

TCGDB: A Compendium of Molecular Signatures of Thyroid Cancer and Disorders

Ankush Bansal* and Jayashree Ramana

Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Solan, Himachal Pradesh, India

Abstract

Thyroid cancer is the most prevalent endocrine malignancy affecting 213,000 people every year worldwide. There are many types of thyroid disorders including hyperthyroidism, hypothyroidism, Goitre, Hashimoto's thyroiditis.

The ocean of information that is available with respect to molecular markers related to Thyroid disorders and cancer is scattered in literature. Thyroid Cancer and Disorder Gene Database (TCGDB) is a compendium of genes and miRNAs experimentally known to be involved in thyroid cancer and disorders. The current release of TCGDB houses a collection of 250 genes and 120 unique miRNAs compiled by manual screening of numerous research articles. The literature references have been provided for each entry to establish its association with thyroid cancer and disorders. TCGDB provides information on miRNA in body fluids which can be used as potential biomarkers for diagnosis of thyroid cancer and disorders. We have also analysed the overlap of thyroid cancer related miRNAs with other type of cancers. The database provides various browse and search options for users to quickly extract required information. TCGDB will help in prioritizing, systematic testing and comparative analysis of candidate biomarkers for cancer diagnosis. It is freely accessible at <http://www.juit.ac.in/attachments/tcgdb/index.php>

Keywords: Thyroid cancer; Disorders; miRNA; Hypothyroidism; Goiter; Graves' disease

Introduction

Thyroid disorders are common across the population of the world and have been recognized for more than a century E.g.: Hypothyroidism, Hyperthyroidism, Hashimoto's Thyroiditis, Graves' disease, Goiter etc.

These can be accurately diagnosed by using various laboratory techniques. Estimating TSH (Thyroid stimulating hormone) along with T3 (Tri-Iodothyronine) and T4 (Tetra-Iodothyronine) are the commonly performed tests to diagnose thyroid dysfunction [1,2].

Though it is not common among the other types of prevalent malignancies there are occurrences of thyroid cancer; they rank first among the endocrine cancers [3]. The annual occurrence rate of thyroid cancer in different parts of the world is reported to be about 0.5 to 10 in every 100,000 person [4]. Literature shows that there is a gradual progress from the neoplastic cell to tumor as a result of the sequential genetic events [5]. Many studies indicate that women are more likely to suffer from thyroid dysfunctions and cancer than men [6].

Thyroid cancers can be classified into four types according to pathological analysis, namely medullary, papillary, anaplastic and follicular thyroid cancers. Medullary carcinoma originates from parafollicular cells and other three are of the follicular cell origin [7]. Follicular and papillary cancer account for about 80-90% of thyroid tumors [8], medullary cancer for 5-10% [9] and anaplastic thyroid cancer for 1-2% of all thyroid cancers [10].

Thyroid cancer is mainly diagnosed by histology, ultrasound elastography [11] and Fine needle aspiration cytology (FNAC) techniques [12]. To overcome the disadvantages of conventional histology and FNAC, several biomarkers are used and their efficiency in diagnosis, treatment and prognosis of thyroid cancer are being evaluated [13].

MicroRNAs (miRNAs) are expressed endogenously and these are 22 nucleotide RNAs which play important regulatory roles in plants and animals by targeting mRNAs [14]. Thyroid tumors show different

clinical behaviors. Extraction of miRNA profiles from tumor tissues and normal tissues exhibited aberrant profiles in tumors [15]. Expression analysis of miRNA could help us in differentiating between benign and malignant thyroid neoplasms that are uncertain by conventional techniques [16,17].

Analysis of expression profiles of the genes involved in carcinogenesis of thyroid cancer provides us a better understanding of the underlying mechanisms of tumor invasion and provides valuable information in the discovery of possible novel molecular targets for the treatment of thyroid cancer and as diagnostic tumor markers [18].

