

# BIOLOGICAL DATABASE

**BIT2002**

**WINTER SEMESTER 2021-22**

***ORAL PRECANCER DATABASE***

### SUBMITTED BY

**LAKSHIT M. SANGHRAJKA - 19BCB0031**

**NITISH JHA – 19BCB0068**

**JAI PARWANI – 20BCB0139**

### ASSIGNED BY

**SATARUPA BANERJEE**

**Introduction**:

* Oral Cancer burden poses a major healthcare challenge, not just in India but also worldwide. It is estimated that nearly 1/3rd of all oral cancer cases worldwide are from India. In our country, cancer of the lips and the oral cavity account for approximately 10.4% of all new cases annually.
* Despite ease of access to the oral cavity for examination, poor outcomes have been largely attributed to late identification of disease, lack of access to tertiary oncology centers and high costs not afforded by many. Most cases of oral cancer are preceded by potentially malignant conditions such as oral leukoplakia, oral sub-mucous fibrosis and erosive lichen planus.
* Prevention through tobacco cessation programmes and early disease identification offer the best long term prospects of mitigation. In light of the above, we have initiated this project to collect data on oral pre cancer from across the country.
* In the post genomic era, efforts to understand oral cancer are aimed at obtaining genomic and proteomic profiles which generate enormous data. Thus the challenge is to analyze the data effectively and utilize the information for diagnosis, prognosis and treatment.

**Project Description:**

The aim of the project is to curate a well-defined database regarding **Oral Precancer** byprovidinga system that is capable of handling details of a genes involved and mutated in oral precancer stage . The objective is to create a website that can store details of these oral precancer genes in the most feasible form of database. This is done to make documentation of this gene records digital to enable easy access to a database in which these affected genes can be entered and updated based on the choices of the user, which will bring down the amount of time spent in searching records compared to that of doing on the other websites. This will also facilitate storage and management of large gene bodies which would be very difficult via general websites. Main motivation to create this project is to create a very rich domain specific web tool for Oral precancer and enable better means for retrieval and submission of data as this would help in increasing the quality of surfing the gene oriented knowledgebase via user. Providing well developed digital systems to automate storing of records will help increase the efficiency of the website to function thereby making it grow. By allowing a greater number of this disease specific genes to be recorded, will result in a larger number of researchers who can work at higher levels and help in developing the knowledge enriched environment.

**Novelty:**

## Since no proper well defined database exist just for oral precancer genes we have tried our best in developing a website just for oral precancer genes and giving various details of their genes which is distributed over multiple platforms and websites like

## **The Human Gene Mutation Database**

## **NCBI dbSNP**

## **UniprotKB**

## **NCBI**

## **KEGG PATHWAY Database**

## **THE HUMAN PROTEIN ATLAS**

## **PUBMED**

We have made an effort to collect data from all this websites and available databases regarding the gene and provide it by assembling it into a single platform for the ease of the user.

**Literature Survey:**

# Paper 1: Oral Cancer and Precancer: A Narrative Review on the Relevance of Early Diagnosis

**Authors:** [Silvio Abati](https://sciprofiles.com/profile/1283233), [Chiara Bramati](https://sciprofiles.com/profile/1356391), [Stefano Bondi](https://sciprofiles.com/profile/1138147), [Alessandra Lissoni](https://sciprofiles.com/profile/1377816), [Matteo Trimarchi](https://sciprofiles.com/profile/1138146)

**Abstract:** Potentially malignant oral epithelial lesions (PMOELs) are a group of oral conditions and diseases that can be present before the onset of OSCC and include a group of clinically suspect oral mucosal lesions such as leukoplakia, erythroplakia, submucosal fibrosis, and lichen planus. The majority of PMOELs, however, do not progress to cancer. Surgical biopsy remains the gold standard for diagnosis of oral cancer. Adjunctive tools have been developed and studied to help clinicians in the diagnostic pathway, such as toluidine blue vital staining and autofluorescence imaging. In the near future other methods, i.e., identification of salivary markers of progression may help in reducing mortality due to oral cancer.

# Paper2: Association between risk of oral precancer and genetic variations in microRNA and related processing genes

**Authors:** Roshni Roy, Navonil De Sarkar, Sandip Ghose, Ranjan R Paul, Anindita Ray, Indranil Mukhopadhyay & Bidyut Roy

### **Abstract:** MicroRNAs have been implicated in cancer but studies on their role in precancer, such as leukoplakia, are limited. Sequence variations at eight miRNA and four miRNA processing genes were studied in 452 healthy controls and 299 leukoplakia patients to estimate risk of disease. Results

showed that variant genotypes at *Gemin3* and *mir-34b* reduced risk of disease in overall patients as well as in smokers. Among chewers, only *mir29a* significantly increased risk of disease. Gene-environment interactions using MDR-pt program revealed that *mir29a*, *mir34b, mir423* and *Xpo5* modulated risk of disease (p < 0.002) which may be related to change in expression of these genes as observed by Real-Time PCR assays. But association between polymorphisms and gene expressions was not found in our sample set as well as in larger datasets from open access platforms like Genevar and 1000 Genome database.

Variations in microRNAs and their processing genes modulated risk of precancer but further in-depth study is needed to understand mechanism of disease process.

# Paper3: DNA methylation markers for oral pre-cancer progression: A critical review

**Author****s:** Krithiga Shridhar, Gagandeep Kaur Walia, Aastha Aggarwal, Smriti Gulati, A.V. Geetha, Dorairaj Prabhakaran, Preet K. Dhillon, Preetha Rajaraman

