ASCPT 2018

ANNUAL MEETING

MARCH 21 - 24, 2018 · HILTON ORLANDO · ORLANDO, FL

Knowledge Representation Standards for Translational Genomics

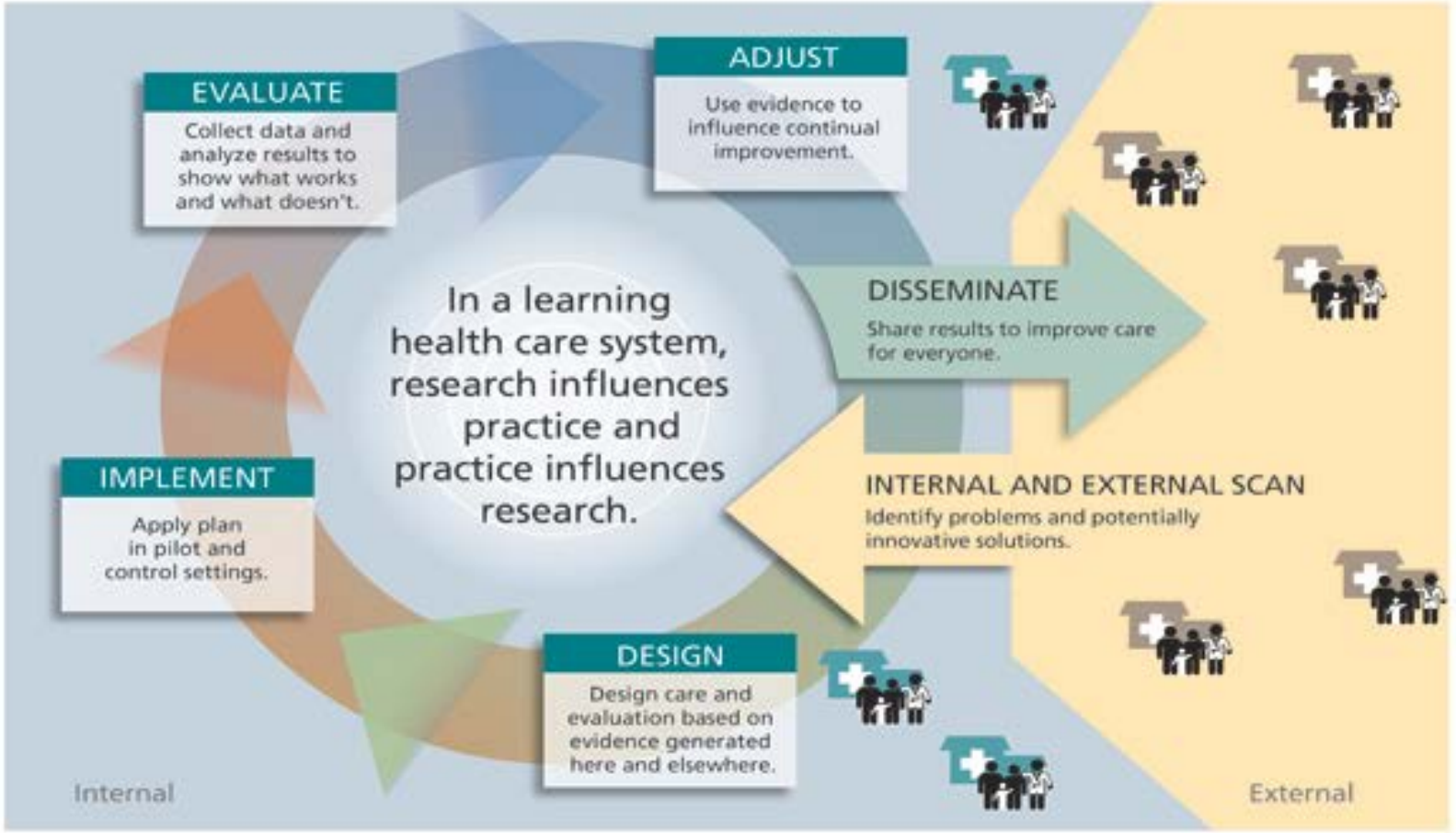
Robert R. Freimuth, Ph.D.
Center for Individualized Medicine
Center for Translational Informatics and Knowledge Management
Department of Biostatistics and Informatics
Mayo Clinic

March 23, 2018

ASCPT 2018 ANNUAL MEETING
MARCH 20-24, 2018 - ORLANDO, FL

Disclaimer

This will *not* be comprehensive!



http://www.nationalacademies.org/hmd/Activities/Quality/LearningHealthCare.aspx

Clinical Pharmacogenetics Implementation Consortium Guidelines for *HLA-B* Genotype and Abacavir Dosing: 2014 Update

MA Martin[1], JM Hoffman[2], RR Freimuth[3], TE Klein[4], BJ Dong[5], M Pirmohamed[6], J K Hicks[7], MR Wilkinson[2], DW Haas[8] and DL Kroetz[1]

+

MAYO CLINIC Mayo Medical Laboratories

Laboratory Service Report **1-800-533-1710**

Patient Name	Patient ID	Age	Gender	Order #
TESTINGRNV,2C19B ATLAS REPORT	SA00234474	32	M	SA00234474
Ordering Phys				**DOB**
CLIENT,CLIENT				03/26/1983
Client Order #	**Account Information**			**Report Notes**
SA00234474				
Collected	C7028846-DLMP Rochester			
10/18/2015 00:00	SDSC 2 - Client Support			
Printed	Rochester, MN 55901			
11/02/2015 15:54				

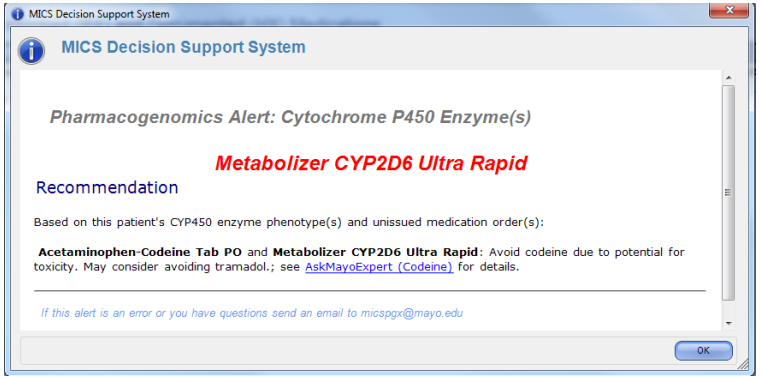
Test	Flag	Results	Unit	Reference Value
CYP2C19 Genotype, B				
CYP2C19 Phenotype		Poor metabolizer		
CYP2C19 Star Alleles		2/2		
CYP2C19 Interpretation				

This individual is likely a poor CYP2C19 metabolizer. Caution should be exercised when treating with drugs metabolized by CYP2C19 as follows:

If this patient is taking a prodrug that is activated by CYP2C19, such as clopidogrel, reduced activation of the drug is expected which may result in decreased efficacy of the drug.

