

Community tools for data harmonization and genetic variant interpretation

Erin M. Ramos, PhD MPH
Division of Genomic Medicine, NHGRI
9.28.17



Community tools for data harmonization and genetic variant interpretation

Erin M. Ramos, PhD MPH
Division of Genomic Medicine, NHGRI
9.28.17

NHGRI's Strategic Plan

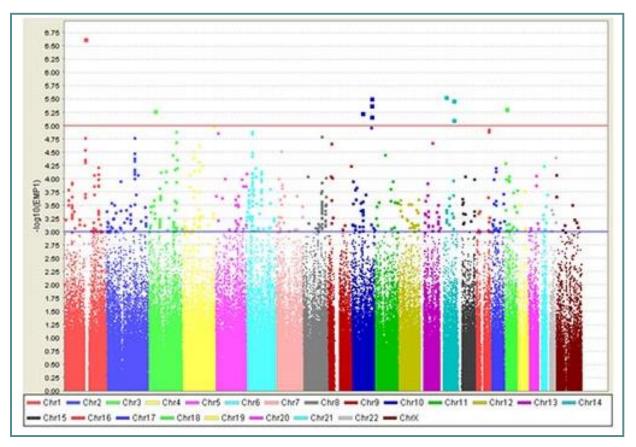


"A complete understanding of disease also requires...high-quality phenotypic data. Obtaining phenotypic data that are both thorough and accurate enough to be analyzed in conjunction with genomic and environmental data requires meticulous application of phenotyping methods, improved definitions of phenotypes, new technologies, and the consistent use of data standards (http://www.phenx.org)."

Green and Guyer, 2011

Standard Measures Needed – c2006

Type 2 Diabetes GWAS (>380K SNPs)



(www.broad.mit.edu/diabetes/scandinavs/type2.html)

- To detect loci with moderate effect size and G x G; G x E interactions, large sample sizes needed.
- Standard (common) measures facilitates replication of study findings.
- Potential for cross-study analysis and replication limited by lack of standardized measures

The PhenX Project

- Goal: Create online resource of standard (common) phenotypic & environmental exposure measures
- PI: Carol Hamilton, RTI International
- 500+ measures (< 620 protocols) addressing 24 domains
- Protocols, data element dictionaries, data collections worksheets



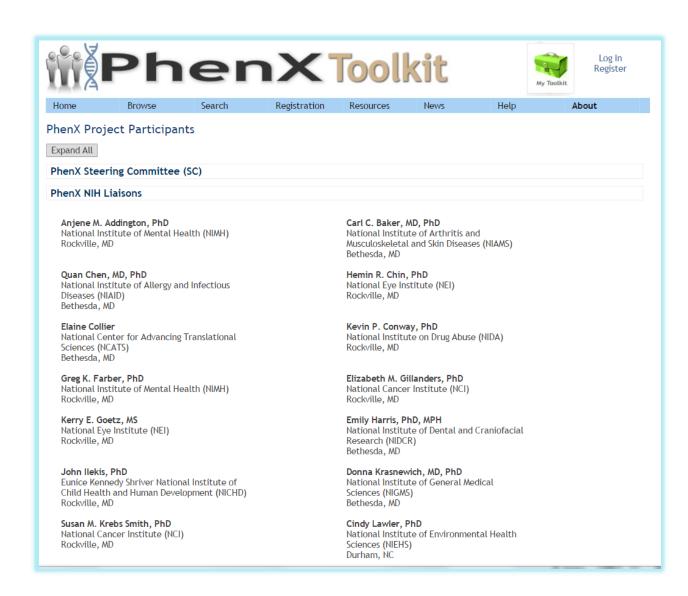
(www.phenxtoolkit.org)



 NHGRI: National Human Genome Research Institute

of Health

- NIDA: National Institute of Drug Abuse
- NIMH: National Institute of Mental Health
- NHLBI: National Heart, Lung, and Blood Institute
- OBSSR: Office of Behavioral and Social Sciences Research
- TRSP: NIH Office of Disease Prevention, Tobacco Regulatory Science Program



PhenX Definitions

 DOMAIN: Topical area with unifying theme (organ system, complex disease)

Substance Use

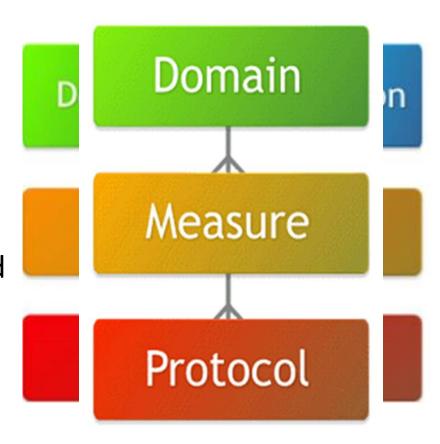
 MEASURE: Certain characteristic of, or relating to, a study subject

