**Analyzing Science Articles: Identifying Genes Associated with Brain Tumors**

Julia Christensen

CSC 450 Senior Research, Due Date: 13th December 2023

**Abstract**

Brain tumors are life-threatening diseases and are one of the leading causes of death in children and adolescents with cancer [1]. Studying the genetic makeup of the tumors helps researchers understand tumor development and look at the similarities and differences when compared to normal tissue. Researchers write their findings in articles published in places like PubMed to share what they learned with others. In this paper, gene names were extracted from some of these articles and abstracts to analyze what genes have been found regarding brain tumor studies. Data was collected through PubTator, where the gene names were extracted and analyzed using Python code. The overall goal of the research was to provide insight into genetic factors linked to brain tumor development and prognosis to help other researchers focus more on certain types of genes that have the potential for better treatment plans. The experiment's outcome concluded that the BRAF and MGMT genes were the most referenced overall. Knowing that those genes are associated with brain tumors and are the most referenced, future researchers can now investigate these specific genes in brain tumors to continue this vital research.

**Introduction**

Brain tumors cause serious, life-threatening diseases, for they are abnormal cell growths [2]. There are several factors that can lead to abnormal cell growth, and researchers are looking at genetic and molecular changes in a person's cell to understand why there was a change in cells that caused them to start growing abnormally, which in turn caused the tumors [3]. *Genes* are the building blocks of life; they are segments of DNA that determine how an organism looks, behaves, and survives in environments or places they are in [4]. Diagnosing brain tumors can be done through various tests, one being an MRI scan with a gadolinium enhancement that examines the brain, looking for any indications of abnormalities [2]. Brain tumors can either be *benign* or *malignant*. Benign tumors stay put and do not travel to other parts of the body, unlike malignant cancerous tumors that can spread (*metastasize*) from one part of the body to another [5]. Since benign tumors do not disperse, they are just classified by their look, size, and area in which it is located; however, malignant tumors can be further classified into specific types or *histologic (main)* categories, which will then sometimes have subcategories which will be placed into a type of *grade* [5]. Grade I or II are *low-grade* tumors that grow slowly and are less likely to cause nearby tissues to become tumors, and *higher-grade* tumors, grade III or IV, will grow quickly and cause other tissues to become tumors [5]. The likely hood on developing a primary cancerous brain tumor is less than 1%, however it is estimated that around 5,000 children under 20 will be diagnosed with it in the year 2023 in the United States [6]. The leading cause of death in children and adolescence is cancer, and brain tumors being one of the top 3 cancers [1]. There are so many different types of brain tumors; however, they are hard to diagnose, for they have symptoms like headaches and nausea or vomiting, which other diseases can also cause. These symptoms are why brain tumors in children are often delayed in diagnosis or misdiagnosed, for it could be other ailments [7].

Studying all tumors' *biological, genetic,* or *mechanical* factors is essential to understand how they behave or grow, spread, and react to other tissues in the human body. Understanding the *molecular mechanisms* of brain cancerous tumors to help improve diagnosis and treatment to help decrease the disease's mortality [8]. Studying the mechanics of the tumors will help researchers learn more about why tumors in the brain happen and learn other things, like how the genetics of the tumors can be different due to *mutations,* which will affect the locations in which they develop in a person [5]. *Mutations* (i.e., changes in DNA sequence) can lead to malformation of proteins that cannot perform their duties as they are meant to, leading to genetic disorders or even diseases like cancer [4]. Studying tumor genes and associated genes will help researchers understand what makes a tumor different from normal tissue. If they find what causes the tumor, they can create more effective treatments that will target the cause of the tumor and save lives.

Researchers share their findings through research articles, which are published and can be accessed through various web platforms such as PubMed (<https://pubmed.ncbi.nlm.nih.gov/>). PubMed is a literature-based resource where scientific papers and other resources can be found and is one of many resource systems in the *National Library of Medicine: National Center for Biotechnology Information* (NCBI) government website. PubMed articles and their abstracts can be accessed through the PubTator Central platform, or PubTator for short, a free online resource that provides bio information using PubMed articles. PubTator is a system that highlights six biomedical concepts mentioned in PubMed article abstracts, which are summaries of research articles [9]. A user can obtain PMIDs (PubMed IDs) of articles by saving an article in a collection by clicking the heart icon underneath the article title and author names. There are well over 30 million pieces of biomedical literature that PubTator can access, and there are various categories under biomedical, with brain tumors being one [9]. The point of this tool is to help researchers in many ways, like making it easier to identify genes, diseases, chemicals, and other biomedical concepts to help researchers analyze important information in various numbers or research articles and find related work on biological topics quickly. Suppose researchers want to download data on the biochemicals. In that case, they can use the PubTator API platform, where following the instructions will obtain data and is the main reason PMIDs are important. PubTator helps provide data that researchers can analyze.