If this patient is taking a drug that is inactivated by CYP2C19, such as citalopram, reduced inactivation is expected which may result in higher blood levels of the drug and potential side effects.

Consideration should be given to using drugs not metabolized by CYP2C19.



- *Data must be discrete, normalized, and unambiguous*
- *Knowledge must be computable*

http://www.mayomedicallaboratories.com/test-updates/attachment.php?id=43252

Standards Needed for Clinical Genomics

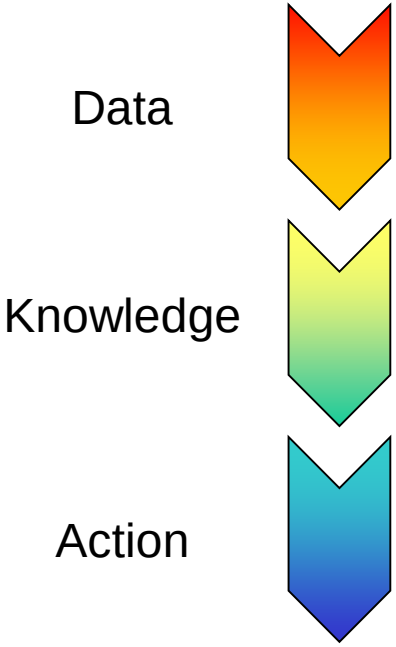
- Data Standards
  - Genetic data (observed result)
  - Test metadata
- Process Standards
  - Variant & haplotype interpretation
  - Genotype-phenotype translation rules
- Terminology Standards
  - Coded results and interpretations
  - Molecular phenotype
- Message/Interface Standards
  - Reporting

Standards Development

- SDOs
  - HL7, CDISC
- Government
  - ONC, FDA
- Consortia
  - CPIC, GA4GH
- Academia
  - Research Networks
    - eMERGE, ClinGen
  - Individual Researchers

Formal	*International (voluntary effort)*
Formal or Semi-Formal	*National (often mandated)*
Semi-Formal	*Domain (de facto std?)*
Informal or Semi-Formal	*Pragmatic, ad hoc (de facto std?)*

PGx Guidelines: Translating Data to Action



Genotype

↓

Diplotype

↓

Molecular Phenotype

↓

Clinical Significance

rs1142345: T/T,
rs1800460: C/T

↓

*TPMT *1/*3B*

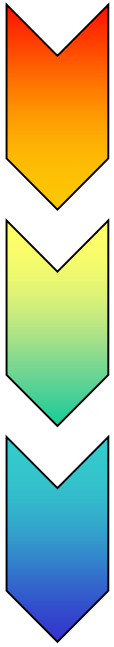
↓

Intermediate Metabolizer

↓

Adjust dose when treated with mercaptopurine

Agenda:
Standards for Genomic Data and Knowledge



Data

Variant Representation: GA4GH, ClinGen Allele Registry

Knowledge

Variant Annotation: ACMG, ClinGen VCI

Translation Tables & Phenotypes: CPIC

Action

EHR Integration: HL7

Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

The Global Alliance for Genomics and Health aims to accelerate progress in genomic science and human health by developing standards and framing policy for responsible genomic and health-related data sharing.

Genomic Knowledge Standards

Global Alliance for Genomics & Health

GKS aims to develop, adopt, and adapt standards to enable the exchange of genomic knowledge

- Establish framework of standards that lower barriers to translating genomic knowledge into clinical practice
- Reusing existing knowledge is essential to future impact
- Knowledge repositories are a key component
- Collaborate with other SDOs on harmonized specifications

Genomic Knowledge Standards

Global Alliance for Genomics & Health

Variant Representation

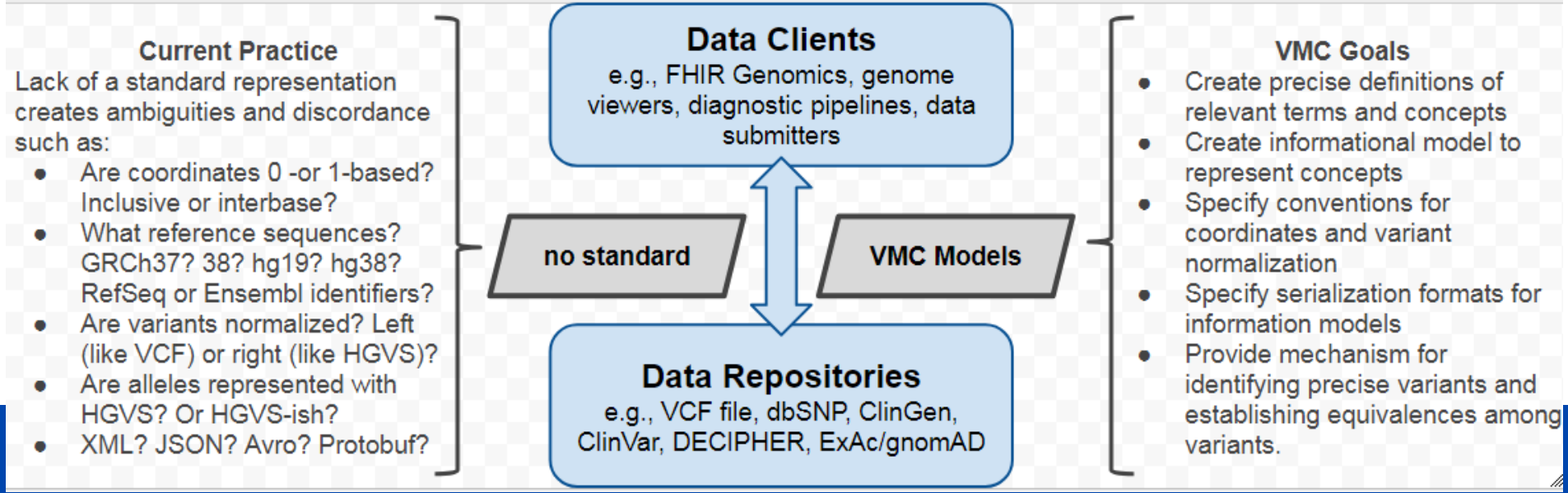
- Driver Project priorities => Work Stream goal
  - Create a standard model for computer readable variant representation (definitional)
- Simple variants (VMC)
  - v 1.0 by Q4 2018
  - Peer-reviewed publication
- Complex variants (structural, CNV, imprecise ends)
  - v 0.1 by Q4 2018

