# BCDB - A database for breast cancer research and information

## Jayashree Mohandass [1], Saranya Ravichandran [1], Koganti Srilakshmi [1], Chinnasamy Perumal Rajadurai [2], Sangeetha Sanmugasamy [2], Gopal Ramesh Kumar [2]*

[1]Department of Bioinformatics, Guru Nanak College, Chennai, India; [2]Bioinformatics Lab, AU-KBC Research Centre, M.I.T Campus of Anna University, Chennai 600044, India; Gopal Ramesh Kumar - E-mail: gramesh@au-kbc.org; Phone: +91-44-2223 2711; Fax: +91-44-2223 1034/ 7073; * corresponding author

**Abstract:**
In pursuit of a better updated source including 'omics' information for breast cancer, Breast Cancer Database (BCDB) has been developed to provide the researcher with the quick overview of the Breast cancer disease and other relevant information. This database comprises of myriad of information about genes involved in breast cancer, its functions and drug molecules which are currently being used in the treatment of breast cancer. The data available in BCDB is retrieved from the biomedical research literature. It facilitates the user to search information on gene, its location in chromosome, functions and its importance in cancer diseases. Broadly, this can be queried by giving gene name, protein name and drug name. This database is platform independent, user friendly and freely accessible through internet. The data present in BCDB is directly linked to other on-line resources such as NCBI, PDB and PubMed. Hence, it can act as a complete web resource comprising gene sequences, drug structures and literature information related to breast cancer, which is not available in any other breast cancer database.

**Availability:** The database is freely available at: http://122.165.25.137/bioinfo/breastcancerdb/

**Keywords:** Breast cancer; database; genes; protein; drug; PDB structure; FASTA

**Background:**
Breast cancer is the second most common types of cancer and it is found to be the fifth major cause of cancer death [1] and it starts from the cells in the breast and spreads to other regions. The initial symptom of breast cancer is a lump formed in the breast which is different from surrounding tissue. The primary risk factors are identified as age, sex, high-fat diet, alcohol intake and obesity. The most common screening method for breast cancer is a combination of X-ray mammography and Breast examination.

According to the American Cancer Society about 1.3 million women are diagnosed annually with breast cancer [2]. Early detection has been developed for breast cancer during the past few years and national screening programmes for breast cancer are introduced in many of the countries. Digital technology and blood test are being performed for early detection of the disease [3]. Herceptin and Tamoxifen are the major drugs used in the treatment of Breast Cancer currently [4 - 5].

The number of breast cancer cases arose tremendously in recent years, thus there is a need to create a comprehensive database which holds all the information pertaining to breast cancer. Breast cancer database by Institute of Bioinformatics, Breast cancer Information core by National Human genome Research Institute and the database by National Cancer Institute are the major databases available in web for Breast cancer Information [6 - 7].

The current release of Breast cancer database (BCDB) provides information about risk factors, symptoms, diagnosis and treatment taken from relevant literatures. Links to all other breast cancer resources in the web are linked. The number of genes responsible for breast cancer is gradually increasing and the main bottleneck in breast cancer database creation is to give the complete information about the disease to the researcher. This challenge has been overcome partly in the BCDB, which holds the summary of genes and their corresponding proteins involved in breast cancer along with the drugs used in the treatment. It helps the researcher to accomplish molecular docking and drug designing studies.

**Methodology:**
It is found that 2948 genes are involved in breast cancer [8]. Literature survey and analysis revealed that 466 genes play an imperative role in causing the disease. Information on those 466 genes which play active role in breast cancer was retrieved from NCBI (National Center for Biotechnology Information) database [9]. The data are normalized to reduce and eliminate redundancy. The protein functional information was extracted from UniProt database which is curated manually [10]. About 60 drug molecules are found to be used in the treatment of breast cancer and this information was obtained from DrugBank database [11]. The structures of proteins were extracted from PDB (Protein Data Bank) which is a world wide repository of information about the three dimensional structures of large biological molecules [12].

PHP (Hypertext preprocessor) has been used as a front-end tool owing to its ability to be used as a general purpose scripting language, especially suited for web applications and it can be embedded in HTML. MySQL was applied to construct a back-end, because of its capability to support very large databases. It is compatible with vast majority of operating systems (UNIX, Windows) and a powerful relational database management system (RDBMS). JavaScript is used for client-side validation purpose. The Database was developed using WAMP5 (Windows Apache MySQL PHP) package **[13].**

**Database feaures:**
BCDB acts as a complete web source for Breast cancer information. The database can be searchable at gene, protein and drug level. The genes can be searched by providing the gene name or taxonomic lineage and the proteins can be queried by offering the Protein name, accession number, PDB ID, and Structure name. The drug can be browsed by providing the drug name, drug bank id, IUPAC name, and molecular weight. Resources page includes Breast cancer clinical trials, risk calculator, tutorials, Microarray datasets for Breast cancer, Hospitals, Research Institutes, and Journals. Breast cancer research community can be browsed through this web page.

