

# Bidirectional data flow from clinic to lab and back

Lawrence Babb  
Genomic Medicine XI  
September 2018

## *Conflicts of interest*

- No conflicts

## *Objective*

- Briefly discuss Clinic->Lab->Clinic primary data flow
- Identify fundamental next steps to realize EHR implementations
- Describe barriers in variant data exchange in clinical systems
- Describe variant specification and service efforts
- Discuss solutions to remove adoption barriers and begin looking ahead.

## *My Experience (clinical genetic data exchange.)*

- Software developer and data modeler for 30 years
- Past 15 yrs full-time clinical genetic test result and federated knowledge management system development, design & operations support
  - Product : Partners Healthcare & Sunquest - GeneInsight/MitoGen Genetics.
- Past 10 yrs participation in development of CG community standards
  - HL7 Clinical Genomics WG (10yrs off and on)
  - DIGITizE AC: Displaying and Integrating Genetic Information Through the EHR
  - Global Alliance for Genomic Health (GA4GH) - Genomic Knowledge Workstream (GKS), VMC, Variant Annotation (1yr)
  - eMERGE: enabling data sharing for Deidentified Case Repo & LMM Clinics
  - ClinGen: data exchange modeling standards development 3yrs

# Clinic → Lab → Clinic

## A. Clinic Sends Order to Lab

- Clinic's ordering physician fills out genetic test requisition and sends along with patient bio-sample to lab.

## B. Lab Receives Order / Performs Services

- Accessioner transfers requisition into lab system.
- Lab Tech performs assay to identify variants (findings)
- Path./Gen/GC/Fellows curate evidence, assesses variants, draft case level interpretation.

## C. Lab Sends Results to Clinic (Sign-out)

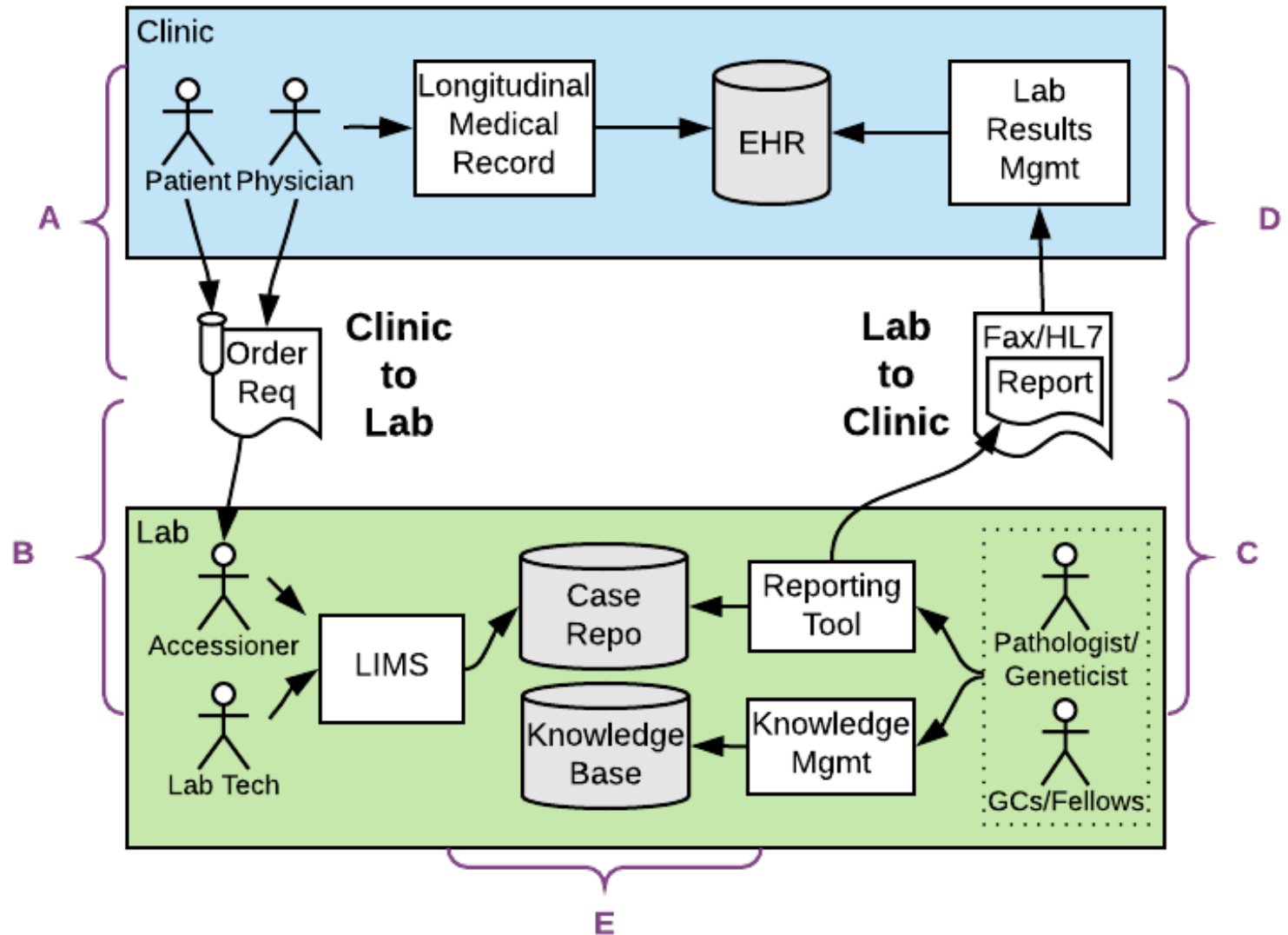
- Pathologist / Geneticist finalizes case report.

