|  |
| --- |
| **Gene to Cancer Correlation Assessment[[1]](#footnote-2)**  Dao Lin 1, Henry Nguyen 1, Patrick Chapagain 1  1 School of Biomedical Engineering, Drexel University, USA  Course   : BMES 550 Advanced Biocomputational Languages  Instructor: Ahmet Sacan  Date       : 2022-12-07 |

**abstract**

There is genetic information about genes associated with cancer but have not yet been utilized for any diagnostics. End stage diagnosis of cancer does not have many treatment options. There is a big need for an early diagnosis/prediction if a person has a risk of cancer. Our model takes patient’s information and desired genes input to assess if they are at risk for certain types of cancer. The application will generate a tree diagram associating the gene types with the cancer types, and store that into the database for future access.

# 1. introduction

There are many human disease databases with genes and variants on the internet, however, they lack proper filtering and categorization for easy access. It would be difficult for physicians and researchers to utilize the databases to find useful information. Time would be wasted in attempting to correlate a gene to different cancers, or the other way around. The focus of the diagnostic is to provide an example of how one might filter different cancer types that are correlated to an inputted cancer gene sequence. Additionally, it allows the user to store voluntary patient information for future contact. Physicians and biomedical researchers would be able to store, and share information.

**Computer science:**

This application uses database interactions and GUI implementations to compare the user input to the database data. All information of the patients is taken from the GUI and stored into a table in the database. The field gene is compared to the database gene to see if the categories match. If they match, the GUI will have a figure showing gene and cancer association for that specific patient and store it in the patient information table. If there is no match, the application will say there is no cancer risk associated with the patient’s gene.

**Goals:**

With the improving technology in genetic mapping, diagnostics can be improved for the human body. We are creating a diagnostics tool that can take patient’s information including the gene data and associate the genetic information with cancer type and risks to alert the patients for early diagnosis to prevent cancer.

The target end users are doctors and medical professionals who can diagnose and give medical advice to patients. When a patient’s genetic information is analyzed, the doctor can use that data to see if there are any cancer associated genes and ask the patients to get a screening for that specific type of cancer. An early diagnosis can help prevent the growth of the cancer with proper treatment and the patient can live a healthy life.

With the increasing medical technology on genomics, a person’s genetic information can help identify potential cancer victims and have an early intervention to prevent it from happening. With the genes associated with cancers analyzed, we can correlate the gene association with a certain type of cancer. This will help identify cancer causing genes and have an early intervention for those at risk. This will revolutionize medical practice as it can save lives, we are currently losing to cancer victims.

**Related works:**

The Cancer Genomes Atlas[5] is a collection of many other databases. It is created by the National Human Genome Research Institution. The website allows users to select the type of cancer, genome, and other important information. It contains a higher-level definition of correlation for in depth research.

**2. DATASET**

The database used for this project was extracted from disgenet.org. They have collected and curated the data from multiple credible sources which includes CTD, CLINGEN, GENOMICS ENGLAND, CGI, LHGDN, BEFREE, UNIPROT, CLINVAR, GWAS catalog, and GWAS db[6]. The database was created by disgenet using DisGeNET Metrics and Data attributes. The data collected was from humans from different databases aggregated into one by DisGeNET[6].

# 3. methods and IMplementation

Database disgenet\_2020.db was downloaded from disgenet.org and loaded into SQLite Studio. SQL query was made to create a table in the database to store patient’s information with gene type and cancer association. Then, to have the medical professional input the data, we created a MATLAB Graphical user interface (GUI) where patient’s information can be added. The input information is stored in the database table named patientinfo. Then, in MATLAB we created three functions, submitpatientinfo.m, diagram.m, and no NoData.m, for submitting patient information and creating the correlation diagram.

GUI was created using MATLAB App Designer, and it was utilized to store patient data into the DisGeNET database. SQL Query statement was constructed to pull out unique cancer related diseases that were associated with a specific gene that was inputted through the GUI. Graphical representation of the gene and cancer association was done through use of a tree diagram, which was coded using MATLAB.

The DisGeNET database was imported into SQLiteStudio and used the query format to filter information that is only correlated to cancer types. The gene name, amount of cancer types correlated to gene, and array of cancer types were stored into another table (compare\_table) for quicker access. A compare\_table2 was also created to contain the cancer name, number of genes associated with the cancer, and an array of gene names.

A GUI was created for data input of the volunteer patient information, and a desired mutated gene to establish correlation with cancer types. A tree diagram will appear if correlation is found.  The origin node will be the desired gene, and branches are the cancer types. Each cancer type will further branch off connecting to other correlating genes. To establish a connection between DisGeNET database and GUI, the two MATLAB functions, submitpatientinfo.m and diagram.m, are called. If the imputed gene has no correlation, NoData.m is called.

Submitpatientinfo.m function creates a new table in the DisGeNET database to store input patient information. Every time new patient info is submitted it will be stored into patientinfo table. Diagram.m function links the input desired gene to compare\_table, which returns the cancers associated with the gene. Each cancer is then linked to compare\_table2, which connect cancers to genes. It takes the correlated cancer types of arrays and creates a tree diagram to visually depict the connection. GUI pulls the diagram from diagram.m and displays it on the interface.