The Breast cancer database is made up of three components: Genes, proteins and drugs. Gene part contains entire set of genes involved in breast cancer. Viewers can also browse through individual chromosomes and view the genes involved in breast cancer for quick reference. The gene section is carefully designed by taking 12 important parameters into consideration that comprises gene id, gene name, gene type, gene symbol, organism, taxonomic lineage, gene function, proteins coded by the gene, gene description, other names and NCBI link for retrieving gene sequence. PubMed reference is also provided for individual genes **(Figure 1 & Figure 2).**

Protein table consists of the protein sequence in FASTA format, PDB file and the protein structure which gives unique appearance to the database. Drug table encloses information about 60 drugs taken from drug bank database which is created based upon 16 different parameters which includes drug id, drugbank id, drug name, drug category, IUPAC name, molecular weight, action, description, logP value, indication, pharmacology, toxicity, protein binding and biotransformation. The chemical structure of the drug and its PDB file is also available in the database. Search box is also provided in BCDB. User can query the database using genename (or) geneid, protein name (or) Proteinid, and drug name.

**Future developments:**
The database will be updated periodically so as to ensure that users get latest information on breast cancer. The database will be linked to other breast cancer databases in near future for easy accessing of information.

**Conclusion:**
Breast Cancer Database (BCDB) gives the comprehensive information about the disease to a researcher and general awareness to a common man. It is a comprehensive database with data related to every aspect of breast cancer and access to genes and proteins involved in breast cancer. Breast cancer glossary is linked with the database which describes all the terms related to breast cancer. Latest news on breast cancer is also provided in the database to help the user.
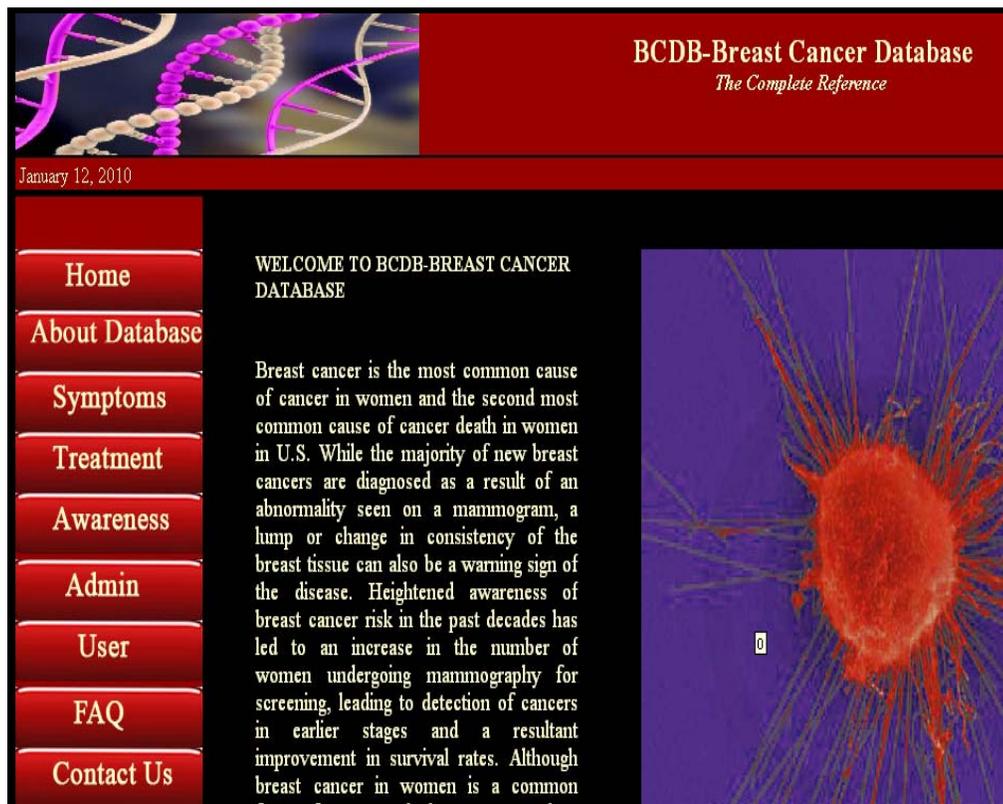