The project aimed to identify genes associated with brain tumors, providing insight into genetic factors linked to brain tumor development and prognosis, through mentions in biological articles and their abstracts. PubTator was the source from which data was obtained for the project, and it contained research articles on studies done regarding brain tumors (**Note:** the experiment used Python in Jupyter Notebook environment to analyze the data). The project's outcome will help other researchers learn what genes are associated with brain tumors, which could lead to future research on the most frequently referenced gene in brain tumors. In addition, it will help educate researchers new to studying brain tumors on what genetic information has been discovered and what has been most focused on based on the number of times the gene name is referenced. Understanding genes associated with brain tumors will help other research topics focus more on certain types of genes that have the potential for better treatment plans.

Other studies are similar to this project, which looked at gene data from the literature. One group of researchers created an online database focused on brain tumor genes. They were looking at brain tumors and wanted to create a database containing genes related to them all in one place. The information gathered was from PubMed abstracts, and they were able to find 1421 unique genes that are associated with brain tumors [8]. Not only does the database show the genes, their properties, and other information, but it also provides links to other works of literature that reference that type of gene.The whole point of the research is to show how the researchers created the database that shows genes associated with a type of disease like brain tumors and why their study was important. It was about making it easier for other researchers to look at biological information on genes and their associated diseases in one spot for easy access.

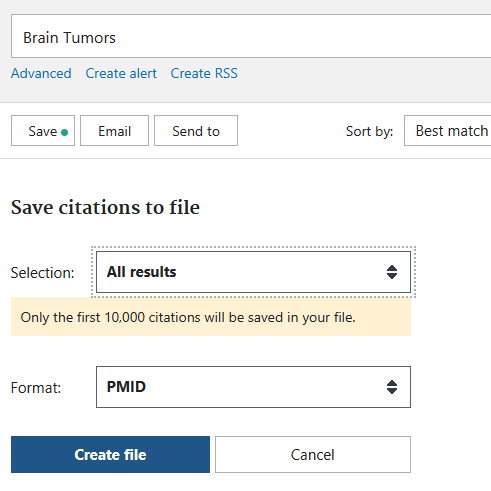
In this paper the goal was to look through biomedical articles and abstracts on brain tumors, to see what genes were mentioned in them, then find the overall mentioned gene associated with brain tumors. For the experiment a combination of abstracts and articles, about 1,100 total, were used to extract gene names, which were then analyzed. At the end of the experiment, it concluded that not one but two genes were the most mentioned, BRAF and MGMT being referenced about 33 times each when looking at both abstracts and full-text articles together. The genes IDH and EGFR following in close second with 31 times being mentioned.

**Methodology (i.e., Materials and Methods)**

The experiment's outcome was to obtain gene names from articles and abstracts on brain tumors in the PubTator Central platform to analyze what genes were mentioned in them. The experiment is broken down into these two steps, “*Obtain the Data*” and “*Analyze the Data,*” before recording the results. Be aware that a Windows operating system computer was used for the experiment. When referencing Python code, it is summarized in the article, but to see the code the researcher did and their data, head over to their web page: <https://fellowship1954.github.io/Analyzing-Science-Articles-Identifying-Genes-Associated-with-Brain-Tumors/>

*Obtaining the Data*

On the PubMed website (link: <https://pubmed.ncbi.nlm.nih.gov/> ), the keyword “brain tumors” was typed into the search box and entered, and around 226,378 results popped up. From there, the save button underneath the search bar was clicked, and the following options were selected: see **Figure 1**. The file was saved in a folder as a text file (i.e., txt).