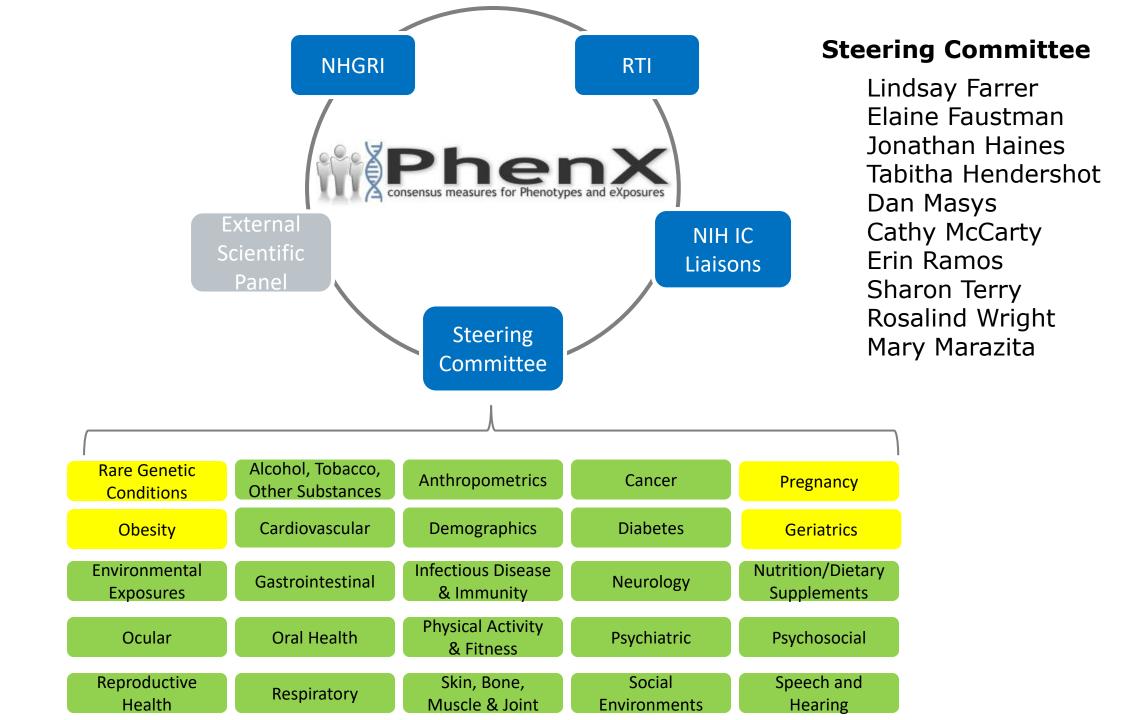
Nicotine Dependence

 PROTOCOL: Standard procedure recommended for collecting and recording a PhenX measure Fagerstrom Test

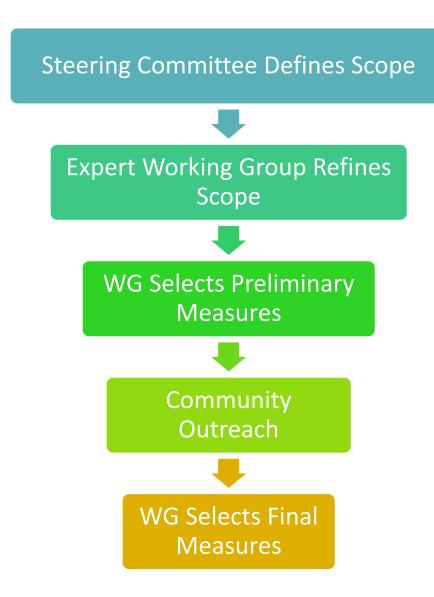
 COLLECTION: A set of measures with a shared characteristic, target population or topic. The measures may cut across research Domains.

Mental Health Research





Selecting PhenX Measures



Criteria for measures selection:

- Well-established
- Low burden to participant & investigator
- Reproducible
- Open-source software
- Useful to investigators who are not domain experts

Data collection mode

- Interviewer- or self-administered
- Bioassays
- Clinical exam, physical measurement
- GIS, Census data

Pregnancy Measures

Working Group

Siobhan Dolan, MD MPH (Chair) Albert Einstein College of Medicine

Russ Kirby, PhD Patrick Catalano, MD Case Western Reserve University

Ann Kinga Malinowski, MD Mt. Sinai Hospital

Mark Klebanoff, MD MPH Nationwide Children's Hospital

Hyagriv Simhan, MD MS McGee Research Insitute

Cande Ananth, PhD **Columbia University**

Erin Hines, PhD US EPA

U of South Florida

John Mulvihill, MD NHGRI

Rosalind Wright, MD Icahn School of Medicine

Lisa Kilpatrick **RTI** International

Mike Phillips, MS **RTI** International



American Journal of Obstetrics and Gynecology

Volume 217, Issue 3, September 2017, Pages 249-262



Special Report

Research standardization tools: pregnancy measures in the PhenX Toolkit

- Considered input from the pregnancy experts and the broader research community
- Recommend 15 protocols
- Complement the existing PhenX domains, including reproductive health, anthropometrics, demographics, and substance use.

Pregnancy Measures



Please take the PhenX Toolkit User Survey.

The PhenX (consensus measures for Phenotypes and eXposures) Toolkit is a catalog of recommended, standard measurements and environmental exposures for use in biomedical research. PhenX measures can be used to expand a strong the primary research focus. Use of PhenX measures facilitates cross-study analysis, potentially increasing the impact of individual studies. The PhenX Toolkit is a Web-based resource and is available for use at no cost. More >>