Variant Modelling Collaboration (VMC)

Global Alliance for Genomics & Health

Participants: GA4GH, HL7, ClinGen, NCBI, SO, EBI, Ensembl

Goal: to develop a technical specification for representing and communicating biological sequence variation



VMC 1.0

Global Alliance for Genomics & Health

Focus: variation that is defined by alleles — assertions of a specific sequence at precise locations on a sequence

Entities: Sequence, Location, Allele, Haplotype, Genotype

- Terminology
- Information Model
  - Minimalistic, use as part of a given use case
- Machine-readable schema definition

ClinGen Allele Registry

Key problem for variant identification and integration:

There are different ways to describe the same variations using HGVS expressions

Indels are defined differently using HGVS expressions

- Reference sequence (RF1): ACTGTCGTG
- ACT~~GTC~~GTG -> *RF1:c.**4**_**6**delGTC*
- ACTG~~TCG~~TG -> *RF1:c.**5**_**7**delTCG*
- ACTGT~~CGT~~G -> *RF1:c.**6**_**8**delCGT*
- Alternate sequence: ***ACTGTG***

Slide courtesy of Ronak Y Patel

How can ClinGen Allele registry help?

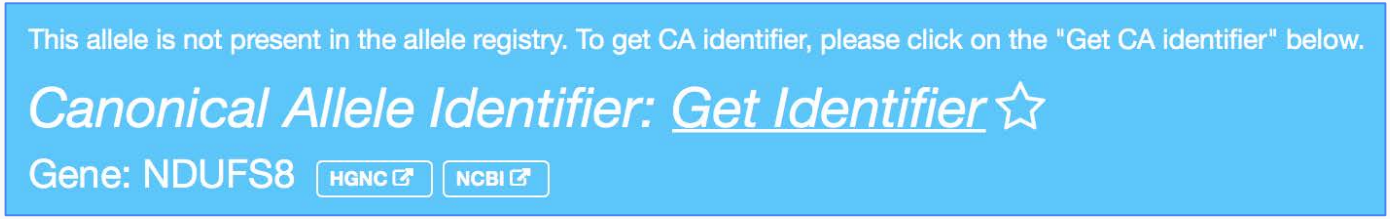
Resolves and provides identity to allele

Allele registry provides services to match needs of wide variety of audience

Variant interpreter/curator:

1. Search using HGVS

2. Get identifiers if not registered in the registry



> 650 million variants are already registered, so likely the variant you are looking for is already registered and have a canonical allele identifier

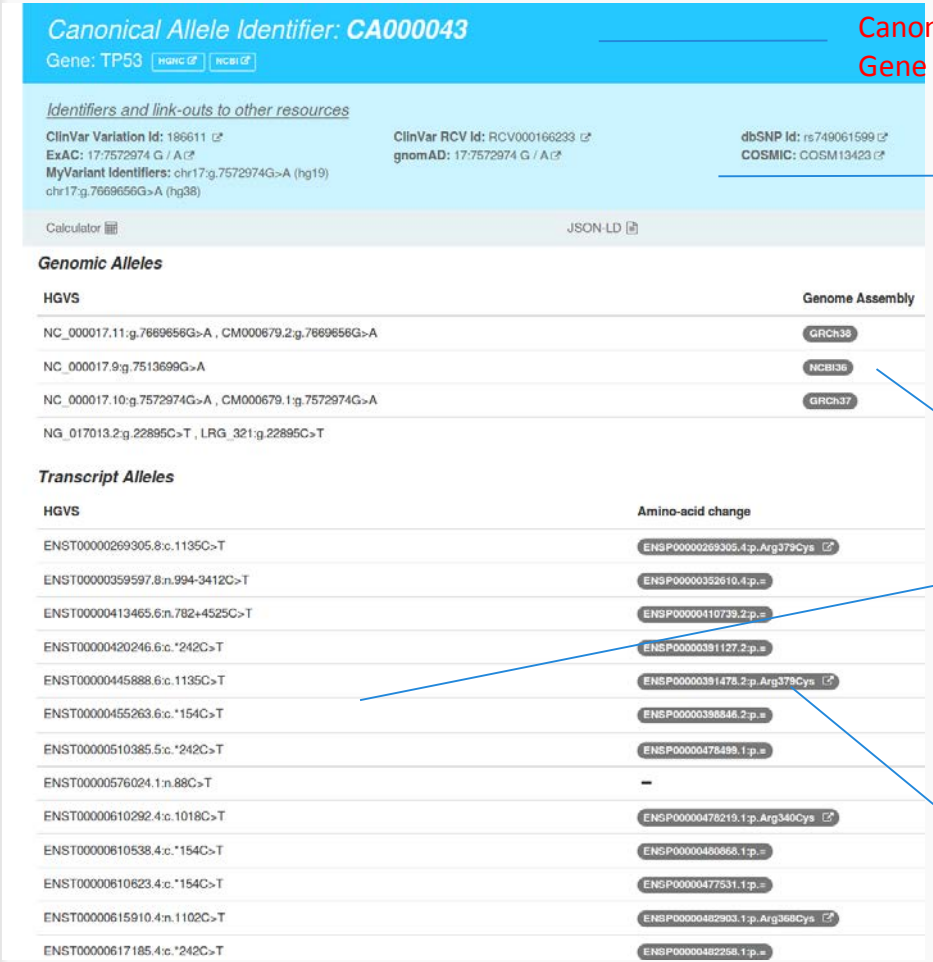
Slide courtesy of Ronak Y Patel

https://reg.clinicalgenome.org

How can ClinGen Allele registry help?

