Bidirectional data flow from clinic to lab and back

Lawrence Babb

Genomic Medicine XI

September 2018

BROAD INSTITUTE

ClinGen
Clinical Genome Resource

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

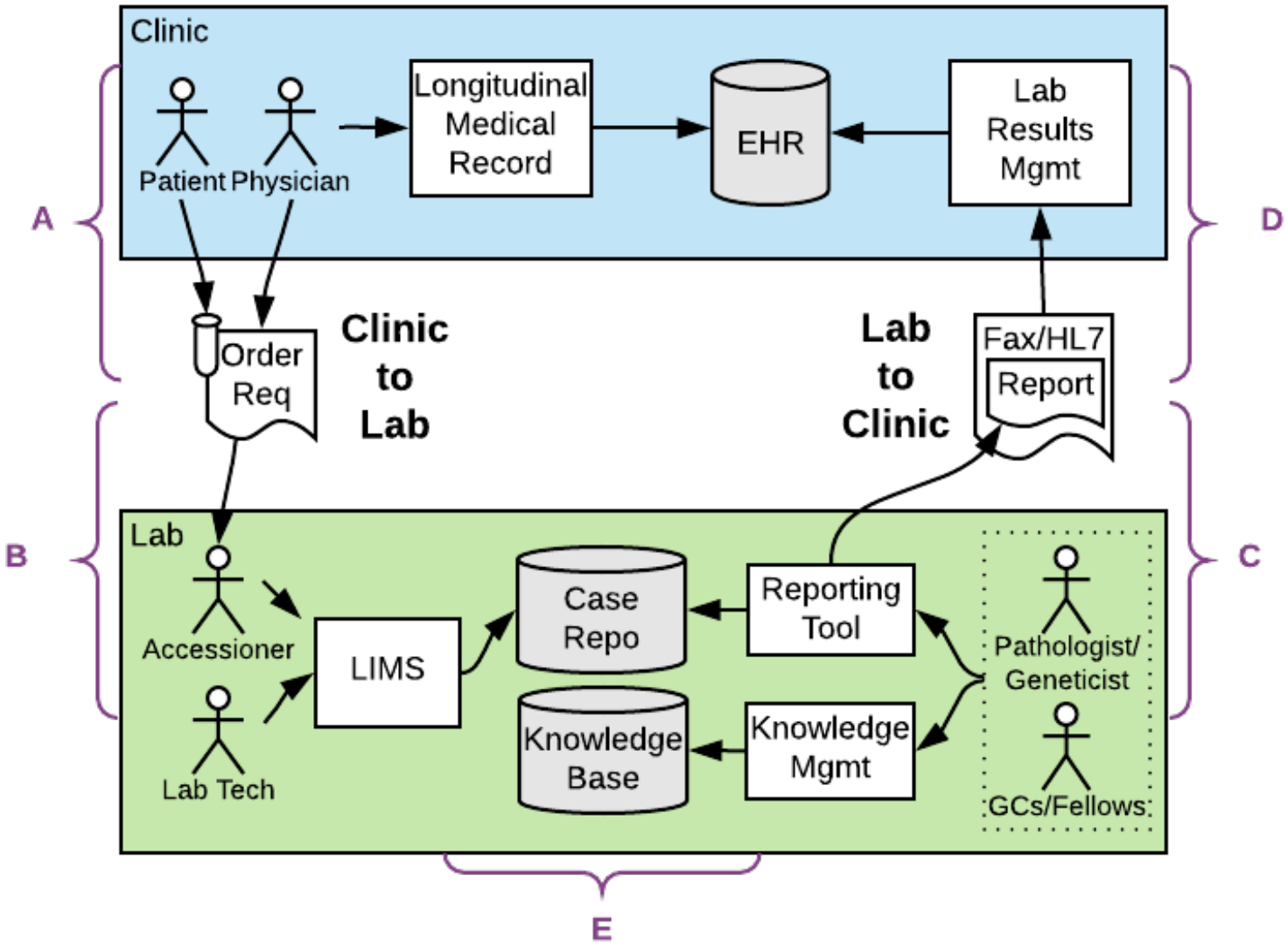
A. Clinic Sends Order to Lab
   - Clinic's ordering physician fills out genetic test requisition and sends along with patient bio-sample to lab.

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 INTERNATIONAL

ClinGen
Clinical Genome Resource

monarch INITIATIVE

EMBL-EBI

Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.

