

Cancer Publication Portal: Identifying Gene-Cancer Associations from Biomedical Literature

Gargano, M.A., Dancik, G.M.

Opportunity

We describe a web application interface capable of summarizing the gene-cancer associations in PubMed, an online database of biomedical research articles with exponential growth of cancer articles in the past 15 years. Cancer is a deadly genetic disease where gene mutations result in abnormal cellular functions. Mutations may include substitutions of one base by another, insertions or deletions of varying sizes, or DNA rearrangements. Often gene expression is drastically effected by these mutations, which can drive abnormal cell activity.

Approach I - Objectives

- Identify gene-cancer associations by performing biomedical text mining
- Creating a user interface that
 - allows a user to select a gene to explore its statistical associations with cancer types
 - provides a user extra information such as statistics, links, graphs, and word clouds
 - provides a user the functionality to review abstracts of associated cancer articles

Approach II - Methods

- **Biopython**, a set of a python tools for retrieving PubMed data
- **MongoDB**, a non-relational database to store the PubMed data and gene-cancer relations.
- **Node.js**, **Express.js**, and **Angular.js** an asynchronous Javascript web application framework.

Results

The figure displays three screenshots of the Cancer Publication Portal interface, illustrating the results of a search for the gene TP53. The interface is divided into sections for searching, reviewing articles, and displaying statistics.

Search Results: The search results show TP53 associated with various cancer types, including Lung Cancer, Prostate Cancer, Colon Cancer, and Pancreatic Cancer. The results are presented in a table with columns for Tumor Type, Articles, Total Articles, % (By Tumor), and % (By Gene).

Tumor Type	Articles	Total Articles	% (By Tumor)	% (By Gene)
Lung	178	27615	0.0064	0.2572
Bladder	165	23050	0.0072	0.2384
Colon	140	28522	0.0049	0.2023
Pancreatic	138	25693	0.0054	0.1994
Prostate	71	26595	0.0027	0.1026

Review TP53 in PubMed Articles: The review section displays a list of articles related to TP53, including titles, authors, and abstracts. The articles are sorted by relevance, and the user can click on the title to view the full article.

Statistics: The statistics section provides a summary of the search results, including the total number of articles, the number of articles by tumor type, and the percentage of articles by tumor type.

Impact

The Cancer Publication Portal will have a significant impact in cancer research. Utilization of the Cancer Publication Portal will drastically reduce the duration of time currently needed when scientists are investigating the biomedical literature for a cancer-related gene. The application will be able to easily identify if a gene of interest has an association with a specific tumor type and through text-mining, help the user to quickly understand the nature of the association.

As we look to the future, research expansion will enable incorporation of other helpful information to scientists such as a gene coding sequence, variations, and common polymorphisms. This portal will be a 'one stop shop' when reviewing genes related to cancer.