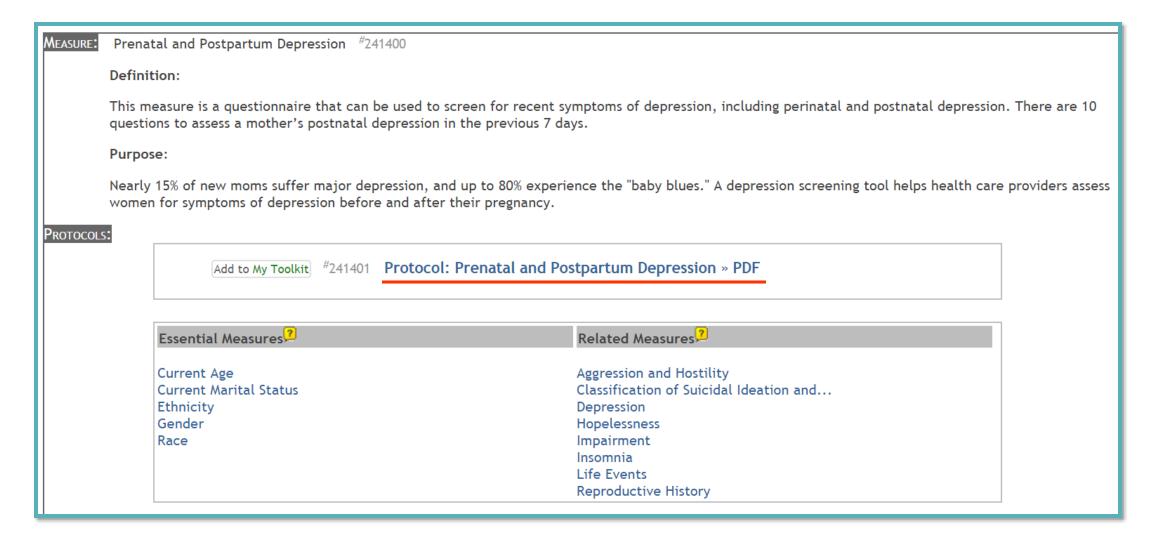






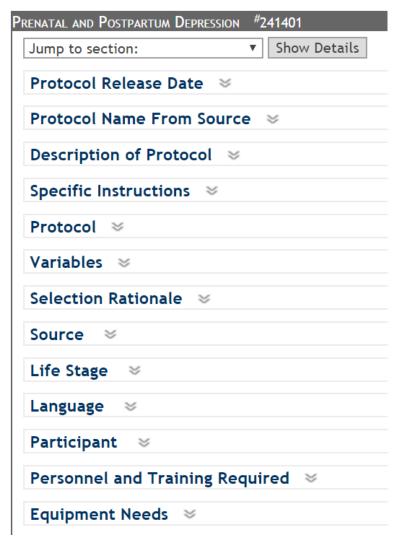
Add to My Toolkit	[#] 240100	Adequacy of Prenatal Care »
Add to My Toolkit	[#] 240200	Concentrations of Flame Retardants »
Add to My Toolkit	[#] 240300	Concentrations of Phenols and Parabens »
Add to My Toolkit	[#] 240400	Concentrations of Polychlorinated Biphenyls
Add to My Toolkit	[#] 240500	Concentrations of Trace Metals »
Add to My Toolkit	[#] 240600	Current Pregnancy Status »
Add to My Toolkit	[#] 240700	Difficulties in Pregnancy »
Add to My Toolkit	[#] 240800	Family History of Pregnancy Complications »
Add to My Toolkit	[#] 241500	Fetal Growth Assessment - Percentiles for U.S. Populations
Add to My Toolkit	[#] 240900	Gestational Age »
Add to My Toolkit	[#] 241000	Gestational Diabetes »
Add to My Toolkit	[#] 241100	Health and Wellness Before, During, and After Pregnancy »
Add to My Toolkit	[#] 241200	Mode of Conception »
Add to My Toolkit	[#] 241300	Mode of Delivery »
Add to My Toolkit	[#] 241400	Prenatal and Postpartum Depression »

Prenatal and Postpartum Depression



Source: Cox, J. et al (1987). Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry*, 150, 782-786.

Prenatal and Postpartum Depression





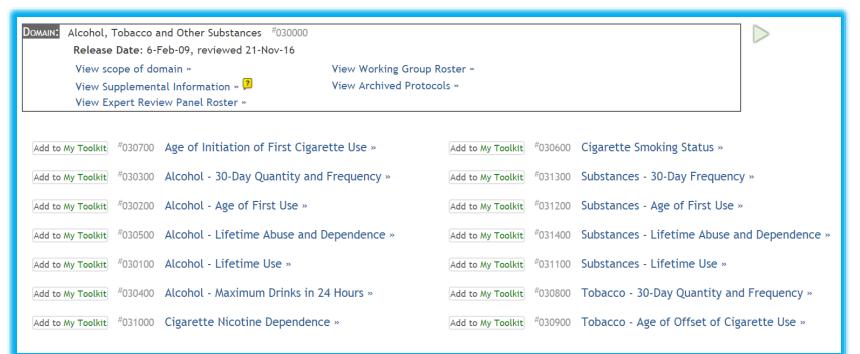
Source: Cox, J. et al (1987). Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry*, 150, 782-786.

Prenatal and Postpartum Depression

Download Report	Download Data Collection Worksheet	Download Dat	a Dictionary	Download REDCap Instrument Zip
Users who chose this Measure also chose: Add to My Toolkit Mode of Conception (Final Diabetes (Add to My Toolkit Health and Wellness Be Add to My Toolkit Gestational Age (Pregion of the Add to My Toolkit Gestation of the Add	Pregnancy) fore, During, and After Pregnancy (Pregnancy)			X
Empty My Toolkit Measure	Protocol	Requirements	Essential Measure	s <mark>.7</mark>
Prenatal and Postpartum Depression Remove Related Measures	Prenatal and Postpartum Depression Remove		Essential Measure Review additional Esser Current Age Current Marital St Ethnicity Gender Race	
<u>∧</u> E	ssential Measures are needed to interpret your dat	t <mark>a. Add all Essentia</mark> l M	easures to your Toolk	it »

Measures: 1 Protocols: 1

Substance Use Measures





 Collaborative data consensus effort from the NIH Tobacco Regulatory Science Program and the FDA Center for Tobacco Products (CTP)

- Social/Cognitive
- Biobehavioral
- Agent
- Vector

Interpersonal factors influencing product use (stress, motivation to quit) Product use, exposure, and health outcomes (use patterns, biomarkers) Assessing tobacco product (toxicology, warning labels, brands) Industry and retailer activity (packaging, price, advertising)

(Project lead = Kay Wanke, TRSP)

Social Environment Measures

DOMAIN: Social Environments #210000 Release Date: 8-Oct-10, reviewed 31-May-16							
	View scope of domain »		View Working Group Ros				
	• •	al Information » 🛂 ew Panel Roster »	View Archived Protocols	>>			
VICW L	KPCTC NCVI	ew Fallet Rostel					
Add to My Toolkit	#210200	Child-Reported Parental Educ	ation Attainment »	Add to My Toolkit	[#] 211500	Life Events »	
Add to My Toolkit	[#] 210100	Childhood Maltreatment »		Add to My Toolkit	[#] 210800	Neighborhood Collective Efficacy - Community	
Add to My Toolkit	[#] 210300	Discrimination »		Add to My Toolkit	[#] 211300	Neighborhood Concentrated Disadvantage »	
Add to My Toolkit	[#] 210400	Family Conflict »		Add to My Toolkit	[#] 210900	Neighborhood Safety »	
Add to My Toolkit	#210500	Family Control and Organizat	on »	Add to My Toolkit	[#] 211400	Race/Ethnic Residential Segregation »	
Add to My Toolkit	[#] 210600	Family Interpersonal Relation	ships »	Add to My Toolkit	[#] 211000	School Social Environment »	
Add to My Toolkit	[#] 210700	Healthy Food Environments »		Add to My Toolkit	[#] 211100	Social Networks »	
Add to My Toolkit	[#] 211200	Job Strain »					

