

# Ontology-supported searches for the PhenX Toolkit: Development of a Smart Query Tool (SQT)

Josh Levy, Ying Qin, Huaqin Pan,  
Carol M. Hamilton  
June 25, 2013

*RTI International is a trade name of Research Triangle Institute*

**New NIH Common Data Element (CDE) Portal, more [here](#)**

Quick Start

QuickStart



Tutorial



Substance Abuse and Addiction

PhenX  
RISING

NEW!

Featured Measure

## Welcome to the PhenX Toolkit

The Toolkit provides *standard* measures related to complex diseases, phenotypic traits and environmental exposures. Use of PhenX measures facilitates combining data from a variety of studies, and makes it easy for investigators to expand a study design beyond the primary research focus. All Toolkit content is available to the public at no cost.

Information about the project is available at [www.phenx.org](http://www.phenx.org)

[More »](#)

### Please Read Toolkit Guidance

#### How to cite use of PhenX measures:

Measures incorporated in this study were selected from the PhenX Toolkit version April 29 2013, Ver 5.4. [More »](#)

#### How to cite the PhenX Toolkit:

Hamilton, et al. (2011) The PhenX Toolkit: Get the Most From Your Measures. *American Journal of Epidemiology*, 174(3), 253-60.

*Funding for PhenX and the PhenX Toolkit was provided by NHGRI 5U01HG004597 and 3U01HG004597-03S3.*

#### PhenX Toolkit Release Notes

Please take a quick survey to tell us how we can improve the Toolkit.

There are a total of 339 measures in the PhenX Toolkit.

#### Top 5 Measures - Release 5.4

1. Current Age
2. Gender
3. Ethnicity
4. Race
5. Weight

[View Top 20 Measures](#)

#### Top 5 Domains - Release 5.4

1. Demographics
2. Anthropometrics
3. Alcohol, Tobacco and Other Substances
4. Cardiovascular
5. Environmental Exposures



Browse



Search



My Toolkit



Registration

### Registration

You may [browse](#) the PhenX Toolkit, but to save your work, you need to [Register](#). See additional [benefits](#) for registered users.

Existing users may login:

User ID:

Password:

[Forgot](#)

[Forgot User ID](#)

[Help With Login](#)



PhenX Updates:  
[Subscribe](#) [Unsubscribe](#)

# PhenX Domains

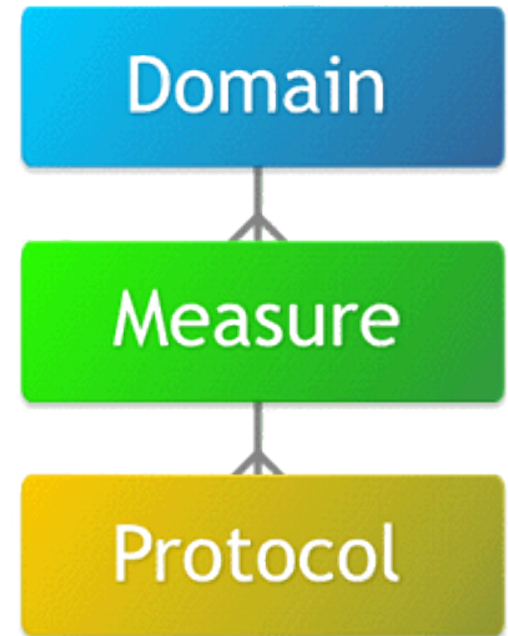
- Alcohol, Tobacco, and Other Substances
- Anthropometrics
- Cancer
- Cardiovascular
- Demographics
- Diabetes
- Environmental Exposures
- Gastrointestinal
- Infectious Diseases and Immunity
- Neurology
- Nutrition and Dietary Supplements
- Ocular
- Oral Health
- Physical Activity and Physical Fitness
- Psychiatric
- Psychosocial
- Reproductive Health
- Respiratory
- Skin, Bone, Muscle, and Joint
- Social Environments
- Speech and Hearing

# PhenX: consensus measures for Phenotypes and eXposures

- Goal: To identify and disseminate broadly relevant, easily implemented, standard measures of phenotypes and exposures suitable for use in genome-wide association studies (GWAS) and other human subjects research
- The PhenX project is led by RTI International and funded by the National Human Genome Research Institute (NHGRI)
  - RTI International Principal Investigator: Dr. Carol Hamilton
  - NHGRI Project Scientist: Dr. Erin Ramos