**Figure 1:** Shows the choices picked to obtain the PMIDs from PubMed and creates a file containing the list of PMIDs.

Next, the PubTator Central platform was accessed (link: <https://www.ncbi.nlm.nih.gov/research/pubtator/>), and on the left-hand side of the webpage under “*Anonymous*” the button labeled “*Collections*” was selected, then following the instructions, a new collection was created, see **Figure 2** for visualization.

|  |  |
| --- | --- |
|  |  |

**Figure 2:** How to create a collection in PubTator. The left column is step 1, and the right is step 2.

The following steps were how the PMIDs of articles containing gene information were obtained.

1. Created a separate tab in a text editor and saved it in the same folder as the file created from PubMed.
2. Highlighted, cut, and pasted PMIDs between 10 to 21 into the search bar of PubTator, and the search button.
3. Reviewed each article:
   1. If the label “Gene” in purple was not present in the column to the left, then the next article was accessed.
   2. Else if “Gene” was in the column on the left, the heart icon labeled “*Add to Collections*” was selected before preceding to the next article.
4. Repeated Step 3 until all the articles were accessed. After that the search bottom at the top was clicked, going back to the beginning before going through the articles.
5. Checked to make sure the number of PMIDs collected showed in the collections. Then cleared the search box.
6. Repeated steps 2-5 until around 50 PMIDs were collected.
7. When around 50 PMIDs were collected, the collections button was selected, and the list of PMIDs appeared in a text box.
8. Highlighted all the PMIDs in the text box and cut them. When collected the “clear” was clicked to reset the collections.
9. Pasted and saved the PMIDs into the text file tab created in Step 1.
10. Repeated Steps 2-9 until collected 1,100 PMID samples for the experiment.
11. Reviewed the text file with the 1,100 PMID samples and went through ensuring they were listed in one column and saved the document (**Note:** Confirmed the number of samples in the text editor like Notepad, which is at the bottom line).

Then went to Jupyter Notebook environment to program in Python. Here is the list of modules used: Requests, Pandas, Json, Openpyxl, Matplotlib.pyplot, Random, WordCloud. A new worksheet was created, where the first coding concept was to check to make sure the text file and data could be accessed using Python. The code opened and read the file outputted how many samples there were, which matched to the number in the text editor; 1,100 samples. Further code was created to check for duplicates, and there were six duplicates, which were deleted. Then Steps 2-9 were repeated until 1,100 samples were again reached, which lead back to the Python code, to ensure the number of samples matched, and no duplicates found. Afterward, went to the PubTator API website (link: <https://www.ncbi.nlm.nih.gov/research/pubtator/api.html>) and followed the directions. After realizing that the code to obtain the information using the PMIDs differs between full-text and abstracts, the researcher had to take their list of 1,100 samples and perform similar steps to what they did before to separate abstracts and full-text. The steps that the researcher did are listed below:

New tabs in the text editor was created and saved, dedicated for abstracts and full-text data respectively.

1. In PubTator, he researcher did a similar Step 2 by obtaining the PMIDs from the samples file, but instead of cutting, it was copied.
2. The PMIDs were pasted in the search bar, entered and a filter for full-text was selected.
3. All the articles were placed in the collects by hitting the heart icon.
4. When the PMIDs were collected, they were cut, pasted, and saved into the text file for full-text and the code.
   1. The code compared the PMIDs from the original sample and full-text to display all the ones that were not matched, which were the abstracts.
   2. The abstract PMIDs were highlighted, copied, pasted, and saved in the text editor tab for abstracts.
5. The PMIDs were deleted, the code reset to empty lists.
6. Steps b-f were repeated until all the samples were checked and separated into abstract or full-text PMIDs.

The Python code was used to check the total samples in case there were any deletions or extra; there were none. Duplicates in the total samples were checked for, and there were none, so the researcher went on and did similar things on the abstracts and full-text. The lengths were checked, looked for duplicates in each file, and then more code was added to compare the abstract and full-text files to ensure no duplicates. There were some, but after checking on what they were in PubTator, they were deleted from the file they were not supposed to be in and then rechecked for more duplicates. After, another code was created to make sure that the samples in both the abstract and full-text added together equal the total number of samples, which it did. The PMIDs were then used to obtain the gene data, by using the template that PubTator API provides (see **Figure 3**) [9]. The following steps lead to using the PMIDs obtained in the PubTator API code to gather the gene data.