Integration by Large Research Programs



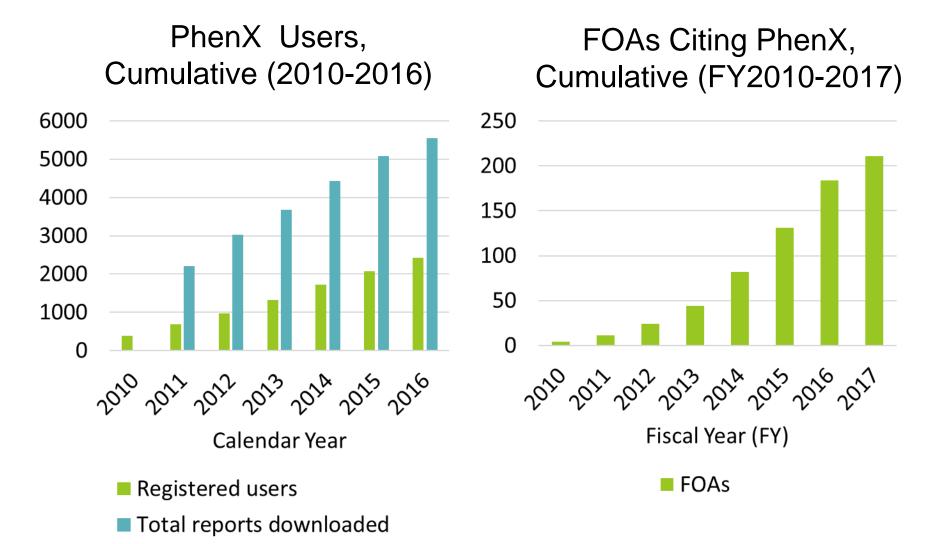




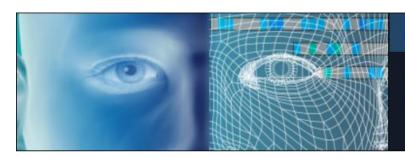


A collaboration between the NIH and FDA

PhenX Toolkit Use



Number of Visits: 1,192,834



dbGaP

The database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the data and results from studies that have investigated the interaction of genotype and phenotype in Humans.

PhenX Variable ID	PhenX Variable Name	Measure Name	Mapping level	dbGAP Study	dbGAP Variable
PX030301010000	PX030301_Alcohol_30Day_Frequency	Alcohol - 30- Day Quantity and Frequency	comparable	NHLBI Cleveland Family Study (CFS) Candidate Gene Association Resource (CARe)	phv00129665
PX030301010000	PX030301_Alcohol_30Day_Frequency	Alcohol - 30- Day Quantity and Frequency	comparable	NEIGHBOR Consortium Glaucoma GWAS	phv00072936
PX030301010000	PX030301_Alcohol_30Day_Frequency	Alcohol - 30- Day Quantity and Frequency	comparable	Genetic Multiple Sclerosis Associations - GeneMSA	phv00090478
PX030301010000	PX030301_Alcohol_30Day_Frequency	Alcohol - 30- Day Quantity and Frequency	comparable	The Vaginal Microbiome: Disease, Genetics and the Environment	phv00081806
PX030301010000	PX030301_Alcohol_30Day_Frequency	Alcohol - 30- Day Quantity and Frequency	comparable	GenADA/LONG/Imaging (Genetic Alzheimer's Disease Associations)	phv00090481
PX030301020000	PX030301_Alcohol_30Day_Quantity	Alcohol - 30- Day Quantity and Frequency	comparable	The Vaginal Microbiome: Disease, Genetics and the Environment	phv00129666

Acknowledgements

- NHGRI
 - Margaret Ginoza
 - Teri Manolio
- PhenX Steering Committee
 - Mary Marazita, Co-Chair
 - Cathy McCarty, Co-Chair
- WG Chairs & BMembers
- NIH Liaisons
- OBSSR lead Bill Riley/Erica Spotts
- NIDA lead Kevin Conway
- TRSP lead Kay Wanke
- NIMH lead Greg Farber
- NHLBI lead Ellen Werner
- NCBI's dbGaP Team

- RTI Team
 - Carol Hamilton (Principal Investigator)
 - Tabitha Hendershot (Co-Investigator)
 - Amanda Riley (Project Manager)
 - Darigg Brown
 - Wayne Huggins
 - Debbie Maiese
 - Lisa Jackson
 - Helen Pan
 - Mike Phillips
 - Toolkit team
 - Communications team
 - Logistics team





Community tools for data harmonization and genetic variant interpretation

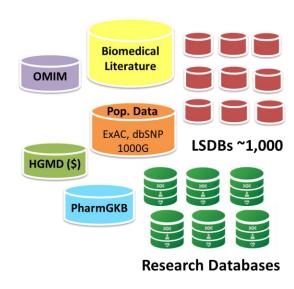
Erin M. Ramos, PhD MPH
Division of Genomic Medicine, NHGRI
9.28.17

The Problem





- Ability to detect DNA variants has greatly surpassed our ability to interpret their clinical impact
- >19,000 genes and > 88 million known variable sites in the human genome



Largely without standard assertions



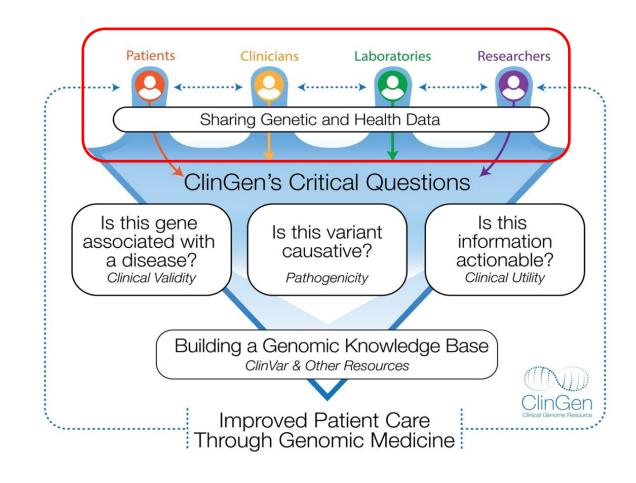
Clinical Testing Lab Databases

Largely absent from public domain

Improving our knowledge of the clinical impact of genomic variation requires a massive effort in data sharing and collaborative curation