## D. Clinic Receives Results from Lab

- Clinic receives result by fax or electronic HL7 v2 Unsolicited Result message.

## E. Lab Knowledge Evolves

- Each new similar case with the same or related variants along with re-assessment of external evidence yield changes to prior interpretations.
- Previously reported cases not updated or notified.



# Exchanging Data v. Sharing Data

- In order **to share data** effectively, **we must first exchange** it effectively
- **Job #1**: standardize the exchange of **variants** and **phenotypes** to support clinical genetic results and variant knowledge management.
- EHR - Lab vendor adoption **won't occur without HL7**
- The domain is **too big for HL7 Clinical Genomics WG** to do alone
- **Supporting resources are essential**: GA4GH, ClinGen, NCBI/EBI, others...

## My Assertion

*Resources in the form of **specifications and services** that support clinical utility will*

*1) **Reduce time to normative standard** by reducing the scope of work, and*

*2) **Increase adoption** by reducing development costs and risk to lab/EHR vendors*

*resulting in a **shortened timeframe** to attain **clinically useful exchange of data to the EHR** from labs while **also creating the opportunity for large-scale data sharing of knowledge.***


# Genetic Data Exchange Resources

(some examples of developing and existing resources)

- HL7 Clinical Genomic
  - Info modeling, FHIR, V2
- ClinGen
  - Data Exchange Work Group
  - Allele Registry
- Ontologies
  - Monarch Disease Ontology – MONDO ( HPO, OMIM, ORDO, DO, MESH, etc..)
  - Scientific Evidence and Provenance Information Ontology – SEPIO
  - EMBL-EBI – Ontology Lookup Service
- Global Alliance for Genomic Health (GA4GH)
  - Genomic Knowledge Standards Workstream (GKS)
    - Variant Representation Specifications – VMC
    - Variant Annotation Specifications – Interpretation Statements (SEPIO?)
  - Clinical & Phenotype Data Capture Workstream (CPDC)
    - Phenopackets, Disease/Phenotypes – Snomed
  - MatchMaker Exchange, Variant Interpretation for Cancer Consortium (VICC)
- NCBI
  - Variation Services SPDI Notation - Sequence Position Deletion Insertion
  - ClinVar, dbSNP
  - MedGen



## *HL7: “Keeping our feet to the FHIR”*

- HL7 FHIR protocol has the promise of significantly increasing the number of innovators in the Health IT domain.
  -  on FHIR is a one example of an open platform which leverages FHIR
- HL7 v2 is the predominate implementation...and will be for years.
  - The v2 developments are moving in a good direction – but it is not clear if the barriers for adoption can be overcome – externalizing complexity can help!
- HL7 Clinical Genomic WG has a daunting task.
  - CG (v2), FHIR, Info Modeling – slow and steady, all volunteers
  - Good engagement, resource constrained, very large scope, timeline suffers



# HL7 Clinical Genomics Challenges

Scope is too broad and deep

Genomics – Variants, Phenotypes, Genomes, Specimens, Assays, Interpretation, ...

Use cases – Sequencing, Cyto, HLA, Somatic, Germline, Pharmacogenomic, WEX, WGS,...