In the recent years, a large number of databases have emerged with a central focus on a specific cancer as exemplified by Renal Cancer Gene Database [19], Cervical Cancer Gene Database [20], Human Lung Cancer Database [21], Breast Cancer Gene Database [22], Oral Cancer Gene Database [23], but there is no database for genes involved in thyroid cancer and disorders. Therefore, we have collected thyroid cancer and disorders related genes to construct an integrated database- Thyroid Cancer and disorder Gene DataBase (TCGDB) that catalogs the genes, miRNA and proteins involved thyroid diseases as evidenced from the literature. We have created a user friendly interface and also have provided open access to our database. In addition, the database provides search facility BLAST, for querying the database for sequence similarity search. Overall, TCGDB is a specialized, value-added database which enables the exploration of relevant information

***Corresponding author:** Ankush Bansal, Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Solan-173234, Himachal Pradesh, India, Tel: +91-7807146646; E-mail: bansal.ankush13@gmail.com

Received June 05, 2015; Accepted July 09, 2015; Published July 13, 2015

Citation: Bansal A, Ramana J (2015) TCGDB: A Compendium of Molecular Signatures of Thyroid Cancer and Disorders. J Cancer Sci Ther 7: 198-201. doi:10.4172/1948-5956.1000350

Copyright: © 2015 Bansal A, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

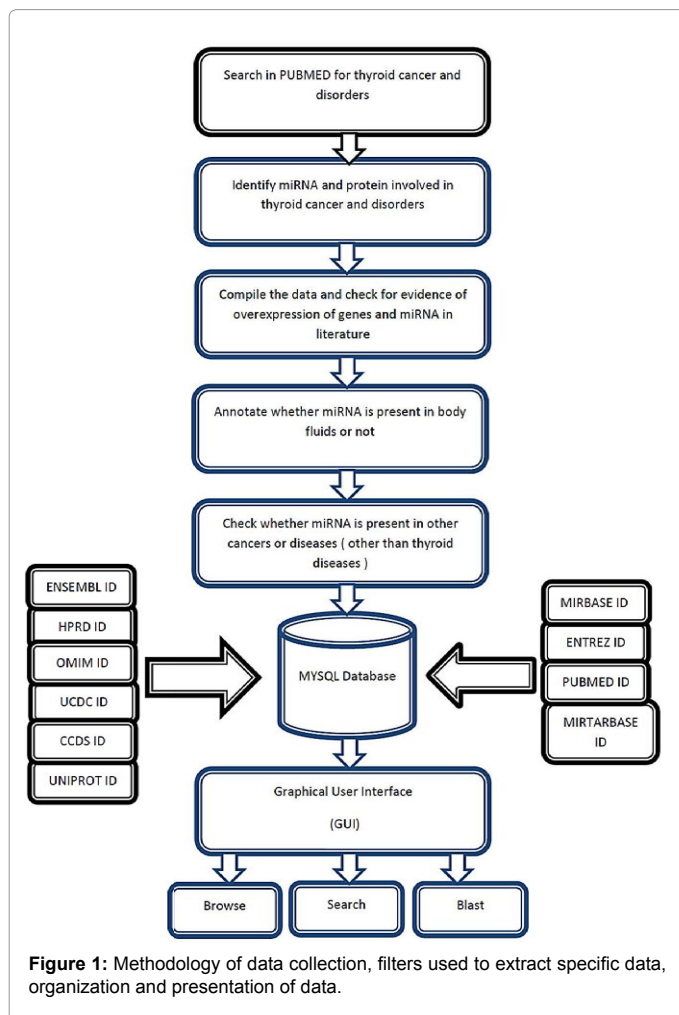
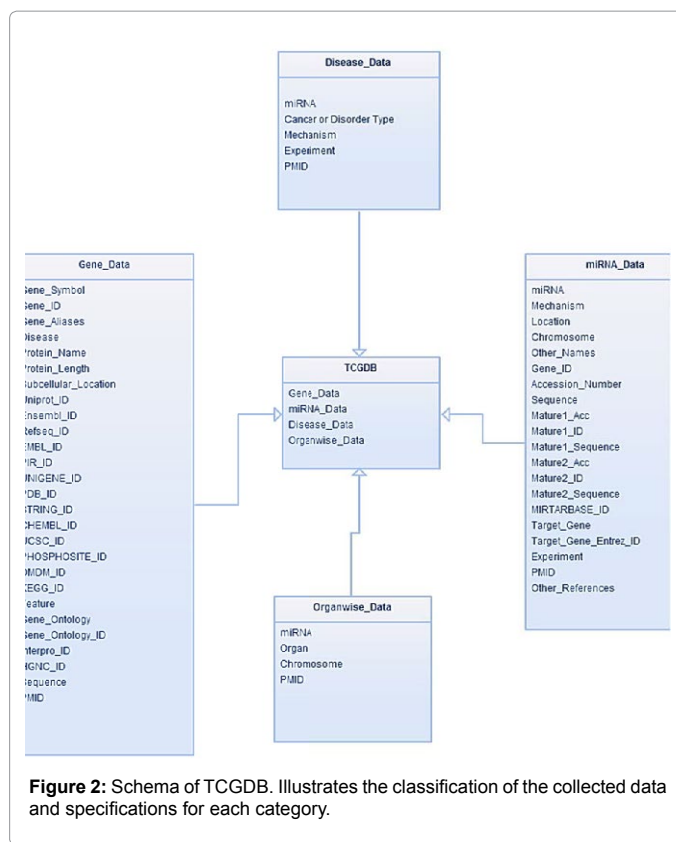
for all experimentally determined human thyroid-cancer-related genes making it a unique resource in the area of thyroid cancer biology.

