

BCGDB: Breast Cancer Gene Database

S. Rajasekaran¹, S. Sathiyabama² and T. S. Gnanendra³

¹Research and Development Centre, Bharathiar University, Coimbatore - 641046, Tamil Nadu, India

²Department of Computer Science, Government Arts College, Rasipuram - 637401, Tamil Nadu, India

³Department of Biotechnology, Mahendra Arts and Science College, Tiruchengode, Namakkal - 637501, Tamil Nadu, India

Abstract

Globally, breast cancer is the most commonly diagnosed tumour that leads to the death in women. The number of genes that leads to the breast cancer is gradually increasing and a pursuit for the better understanding of the genes and their products that serves as the potential drug targets of breast cancer are demanding. In this regard, the development of potential resources that can employ the key genes and their products that plays a crucial role in the cause of breast cancer, may serve as a potential resource for the identification of novel drug targets and also in the development of novel drugs. Thus Breast Cancer Gene Database (BCGDB) holding 2948 over expressed genes that involve in breast cancer and 487 genes that play significant role in breast cancer were provided. The structural and functional aspects of these genes and their products provided in this database can serve as an ecstasy resource for the breast cancer researchers to pursue their investigations for the development of novel therapeutics for the cure of breast cancer. This database can be accessed online (<http://www.bcgdb.in>).

Keywords: Adenocarcinomas, Breast Cancer, Database, Data Mining, Malignant Tumour

1. Introduction

Globally, the most frequently diagnosed cancers among women is reported to be Breast cancer¹ and found to be steadily increased since the 1970s resulting in the disease becoming a worldwide major health burden². Mortality rates for this disease have remained constant or declined slightly in the past 20-25 years in the developed and developing countries^{3,4}. Across the world, the most commonly diagnosed tumour that leads to the death in women are due to breast cancer. The adenocarcinoma is a malignant tumour that starts in the glandular tissues of breast and invades in to other parts of the body.

Based on the histopathological features and microarray based molecular profiling the breast cancer can be grouped. Among the types of cancers in females, breast cancer is found to be the most common one that accounts for 27% and is reported as 15% responsible for cancer fatality in females. In United States during 2013, nearly 226,870 and 2,190 new cases of breast cancers were

diagnosed in females and males respectively. In 2014, approximately 2.9 million breast cancer survivors in the United States were reported⁵.

Adenocarcinoma is a multi-factorial disease, with genetic, endocrinological, environmental and lifestyle factors contributing directly or indirectly to carcinogenesis⁶. Even though adenocarcinoma is the most studied diseases in the world, the aetiology for over 70% of breast cancers remains unknown. There are a number of ways genetic alterations may lead to breast cancer. These include point mutations, gene amplifications and chromosomal alterations such as gains, losses and recombinations. It has been extensively documented that the inheritance of deleterious mutations in the high penetrance breast cancer susceptibility genes, Breast Cancer 1 (BRCA1) and Breast Cancer 2 (BRCA2) account for about 5-10 % of hereditary breast and ovarian cancers⁷. In addition, breast cancers that exhibit molecular and clinical characteristics that are shared by breast cancers that arise from BRCA1 mutations can arise

* Author for correspondence

sporadically; this concept is termed “BRCAness”⁸. Breast tumours that exhibit BRCAness include basal-like breast cancers and triple negative breast cancers, Progesterone Receptor (PGR) and v-erb-b2 erythroblastic leukaemia viral oncogene homologue 2 (ERBB2; more commonly known as HER2)⁹.

As the number of genes that were responsible for the breast cancer was gradually increasing and the paucity on the breast cancer gene information emphasized us for the development of Breast Cancer database. This BCGDB holds the breast cancer gene and their related information collected from multiple sources. This warehouse is devoted to envisage the structure and function of the genes and their corresponding proteins which can ignite researchers to elucidate its role as an important target of breast cancer and also to provide the complete information about the disease to the researcher. These understandings can ecstasy the researchers to attain the urgent exploration for the design of novel drugs to treat and prevent the breast cancer.

2. Data Source

Among the available 2948 over expressed genes that involve in breast cancer, 487 genes that play significant role in breast cancer were analysed by oncomine webserver. The gene summary of these 487 genes was retrieved NCBI database¹⁰. The protein sequences, functions and their Post Translational Modifications (PTM’s), structures, interpro signatures and the related publications were retrieved from UniProt database and manually curated based on their Gene ontology and keywords¹¹. The protein synonyms, families, location on chromosome and the homologous gene information based on the primary data source Human Gene Nomenclature Committee (HGNC) id were presented in the database¹².

3. Database Implementation

BCGDB was designed by using MySQL as background relational database to store the information of the gene sequences and their corresponding proteins. The PHP programming and certain modules of DBI and CGI, were used for query processing. Client-side JavaScript were employed for making it user-friendly.

4. Features

BCGDB is designed as an ecstasy database to provide significant information on Breast cancer genes that will augment the researchers to design the novel inhibitors for breast cancer treatment. The database is designed as such one can search the gene and its corresponding protein information. The database can be queried by using gene ontology term or keyword or gene name or protein name. The search based on gene provides the information such as gene summary, gene name, gene ontology terms, gene ontology ids, chromosome locations and keywords, NCBI Id, Variants and Sequences in FASTA format.

The protein search provides the information such as protein name, alternate names, associated disease, interactions, molecular class, molecular function, biological process, protein families, domains and motifs, post translation modifications, molecular weight, amino acid composition, function, Gene ontology terms based on Biological Process, Molecular function, cellular component, interpro signatures, associated pathways, 3D structures and sequences in FASTA format. The database is also provided with an option to visualize the 3D structures of the respective proteins.