NCBI
National Center for Biotechnology Information

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,…

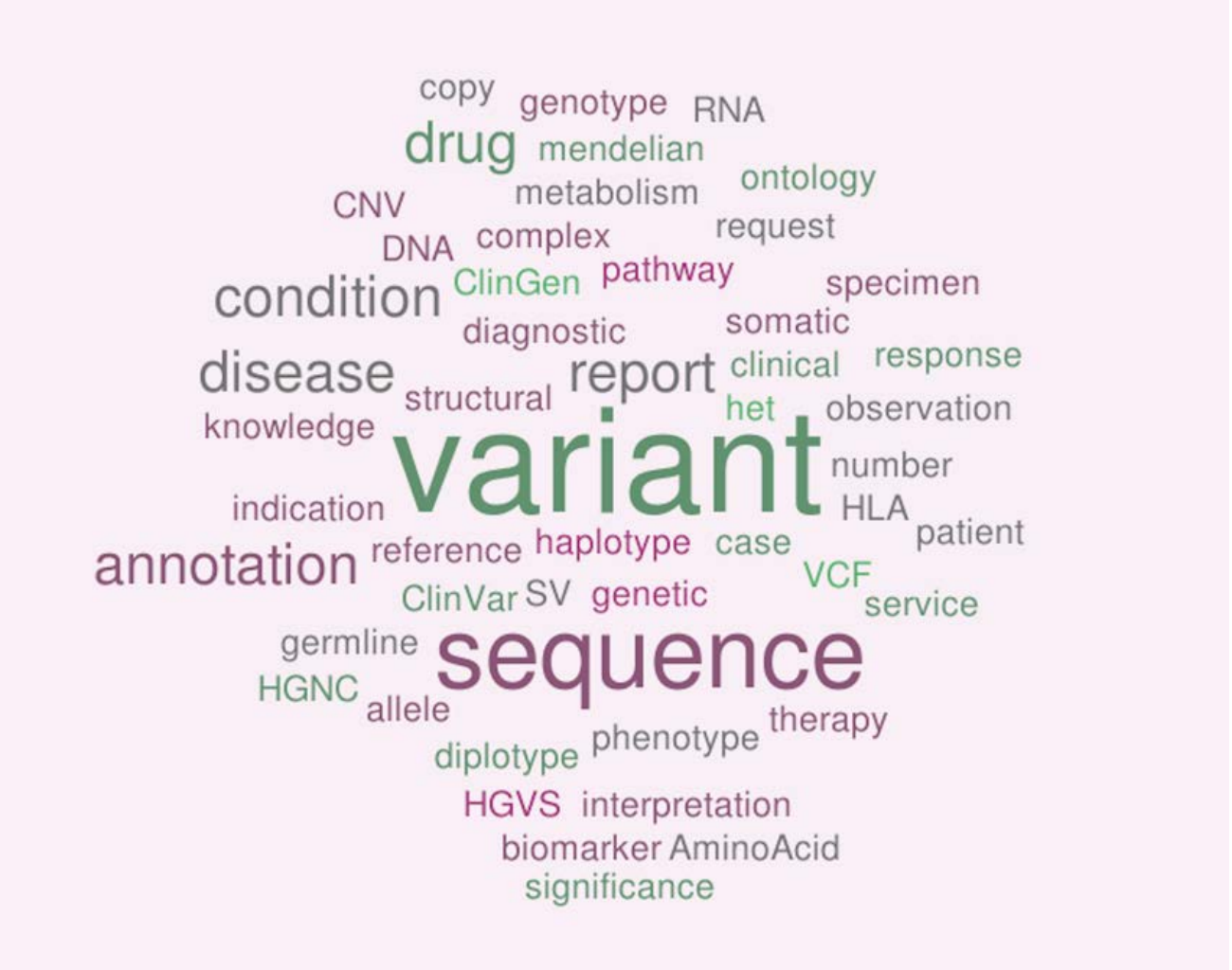
Defining Concepts / Models – Everything is an Observation or Sequence!

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 HUGO Gene Nomenclature Committee 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…
- GRC 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.

