

GICDB: an oncogenomic database of gastrointestinal cancer

Jyotsna Choubey*, Jyoti Kant Choudhari, Ashish Patel and M. K. Verma

Gastrointestinal Cancer Database (GICDB) is a repository of genetic association studies on gastrointestinal cancer which allows researchers to interpret the risk factors of the disease accurately and provide new insights towards further work. GICDB has been developed to provide a platform to clinicians and researchers for the easy retrieval of appropriate information. All data are stored and managed in MySQL and web interface has been developed in ASP. Currently, the database integrates information of 463 genes and 713 miRNA; these genes are implicated in the diverse phases of various gastrointestinal (GI) cancers. GICDB is distinct from other databases in that: (i) it consists of information on genes and miRNAs involved in various GI cancers; (ii) data in GICDB are linked to other on-line resources such as Entrez, Ensemble, UniGene, Swiss-Prot and On-Line Mendelian Inheritance in Man; (iii) there is information on GI cancer-associated drugs retrieved from DrugBank, and (iv) it also integrates the information related to medicinal plants and their use in the treatment of different GI cancers with the structure of their chemical constituents. Unambiguously this database would provide a platform to retrieve genuine information for researchers working on GI cancer. GICDB can be freely accessed at <http://gidb.subdic-bioinformatics-nitr.in/>

Keywords: Gastrointestinal cancer, oncogenomic database, genes and miRNAs.

GASTROINTESTINAL (GI) cancer is common in humans. The increasing incidence of this disease causes mortality of patients above 70 years. However, elderly patients were not older patients included in a clinical trial by Leo *et al.*¹. GI cancer refers to cancer of the gastrointestinal tract (GI tract) and accessory organs of digestion (pancreas, liver and gall bladder). The symptoms are abnormal bleeding or other associated problems. It is diagnosed by endoscopy and biopsy of suspected tissue. The treatment of GI cancer is location-specific along with cancer cell type. Mortality rate is higher in cancer of GI tract and the accessory organs of digestion as compared to the cancer of any other system in the body^{2,3}. A total of 291,150 new GI cancer cases and 149,300 GI cancer deaths were projected for the United States in 2015 (ref. 4).

Cancer is a genetic disease which is caused by certain changes in the genes which control normal cell functioning. These changes in the genes may include DNA mutation. Two major classes of genes are associated in tumour augmentation – the proto-oncogenes and tumour suppressor genes (TSGs). As a result of mutation or dysregulation, proto-oncogenes transform into oncogenes which promote abnormal cellular proliferation. TSGs regulate

normal cell proliferation or mediate cell-specific differentiation. Inactivation of TSGs results in neoplastic growth. Several important genes are involved in GI carcinogenesis⁵.

At present, the cancer database is available organ-wise, e.g. liver cancer⁶, cancer of pancreas⁷, esophageal cancer⁸, etc. A particular organ system accomplishes a specific task, like respiratory system for respiration, digestive system for digestion, reproductive system for reproduction, etc. It is quite obvious that there may be fundamental peculiarity and similarity in biological functioning of organ system. Similarly, disease of organ of a particular organ system must also show peculiarity and similarity in their causes, characteristics and their possible remedial measure. With the above viewpoint in view, the gastrointestinal cancer database (GICDB) has been developed. This is a comprehensive oncogenomic database of GI cancer. It integrates information on protein coding genes, miRNAs, gene ontology, drug and drug targets and on plants with medicinal values that are utilized in the treatment of several types of GI cancer.

Description and construction of the database

GICDB consists of information on genes involved in various GI cancers, identified from the published literature. PubMed has been queried using different keywords

The authors are in the Sub-Distributed Informatics Centre of Bioinformatics, National Institute of Technology, Raipur 492 010, India.

*For correspondence. (e-mail: joys.choubey@gmail.com)

GICDB
Gastrointestinal Cancer Database

Home Gene miRNA Drug Medicinal Plant Statistics Research Team Contact Us Feedback

Search

Gastric Cancer
Gallbladder Cancer
Pancreatic Cancer
Esophageal Cancer
Liver Cancer
Colorectal Cancer

Welcome to GICDB

Welcome to GICDB! Gastrointestinal cancer (cancer of the digestive system) includes cancers of the esophagus, gallbladder, liver, pancreas, stomach, small intestine, large intestine (colon) and rectum. Some form of gastrointestinal cancer is newly diagnosed in more than 250,000 patients annually in the United States. This amounts to about 20 percent of all newly diagnosed cancers every year. The Gastrointestinal Cancer Database (GICDB) is a manually curated catalog of experimentally validated genes that are thought to be involved in the different stages of gastrointestinal cancer. GICDB contains information about genes and miRNA which can be used as potential biomarkers for gastrointestinal cancer. The database also contains the information of medicinal plants with structure of its active constituents used in the treatment of gastrointestinal cancer. The data in GICDB is extracted from the published research literature and stored as a collection of facts. GICDB is freely accessible for academic and non-profit users.

