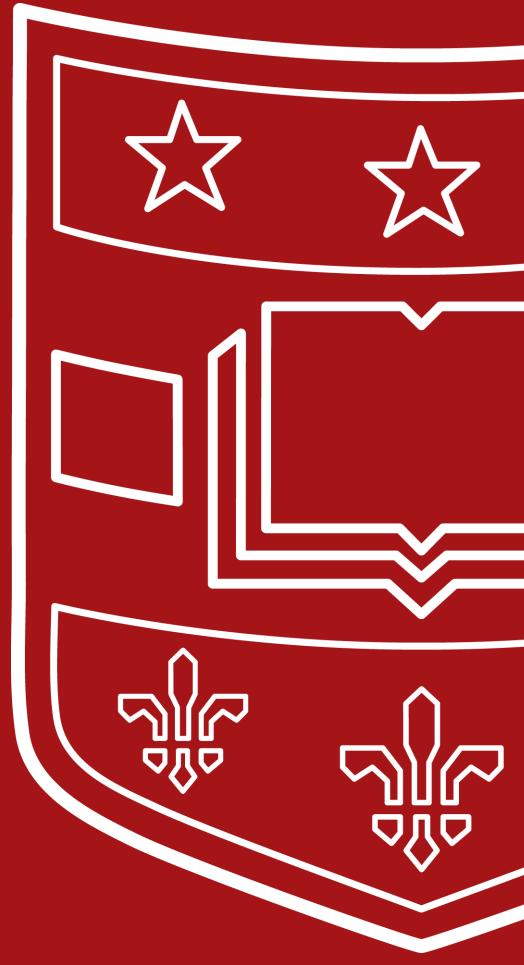


Knowledge standards drive clinical interpretation of genomic variants in cancers

Alex H Wagner, PhD
December, 2019 - UZH Irchel

 Washington University in St. Louis





Introduction



Precision Oncology



Tumor DNA



Clinical Genomic Variant Report



Patient Care



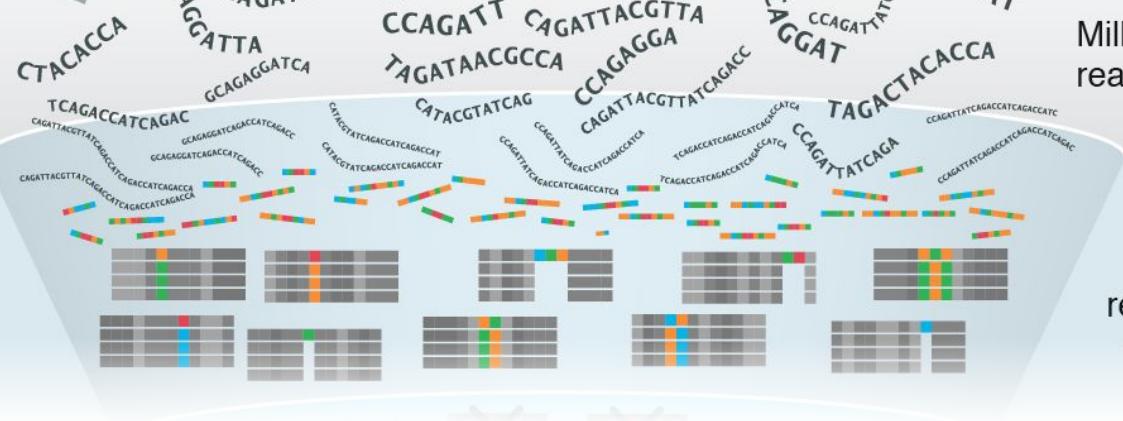
Precision Oncology Bottleneck

1. Data Production



Millions of raw sequence reads are produced for a patient tumor.

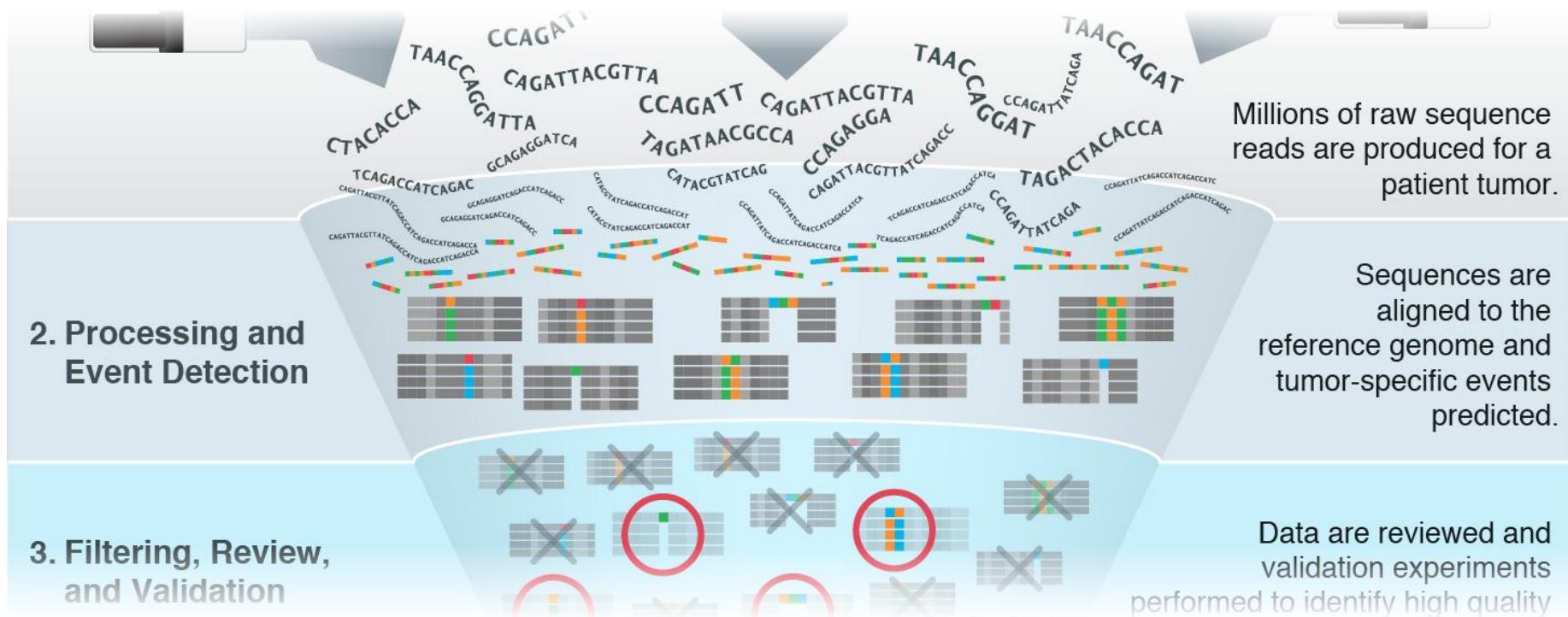
2. Processing and Event Detection



Sequences are aligned to the reference genome and tumor-specific events predicted.



Precision Oncology Bottleneck



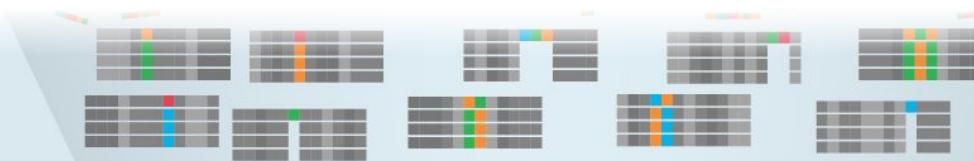
@handlerwagner

Good BM, Ainscough BJ, McMichael JF, Su AI†, Griffith OL†. 2014. Genome Biology. 15(8):438.



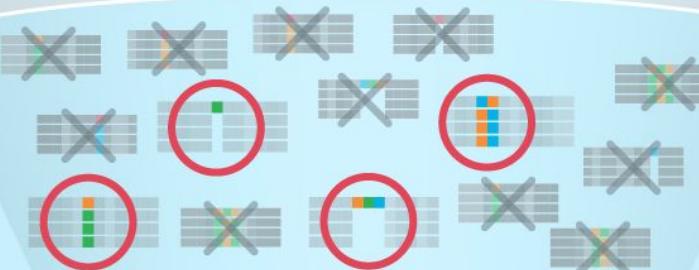
Precision Oncology Bottleneck

2. Processing and Event Detection



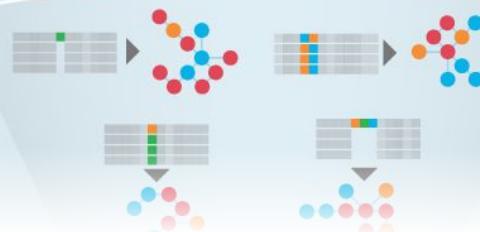
Data are aligned to the reference genome and tumor-specific events predicted.

3. Filtering, Review, and Validation



Data are reviewed and validation experiments performed to identify high quality events.

4. Annotation and Functional Prediction

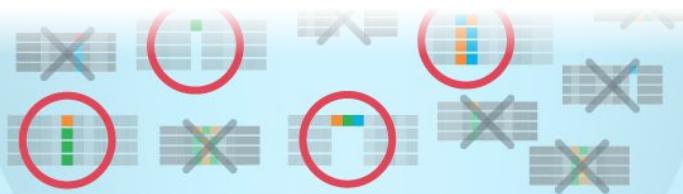


Events are annotated and scored in an effort to predict events of functional significance.



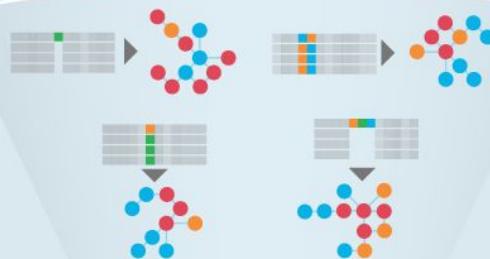
Precision Oncology Bottleneck

3. Filtering, Review, and Validation



Data are reviewed and validation experiments performed to identify high quality events.

4. Annotation and Functional Prediction



Events are annotated and scored in an effort to predict events of functional significance.

5. Interpretation and Report Generation



A genome analyst attempts to interpret, prioritize, and summarize functionally significant events in the context of published literature, clinical trials, and a



Precision Oncology Bottleneck

4. Annotation and Functional Prediction



Events are annotated and scored in an effort to predict events of functional significance.

5. Interpretation and Report Generation



A genome analyst attempts to interpret, prioritize, and summarize functionally significant events in the context of published literature, clinical trials, and a multitude of knowledgebases.

6. Clinical Application

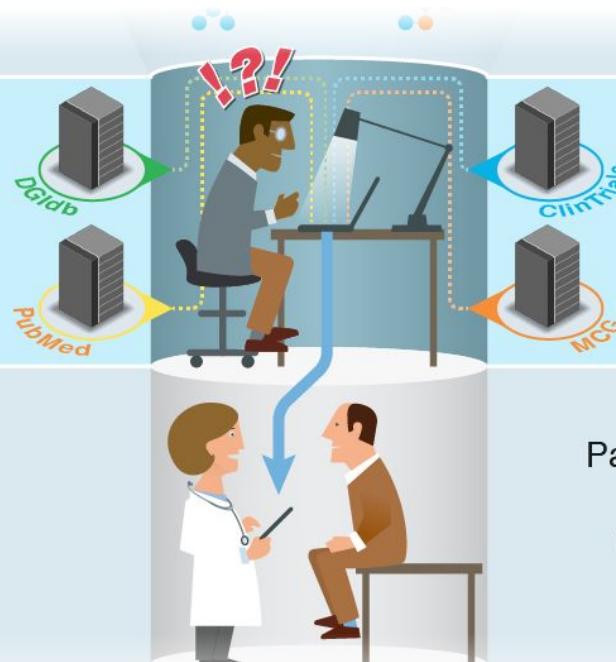


Pathologists and oncologists evaluate the significance of potentially clinically actionable events, and incorporate their research into patient care.



Precision Oncology Bottleneck

5. Interpretation and Report Generation



A genome analyst attempts to interpret, prioritize, and summarize functionally significant events in the context of published literature, clinical trials, and a multitude of knowledgebases.

6. Clinical Application

Pathologists and oncologists evaluate the significance of potentially clinically actionable events, and incorporate their research into patient care.



Precision Oncology Bottleneck



@handlerwagner

Good BM, Ainscough BJ, McMichael JF, Su AI†, Griffith OL†. 2014. Genome Biology. 15(8):438.



Precision Oncology Bottleneck



How do we alleviate this bottleneck?



Roadmap from Evidence to Action

Evidence

Publications

Genomic Information Databases

Structured Knowledge

Clinical Interpretations of Variants

Consensus Knowledge

Standardized Clinical Interpretations of Variants

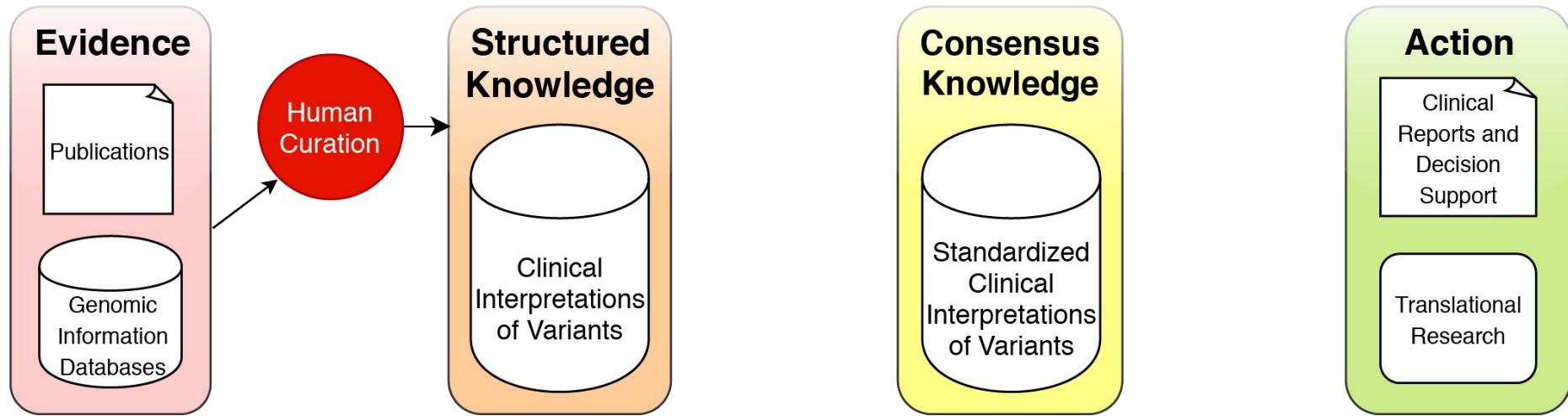
Action

Clinical Reports and Decision Support

Translational Research



Roadmap from Evidence to Action





Alleviating the Curation Burden with CIViC

Commentary | [Open Access](#) | Published: 31 January 2017

CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer

Malachi Griffith , Nicholas C Spies [...] Obi L Griffith 

Nature Genetics **49**, 170–174 (2017) | [Download Citation](#) 

Public contributions, open discussion, curation standards and expert review

Researchers, clinicians,
patient advocates and
others

Public domain (CC0)
license



Content provenance
and creator
acknowledgement

Structured data and
APIs

No fees, anonymous access



Data and Knowledge Production

Millions of raw sequence reads are produced for a patient tumor.



TAACCA
GATC
CCAG
GATAACGCCA
TAACCA
GAGGAT

Sequences are aligned to the reference genome and tumor-specific events predicted.



Data are reviewed and validation experiments performed to identify high quality events.

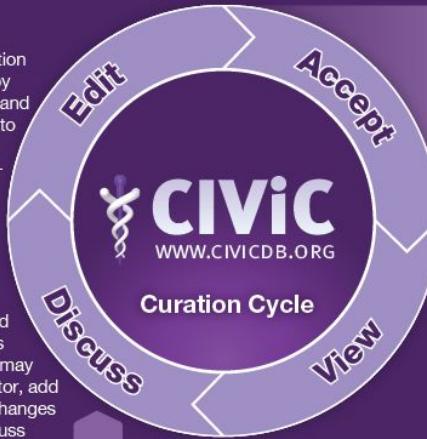


Events are annotated and scored in an effort to predict events of functional significance.



CIViC Curation

Crowdsourced curation efforts, moderated by experts in oncology and bioinformatics, help to build a knowledge-base of clinical interpretations of variants in cancer, describing the therapeutic, prognostic, diagnostic, and predisposing relevance of inherited and somatic variants of all types. Anyone may sign up to be a curator, add evidence, suggest changes to records, and discuss ongoing curation efforts.

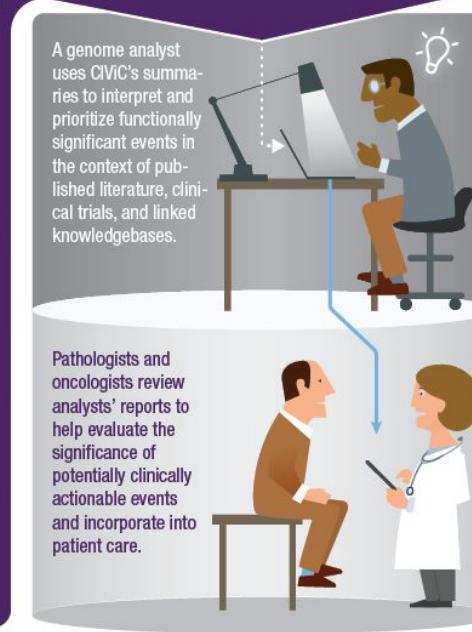


Add New Evidence

Review and Discuss Edits

Research Gene, Variant, & Evidence Summaries

A genome analyst uses CIViC's summaries to interpret and prioritize functionally significant events in the context of published literature, clinical trials, and linked knowledgebases.



Pathologists and oncologists review analysts' reports to help evaluate the significance of potentially clinically actionable events and incorporate into patient care.



Last Modified by kkrysiak

Last Reviewed by obigriffith

Last Commented On by obigriffith

Aliases: RS113488022 and VAL600GLU**Allele Registry ID:** CA123643

BRAF V600E has been shown to be recurrent in many cancer types. It is one of the most widely studied variants in cancer. This variant is correlated with poor prognosis in certain cancer types, including colorectal cancer and papillary thyroid cancer. The targeted therapeutic dabrafenib has been shown to be effective in clinical trials with an array of BRAF mutations and cancer types. Dabrafenib has also shown to be effective when combined with the MEK inhibitor trametinib in colorectal cancer and melanoma. However, in patients with TP53, CDKN2A and KRAS mutations, dabrafenib resistance has been reported. Ipilimumab, regorafenib, vemurafenib, and a number of combination therapies have been successful in treating V600E mutations. However, cetuximab and panitumumab have been largely shown to be ineffective without supplementary treatment.