**Abstract:** Although oral cancers are generally preceded by a well-established pre-cancerous stage, there is a lack of well-defined clinical and morphological criteria to detect and signal progression from pre-cancer to malignant tumours. We conducted a critical review to summarize the evidence regarding aberrant DNA methylation patterns as a potential diagnostic biomarker predicting progression. We identified all relevant human studies published in English prior to 30th April 2015 that examined DNA methylation (%) in oral pre-cancer by searching PubMed, Web-of-Science and Embase databases using combined key-searches. Twenty-one studies (18-cross-sectional; 3-longitudinal) were eligible for inclusion in the review, with sample sizes ranging from 4 to 156 affected cases. Eligible studies examined promoter region hyper-methylation of tumour suppressor genes in pathways including cell-cycle-control (n = 15), DNA-repair (n = 7), cell-cycle-signalling (n = 4) and apoptosis (n = 3). Hyper-methylated loci reported in three or more studies included p16, p14, MGMT and DAPK. Two longitudinal studies reported greater p16 hyper-methylation in pre-cancerous lesions transformed to malignancy compared to lesions that regressed .The one study that explored epigenome-wide methylation patterns reported three novel hyper-methylated loci (TRHDE; ZNF454; KCNAB3). The majority of reviewed studies were small, cross-sectional studies with poorly defined control groups and lacking validation. Whilst limitations in sample size and study design preclude definitive conclusions, current evidence suggests a potential utility of DNA methylation patterns as a diagnostic biomarker for oral pre-cancer progression. Robust studies such as large epigenome-wide methylation explorations of oral pre-cancer with longitudinal tracking are needed to validate the currently reported signals and identify new risk-loci and the biological pathways of disease progression.

**Methodology:**

According to our topic, Oral Precancer Database we started by understanding the prime difference between oral cancer and oral precancer i.e. Oral precancer is just precursor to oral cancer and knowing it symptoms, cause, remedies and treatment .

We indexed the genes which were reported to have contributed to oral precancer and extracted its details on various parameters .

So we took help from various pre-existing databases platforms like NCBI, Pubmed, Uniprot etc and identified the genes associated with it by using appropriate keywords like “oral squamous cell carcinoma”, “oral precancerous”, etc.

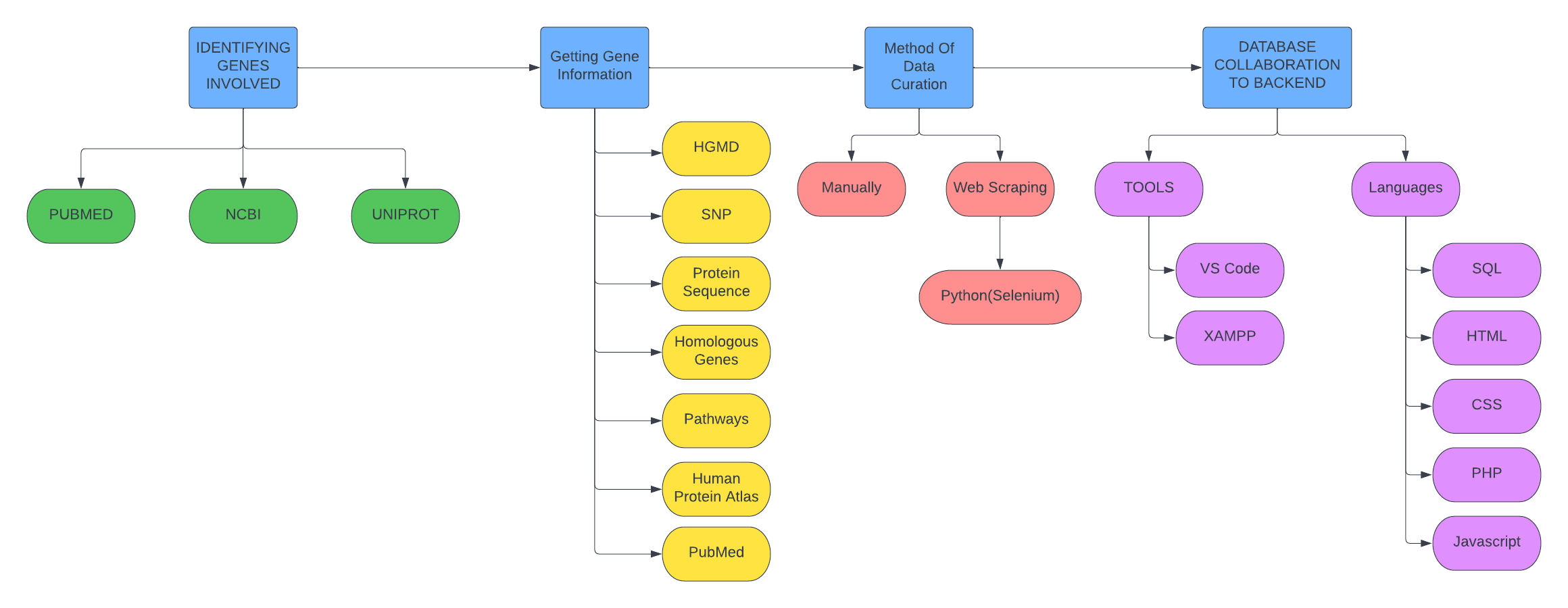
Due to no prior proper and well defined database availability in this area we had to rely more on manual curation but we also used python code for web scraping. Furthemore an important part after data extraction is to remove data redundancy which is an important aspect as data would be inputted in the backend and it is very important for the data to be error free.

## After identifying the set of genes in order to give various information about the gene we manually started extracting it details from different available databases like The Human Gene Mutation Database, NCBI dbSNP, UniprotKB, NCBI, PATHWAY Database, THE HUMAN PROTEIN ATLAS, PUBMED and started storing the information for creation of our one final database.

Our front end was decided upon to be a website with simple UI to facilitate exchange of knowledge between researchers with the help of our database.

You will be able to search a gene and it retrieve various details about it available across various other databases

FLOWCHART:



## **SOFTWARE/LANGUAGES USED:**

* VS IDE
* PYTHON (SELENIUM FOR AUTOMATION AND WEB SCRAPING)
* HTML
* CSS
* JAVA SCRIPT
* BOOTSTRAP
* PHP/SQL

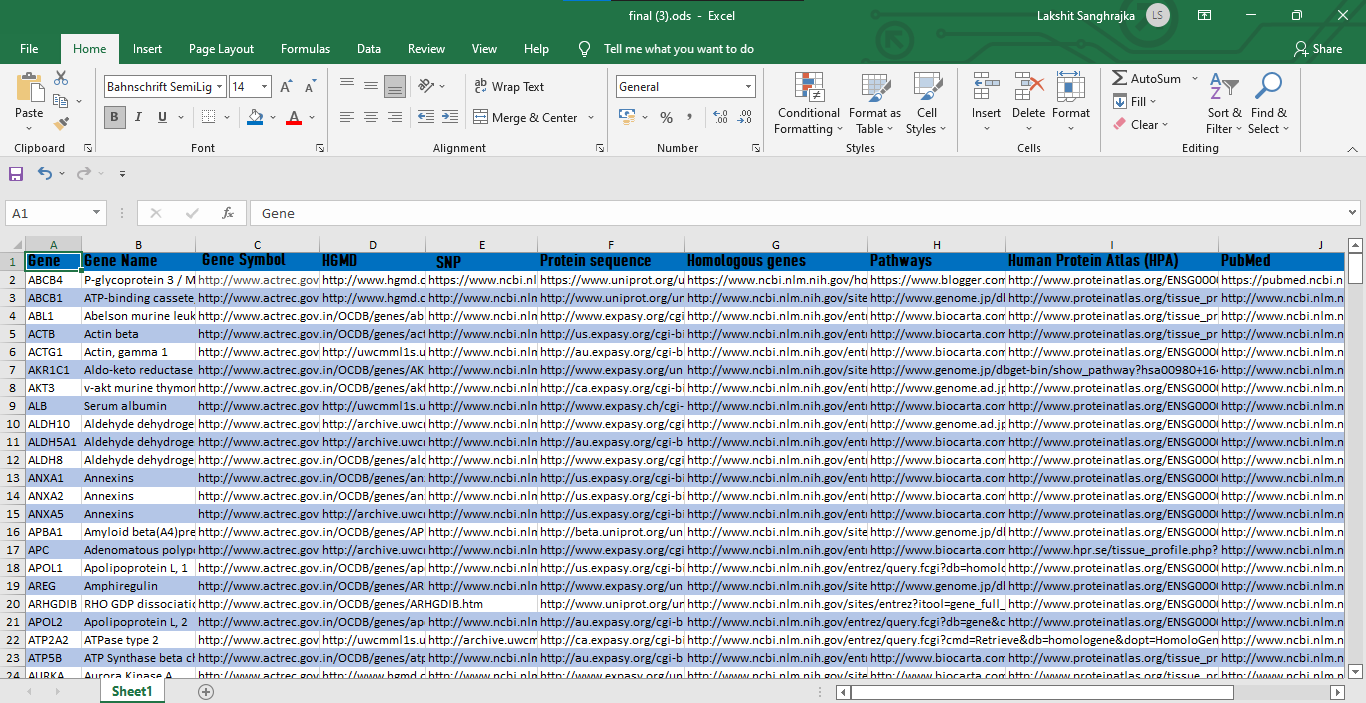
**References:**

**Websites Referred for Data Creation and Extraction:**

<https://pubmed.ncbi.nlm.nih.gov/?term=oral+precancer>

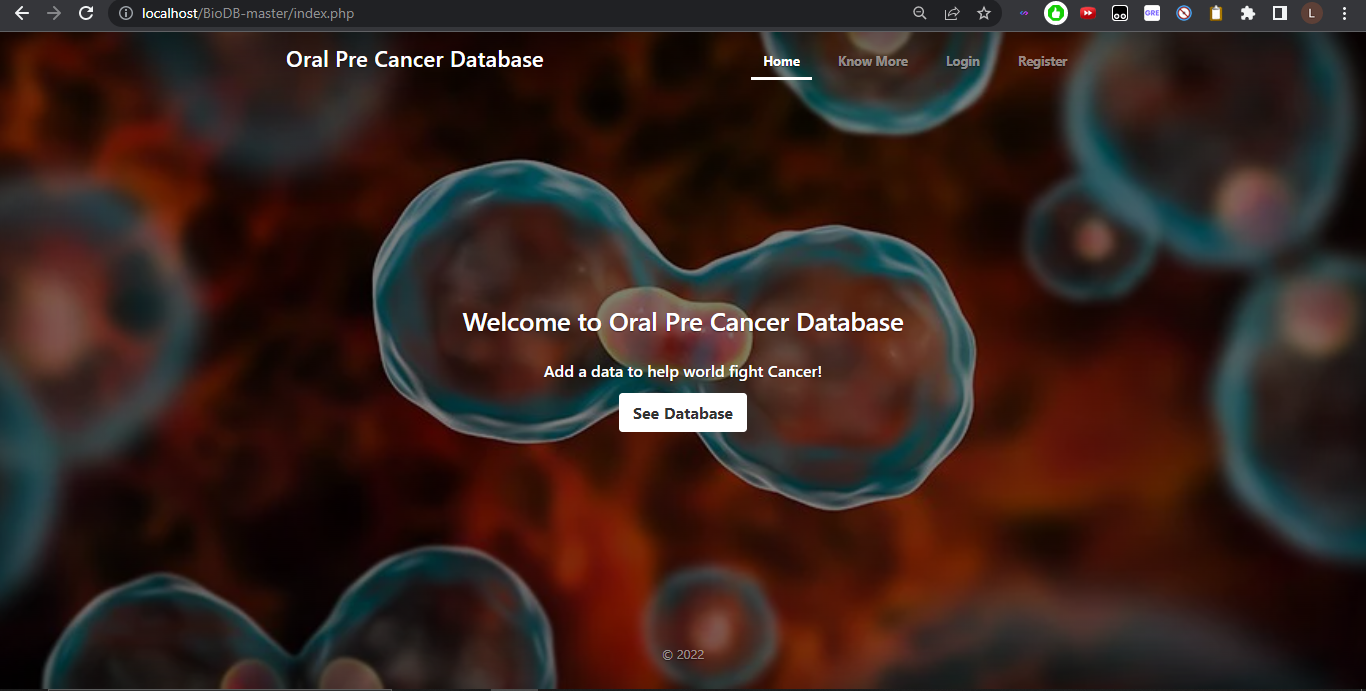
<https://www.uniprot.org/uniprot/?query=oral+cancer&sort=score>