Liver Bile ducts Pancreas Small intestine Gallbladder Cystic duct Common hepatic duct Common bile duct Small intestine

Figure 1. Screenshot of database GICDB.

like gastric carcinoma, esophageal cancer, hepatocellular carcinoma, etc. and abstracts were retrieved. These PubMed-indexed articles were studied by biologists to identify the genes responsible for GI cancers. Then gene-related information such as gene name, gene id, chromosome location, etc. was retrieved from various sources like Entrez Gene⁹, GeneCards¹⁰ and OMIM¹¹ (Online Mendelian Inheritance in Man, 2007). miRNAs play an important role during the regulation of most pathways, although changes in the expression of miRNAs are connected with many pathologies of humans including cancer. When TSGs and oncogenes are being targeted, the key cellular process which determines the cell phenotype has been modulated by miRNAs¹². After key modulation, cell phenotype may be the potential therapeutic target. GICDB includes information on miRNA misregulation in GI cancer identified from the published literature.

Drug-related information has also been included in this database, obtained from DrugBank¹³. It provides chemical and pharmacological properties of drugs, drug targets, coordinate files for proteins and links of databases such as RxList (<http://www.rxlist.com/script/main/hp.asp>)¹⁴, KEGG drug (<http://www.genome.jp/kegg/drug/>)¹⁵, KEGG compound (<http://www.genome.jp/kegg/compound/>)¹⁵, PubChem compound (<http://www.ncbi.nlm.nih.gov/pccompound>)¹⁶, PubChem substance (<http://www.ncbi.nlm.nih.gov/pcsubstance>)¹⁵ and PharmGKB (<http://www.pharmgkb.org/>)¹⁷.

Information on plants with medicinal value used in the treatment of different GI cancers is also included in the database. Allopathic system of medicine is associated with side effects, high cost and resistance developed

against the recently used drugs. This has encouraged the utilization of plant products for various human diseases. Due to their efficacy and less side effects, herbal medicines have drawn attention of scientists to use plant products for the treatment of a variety of human ailments, including cancer¹⁸.

The GICDB database has been built in ASP and information translated into MySQL database management system, which works at the backend of the database. Web interface was designed by HTML, Java script and CSS¹⁹. Currently, information on 463 genes and 713 miRNA genes is available in the database (Figure 1).

Data access

The web interface query form of GICDB (<http://gidb.subdic-bioinformatics-nitr.in>) permits researchers to search the desired genes and proteins from the database. A list of genes is retrieved and these genes are further linked with specific details. The initial page of every gene contains the following details: (i) specific information of gene with symbol, id of gene, gene name, gene aliases, contig, chromosome location; (ii) link to other public databases, HGNC id, GeneBank id, UniGene id, HPRD id, Ensemble id; (iii) clicking on gene product provides information related to CDS, mRNA and protein sequence, UniProt id and PDB id details; (iv) link to associated disease; (v) link to homologous gene entries, and (vi) linking the reference for confirmation that the gene belongs to a particular GI cancer. Information on miRNAs can be search in a similar manner. Then, link to miRBase is provided from the miRNAs (Figure 2)²⁰.