Variant Type:

Missense Variant

Assertions: Show rejected: **HGVS Expressions:**

NM_00433.4:c.1799T>A, NP_004324.2:p.Val600Glu,
NC_000007.13:g.140453136A>T, and
ENST00000288602.6:c.1799T>A

ClinVar ID:

13961

CIViC Variant Evidence Score:

1019

Evidence for V600E 161 total items (showing 154)

Variant Summary

Variant Talk

**Representative Variant Coordinates**

Ref. Build: GRCh37 Ensembl Version: 75

Chr.	Start	Stop	Ref. s	Var. Bases
7	140453136	140453136	A	T

Transcript

ENST00000288602.6

[Edit Coordinates](#)**ClinVar ID**

13961

ClinVar Clinical Significance

Pathogenic

COSMIC ID (v68)

COSM476

dbSNP RSID

rs113488022

HGVIS ID

chr7:g.140453136A>T

SnpEff Effect

missense variant

SnpEff Impact

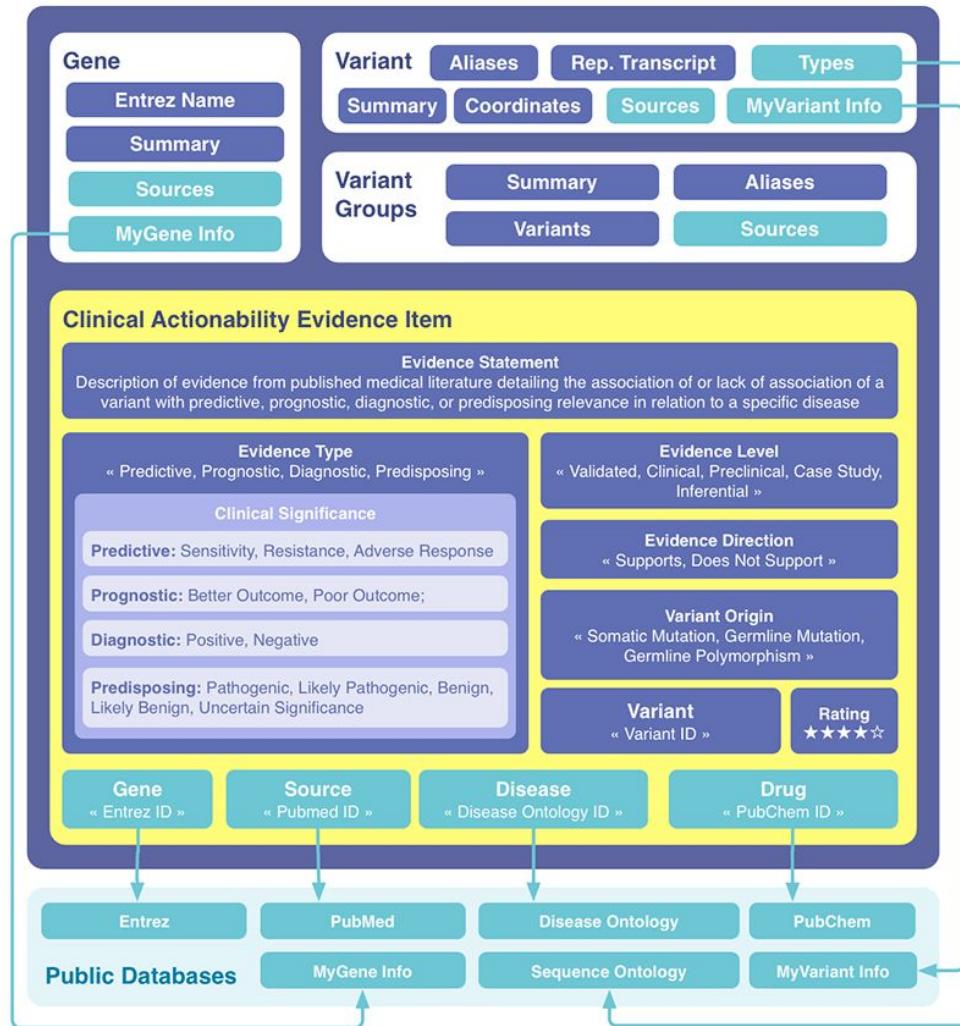
MODERATE

gnomAD Adj. AF

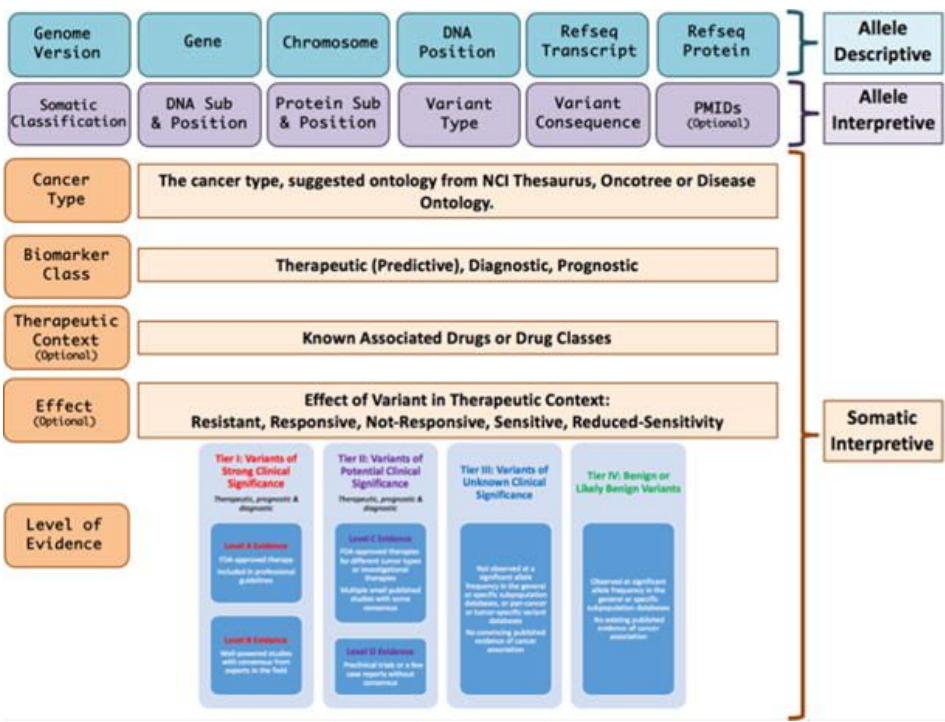
0

[View MyVariant.info Details](#)

EID	DIS	DRUGS	DESC	EL	ET	ED	CS	VO	ER	☰
7612	Colorectal Cancer	Binimetinib, Encorafenib, Cetuximab (Co...)							5	
1409	Skin Melanoma	Vemurafenib							5	
3017	Lung Non-small Cell Carcinoma	Dabrafenib, Trametinib (Combination)							4	
102	Thyroid Gland Papillary Carcinoma	N/A							5	

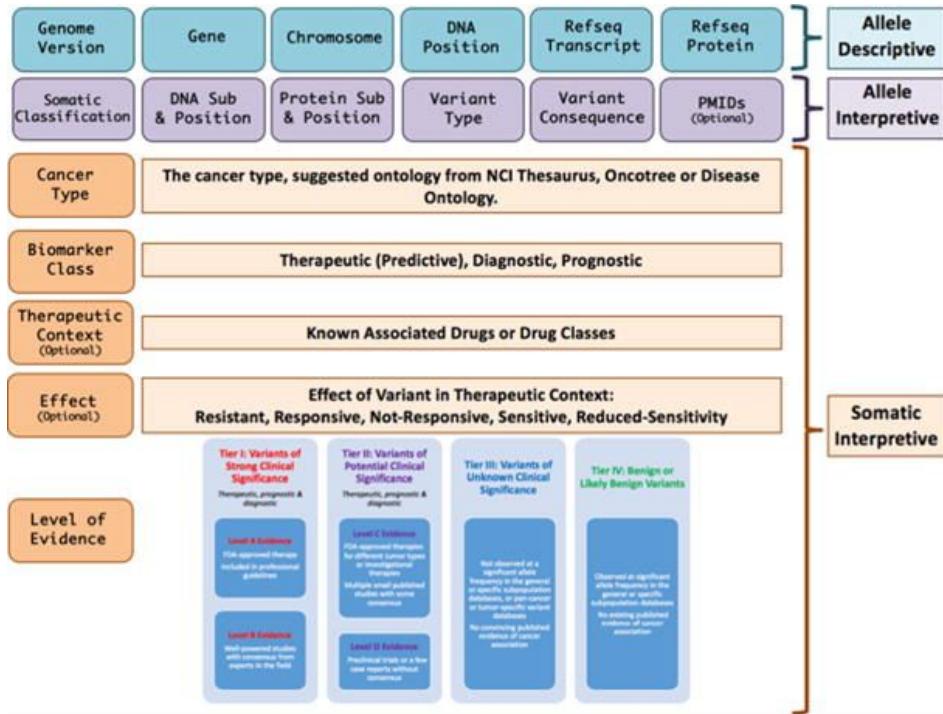
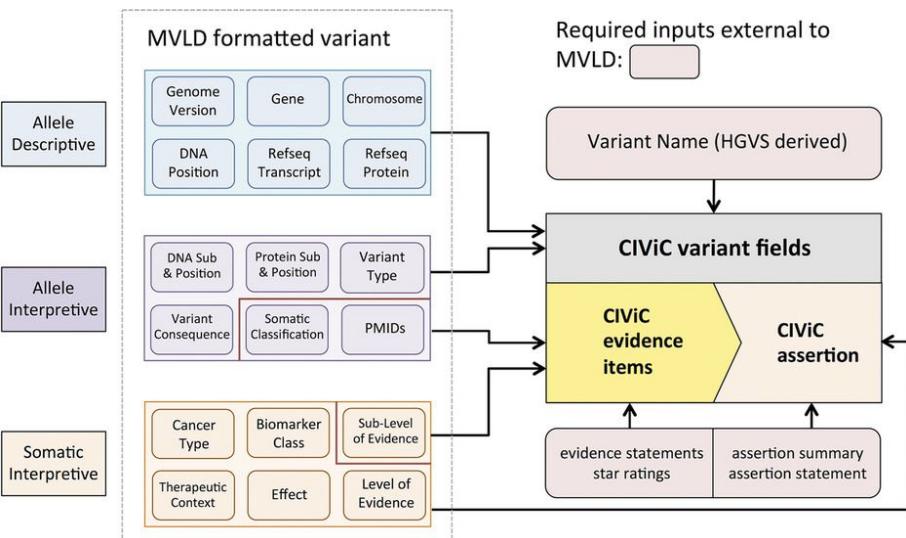


ClinGen Somatic WG: MVLD Guidelines



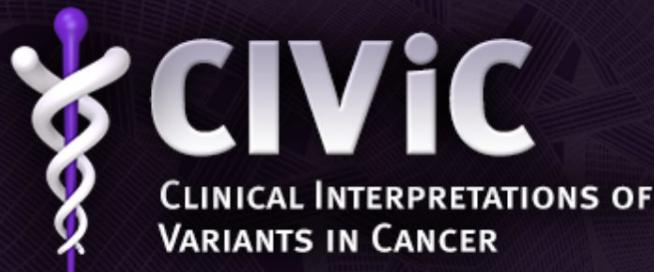
Madhavan S, et al. *Pac. Symp. Biocomput.* 2018

ClinGen Somatic WG: MVLD Guidelines



Madhavan S, et al. *Pac. Symp. Biocomput.* 2018

Danos AM, et al. *Hum. Mut.* 2018



About Participate Community Help FAQ [Sign In/Sign Up](#)

[Go to Genes & Variants](#)

Go!

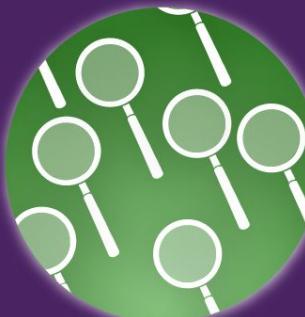
BROWSE

SEARCH

ACTIVITY



Discover supported
clinical interpretations of
mutations related to
cancer.



Participate with
colleagues to add
variants and support for
cancer-related
mutations.

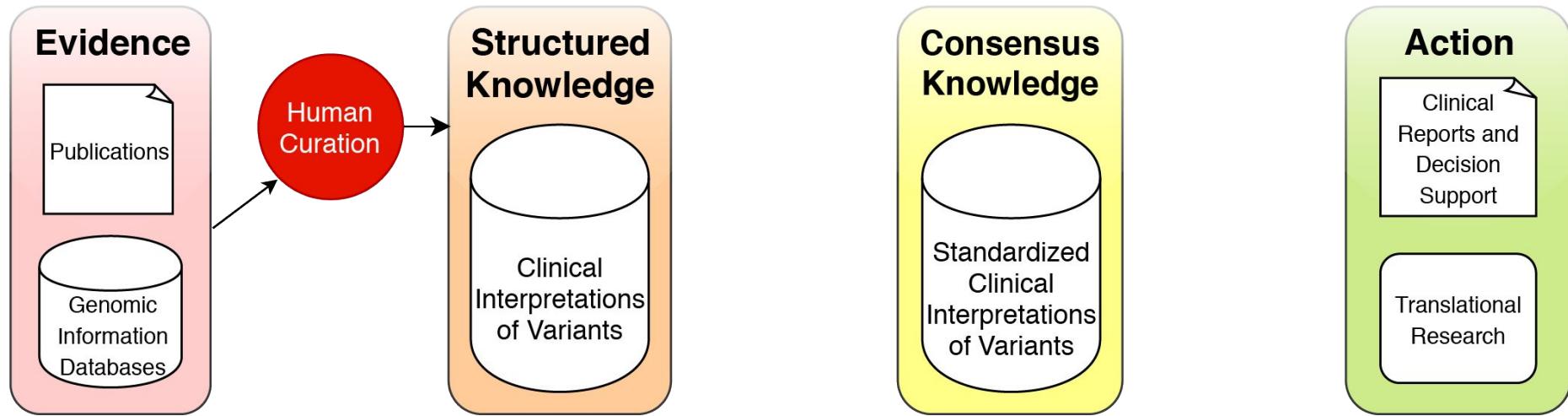
<https://civicdb.org>



@handlerwagner



Roadmap from Evidence to Action





Problem: CIViC is a Silo in a Diverse Knowledge Ecosystem



Variant Interpretation Knowledge is Siloed

Established Interpretation Knowledgebases

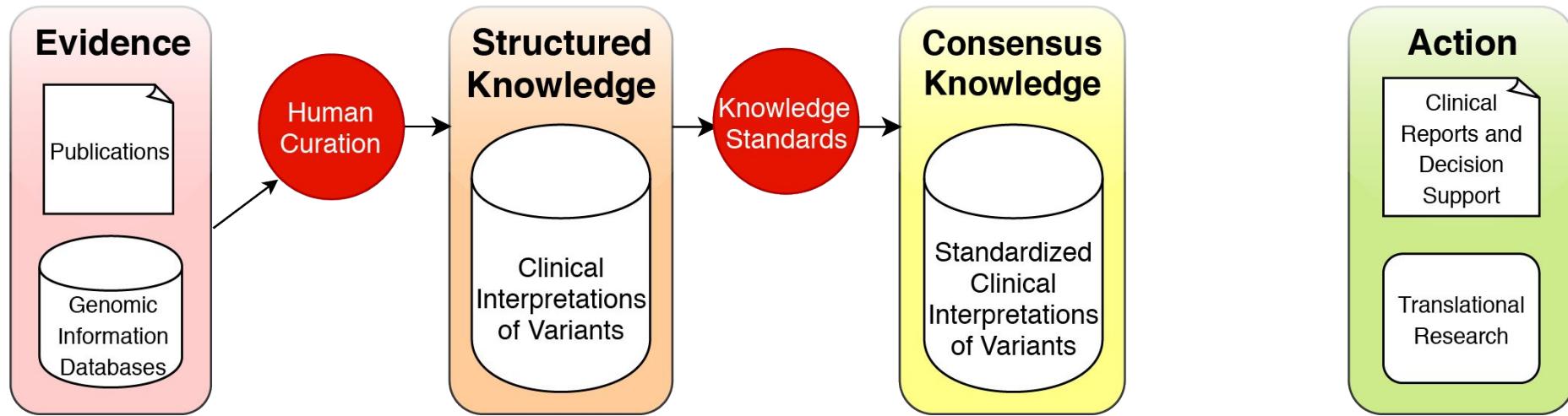
- [CIViC \(WashU\)](#)
- [Cancer Genome Interpreter \(Barcelona\)](#)
- [OncoKB \(MSKCC\)](#)
- [PMKB \(Cornell\)](#)
- [JAX-Clinical Knowledgebase \(Jackson lab\)](#)
- [CanDL \(Ohio State\)](#)
- [COSMIC \(Sanger\)](#)
- [MyCancerGenome \(Vanderbilt\)](#)
- [KnowledgeBase for Precision Oncology \(MD Anderson\)](#)
- [BRCA Exchange](#)
- [Gene Drug Knowledge Database](#)
- [PharmGKB](#)
- [ClinVar](#)
- [Pecan](#)

Additionally...

- ad hoc “databases” (acad. centers / hospitals)
- Industry (FoundationMedicine, Agilent, etc)



Roadmap from Evidence to Action





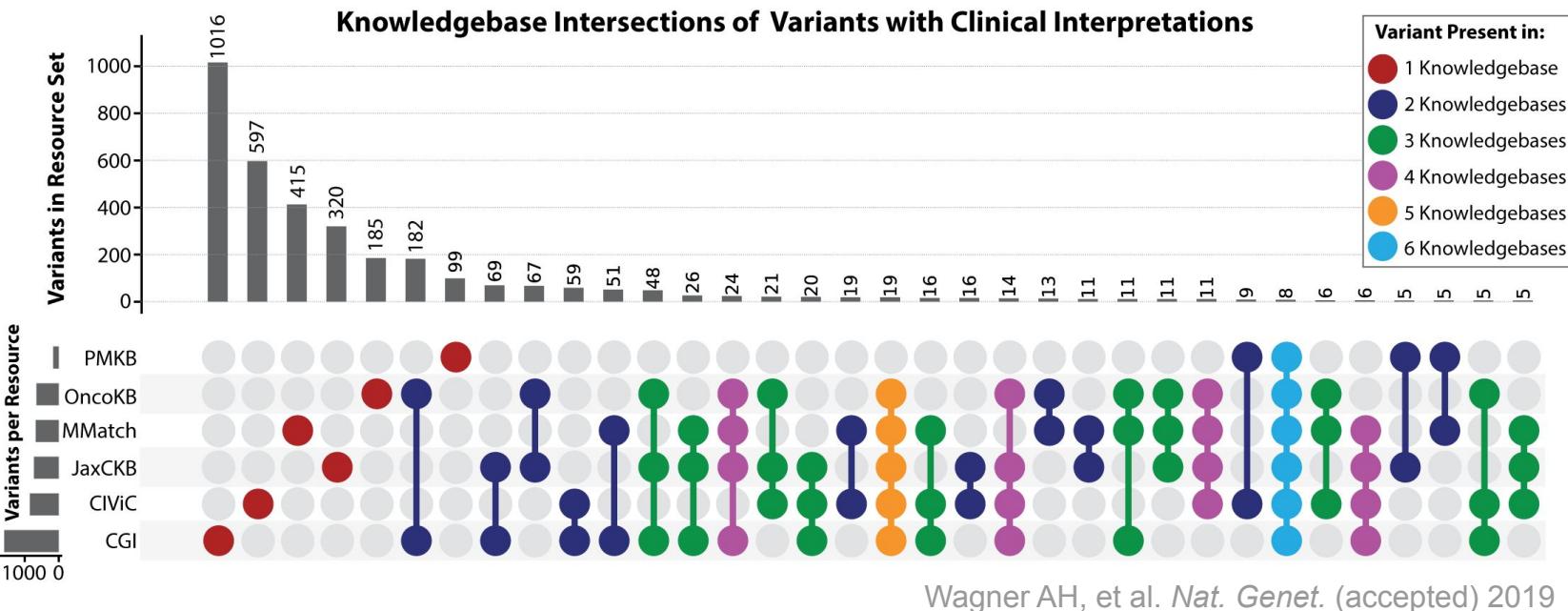
- Year Started: 2016
- Country: Global – USA, Barcelona, UK
- Institutions: WashU, MSKCC, DFCI, OHSU, IRB Barcelona, Cornell, Jackson Labs
 - Consortium of Clinical Interpretations KBs: OncoKB, Jax-CKB, CIViC, PMKB, CGI, ...
- Mission:
 - Global integration of knowledgebases for clinical interpretation of cancer variants
- Clinical Focus
 - Ultimate goal – **consistent, actionable clinical reports**



Variant Interpretations are **Heterogeneous**



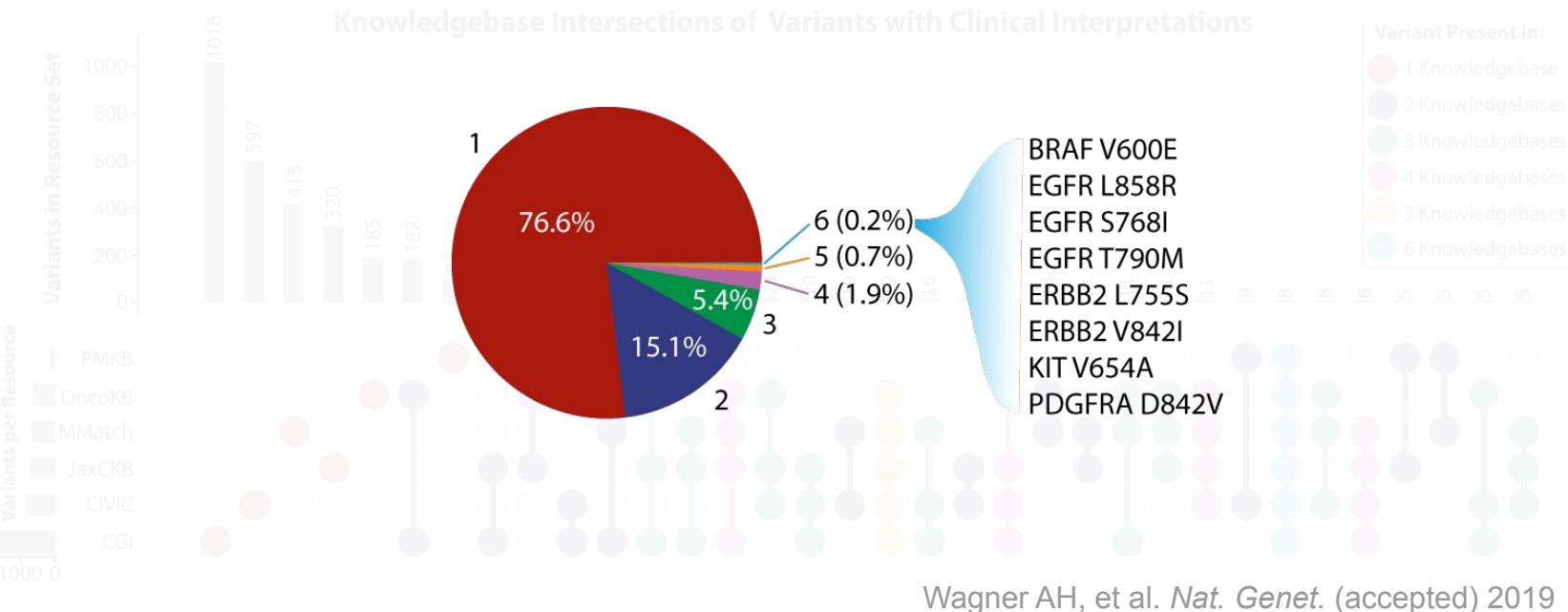
Variant Representation Across Knowledgebases



Wagner AH, et al. *Nat. Genet.* (accepted) 2019



Variant Representation Across Knowledgebases





Poor overlap of variants creates
inconsistent clinical interpretations

This can be addressed by
aggregating curated knowledge





Variant Interpretations are Structurally Disparate

Diversity in Structure of Variant Interpretations



Gene	BRAF	BRAF (Entrez ID: 673)	BRAF
Isoform	ENST00000288602 / NM_004333.4	ENST00000288602.6	ENST00000288602
Variant	V600E (?????)	V600E (chr7:g.140453136A>T)	V600E (7:140453136:140453136)
Disease	Melanoma	Skin Melanoma (DOID:8923)	Tumor: Melanoma / Tissue: Skin
Drug	Dabrafenib	Dabrafenib + Trametinib	?
Clinical Significance	Known Effect: Sensitive	Supports Sensitivity	?
Evidence Level	2B	A - Validated	Tier 1
Statement	Approved Indications: Dabrafenib is FDA-approved for BRAF V600E mutant unresectable or metastatic melanoma	Open-label, randomized phase 3 trial with 704 patients with metastatic melanoma with a BRAF V600 mutation. Patients were randomized Various B-Raf inhibitors (Vemurafenib, Dabrafenib) have been FDA approved for melanoma therapy in certain settings.