**Figure 3**: The URL template from PubTator API to obtain the gene names.

The URL template was used in the code; however, it was modified by the following:

* biocjson replaced [Format] in the URL, meaning that the data will be in JSON format.
* pmids replaced [Type] in the URL
* The PMIDs that were collected replaced [Identifiers] in the URL. The code did batches of 100 PMIDs at a time, each ID separated by a coma.
* gene replaced [Bioconcepts] in the URL

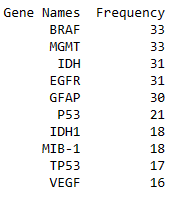
The researcher then placed the PMIDs from the text editor in an Excel worksheet in the respective tabs. The worksheet contained the PMIDs for the abstract and the full-text to keep track of what was looked at and what data was obtained. Then, in the Jupyter Notebook, more code was added, so it went into the specified tab, got a copy of 100 PMIDs at a time, placed it into the URL, then gathered the ID number of the article/abstract, identification number, biochemical type “gene,” and the gene names and placed that all in a new separate tab on the worksheet. After a PMID was used, the code also highlighted the cell so as a visualization that the code did check and used that PMID. The code ran through all the PMIDs of the abstracts first, then the researcher duplicated the code and tweaked it a bit so the full-text could run and the information gathered and placed into a new tab on the worksheet. Once all the data was obtained, the researcher created a code to get each unique gene name from each PMID, ensuring no duplicates of the same gene were added under a different name. As stated before, all the code done during the experiment can be seen on the web page (Link: <https://fellowship1954.github.io/Analyzing-Science-Articles-Identifying-Genes-Associated-with-Brain-Tumors/>).

*Analyzing the Data*

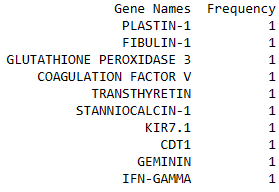
Once the gene names were gathered and placed into another tab in the worksheet, the code was created to grab and stored in a Python list. From there, the Pandas library counted the number of times a gene name was mentioned and stored that number by creating a frequency table. Other Python code was created to analyze the data, through bar graphs and word clouds in which the results are shown below, and again, the code, abstracts, and full-text tables, bar graphs and word clouds are available on the web page (Link: <https://fellowship1954.github.io/Analyzing-Science-Articles-Identifying-Genes-Associated-with-Brain-Tumors/>).

**Results**

For the experiment, 1,100 samples were examined, and around a total of 736 gene names were referenced. However, split between abstracts and full-text around 656 genes were mentioned in the abstracts and 136 in full-text. **Tables 1** and **2** show the top and bottom ten gene names with combined abstracts and full-text.

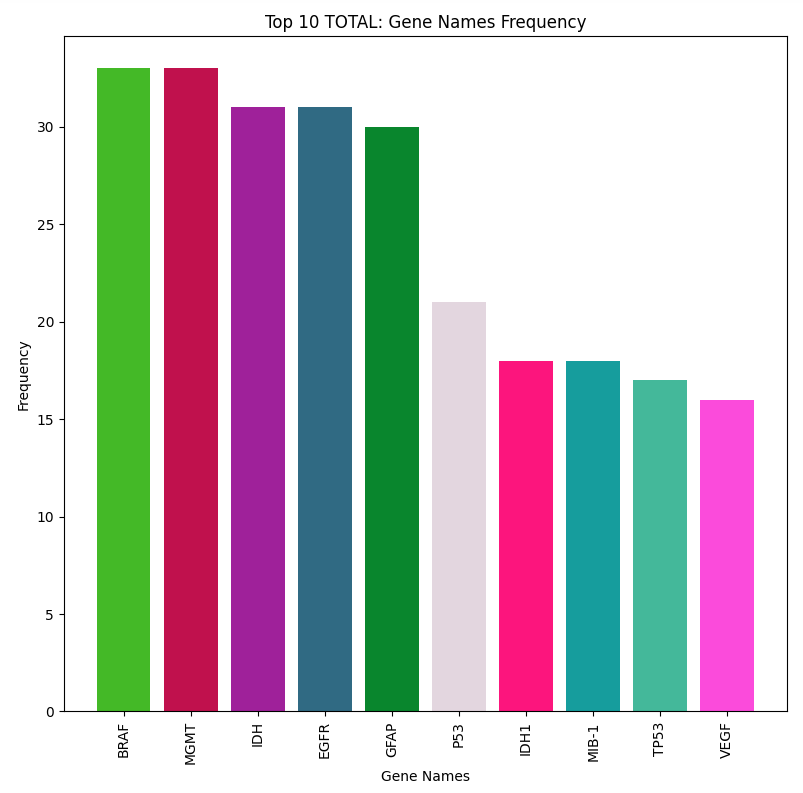
****

**Table 1:** Frequency table of the top 10 most frequently mentioned genes for abstracts and full-text. (**Note:** for a full table, see web page: <https://fellowship1954.github.io/Analyzing-Science-Articles-Identifying-Genes-Associated-with-Brain-Tumors/>)