Methods

We have comprehensively collected information from PubMed about the genes that are playing a pivotal role in causing and sustaining thyroid cancer and disorders. Genes and miRNAs involved in the diseases progression were extracted after a thorough study of the full text of each research article that we obtained (Figure 1).

Further information which includes (1) Gene Symbol (2) Gene ID (3) Gene Aliases (4) Disease description (5) Protein Name (6) Protein Length (7) Subcellular Location (8) UNIPROT ID (9) Ensembl ID (10) RefSeq ID (11) EMBL ID (12) PIR ID (13) Unigene ID (14) PDB ID (15) String ID (16) ChEMBL ID (17) UCSC ID (18) Phosphosite ID (19) DMDM ID (20) KEGG ID (21) Gene Ontology (22) Gene Ontology ID (22) Interpro ID (23) HGNC ID (24) Sequence (25) Features and (26) Reference was obtained from various databases and integrated in the form of tables to give a detailed description and insight into each entry that is present in TCGDB (Figure 2).

MicroRNAs with experimentally validated information from the literature and documented information regarding gene that codes for miRNA, mechanism, chromosome number, chromosomal location, gene ID, accession number, miRNA sequence, miRTarBase ID,



target gene, experiments and its PubMed ID for reference were also incorporated.

The entire information was imported in MySQL (A Relational Database Management System). A web interface was built in PHP which relates our User Interface and our integrated database.

Currently, the data we have accumulated includes a repository of 250 genes and 120 unique miRNAs which are involved in the various stages of thyroid cancer.

Hence, the data collected would be a unified information portal for the thyroid disorders research community.

Results

The collated data present in TCGDB has been classified in a way which helps the user to easily and efficiently browse through the database. Input query can be in the form of Gene symbol, Gene ID, Uniprot ID and PDB ID which converge and display information about the particular query in detail. Each entry in the database has gene-centered information. PMID or Tracking number that is provided in the main page of an entry would relate each entry to literature which establishes the relationship between thyroid diseases and the particular gene or miRNA.

Also TCGDB provides three other ways to view and retrieve all thyroid-cancer-related genes or miRNA. First, a user can query the database for human chromosome number to display the genes present on each chromosome and then browse them individually.

Secondly, TCGDB has a browsing section that allows the user to access the entire gene data set or search gene by cancer type and

respective categories or search the genes by their names or in an ordered alphabetical manner. Finally, the entries are categorized into biological process whose alteration leads to thyroid cancer disease as evident from literature.

Thus, TCGDB provides a gateway through which the scientific community can easily access the latest information on the genes involved in thyroid diseases. Further, a customized BLAST tool has been incorporated which searches a user-defined query against the sequences available in the database. It will come handy in characterization of orphan sequences or identifying homologous sequences from TCGDB. Also, an online submission portal has been provided to include new gene entries that are associated with thyroid disorders. Once the new gene information is received with specified fields, the database would be updated after validation.

In addition to the genes involved in the disease progression, we have also included miRNAs, which are present in other cancers or disorders. Therefore, TCGDB will supplement the existing databases in serving the scientific community. Also, TCGDB allows the researchers to make a thorough comparison of miRNAs that are common in most of the cancers as well as detect the ones that are unique and demonstrate aberrant behavior in thyroid cancer.