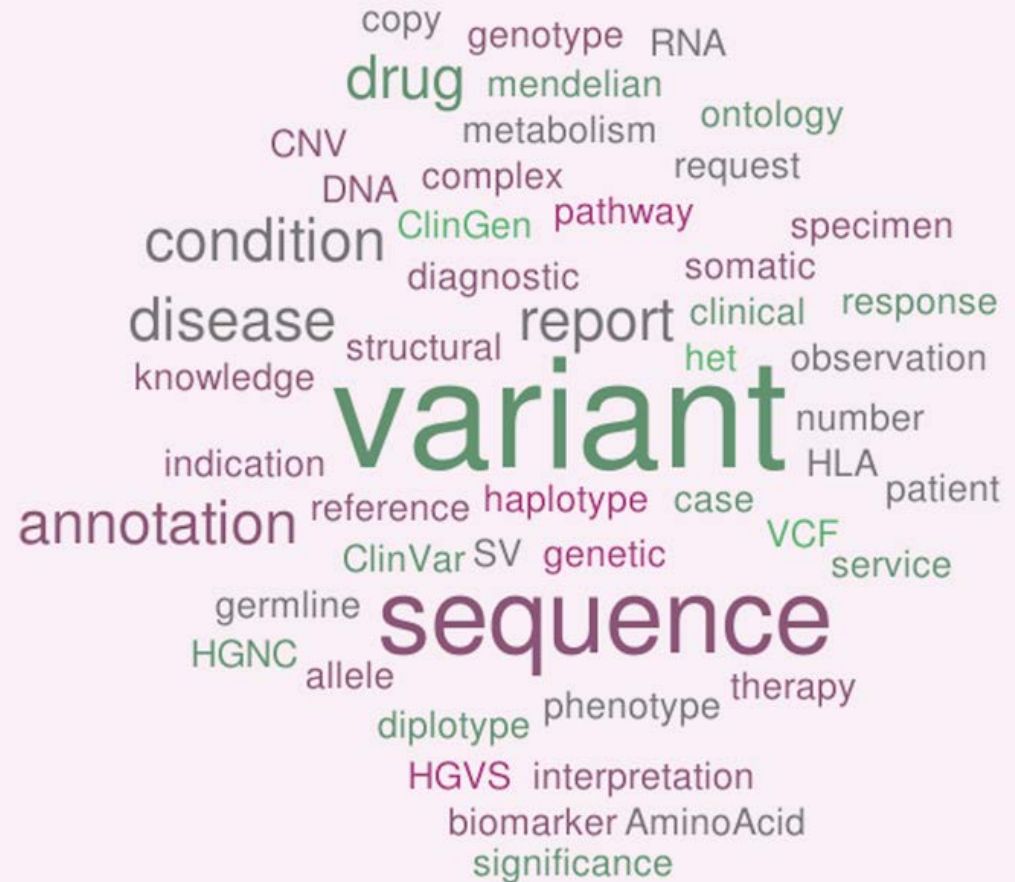
Defining Concepts / Models – Everything is an Observation or Sequence!

Applying Terminologies – Genes=HGNC, Disease=SNOMED, etc..

Modeling vs. Implementation Guide (IG)

[CG IG](#) – working to build profiles to support all of the above.



Resources – Volunteers, growing meeting participation, enlisting help when possible



## *HL7 – There's still hope!*

- FHIR is a considerably a more **open and innovative** platform in comparison to previous HL7 protocols (v2, v3, CDA).
- Any implementer/innovator can define **profiles and extensions** and implement it into a working solution.
- Standards driven by adoption - Working solutions that gain adoption will be used to drive evolving specification and implementation guidance.
- So, is anyone doing this?
  - The FHIR platform is getting close to being a Normative standard (Observation is there - but not much else) – 2019?.
  - It's still a pretty big job – and genetics is still a very complex model and insufficient services exist to reduce the cost of implementation.

## *Examples of standard enabling resources*

-  **HGNC** is a great example of a resource / standard that has broad adoption.
  - It removes the technical and conceptual data sharing/exchange barriers to adoption related to genes – a fundamental concept to genetics.
  - An immutable gene identifier, official/alias symbols, descriptions, classifications, synonyms, homologs, ext gene resources, etc...
-  **Genome Reference Consortium** a collaborative effort to provide standard stable versioned genome assemblies.

*Now we need clinically useful Variant and Phenotype resources to lower the barriers to adoption!*

## *Exchanging Variants is Step 1.*

- Variant Representation accuracy in exchange is paramount for clinical systems and computational consistency is essential to associate knowledge, compare findings, and scale discovery.
- Everyone benefits from open services. Providing these fundamental services significantly reduces the barriers for adoption and fuels investment and innovation.
- GA4GH GKS – Variant Representation – VMC Specification
  - Objective - A comprehensive specification to provide the exchange format of all forms of variant representation – not just sequence variants.
- Allele/Variant Services like a registry, validator, normalizer is essential.