****

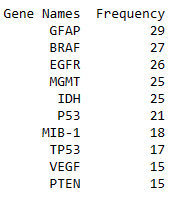
**Table 2:** Frequency table of the bottom 10 least frequently mentioned genes in total for both abstracts and full-text. (**Note:** for a full table, see web page: <https://fellowship1954.github.io/Analyzing-Science-Articles-Identifying-Genes-Associated-with-Brain-Tumors/>)

According to **Tables 1** and **2**, there were two genes that were the most mentioned, and all top ten least referenced had the same number. Overall, the BRAF and MGMT were the most referred to in total which is the combined data of abstracts and full-text. For a better view of the most mentioned genes overall, a bar graph (see **Graph 1**) was made, and it is a great visualization of BRAF and MGMT being referenced 33 times. The first 5 gene names mentioned were all in the thirties, with the first and second (BRAF and MGMT) having the same number, 33, and the third and fourth (IDH and EGFR) having the frequency of 31.

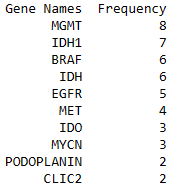


**Graph 1:** A bar graph that displays the top 10 most frequently mentioned genes in both the abstract and full-text.

However, when abstracts and full-text are separated the gene names most mentioned changes, and can be seen in **Table 3**. For abstracts, GFAP was mentioned 29 times and is ahead of BRAF by only 2. Additionally, in full-text alone, the MGMT was the most referenced with being mentioned only 8 times (see **Table 4**).

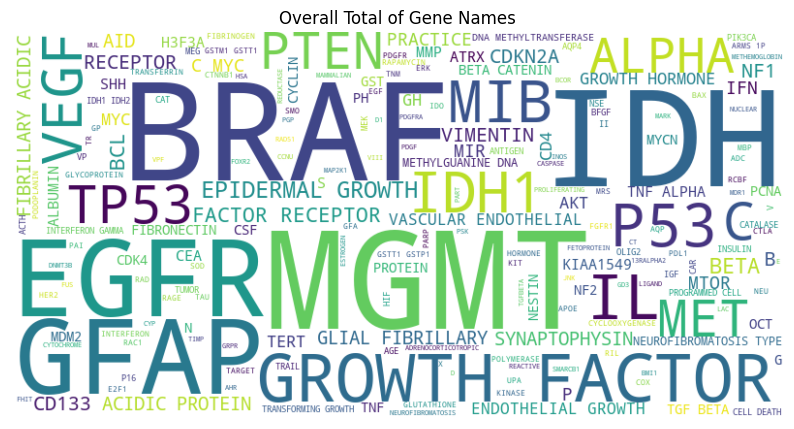
****

**Table 3:** Top 10 most frequently mentioned genes in only the abstracts.



**Table 4:** Top 10 most frequently mentioned genes in only the full-text.

For better visualization, a word cloud was created to show a visually pleasing graph showing BRAF and MGMT being the most mentioned overall due to its size (see **Graph 2**).



**Graph 2:** Word cloud that shows all the genes mentioned in both abstracts and full-text, with the bigger text BRAF and MGMT referred to most frequently.

Both the word cloud and frequency table show that BRAF and MGMT is the most mentioned in brain tumor articles and abstracts.

**Discussion**

*Overview*

Understanding genes associated with brain tumors provided insight into genetic factors linked to brain tumor development and prognosis through how much they were mentioned in biological articles and their abstracts. Looking at the results, it showed the most referred to gene along with the other top 10 frequently mentioned genes (see **Tables 1**), which will help researchers new to studying brain tumors on what genes are associated with this disease had been discovered and most focused on based on the number of times the gene name is referenced.

