Journal of Computational Intelligence in Bioinformatics
ISSN 0973-385X Volume 5, Number 3 (2012) pp. 161-167

http://www.ripublication.com/jcib.htm

ECDB: A Database for Endometrial Cancer

[*]Shyam Perugu, Purushottam Naidu Bade and Shaik Sarafaraz Nawaz

[*]Assistant Professor, School of Information Technology,
Jawaharlal Nehru Technological University,
Hyderabad, Andra Pradesh-500085, India
[*]Corresponding Author E-mail: shyamperugu@jntuh.ac.in

Abstract

Endometrial Cancer Database in short (ECDB) is an integrated knowledge database has been developed to provide the researcher with quick overview of the Endometrial Cancer Disease and other relevant information. This database comprises of myriad of information about genes, protein and its functions involved in endometrial cancer. The data available in ECDB is retrieved from the biomedical research literatures; it facilitates the user to search information on gene, its location on chromosomes, functions and its importance in cancer disease. This database is platform independent, user friendly and freely accessible through internet. The data present in ECDB is directly linked to other online resources such as NCBI, PDB, and Pubmed. ECDB provides researchers with easy access to the latest information on genes, protein and drugs involved in endometrial cancer.

Keywords: Endometrial Cancer, Database, Genes, Proteins, Drug

Introduction

Endometrial cancer is one of the most common cancers in women, placing 4th in incidence just after Breast cancer, Lung cancer and Colon cancer. Endometrial cancer starts from the cell in the endometrium, the inner lining of the *uterus* (womb) and spreads to other regions. It can affect women of all ages, but it is most often occurs after the reproductive years. Endometrial cancer is the most common gynecologic cancers in the United States, with over 35,000 women diagnosed each year. [1-3]

There are numerous factors affecting the development of endometrial cancer, often linked to the hormonal imbalance. The first sign of disease is usually an abnormal vaginal bleeding [2] after menopause (cessation of menstrual cycles). In

addition, symptoms may include pelvic pain, pain during intercourse and weight loss, null parity (women who never gave birth), irregular menstrual periods, obesity, diabetes mellitus, hormone replacement therapy and ovarian cancer. In addition, age, race, therapy with tamoxifen (Endometrial cancer medication) and the presence of Hereditary No Polyposis Colorectal Cancer (HNPCC) also influence the development of endometrial cancer [3-4].

The number of endometrial cancer cases arose tremendously in recent years, thus there is a need to create a comprehensive database which holds all the information pertaining to endometrial cancer. Endometrial cancer information core by National Human Genome Research Institute (NHGR) and the database by National Cancer Institute (NCI) are the major databases available in web for endometrial cancer information [5-12].

The current release of endometrial cancer database (ECDB) provides information about risk factors, symptoms, diagnosis and treatment taken from relevant literatures. Links to all other cancer resources in the web are linked. The number of genes responsible for endometrial cancer is gradually increasing and the main bottleneck in endometrial cancer database creation is to give the complete information about the disease to the researcher. This challenge has been overcome partly in the ECDB, which holds the summary of genes and their corresponding proteins involved in endometrial cancer along with the drugs used in the treatment. It helps the researcher to accomplish drug design and molecular docking studies [13-18].

Methodology

Data Collection

For genes involved in Endometrial Cancer we have extensively searched pubmed database and collected relevant literature. We have identified 106 genes which play an imperative role in causing the diseases. A national wide cohort study in Sweden by *Moradi et al* found that 178 genes involved in endometrial cancer. The protein functional information was extracted from UniPort database which is curated manually (19). About 11 drugs molecules are found to be used in the treatment of endometrial cancer and this information was obtained from Drug Bank Database (20). The structures of protein were extracted from PDB (Protein Data Bank) which is a worldwide repository of information about the three dimensional structures of large biological molecules.

Hypertext Markup Language (HTML) 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. 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 jsp, servlet package and Tomcat server.