Inconsistent Variant Normalization

ERBB2 (NP_004439.2) reference protein sequence



Non-standard HGVS: ERBB2 p.E770delinsEAYVM



Standard HGVS: NP_004439.2:p.Y772_A775dup



Wagner AH, et al. *Nat. Genet.* (accepted) 2019



Poor overlap of variants creates
inconsistent clinical interpretations

We can aggregate, but...

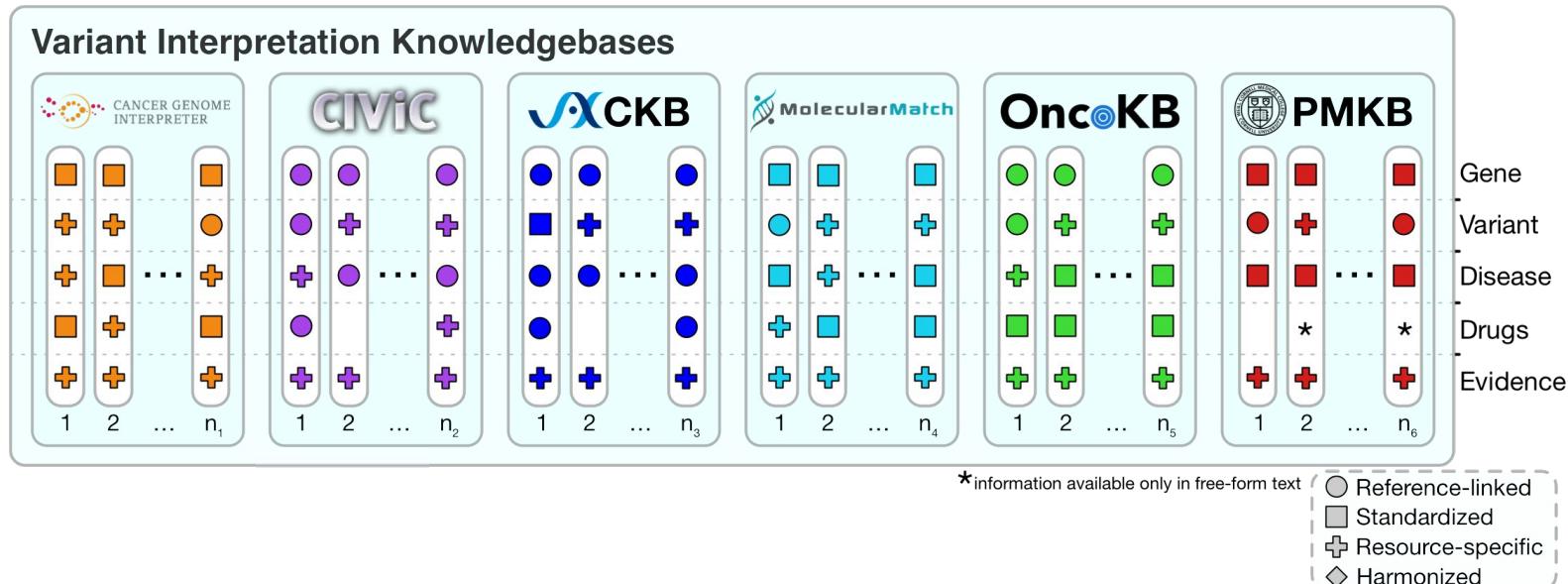


Clinical variant interpretation
knowledgebases are not interoperable

This can be addressed by harmonizing interpretations



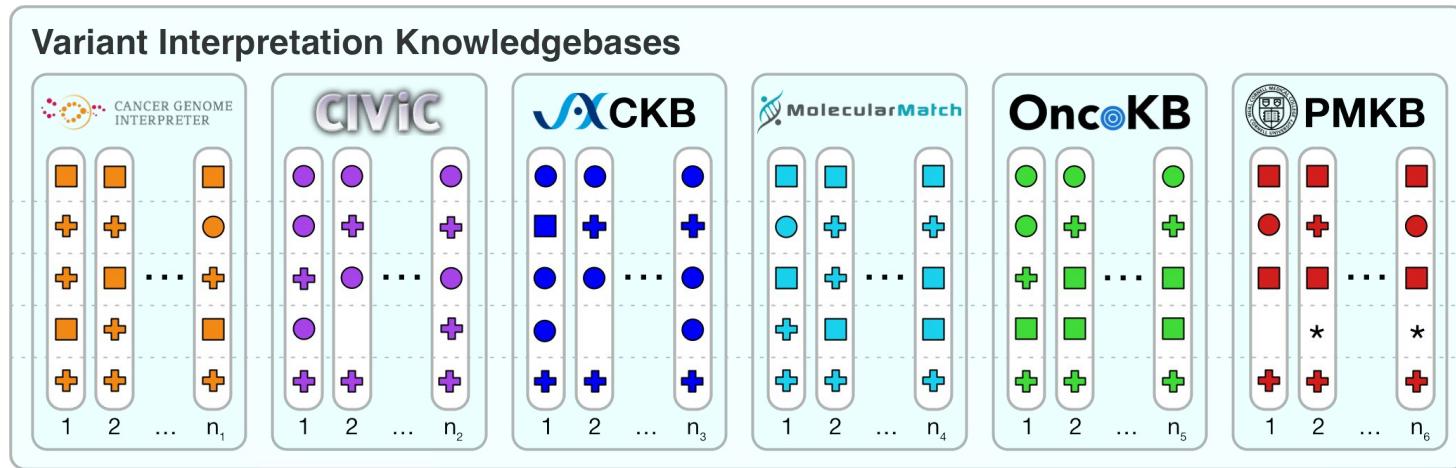
Structure of a Variant Interpretation



Wagner AH, et al. *Nat. Genet.* (accepted) 2019



Structure of a Variant Interpretation



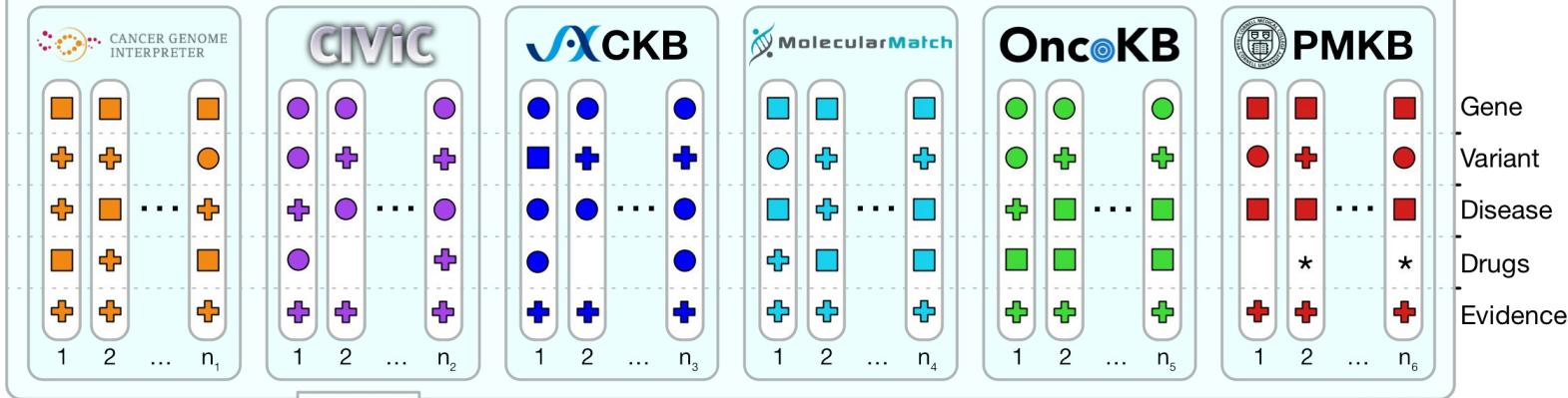
Wagner AH, et al. *Nat. Genet.* (accepted) 2019

AMP/ASCO/CAP Variant Evidence Guidelines							
Evidence Level	Defining Characteristics	CIViC	OncoKB	JAX-CKB	CGI	MMatch	PMKB
Level A (Tier I)	<i>Evidence from professional guidelines or FDA-approved therapies relating to a biomarker and disease.</i>	Level A	Level 1 / 2A /R1	Guideline / FDA Approved	Clinical Practice	Level 1A	Tier 1
Level B (Tier I)	<i>Evidence from clinical trials or other well-powered studies in clinical populations, with expert consensus.</i>	Level B	Level 3A	Phase III	Clinical Trials III-IV	Level 1B	
Level C (Tier II)	<i>Evidence for therapeutic predictive markers from case studies, or other biomarkers from several small studies. Also evidence for biomarker therapeutic predictions for established drugs for different indications.</i>	Predictive Level C	Level 2B, Level 3B	Clinical Study/ Phase I / Phase II	Clinical Trials I-II, Case Reports	Level 2C	Tier 2
Level D (Tier II)	<i>Preclinical findings or case studies of prognostic or diagnostic biomarkers. Also includes indirect findings.</i>	Non-predictive Level C / Level D / Level E	Level 4	Phase 0, Pre-clinical	Pre-clinical Data	Level 2D	

Wagner AH, et al. *Nat. Genet.* (accepted) 2019



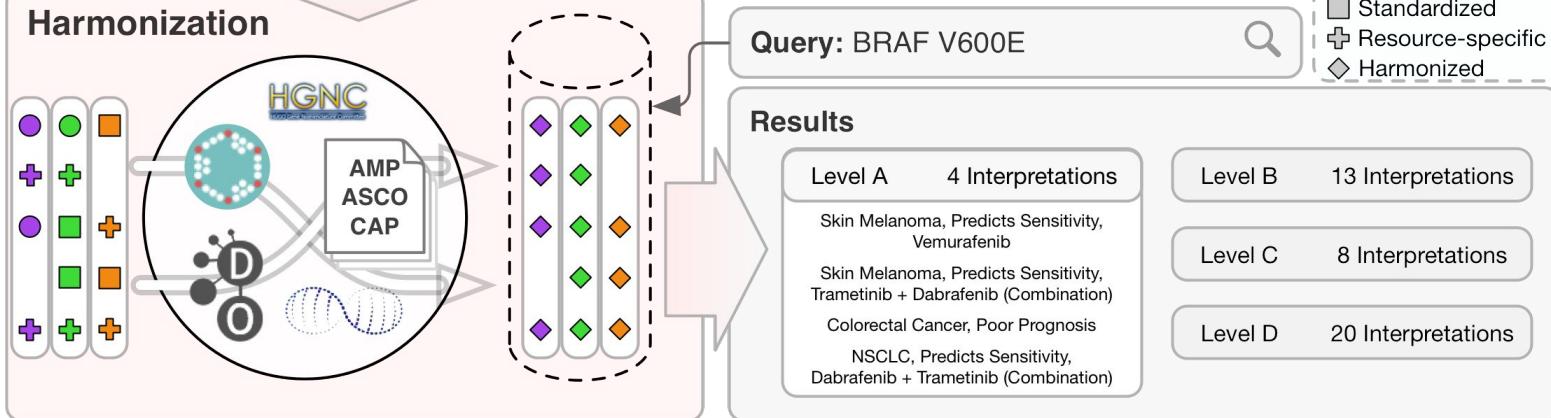
Variant Interpretation Knowledgebases



*information available only in free-form text

- Reference-linked
- Standardized
- + Resource-specific
- ◊ Harmonized

Harmonization



Wagner AH, et al. *Nat. Genet.* (accepted) 2019

Harvested Record

```
'variant': {'alteration': 'L2230V',
'consequence': {
'isGenerallyTruncating': False,
'term': 'missense_variant'},
'gene': {
'hugoSymbol': 'MTOR',
'oncogene': True,
'tsg': False,
'name': 'L2230V '},
'cancerType': 'Renal Clear Cell Carcinoma',
'drug': 'Temirosiromus, Everolimus',
'pmids': '27482884',
'level': '4',
'level_label': 'Compelling biological evidence supports the biomarker as being predictive of response to a drug but neither biomarker and drug are standard of care'}
```

Gene Harmonizer

HGNC Symbol Table

Variant Harmonizer

COSMIC Lookup and Variant Type Rules

ClinGen Allele Registry

Disease Harmonizer

DOID search from EBI, Bioontology

Drug Harmonizer

Pubchem, ChEMBL from Biothings

Evidence Harmonizer

Source-Specific Rules for AMP/ASCO/CAP Guidelines

Normalized Record

```
'gene_identifiers': {
'symbol': 'MTOR',
'ensembl_gene_id': 'ENSG00000198793',
'entrez_id': '2475'},
```

```
'features': [
'referenceName': 'GRCh37',
'chromosome': '1',
'start': 11182158,
'end': 11182158,
'ref': 'A',
'alt': 'C',
'name': 'L2230V',
'sequence_ontology': [
{name: 'missense_variant',
'soid': 'SO:0001583'}]]},
```

```
'disease': {
'source': 'DOID',
'id': 'DOID:4467',
'term': 'renal clear cell carcinoma'
},
```

```
'drugs': [
{id: 'CID6918289',
'source': 'pubchem/compound',
'term': 'TEMSIROLIMUS'},
{id: 'CID6442177',
'source': 'pubchem/compound',
'term': 'EVEROLIMUS'}
],
```

```
'description': 'Compelling biological evidence supports the biomarker as being predictive of response to a drug but neither biomarker and drug are standard of care',
'evidence_label': 'D',
'publication_url': 'http://www.ncbi.nlm.nih.gov/pubmed/27482884',
'response_type': 'sensitive'
```

Store Harvested Record
as entity within
Normalized Record



Harmonization **Increases** Breadth of Interpretation



AACR Project GENIE

AACR Project Genomics Evidence Neoplasia Information Exchange

- Registry for aggregating cancer genomic data with clinical outcomes
- Represents “real-world” clinical sequencing efforts
 - Data is CLIA-/ISO-certified
 - Obtained through participation of 8 major international cancer genomics centers
 - Enriched in examples of late-stage disease

<https://www.aacr.org/RESEARCH/RESEARCH/PAGES/AACR-PROJECT-GENIE.ASPX>



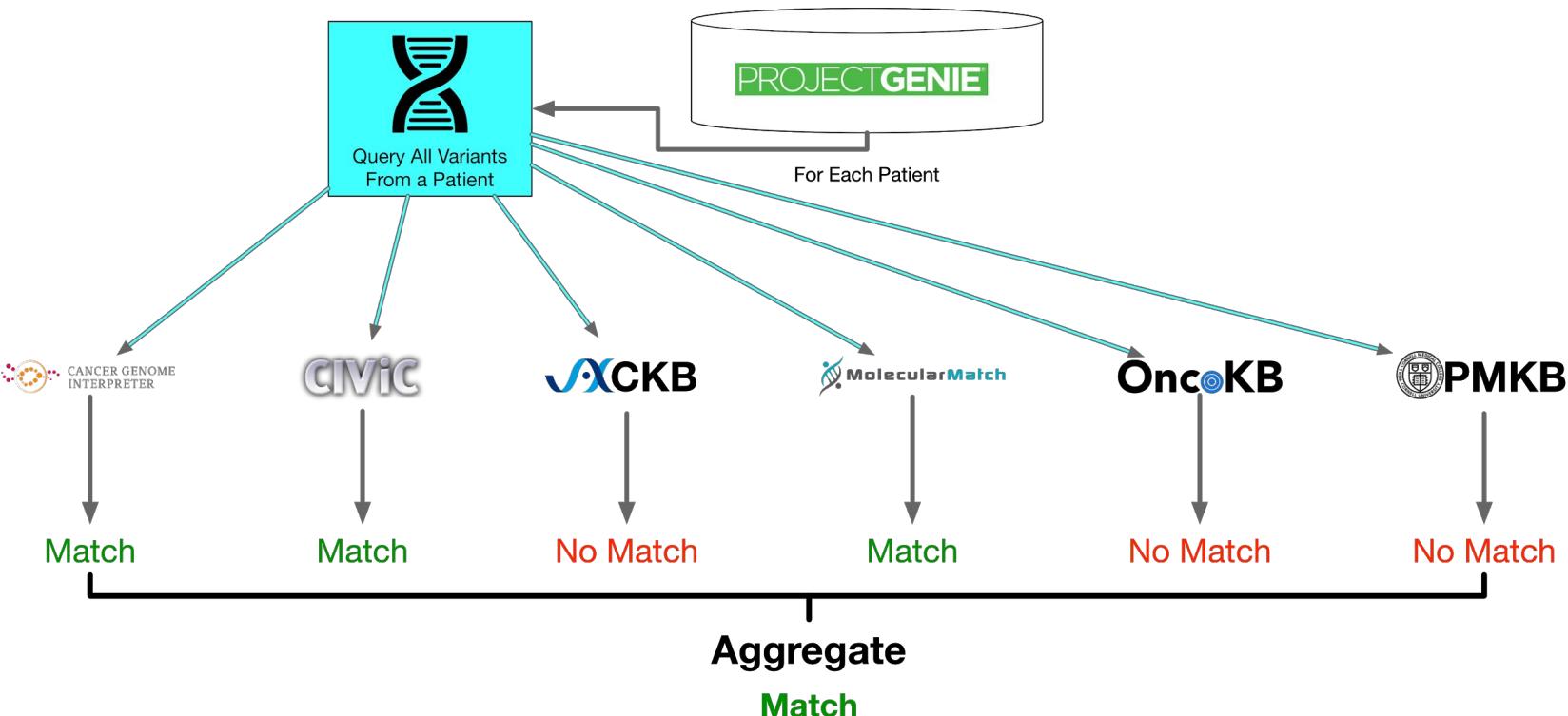
Project GENIE Cohort

- v3.0.0, released January 2018
- 38,207 patients
- 263,592 variants
- 539 distinct disease terms (Oncotree)

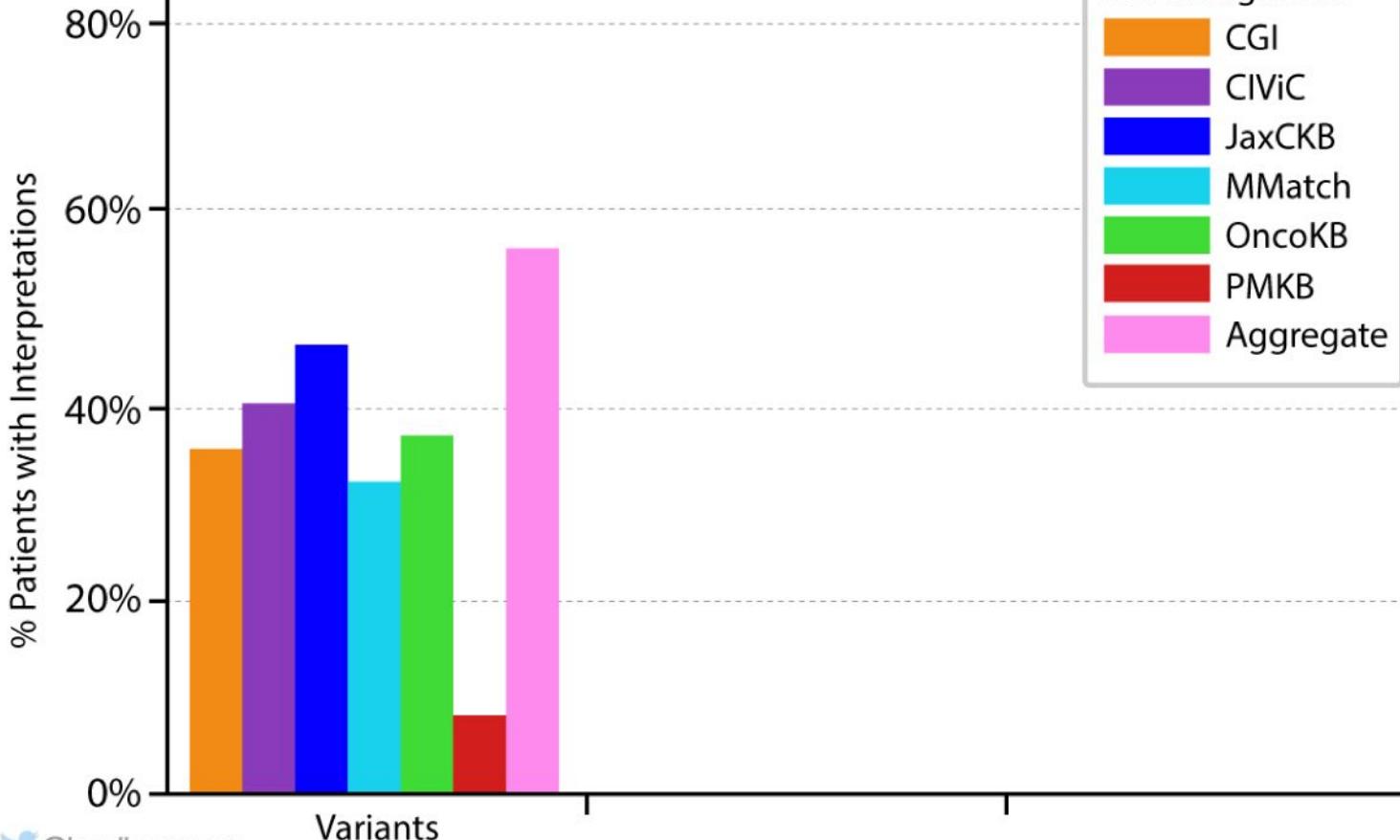
<https://www.synapse.org/#!Synapse:syn7222066/wiki/410924>