**Database Manually curated:**



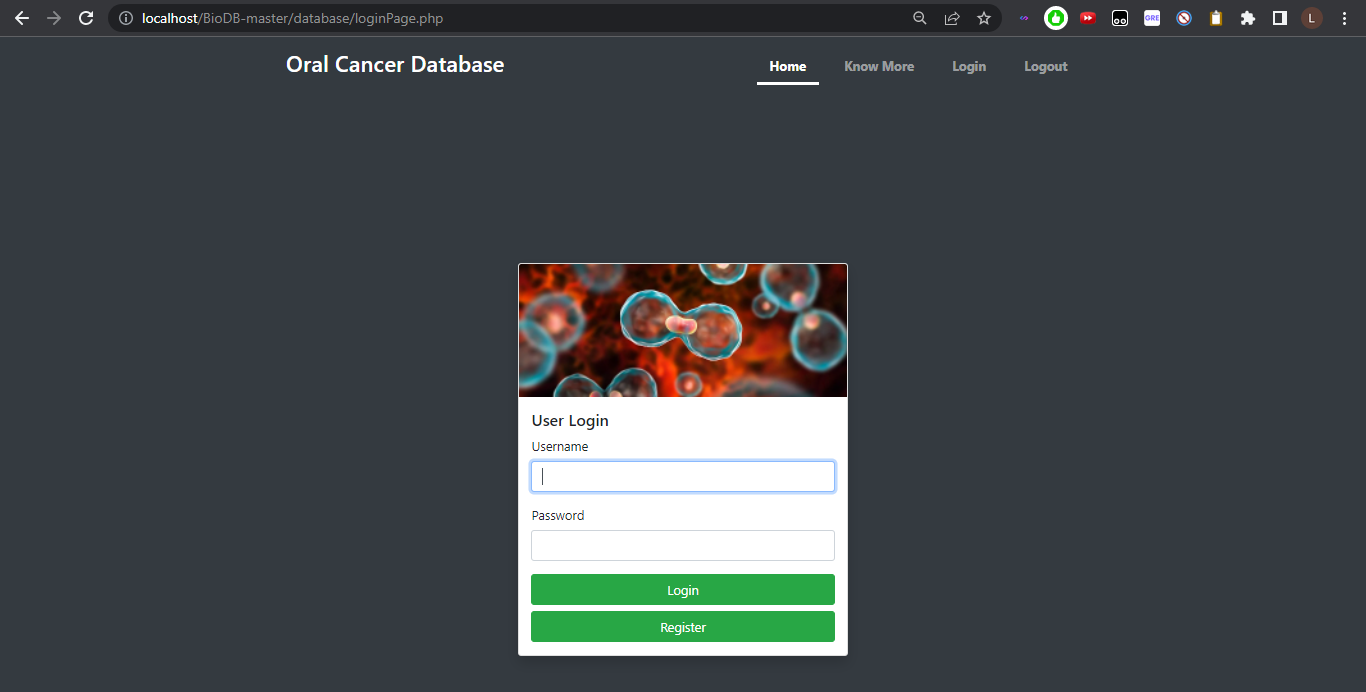
**Website Basic Frontend:**

HOME PAGE:

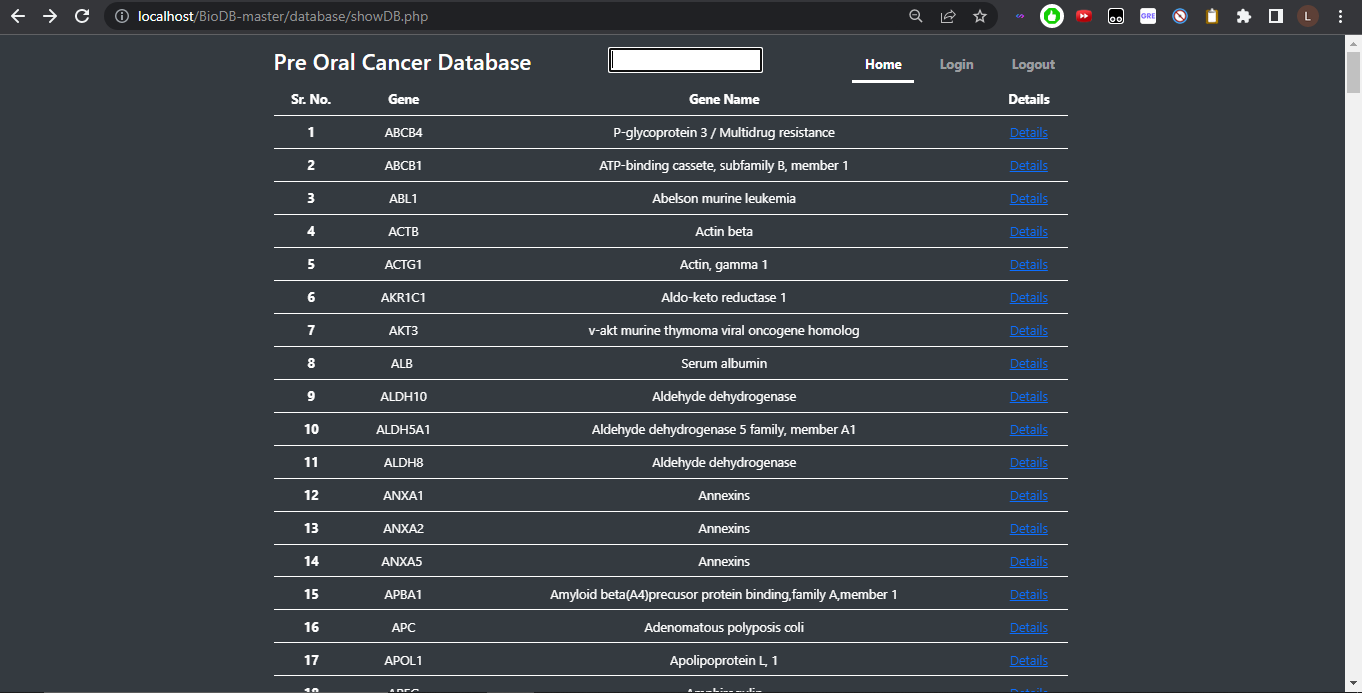


Login Page:

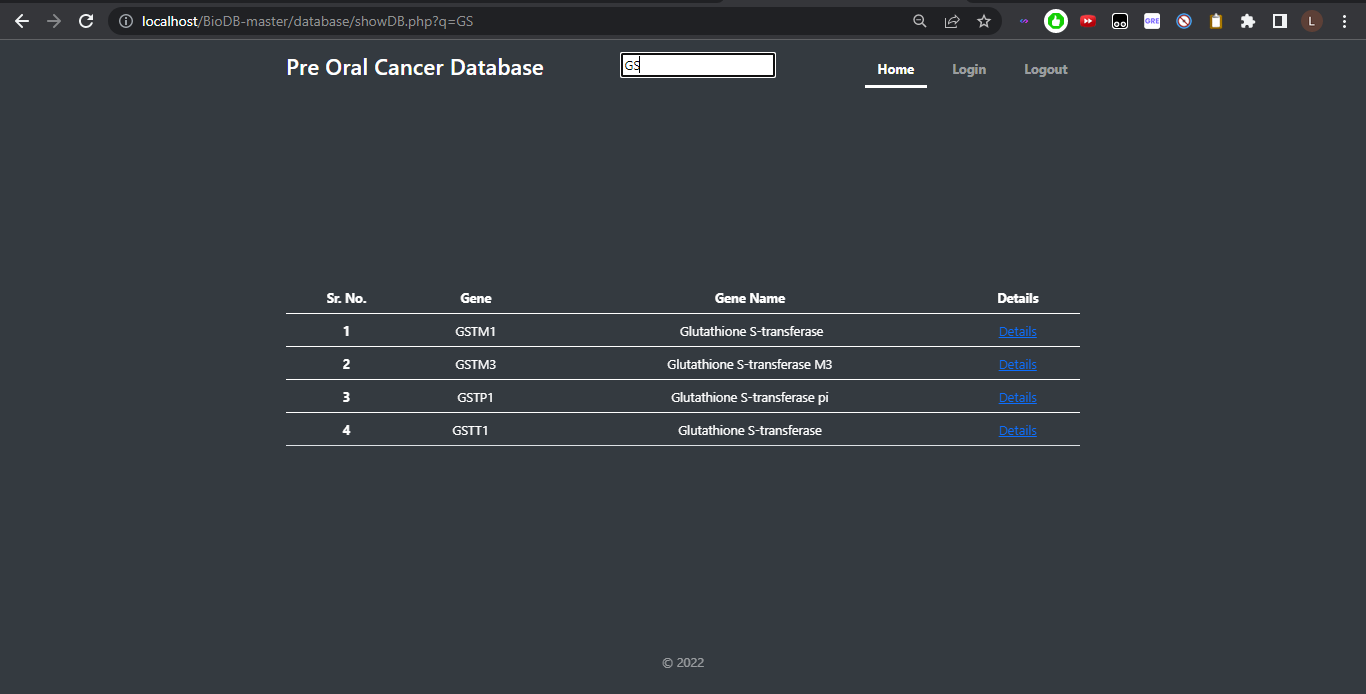
For adding security and especially for adding records which would be evaluated and can be approved by the administrator



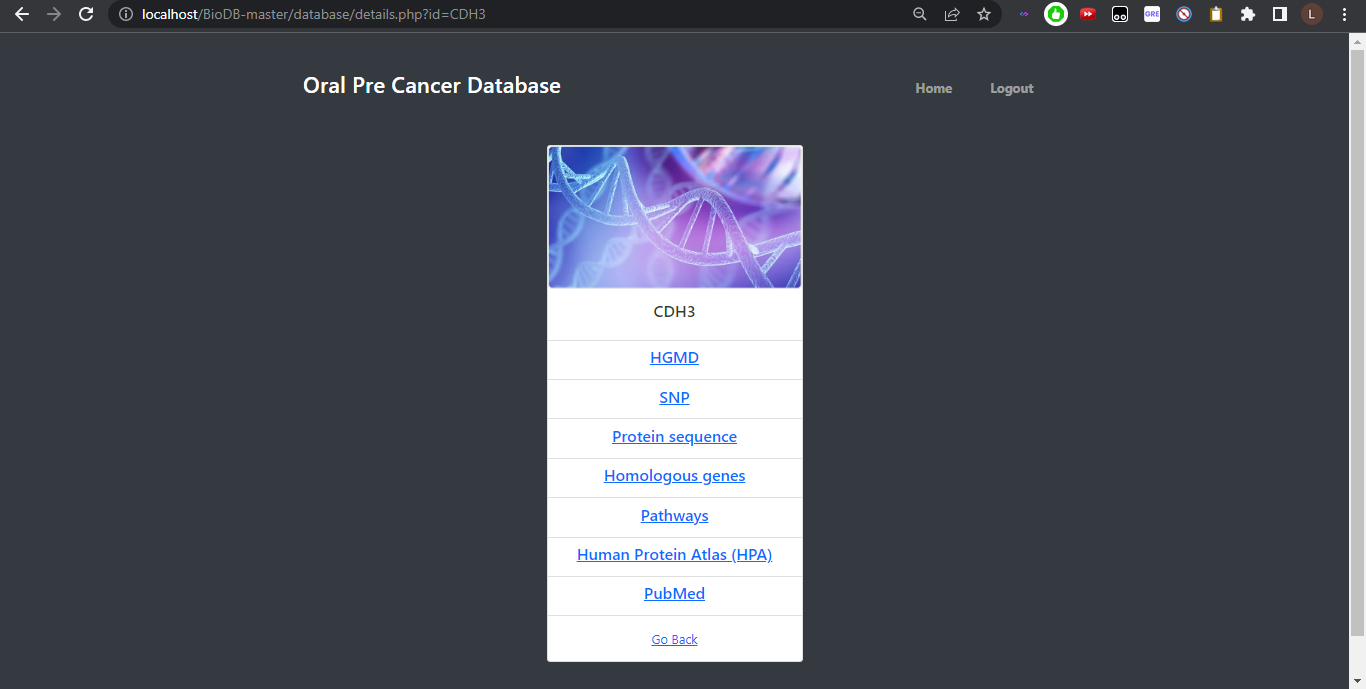
SHOW DATABASE:



SEARCH DATABASE BY GENE:

