related to diabetes mellitus is needed. DMBase is an integrated web-based genetic information resource for diabetes mellitus designed to service genomic variants, genes, and secondary information derived for diabetes mellitus genetics researchers. The current version of DMBase documents 754 genes with 3056 genetic variants and 66 pathways. It provides many effective search interfaces for retrieving diabetes mellitus and genetic information. A web interface for the DMBase is freely available at http://sysbio.kribb.re.kr/dmBase.
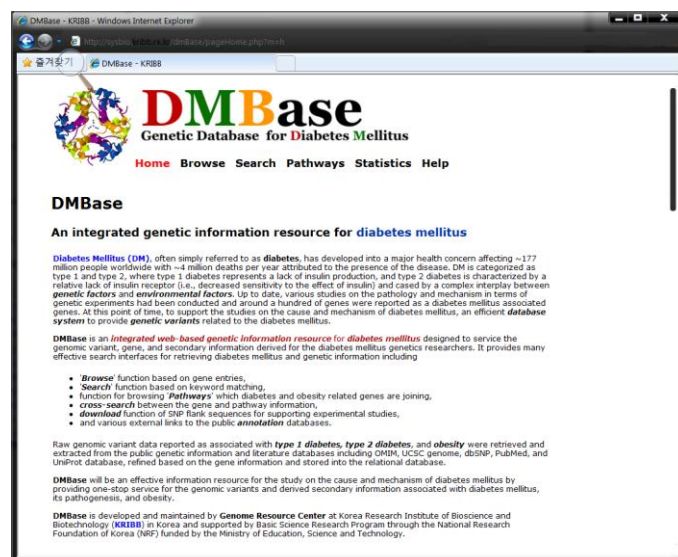
Keywords: diabetes mellitus, genetics, web interface, obesity, systems biology

Introduction

DMBase is an integrated genetic information resource for diabetes mellitus. Diabetes Mellitus is an increasing health concern affecting more than 200 million people worldwide with approximately 4 million deaths per year attributed to the presence of the disease (http://www.who.int/diabetes/facts/en). Diabetes mellitus is categorized as Type 1 and Type 2, where Type 1 diabetes (T1D) represents a lack of insulin production, and type 2 diabetes (T2D) is characterized by a relative lack of insulin receptor (i.e., decreased sensitivity to the effect of insulin). T1D is a chronic disease that typically manifests itself in childhood[1,2]. T2D is a common complex disease, thought to be caused by a complex interplay between genetic factors and environmental factors[3,4].

Up to date, various studies on the pathology and mechanism in terms of genetic experiments have been conducted and approximately hundreds of genes were reported as diabetes mellitus associated genes[5-7]. At this point, to support studies on the cause and mechanism of diabetes mellitus, we've developed an efficient database system that provides many effective search interfaces for retrieving diabetes mellitus and genetic information related to diabetes mellitus and obesity.

Results and Discussion

DMBase is an integrated web-based genetic information resource for diabetes mellitus designed to service genomic variants, genes, and secondary information derived for diabetes mellitus genetics researchers, supplementary to understand the relation with obesity. A web interface for the DMBase is freely accessible at http://sysbio.kribb.re.kr/dmBase[9] (Figure 1). We used OMIM[8] and UniProt[9] to retrieve information on Type 1 diabetes, Type 2 diabetes, and obesity related genomic variants, dbSNP[10] for SNP information, UCSC genome[11] for the gene structure information. HGNC database[12] for standard gene names, MSigDB[13] for pathway and functional gene set information, and PubMed links as raw data sources were utilized (Figure 2).

The statistics of data stored DMBase are summarized in Table 1. The current version of DMBase documents 754 genes with 3056 genetic variants and 66 pathways. DMBase provides user interfaces to retrieve genetic variants and biological pathway information, such as i) *Browser* function based on gene entries, ii) *Search* function based on keyword matching, iii) *Pathways* function which diabetes

Table 1. Statistics of the DMBase

No.	Disease	# of Genes	# of Variants	# of Sig. Pathways[a]
1	Diabetes mellitus, Type 2	209	1665	19
2	Diabetes mellitus, Type 1	131	266	20
3	Obesity	414	1125	27
Total	ALL	754(600[b])	3056	66(47[b])

a) Pathways with the FDR q-values less than 0.05.
b) Count for non-redundant genes or pathways over the etiologies.

mellitus and obesity related genes are joining, and iv) various external links to the public annotation databases (Figure 3). To retrieve information related to diabetes and obesity, users can use the two search flows such as gene-based variant search and pathway-based variant search. The gene-based variant search uses the genes as an entry point of the variant information by using gene table browsing or keyword search (Figure 4). The pathway-based variant search uses pathways as an entry point of the variant information by using pathway table browsing (Figure 5).