Interpretation Coverage of GENIE Cohort

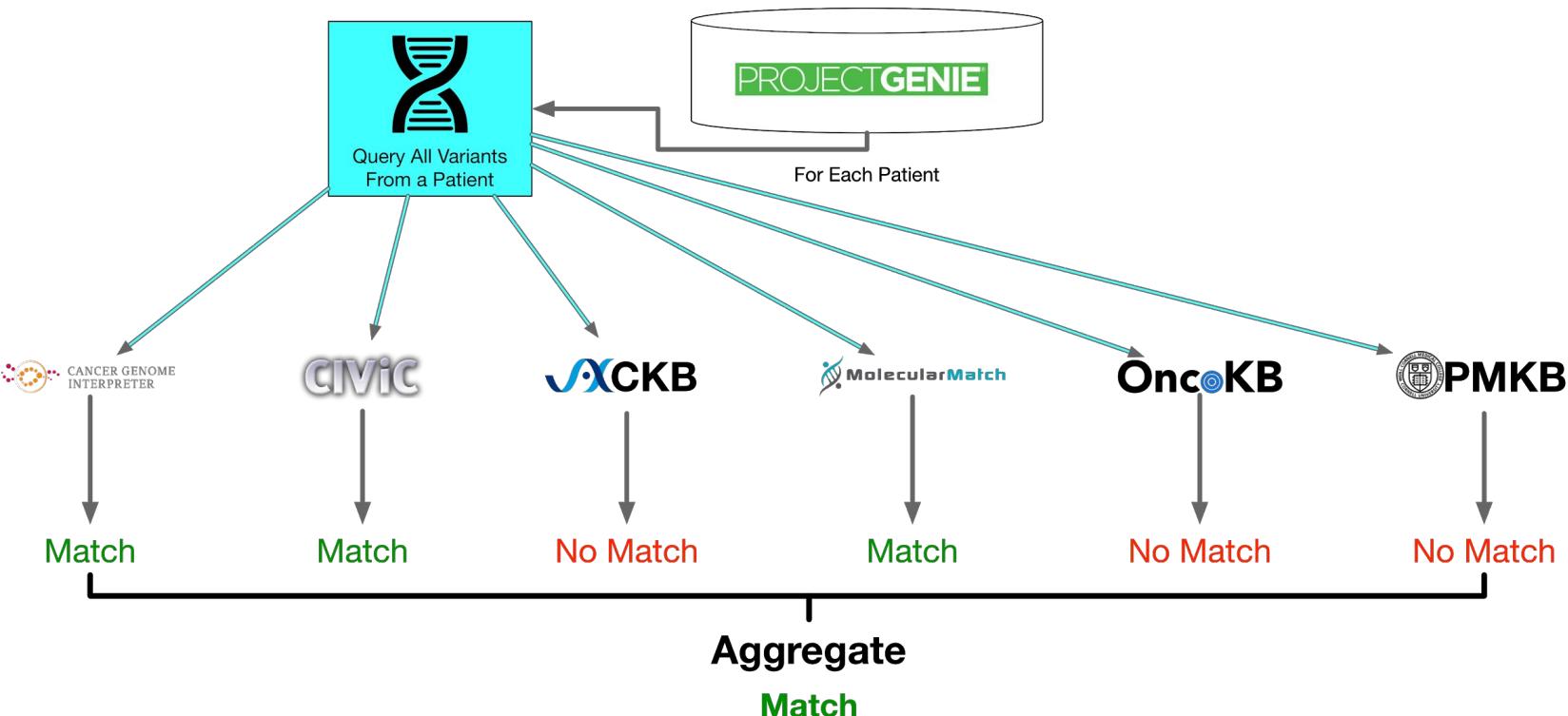


Project GENIE Cohort Interpretations



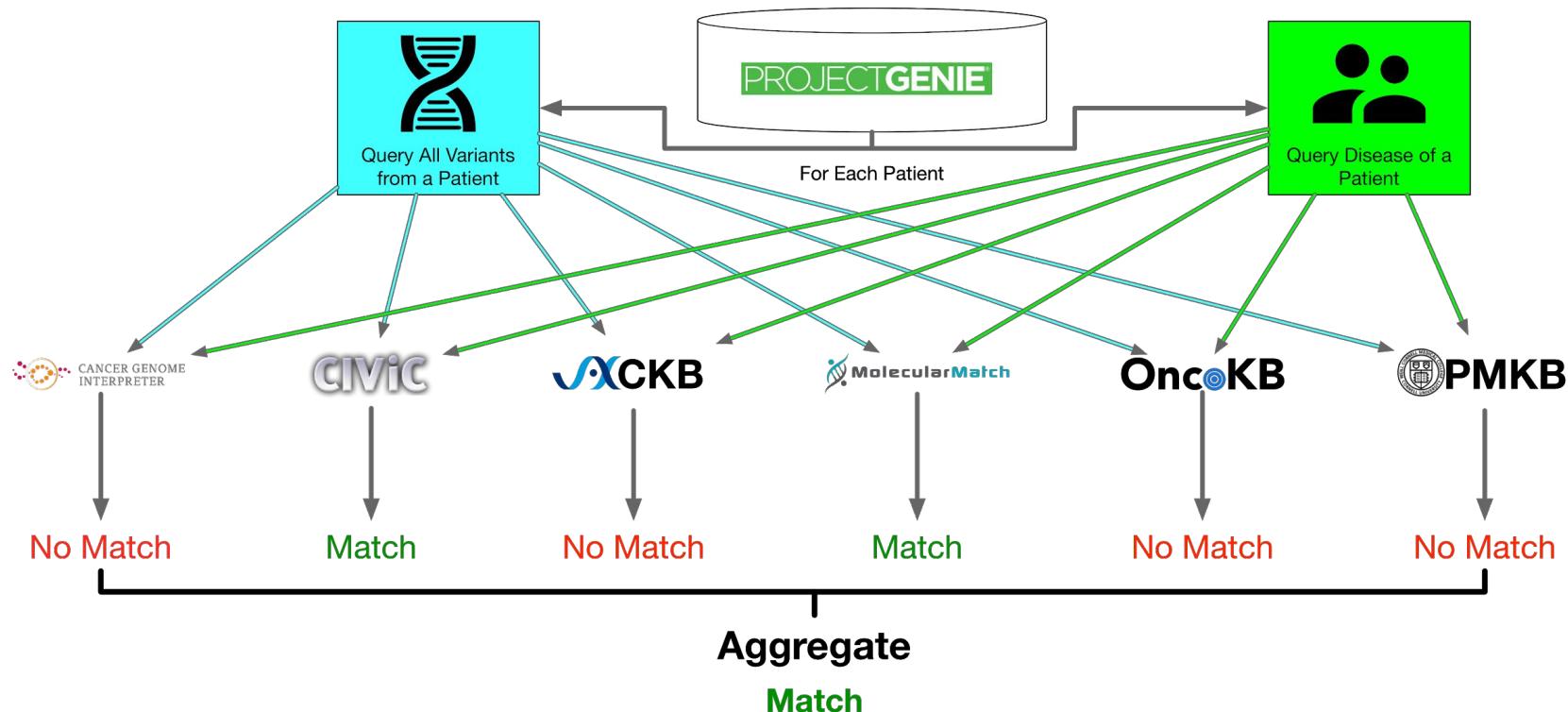


Interpretation Coverage of GENIE Cohort

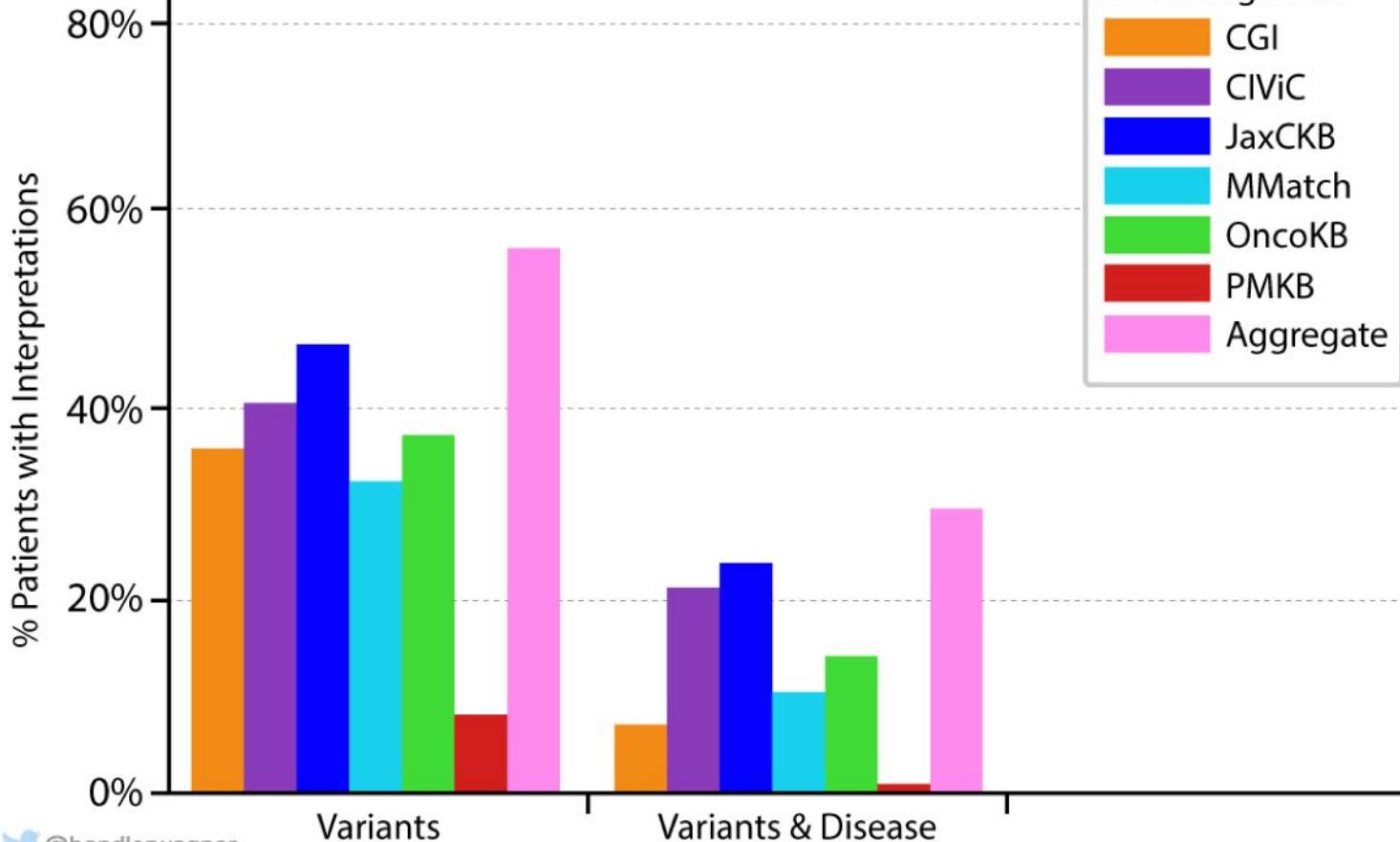




Interpretation Coverage of GENIE Cohort

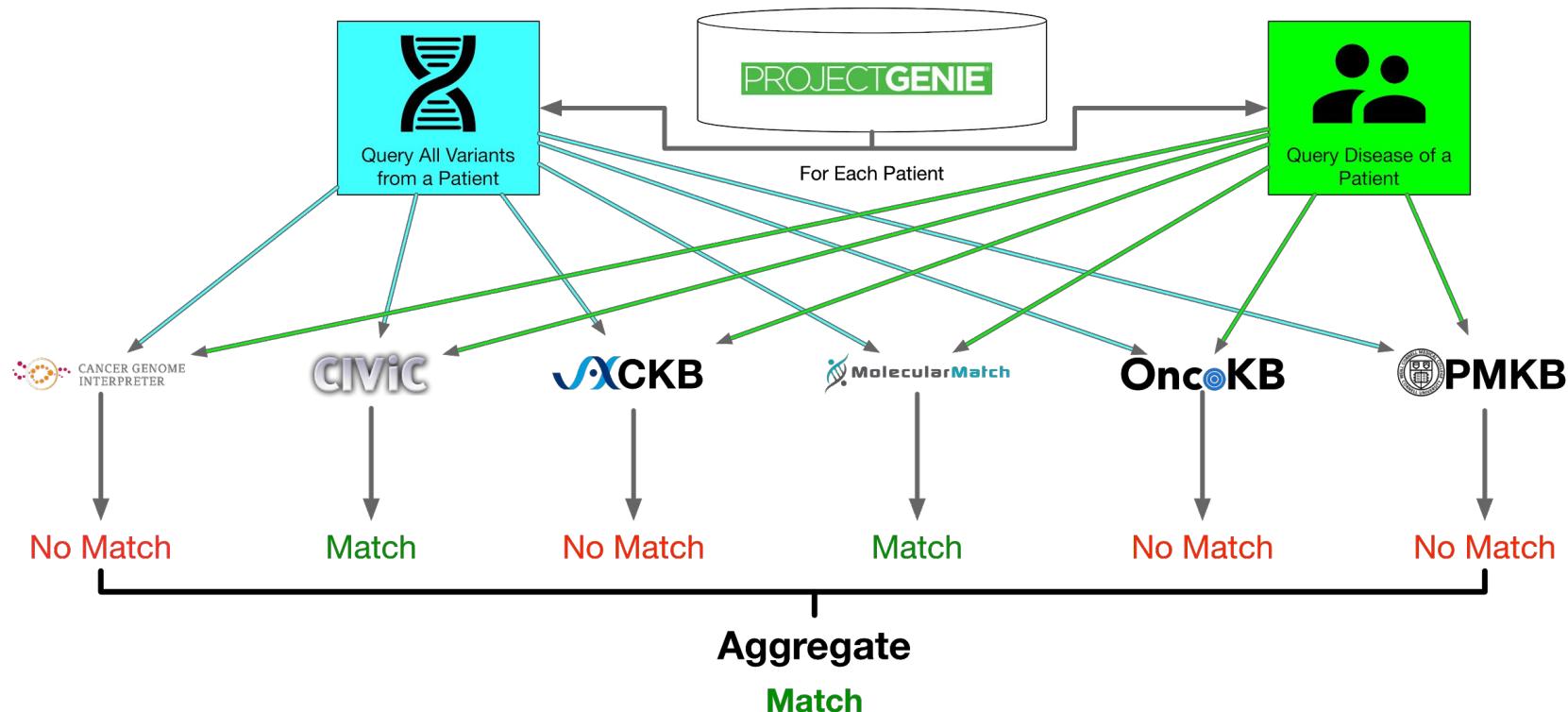


Project GENIE Cohort Interpretations



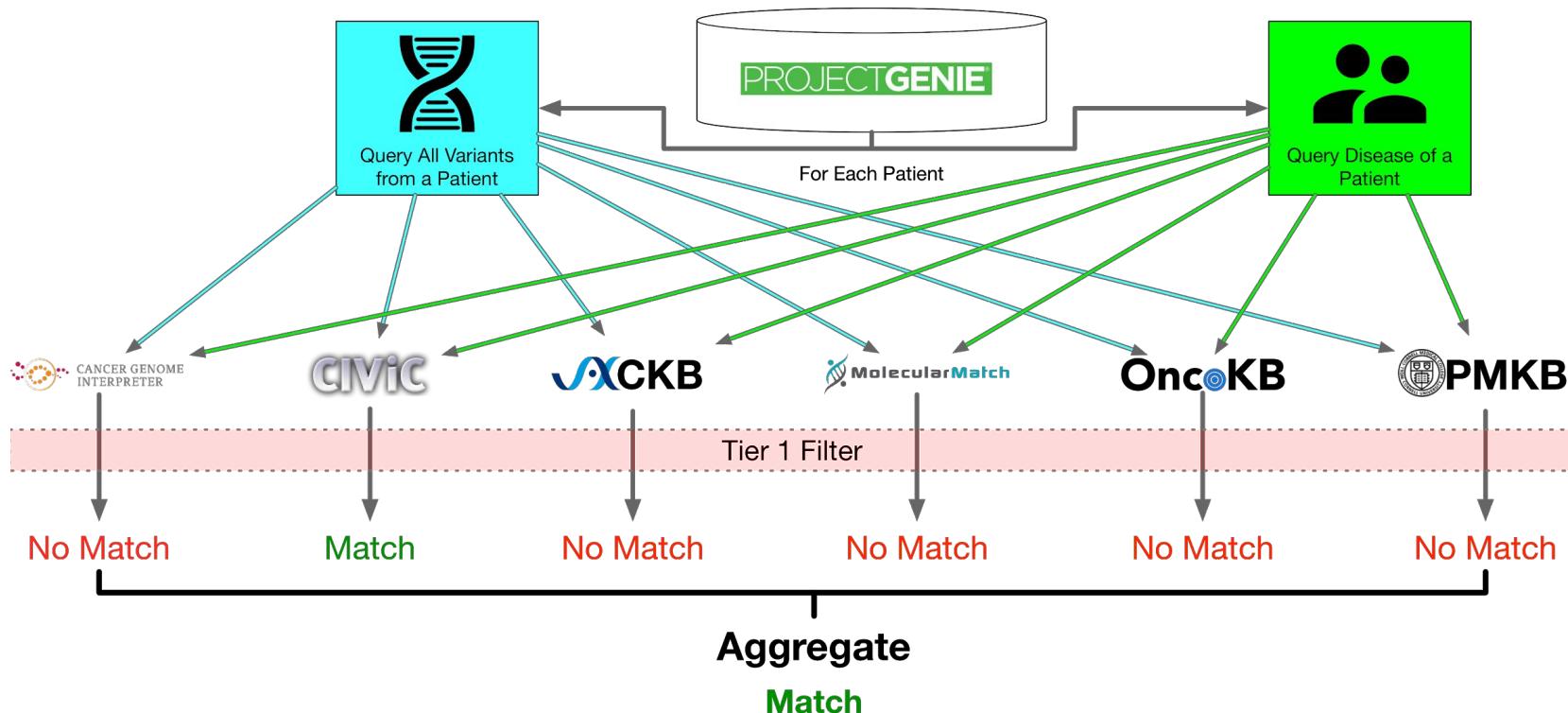


Interpretation Coverage of GENIE Cohort

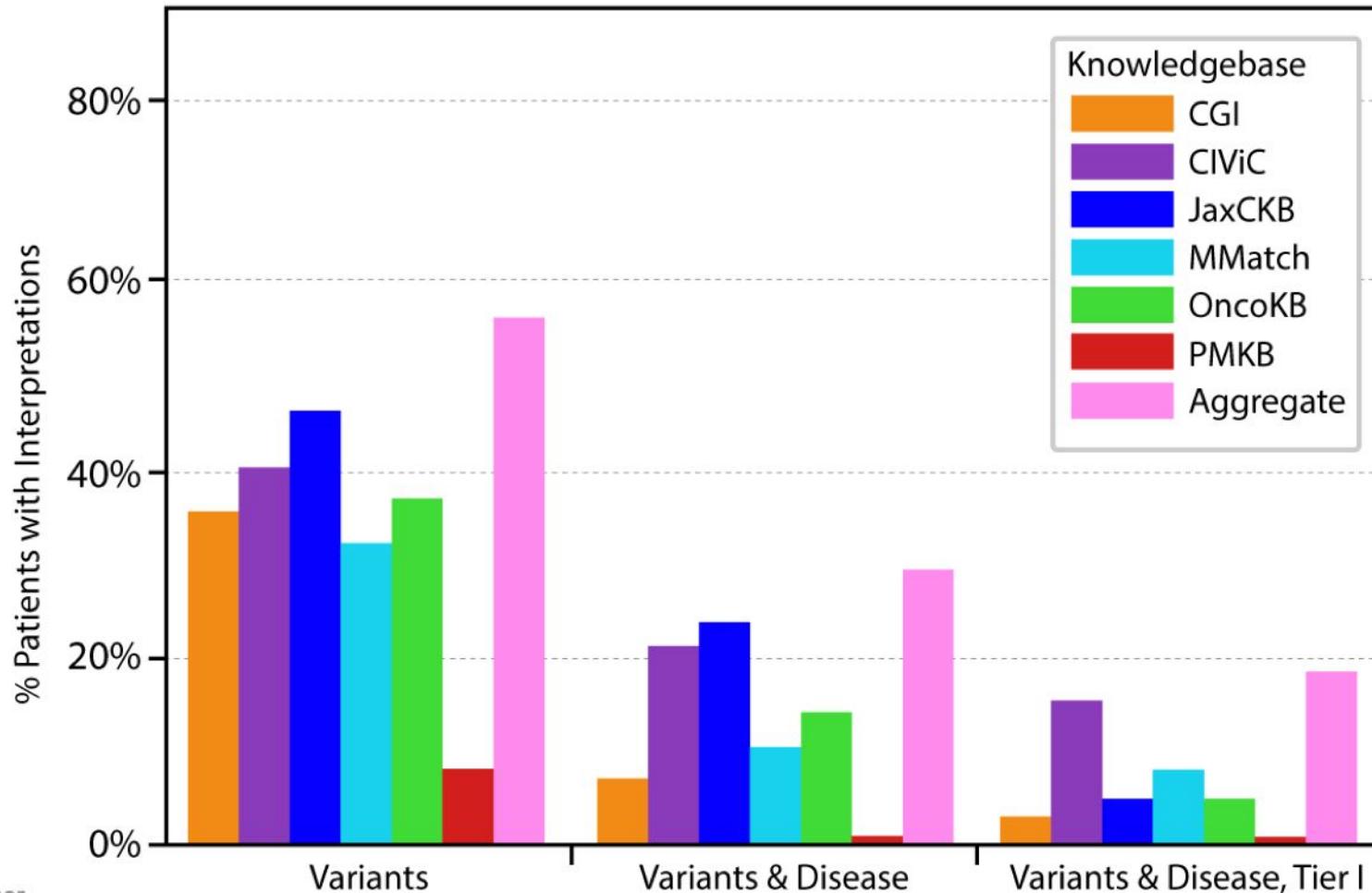




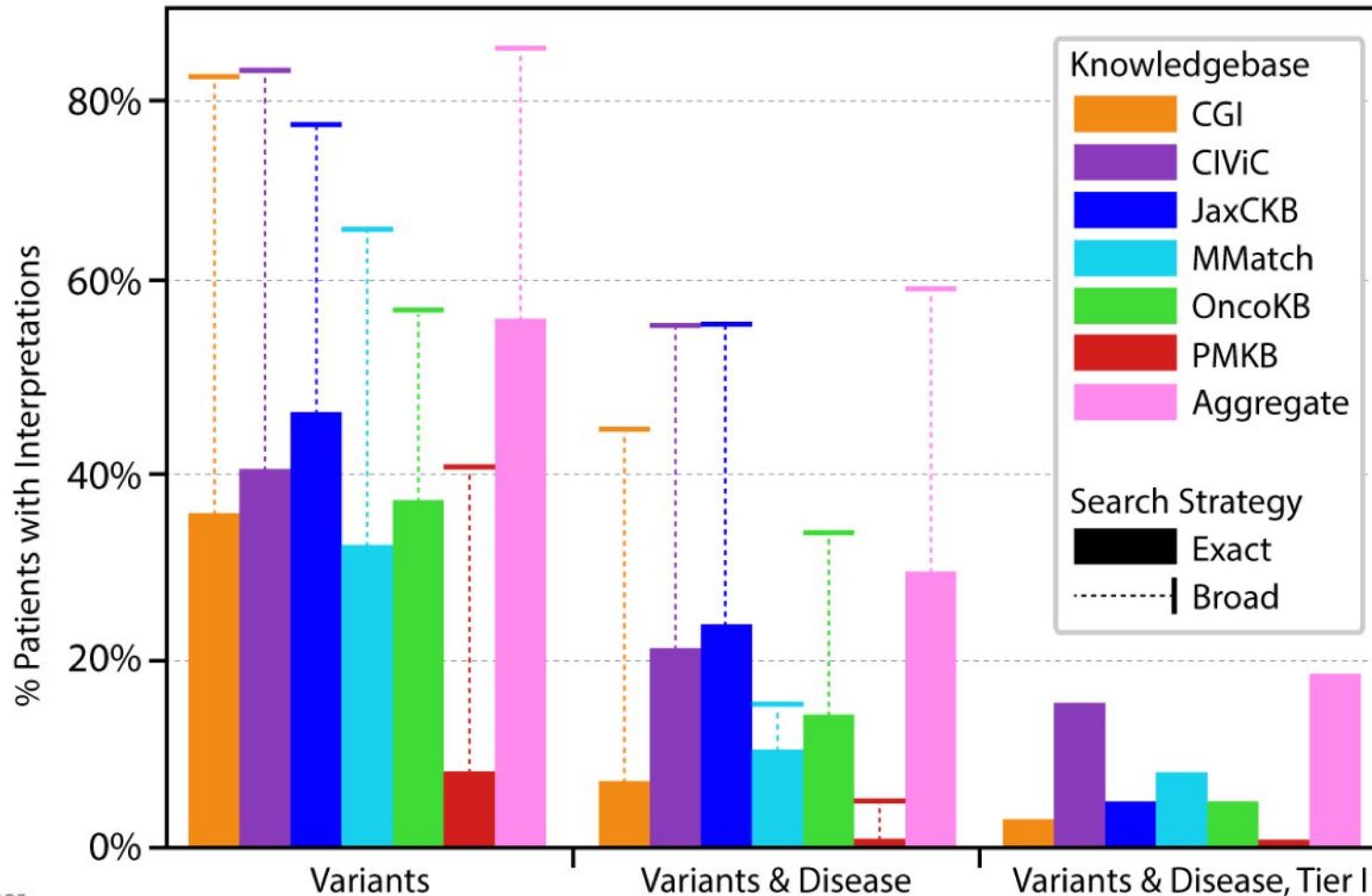
Interpretation Coverage of GENIE Cohort



Project GENIE Cohort Interpretations



Project GENIE Cohort Interpretations





Meta-KB Search Strategies

a



Chr12 25398282-25398282 C>T

KRAS Amplification



Chr12 25362365-25403737

Exon 2 Mutation



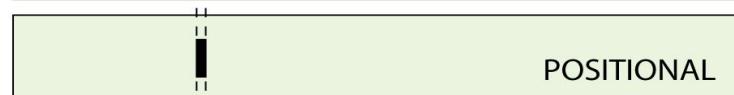
Chr12 25398208-25398329

p.G13



Chr12 25398281-25398283

p.G13C



Chr12 25398282-25398282 C>A

p.G13S

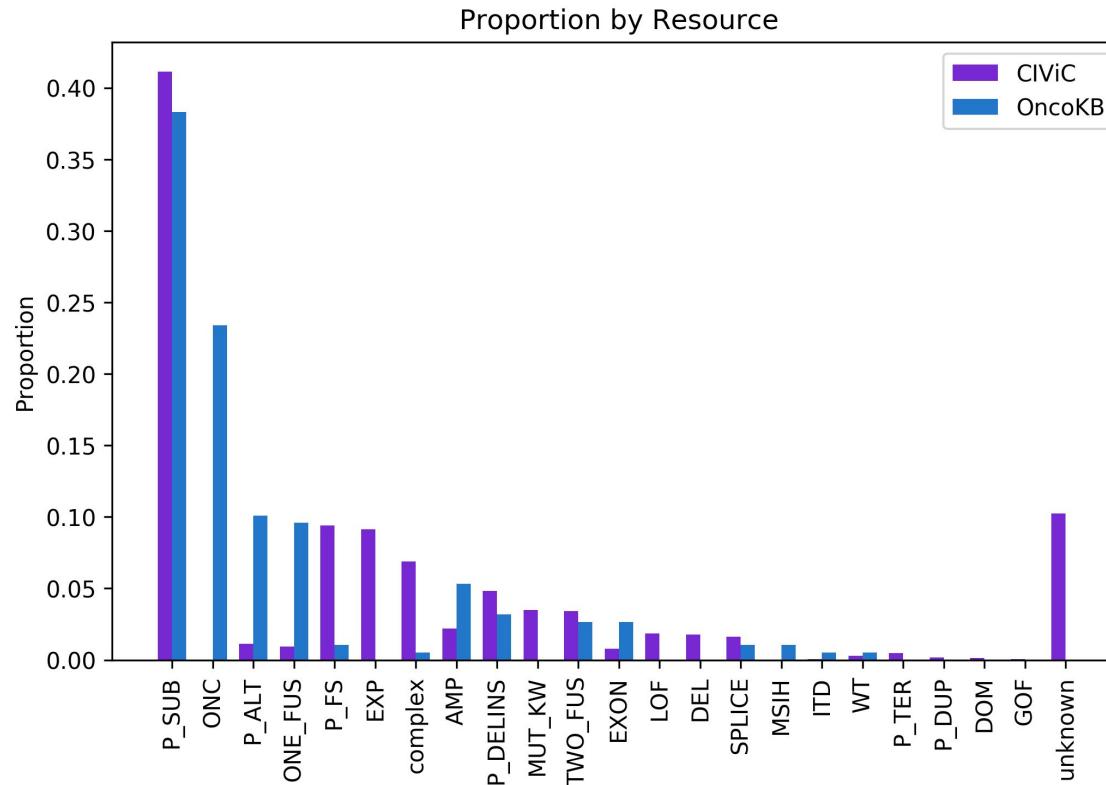


Chr12 25398282-25398282 C>T

Wagner AH, et al. *Nat. Genet.* (accepted) 2019



The Somatic Variant Landscape





Interoperability through **Expressive** and **Computable** Variation Representation



Interoperability through **Expressive** and **Computable** Variation Representation

Current and Future Work



Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.



GA4GH: Mission

“