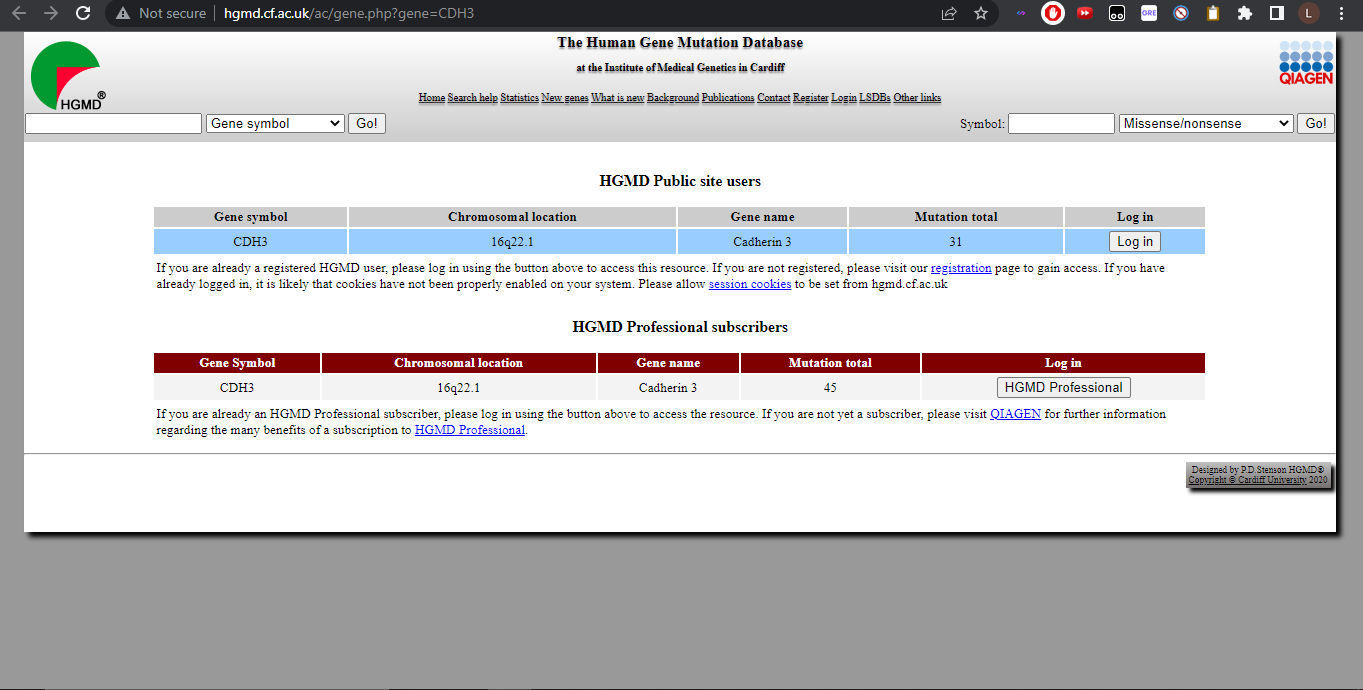


GENE DETAILS:



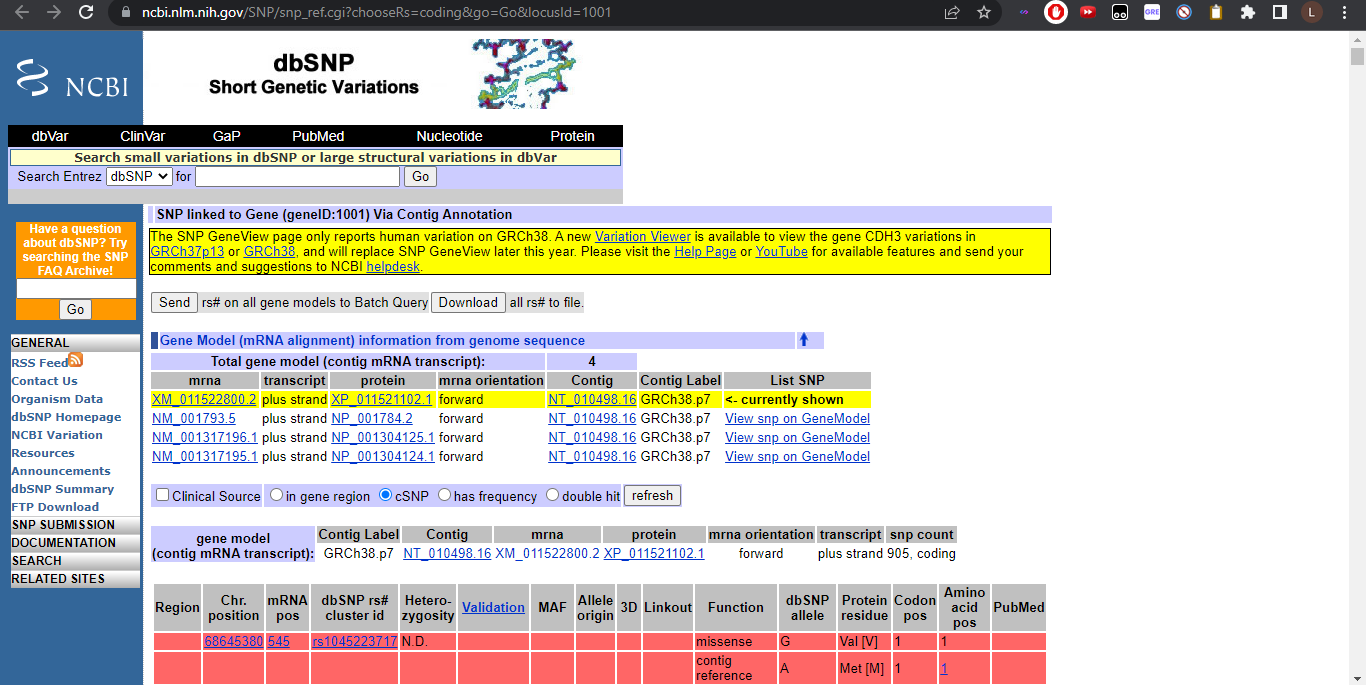
HGMD:

On clicking HGMD redirects to The Human Gene Mutation Database for that gene



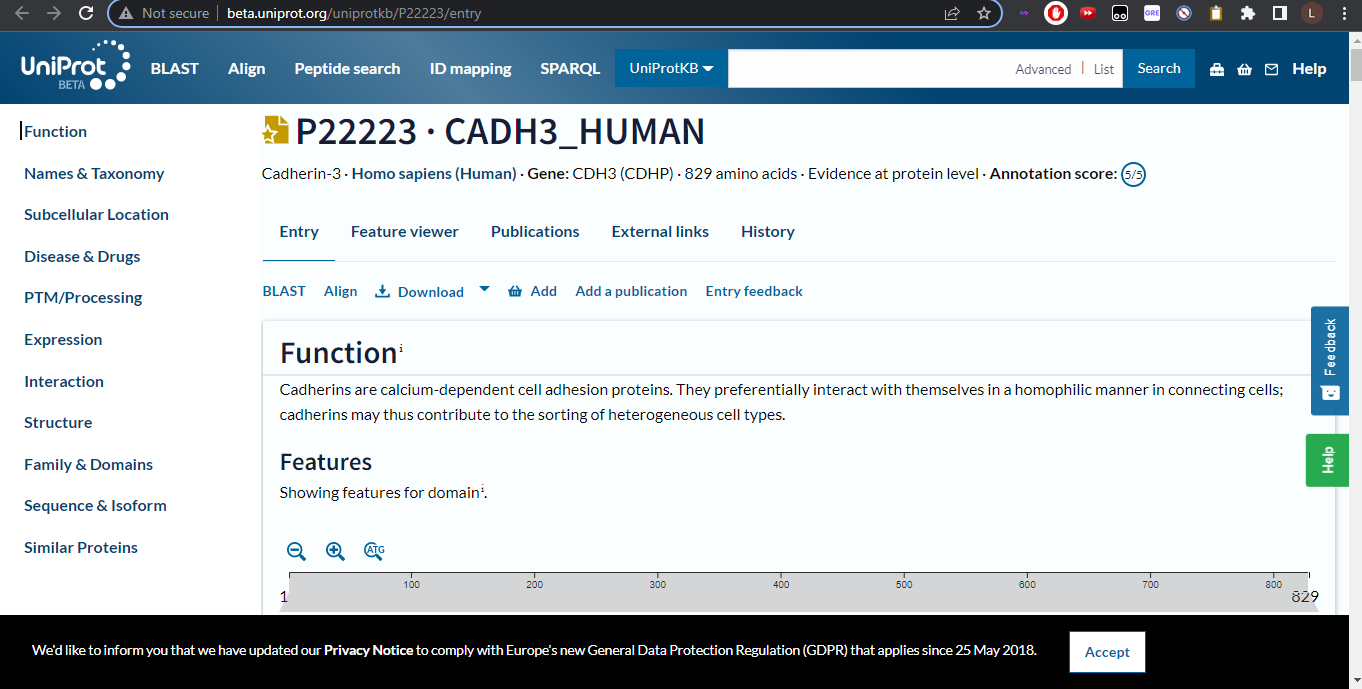
SNP:

On clicking SNP redirects to NCBI dbSNP which gives information about single nucleotide polymorphism i.e. the most common genetic variation for that particular gene



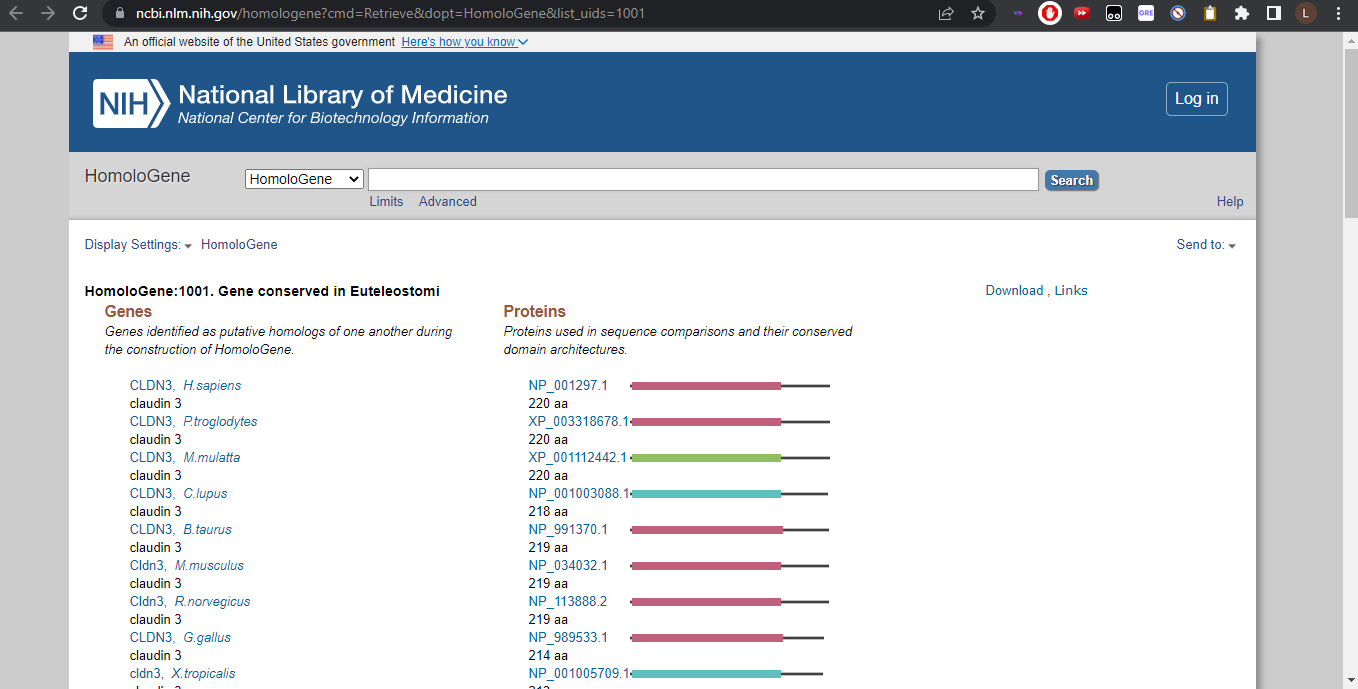
Protein Sequence:

On clicking Protein Sequence redirects to Uniprot which shows protein sequence and functional annotation for the gene