5. Database Accessibility

The BCGDB is made available at <http://www.bcgdb.in>. The users are allowed to query the database by gene symbol (ABHD5) gene name (abhydrolase domain containing 5) or its alias names (CDS; CGI58; IECN2; NCIE2) or by HGNC id (HGNC:21396) or Gene Ontology (1-acylglycerol-3-phosphate O-acyltransferase activity; cell differentiation) or GO-ID (GO:0003841; GO:0030154) or chromosome location (3p21), keywords (Acetylation; Cataract; Disease mutation, Ichthyosis) and the protein search by Uniprot ID (Q8WTS1) or protein name (1-acylglycerol-3-phosphate O-acyltransferase) or Human Protein Reference Database (HPRD) ID (HPRD:05308; HPRD:11986) or Online Mendelian Inheritance in Man id(MIM:604780). To obtain the information of breast cancer genes and related protein sequences the Hypertext Pre-processor Scripting language is used to fetch data from MySQL database and the database is also embedded with server-side HTML

scripting language to respond on the user's query to build dynamic pages. The BCGDB Home page, query based search on gene-ontology terms, gene name and protein name to retrieve the information and the result page were shown in Figure 1.

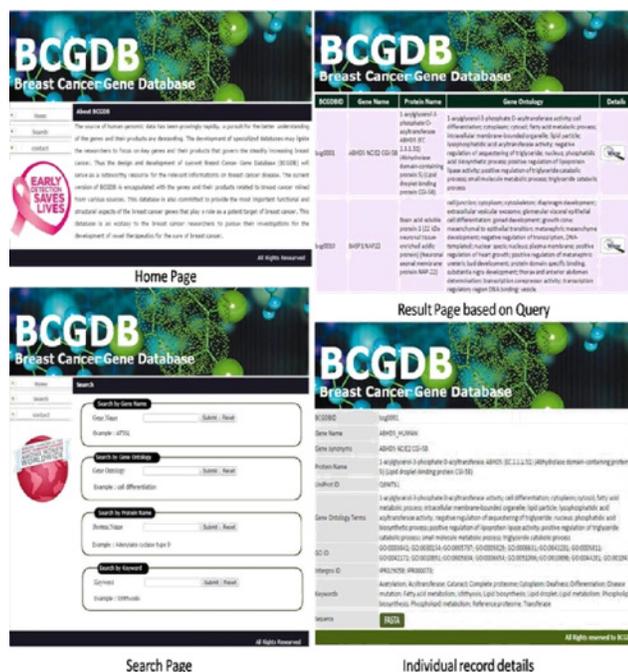


Figure 1. BCGDB Home page with sample search result.

6. Future Directions

The present database BCGDB is committed to provide the rational information on most vital genes that play a significant role in Breast cancer. The current release is focused on only 487 over expressed gene information. The database will be regularly updated and in the near future it is aimed to update with all the 2287 over expressed genes and 1253 under expressed gene information. Further the database will be enhanced to accept the new submissions by researchers from worldwide. In future the database is aimed to serve as a specialised database that focuses on the differential gene expression data analysis for breast adenocarcinoma. As the expression of these genes and their associated pathways plays a vital role in the cause of adenocarcinoma, we have future plan to include much

details on gene regulation and drug information and their interactions.

7. Conclusion

Globally, breast cancer is the most frequently diagnosed tumours among women, the development of Breast Cancer Gene Database (BCGDB) will serve as a remarkable resource. This ecstasy database provided with genes and their corresponding protein that was associated with breast cancer can serve as comprehensive database for the cancer researchers. In future this database might serve as a nodal resource for the selection of novel targets for the design of novel inhibitors for the treatment of Breast cancer.

8. References

1. Jemal A, Bray F, Centre MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin.* 2011; 61:69–90.
2. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics. *CA Cancer J Clin.* 2005; 55:74–108.
3. Hery C, Ferlay J, Boniol M, Autier P. Changes in breast cancer incidence and mortality in middle-aged and elderly women in 28 countries with Caucasian majority populations. *Ann Oncol.* 2008; 19:1009–18.
4. Auber J-bar-hen A, Daudin JJ, Robin S. Determination of differentially expressed genes in microarray experiments using local FDR. *BMC bioinformatics.* 2004; 5:1–9.
5. Susan G. Komen Foundation. Breast cancer facts PDF. 2014.
6. Benson JR, Jatoi I, Keisch M, Esteva FJ, Makris A, Jordan VC. Early breast cancer. *Lancet.* 2009; 373:1463–79.
7. Oldenburg RA, Meijers-Heijboer H, Cornelisse CJ, Devilee P. Genetic susceptibility for breast cancer: How many more genes to be found? *Crit Rev Oncol Hematol.* 2007; 63:125–49.
8. Turner N, Tutt A, Ashworth A. Hallmarks of 'BRCAness' in sporadic cancers. *Nat Rev Cancer.* 2004; 4:814–9.
9. Rakha EA, Reis-Filho JS, Ellis IO. Basal-like breast cancer: A critical review. *J Clin Oncol.* 2008; 26:568–81.
10. Available from: <http://www.ncbi.nlm.nih.gov>
11. The Gene Ontology Consortium. Gene ontology: Tool for the unification of biology. *Nat Gesnet.* 2000 May; 25(1):25–9.
12. Eyre TA, Ducluzeau F, Sneddon TP, Povey S, Bruford EA, Lush MJ. The HUGO gene nomenclature database. *Nucleic Acids Res.* 2006 Jan 1. 34 (Database issue):D319–21.