The web interface of DMBase has been developed using HTML and PHP language and all data in the database has been stored in MySQL tables.

Conclusion and Prospects

We developed the DMBase to provide genetic variants related to diabetes mellitus and obesity. DMBase will be an effective
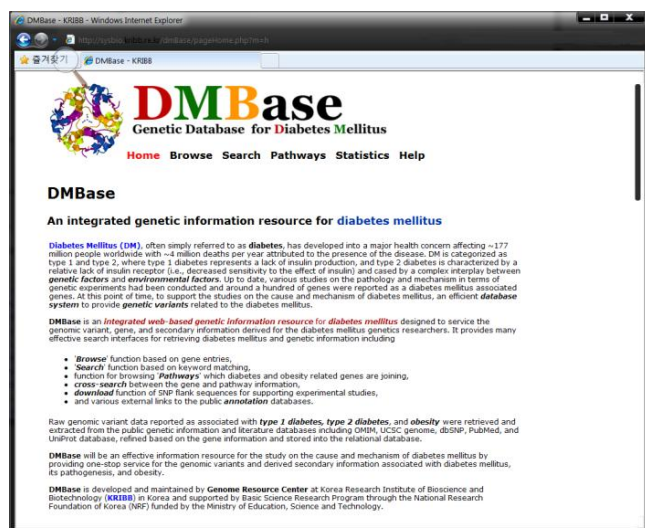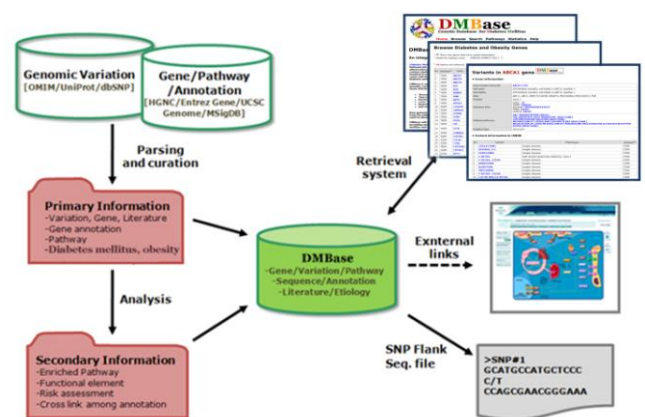

Figure 2. Schematic diagram for the construction and service of DMBasae.

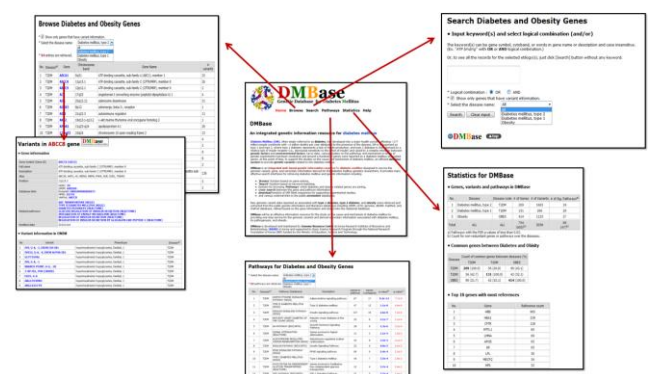**Figure 1.** Main home page of DMBase (http://sysbio.kribb.re.kr/dmBase/).

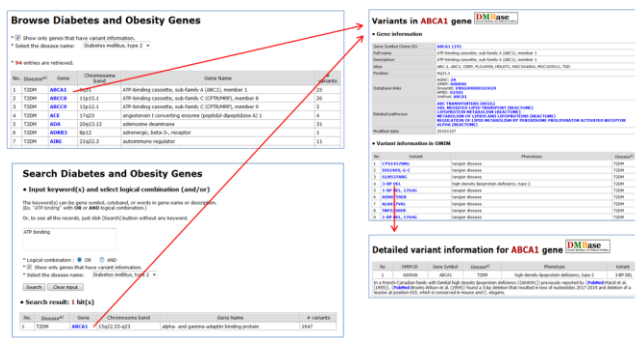**Figure 3.** User interface of DMBase.

