Northeastern University

Cancer Publication Portal: Identifying Gene-Cancer Associations from Biomedical Literature Gargano, M.A., Dancik, G.M.

Opportunity

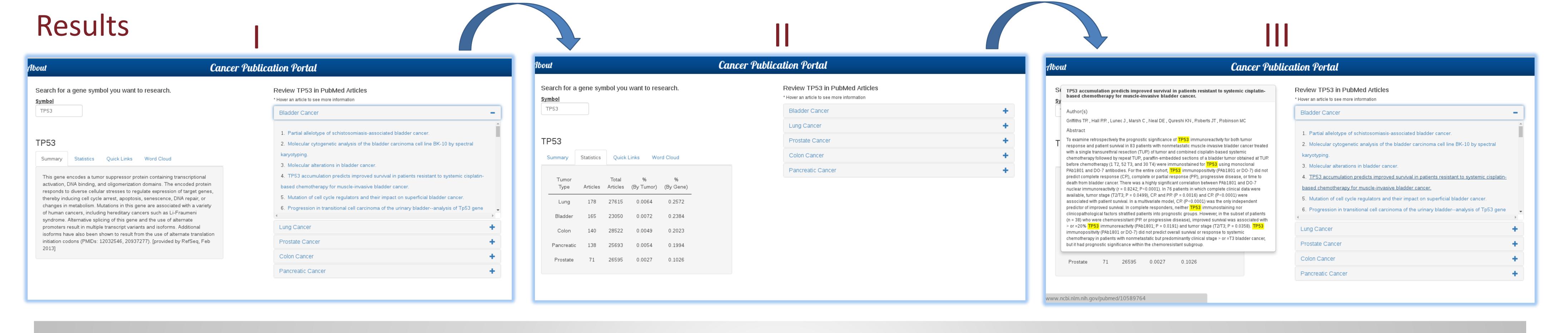
We describe a web application interface capable of summarizing the gene-cancer associations in PubMed, an online database of biomedical research articles with exponential growth of cancer articles in the past 15 years. Cancer is a deadly genetic disease where gene mutations result in abnormal cellular 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
 - I. allows a user to select a gene to explore its statistical associations with cancer types
 - II. provides a user extra information such as statistics, links, graphs, and word clouds
 - III. 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.



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.