BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Database

Volume 6(4)

Database and interaction network of genes involved in oral cancer: Version II

Nikhil Sureshkumar Gadewal[1, 2], Surekha Mahesh Zingde[2]*

[1]Bioinformatics centre (BTIS Sub-DIC), ACTREC, Kharghar, Navi Mumbai 410210, India; [2]Advanced Centre for Treatment Research and Education in Cancer, Tata Memorial Centre, Kharghar, Navi Mumbai, 410210, India; Surekha Mahesh Zingde - Email: szingde@actrec.gov.in; Phone: +91-22-27405034; Fax: +91-22-27405095; *Corresponding author

Received April 13, 2011; Accepted April 25, 2011; Published May 07, 2011

Abstract:
The oral cancer gene database has been compiled to enable fast retrieval of updated information and role of the genes implicated in oral cancer. The first version of the database with 242 genes was published in *Online Journal of Bioinformatics 8(1), 41-44, 2007*. In the second version, the database has been enlarged to include 374 genes by adding 132 gene entries. The architecture and format of the database is similar to the earlier version, and includes updated information and external hyperlinks for all the genes. The functional gene interaction network for important biological processes and molecular functions has been rebuilt based on 374 genes using 'String 8.3'. The database is freely available at http://www.actrec.gov.in/OCDB/index.htm and provides the scientist information and external links for the genes involved in oral cancer, interactions between them, and their role in the biology of oral cancer along with clinical relevance.

Keywords: Oral cancer, Database, Gene interaction network

Background:
Head and neck cancer is the sixth common malignancy and is the major cause of cancer morbidity and mortality worldwide. In India, cancers of Head and Neck comprise ~24.1 % of total cancers seen at Tata Memorial Centre, Mumbai of these ~13.2 % are from the oral cavity [K.A. Dinshaw and B. Ganesh, Annual Report 2002-2005, Hospital based cancer registry, Tata Memorial Hospital, 2008]. Research on oral cancer is receiving increasing attention and it is often reported under Head and Neck cancer. In the post genomic era, efforts to understand oral cancer are aimed at obtaining genomic and proteomic profiles and correlating them with clinical presentation. This has generated enormous data, which needs to be organized effectively in the form of a database, to optimally utilize the information for diagnosis, prognosis and treatment. In the present database the genes which are listed are reported to be altered in oral cancer.

Methodology:
The database published in 2007 **[1]** was updated by adding 132 new genes searched from the PubMed database using the MESH words "genes AND oral cancer". Genes which are not mentioned in the PUBMED abstracts were obtained from 10 full-text articles for proteomics studies retrieved using MESH words "oral cancer AND proteomics" and "oral cancer and autoantibodies" **[2-8]** as well as from "head and neck cancer AND proteomics" and "head and neck cancer AND autoantibodies" **[9-11]**. The genes are presented in alphabetical order in the *gene-list*, with links to the *gene-info* page wherein detailed information of the gene is available through hyperlinks which connect to specific databases for complete information. **Figure 1** summarizes the databases used for mining information presented in the gene-info page. The detailed procedure for mining is provided as supplementary information on the website http://www.actrec.gov.in/OCDB/ Supp_Info.htm. The searchable content of all the genes is stored in the MySQL database at the back-end and queries are handled by PHP at the front-end. The database is hosted on Linux operating system run by an Apache server. The interaction network of the 374 genes was obtained by submitting the genes to the FATIGO tool which classifies the genes according to biological processes and molecular functions on the basis of gene ontology **[12]**. The genes involved in particular biological processes and molecular functions were submitted to 'String 8.3' tool to generate functional protein-protein interaction networks using the 'Text mining' parameter and the default score was increased so as to obtain interaction networks with higher confidence **[13]**. The interaction network for each biological process and molecular function was downloaded in SVG format and edited using Inkscape software to add the hyperlink for the PubMed abstract depicting the functional relationship between two interacting genes / proteins.

Database features:
The *gene-list* page shows the list of all 374 genes alphabetically arranged with gene symbol, gene description and two hyperlinks to the PubMed references for each gene. The first link has the keyword 'Oral Cancer' and second link is with 'Head and Neck Cancer'. This ensures that information related to the molecule under consideration is retrieved from both the headings where it is generally reported. Clicking on the name of the gene on the *gene-list* page, opens up the gene-info page, which provides detailed information on aliases, description, chromosomal location, mutations and SNPs, mRNA expression, protein information, pathways involved and interacting proteins, expression of genes in different tissues, and clinical correlates.

The second part of the database is *Keyword search* from which specific features of the genes can be retrieved by querying the database and the results are displayed dynamically. Features include 1) *Gene name*: Since there are alternative names for the genes, a search can be performed by any gene symbol or alias or gene name of a particular gene. 2) *Chromosomal location*: On the basis of the chromosome number a search can be performed which provides the

 © 2011 Biomedical Informatics

list of genes with their location on a particular chromosome. 3) *CGH*: The percentage of gains and losses on a chromosomal region can be used as input parameters (in a given range) to locate the genes responsible for the aberrations. 4) *Molecular weight*: A range of molecular weights (in dalton) can be used to list out the genes in the required range. 5) *Advanced search*: A multiple search can be performed using chromosomal region, CGH data and molecular weight options to obtain a list of genes matching all options.