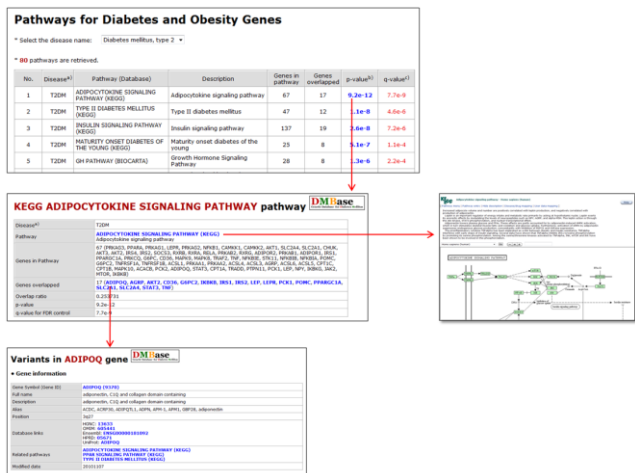**Figure 4.** The workflow for gen-based variant search.

Figure 5. The workflow for pathway-based variant search.

information resource for the study on the cause and mechanism of diabetes mellitus by providing one-stop service for the genomic variants and derived secondary information associated with diabetes mellitus, its pathogenesis, and obesity.

Acknowledgements

References

1. Grant, S.F., and Hakonarson, H. (2009). Genome-wide association studies in type 1 diabetes. *Curr Diab Rep* 9, 157-163.
2. Hulbert, E.M., Smink, L.J., Adlem, E.C., Allen, J.E., Burdick, D.B., Burren, O.S., Cassen, V.M., Cavnor, C.C., Dolman, G.E., Flamez, D., et al. (2007). T1DBase: integration and presentation of complex data for type 1 diabetes research. *Nucleic Acids Res* 35, D742-746.
3. McCarthy, M.I., and Zeggini, E. (2009). Genome-wide association studies in type 2 diabetes. *Curr Diab Rep* 9, 164-171.
4. Zeggini, E., and McCarthy, M.I. (2007). Identifying susceptibility variants for type 2 diabetes. *Methods Mol Biol* 376, 235-250.
5. Agrawal, S., Dimitrova, N., Nathan, P., Udayakumar, K., Lakshmi, S.S., Sriram, S., Manjusha, N., and Sengupta, U. (2008). T2D-Db: an integrated platform to study the molecular basis of Type 2 diabetes. *BMC Genomics* 9, 320.
6. Mazzarelli, J.M., Brestelli, J., Gorski, R.K., Liu, J., Manduchi, E., Pinney, D.F., Schug, J., White, P., Kaestner, K.H., and Stoeckert, C.J., Jr. (2007). EPConDB: a web resource for gene expression related to pancreatic development, beta-cell function and diabetes. *Nucleic Acids Res* 35, D751-755.
7. Perry, J.R., McCarthy, M.I., Hattersley, A.T., Zeggini, E., Weedon, M.N., and Frayling, T.M. (2009). Interrogating type 2 diabetes genome-wide association data using a biological pathway-based approach. *Diabetes* 58, 1463-1467.
8. Amberger, J., Bocchini, C.A., Scott, A.F., and Hamosh, A. (2009). McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res* 37, D793-796.
9. UniProt consortium. (2009). The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res* 38, D142-148.
10. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29, 308-311.
11. Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Diekhans, M., Smith, K.E., Rosenbloom, K.R., Raney, B.J., et al. (2010). The UCSC Genome Browser database: update 2010. *Nucleic Acids Res* 38, D613-619.
12. Povey, S., Lovering, R., Bruford, E., Wright, M., Lush, M., and Wain, H. (2001). The HUGO Gene Nomenclature Committee (HGNC). *Hum Genet* 109, 678-680.
13. 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 U S A* 102, 15545-15550.
14. HUGO Gene Nomenclature Committee (HGNC). http://www.genenames.org.
15. The Molecular Signatures Database (MSigDB). http://www.broadinstitute.org/gsea/msigdb.
16. UCSC Genome Bioinformatics. http://genome.ucsc.edu.