Implementing PheWAS in i2b2

Huan Mo, MD, MS¹
Jacqueline Kirby¹, Zhiao Shi¹
Robert Carroll, PhD², Lisa Bastarache¹
Richard Kiefer³, Guoqian Jiang, MD, PhD⁵
Luke Rasmussen¹, Jennifer Pacheco⁴
William Thompson, PhD², Firas Wehbe, MD, PhD¹
Jyotishman Pathak, PhD², Joshua Denny, MD, MS¹

OBJECTIVE:
1. Transform and import PheWAS ontology (ICD-9 based) to i2b2 ontology cell.
2. Connect PheWAS R package to i2b2 CRC cell.
3. (Next step) Create PheWAS cell that runs on i2b2 framework. Seek/establish i2b2 solution for PheWAS genotype data.

BACKGROUND
Phenome-wide association studies (PheWAS) look for associations between many phenotypes and a single nucleotide polymorphism (SNP), or other predictors (e.g., use of a medication, disease subtypes). In a standard PheWAS, a phenotype is defined by the presence of certain ICD-9 codes in a patient’s electronic health record (EHR); a control for the phenotype is defined as the absence of these ICD-9 codes as well as other, more broadly related ICD-9 codes. In our previous efforts, we have grouped ICD-9 codes to ~1,600 PheWAS codes hierarchically (as a PheWAS ontology) according to the similarity of their underlying pathophysiological processes.

METHODS
PheWAS Ontology for i2b2: We used a KNIME workflow to transform the mapping table of PheWAS codes to an i2b2 metadata table (for the ontology cell). In metadata SELECT SQL query construction fields, we adopted an “IN” clause (instead of “LIKE”) with a list of ICD-9 codes that mapped to a PheWAS concept unit. No changes in the CRC cell or its database are needed.

PheWAS ontology for i2b2

Connect PheWAS to CRC cell: We developed an R package (i2b2-PheWAS) to generate phenotype files from CRC cells that can be used in Carroll’s 2014 PheWAS R package. For each phenotype, i2b2-PheWAS issues two XML queries (for cases and controls) to CRC cells, and transforms the response XML forms to a phenotype column in R. (A PhEMA effort.)

Availability of Software
PheWAS ontology for i2b2 can be downloaded from ProjectPhEMA.org. i2b2-PheWAS R package is available upon request, and will be publicly available in the near future.

FUTURE DIRECTIONS
1. We will test the i2b2-PheWAS package with i2b2 instances for production with real clinical data.
2. We will integrate i2b2-PheWAS and PheWAS R package into an i2b2 cell with a graphical user interface for “Analysis Tools” on the Web Client.
3. As an iPGx effort, we will seek or build a solution to allow i2b2 to manage genotype data (similar to PLINK), so that we can perform GWAS within the i2b2 framework.

REFERENCES