Summer 2015 Internship Project Descriptions

Melissa Haendel PhD & Nathan Dunn PhD. Work closely with biologists and programmers to integrate aspects of the JBrowse genomic browser (http://jbrowse.org/) with the Monarch Initiative (http://monarchinitiative.org). Monarch is a portal that uses computational reasoning to link phenotype to genetic information across species and models. Integration of JBrowse will allow users to view and browse relevant gene sets across multiple chromosomes, and visualize the relationships to phenotypes. Student should have a good understanding of Javascript and scripting. An understanding of bioinformatics and Perl will be helpful, but not necessary.

Stephen Wu, PhD, and Bill Hersh, MD. We’re exploring how best to search EHR data (especially text documents) to find groups of patients that fit a criteria. Develop and refine an IR research software platform with us, based on Elasticsearch/Lucene. Programmatically analyze large-scale EHR text and structured data. Skills: Some programming/software engineering coursework or experience.

Aaron Cohen, MD. Develop a general biomedical named entity recognition (NER) engine in Python using the Natural Language Processing Toolkit (NLTK) or other framework. This NER engine should be data driven, using a domain specific terminology or ontology, and automatically annotate text documents for desired concepts. There are many ways to approach this task, Alternative methods will be discussed and evaluated with the mentor, and an implementation and testing plan will be developed before the software is written.

Karen Eden, PhD - Seeking interested future graduate student in biomedical informatics. Assist with summarizing results from current evaluation study of women using a mammography decision aid (iPad app). Women with average risk for breast cancer are sent conflicting messages by providers about when to begin mammograms. This app is designed to help them understand the medical evidence and make an informed choice. Assist in updating the risk assessment in the breast screening app. Assist in drafting a manuscript on portfolio of women’s health decision aids. As time allows, assist with summarizing results of a biomedical informatics graduate program evaluation. The purpose will be to compare the various programs for curriculum required at the PhD level. Skills: scientific writing expertise, interpretation, medical literature searching, understanding of health risk assessment.

David Dorr, MD - TOPMED – The Transforming Outcomes for Patient-centered Medical home Evaluation and reDesign trial is in its evaluation stage. The trial is studying how primary health care reform can be enhanced with more focused use of incentives, HIT, and practice facilitation. Final data cleaning, analysis, and synthesis will be completed. Specific subprojects for interns under this project are the effect of achievement of High Value Elements on cost, patient satisfaction, and quality; use of information to improve outcomes for patients; and clinic HIT use and team performance on sustainability. Some familiarity with interpreting data is helpful.

David Dorr, MD - Big data to knowledge skills - Big data – the proliferation of massive amounts of clinical, observational, and genomic data – has the potential to revolutionize the way we discover new knowledge and deliver health care. However, significant issues remain – irreproducibility, data quality, and lack of metadata. New data scientists need skills to understand and meet these challenges. This internship would help develop use cases for big data issues by creating synthetic data sets around particular concepts and develop scripts to understand and mitigate the issues. Must be willing to learn R (statistical package).
**Lucia Carbone PhD** - Seeking interested graduate students in bioinformatics. Our project will involve analysis of next-generation sequencing data with existing software to identify insertions of retrotransposons. These are repeated sequences or “jumping genes” that have been colonizing primate genomes and might have a functional impact. We are particularly interested in differentiating insertions occurred in the germline (i.e. present in all cells of a given organism) from “somatic” insertions (i.e. occurred only in a subset of cells). Ultimately, we aim to study the association between environmental insults (i.e. alcohol use, high fat diet) and germline and somatic retrotransposition.

**Michael Chiang MD & Thomas Yackel MD** - Evaluation of EHR system implementation and workflow at an academic medical center. Ongoing projects involve analysis of outcome measures such as speed, efficiency, cost-effectiveness, and user satisfaction; as well as optimization of clinical workflow using EHR data, indoor positioning systems, and computer simulations. Skills: statistical analysis.

**Michael Chiang MD & Jayashree Kalpathy-Cramer PhD** - Computer-based image analysis for diagnosis of retinal disease. This would involve design and/or evaluation of diagnostic algorithms, and comparison to interpretation by human experts. Skills: knowledge of computer programming & database architecture.