The current version (v. 1.0) of TCGDB contains 250 unique genes and 120 miRNAs that are validated by the literature. The data is presented in an organized way. Apart from search facility; various browsing options assist efficient, fast and user-friendly retrieval of information (Figure 3). We propose to update TCGDB on a regular basis that includes new data from the user inputs, literature, as well

as other data exploration tools that help in improving our knowledge about thyroid diseases and contribute towards the development of novel therapeutic approaches. Database is freely accessible at <http://www.juit.ac.in/attachments/tcgdb/index.php>

Discussion

Our efforts at creating TCGDB represent the first attempt in dealing the omics data at genomic, proteomic, miRNA omics level for thyroid cancer and disorders in an extensive and organized way. There are a number of cancer specific databases developed and published, still at present there is no database that catalogues genes involved in thyroid cancer and disorder.

Thyroid cancer and disorder gene database comprises gene information for different types of thyroid cancers (papillary thyroid cancer, follicular thyroid cancer, medullary thyroid cancer, anaplastic thyroid cancer), miRNA data involved in other cancers and diseases, miRNA which is present in body fluid. This database was especially developed for selecting biomarkers for effective diagnosis of thyroid cancer and disorders. Flexible design lets the user to know miRNA information about other cancers as well as for prioritizing and systematic testing of candidate biomarkers.

As the whole system depends on the data layer, the development of it was the crucial point. It is obvious that the design of databases for gene as well as miRNA data has to be very flexible, to be able to adapt or even upgrade the whole system to new scientific insights without major changes. The usage of relational databases together with compatible technologies (MYSQL, PHP etc.) on the server side enables a fulfillment of all these demands. As this database is designed for scientists and researchers who are working in different locations, it is important to provide an easy way to give the users access to the stored information. Thus a web application is the best solution to provide an easy data accession. The flexible user management and the secure web connection allow easy accession and insights to the data. For users who need an evaluation and analysis of their data, appropriate query methods are given, which present data in desired formats. The best option for a quick data exploring are the basic queries. To get deeper insights into the data and create queries on parameters to which one is interested in, the custom queries represent the best solution. The custom queries further provide save options for each user in order to store specific query parameters in the database. A sequence similarity search tool that will enable the exploration of relevant information for all experimentally determined genes and proteins present in the database.

In future work, clinical trial datasets and functional data will be integrated with microarray data in order to explore new relations between expression patterns. Hence, this database should become a valuable mean in thyroid cancer and disorder studies.

Conclusion

TCGDB has been developed as an integrated information resource to assist the research efforts of scientists and clinicians working on various thyroid disorders and cancers. It serves as a comprehensive repository of information related to thyroid cancer and disorders as well as facilitates thorough exposition of each gene by providing hyperlinks to relevant PubMed records. In future, TCGDB would be updated on a regular basis. It is anticipated that TCGDB would serve as a valuable resource to the scientific community.

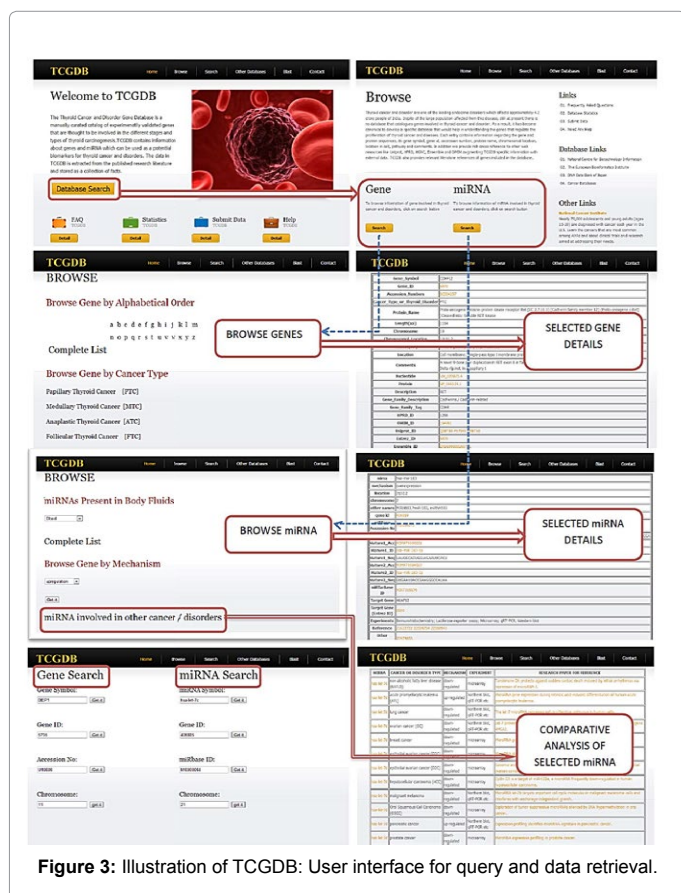


Figure 3: Illustration of TCGDB: User interface for query and data retrieval.