The Global Alliance for Genomics and Health aims to accelerate progress in genomic research and human health by cultivating a common **framework of standards** and harmonized approaches for effective and responsible **genomic and health-related data sharing**.

”

<https://www.ga4gh.org/about-us/>

Introduction

Terminology & Information Model

Schema

Implementation Guide

GA4GH Variation Representation Specification

The Variation Representation Specification (VR-Spec) is a standard developed by the Global Alliance for Genomic Health to facilitate and improve sharing of genetic information. The Specification consists of a JSON Schema for representing many classes of genetic variation, conventions to maximize the utility of the schema, and a Python implementation that promotes adoption of the standard.

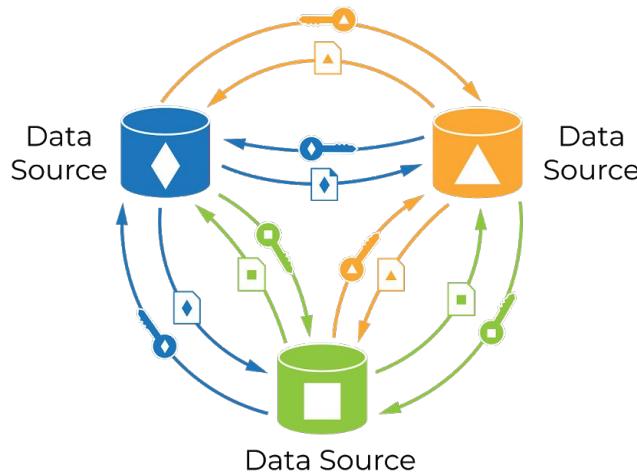
- [Introduction](#)
- [Terminology & Information Model](#)

<https://vr-spec.readthedocs.io>



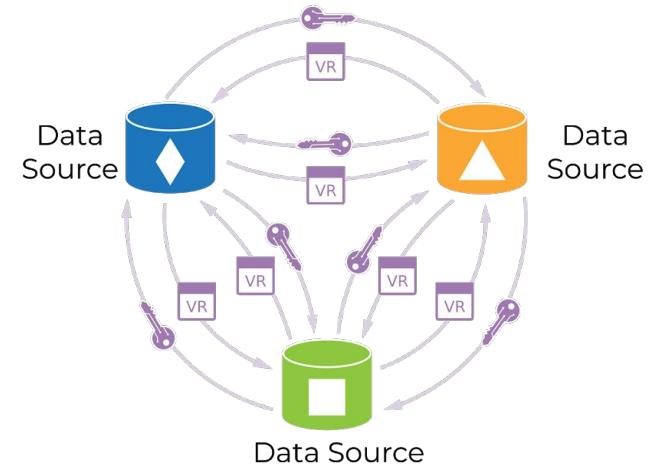
VR Specification Driving Interoperability

CURRENTLY...



PAIRS OF SYSTEMS COORDINATE KEYS AND FORMATS IN ORDER TO SHARE VARIATION DATA. ADDING A NEW SYSTEM IS DIFFICULT.

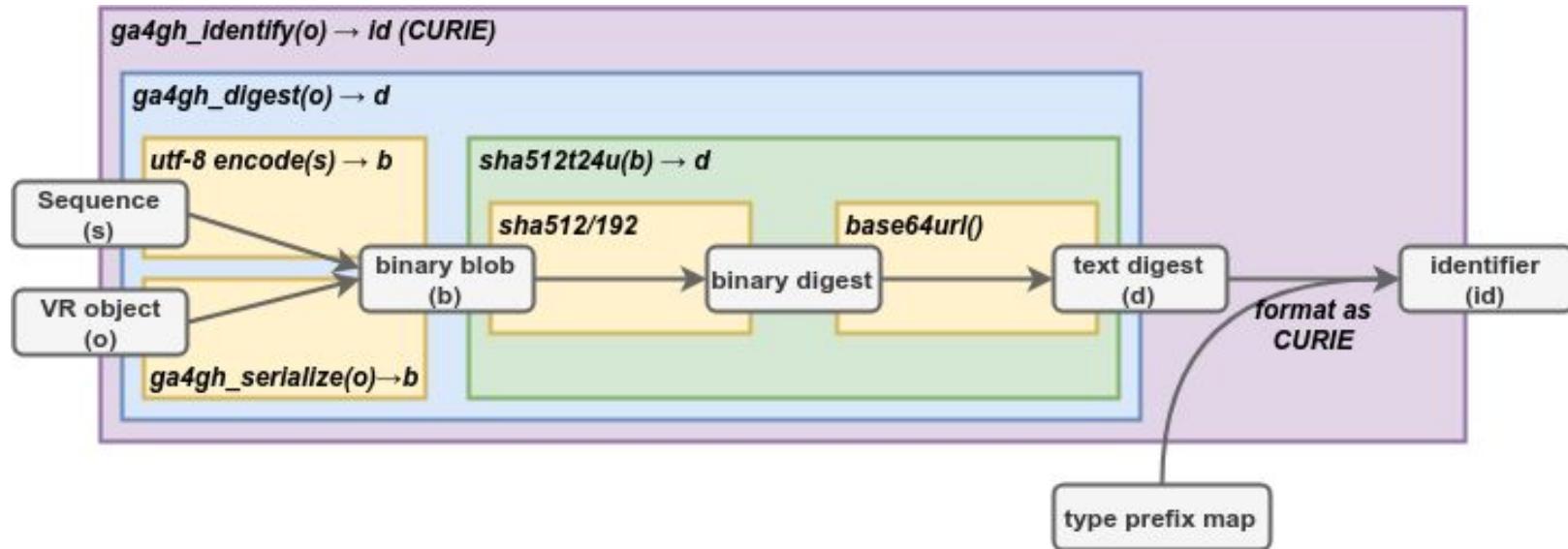
WITH THE VR SPECIFICATION...



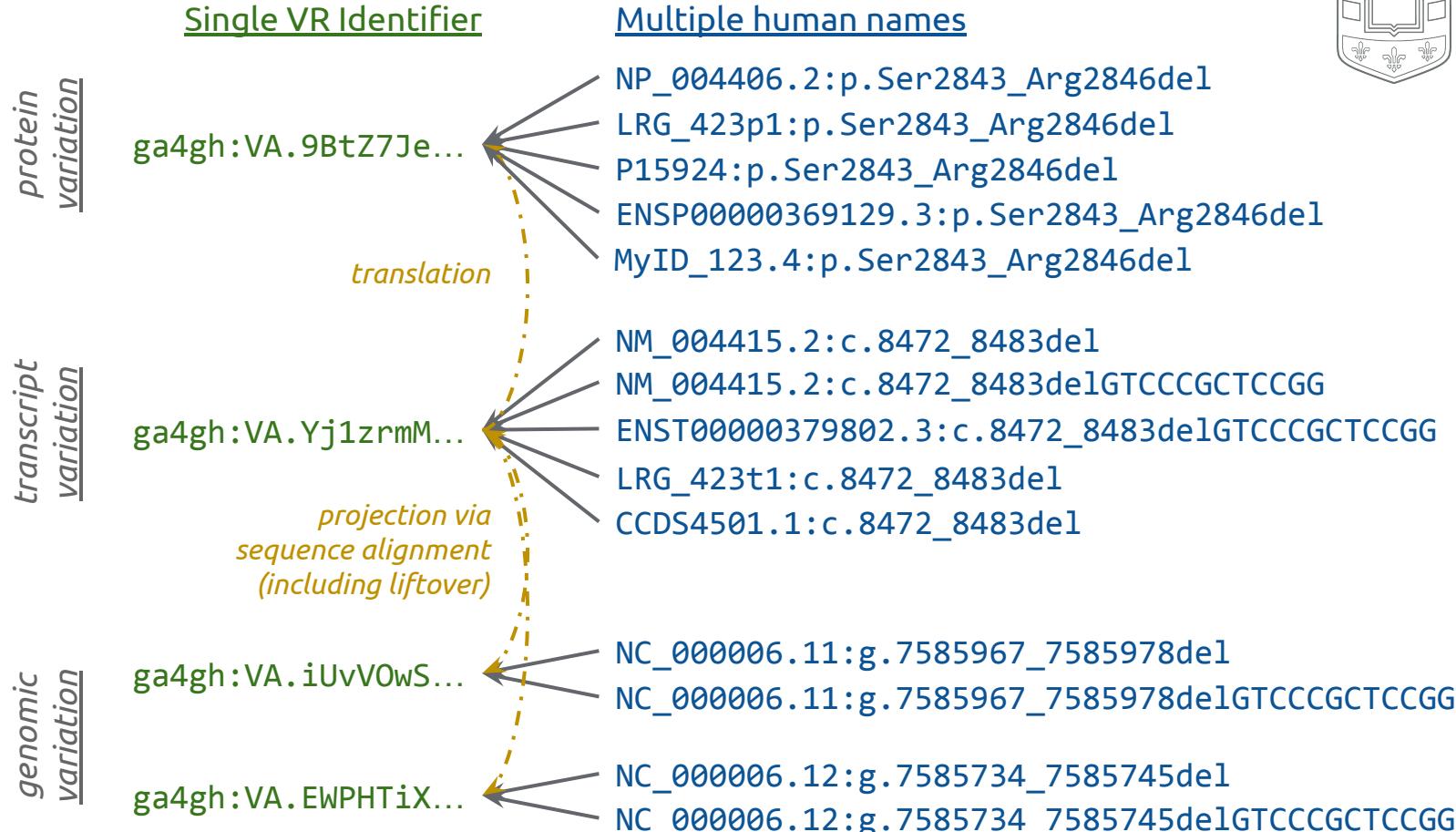
SYSTEMS USE A COMMON IDENTIFIER, COMPUTED FROM THE DATA ITSELF, AND A COMMON DATA FORMAT. ADDING A NEW SYSTEM IS MUCH EASIER.