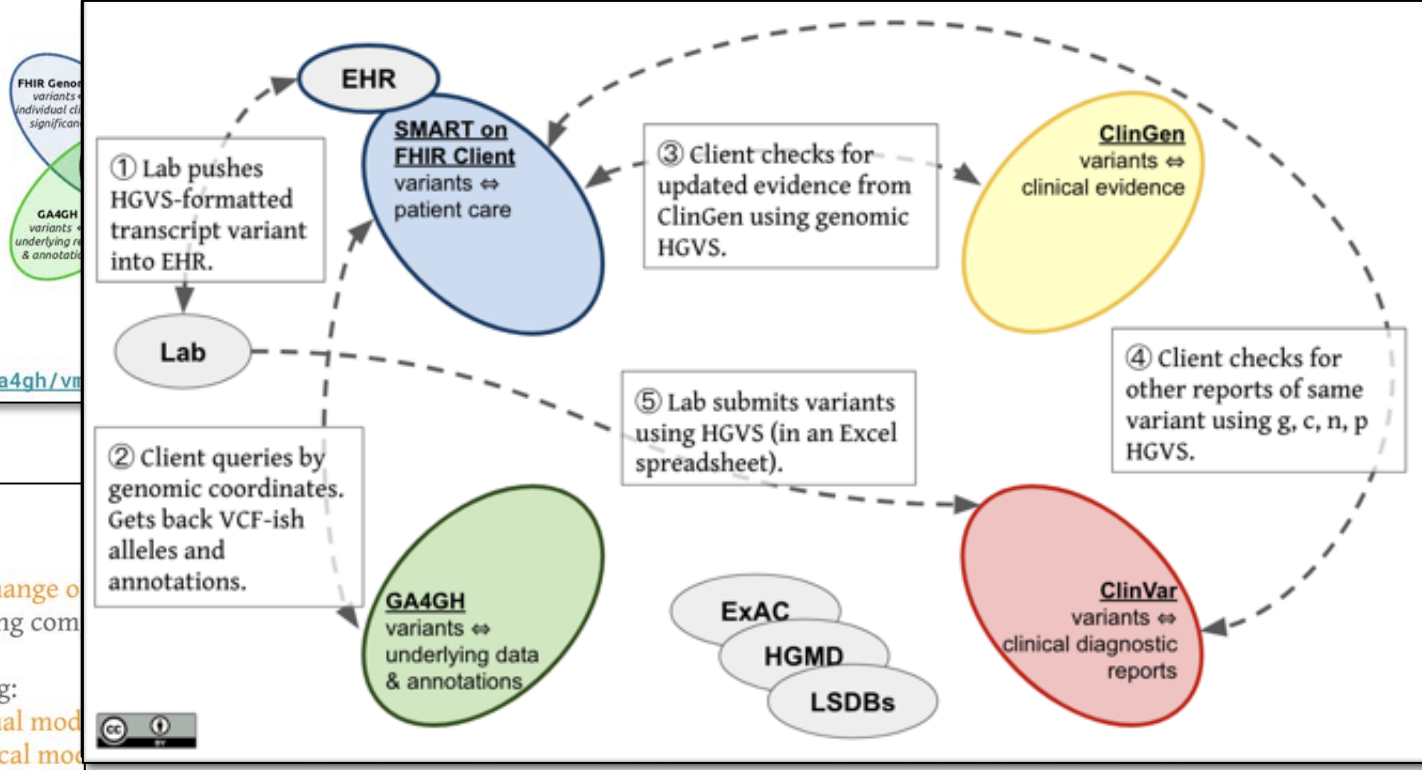
# VMC in 1 minute! - now under GA4GH GKS

## Variation Modelling Collaboration

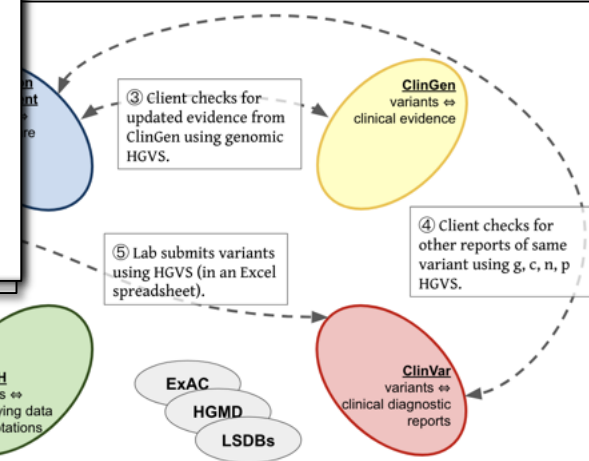
Reece Hart, Invitae, GA4GH  
 on behalf of  
 Gil Alterovitz, HMS/BCH, FHIR Genomics  
 Larry Babb, Sunquest, ClinGen  
 Karen Eilbeck, University of Utah  
 Bob Freimuth, Mayo Clinic, ClinGen, HL7/FHIR  
 Sarah Hunt, Ensembl  
 Brett Johnson, BCH, GA4GH, HL7/FHIR  
 David Kreda, Harvard Medical School, FHIR Genomics  
 Jennifer Lee, NCBI, ClinVar  
 Peter Robinson, Jackson Laboratory, HPO  
 Shawn Rynearson, University of Utah, UCGD

<https://github.com/ga4gh/vm>

All of these are the same variant. Or not.



if we do nothing?