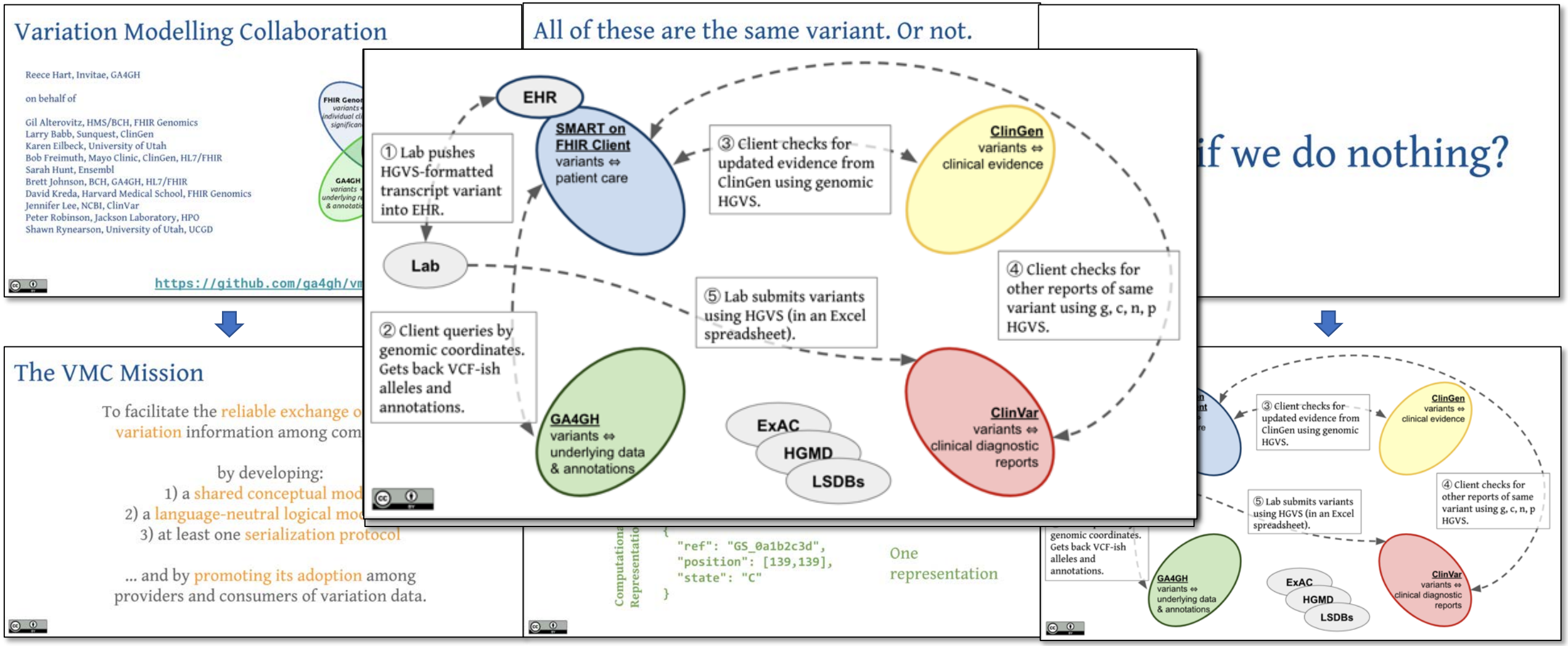
- Variant Representation accuracy in exchange is paramount for clinical systems and computational consistency is essential to associate knowledge, compare findings, and scale discovery.

- 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



R. Hart – Variant Detection 2017 mtg, VMC presentation, Barcelona, Spain June 2017

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 inconsisten 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

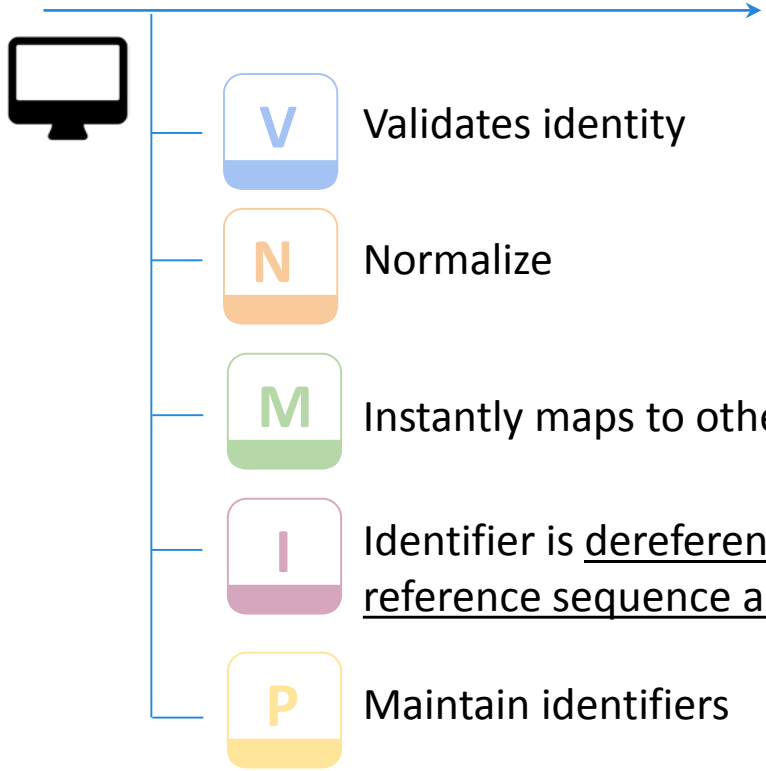
- NCBI supported the concept and called for the establishment of a community organization to set policy (Variant Reference Consortium?)
- Pilot work began on devising improved method for variant normalization within NCBI archives (ClinVar, dbSNP, …)
  - Variation Service - Sequence Position Deletion 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