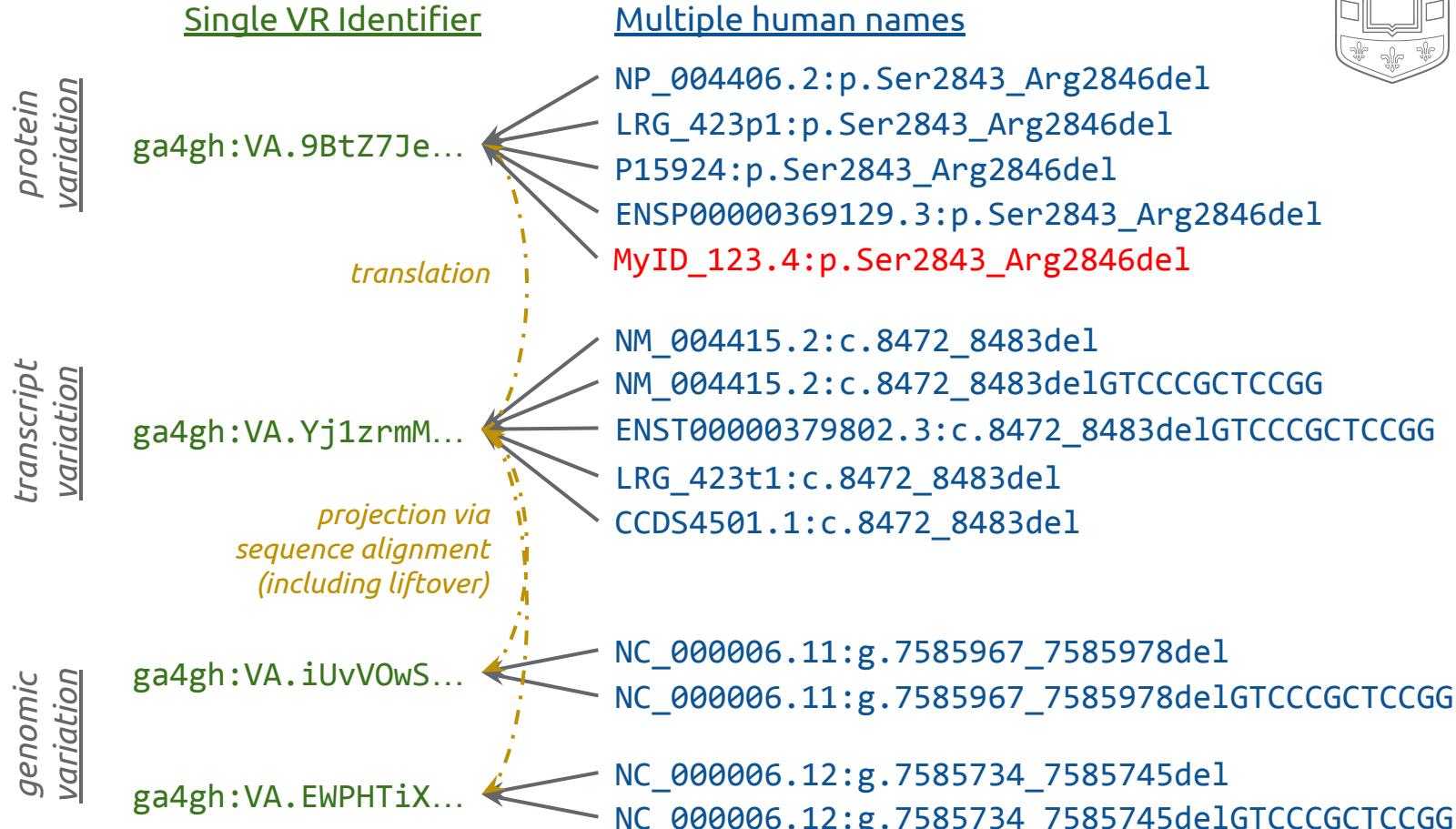
VR Computed Identifiers



VR identifiers: precise, unique names for variation



VR identifiers: precise, unique names for variation

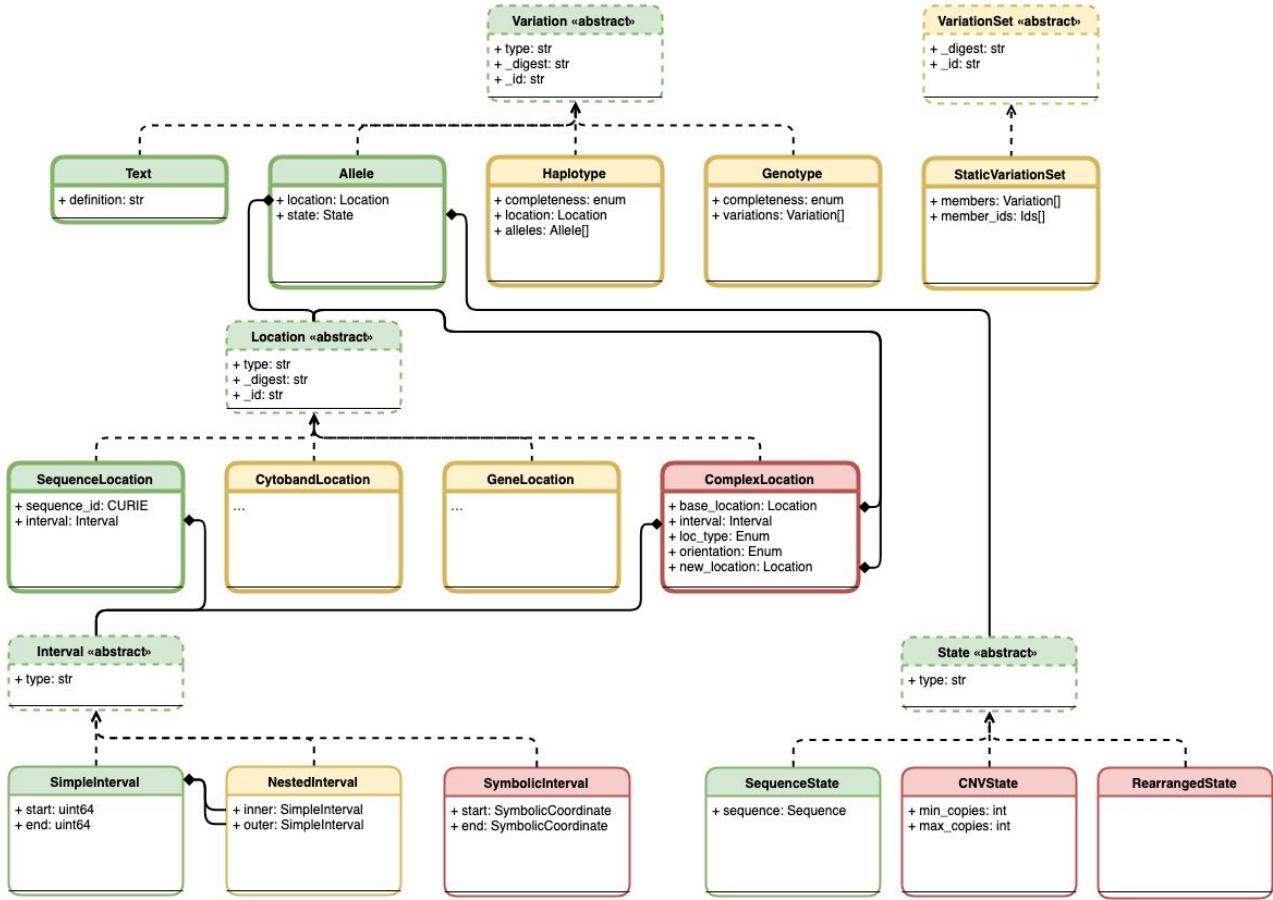


Upcoming milestone goals:

- Variation Sets
- SVs
- Fusions
- CNVs
- Cytoband Locations
- Gene Locations

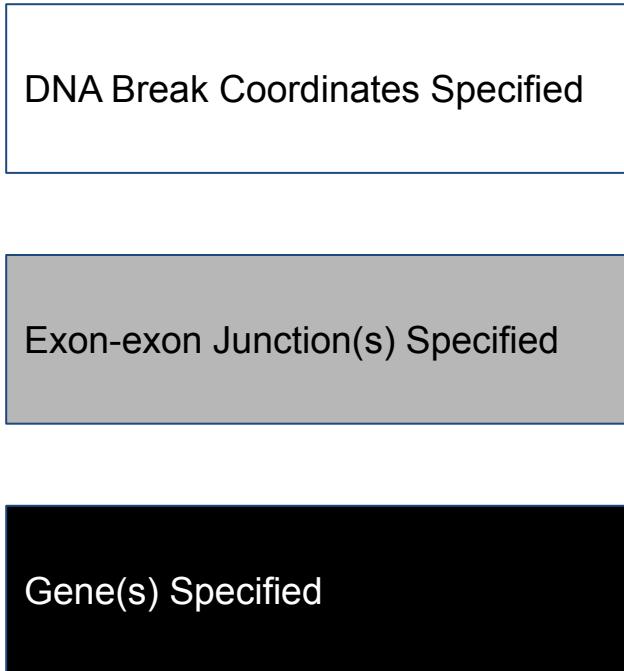
All classes have a string 'type'.
Dashed lines denote abstract classes. Abstract classes are not instantiated.
Thin solid lines denote classes that may be instantiated but are not identifiable.
Bold lines denote identifiable objects (i.e., may be serialized and identified by computed identifier).
Solid arrow lines denote inheritance. Subclasses inherit all attributes from their parent. Inherited attributes are not shown.

v1.0 | Previously developed objects in v1.1 | Newly developed objects in v1.1

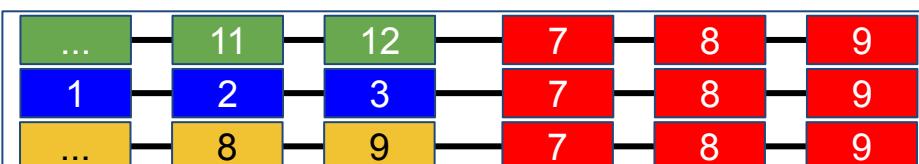
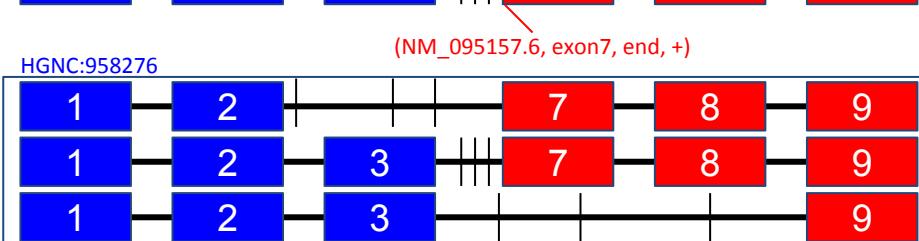
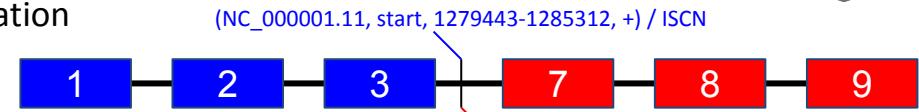




Fusion Specificity Needs Vary



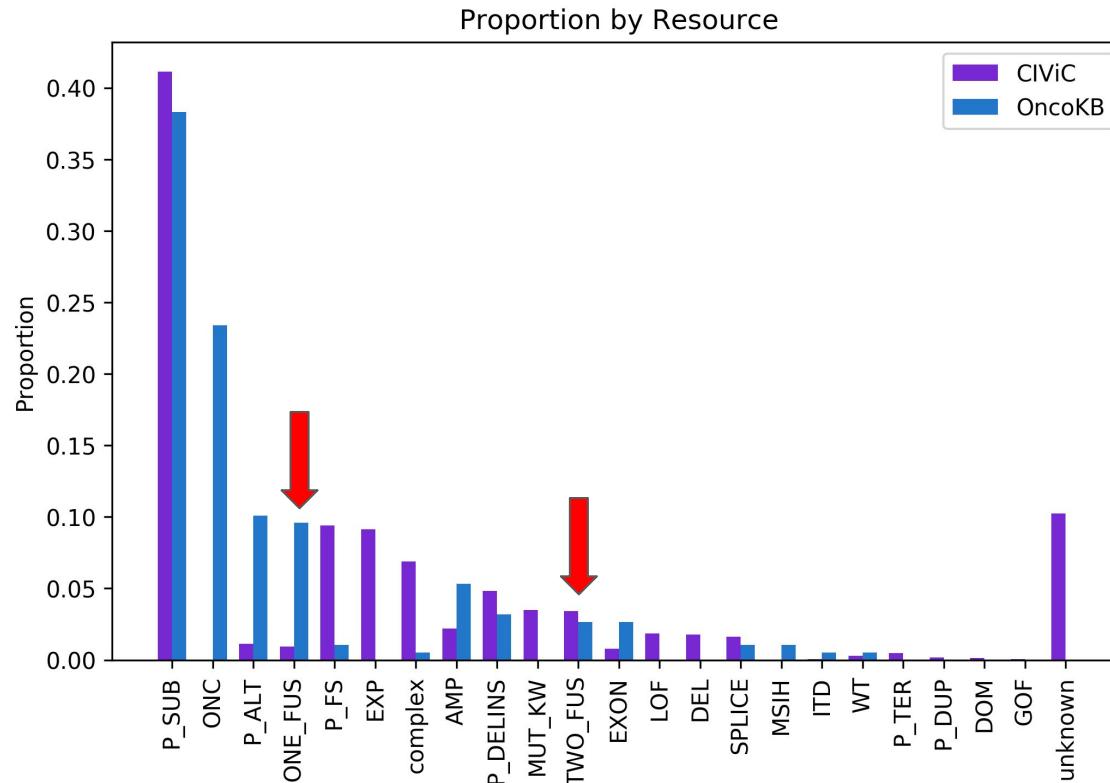
Discrete Variation



Categorical Variation

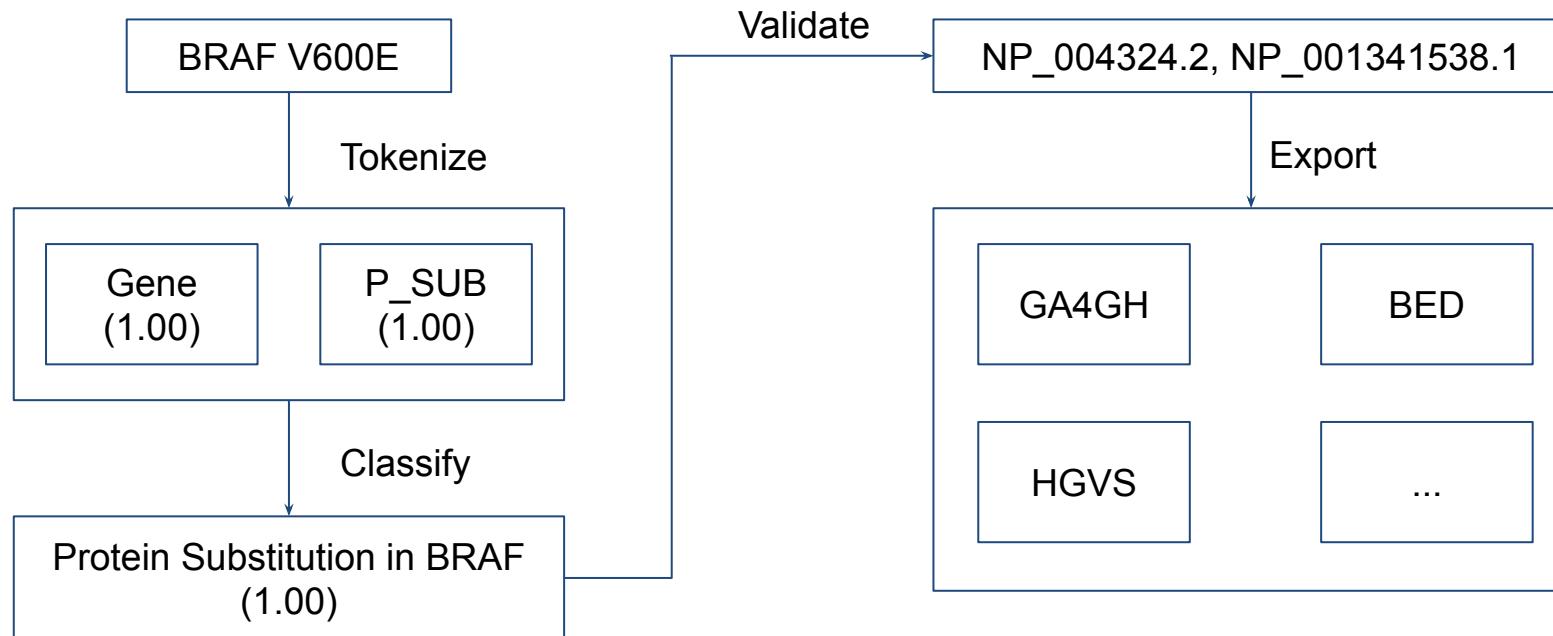


Fusions in the Somatic Variant Landscape





Lexing and Classification of Categorical Variants



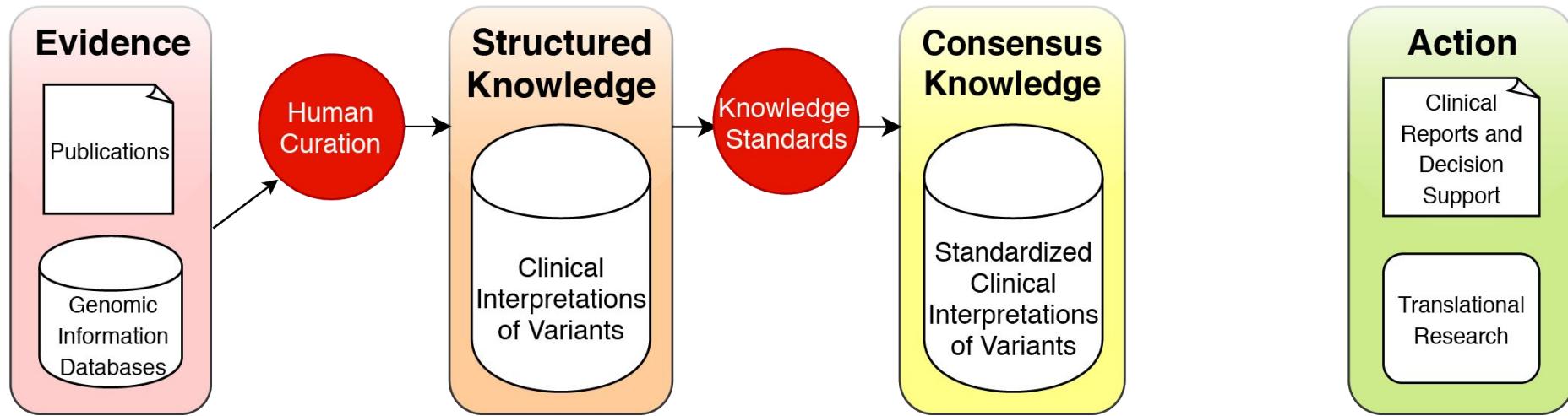


Improving Normalization of Drugs and Diseases

- Current efforts are focused on improving existing strategies for normalizing drug terms (K99/R00 Aim 1)
- Several useful services currently exist
 - e.g.: NCI SEER, NCI EVS, RxNorm
- Complex strategies needed for resolving difficult terms
- For diseases, inter-ontology mapping is a pain
 - working with Monarch Initiative (MONDO) and Human Disease Ontology (DO) to drive solutions
 - NCI-Thesaurus is being adopted as core for semantic precision