Database features

ECDB acts as a complete web source for endometrial 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 endometrial cancer clinical trials, symptoms, medication, causes, prevention, awareness, and patient discussions for endometrial cancer, hospitals, research institutes, and journals. Endometrial cancer research community can be browsed through this web page.

The Endometrial cancer database is made up of three components: Genes, Proteins and Drugs. Gene part contains entire set of genes involved in endometrial cancer. Viewers can also browse through individual gene by using search and view the genes involved in endometrial 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 and 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 part consists of the protein name, PubMed abstract, molecule, polymer, protein-ligand structure which gives unique appearance to the database. Drug part encloses information about 11 drugs taken from drug bank database which is created based upon 15 different parameters which includes drug bank id, drug name, drug category, IUPAC name, molecular weight, action, description, brand name, side effects, chemical name, toxicity, protein binding and other attributes. The chemical structure of the drug and its PDB file is also available in the database. Search box is also provided in ECDB. User can query the database using gene name (or) gene id, protein name (or) protein id, and drug name. Endometrial cancer glossary is linked with the database which describes all the terms related to endometrial cancer. Latest news on endometrial cancer is also provided in the database to help the user on latest developments in this field.

Future developments

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

Discussion and Conclusion

Large number of cancer databases has been published, yet to date to our knowledge, there is no resource available that provides detailed information about the genes, protein and drugs known to be associated with endometrial cancer. Therefore we have developed ECDB, where manual curation along with information from other resources have been linked to provide a information related to genes and proteins involved in endometrial cancer. Information present in our database is supported by many literatures references as documented in Pubmed. Therefore, we feel that our database will complement the existing databases.