	Database	Information provided
Alias	**NCBI** [http://www.ncbi.nlm.nih.gov/]	Other abbreviations and names for a gene.
Description	**SwissProt** [http://us.expasy.org/sprot/]	Description on the function of the gene.
Chromosomal location	**NCBI** [http://www.ncbi.nlm.nih.gov/] and **Progenetix** [http://www.progenetix.org/cgi-bin/pgHome.cgi]	Chromosome location, size and DNA sequence, the percent gains and losses of the chromosome regions by CGH technique.
Mutations and SNPs	**HGMD** [http://www.hgmd.org/] and **dbSNP** [http://www.ncbi.nlm.nih.gov/SNP/]	External link for mutation and SNPs reported for the gene.
mRNA	**NCBI** [http://www.ncbi.nlm.nih.gov/] and **CGAP** [http://cgap.nci.nih.gov/]	Sequences of known transcripts, variants and cDNA libraries for the expression in different tissues.
Protein	**SwissProt** [http://us.expasy.org/sprot/], **Swiss2D** [http://us.expasy.org/ch2d/], **PBD** [http://www.rcsb.org/pdb/home/home.do], and **dbPTM** [http://dbptm.mbc.nctu.edu.tw/]	Size, molecular weight, domains, sequences, estimated location of the protein on 2D-PAGE maps, 3D structure and PTMs.
Pathway and Interactions	**BioCarta** [http://www.biocarta.com/genes/index.asp] or **KEGG** [http://www.genome.jp/kegg/pathway.html] and **DIP** [http://dip.doe-mbi.ucla.edu/] or **IntAct** [http://www.ebi.ac.uk/intact/main.xhtml]	Role of gene in metabolic pathway and its interaction with different proteins.
Tissue	**GeneNote** [http://bioinfo2.weizmann.ac.il/cgi-bin/genenote/home_page.pl], **PubMed** [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed] and **HPA** [http://www.proteinatlas.org/index.php]	Whole genome expression profiles in normal and cancer tissue, *PubMed* references for the gene expressed in tissues, antibody information for normal and cancer tissues and cell lines.
Clinical correlates	**OMIM** [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM] and **PubMed** [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed]	Genetic disorder associated with gene, *PubMed* references presenting the potential of gene to be a marker for early detection, diagnosis, prognosis and therapy.

Figure 1: The list of databases used for mining the information for the genes on the gene-info page.

The special feature of the database is the interaction network. The protein-protein interaction networks of the genes involved in various biological processes and molecular functions provide clues to genes / proteins which regulate a given biological process. The earlier database **[1]** has been cited for its application in predicting the possible role of differentially expressed markers in cell transformation identified by Govekar *et al.* **[14]**. The markers identified have now been analyzed using String 8.3 database alone and in combination with the genes listed in the apoptosis process. The interaction networks so obtained are provided as supplementary information on the website http://www.actrec.gov.in/OCDB/Supp_Info.htm. Each node represents a gene and the lines connecting the nodes indicate the probable relationship between them in the apoptosis process. By clicking on the line connecting the genes, a link is provided to one PubMed reference as an example, although a similar search through the String database will fetch more articles. This network shows several interacting proteins involved in the apoptosis pathway and broadens the scope for further investigations related to oral carcinogenesis. In conclusion, this database provides the scientist information and external links for the genes involved in oral cancer, interactions between them, and their role in the biology of oral cancer along with clinical relevance. The external links ensure that new information is continuously available.

Acknowledgements:
We acknowledge Cynthia Martin, Ashwathy Sebastian, Jissy Jose, and Poornima Rathod who helped in the data mining. The database was generated using the infrastructure available in the BTIS facility, funded by Department of Biotechnology, Govt. of India, at ACTREC.

References:

[1] Gadewal NS & Zingde SM. *Online Journal of Bioinformatics* 2007 **8:** 41
[2] Patel V *et al. Crit Rev Oral Biol Med.* 2001 **12**: 55 [PMID:11349962]
[3] Chen J *et al. Proteomics* 2004 **4**: 2465 [PMID: 15274141]
[4] He QY *et al. Proteomics* 2004 **4**: 271 [PMID: 14730689]
[5] Lo WY *et al. Clin Chim Acta.* 2007 **376**: 101 [PMID: 16889763]
[6] Shukla S *et al. Proteomics Clin Appl.* 2007 **1**: 1592 [PMID: 21136657]
[7] Patel V *et al. Clin Cancer Res.* 2008 **14**: 1002 [PMID: 18281532]
[8] Wang Z *et al. BMC Genomics.* 2009 **10**: 383 [PMID: 19691830]
[9] Gires O *et al. Cell Mol Life Sci.* 2004 **61**: 1198 [PMID: 15141305]
[10] Roesch-Ely M *et al. Oncogene* 2007 **26**: 54 [PMID:16819514]
[11] Lin HS *et al. Cancer Epidemiol Biomarkers Prev.* 2007 **16**: 2396 [PMID: 18006929]
[12] Al-Shahrour F *et al. Bioinformatics* 2004 **20**: 578 [PMID: 14990455]
[13] Jensen LJ *et al. Nucleic Acid Res.* 2009 **37**: 412 [PMID: 18940858]
[14] Govekar RB *et al. Proteomics Clin Appl.* 2009 **3**: 1451 [PMID: 21136964]

Edited by P Kangueane
Citation: Gadewal & Zingde. Bioinformation 6(4): 169-170 (2011)
License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.