# PhenX Definitions

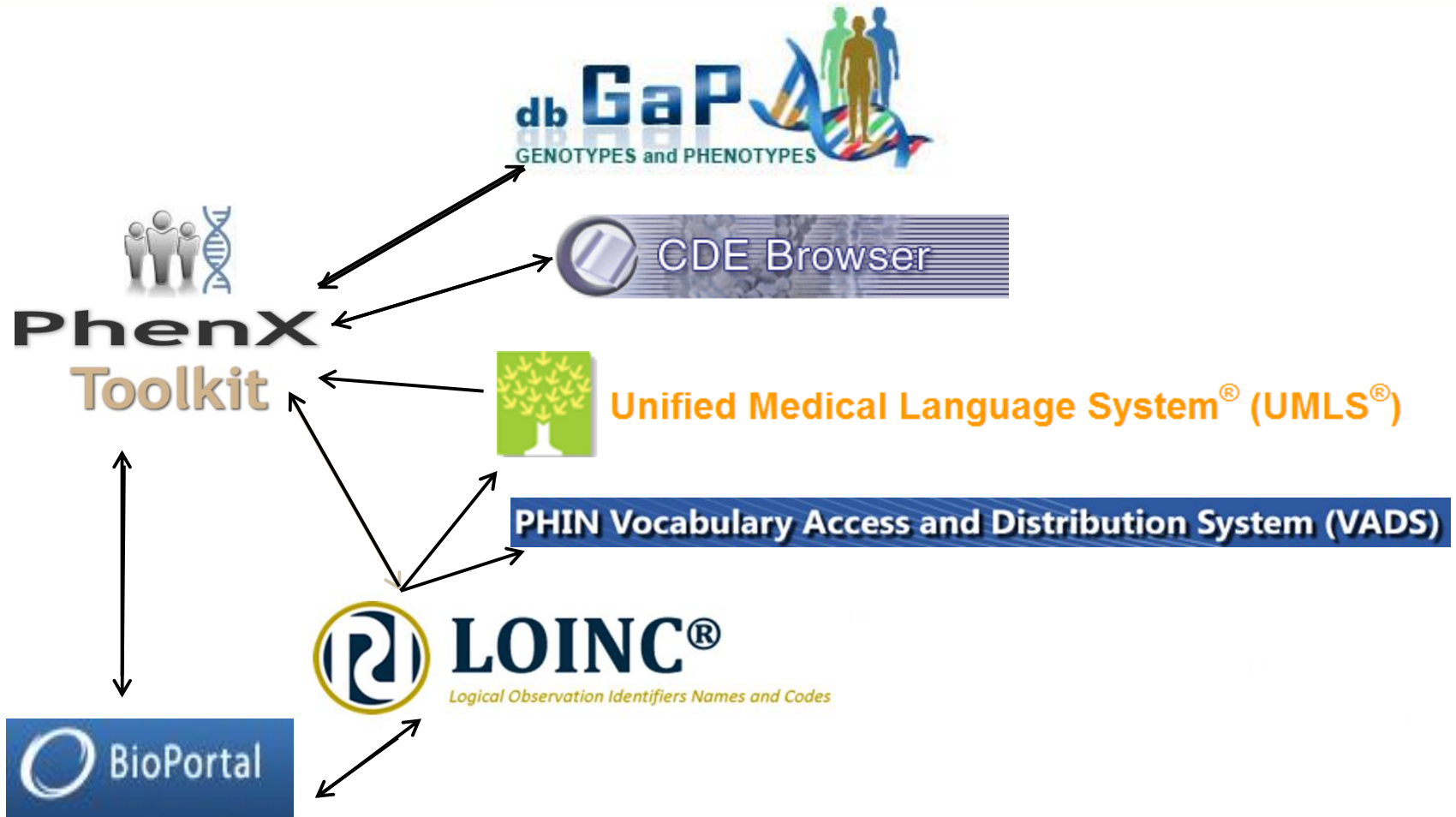
- **DOMAIN:** Topical area with a unifying theme  
Alcohol, Tobacco and Other Substances
- **MEASURE:** A certain characteristic of, or related to a study subject  
Tobacco - Nicotine Dependence
- **PROTOCOL:** Standard procedure recommended by a Working Group for collecting a PhenX measure  
Fagerstrom Test for Nicotine Dependence