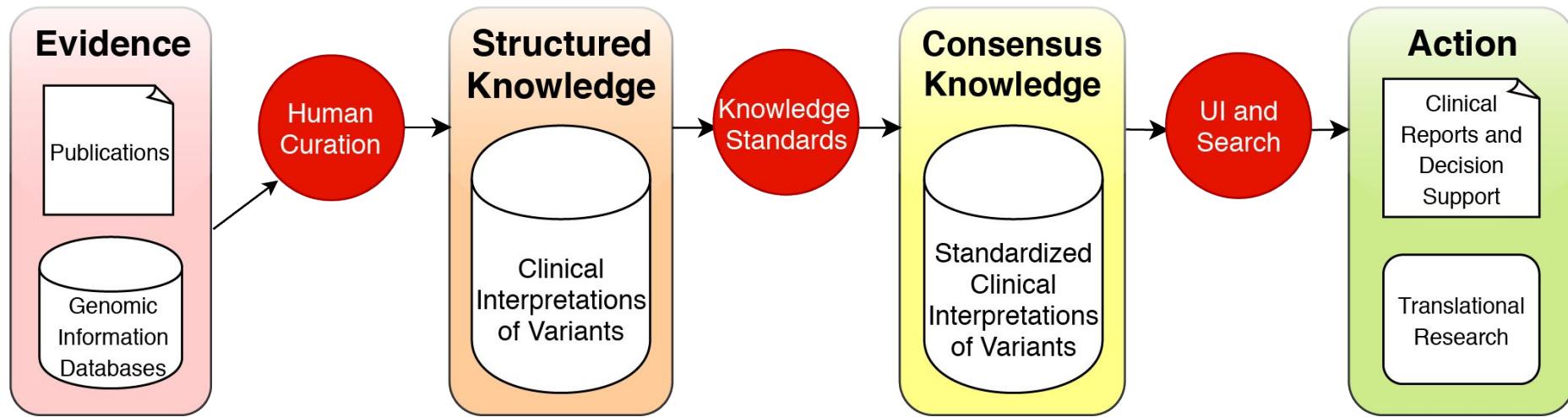


Roadmap from Evidence to Action

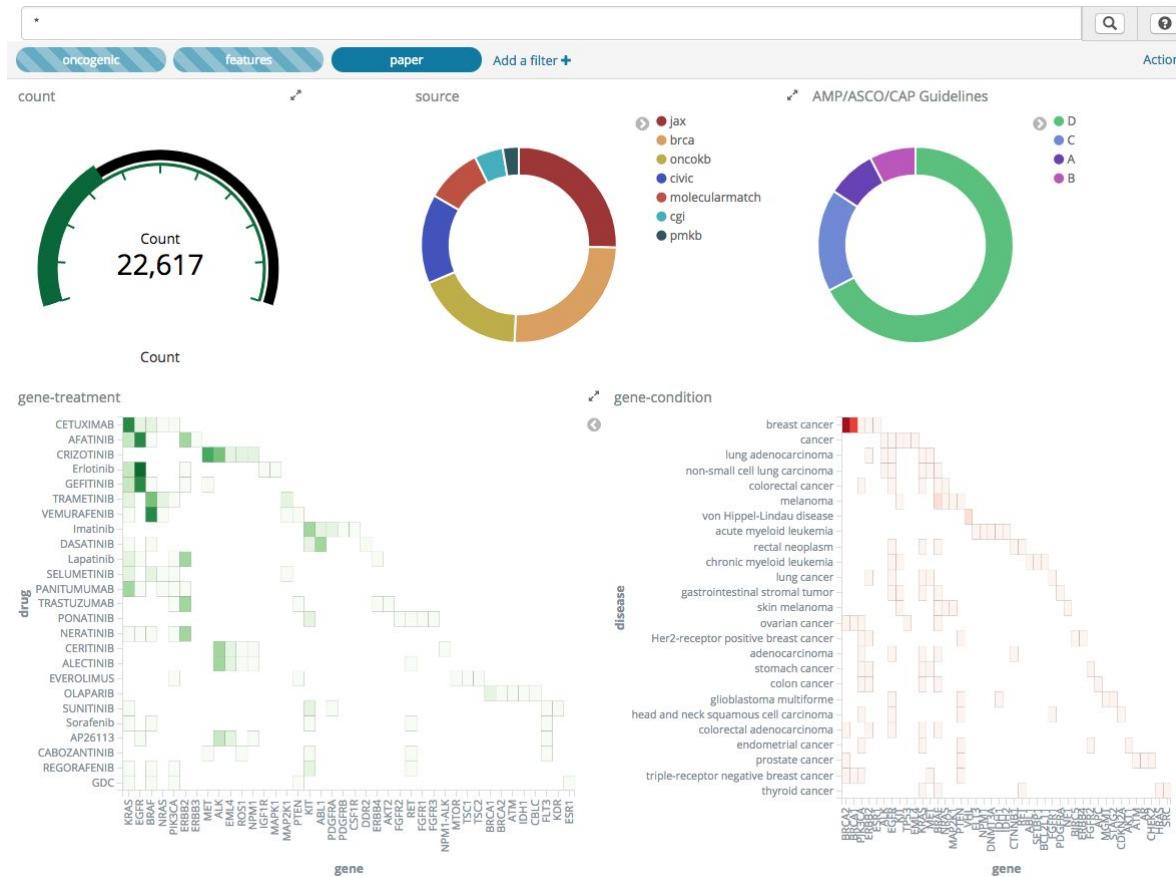




Roadmap from Evidence to Action



Public Search Tool



 @handlerwagner

Public Search Tool

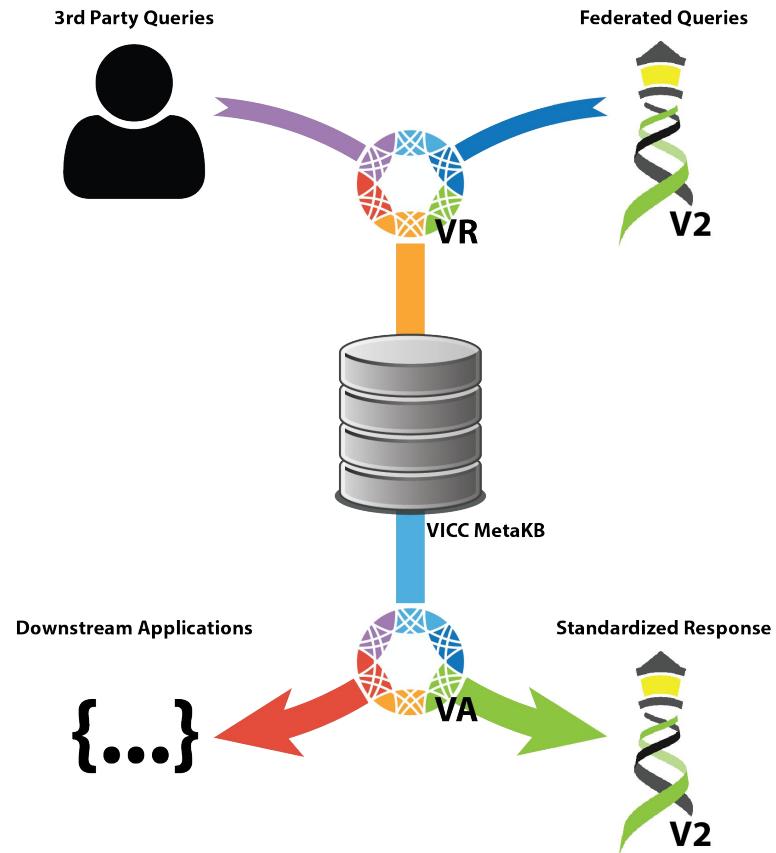


Documentation:
docs.cancervariants.org

Web Application:
search.cancervariants.org

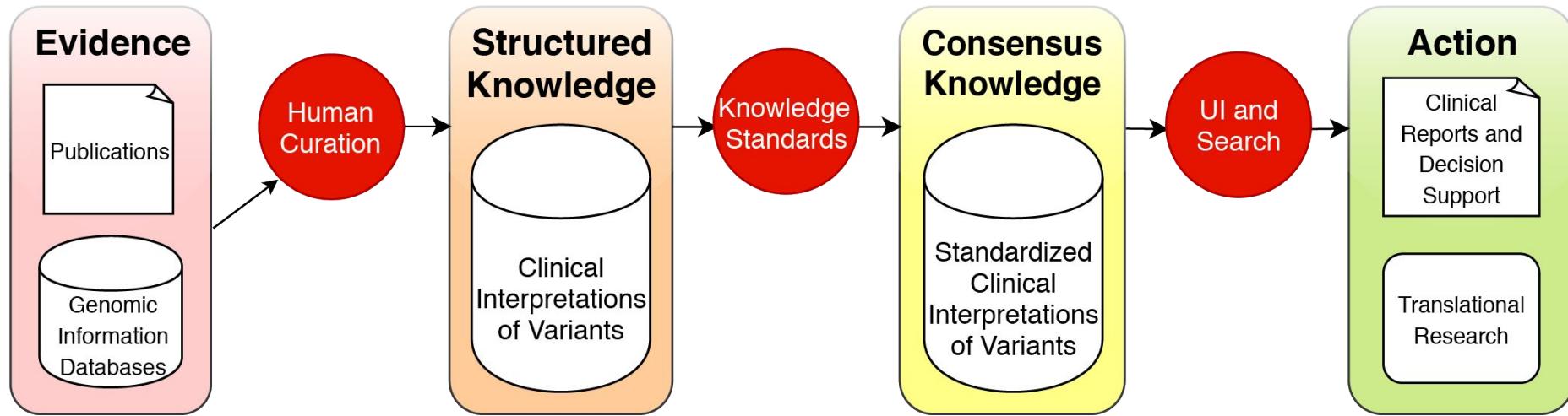


VR, VA, and Beacon v2





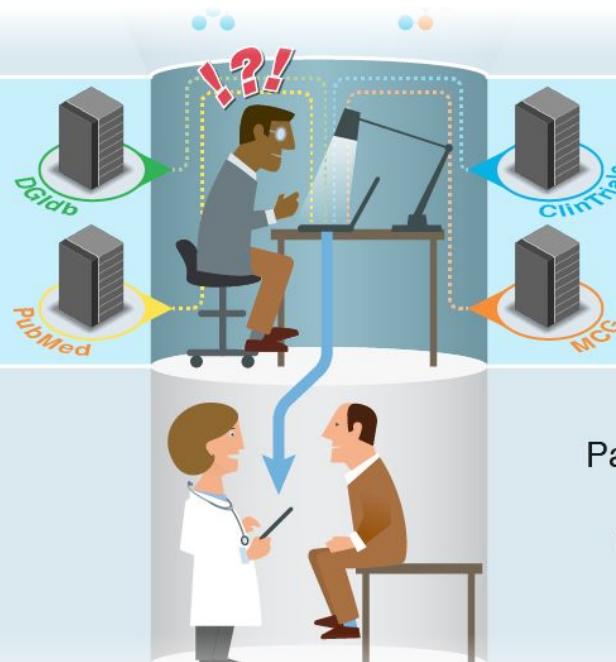
Roadmap from Evidence to Action





Precision Oncology Bottleneck

5. Interpretation and Report Generation



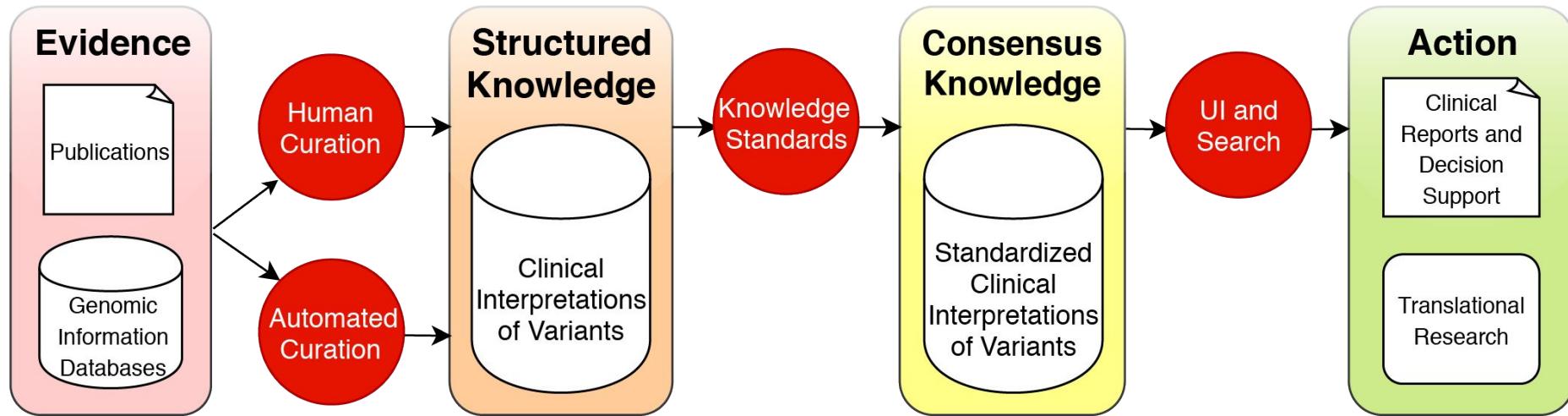
A genome analyst attempts to interpret, prioritize, and summarize functionally significant events in the context of published literature, clinical trials, and a multitude of knowledgebases.

6. Clinical Application

Pathologists and oncologists evaluate the significance of potentially clinically actionable events, and incorporate their research into patient care.



Roadmap from Evidence to Action





Automated Curation Using Domain Knowledge

- **Extension goal** of K99/R00
- Deep learning approaches (e.g. Watson) have had difficulty in the area of clinical decision support
- ML strategies utilizing domain-specific knowledge can outperform deep learning on sparse, low-quality (e.g. un-normalized) data
- **Hypothesis:** machine learning models using domain-specific tools (such as those in K99/R00) and trained on normalized curated knowledge can drive rapid curation of predicted high-impact knowledge sources (e.g. publications, clin. trials)



Final Comments



Community Input is Welcome and Critical!

VICC is an open consortium: <https://cancervariants.org/join>

Our Variant Harmonization Working Group is driving this effort

The VR specification is open for public use and contributions:

Specification: <https://vr-spec.readthedocs.org>

Repository: <https://github.com/ga4gh/vr-spec>

SV requirements: <https://github.com/ga4gh/vr-spec/issues/28>



Acknowledgements

Consortium Leadership



Obi Griffith
Washington U
VICC Founder
VICC Co-director



Malachi Griffith
Washington U
VICC Founder
VICC Co-director



Debyani
Chakravarty
MSKCC
VICC Co-director

GA4GH GKS Work Stream

Reece Hart
Larry Babb
Bob Freimuth
Andy Yates
Matthew Brush
Javier Lopez
Melissa Konopko

Key Knowledgebase Partners



Nikolaus Schultz
MSKCC



Debyani
Chakravarty
MSKCC



Rodrigo
Dienstmann
Vall D'Hebron



Olivier Elemento
Weill Cornell



Ryan Duren
Molecular Match



Susan Mockus
Jackson Lab



Ethan Cerami
DFCI



Steven Jones
BC Cancer

VICC and ClinGen Somatic

Subha Madhavan
Gordana Raca
Angshumoy Roy
Beth Pitel
Shruti Rao
Dmitriy Sonkin
Ian King

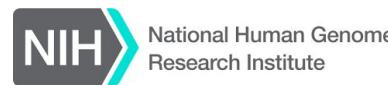
Consortium Participants (Past and Present)

Benjamin Ainscough, Tero Aittokallio, Larry Babb, Michael Baudis, Jacques Beckmann, Anas Belouali, Andrew Biankin, Michael Bouziner, Lynn Brazil, Steven Brenner, Alberto Cambrosio, Jonah Campbell, Ethan Cerami, Debyani Chakravarty, David Chang, Brad Chapman, Justina Chung, Chris Corless, Melanie Courtot, Fiona Cunningham, Rob Currie, Catherine del Vecchio Fitz, Jordi Deu Pons, Rodrigo Dienstmann, Kenneth Doig, Lena Dolman, Ryan Duren, Daniel Durkin, Olivier Elemento, Jonathan Ellis, Kyle Ellrott, Robert Freimuth, Jianjiang Gao, Moritz Gerstung, Bailey Glen, William Glen, Jeremy Goecks, Santiago Gonzalez, Sara Gosline, Malachi Griffith, Obi Griffith, Melissa Haendel, Maximilian Haeussler, David Haussler, David Heckerman, Oliver Hofmann, Peter Horak, Sarah Hunt, Mark Jensen, Peter Keating, Ian King, Kilannin Krysiak, Melissa Landrum, Mark Lawler, Michele LeNoue-Newton, Aitana Lebrand, Paul Leo, Koh Liang Kai, Rachel Liao, Nuria Lopez-Bigas, Subha Madhavan, Adam Margolin, David Masica, Georgia Mayfield, Julie McMurry, Christine Micheal, Susan Mockus, Chris Mungall, Kevin Osborn, Ravi Pandya, Sara Patterson, Beth Pitel, Gordana Raca, Erin Ramos, Shruti Rao, Damian Rieke, Deborah Ritter, Peter Robinson, Peter Rogan, Jeffrey Rosenfeld, Sameek Roychowdhury, Gabe Rudy, Chris Sander, Andrea Sboner, Lynn Schriml, Nikolaus Schultz, Alexander Senf, Ozman Ugur Sezerman, Mamatha Shekar, Xuan Shirley Li, Lillian Siu, Heidi Sofia, Dmitriy Sonkin, Vipin Sreedharan, Daniel Stekhoven, Greg Stupp, Andrew Su, David Tamborero, Bin Tean Teh, Nora Toussaint, Eli Van Allen, Etienne Vignola-Gagné, Ioannis Vlachos, Andra Waagmeester, Alex Wagner, Brian Walsh, Jeremy Warner, Joachim Weischenfeldt, Trish Whetzel, Julia Wilson, Chunlei Wu, Andy Yates, Andrey Zapariy, Alexander Wait Zarank, Zhenyu Zhang.

F32 CA206247



K99 HG010157





Thank You

Questions?

Email: awagner24@wustl.edu

This slide deck is at: <http://bit.ly/WagnerUZH19> (case sensitive)



Resources

ClViC: civicdb.org

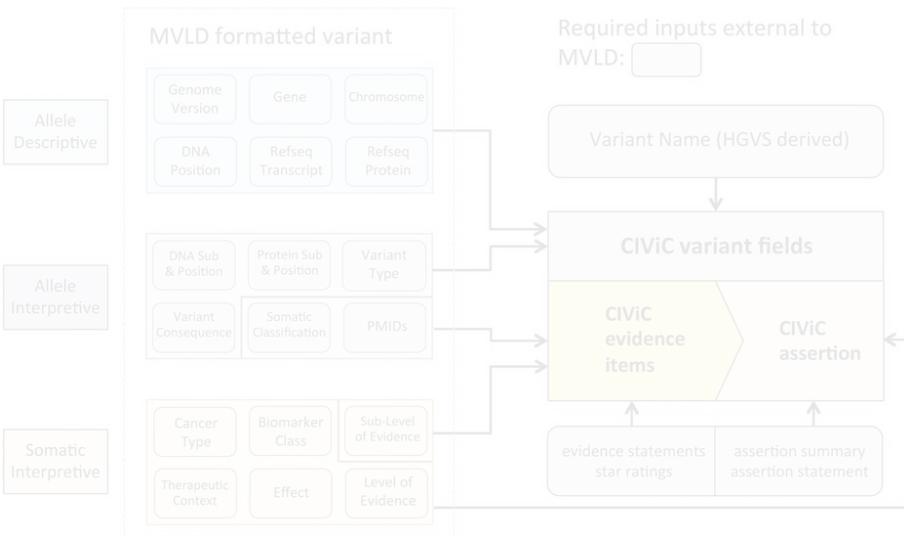
ClViCpy: civicpy.org

VICC: cancervariants.org

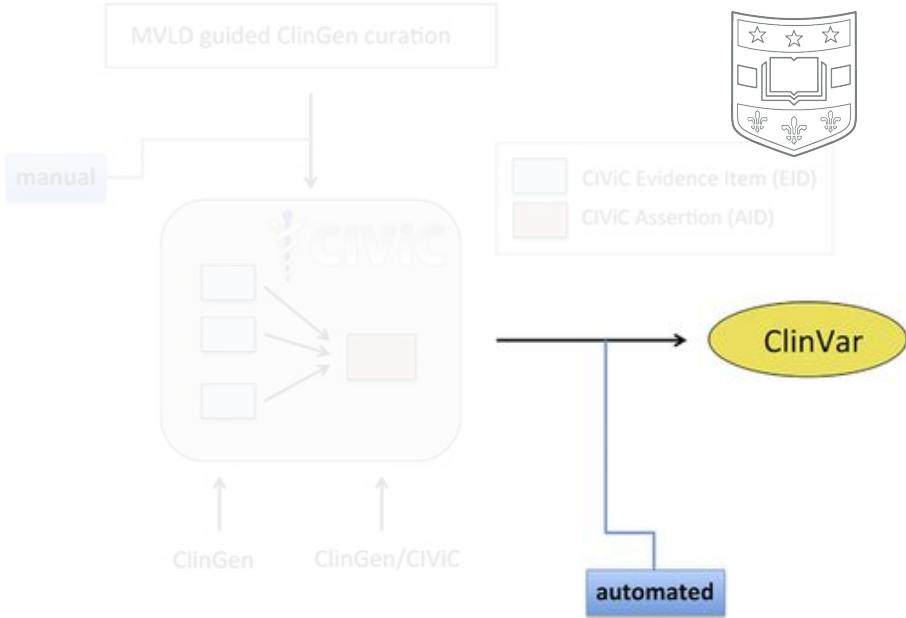
GA4GH-VR Specification: vr-spec.readthedocs.io

