

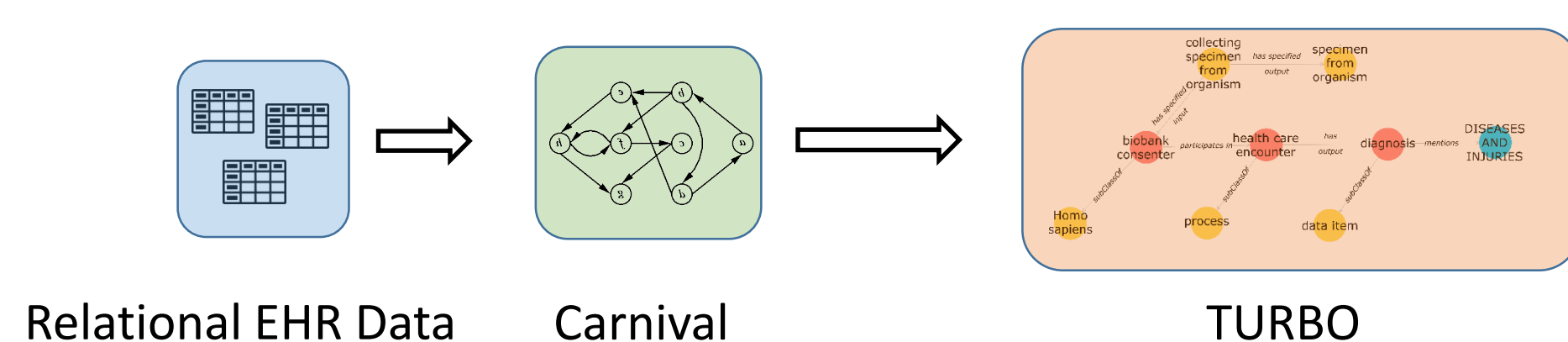
# TURBO: Transforming and Unifying Research with Biomedical Ontologies

Mark A Miller, MS<sup>1</sup>, Hayden Freedman, BA<sup>1</sup>, Heather Williams, MS<sup>1</sup>, David Birtwell, MS<sup>1</sup>, Chris Stoeckert, PhD<sup>1,2</sup>

<sup>1</sup> Institute for Biomedical Informatics, <sup>2</sup> Department of Genetics, Perelman School of Medicine, University of Pennsylvania

## Introduction to Penn TURBO

- Transforms relational EHR data into a semantic graph
  - Flexible. Not constrained by a schema.
  - Made possible by the Carnival property graph
- Unifies data from disparate sources
- Reifies data as being *about* real things and processes
  - Uses well-defined terms from OBO Foundry ontologies
  - Doesn't just add semantic "tags" to relational data
- Gracefully addresses missing or conflicting information.
  - The ground truth "humans have a biological sex" holds, even when the data about one patient's sex are unclear.
- Draws intelligent conclusions, based on
  - Knowledge embedded in the ontologies
  - Rules guided by domain experts



## TURBO includes a whole-exome sequencing cohort

- Demographics for 11,237 patients: Date of birth; height, weight and BMI; gender and/or biological sex; race
- Health care encounters in which the patients participate
  - Encounter dates
  - Diagnosis codes, with links to classes of diseases
  - Prescriptions, with links to active ingredients and pharmaceutical roles
- Biobank encounters
  - Encounter dates
  - Specimens used for whole-exome sequencing
  - Predicted loss of function (pLOF) calls on protein-coding genes, based on exome sequence data
  - biological processes linked to pLOF genes

TURBO enables **visual** exploration of the EHR's reality

## TURBO uses *OBO ontologies* to answer questions and requests like:

- Tabulate and plot demographic data, grouped at any level of semantic specificity
- What diagnosis codes have been assigned to the patients?
- What diseases are the codes indicative of? (*MonDO*)
- What drug products & active pharmaceutical ingredients have the patients been prescribed? (*The Drug Ontology*)
- What therapeutic roles do those drugs have? (*ChEBI*)
- List patients who have been prescribed clopidogrel, or any other P2Y<sub>12</sub> antagonist or platelet aggregation inhibitor
- Count or list patients with pLOF for a biomarker like CYP2C19, or any other xenobiotic catabolism gene (*GO*)
- List all pLOFs for a given patient