References

1. Liewendahl K (1983) Assessment of thyroid function with hormone assays. *Ann Chir Gynaecol* 72: 84-89.
2. Lopresti JS (1996) Laboratory tests for thyroid disorders. *Otolaryngol Clin North Am* 29: 557-575.
3. Correa P, Chen VW (1995) Endocrine gland cancer. *Cancer* 75: 338-352.
4. Siegel R, Naishadham D, Jemal A (2013) Cancer statistics, 2013. *CA Cancer J Clin* 63: 11-30.
5. Fagin JA (2002) Minireview: branded from the start-distinct oncogenic initiating events may determine tumor fate in the thyroid. *Mol Endocrinol* 16: 903-911.
6. Surks MI, Ortiz E, Daniels GH, Sawin CT, Col NF, et al. (2004) Subclinical thyroid disease: scientific review and guidelines for diagnosis and management. *Jama* 291: 228-238.
7. Samimi H, Zaki DM, Ghadami M, Fazeli AS, Khashayar P, et al. (2013) Essential genes in thyroid cancers: focus on fascin. *J Diabetes Metab Disord* 12: 32.
8. Gimm O (2001) Thyroid cancer. *Cancer letters* 163: 143-156.
9. Orlandi F, Caraci P, Berruti A, Puligheddu B, Pivano G, et al. (1994) Chemotherapy with dacarbazine and 5-fluorouracil in advanced medullary thyroid cancer. *Ann Oncol* 5: 763-765.
10. Reynolds RM, Weir J, Stockton DL, Brewster DH, Sandeep TC, et al. (2005) Changing trends in incidence and mortality of thyroid cancer in Scotland. *Clin Endocrinol (Oxf)* 62: 156-162.
11. Carneiro-Pla D (2013) Ultrasound elastography in the evaluation of thyroid nodules for thyroid cancer. *Curr Opin Oncol* 25: 1-5.
12. Pacini F, Fontanelli M, Fugazzola L, Elisei R, Romei C, et al. (1994) Routine measurement of serum calcitonin in nodular thyroid diseases allows the preoperative diagnosis of unsuspected sporadic medullary thyroid carcinoma. *J Clin Endocrinol Metab* 78: 826-829.
13. Sethi K, Sarkar S, Das S, Mohanty B, Mandal M (2010) Biomarkers for the diagnosis of thyroid cancer. *J Exp Ther Oncol* 8: 341-352.
14. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281-297.
15. Visone R, Pallante P, Vecchione A, Cirombella R, Ferracin M, et al. (2007) Specific microRNAs are downregulated in human thyroid anaplastic carcinomas. *Oncogene* 26: 7590-7595.
16. Chen X, Ba Y, Ma L, Cai X, Yin Y, et al. (2008) Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell research* 18: 997-1006.
17. Vriens MR, Weng J, Suh I, Huynh N, Guerrero MA, et al. (2012) MicroRNA expression profiling is a potential diagnostic tool for thyroid cancer. *Cancer* 118: 3426-3432.
18. Nikolova DN, Zembutsu H, Sechanov T, Vidinov K, Kee LS, et al. (2008) Genome-wide gene expression profiles of thyroid carcinoma: Identification of molecular targets for treatment of thyroid carcinoma. *Oncol Rep* 20: 105-121.
19. Ramana J (2012) RCDB: Renal Cancer Gene Database. *BMC research notes* 5: 246.
20. Agarwal SM, Raghav D, Singh H, Raghava GP (2011) CCDB: a curated database of genes involved in cervix cancer. *Nucleic acids res* 39: D975-979.
21. Wang L, Xiong Y, Sun Y, Fang Z, Li L, et al. (2010) HLungDB: an integrated database of human lung cancer research. *Nucleic acids res* 38: D665-669.
22. Baasiri RA, Glasser SR, Steffen DL, Wheeler DA (1999) The breast cancer gene database: a collaborative information resource. *Oncogene* 18: 7958-7965.
23. Levine AE, Steffen DL (2001) OrCGDB: a database of genes involved in oral cancer. *Nucleic acids res* 29: 300-302.