- HGVS
- VCF
- Interactively provided

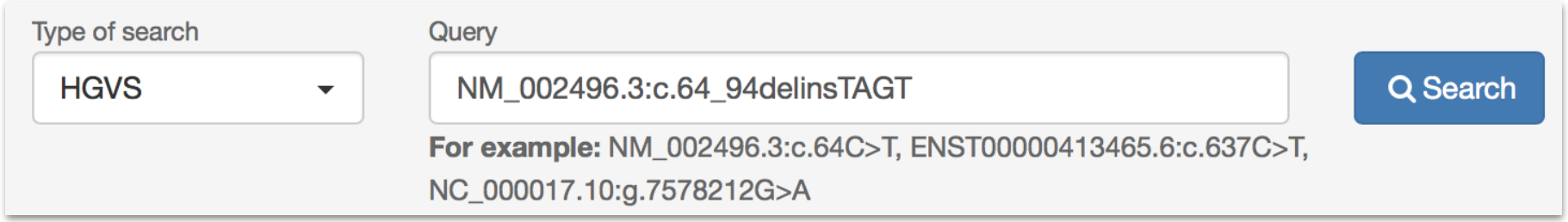
Provides identifier (URI) instantly



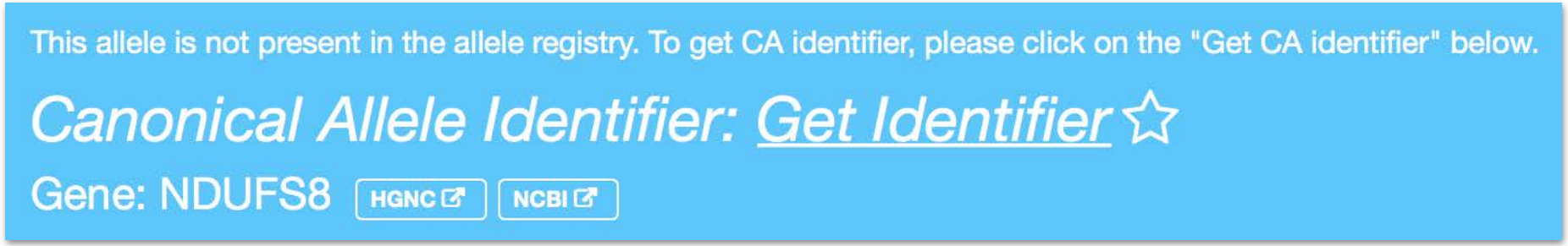
Visit us at
http://reg.clinicalgenome.org

Allele Registry resolves and provides identity to allele

1. Search using HGVS



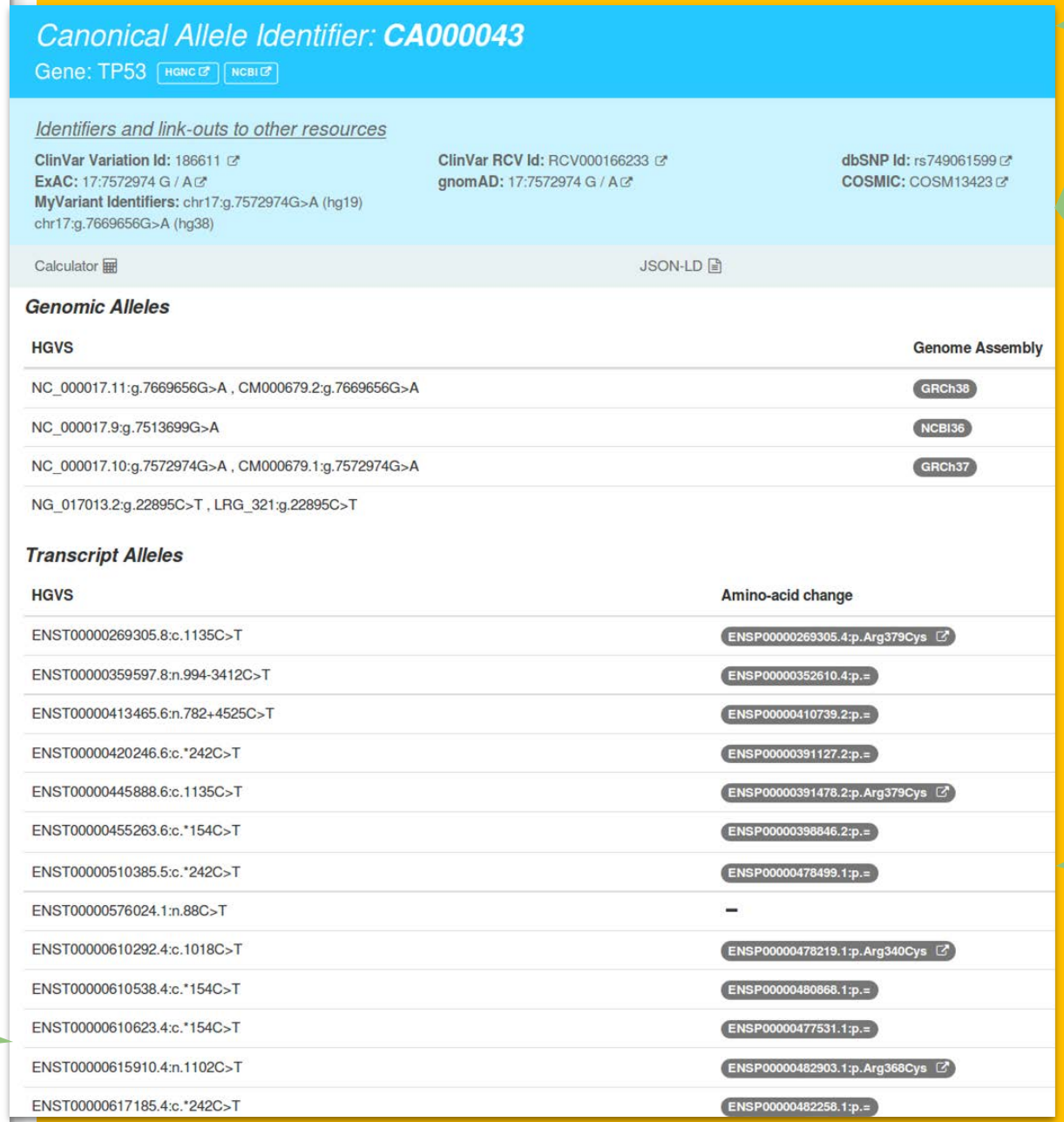
2. Get identifiers if not registered in the registry