Resolves and provides identity to allele

Single Allele view



Slide courtesy of Ronak Y Patel

How can ClinGen Allele registry help?

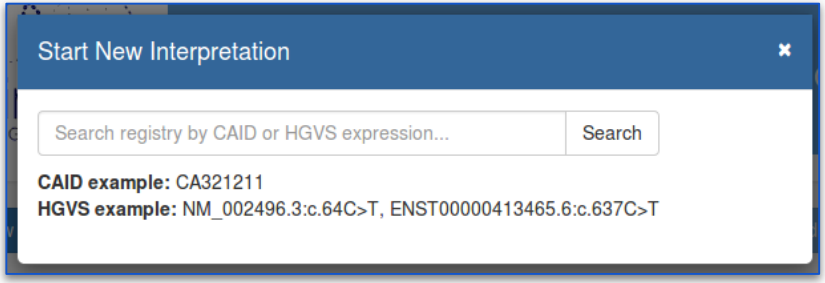
Help integrate variant annotations from various resources

- Simple, well-documented REST-APIs [Backward compatible]
- Simple GET/PUT/POST requests make it easy to integrate
- Pathogenicity Calculator/Variant curation Interface/Data exchange

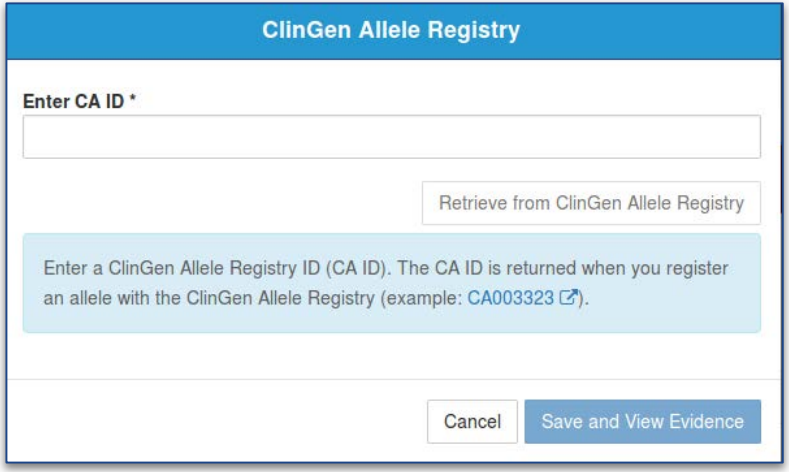
ClinVar

MyVariant.info

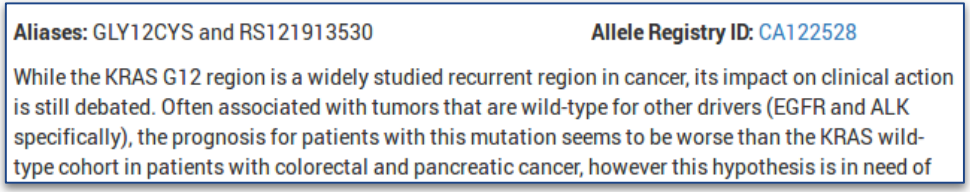
ClinGen Pathogenicity Calculator



ClinGen Variant Curation Interface



CIViC



Slide courtesy of Ronak Y Patel

https://reg.clinicalgenome.org

ACMG STANDARDS AND GUIDELINES | Genetics inMedicine

May 2015

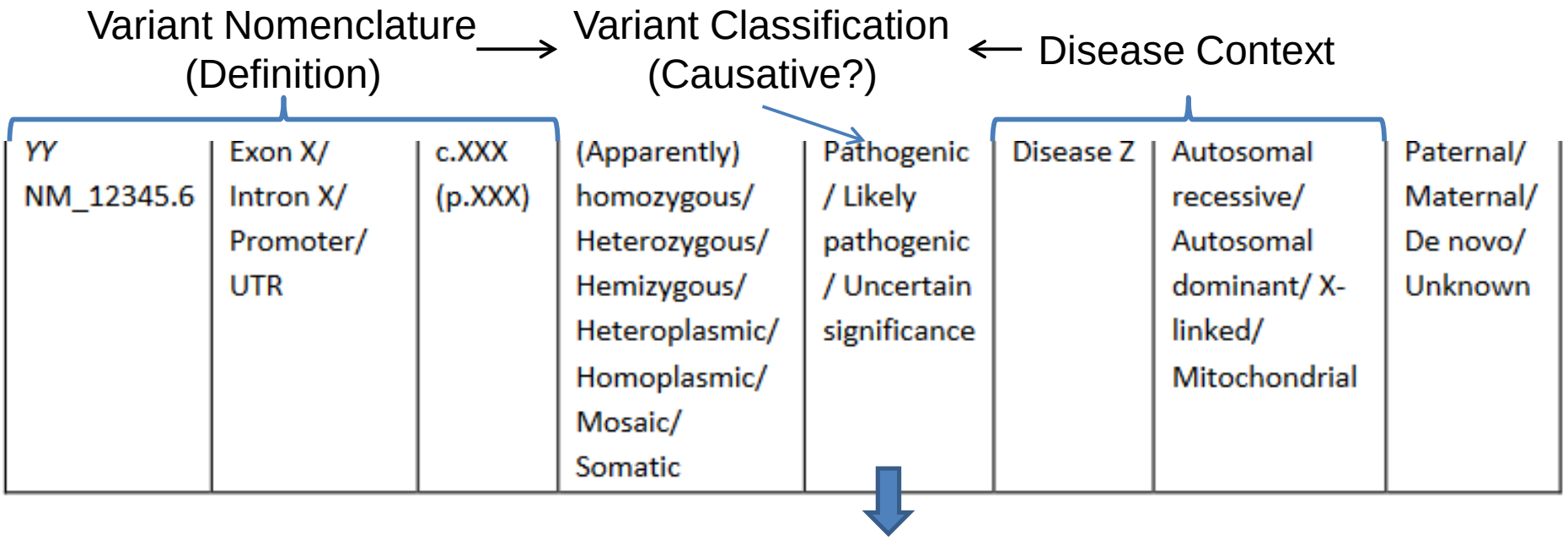
PMID 25741868

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Sue Richards, PhD[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

"By adopting and leveraging next-generation sequencing, clinical laboratories are now performing an ever-increasing catalogue of genetic testing spanning genotyping, single genes, gene panels, exomes, genomes, transcriptomes, and epigenetic assays for genetic disorders. By virtue of increased complexity, this shift in genetic testing has been accompanied by new challenges in sequence interpretation. In this context the ACMG convened a workgroup in 2013 comprising representatives from the ACMG, the Association for Molecular Pathology (AMP), and the College of American Pathologists to revisit and revise the standards and guidelines for the interpretation of sequence variants."