The results show out of the total 739 genes total, the BRAF and MGMT genes were the most referenced overall. Both genes had the same frequency of 33, with IDH and EGFR coming in second place with a frequency of 31. BRAF stands for B-Raf proto-oncogene which encodes for a protein that plays a role in regulating MAP kinase which affects cell division and other things along the line and if it is mutated it is most commonly found in cancer causing mutations [10]. The MGMT gene also called O-6-methylguanine-DNA methyltransferase, which codes for a protein that repairs and protects cells from mutagenesis and toxicity, also seen in cancer causing mutations [10 a]. When looking at the data separately, the abstracts and the full-text there is a slight difference in the ranking of the gene names. For in the abstracts only the gene GFAP is at the top with 29 frequencies or mentions, and BRAF in second with 27 frequencies. When just looking at the full-text the gene MGMT was first with the highest number, being 8. In comparing the number of frequencies between abstracts and full-text there is a considerable difference for the abstracts’ top mentioned genes were in the twenties, the full-text was in the single digits (see **Tables 3 and 4**). It is good to note that majority of the genes found in total only were mentioned once, about 540 to be exact, which can be seen in the full total table (see web page: <https://fellowship1954.github.io/Analyzing-Science-Articles-Identifying-Genes-Associated-with-Brain-Tumors/>).

*Reflection and Limitations*

The methods used during the experiment were more complex, or in other words, a long way to obtain the information was used. Being new to PubTator and PubMed, the researcher did not initially realize there were other, more straightforward methods to get the data but learned them along the way. One lesson is that although PubTator API states that abstract and full-text code to obtain the data are slightly different, they are not. Using just the code for abstracts will work for both. Another lesson learned was that PubMed has filters for abstracts and full-text, so instead of downloading the list of samples, splitting them into abstracts and full-text was unnecessary. Of course, looking through each chunk of PMIDs would still be the same steps to see what documents contain gene information, for not all of them would. As a precaution, to ensure there are no duplicates, the steps on comparing the PMIDs for abstracts and full-text are still done. Other than those lessons, the rest of how the data was obtained would stay the same.

Although the experiment went over well and lots of precautions were taken to obtain sufficient data to make a scientific reasoning or conclusion based on the information gathered that BRAF and MGMT are the most known gene associated with brain tumors, there were a few limitations to the study. One limitation of the experiment is a small data sample size was used. Although 1,100 samples might seem sufficient, remember there were around 226,378 results but only 10,000 PMIDs could be downloaded and saved at a time. After the samples, there were still 8,556 more PMIDs that could have been looked through if there was no deadline for the project. Time is another limitation, for a deadline was required for the project because it is a senior research project. When using PubTator, the way the website operated made it hard for quick keyboard shortcuts to be used and splitting the screen harder than it should have been. For splitting the screen, PubTator had to take up most of the screen to see the number of collections and other important aspects of the website, so going back and forth with it took time. Also, the enter button does not work in PubTator for the search bar, which made it an inconvenience when copying and pasting the PMIDs in the search bar by using the keyboard shortcuts worked; however, then the mouse cursor had to be moved and click on the search icon to start the request which affected the time it took to get the necessary data. Relying on the accuracy of PubTator code, looking at biochemical genes being properly labeled was another limitation found during the experiment. Like all code, it sometimes makes mistakes, especially when finding text. For example, PRODUCT and AID were counted as genes in articles since the letters are similar to those of a specific gene name, such as PRAC1 and ACIDA, so some mentioned genes might not have been in an article [9]. The code only made some of these errors due to the inconsistency of scientists calling genes by different names. PubTator’s gene code contains a section when trying to find genes called “alias,” which lists all the different alias names that scientists call a gene [9]. Another limitation is that the code created might not have captured all of the genes for a few shared the same or more than one identifier number (see **Figure 4**). If more time was given then it would be used to check the code more thoroughly one a much smaller sample size before transitioning to a sample size that is over 1,000.











**Figure 4:** Screen shots of just a few gene names with more than one identifier.

*Last Remarks*

Despite the limitation, PubTator is still a powerful tool for researchers to study types of biochemicals in diseases, such as brain tumors. It can be seen in similar research studies where researchers used PubTator to help provide data. One group of researchers created a website containing data collection on tumor suppressor gene (TSG) biological functions associated with various cancers [11]. They developed an updated database of their old model, a literature base database that provides resources for cancer research, and the researchers looked for keywords in abstracts from PubMed and extracted the gene names [11]. The researchers continued doing other things while collecting the data, with the results being a database that contains information on various types of tumor suppressor genes, like how they are expressed and mutated in different types of cancer, and more biological information. The TSG research is similar to this project, taking biological information about genes and their association with diseases like cancer and brain tumors from scientific literature.