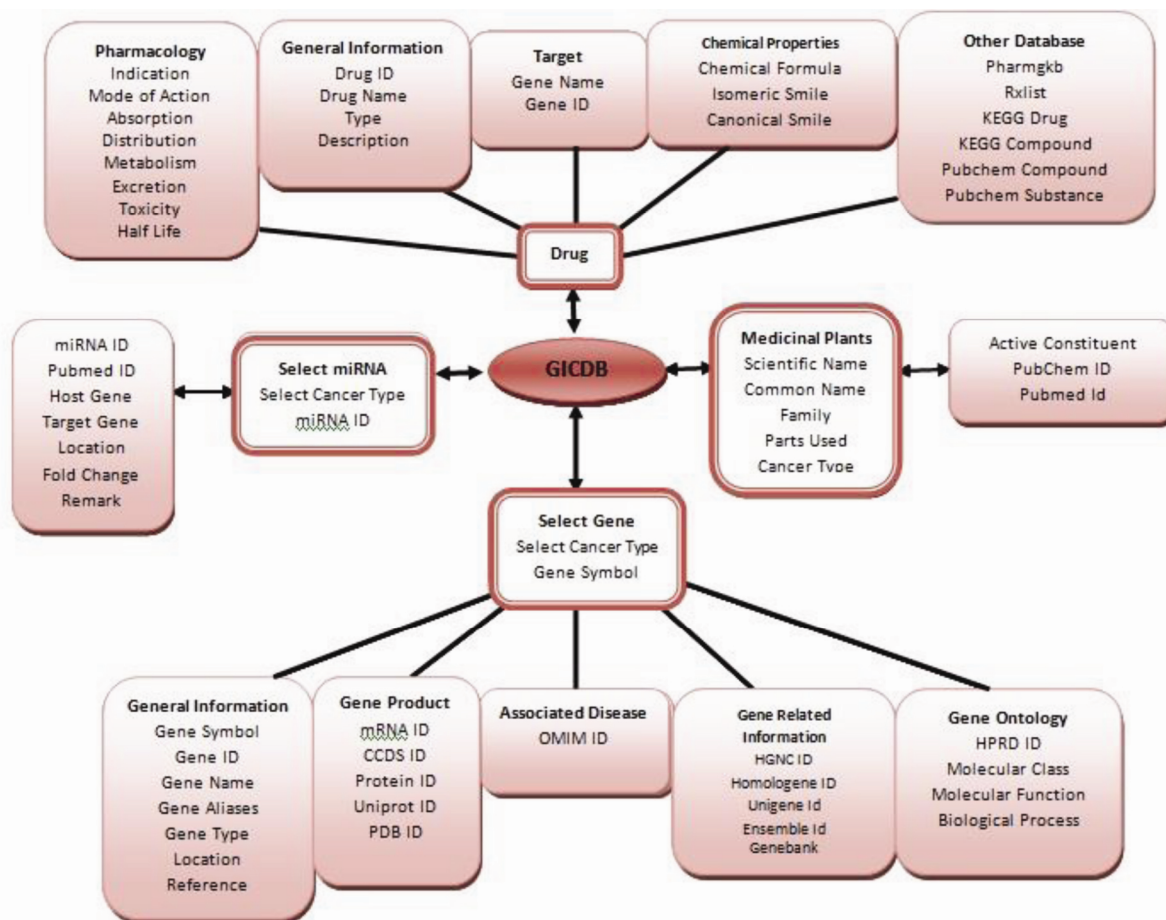


Figure 2. Database structure.

Analysis and significance of GICDB

GICDB is not a substitute for the currently available resources. It serves as complementary information to the already available resources. Data in GICDB are manually curated by biologists along with link to information from other resources. Gene-related information includes Entrez gene id, chromosome location, nucleotide sequence, protein sequence and PubMed literature id, which illustrates the involvement of genes in GI cancer. This database is useful for researchers to compare the genes systematically and provide information on distinctive genes and demonstrate the abnormal characteristics in various GI cancers.

One of the unique features of GICDB is that it includes information on drugs related to GI cancers from Drug-Bank. It contains modes of action, plasma protein-binding percentage, pharmacological properties, structural properties and chemical properties of each drug molecule. It also provides absorption, distribution, metabolism, excretion and toxicity (ADMET) properties of compounds that are important for new drug development. This drug-related information will help researchers and clinicians in the discovery of targets for new drugs, computational designing of drugs, *in silico* docking of drugs, and identifi-

cation of ADMET properties of concordants of these drugs molecules²¹.

Scientists are exploring new anticancer compounds from plants to reduce undesirable side effects associated with chemotherapy drugs. Hence GICDB includes information on medicinal plants with the structure of their active constituents. These data may be helpful for researchers to further explore the medicinal value of these plants. This may lead to the discovery of potential anticancer drugs²².

Going through above database, it has been observed that *EGFR*, *GSTP1*, *GSTT1*, *MTHFR*, *ABCB1* genes are common among various GI cancers. These genes may serve as targets for pharmacogenomics in GI cancer. Similar in-depth efforts from the community of researchers in this area can be initiated and explored. The outcome of systematic research is expected to open new ventures towards identification, diagnosis and therapeutic measures for GI cancer.

Conclusion

GICDB is an integrated-type panoptic database. Information on drugs and drug targets helps identify new targets

for future evaluation. This database may serve as a repository for GI cancer providing important information for researchers and clinicians working in this field. It would significantly save time and effort of researchers to retrieve information on genes involved in GI cancer and helps in answering pertinent biological questions involving GI cancer.