2015 ACMG Guidelines: Structured Variant Data



PMID 25741868, Supplemental Data

2015 ACMG Guidelines: Evidence Framework

Categories and Weight of Evidence

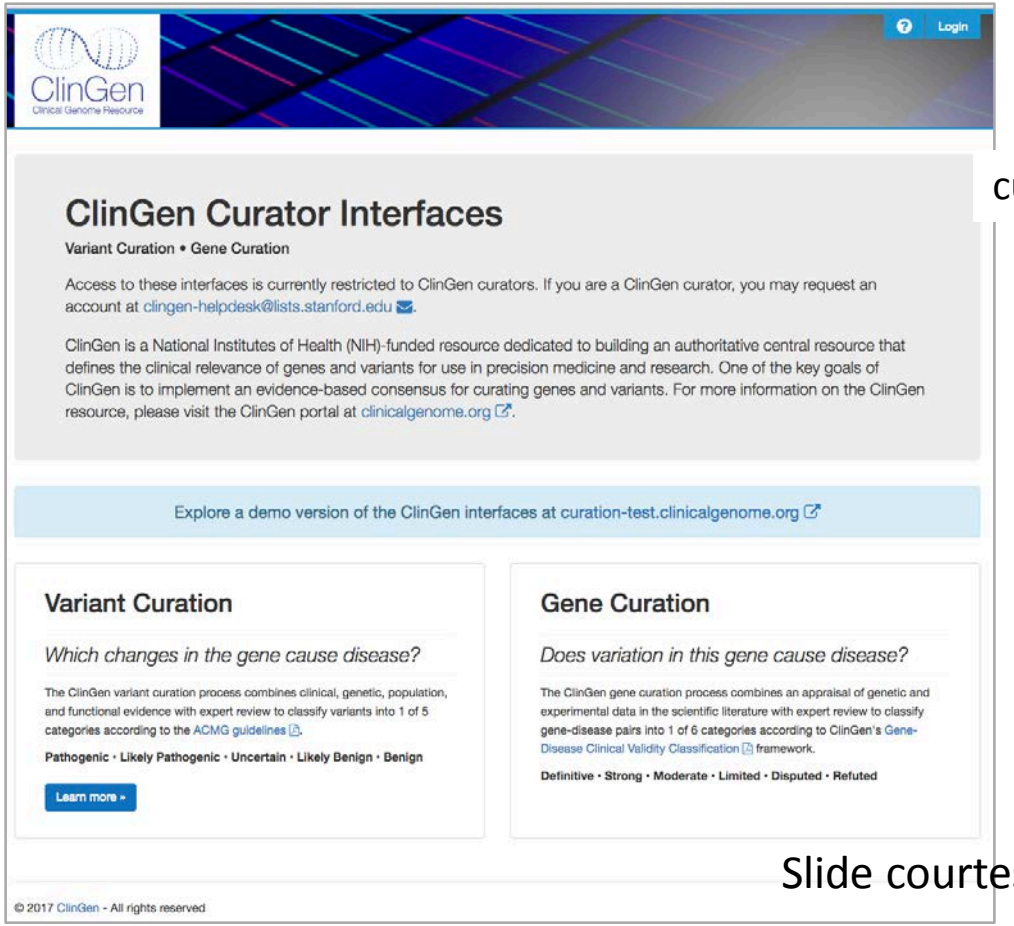
Types of Evidence

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Criteria

PMID 25741868

ClinGen Variant Curation Interface Landing Page



Login

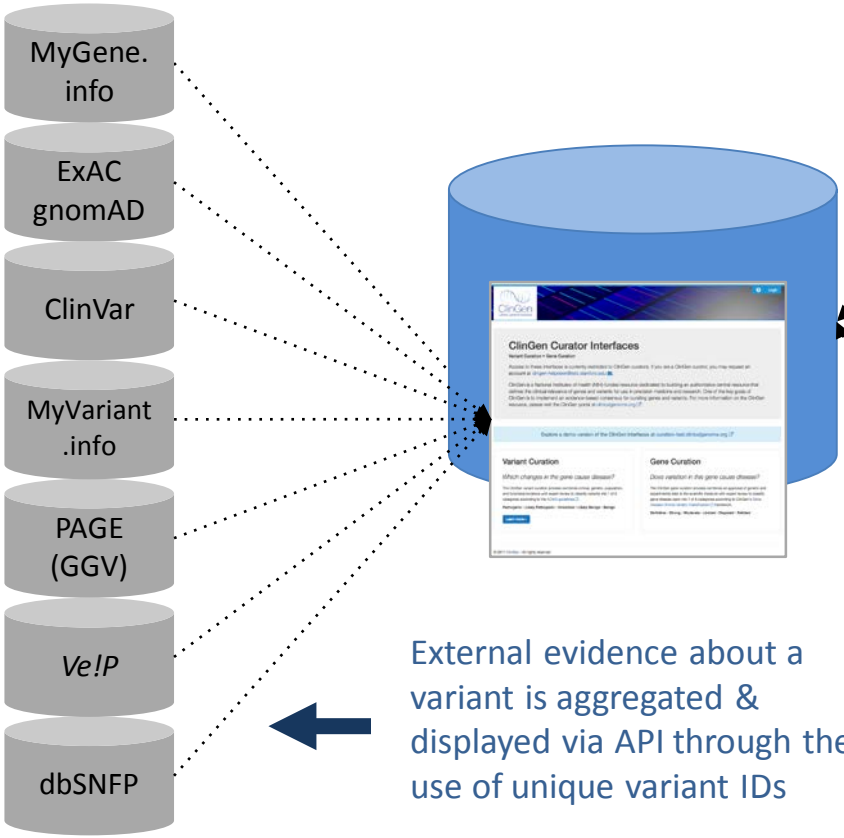
curation.clinicalgenome.org

email helpdesk to register

Explore demo version

Slide courtesy of Selina S. Dwight

The Variant Curation Interfaces rely on ClinGen Allele Registry and ClinVar IDs for aggregating evidence and supporting manual curation of individual variants