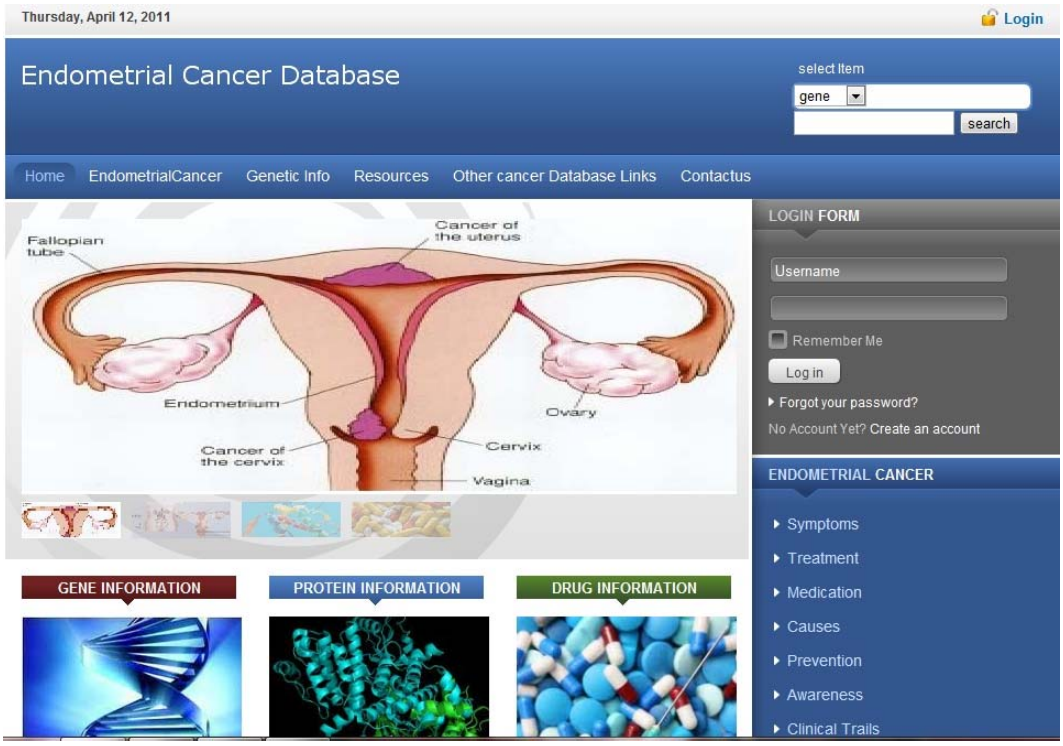
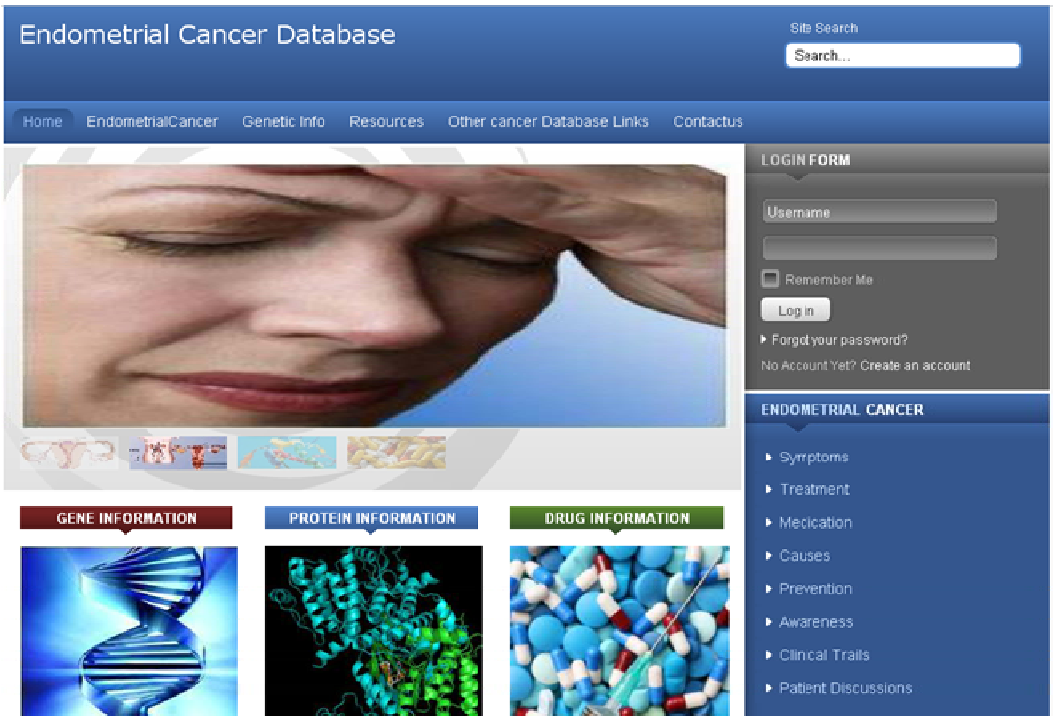