> 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



Canonical Allele Identifier: **CA000043**

Gene: TP53 HGNC NCBI

Identifiers and link-outs to other resources

ClinVar Variation Id: 186611
ExAC: 17:7572974 G / A
MyVariant Identifiers: chr17:g.7572974G>A (hg19)
chr17:g.7669656G>A (hg38)

ClinVar RCV Id: RCV000166233
gnomAD: 17:7572974 G / A

dbSNP Id: rs749061599
COSMIC: COSM13423

Calculator JSON-LD

Genomic Alleles

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

Transcript Alleles

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

Canonical Allele Identifier

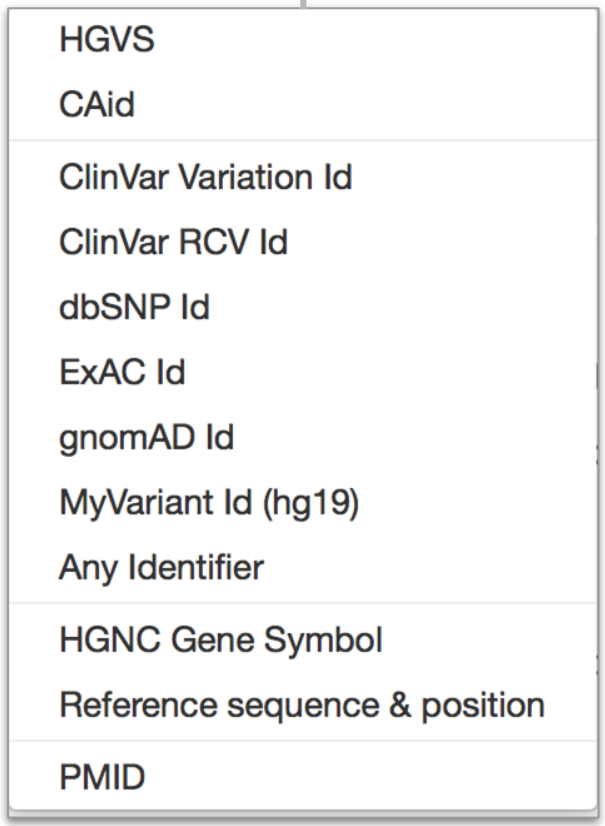
Identifiers and links outs to various resources
ClinVar, dbSNP, ExAC, gnomAD, COSMIC, myVariant.Info