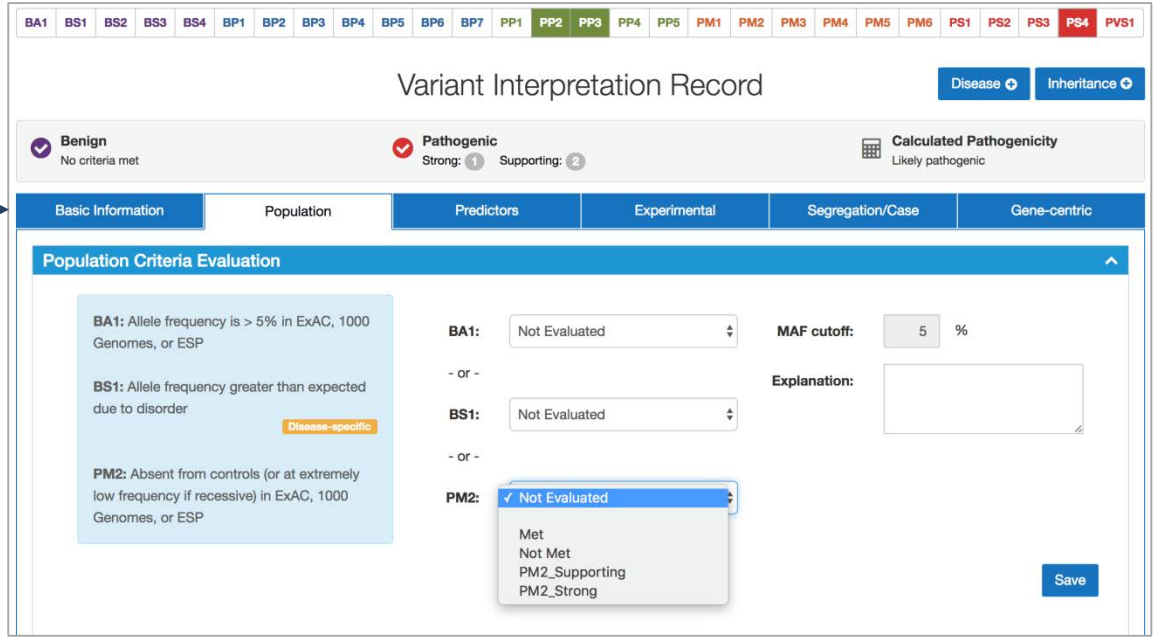
The API feature of the ClinGen Allele Registry is used to identify the correct variant during manual curation

External evidence about a variant is aggregated & displayed via API through the use of unique variant IDs

ClinGen Allele Registry

Slide courtesy of Selina S. Dwight

Interpretation Mode



ACMG criteria appear alongside relevant evidence categories

Ability to evaluate criteria at strength or modified strength

Criteria bar indicates status of each criterion (Met, Met at Modified strength, Not Met, Not evaluated)

Pathogenicity calculated “on the fly” according to ACMG rules

Slide courtesy of Selina S. Dwight

Acknowledgements

GA4GH GKS

- Andy Yates
- Bob Freimuth
- Larry Babb
- Matt Brush
- Michael Baudis
- Javier Lopez
- Reece Hart

ClinGen AR

- Piotr Pawliczek
- Ronak Patel
- Lillian Ashmore
- Sameer Paithankar
- Andrew Jackson
- Neethu Shah
- Sharon Plon
- Aleks Milosavljevic

ClinGen DM WG

- Larry Baab
- Chris Bizon
- Tristan Nelson
- Bradford Powell
- Bob Freimuth
- Danielle Azzariti

ClinGen VCI

- Selina Dwight
- Matt Wright
- Jimmy Zhen

CPIC™
Clinical Pharmacogenetics Implementation Consortium

- CPIC guidelines are designed to help clinicians understand HOW available genetic test results should be used to optimize drug therapy.
  - Not WHETHER tests should be ordered.
- Key Assumption:
  - Clinical high-throughput and pre-emptive genotyping will become more widespread.
  - Clinicians will be faced with having patients' genotypes available even if they did not order test with drug in mind.

CPIC Informatics Working Group

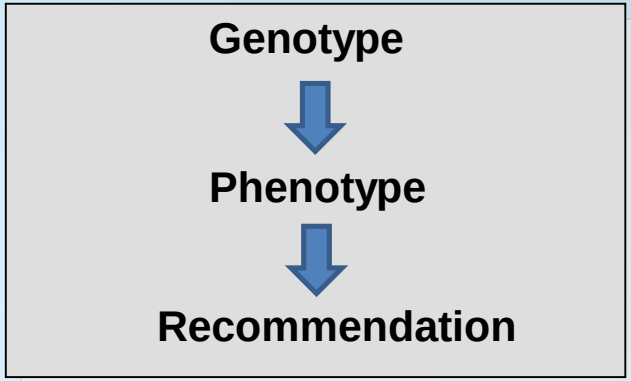
- Growing interest in informatics aspects of CPIC guidelines and clinical implementation of pharmacogenetics
- Goal: To support the adoption of the CPIC guidelines by identifying, and resolving where possible, potential technical barriers to the implementation of the guidelines within a clinical electronic environment.
- Working group leaders
  - Bob Freimuth (Mayo Clinic)
  - James Hoffman (St. Jude)
  - Michelle Carrillo (PharmGKB)

Slide courtesy of James Hoffman

CPIC Informatics Working Group: Initial Focus

- Create comprehensive translation tables from genotype to phenotype to clinical recommendation for CPIC guidelines
  - Define structure and process to efficiently develop and maintain in the most useful format(s)
  - Publish as part of CPIC guidelines

