Target Locked

Make the life of a genomic data analyst a little more enjoyable by building a tool that can complete real time searches, prioritizations, and digests of the academic literature needed to support the diagnosis of rare disease. Bring your best weapons, as this challenge requires both precision and accuracy.

Long Description

One of the most time-consuming processes in analyzing genomic data is the research required to determine if a finding should be classified as Pathogenic (symptom causing), Likely Pathogenic (likely symptom causing), Uncertain Significance (some evidence but not enough for a diagnosis), or Benign (not causing symptoms). This process can take an analyst days or weeks to do on a newly found genetic change. Databases of academic literature exist and are searchable by the gene name and/or specific genetic change found, but different databases have different information. Currently, analysts spend time google searching the genetic change to find those academic articles and use established databases of literature like NCBI. Sorting through hundreds or thousands of search results is often time consuming and is exhausting. This challenge asks teams to put on the role of an analyst going through this process and creating a tool that can curate search results from various sources into a single display. The tool will need to be robust and include both a reliable and accurate search with a confidence interval for each search result based on its relevance to search terms; and a clean and user-friendly UI by which analysts can sort, tag, and filter results. Other tools that need to be included are an export of an APA style reference for each search result, a feedback system that collects analyst's use of a search result and uses it to inform future searches, and both a global and user specific tracking of literature used to support a genetic finding. To complete the entire challenge, participants should consider creating a functional product that contains all three parts; UI/UX, Optimization, and Prioritization. However, participants may also choose to focus on only one or two of the sub-challenges.

1. UI/UX
   This is a tool for analysts looking at dozens of result reports each day, which requires finding relevant literature to back up a decision on the classification of the genetic change. Develop an easy-to-digest UI/UX that can handle the display of hundreds of academic articles in an easily searchable, scannable, and prioritized manner. The UI should include CSS/HTML/JS components that represent sorting, tagging, and filtering results, as well as a confidence interval for each search result based on its relevance to search terms. Other tools that need to be included are exporting of an APA style reference for each search result, UI components that represent a feedback system that collects analyst's use of a search result and uses it to inform future searches, and a UI that can display both a global/whole system and user specific tracking of literature used to support a genetic finding.

2. Optimization

Searching multiple databases online is both time consuming and tedious, especially when there are hundreds of results for a given search and those searches have significant overlap. For the sub-challenge, participants are to optimize a search that includes an indexing of results from PubMed, Google Scholar, BioArchive, and MedArchive. Use the provided search terms to demonstrate a system that can efficiently obtain results (under 2000ms total response time), remove overlaps, and identify terms within search results and tag appropriately with Gene, Disease, Mutation, and Species.

3. Prioritization

Sorting through hundreds of research articles increases the time to diagnosis of patients and is currently a very manual process. In order to make this process more efficient, participants are asked to create a system by which search results for academic literature from PubMed, Google Scholar, BioArchive, and MedArchive are prioritized for use in genomic sequence analysis. Prioritization should be keyed off of the following: 1- gene/variant/DNA change to disease association, 2- gene/variant/DNA change function, 3- gene/variant/DNA change in animal models, 4- date of publication, and 5- user selection/star rating.