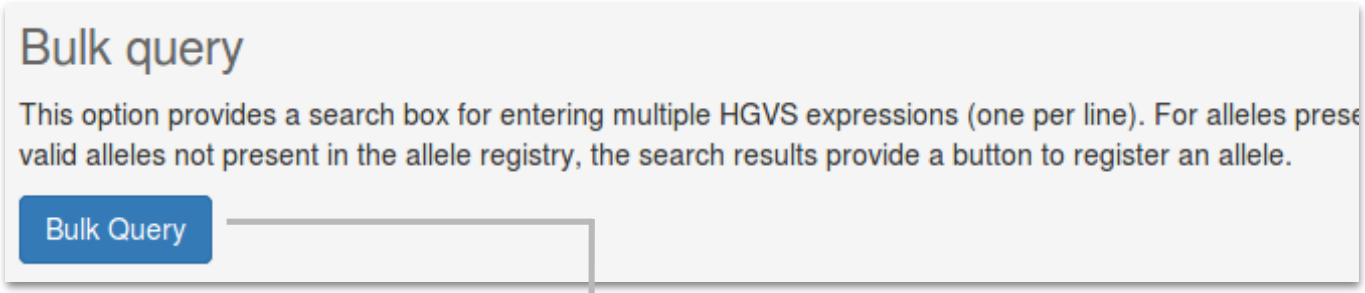
Genomic HGVS

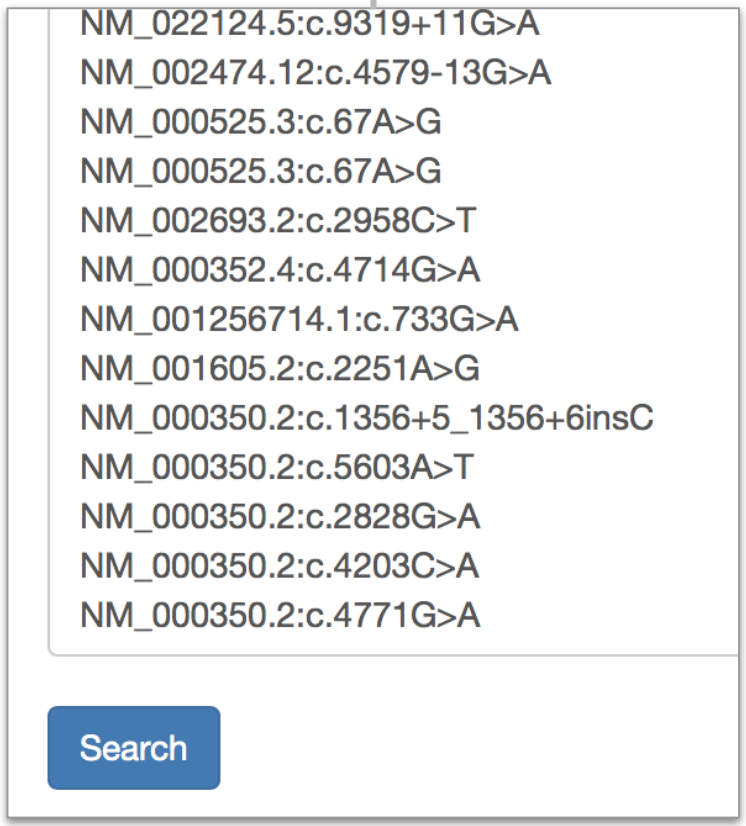
protein HGVS

Transcript HGVS

Allele Registry resolves and provides identity to allele







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

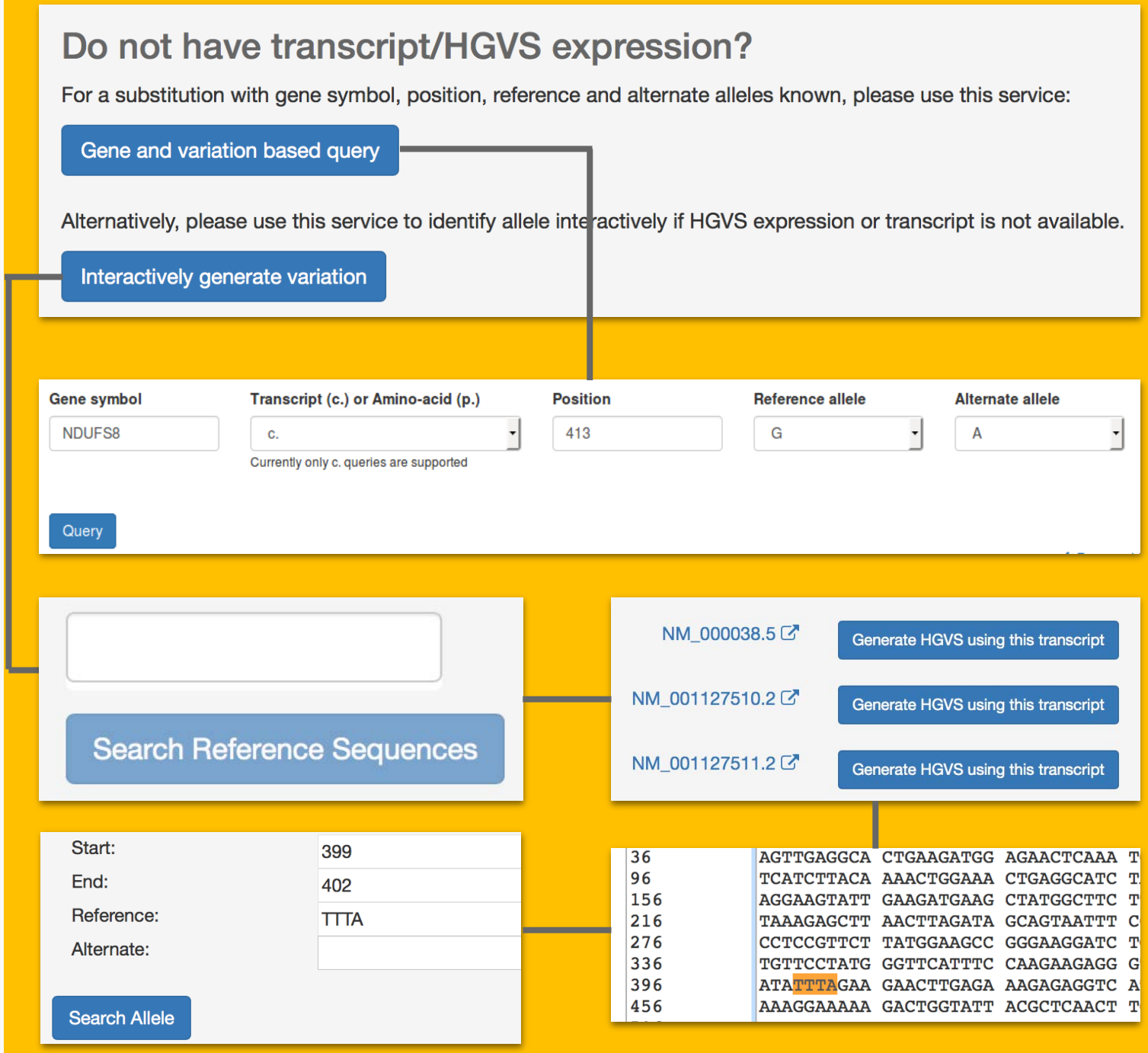
Chr	Mutation (hg38)	CAid	Gene	Transcript	Linkouts
13	g.32319099T>A	*CA387754103*	BRCA2	c.90T>A (p.Asn30Lys) n.288T>A n.90T>A	
13	g.32319099T>C	*CA025976*	BRCA2	c.90T>C (p.Asn30=) n.288T>C n.90T>C	ClinVar dbSNP ExAC gnomAD
13	g.32319101_32319115del	*CA10579445*	BRCA2	c.92_106del (p.Trp31_Leu35del) n.290_304del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