# Challenges

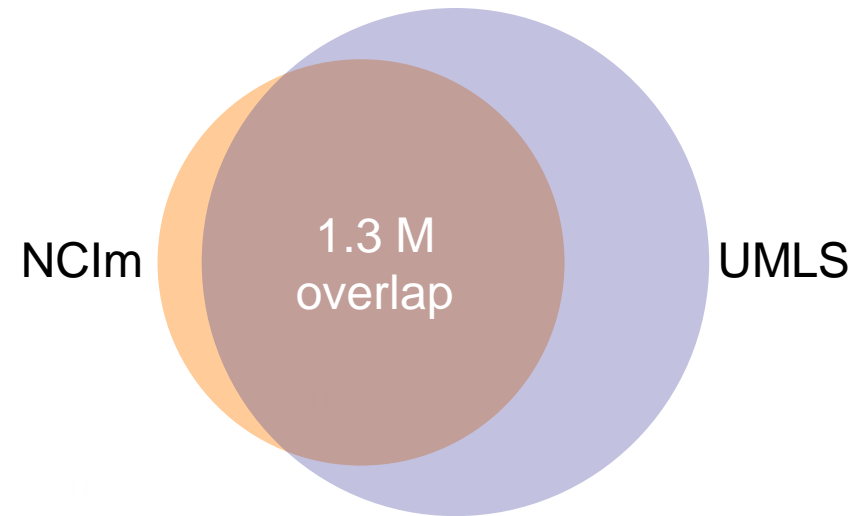
- What can we do to make it easier for Toolkit users to find measures of interest?
- What can we do to more fully support data interoperability?

# PhenX and Existing Standards



# The PhenX SQT - ontology backend

- NCI Metathesaurus (NCIm)
  - Excellent collection of aliases
  - Well-developed concept associations
  - Based on UMLS semantic network
    - *Semantic network is not an ontology*





# Working with NCIIm

- Eliminating cycles
  - UMLS and NCIIm **Semantic networks** contain cycles, which break a recursive traversal search
  - We use strategies to break these cycles
    - Visited nodes list
    - Visited paths lists
- Ontology size
  - Need to filter paths that do not end in a PhenX node
  - Recursive path building starting with PhenX measures

# Resources

- [www.phenx.org](http://www.phenx.org)
  - Provides general information about the PhenX project
  - Register to receive periodic updates via e-mail of the PhenX Newsletter and notification of new surveys
- [www.phenxtoolkit.org](http://www.phenxtoolkit.org)
  - Find PhenX measures for inclusion in your study
- [www.genome.gov/gwastudies/](http://www.genome.gov/gwastudies/)
  - A catalog of published Genome-Wide Association Studies (Hindorff et al. PNAS 2009)