Figure 1: Home Page of ECDB database



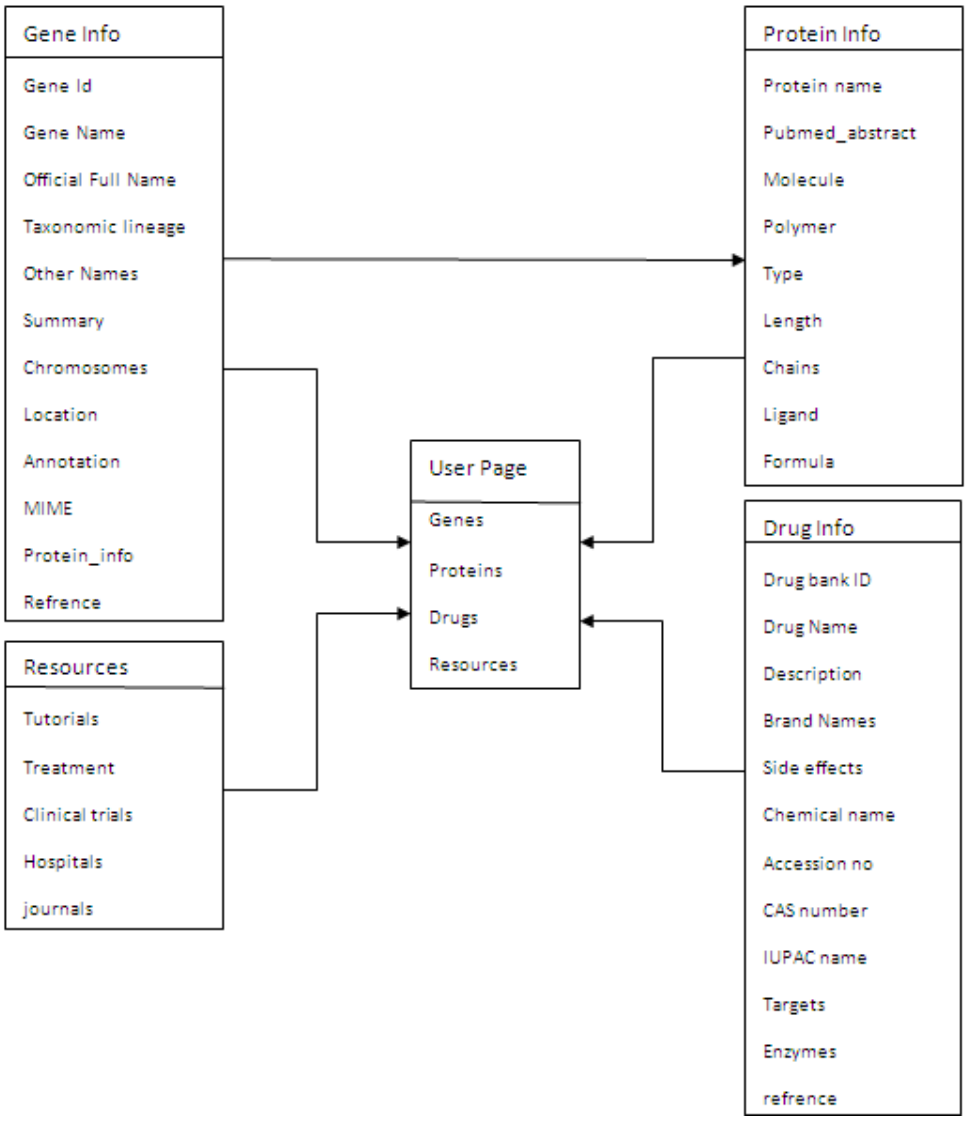


Figure 2: Schematic representation of ECDB

References

[1] Kari Hemminki, Pauli Vaittinen and Chuanhui Dong, 1999, "Endometrial Cancer in the Family-Cancer Database," *Cancer Epidemiol Biomarkers Prev*, 8(1), pp.1005-1010.

[2] Sara Wedrén, Lovisa Lovmar, Keith Humphreys, Cecilia Magnusson, Håkan Melhus, Ann-Christine Syvänen, Andreas Kindmark, Ulf Landegren, 2008, "Estrogen receptor alpha gene polymorphism and endometrial cancer risk – a case-control study,".

[3] Burke, T., Eifel, P., and Muggia, F., 1997, "Cancer of the uterine body," Cancer: Principles and Practice of Oncology, pp. 1478–1499.