The Clinical Genome Resource

- ClinGen is creating standard curation approaches and a central resource of clinically relevant genes and variants for use in medicine and research
- NIH-funded program launched Sept.
 2013
 - NHGRI, NCI, NICHD, All of Us
 - –3 Genomic Resource Grants (U41)& NCBI's ClinVar
 - -> 570 researchers & clinicians from 230 institutions









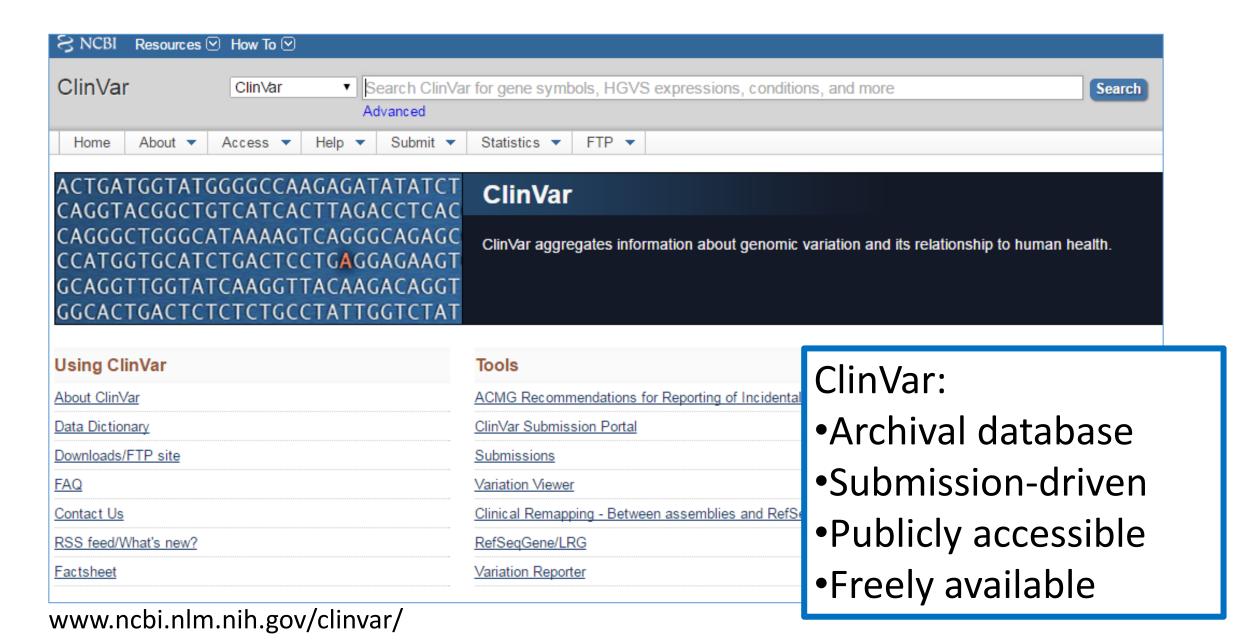




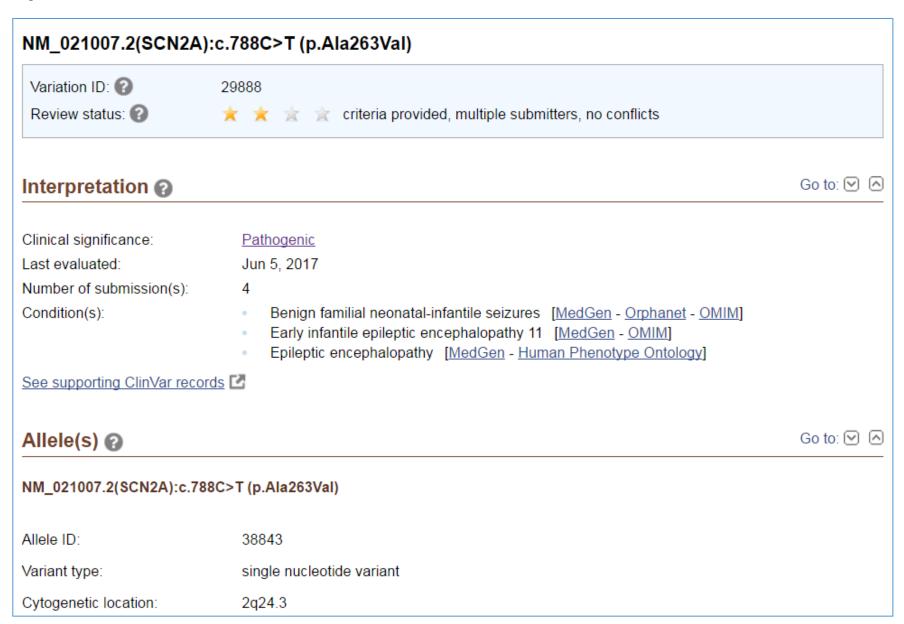




ClinVar



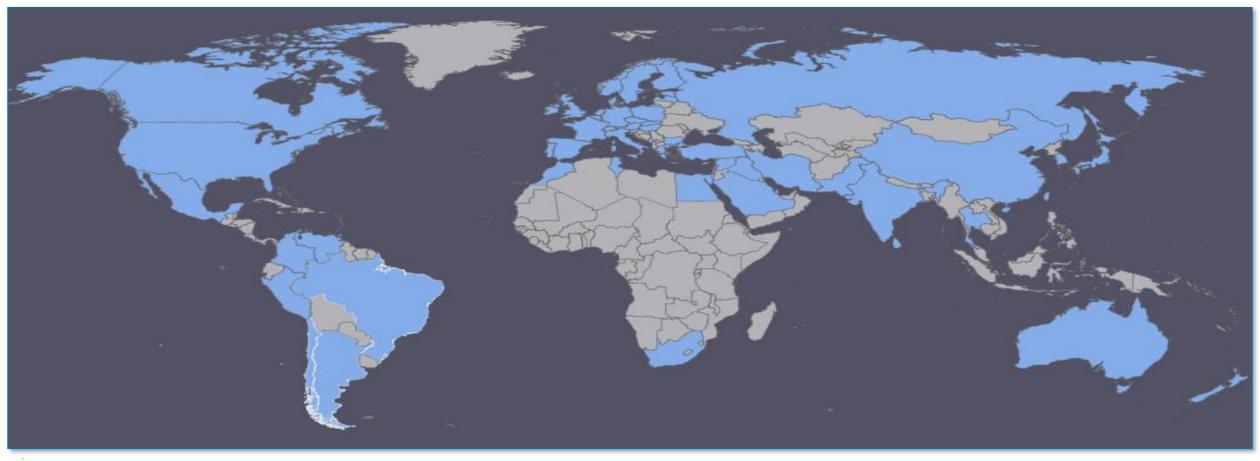
Example ClinVar Record



Example ClinVar Record

Clinical significance (Last evaluated)	Review status (Assertion method)	Collection method	Condition(s) (Mode of inheritance)	Origin	Citations	Submitter - Study name
Pathogenic (Aug 6, 2013)	criteria provided, single submitter • ACMG Guidelines, 2007	clinical testing	Benign familial neonatal-infantile seizures (Autosomal dominant inheritance) [MedGen Orphanet OMIM]	germline		Genetic Services Laboratory, University of Chicago
Pathogenic (Jun 5, 2017)	criteria provided, single submitter GeneDx Variant Classification (06012015)	clinical testing	not provided [MedGen]	germline		GeneDx
Pathogenic (Nov 16, 2016)	criteria provided, single submitter • ACMG Guidelines, 2015 • ACMG Guidelines, 2015	clinical testing	Epileptic encephalopathy [MedGen Human Phenotype Ontology]	de novo		Neurogenetics Laboratory - MEYER,AOU Meyer

801 ClinVar Submitters from 60 Countries





August 2017

- Submission from clinical testing labs, healthcare providers, patient registries, researchers, locus-specific databases, expert panels and professional societies
- 328,851 unique variants

ClinGen Curation Standards



- Gene-Disease Validity
 - Can variation in this gene cause disease?



- Dosage Sensitivity
 - Is haploinsufficiency or triplosensitivity an established disease mechanism for this gene?