Chromosome

Canonical Allele Identifiers

Transcript and amino-acid variations

Query ClinGen Allele Registry with partial information about variation

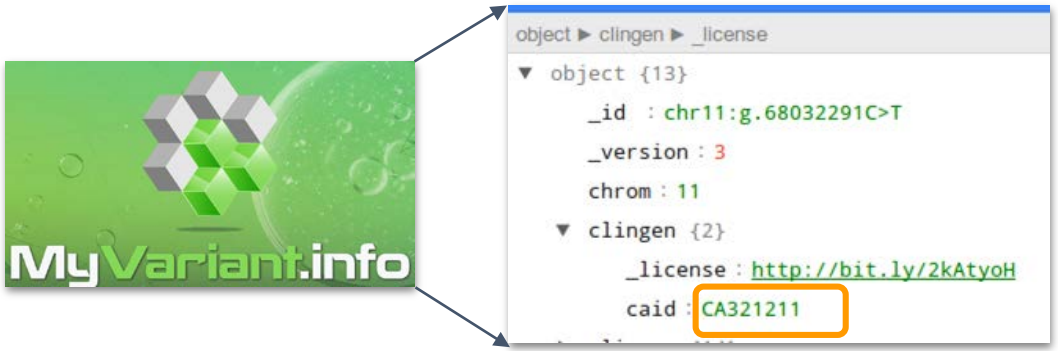


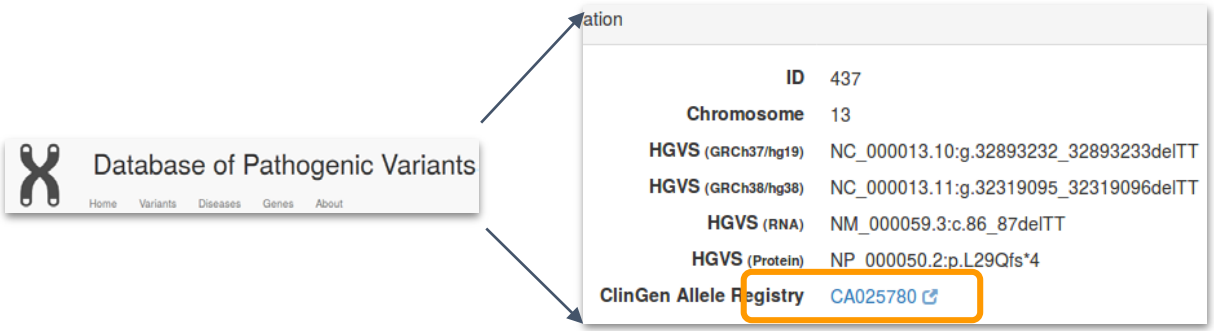
R. Patel – ClinGen Allele Registry Presentation – ClinGen Consortium Call – Aug 17, 2018