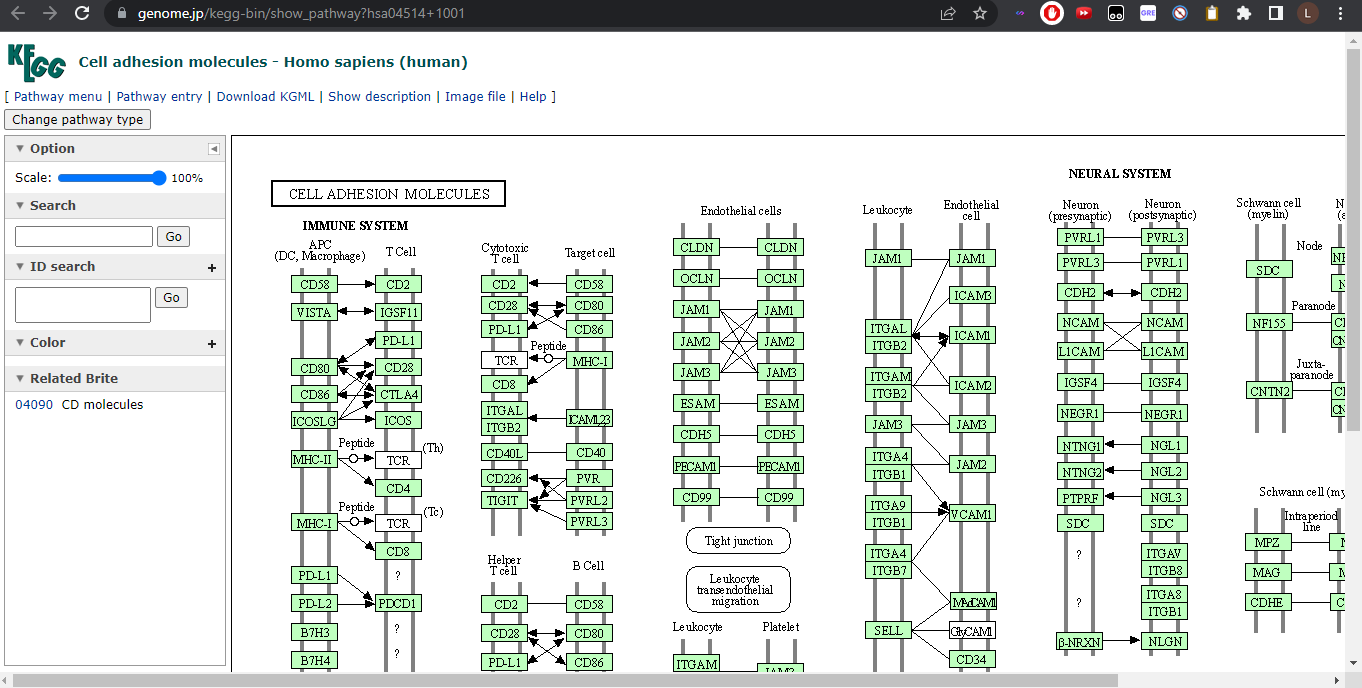
Homologous Genes:

On clicking Homologous Gene it redirects to the NCBI Homologene filter for that particular Gene



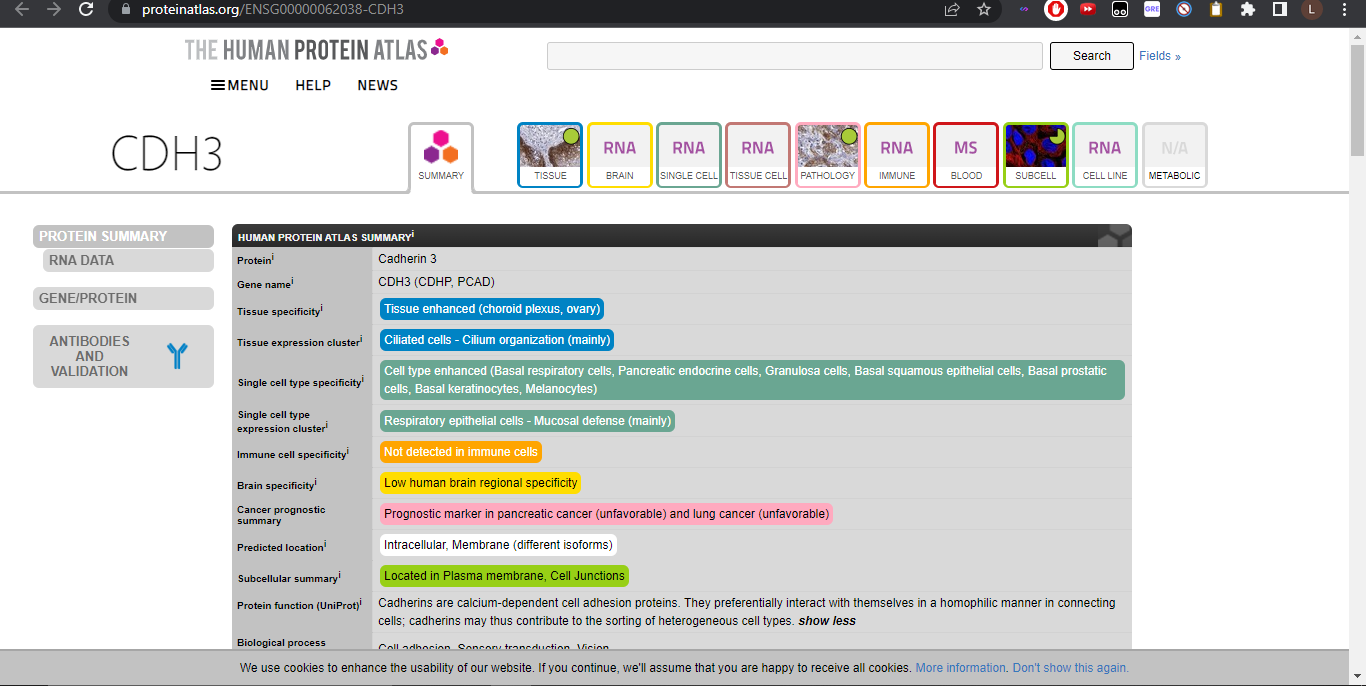
Pathways:

Pathways redirects to the KEGG and provides KEGG pathway analysis for that gene



Human Protein Atlas (HPA):

Human Protein Atlas (HPA) redirects to the Human Protein Atlas (HPA) website and provides the protein summary along with some visualizations



Pubmed:

Pubmed references provides the pubmed references for that gene from PubMed NCBI website