- Variant Pathogenicity
 - Which changes in this gene cause disease?



- Clinical Actionability
 - Are there actions that could be taken to improve outcomes for patients with this genetic risk?

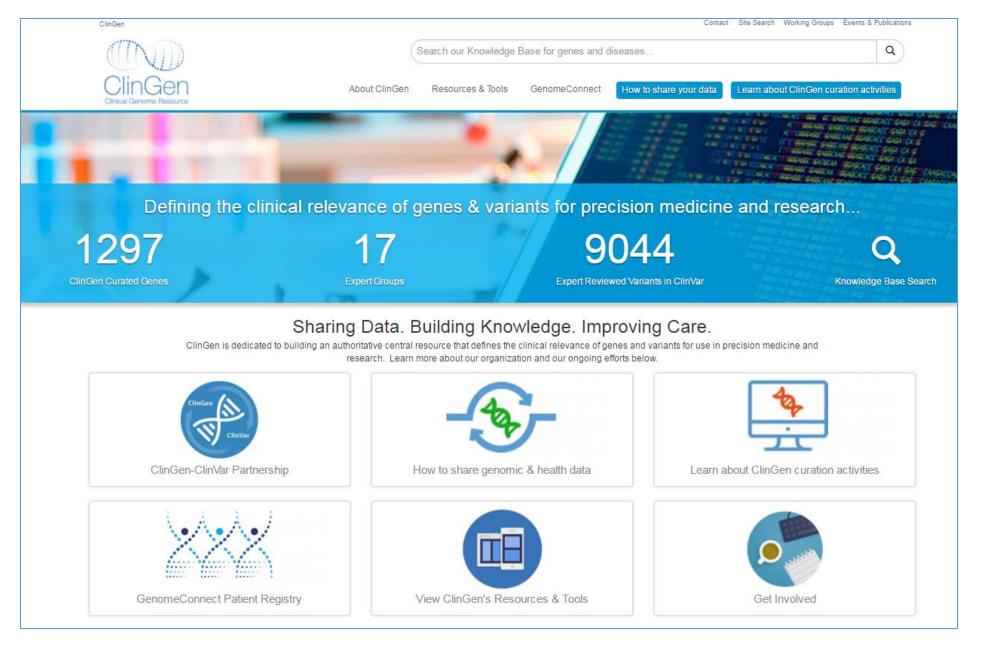
Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource

Marina DiStefano,⁴ Selina S. Dwight,⁸ Jenny Goldstein,¹ Rajarshi Ghosh,⁹ Bryce A. Seifert,¹ Tam P. Sneddon,⁸ Matt W. Wright,⁸ Laura V. Milko,¹ J. Michael Cherry,⁸ Monica A. Giovanni,³ Michael F. Murray,³ Julianne M. O'Daniel,¹ Erin M. Ramos,¹⁰ Avni B. Santani,^{11,12} Alan F. Scott,¹³ Sharon E. Plon,⁹ Heidi L. Rehm,^{4,5,6,7} Christa L. Martin,^{2,3,*} and Jonathan S. Berg^{1,*}

Evidence Type		vidence Type	Case Information		Sugg Points		Points Given	Max Score	
					Default	Range	Given	Score	
Case-Level Data¹ Variant Evidence		Autosomal	Variant is de novo ³			2	0-3		12
	nce	Dominant OR X-Linked	Proband with predicted or proven null variant ⁴			1.5	0-2		10
	Evide	Disorder ²	Proband with other v some evidence of			0.5	0-1.5		7
	/ariant	Autosomal Recessive	Two variants in <i>trans</i> and at least one de novo ³ or a predicted/proven null variant ⁴			2	0-3		- 12
			Two variants (not predicted/proven null) with some evidence of gene impact ⁵ in trans			1			
	•			e s 3	3	5			
		Segregation	Evidence of segregation in one	Sco	2	4	0-7		7
	Evidence		or more families ⁶	0.5	1.5	3	- 0-7		
					1	1.5			
trol	Case-Control Study Type ⁸ Single Variant Analysis ^{8a} Aggregate Variant		Case-Control Quality Criteria9		Suggested Points/Study		Points Given	Max Score	
e-Con Data ⁷			Variant Detection Methodology ^{9a} Power ^{9b}		0-6			-	
Cas	Ag	gregate Variant Analysis ^{8b}	 Bias and Confounding Factors^{9c} Statistical Significance^{9d} 			0-6			12
TOTAL ALLOWABLE POINTS for Genetic Evidence							12		