CPIC™
Clinical Pharmacogenetics Implementation Consortium



CPIC Translation Tables: Allele Function to Phenotype

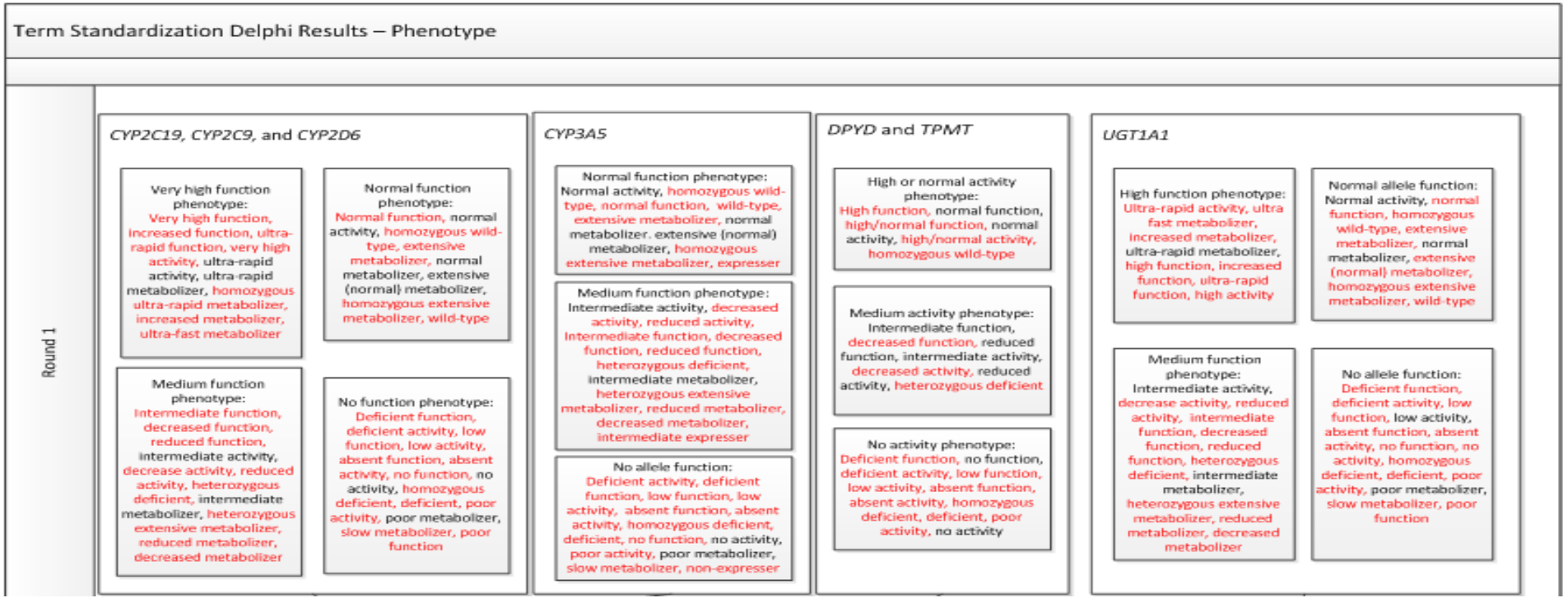
CYP2C19 Allele	Allele Functional Status
*1	Normal function
*2	No function
*3	No function
*4A	No function
*4B	No function
*5	No function
*6	No function
*7	No function
*8	No function
*9	Decreased function
*10	Decreased function
*11	Normal function

	CYP2C19 Diplotype	Coded Genotype/Phenotype Summary[b]
1		
2	*1/*1	Normal Metabolizer
3	*1/*2	Intermediate Metabolizer
4	*1/*3	Intermediate Metabolizer
5	*1/*4A	Intermediate Metabolizer
6	*1/*4B	Intermediate Metabolizer
7	*1/*5	Intermediate Metabolizer
8	*1/*6	Intermediate Metabolizer
9	*1/*7	Intermediate Metabolizer
10	*1/*8	Intermediate Metabolizer
11	*1/*9	Likely Intermediate Metabolizer
12	*1/*10	Likely Intermediate Metabolizer

Table 1 Assignment of likely thiopurine methyltransferase phenotypes based on genotypes

Likely phenotype	Genotypes	Examples of diplotypes
Homozygous wild-type or normal, high activity (constitutes ~86–97%[a] of patients)	An individual carrying two or more functional (*1) alleles	*1/*1
Heterozygote or intermediate activity (~3–14%[a] of patients)	An individual carrying one functional allele (*1) plus one nonfunctional allele (*2, *3A, *3B, *3C, or *4)	*1/*2, *1/*3A, *1/*3B, *1/*
Homozygous variant, mutant, low, or deficient activity (~1 in 178 to 1 in 3,736 patients[a])	An individual carrying two nonfunctional alleles (*2, *3A, *3B, *3C, or *4)	*3A/*3A, *2/*3A, *3C/*3A,

The Need for Term Standardization



Slide courtesy of K. Caudle

CPIC Phenotype Term Standardization Project

Purpose:

- To standardize phenotype terms in the CPIC guidelines and harmonize terms with external groups (e.g., ClinGen, IOM, etc.)
  - Allele functional status terms
    - Low, absent, high, intermediate
  - Phenotype
    - UM, EM, IM, PM

Standardizing terms for clinical pharmacogenetic test results: consensus terms from the Clinical Pharmacogenetics Implementation Consortium (CPIC)