## The VMC Mission

To facilitate the reliable exchange of variation information among consumers and providers

by developing:

- 1) a shared conceptual model
- 2) a language-neutral logical model
- 3) at least one serialization protocol

... and by promoting its adoption among providers and consumers of variation data.

Computational Representation

```
{
  "ref": "GS_0a1b2c3d",
  "position": [139,139],
  "state": "c"
}
```

One representation

genomic coordinates. Gets back VCF-ish alleles and annotations.

## *Allele/Variant Registry*

- Exchanging Variants with external organizations requires variant validation and normalization.
- Building a variant knowledgebase requires variant normalization to prevent duplication of variant concepts and splintering evidence.
- Implementing a variant registry or validation service is costly to build and maintain, prone to clinically harmful errors, and inconsistent with community (if there was one).
- ClinVar is an emerging example of a Variant Registry (variation id). But it is only useful for the variants that have been submitted.

***In May 2015, NCBI began investigating a collaborative project with ClinGen to build an Allele Registry.***

# *What an Allele Registry Would Do?*

- Core Goal – Provide a universal allele identifier
  - Maintain all information required to unambiguously define canonical alleles
- Services
  - Provide publicly available user interface for browsing the registry
  - Expose programmatically accessible services for registering potentially novel alleles
  - Enable bulk download of all registry data
- The Need for Real Time Support
  - Allele registry should support both clinical and human subjects research systems
  - These systems will need access to variant registration services that provide real time response

## *Results of ClinGen / NCBI Collaboration*

- NCBI supported the concept and called for the establishment of a community organization to set policy (Variant Reference Consortium?)
- Pilot work began on devising improved method for variant normalization within NCBI archives (ClinVar, dbSNP, ...)
  - Variation Service - Sequence Position Deletion Insertion (SPDI)
  - Funding constraints indefinitely postponed project consideration for a NCBI Allele Registry
- ClinGen (Baylor) developed a fully functional and performant Allele Registry for ClinGen and community (public access).
- Community adoption is picking up.