Gene-Disease Validty

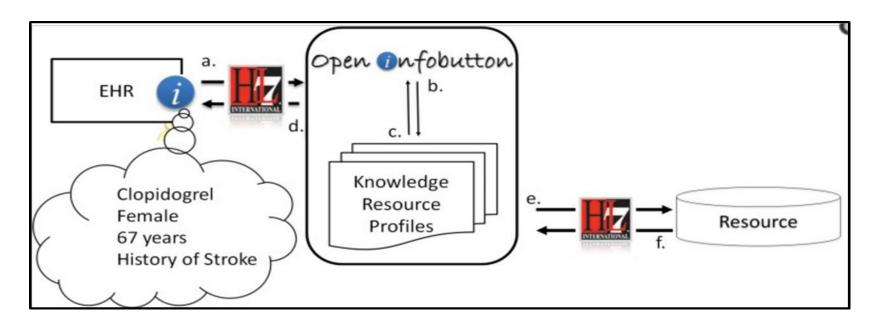
Clinical Valid	ity Classification Summaries			
HGNC Gene Symbol	Disease curated	Orphanet ID	OMIM ID	Clinical Validity Classification
ď	Hypertrophic cardiomyopathy	ORPHA217569 ♂	608751 ්	■ Definitive - 10/04/2016
ACTA2 ♂	Familial thoracic aortic aneurysm and aortic dissection	ORPHA91387 ♂	102620 ಆ	■ Definitive - 09/27/2016
ADCY1 &	Nonsyndromic sensorineural hearing loss	ORPHA90636 ♂	610154 ଫ	Limited - 05/10/2017
ADGRV1 &	Usher syndrome type 2	ORPHA231178 &	605472 ්	■ Definitive - 02/15/2017
AGTR2 ♂	X-linked non-syndromic intellectual disability	ORPHA777 &	N/A	Contradictory (disputed) - 11/16/2016
AKAP9 ♂	Long QT syndrome	ORPHA101016 &	611820 ී	Limited - 12/15/2016
ALMS1 &	Alstrom syndrome	ORPHA64 앱	203800 ♂	■ Definitive - 02/10/2017
ARSD ♂	Chondrodysplasia punctata	N/A	N/A	No Reported Evidence - 11/15/2016
ATF6 ♂	Achromatopsia	ORPHA49382 ♂	616517 ଫ	Strong - 11/16/2016
AXIN2 &	Hereditary colorectal cancer	ORPHA:443909 ♂	604025 ♂	Moderate - 06/08/2017
BAG3 &	Myofibrillar Myopathy	ORPHA593 &	612954 &	■ Definitive - 12/18/2016
BAP1 &	Hereditary pheochromocytoma-paraganglioma	29072 ਫ	603089 &	Limited - 07/22/2017
BARD1 &	Colorectal cancer	ORPHA443909 ♂	601593 ජ	Limited - 06/08/2017



(www.clinicalgenome.org)

EHR HL7 Readiness

- HL7 Compliance
 - Bring all ClinGen resources into HL7 compliance
 - Work with one resource, GTR, to move it to HL7 compliance
- Create a library of indexed clinical genomic resources using OpenInfobutton's Librarian Infobutton Tailoring Environment



Get Involved!

- Incorporate ClinVar into your daily practice and provide feedback on it's utility
- Submit data to ClinVar
- Support laboratories that make their data publicly available
- Join the monthly ClinVar Community Call
- Nominate topics for Clinical Actionability or Gene-Disease Validity curation
- Volunteer as a ClinGen curator
- Apply to be an Expert Panel in ClinVar
- Participate in continuing education activities involving variant reassessment

www.clinicalgenome.org/about/get-involved/

Contact us:

clingen@clinicalgenome.org
@ClinGenResource

Childen Steering Committee						
Jonathan Berg, UNC	Aleks Milosavljevic, Baylor	Brandi Kattman, NCBI				
Carlos Bustamante, Stanford	Kelly Ormond, Stanford	Lisa Brooks, NHGRI				
James Evans, UNC	Sharon Plon, Baylor	Andy Freedman, NCI				
Andy Faucett, Geisinger	Heidi Rehm, Harvard	Danuta Krotoski , NICHD				
Katrina Goddard, Kaiser Permanente	Michael Watson, ACMG	Melissa Landrum, NCBI				

Program Coordinators Danielle Azzariti, Erin Currey, Robert Fullem, Miranda Hallquist, Jenny Goldstein, Kristy Lee, Lisa Kurtz, Laura Milko, Jules Savatt, Meredith Weaver

Actionability WG

Jim Evans, Katrina Goddard

Biocurators WG

Jenny Goldstein

Consent and Disclosure

Recommendations WG

Andy Faucett, Kelly Ormond

Data Model WG

Larry Babb, Chris Bizon

Dosage Sensitivity WG

Erica Andersen, Erik Thorland

Education, WG

Danielle Azzariti, Erin Rooney Riggs

ClinGen Steering Committee

David Ledbetter, Geisinger Kirk Wilhelmsen, UNC Christa Lese Martin, Geisinger Marc Williams, Geisinger

Clinical Domain Oversight Committee

Jonathan Berg, Sharon Plon, Heidi Rehm

Hereditary Cancer: Ken Offit, Sharon

Plon

Somatic Cancer: Shashi Kulkarni,

Subha Madhavan

Cardiovascular: Birgit Funke, Ray

Hershberger

Inborn Errors of Metabolism: Rong

Mao, Robert Steiner, Bill Craigen

Pediatric Neurology: Michael Friez, Heather Mefford, Scott Myers

Hearing Loss: Sami Amr, Ahmad Abou Tayoun, Heidi Rehm

ClinGen Working Groups (WG) and WG Chairs

Erin Ramos, NHGRI

Electronic Health Record WG

Marc Williams

Gene Curation WG

Jonathan Berg, Christa Martin

Genomic Variant WG

Christa Martin, Sharon Plon, Heidi Rehm

Informatics WG

Carlos Bustamante

Sequence Variant Interpretation WG

Leslie Biesecker, Steven Harrison

Structural Variant WG

Swaroop Aradhya, Daniel Pineda-Alvarez

THANK YOU!

