Abstract

Rapidly evolving genomics tools and technologies have contributed to a dramatic rise in the volume and complexity of cancer precision medicine literature. This makes it challenging for an oncologist to search, perceive and use the information for personalized treatment of patients. Moreover, manual curation from literature is time-consuming and expensive. Hence, there is an urgent need to automatically extract information from literature to assist curators and clinical researchers. In this work, we have developed a text mining system to address this need.

Our natural language processing (NLP) tool detects different entities from PubMed abstracts and extracts the entity relationships that indicate the impact of genomic anomalies on cancer therapeutics. Our tool accounts for a variety of ways such relationships can be described in text. In addition to assist with such extraction of information, we have extended and repurposed multiple in-house and public tools for expert validation and ranking. All extracted results are stored in a database and available for curators and clinical researchers via an interactive web interface. The extracted information relating genomic anomalies to drug responses will enable researchers to readily generate hypotheses for new precision medicine based clinical trials.

Aims

- Identify impact of genomic anomalies on drug responses from biomedical text.
- Genomic anomalies:
  - Mutations
  - Change in expression level
- "Drug responses" is a broader term which incorporates ideas such as
  - Survival rate after drug treatment
  - Response rates after treatment
  - Sensitivity to drugs etc.
- Different ways to convey the relationship in text:
  - Association
    - Relation between mutation/expression and drug response
    - A correlation was observed between BRCA-1 expression and benefit from Feedback & Endocrine treatment (M201238899)
  - Sensitization to drugs
    - Mutation/expression affect the sensitivity to drugs
    - low expression of BRCA1 is a significant independent predictor of poor OS (p<0.0001, multivariate)
  - Biomarker
    - Mutation/expression is biomarker/predictor for drug responses
    - Gene expression is beneficial for therapy with gemcitabine+platinum in NSCLC
- Our NLP based system accounts for a variety of ways such relationships can be expressed in biomedical text.

Other information

- Is it a patient or cell line study?
  - Patient
- How many patients?
  - 40 patients
- Race/demography of patients
  - Chinese
- What type of study?
  - Clinical trial
  - Retrospective study
  - in vitro/vivo etc.
- What type of document?
  - Meta-analysis
  - Review etc.

Suggestive sentences

Suggests the potential of gemcitabine+platinum being used to treat NSCLC patients with low RRM1 expression levels.

Reference sections

- Genomic anomaly - Drug response
  - Anomaly type: Expression
  - Gene: RRM1
  - Disease: NSCLC
  - Benefit: Implantation for NSCLC
  - Drug: gemcitabine, platinum

- Genomic anomaly - Drug response
  - Anomaly type: Expression
  - Gene: BRCA1
  - Disease: NSCLC
  - Benefit: Implantation for NSCLC
  - Drug: gemcitabine, platinum

- Evaluation of the system
- Improve entity detection (e.g., disease, drugs, genes)
- More comprehensive detection of outcomes, responses and impact
- Going beyond abstracts:
  - full-length articles
  - Conference proceedings
  - Clinical Trials


Funding

This effort is funded by the NIH BD2K Data Wrangling Awards – Grant Number: 1 U01 HG008390-01