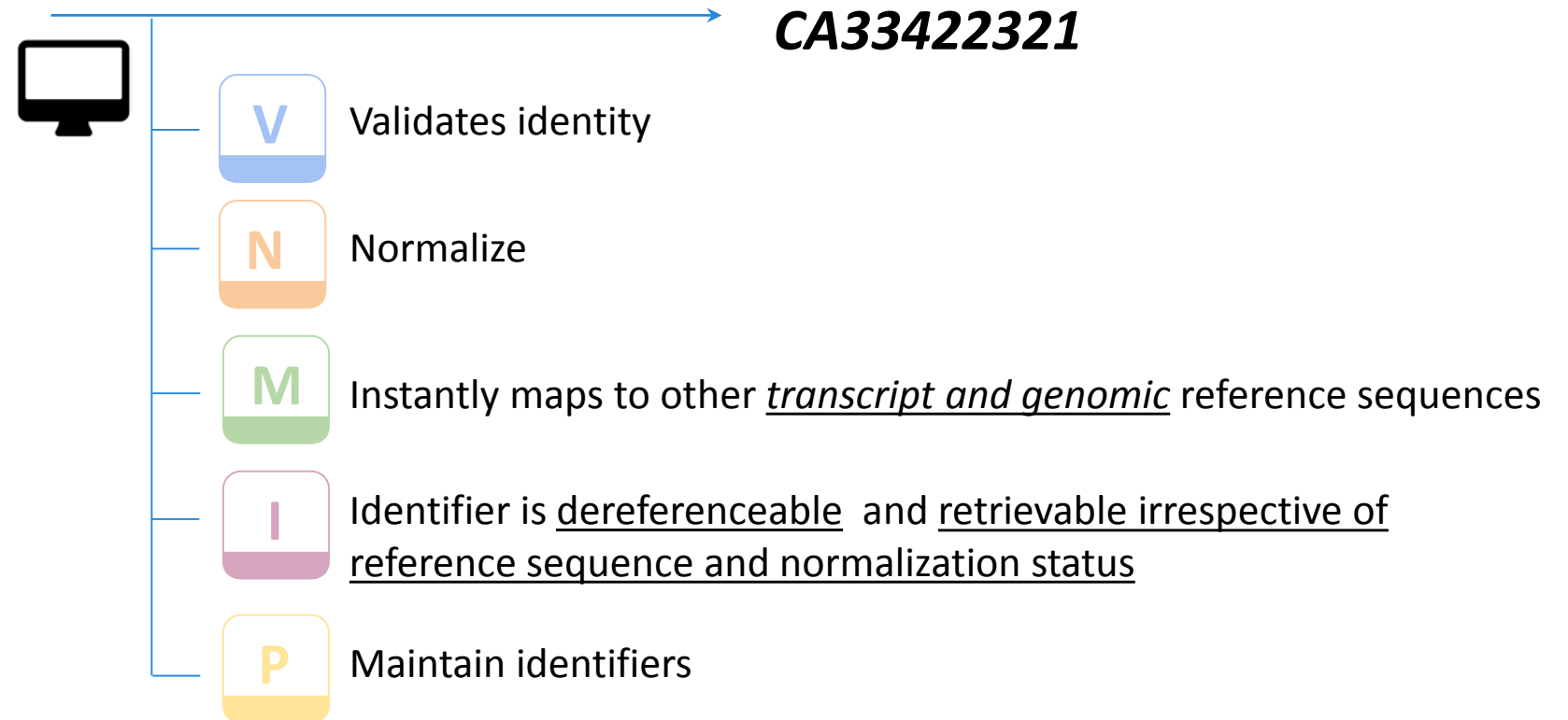


# The ClinGen Allele Registry!

Identifying information about allele

Provides identifier (URI) instantly

- HGVS
- VCF
- Interactively provided



Visit us at

<http://reg.clinicalgenome.org>

# Allele Registry resolves and provides identity to allele

## 1. Search using HGVS

Type of search:

Query:

[Q Search](#)

For example: NM\_002496.3:c.64C>T, ENST00000413465.6:c.637C>T, NC\_000017.10:g.7578212G>A

## 2. Get identifiers if not registered in the registry

This allele is not present in the allele registry. To get CA identifier, please click on the "Get CA identifier" below.

**Canonical Allele Identifier: [Get Identifier](#) ☆**

Gene: NDUFS8 [HGNC](#) [NCBI](#)

[> 650 million variants](#) are already registered, so it is likely the variant you are looking for is already registered and has a canonical allele identifier

Allele Registry resolves and provides identity to allele

Single Allele view

Transcript HGVS

**Canonical Allele Identifier: CA000043**  
Gene: TP53 [HGNC](#) [NCBI](#)

Identifiers and link-outs to other resources

ClinVar Variation Id: 186611 [↗](#)      ClinVar RCV Id: RCV000166233 [↗](#)      dbSNP Id: rs749061599 [↗](#)  
ExAC: 17:7572974 G / A [↗](#)      gnomAD: 17:7572974 G / A [↗](#)      COSMIC: COSM13423 [↗](#)  
MyVariant Identifiers: chr17:g.7572974G>A (hg19)  
chr17:g.7669656G>A (hg38)

Calculator [📄](#)      JSON-LD [📄](#)

**Genomic Alleles**

HGVS	Genome Assembly
NC_000017.11:g.7669656G>A , CM000679.2:g.7669656G>A	<a href="#">GRCh38</a>
NC_000017.9:g.7513699G>A	<a href="#">NCBI36</a>
NC_000017.10:g.7572974G>A , CM000679.1:g.7572974G>A	<a href="#">GRCh37</a>
NG_017013.2:g.22895C>T , LRG_321:g.22895C>T	

**Transcript Alleles**

HGVS	Amino-acid change
ENST00000269305.8:c.1135C>T	<a href="#">ENSP00000269305.4:p.Arg379Cys <a href="#">↗</a></a>
ENST00000359597.8:n.994-3412C>T	<a href="#">ENSP00000352610.4:p.=</a>
ENST00000413465.6:n.782+4525C>T	<a href="#">ENSP00000410739.2:p.=</a>
ENST00000420246.6:c.*242C>T	<a href="#">ENSP00000391127.2:p.=</a>
ENST00000445888.6:c.1135C>T	<a href="#">ENSP00000391478.2:p.Arg379Cys <a href="#">↗</a></a>
ENST00000455263.6:c.*154C>T	<a href="#">ENSP00000398846.2:p.=</a>
ENST00000510385.5:c.*242C>T	<a href="#">ENSP00000478499.1:p.=</a>
ENST00000576024.1:n.88C>T	-
ENST00000610292.4:c.1018C>T	<a href="#">ENSP00000478219.1:p.Arg340Cys <a href="#">↗</a></a>
ENST00000610538.4:c.*154C>T	<a href="#">ENSP00000480868.1:p.=</a>
ENST00000610623.4:c.*154C>T	<a href="#">ENSP00000477531.1:p.=</a>
ENST00000615910.4:n.1102C>T	<a href="#">ENSP00000482903.1:p.Arg368Cys <a href="#">↗</a></a>
ENST00000617185.4:c.*242C>T	<a href="#">ENSP00000482258.1:p.=</a>

Canonical Allele Identifier

Identifiers and links outs to various resources  
ClinVar, dbSNP, ExAC, gnomAD, COSMIC, myVariant.Info

Genomic HGVS

protein HGVS

# Allele Registry resolves and provides identity to allele

## Search Variants in ClinGen Allele Registry

Type of search

Select One

Query

For example: Select type of search to load examples.