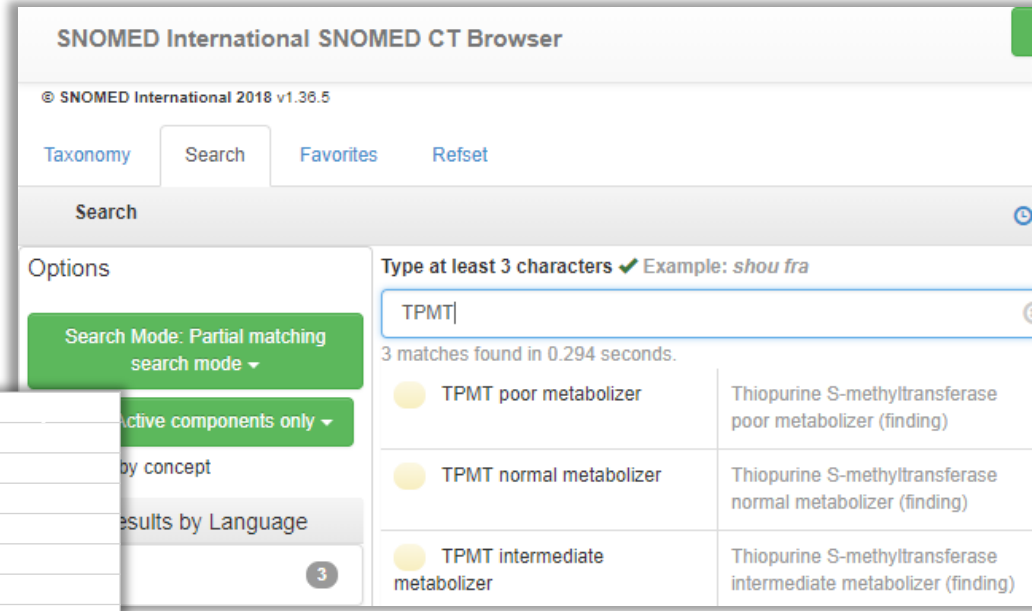
Kelly E. Caudle, PharmD, PhD[1], Henry M. Dunnenberger, PharmD[2], Robert R. Freimuth, PhD[3], Josh F. Peterson, MD[4,5], Jonathan D. Burlison, PhD[1], Michelle Whirl-Carrillo, PhD[6], Stuart A. Scott, PhD[7], Heidi L. Rehm, PhD[8], Marc S. Williams, MD[9], Teri E. Klein, PhD[6], Mary V. Relling, PharmD[1], James M. Hoffman, PharmD, MS[1]

Genet Med. 2016 Jul 21. PMID: 27441996

PGx in LOINC and SNOMED CT

LOINC	LOINC Component
50956-2	HLA-B*57:01
57979-7	HLA-B*15:02
79711-8	HLA-B*58:01
79712-6	HLA-A*31:01
79713-4	TPMT gene product metabolic activity interpretation
79714-2	CYP2C19 gene product metabolic activity interpretation
79715-9	CYP2D6 gene product metabolic activity interpretation
79716-7	CYP2C9 gene product metabolic activity interpretation
79717-5	CYP3A5 gene prod interpretation
79718-3	UGT1A1 gene prod interpretation
79719-1	DPYD gene produc interpretation
79720-9	CYP2B6 gene prod interpretation
79721-7	CYP4F2 gene prod interpretation
79722-5	SLCO1B1 gene pro interpretation

LA6576-8	Positive
LA6577-6	Negative
LA10315-2	Ultrarapid metabolizer
LA25390-8	Rapid metabolizer
LA25391-6	Normal metabolizer
LA10317-8	Intermediate metabolizer
LA9657-3	Poor metabolizer
LA25392-4	Increased function
LA25393-2	Normal function
LA25395-7	Decreased function
LA25394-0	Poor function



Acknowledgements - CPIC

- CPIC Leaders
  - Mary Relling
  - Teri Klein
- CPIC Coordinators
  - Kelly Caudle
  - Roseann Gammal
- PGRN
- PharmGKB
  - Russ Altman
  - Teri Klein
  - Michelle Whirl-Carrillo
- CPIC Informatics WG
  - James Hoffman
  - Michelle Whirl-Carrillo
  - Bob Freimuth

- Develops international standards for transfer of clinical and administrative data between software applications used by healthcare providers and clinical systems
- Examples
  - v2/v3 messaging
  - CDA, CCD
  - Arden Syntax
  - FHIR
- > 60 work groups

HL7 Clinical Genomics Work Group

- Goal: to facilitate the development of common standards for the exchange of clinical and translational information related to an individual's genomic data and family health history, and (its linkage to) relevant clinical information

- Organizational perspectives (not exclusive)
  - Academic medical centers, hospitals
  - Genomic testing labs
  - Public data repositories and knowledge-bases serving as references to clinical genetics/genomics
  - *Translational research programs??*

HL7 Clinical Genomics Working Group Overview of Activities

v2

- Implementation Guides

  Genetic Variation (Fully LOINC-Qualified Genetic Variation Model)

  Laboratory Results Interface (LRI)

v3

- Family History (Pedigree)
- Genetic Variations
- Gene Expression
- CMETs defined by the Domain

CDA

- Implementation Guide for Genetic Testing Reports

FHIR STU3

- Sequence Resource
- Profiles:

  Genetic Observation

  DiagnosticReport

  ProcedureRequest

  HLA Reporting

  Family member history for genetics analysis

Common:

- Domain Analysis Models for the various topics
- A Domain Information Model describing the common semantics
- Semantic a*lignment among the various specs*

10.6 Resource Sequence - Content

Clinical Genomics Work Group	Maturity Level: 1	Trial Use	Compartments: Not linked to any defined compartments

Raw data describing a biological sequence.

10.6.1 Scope and Usage

The Sequence resource is designed to describe an atomic sequence which contains the alignment sequencing test result and multiple variations. Atomic sequences can be connected by link element and they will lead to sequence graph. By this method, a sequence can be reported. Complete genetic sequence information, of which specific genetic variations are a part, is reported by reference to the GA4GH repository. Thus, the FHIR Sequence resource avoids large genomic payloads in a manner analogous to how the FHIR ImagingStudy resource references large images maintained in other systems. For use cases, details on how this resource interact with other Clinical Genomics resources or profiles, please refer to implementation guidance document here .

10.6.1.1 Genetic Standards and Resources include:

- Variant Databases: dbSNP , ClinVar, and COSMIC
- Reference Sequences: RefSeq and ENSEMBL

FHIR®© Current Build

Home | Getting Started | Documentation | Resources | Profiles | Extensions | Operations | Terminologies

Diagnostics > **Genomics Implementer Guidance**

10.10 Genomics Implementation Guidance

Clinical Genomics Work Group	Maturity Level: 1	Ballot Status: Trial Use

Table of Contents

Standards Development

Development

- Evaluate existing standards
- Gap analysis
- Define scope of new
- Develop draft spec
- Ballot (vote)
- Reconciliation
- Approval

Global Alliance
for Genomics & Health

- Implementation

Adoption

- Define domain
- Define use cases and requirements

Motivation

Standard Development – Take Home Points

- Driven by real-world use cases and implementations
- Standards are developed by those that show up
  - Nearly all participants are volunteers
- Robust standards require diverse expertise
  - Domain SMEs, modeling, technical
  - Clinical, academic, industry, government
- Standards are not perfect or perfectly comprehensive
  - Gaps in structure and/or content
  - Collaborate to extend, don't create Yet Another Standard