[4] Lyon, 1996 “IARC Monographs on the Evaluation of Carcinogenic Risks to Humans,” IARC, 66, pp. 260–280.
[5] Altekruse SF, Kosary CL, Krapcho M, 2010, “SEER Cancer Statistics Review,” , National Cancer Institute. Bethesda.
[6] Shoff, S., and Newcomb, P., 1998, “Diabetes, body size, and risk of endometrial cancer,” Am. J. Epidemiol., *148, pp.* 234–40.
[7] Olson, S., Vena, L., Dorn, J., Marshall, J., Zielezny, M., Laughlin, R., and Graham, S., *1997,* “Excercise, occupational activity, and risk of endometrial cancer,” Ann. Epidemiol., *pp.*46–53.
[8] Moradi, T., Nyren, O., Bergstro¨m, R., Gridley, G., Linet, M., Wolk, A., Dosemeci, M., and Adami, H-O., 1998, “ Risk for endometrial cancer in relation to occupational physical activity: a nationwide cohort study in Sweden,” Int. J. Cancer, *76, pp.* 665–670.
[9] World Cancer Research Fund. Food, Nutrition and the Prevention of Cancer: a global perspective. Washington, DC: American Institute of Cancer Research, 1997.
[10] Lynch, H., and Smyrk, T., 1996 “Hereditary nonpolyposis colorectal cancer (Lynch syndrome),” Cancer, *78, pp.* 1149–1167.
[11] Aaltonen, L., Salovaara, R., Kristo, P., Canzzian, F., Hemminki, A., Peltoma ¨ki, P., Chadwick, R., Ka¨a¨ria¨inen, H., Eskelinen, M., Ja¨rvinen, H., Mecklin, J-P., and De La Chapelle, A., 1998 “Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease,” N. Engl. J. Med., *338, pp.*14811–14817.
[12] Aarnio, M., Mecklin, J-P., Aaltonen, L., Nystro¨m-Lahti, M., 1995, “Life-time risk of different cancers in hereditary non-polyposis colorectal cancer(HNPCC) syndrome,” Int. J. Cancer, *64,pp.* 430–433.
[13] Boland, C, B. Vogelstein and K. Kinzler , 1998, “The Genetic Basis of Human Cancer,” New York:McGraw-Hill, pp.333–346.
[14] Aarnio, M., Sankila, R., Pukkala, E., Salovaara, R., Aaltonen, L., De La Chapelle, A., Peltoma¨ki, P., Mecklin, J-P., and Ja¨rvinen, H., 1999, “Cancer risk in mutation carriers of DNA-mismatch-repair genes,” Int. J. Cancer, *81,pp.* 214–218.
[15] Pal, T., Flanders, T., Mitchell-Lehman, M., MacMillan, A., Brunet, J., Narod, S., and Foulkes, W., 1998 “Genetic implications of double primary cancers of the colorectum and endometrium,” J. Med. Genet., *35, pp.* 978–984.
[16] Risinger, J., Haayes, A., Berchuck, A., and Barrett, J., 1997, “PTEN/MMAC1 Mutations in endometrial cancer,” Cancer Res., *57, pp.* 4736–4738.
[17] Gruber, S., Thompson, W., 1996, “A population-based study of endometrial cancer and familial risk in younger women,” Cancer Epidemiol. Biomarkers Prev., *5, pp.* 411–417.
[18] Fornasarig, M., Campagnutta, E., Talamini, R., Franceschi, S., Boz, G.,Scarabelli, C., Andreaus, C. M., Scozzari, G., and Valentini, M. 1998, “Risk factors for endometrial cancer according to familial susceptibility,”Int. J. Cancer, *77, pp.*29–32.

[19] Hemminki, K., and Vaittinen, P., 1998, “Familial breast cancer in the Family-Cancer Database,”. Int. J. Cancer, *77*, *pp.*386–391.
[20] Hemminki, K., Vaittinen, P., and Kyyro¨nen, P., 1998, “Age-specific familial risks in common cancers of the offspring,”. Int. J. Cancer, *78*, *pp.* 172–175.

Web References

[21] http://en.wikipedia.org/wiki/Endometrial_cancer
[22] http://www.cancer.gov/
[23] http://www.medicinenet.com/endometriosis/article.htm#tocb
[24] http://www.emedicinehealth.com/endometriosis/article_em.htm
[25] http://www.endometrialdatabase.com/
[26] http://www.ncbi.nlm.nih.gov/
[27] http://www.uniprot.org/
[28] http://www.drugbank.ca/
[29] http://www.rcsb.org/