Diagram

Description automatically generated

**Referenced Code:**

Utilized bmes.tempir and bmes\_db functions from bmes.Ahmet folder in order to store the SQLite database into a temporary file on the computer and connect to it.

# 4. Experiments and Results

When a patient's information is inputted into the GUI, it will compare the patient’s gene to the gene in the database that is associated with cancer. If there is a match, it will associate the patient information with the type of cancer they are at risk of. If there are no matches, it will tell the user that there is no gene associated cancer risk.

Graphical user interface, application

Description automatically generated

**Figure 1: Application with user input**

A picture containing chart

Description automatically generated

**Figure 2: Application showing tree of cancer association and gene type**

The figures above show a case for a patient where the gene is inputted into the GUI fields. When hit submit, the gene is compared with database genes and if there is a match, it will create a tree diagram to show cancer associated with those genes.

The main parameter for our application is the gene type input. The user must have the patient’s gene to compare to our database model’s gene.

Diagram

Description automatically generated

**Figure 3: Diseases associated with T2D gene [6]**

Graphical user interface, text, application, email

Description automatically generated

**Figure 4: Application when gene T2D is inputted**

Figure 3 shows all the diseases that are linked to the T2D gene and is based on the DisGeNET database that we also used in our GUI. We see that of the many diseases, there are none that are cancer related. Figure 4 shows our GUI application when the same gene is inputted. We see that the gene is not associated with cancer, which is consistent with the findings from figure 3.

# 5. DISCUSSION

With the genetic information from a patient, a doctor can give a risk assessment diagnosis of a patient with gene associated cancer risk.

   To determine the accuracy of the diagnostic, a journal publication relevance test was conducted. The test would be to verify if there is a correlation between an input gene to the cancer types found by the diagnostic. Gene ABCG2, ATP-binding cassette sub-family G member 2, was chosen, and the correlating cancers were hormone refractory prostate cancer, androgen independent prostate cancer, and metastatic renal cell cancer. ABCG2 is a semi-transport protein that plays a major role in multidrug resistance [5]. All three cancer types have publications correlated to ABCG2. The gene transports different types of hormones to the prostate that increases cancer risk [1]. Furthermore, ABCG2 mediates androgen efflux in the prostate region [2]. It plays a role in prostate stem cell maintenance, and growth. Therefore, the prostate region has some dependence on androgen for control. As a result, when it becomes independent of androgen because of ABCG2 mutation, the growth is not controlled, leading to prostate cancer [2][3]. The risk is low because the tumor growth is very slow. Additionally, ABCG2 is also associated with sunitinib-induced toxicity that can cause metastatic renal cell carcinoma [5].

From the test, we can say that the cancer correlation diagnostic tool is accurate in listing the cancer type from a determined gene in the DisGeNET database. The gene has many publications related to the type of cancer it is associated with.

There are few limitations in the diagnostic tool. It may not work with other databases because the SQLite query is made specifically for DisGeNET. Furthermore, there are unidentified cancer types in the database, such as ‘cancer pain’. The diagnostic tool is unable to remove those descriptions from the table. Unfortunately, if there are too many correlations, the chart will not display the names of cancer types and gene names. The database is restricted to cancer gene discoveries up to 2020.

Better chart programming tools can be used to display larger correlations. A future improvement is to add a filtration query to remove unidentified cancer disease types to make the database cleaner.

# 6. References

[1]Gardner, E. R., Ahlers, C. M., Shukla, S., Sissung, T. M., Ockers, S. B., Price, D. K., Hamada, A., Robey, R. W., Steinberg, S. M., Ambudkar, S. V., Dahut, W. L., & Figg, W. D. (2008). Association of the ABCG2 C421A polymorphism with prostate cancer risk and survival. *BJU international*, *102*(11), 1694–1699. <https://doi.org/10.1111/j.1464-410X.2008.07913.x>

[2]Sabnis NG, Miller A, Titus MA, Huss WJ. The Efflux Transporter ABCG2 Maintains Prostate Stem Cells. *Mol Cancer Res*. 2017;15(2):128-140. doi:10.1158/1541-7786.MCR-16-0270-T

[3]Feldman, B., Feldman, D. The development of androgen-independent prostate cancer. *Nat Rev Cancer* 1, 34–45 (2001). https://doi.org/10.1038/35094009

[4]Wang, H., Luo, F., Zhu, Z. *et al.* ABCG2 is a potential prognostic marker of overall survival in patients with clear cell renal cell carcinoma. *BMC Cancer* 17, 222 (2017). https://doi.org/10.1186/s12885-017-3224-6

[5]Pavlopoulou, A., Spandidos, D. A., & Michalopoulos, I. (2015). Human cancer databases (review). *Oncology reports*, *33*(1), 3–18. https://doi.org/10.3892/or.2014.3579

[6]Piñero, J., et al., *The DisGeNET knowledge platform for disease genomics: 2019 update.* Nucleic acids research, 2020. **48**(D1): p. D845-D855.

1. [↑](#footnote-ref-2)