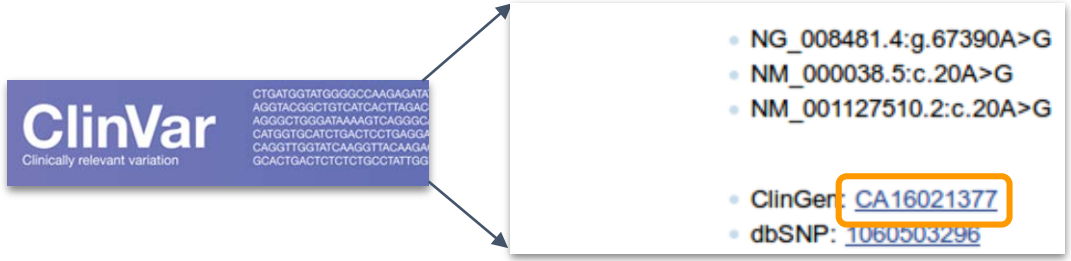
Allele Registry services are accessible through high availability and fast REST-APIs

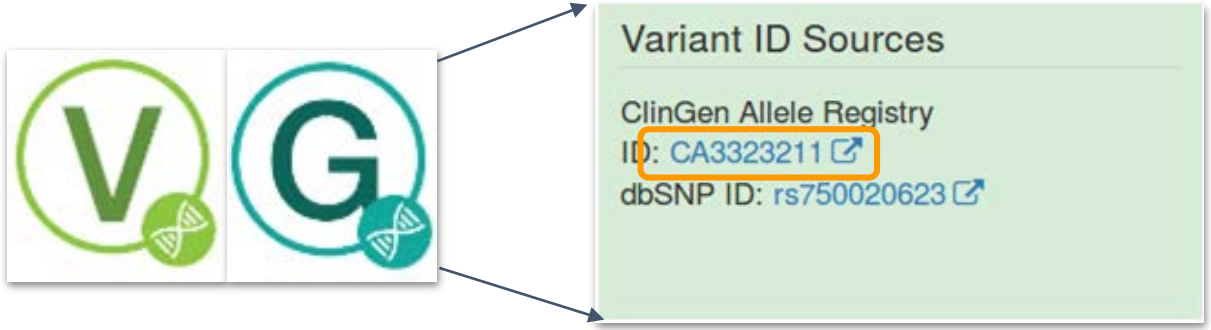
- Simple, well-documented REST-APIs [Backward compatible]
- Simple GET/PUT/POST requests make it easy to integrate
- Current registration/query bandwidth: 1K-50K variants per second

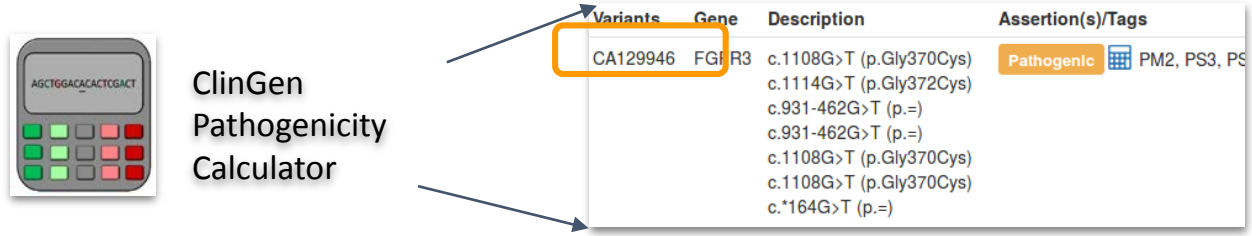
Resources using canonical allele identifiers







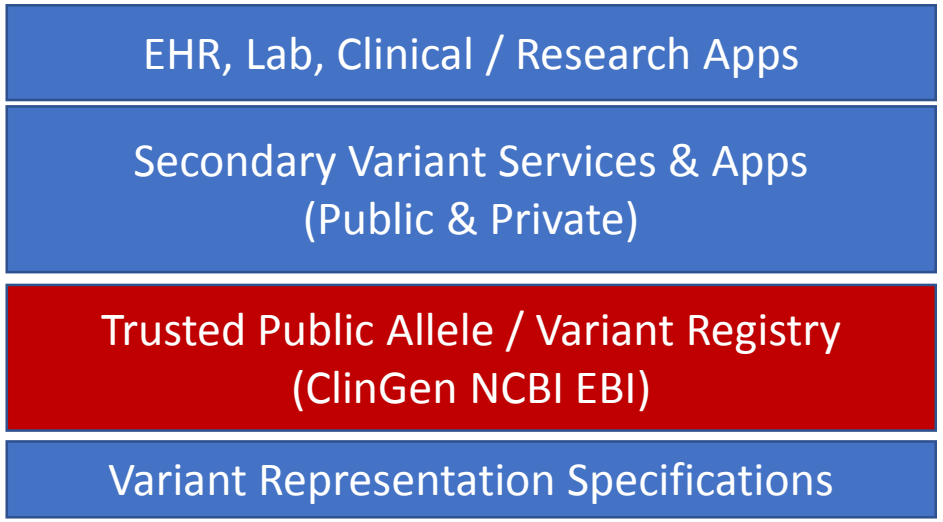




Clinic (EHR) and Lab data flows enabled!

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.



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 Speicific 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
- 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/12E8WbQlvfZWk5NrxwLytmympPby6vsv60RxCeD5wc1E
  - 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