[Supported searches](#)

HGVS

CAid

ClinVar Variation Id

ClinVar RCV Id

dbSNP Id

ExAC Id

gnomAD Id

MyVariant Id (hg19)

Any Identifier

HGNC Gene Symbol

Reference sequence & position

PMID

## Bulk query

This option provides a search box for entering multiple HGVS expressions (one per line). For alleles present in the allele registry, the search results provide a button to register an allele.

Bulk Query

```
NM_022124.5:c.9319+11G>A
NM_002474.12:c.4579-13G>A
NM_000525.3:c.67A>G
NM_000525.3:c.67A>G
NM_002693.2:c.2958C>T
NM_000352.4:c.4714G>A
NM_001256714.1:c.733G>A
NM_001605.2:c.2251A>G
NM_000350.2:c.1356+5_1356+6insC
NM_000350.2:c.5603A>T
NM_000350.2:c.2828G>A
NM_000350.2:c.4203C>A
NM_000350.2:c.4771G>A
```

Search

## Allele Registry resolves and provides identity to allele


































### Multi-allele view

- Bulk query
- Gene
- Reference sequence and position
- dbSNP

Sorted based on genomic coordinates

Gene Symbol

Linkouts to originating resources

Chr	Mutation (hg38)	CAid	Gene	Transcript	Linkouts
13	g.32319099T>A	<a href="#">CA387754103</a>	BRCA2	c.90T>A (p.Asn30Lys) n.288T>A n.90T>A	  
13	g.32319099T>C	<a href="#">CA025976</a>	BRCA2	c.90T>C (p.Asn30=) n.288T>C n.90T>C	      
13	g.32319101_32319115del	<a href="#">CA10579445</a>	BRCA2	c.92_106del (p.Trp31_Leu35del) n.290_304del n.92_106del	    
13	g.32319100T>A	<a href="#">CA387754105</a>	BRCA2	c.91T>A (p.Trp31Arg) n.289T>A n.91T>A	  
13	g.32319100T>C	<a href="#">CA026029</a>	BRCA2	c.91T>C (p.Trp31Arg) n.289T>C n.91T>C	    
13	g.32319100T>G	<a href="#">CA387754107</a>	BRCA2	c.91T>G (p.Trp31Gly) n.289T>G n.91T>G	  
13	g.32319101G>A	<a href="#">CA026096</a>	BRCA2	c.92G>A (p.Trp31Ter) n.290G>A n.92G>A	      

Chromosome

Canonical Allele Identifiers

Transcript and amino-acid variations

# Query ClinGen Allele Registry with partial information about variation

## Do not have transcript/HGVS expression?

For a substitution with gene symbol, position, reference and alternate alleles known, please use this service:

Gene and variation based query

Alternatively, please use this service to identify allele interactively if HGVS expression or transcript is not available.

Interactively generate variation

Gene symbol	Transcript (c.) or Amino-acid (p.)	Position	Reference allele	Alternate allele
NDUFS8	c.	413	G	A

Currently only c. queries are supported

Query

**Search Reference Sequences**

<a href="#">NM_000038.5</a>	Generate HGVS using this transcript
<a href="#">NM_001127510.2</a>	Generate HGVS using this transcript
<a href="#">NM_001127511.2</a>	Generate HGVS using this transcript

Start:	399
End:	402
Reference:	TTTA
Alternate:	

Search Allele

36	AGTTGAGGCA	CTGAAGATGG	AGAACTCAAA	T
96	TCATCTTACA	AAACTGGAAA	CTGAGGCATC	T
156	AGGAAGTATT	GAAGATGAAG	CTATGGCTTC	T
216	TAAAGAGCTT	AACTTAGATA	GCAGTAATTT	C
276	CCTCCGTTCT	TATGGAAGCC	GGGAAGGATC	T
336	TGTTCCCTATG	GGTTCATTTT	CAAGAAGAGG	G
396	ATA <b>TTT</b> A	GAACTTGAGA	AAGAGAGGTC	A
456	AAAGGAAAAA	GACTGGTATT	ACGCTCAACT	T

## *Allele Registry services are accessible through high availability and fast REST-APIs*

- Simple, well-documented REST-APIs [Backward compatible]
- Simple GET/PUT/POST requests make it easy to integrate
- Current registration/query bandwidth: 1K-50K variants per second

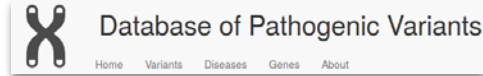


# Resources using canonical allele identifiers



```

object ▶ clingen ▶ _license
  ▼ object {13}
    _id : chr11:g.68032291C>T
    _version : 3
    chrom : 11
    ▼ clingen {2}
      _license : http://bit.ly/2kAtyoH
      caid : CA321211
  