1. Leo, S. *et al.*, Carr systemic treatment of gastrointestinal cancer in elderly patients. *J. Gastrointest. Cancer*, 2013, **44**(1), 22–32.
2. Yamada, T. *et al.*, *Textbook of Gastroenterology*, Blackwell Chichester, West Sussex, 2009, 5th edn, pp. 603, 1028; ISBN 978-1-4051-6911-0. OCLC 404100761 (<https://www.worldcat.org/oclc/404100761>).
3. Bjelakovic, G. I., Nikolova, D., Simonetti, R. G. and Gluud, C., Antioxidant supplements for prevention of gastrointestinal cancers: a systematic review and meta-analysis. *Lancet*, 2004, **364**(9441), 1219–1228.
4. Rebecca, L., Siegel, MPH1* and Kimberly, D., Miller, MPH2; Ahmedin Jemal, DVM, PhD3 Cancer Statistics, *CA: Cancer J. Clin.*, 2015, **65**(1), 5–29.
5. Boland, C. R., The molecular biology of gastrointestinal cancer: implications for diagnosis and therapy. *Gastrointest. Endosc. Clinics North Am.*, 2008, **18**(3), 401–vii; doi:10.1016/j.giec.2008.03.003.
6. Langho, L. *et al.*, Liverome: a curated database of liver cancer-related gene signatures with self-contained context information. *BMC Genom.*, 2011, **12**, S3; doi:10.1186/1471-2164-12-S3-S3.
7. Chelala, C. *et al.*, Crnogorac-Jurcevic pancreatic expression database: a generic model for the organization, integration and mining of complex cancer datasets. *BMC Genom.*, 2007, **8**, 439.
8. Essack, M. I. *et al.*, DDEC: Dragon database of genes implicated in esophageal cancer. *BMC Cancer*, 2009, **9**, 219; doi:10.1186/1471-2407-9-219.
9. Maglott, D., Ostell, J., Pruitt, K. D. and Tatusova, T., Entrez Gene: gene centered information at NCBI. *Nucleic Acids Res.*, 2007, **35**, D26–D31.
10. Safran, M. *et al.*, GeneCards: towards a complete, object-oriented, human gene compendium. *Bioinformatics*, 2002, **18**, 1542–1543.
11. Amberger, J., Bocchini, C. A., Scott, A. F. and Hamosh, A., McKusick's Online mendelian inheritance in man (OMIM). *Nucleic Acids Res.*, 2009, **37**, D793–D796.
12. Pedro, M., Costa 1, 2 and Maria, C. Pedroso de Lima MicroRNAs as Molecular Targets for Cancer Therapy: On the Modulation of MicroRNA Expression. *Pharmaceuticals*, 2013, **6**, 1195–1220.
13. Wishart, D. S. *et al.*, DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.*, 2008 **36**, D901–D906.
14. Hatfield, C. L., May, S. K. and Markoff, J. S., Quality of consumer drug information provided by four Web sites. *Am. J. Health Syst. Pharm.*, 1999, **56**, 2308–2311.
15. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. and Hirakawa, M., KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, 2010, **38**, D355–D360.
16. Sayers, E. W. *et al.*, Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, 2010, **38**, D5–D16.
17. Hodge, A. E., Altman, R. B. and Klein, T. E., The PharmGKB: integration, aggregation, and annotation of pharmacogenomic data and knowledge. *Clin. Pharmacol. Ther.*, 2007, **81**, 21–24.
18. Mouli, K. C., Vijaya, T. and Rao, S. D., Phytoresources as potential therapeutic agents for cancer treatment and prevention. *J. Global Pharma Technol.*, 2009, **1**(1), 4–18.
19. Barbosa, L. C., Garrido, S. S. and Marchetto, R., BtoxDB: a comprehensive database of protein structural data on toxin-antitoxin systems. *Comput. Biol. Med.*, 2015, **58C**, 146–153.
20. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. and Enright, A. J., miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, 2005, **34**, D140–D144.
21. Maqungo, M. *et al.*, Bajic DDP: dragon database of genes associated with prostate. *Cancer Nucleic Acids Res.*, 2010, 1–6.
22. Dhanamani, M., Lakshmi Devi, S. and Kannan, S., Ethnomedicinal plants for cancer therapy – a review. *J. Drugs Med.*, 2011, **3**(1), 1–10.

ACKNOWLEDGEMENTS. We acknowledge the databases NCBI, HPRD, Ensemble, UniProt, RCSB and mirBASE for permission to link information in GICDB. This project is funded by Biotechnology Information System, Department of Biotechnology, Government of India.

Received 19 February 2015; revised accepted 22 July 2016

doi: 10.18520/cs/v112/i09/1822-1825