**OBO:** Open  
Biological and  
Biomedical  
Ontology Foundry



**MonDO:**  
Monarch  
Disease  
Ontology



**ChEBI** ontology:  
Chemicals of  
biological interest



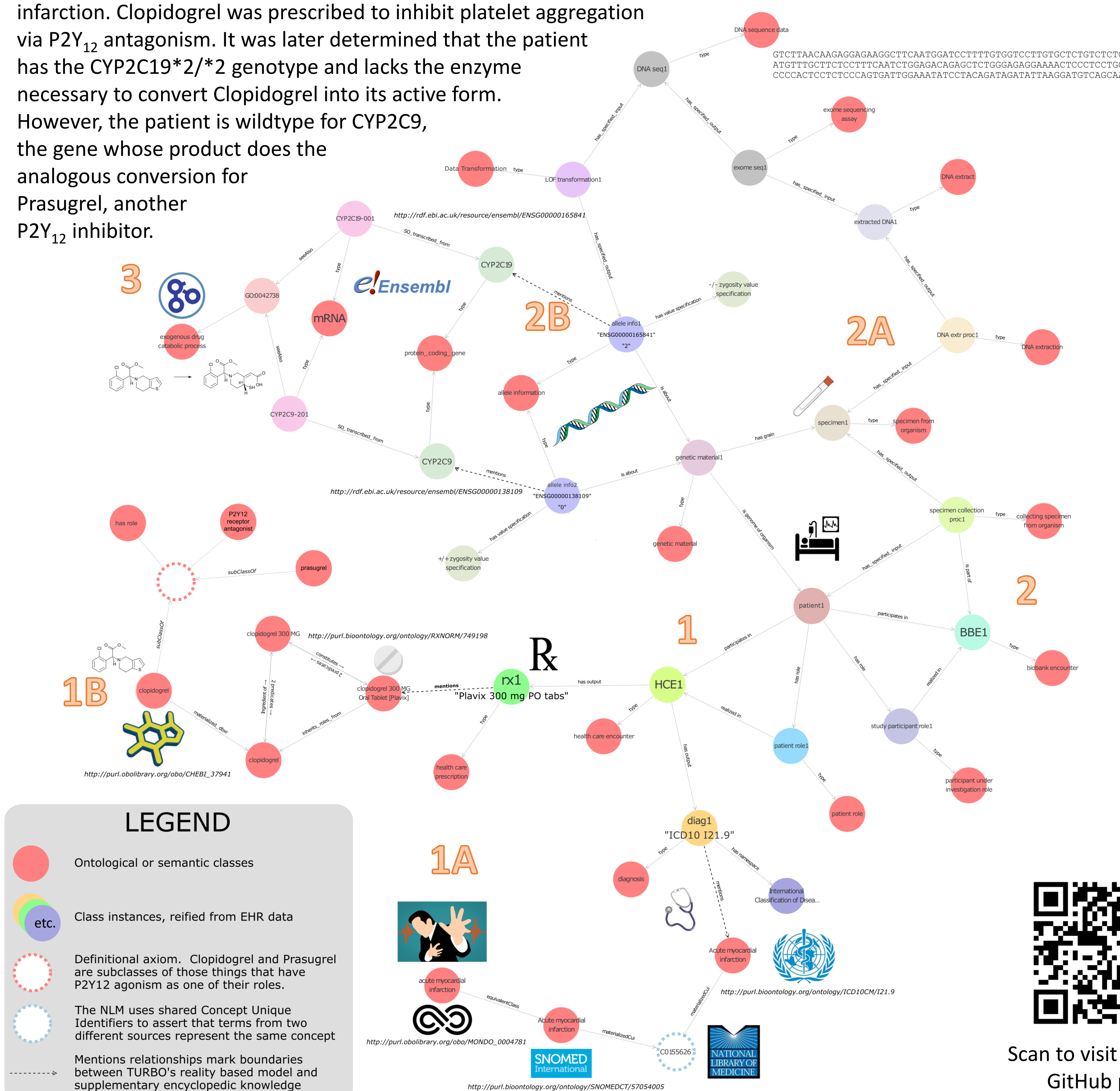
**GO:** Gene  
Ontology



## CYP2C19\*2/\*2 MI patient taking Clopidogrel

### Visualization of TURBO graph patterns with mock EHR data

**Clinical picture and therapeutic options:** Patient had a recent myocardial infarction. Clopidogrel was prescribed to inhibit platelet aggregation via P2Y<sub>12</sub> antagonism. It was later determined that the patient has the CYP2C19\*2/\*2 genotype and lacks the enzyme necessary to convert Clopidogrel into its active form. However, the patient is wildtype for CYP2C9, the gene whose product does the analogous conversion for Prasugrel, another P2Y<sub>12</sub> inhibitor.



