VANDERBILT UNIVERSITY **BIOMEDICAL INFORMATICS**

# Implementing PheWAS in i2b2


**Huan Mo, MD, MS[1]**

**Jacqueline Kirby[1], Zhiao Shi[1]**

**Robert Carroll, PhD[1], Lisa Bastarache[1]**

**Richard Kiefer[5], Guoqian Jiang, MD, PhD[5]**

**Luke Rasmussen[3], Jennifer Pacheco[4]**

**William Thompson, PhD[2], Firas Wehbe, MD, PhD[1]**

**Jyotishman Pathak, PhD[5], Joshua Denny, MD, MS[1]**

[1]Dept. Biomed. Inform., Vanderbilt Univ., TN;

[2]Cntr. Biomed. Res. Inform., NorthShore Univ. HS, IL;

[3]Dept. Preventive Med. & [4]Cntr Genet. Med., Feinberg SOM, Northwestern Univ., IL;

[5]Dept. Health Sci. Res., Mayo Clinic, MN


**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 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.

**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.


## FUNDING

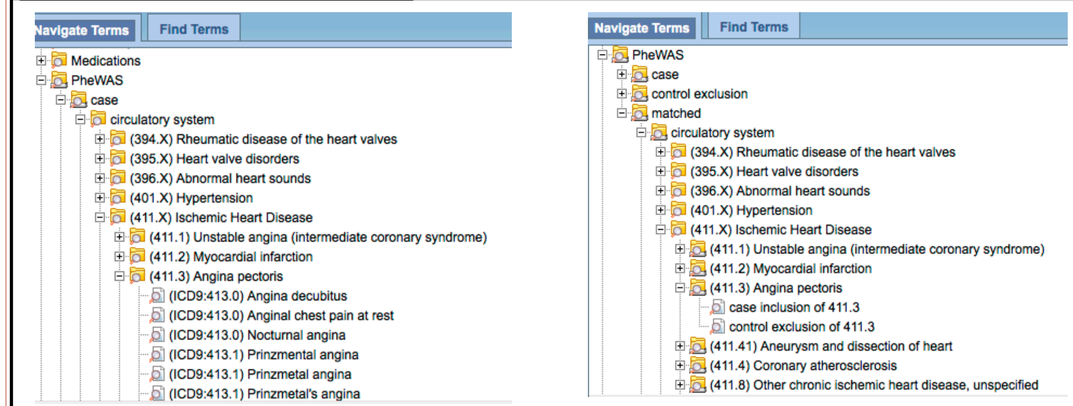
PheWAS: R01 LM010685; PhEMA: R01 GM105688; iPGx: R01 GM 103859


## REFERENCES


Denny JC, Bastarache L, Ritchie MD, et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nature Biotechnology*, *31*(12), 1102-10.

Carroll RJ, Bastarache L, & Denny JC (2014). R PheWAS: Data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics*, 30(16), 2375-6.


### PheWAS Ontology for i2b2




C_TOOLTIP =
PheWAS \ case \ circulatory system \ (411.X) Ischemic Heart Disease \ (411.3) Angina pectoris
C_COLUMNNAME = concept_cd; C_OPERATOR = IN
C_DIMCODE = 'ICD9:413','ICD9:413.0','ICD9:413.1','ICD9:413.9'

### i2b2-PheWAS services