A Web-based Data Management Infrastructure for Curation, Annotation and Dissemination of Biomarker Research results for the Early Detection of Cancer

Andrew Hart[1], Dan Crichton[1], Don Johnsey[2], Chris Mattmann[1], Christos Patriotis[2], Heather Kincaid[1], Sudhir Srivastava[2], Mark Thornquist[3]

[1]Jet Propulsion Laboratory, California Institute of Technology Pasadena, CA 91109, USA
[2]National Cancer Institute, Bethesda, MD 20892, USA
[3]Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

The Early Detection Research Network (EDRN) is an NCI program focused on the discovery and validation of biomarkers. Bioinformatics plays a critical role in defining, capturing and managing the biomarker information that is produced within the EDRN. In 2006, the EDRN began developing plans for capturing and managing information about cancer biomarkers as part of the information captured by informatics systems within the EDRN enterprise. The specific goal of capturing this information is to provide the research community access to information about biomarkers that have been studied along with their results. This includes access to annotations about the biomarkers studied, organized by organ-site, as well as links to the studies and publications.

The Biomarker Database (BMDB) is a web-based data management tool designed to allow a cancer researcher to quickly grasp the current state of EDRN (and non-EDRN) research pertaining to a particular biomarker. Its features include the ability to aggregate and present heterogeneous biomarker information that is captured from a number of sources including publications (available via PubMed), study information, as well as organ-specific information – including pre-computed and dynamically computed sensitivity, specificity and predictive values.

Included in the Biomarker Database is a curation interface which provides tools for managing and aggregating data from multiple sources into an integrated view that allows several of these sources to remain distributed, but related to a biomarker. The curator builds a representation for each biomarker by indicating connections and associations between organ sites, studies, related publications, and external resources. While the current Biomarker Database is deployed in a prototype, it is being extended to support a peer review process. The peer review will screen biomarkers to validate the quality of the annotated data prior to making it available for access. In addition, the database will support specific security measures to ensure there is restricted access to biomarker information until it released for public view.

Long term, the biomarker database will be fully integrated into the EDRN Knowledge Environment. This will provide scientific access to a variety of information, including biomarkers, allowing scientists to locate and access data across the EDRN enterprise from a single web-based portal.

Acknowledgement: This material is based upon work supported by the Jet Propulsion Laboratory, managed by the California Institute of Technology. Effort also supported by the National Cancer Institute's Early Detection Research Network (EDRN) program.