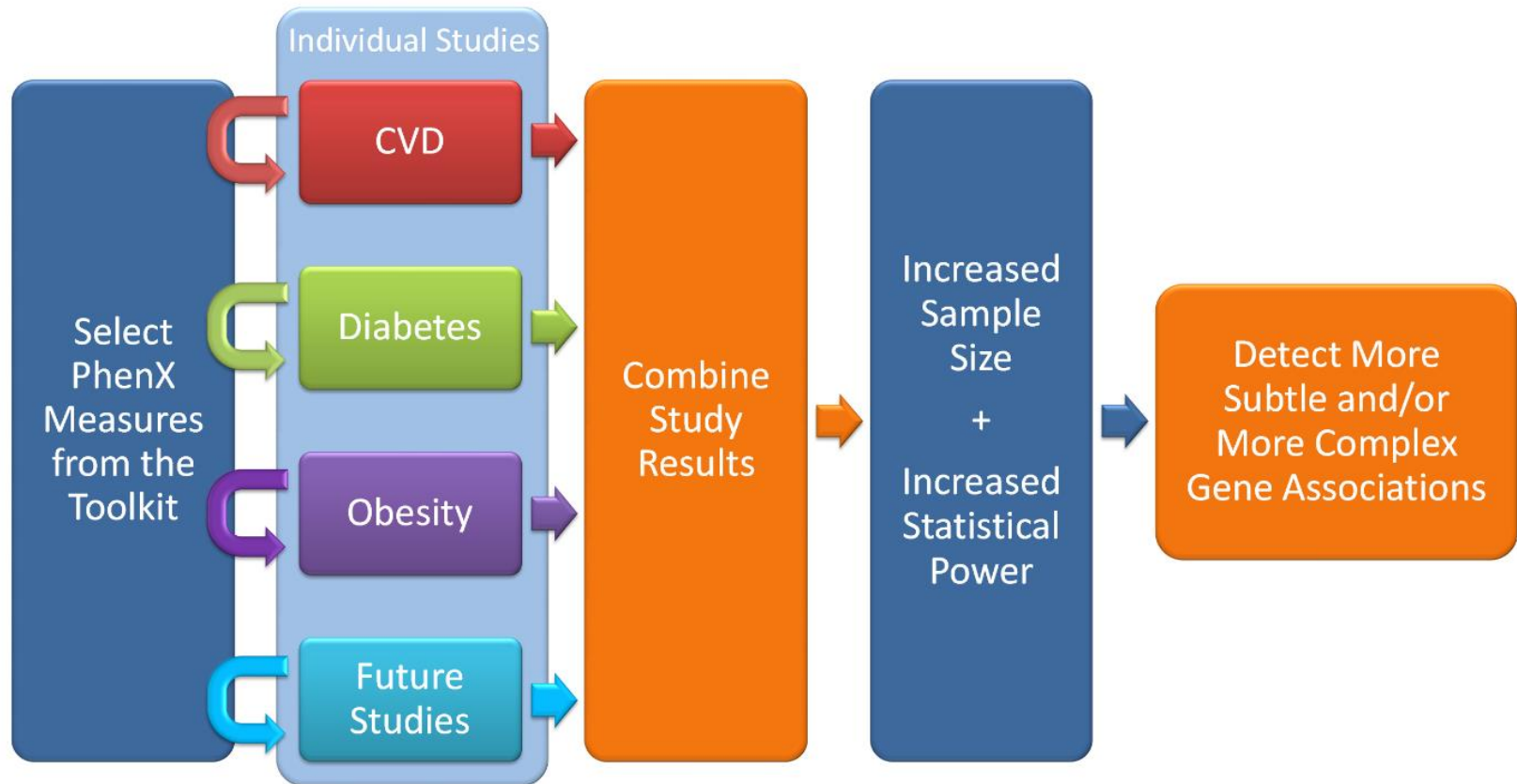
# Acknowledgements

- NHGRI
  - Erin Ramos (Project Scientist)
  - Heather Junkins (Project Analyst)
  - Teri Manolio
- SC Members (12)
  - Jonathan Haines, Chair
  - Bill Harlan, Vice-Chair
- WG Chairs / Members
- SSP (SAA Scientific Panel)
  - Kenneth Sher, Chair
  - Kevin Conway, NIDA
- IC Liaisons
- dbGaP
  - Kim Tryka
  - Mike Feolo
- LOINC
  - Clem McDonald
  - Daniel Vreeman
- eMERGE
  - Rex Chisolm, Chris Chute, Dan Masys, Cathy McCarthy
- NCBO BioPortal
  - Trish Whetzel, Mark Musen
- RTI Team
  - Carol M. Hamilton (PI)
  - Tabitha Hendershot (Co-Investigator)
  - Joe Pratt (Project Manager)
  - Dana Hancock
  - Jane Hammond
  - Wayne Huggins
  - Dean Jackman
  - Debbie Maiese
  - Destiney Nettles
  - Helen Pan
  - Lisa Strader
  - WG Managers
  - Toolkit team
  - Communications team
  - Logistics team

# Benefits of Using Standard Measures

- Enables cross-study analysis
  - Increase statistical confidence in results
  - Identify moderate associations and more complex interactions
- Facilitates validation of findings
  - Replicated in independent populations
- Increases the impact of individual studies
  - Many diseases and conditions share common risk factors

# Facilitating Cross-Study Analysis



# PhenX Collaborations



- National Center for Biotechnology Information (NCBI) database of Genotypes and Phenotypes (dbGaP)



- Logical Observation Identifiers Names and Codes (LOINC)



- Cancer Data Standards Registry and Repository (caDSR) Common Data Element (CDE) at cancer Biomedical Informatics Grid (caBIG)



- National Center for Biological Ontologies (NCBO) BioPortal

# Other NIH Measure Initiatives

- **NIH Toolbox for Assessment of Neurological and Behavioral Function (NIH Toolbox)**  
<http://www.nihtoolbox.org/Pages/default.aspx>
- **Patient-Reported Outcomes Measurement Information System (PROMIS®)**  
<http://www.nihpromis.org/?AspxAutoDetectCookieSupport=1>
- **Quality of Life in Neurological Disorders (Neuro-QOL)**  
<http://www.neuroqol.org/default.aspx>
- **Executive Abilities: Methods and Instruments for Neurobehavioral Evaluation and Research (EXAMINER)**  
<http://examiner.ucsf.edu/index.htm>
- **NINDS Common Data Elements**  
<http://www.commondataelements.ninds.nih.gov/#page=Default>
- **NIH Common Data Elements Portal**  
<http://cde.nih.gov/>

# PhenX SQT Summary

- Search functionality
  - Backend ontology
  - Custom mappings (keyword)
  - Full-text searches
- Browsing via tree
  - PhenX Research Domains
  - Conceptual group
- Flexibility in finding measures
  - Choice to browse or search in different ways accommodates scientists with different backgrounds