Similar to the other related work, a research group also created their database website; however, it was based on genes associated with leukemia. These researchers also used leukemia-based literature from the NCBI website to obtain the proper gene information for their database, which contains around 1805 genes associated with leukemia [12]. PubTator made it easier for researchers for this project or others to look at biological information and their associated diseases in one spot for easy access.

Overall, the experiment was successful, showing that BRAF and MGMT were the most talked about gene concerning brain tumors. If the experiment could be repeated, a few methods would be changed, such as more time, leading to a bigger sample size and an easier way to obtain the data, as described above. In any case, the experiment should help provide researchers with necessary information on the genes that are associated the most with brain tumors, which has the potential to help further other research in brain tumors to understand the disease in the hope that one day, more efficient treatments will develop and lives will be saved.

[1] “Cancer Facts & Figures 2023,” *Am. Cancer Soc.*, 2023. Pages 12 and 14. [Online]. Available: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2023/2023-cancer-facts-and-figures.pdf

[2] L. M. DeAngelis, “Brain Tumors | NEJM,” The New England Journal of Medicine. Accessed: Oct. 03, 2023. [Online]. Available: https://www.nejm.org/doi/full/10.1056/NEJM200101113440207

[3] “What causes brain tumours?,” The Brain Tumour Charity. Accessed: Oct. 12, 2023. [Online]. Available: https://www.thebraintumourcharity.org/brain-tumour-diagnosis-treatment/how-brain-tumours-are-diagnosed/brain-tumour-biology/what-causes-brain-tumours/

[4] “Genes: Function, makeup, Human Genome Project, and research.” Accessed: Oct. 12, 2023. [Online]. Available: https://www.medicalnewstoday.com/articles/120574

[5] “Types of Brain and Spinal Cord Tumors in Children.” Accessed: Oct. 03, 2023. [Online]. Available: https://www.cancer.org/cancer/types/brain-spinal-cord-tumors-children/about/types-of-brain-and-spinal-tumors.html

[6] “Brain Tumor - Statistics,” Cancer.Net. Accessed: Dec. 05, 2023. [Online]. Available: https://www.cancer.net/cancer-types/brain-tumor/statistics

[7] A. L. Albright, “Pediatric brain tumors,” *CA. Cancer J. Clin.*, vol. 43, no. 5, pp. 272–288, 1993, doi: 10.3322/canjclin.43.5.272.

[8] M. Zhao, Y. Liu, G. Ding, D. Qu, and H. Qu, “Online database for brain cancer-implicated genes: exploring the subtype-specific mechanisms of brain cancer,” *BMC Genomics*, vol. 22, no. 1, p. 458, Jun. 2021, doi: 10.1186/s12864-021-07793-x.

[9] C.-H. Wei, A. Allot, R. Leaman, and Z. Lu, “PubTator central: automated concept annotation for biomedical full text articles,” *Nucleic Acids Res.*, vol. 47, no. W1, pp. W587–W593, Jul. 2019, doi: 10.1093/nar/gkz389.

[10] National Center for Biotechnology Information. “BRAF B-Raf proto-oncogene, serine/threonine kinase [ Homo sapiens (human)],” Accessed: Dec. 05, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/gene/673#summary

[10 a] National Center for Biotechnology Information. “MGMT O-6-methylguanine-DNA methyltransferase [ Homo sapiens (human)],” Accessed: Dec. 05, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/gene/673#summaryhttps://www.ncbi.nlm.nih.gov/gene/4255

[11] M. Zhao, P. Kim, R. Mitra, J. Zhao, and Z. Zhao, “TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes,” *Nucleic Acids Res.*, vol. 44, no. D1, pp. D1023–D1031, Jan. 2016, doi: 10.1093/nar/gkv1268.

[12] Y. Liu, M. Luo, Z. Jin, M. Zhao, and H. Qu, “dbLGL: an online leukemia gene and literature database for the retrospective comparison of adult and childhood leukemia genetics with literature evidence,” *Database*, vol. 2018, p. bay062, Jan. 2018, doi: 10.1093/database/bay062.