Erin M. Ramos, PhD MPH

Program Director, Division of Genomic Medicine National Human Genome Research Institute

Tel: (301) 480-3288

RAMOSER@MAIL.NIH.GOV









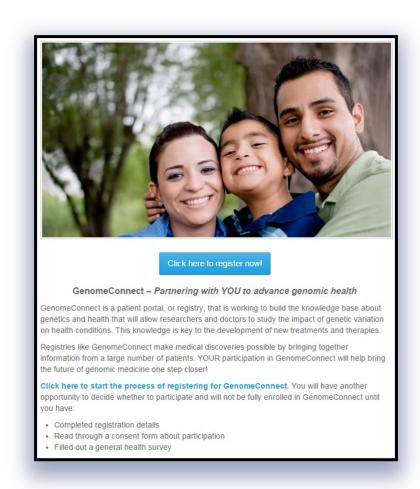
EXTRA SLIDES



Patients and Patient Advocacy Groups

A web-based patient registry that aims to:

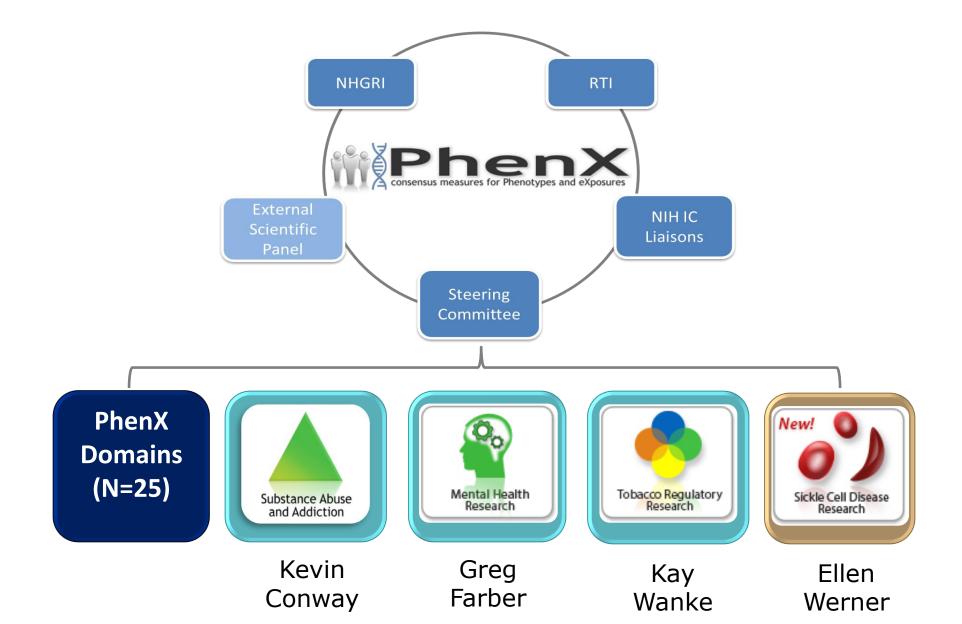
- Provide a mechanism for patients to securely share their genetic and health information
- Support patients as a valuable partner in advancing genomic medicine
- Enable patients to connect with other individuals, genetic testing laboratories healthcare providers, and researchers



www.genomeconnect.org

(Slide modified from Juliann Savatt, MS, LGC, Geisinger)

NIH Collaborations: PhenX Evolution



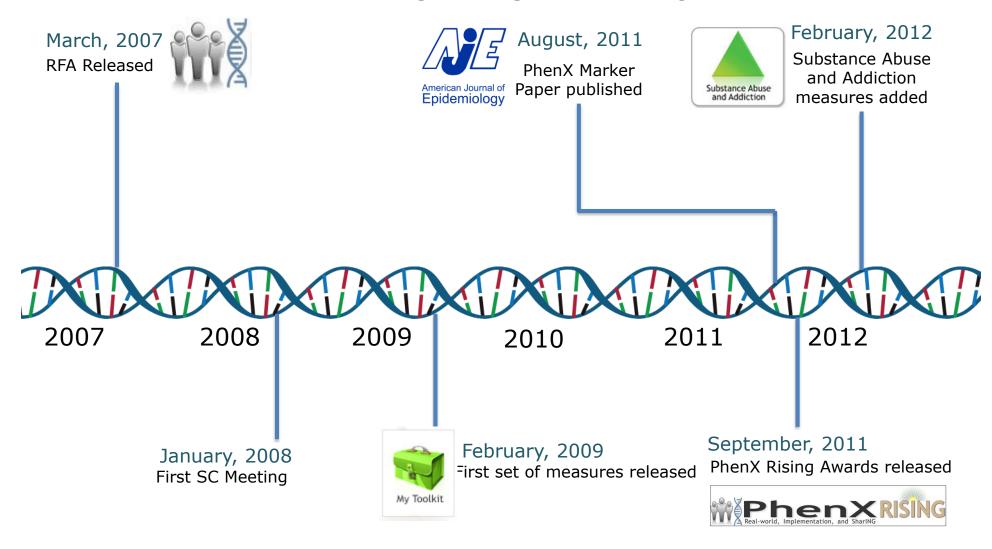
2017-2020 Goals

- Expand PhenX Toolkit research domains
 - Cancer, Environmental Exposures, Pediatrics, Social Determinants of Health
- Link PhenX measures to established ontologies
- Provide new tools that investigators need
 - Updated interface
 - Novel tools for comparing measures and variables.

Social Determinants of Health

- Possible Area of Expansion
 - Some coverage by existing measures, but several gaps identified
- Topic areas include
 - Economic Stability
 - Education
 - Social and Community Context
 - Neighborhood and Built Environment
 - Health and Health Care
 - Protective Factors
 - Risk Factors
 - Outcomes

10th Anniversary: Key Accomplishments





10th Anniversary: Key Accomplishments