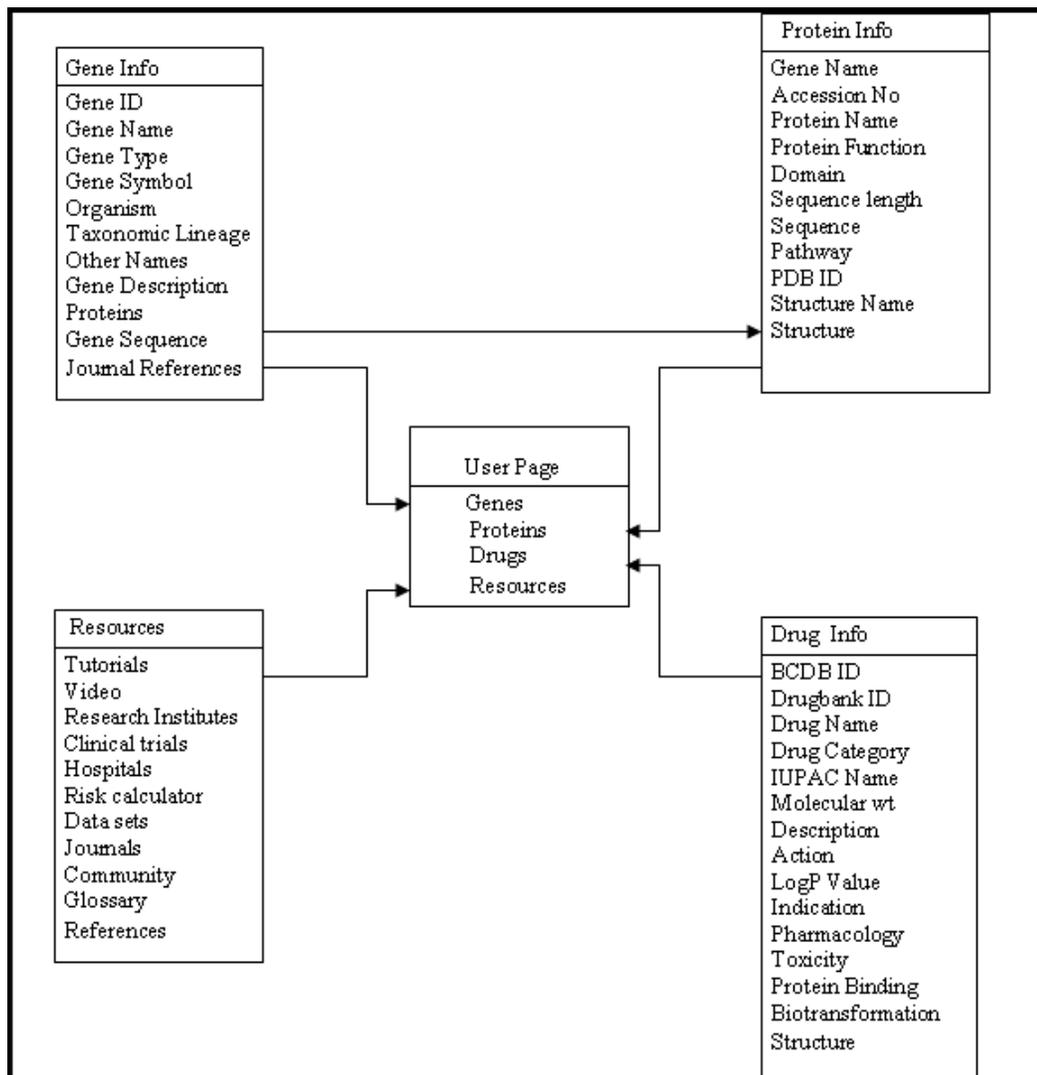**Figure 1:** Home Page of BCDB database

**Figure 2:** Schematic diagram representing BCDB

**References:**
[1]    http://en.wikipedia.org/wiki/Breast_cancer
[2]    http://www.imaginis.com/breasthealth/statistics.asp
[3]    http://www.netdoctor.co.uk/womenshealth/breastcancernews.htm
[4]    L Gu *et al*. *PLoS One* (2009) **4:**7.
[5]    R Hattar *et al*. *Breast Cancer Research* (2009) **11:**1.
[6]    http://research.nhgri.nih.gov/bic/
[7]    http://www.cancer.gov/cancertopics/types/breast
[8]    http://www.breastcancerdatabase.org/
[9]    http://www.ncbi.nlm.nih.gov/
[10]   http://www.uniprot.org/
[11]   http://www.drugbank.ca/
[12]   http://www.rcsb.org/pdb/home/home.do
[13]   http://wamp5.en.softonic.com