```



ID	437
Chromosome	13
HGVS (GRCh37/hg19)	NC_000013.10:g.32893232_32893233delTT
HGVS (GRCh38/hg38)	NC_000013.11:g.32319095_32319096delTT
HGVS (RNA)	NM_000059.3:c.86_87delTT
HGVS (Protein)	NP_000050.2:p.L29Qfs*4
ClinGen Allele Registry	CA025780



CTGATGGTATGGGGCCAGAGATA  
 AGGTACGGCTGTCACTTACAGAC  
 AGGGCTGGGATAAAAGTCAGGGC  
 CATGGTGCATCTGAATCCTGAGG  
 CAGGTTGGTATCAGGTTACAAAG  
 GGACTGACTCTCTGCGTATTGG

- NG\_008481.4:g.67390A>G
- NM\_000038.5:c.20A>G
- NM\_001127510.2:c.20A>G
- ClinGen: CA16021377
- dbSNP: 1060503296



### Variant ID Sources

ClinGen Allele Registry  
 ID: CA3323211  
 dbSNP ID: rs750020623



...NT G93D (c.278G>A)

Last Reviewed by: [kkrysiak](#) Last Commented On by: [kkrysiak](#)

...GLY93ASP Allele Registry ID: CA351750781

...currently have a Summary.



ClinGen Pathogenicity Calculator

Variants	Gene	Description	Assertion(s)/Tags
CA129946	FGFR3	c.1108G>T (p.Gly370Cys) c.1114G>T (p.Gly372Cys) c.931-462G>T (p.=) c.931-462G>T (p.=) c.1108G>T (p.Gly370Cys) c.1108G>T (p.Gly370Cys) c.*164G>T (p.=)	Pathogenic PM2, PS3, PS4



## *Clinic (EHR) and Lab data flows enabled!*

If a **trusted public variant registry built on a standard specification** existed then...

- EHR & Lab vendors would not need to build custom variant validation, comparison logic. Instead focusing on patient care concerns.
- Secondary variant services and applications (public & private) would further enhance EHR, Lab and research capabilities.
- Variant data is now reliable for Clinical Decision Support
- Discovery accelerates.
- A VRC-like organization would manage challenges and releases.

EHR, Lab, Clinical / Research Apps

Secondary Variant Services & Apps  
(Public & Private)

Trusted Public Allele / Variant Registry  
(ClinGen NCBI EBI)

Variant Representation Specifications

## *What about Steps 2 and 3?*

- Step 2 : Phenotypes and Diseases
  - Terminology gaps, mapping, and hierarchy/ontology must be resolved
  - SNOMED, ICD – needed for EHR
  - Monarch Initiative (OMIM, HPO, Orphanet, DO, NCIT, ...) – needed for Lab Case and Knowledge Repos.
  - Specifications and Services needed.
- Step 3: Variant Knowledge and Case Level Phenotypic Data
  - Interpretation Guidelines – ACMG/AMP, AMP/ASCO/CAP, PharmGKB/CPIC
  - ClinGen Expert Curation – SVI WG, Domain Specific Guidance, MVLD
  - Specifications and Services – GA4GH Var Anno, ClinGen-SEPIO
- (not enough time today...next time?)

## *Related web sites and resources*

- HL7 CG Draft Impl. Guide - <http://hl7.org/fhir/uv/genomics-reporting/index.html>
- ClinGen Website – <http://clinicalgenome.org>
- ClinGen Data Model – <http://dataexchange.clinicalgenome.org/>
- ClinGen Allele Registry – <http://reg.clinicalgenome.org/>
  - Test instance – <http://reg.test.genome.network>
  - API docs - [http://reg.clinicalgenome.org/doc/AlleleRegistry\\_0.12.xx\\_api\\_v2.pdf](http://reg.clinicalgenome.org/doc/AlleleRegistry_0.12.xx_api_v2.pdf)
- SEPIO wiki - <https://github.com/monarch-initiative/SEPIO-ontology/wiki>
- ClinVar – <http://www.ncbi.nlm.nih.gov/clinvar/>
- GA4GH VMC Specification - <https://docs.google.com/document/d/12E8WbQlvfZWk5NrxwLytmymppPby6vsv60RxCeD5wc1E>
  - VMC github project - <https://github.com/ga4gh/vmc>

# Acknowledgements

- ClinGen
  - Heidi Rehm
  - Sandy Aronson
  - Danielle Azzariti
  - Steven Harrison
  - Allele Registry - Ronak Patel
  - Data Exchange WG
- HL7 Clinical Genomics WG
- NCBI ClinVar & Variation Service Team
- VMC – Reece Hart
- GA4GH GKS
- eMERGE EHRI WG