Scan to visit TURBO's  
GitHub page

### Reporting diagnosis codes and semantically related diseases

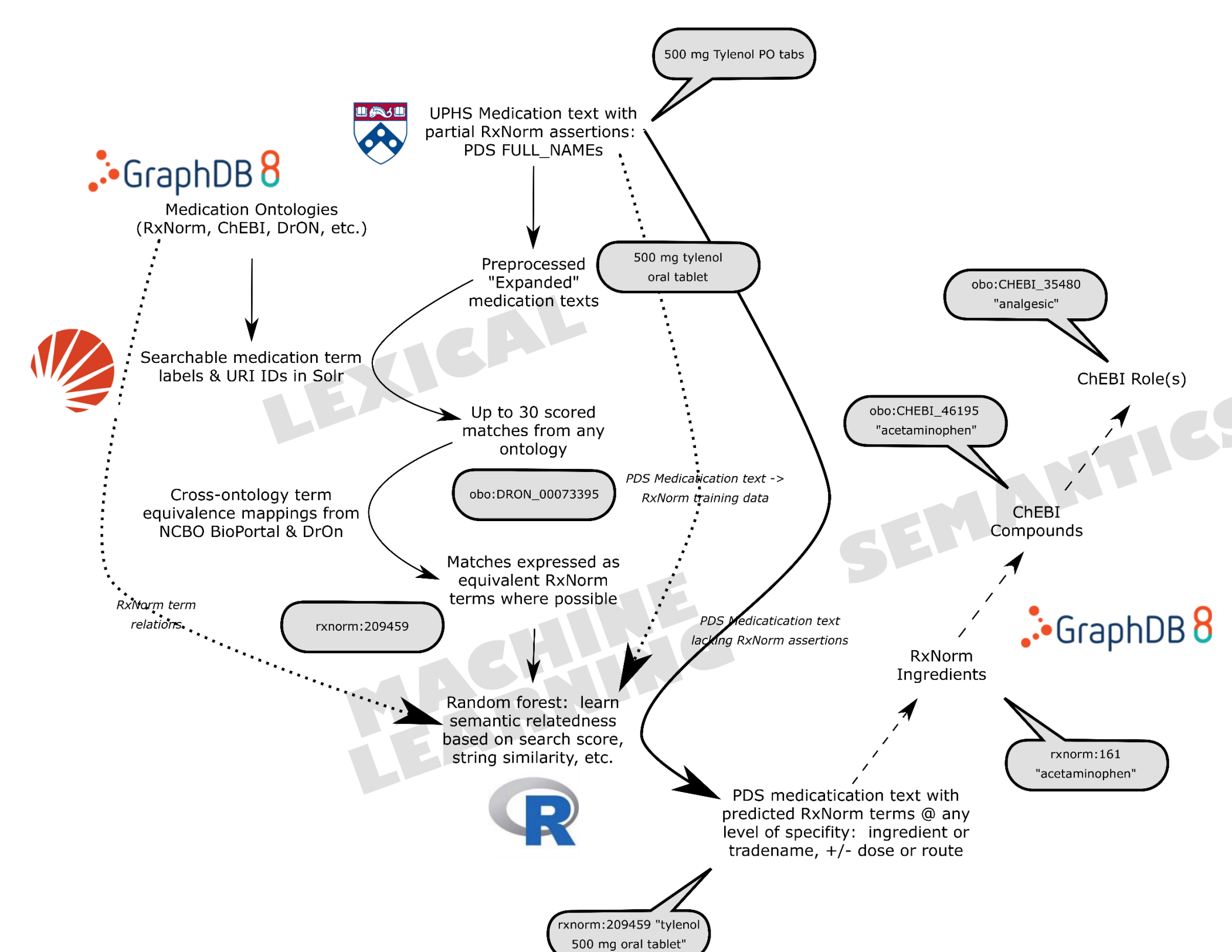
Frequently, TURBO can just rewrite data values into URL-like semantic terms, which become entry points for reality-based queries. For example, "ICD10 I21.9" → <http://purl.bioontology.org/ontology/ICD10CM/I21.9> "Acute myocardial infarction, unspecified"

### Defining disease pattern-based cohorts with diagnosis codes, prescribed drugs, etc.

Disease Pattern Project	Technical Collaborators	PI
Prostate cancer patients not yet in Cancer Registry	DAC	K. Maxwell
IBD patients who transitioned from CHOP to HUP	DAC	A. Rustgi
Candidates for autism spectrum screening	S. Lynch, IBI CIC	W. Guthrie/ R. Schultz

### Reporting PDS/EPIC drugs by pharmaceutical role

TURBO uses lexical and machine learning techniques to double the number of EPIC medications mapped to RxNORM



## Future TURBO development

- Web interface for Phenotype Storefront
- Improved disease patterns by modeling additional data types
  - Ontological representation of lab tests
  - Specialties and roles of prescribers and diagnosticians
- Generation of output for common data models, like OMOP
- Integrations with OBO-based linked data at other institutions

**TURBO is a demonstration project supported by the IBI.**

Support has also been provided by the Institute for Translational Medicine and Therapeutics, and the Penn Medicine BioBank.