Remaining Content

ClinGen Somatic WG: MVLD Guidelines

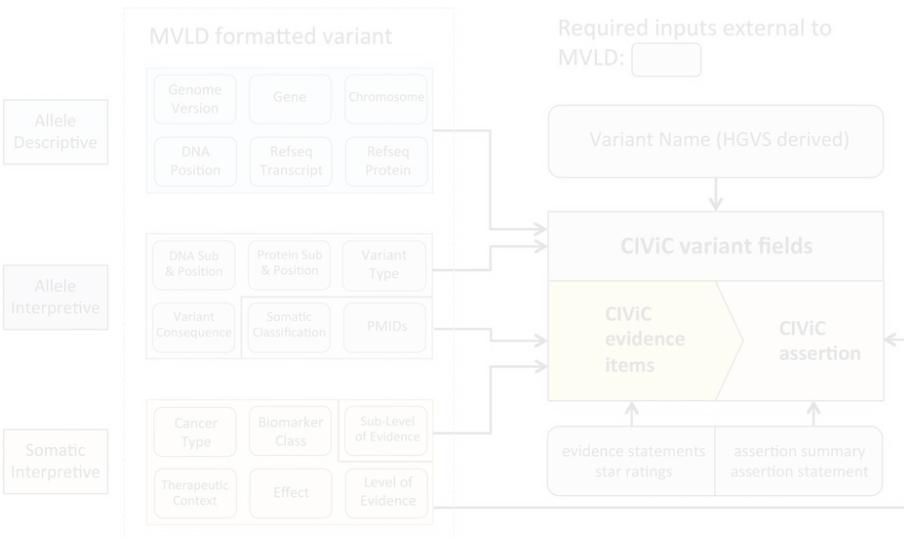


Danos AM, et al. *Hum. Mut.* 2018

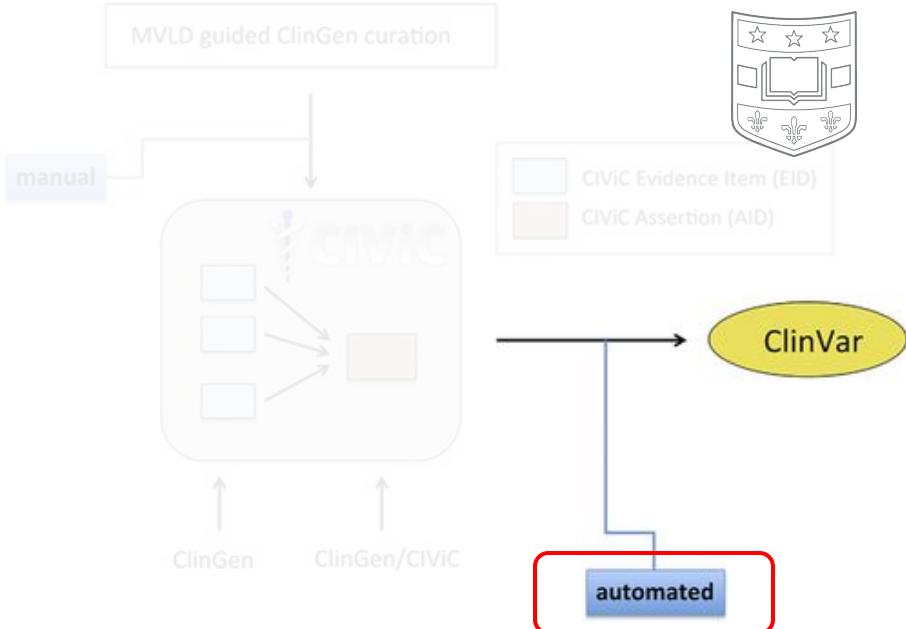


Danos AM, et al. *Hum. Mut.* 2018

ClinGen Somatic WG: MVLD Guidelines

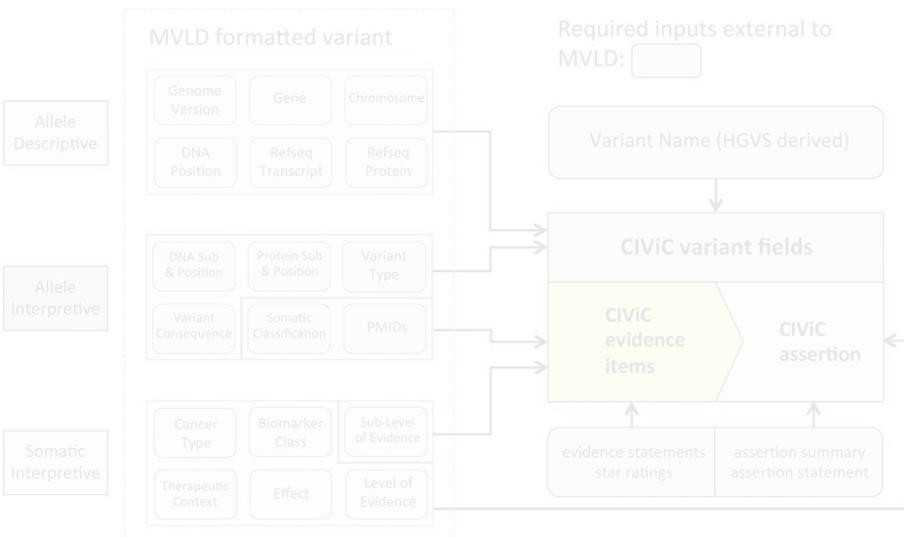


Danos AM, et al. *Hum. Mut.* 2018

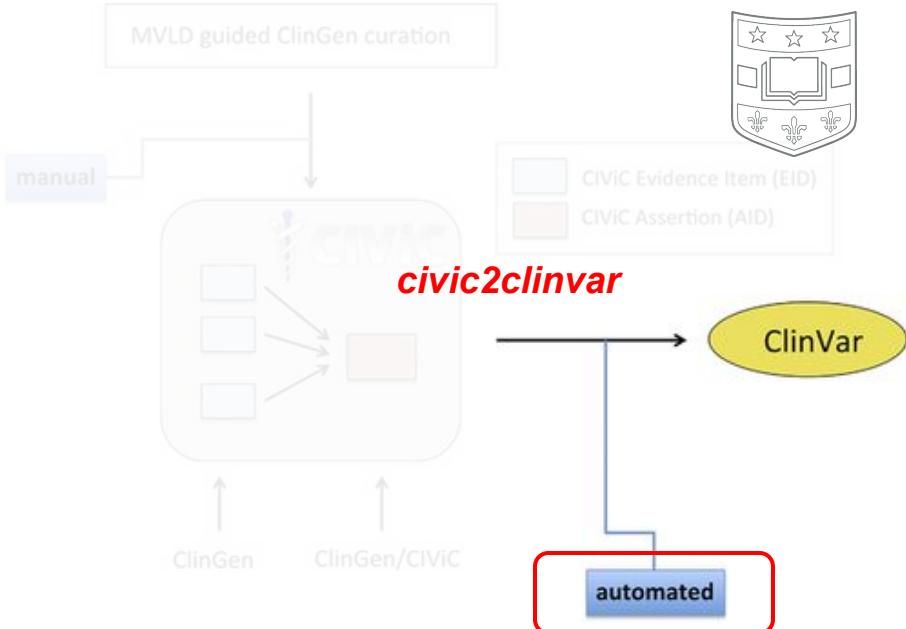


Danos AM, et al. *Hum. Mut.* 2018

ClinGen Somatic WG: MVLD Guidelines



Danos AM, et al. *Hum. Mut.* 2018

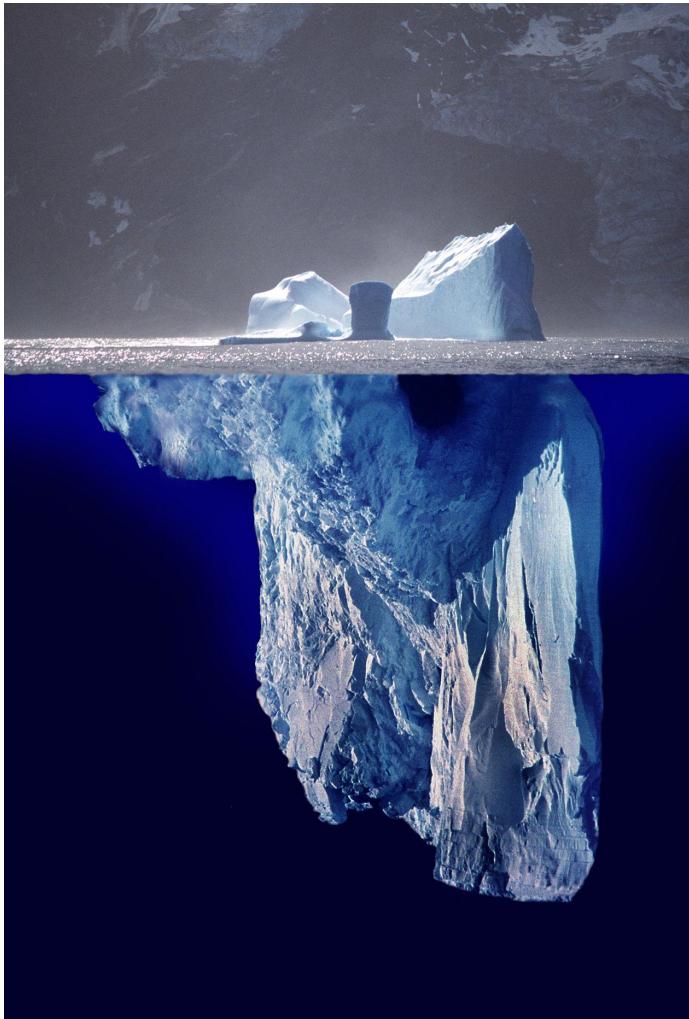


Danos AM, et al. *Hum. Mut.* 2018



civic2clinvar

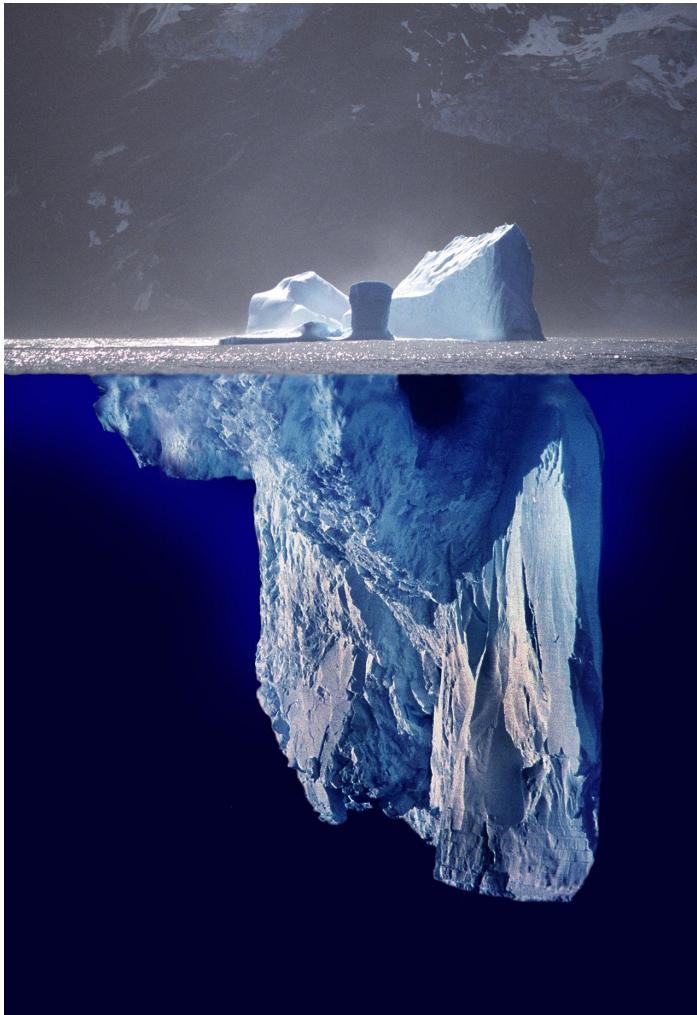




civic2clinvar



@handlerwagner



civic2clinvar



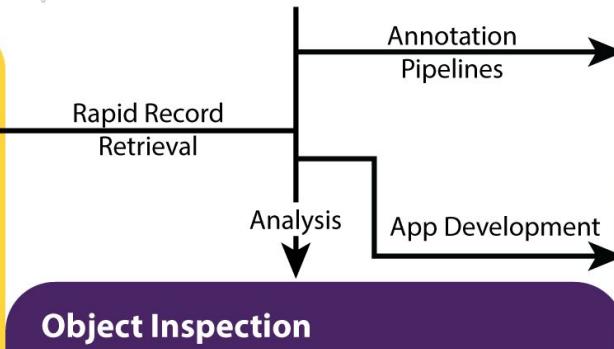
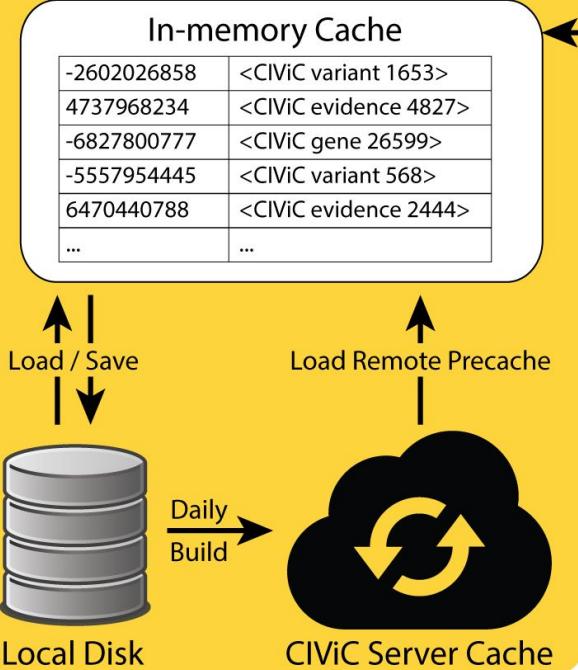
CIViCpy
The Python SDK for CIViC



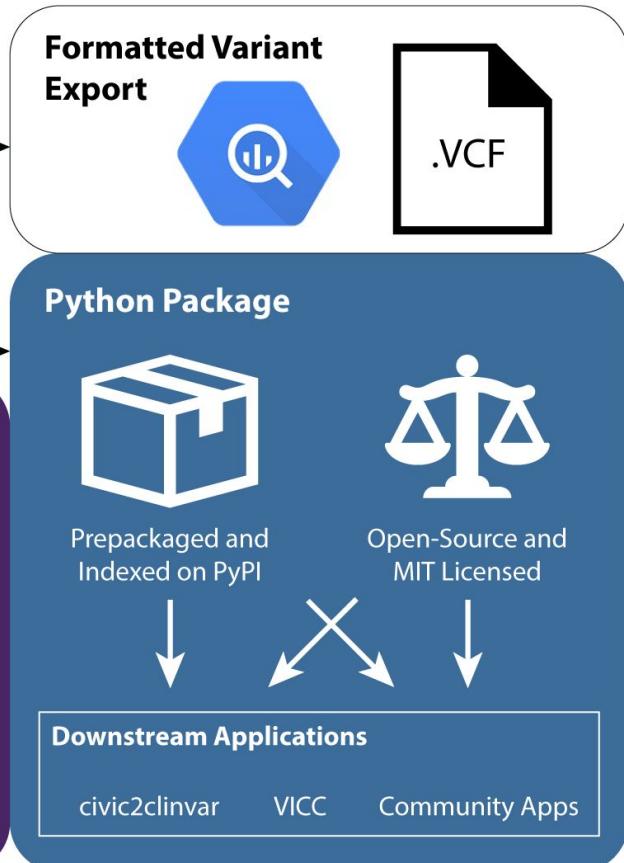
@handlerwagner



Caching



```
>>> from civicpy import civic
>>> assertion = civic.get_assertion_by_id(3)
>>> assertion.variant
<CIViC variant 499>
>>> assertion.variant.name
'ALK FUSIONS'
>>> assertion.amp_level
'Tier I - Level A'
>>> assertion.status
'accepted'
```





A) On Cache Update / Load

Extract Variants
Filter Missing Coordinates
Split Compound Coordinates
Convert to DataFrame
Sort

Sort Criteria

- 1) Chromosome
- 2) Start Coordinate
- 3) Stop Coordinate
- 4) Alternate Sequence

Variant Coordinate Index (VCI)

Chr	Start	Stop	Alt	v_hash
1	1000	1100	None	4398339850
1	50000	60000	None	251528793
1	51075	51075	C	-5647252664
1	68000	68000	T	3297066642
...

B)

Variants to Query

Extract Coordinates
Convert to CoordinateQuery Objects

Requirements

- 1) GRCh37 reference
- 2) Genomic coordinates
- 3) 1-based coordinates

Bulk Coordinate Query (Q)

Chr	Start	Stop	Alt	Key
1	100	150	None	MyRange1
1	45000	55000	None	MyRange2
1	51075	51075	A	MyVariant1.2
1	51075	51075	None	MyVariant1
...

C)

State 1:

VCI	Q
Record 1	Query 1
Record 2	Query 2
Record 3	Query 3
Record 4	Query 4
...	...

Compare Values:
Record 1 > Query 1

Actions:
Increment vci_ptr

State 2:

VCI	Q
Record 1	Query 1
Record 2	Query 2
Record 3	Query 3
Record 4	Query 4
...	...

Compare Values:
Record 1 < Query 2

Actions:
Increment vci_ptr

State 3:

VCI	Q
Record 1	Query 1
Record 2	Query 2
Record 3	Query 3
Record 4	Query 4
...	...

Compare Values:
Partial overlap

Actions:
Report match for some modes
Store vci_ptr to tmp
Increment vci_ptr

State 4:

VCI	Q
Record 1	Query 1
Record 2	Query 2
Record 3	Query 3
Record 4	Query 4
...	...

Compare Values:
Query 2 encompassing

Actions:
Report match for some modes
Increment vci_ptr

State 5:

VCI	Q
Record 1	Query 1
Record 2	Query 2
Record 3	Query 3
Record 4	Query 4
...	...

Compare Values:
Record 4 > Query 2

Actions:
Reset vci_ptr to tmp
Increment q_ptr

D)

Search Mode Behavior

Scenario

Search Mode*

Partial Overlay

Record	Query	Any	QE	RE	Exact
Record	Query	✓	✗	✗	✗

Query Encompassing

Record	Query	Any	QE	RE	Exact
Record	Query	✓	✓	✗	✗

Record Encompassing

Record	Query	Any	QE	RE	Exact
Record	Query	✓	✗	✓	✗

Coordinate Match

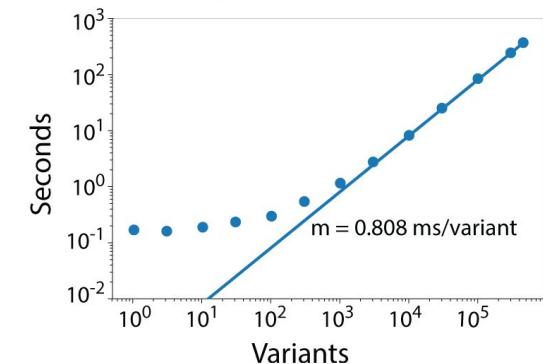
Record	Query	Any	QE	RE	Exact
Record	Query	✓	✓	✓	✗

Allele Match

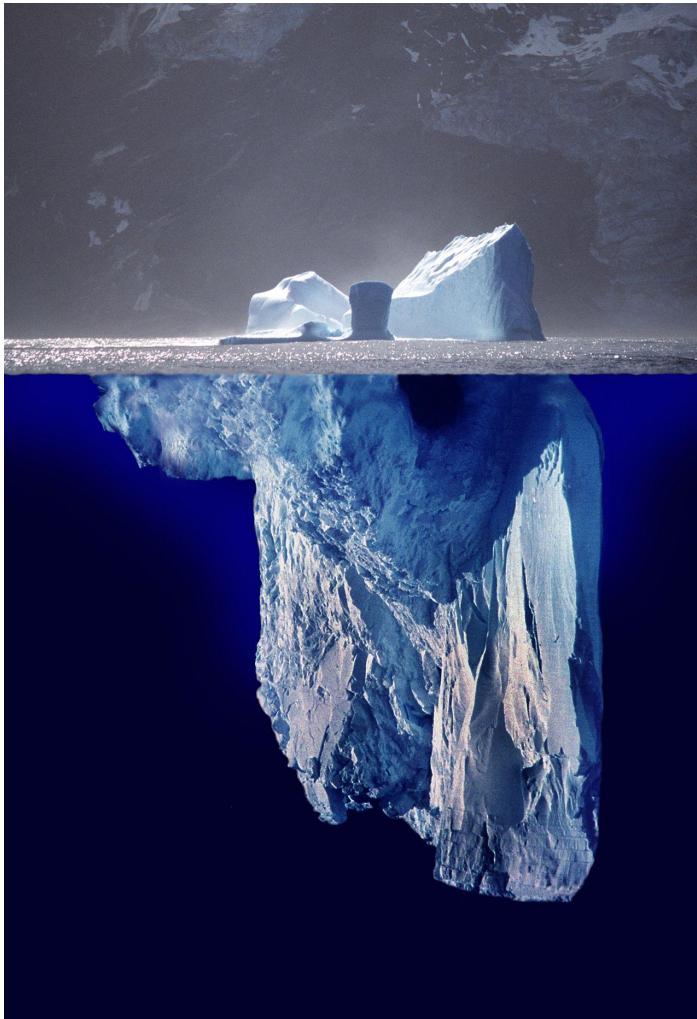
Record	Query	Any	QE	RE	Exact
Record	Query	✓	✓	✓	✓

Searches **>1,200 variants/second**

CViCpy Search Performance



@handlerwagner



CIViCpy

The Python SDK for CIViC

civicpy.org



@handlerwagner



Precision Cancer Treatment Cycle

Sample Collection → Sequencing → Analysis → Interpretation → Treatment

Users browse CIViC in order to obtain the latest collaboratively-sourced clinical interpretation of cancer variants.

In addition to the web app, CIViC users may download the database or make queries via the API.

All registered users may add, rate, discuss, and edit evidence, genes, and variants.

Researchers, clinicians, patient advocates and other members of the public contribute to CIViC.

Evidence statements must be derived from published research. Standard ontologies are used.

CIViC Data Exports

CIViC User Interface

Add Evidence Rate Evidence Edit Summaries
Discuss Collaborate

CIViC Collaboration

Research Scientists

Clinical Scientists

Patient Advocates

Contributors

PubMed

Disease Ontology

PubChem

Public Databases



Precision Cancer Treatment Cycle

Sample Collection → Sequencing → Analysis → Interpretation → Treatment

Users browse CIViC in order to obtain the latest collaboratively-sourced clinical interpretation of cancer variants.

In addition to the web app, CIViC users may download the database or make queries via the API.

All registered users may add, rate, discuss, and edit evidence, genes, and variants.

Researchers, clinicians, patient advocates and other members of the public contribute to CIViC.

Evidence statements must be derived from published research. Standard ontologies are used.

CIViC Data Exports

CIViC User Interface

Add Evidence Rate Evidence Edit Summaries
Discuss Collaborate

CIViC Collaboration

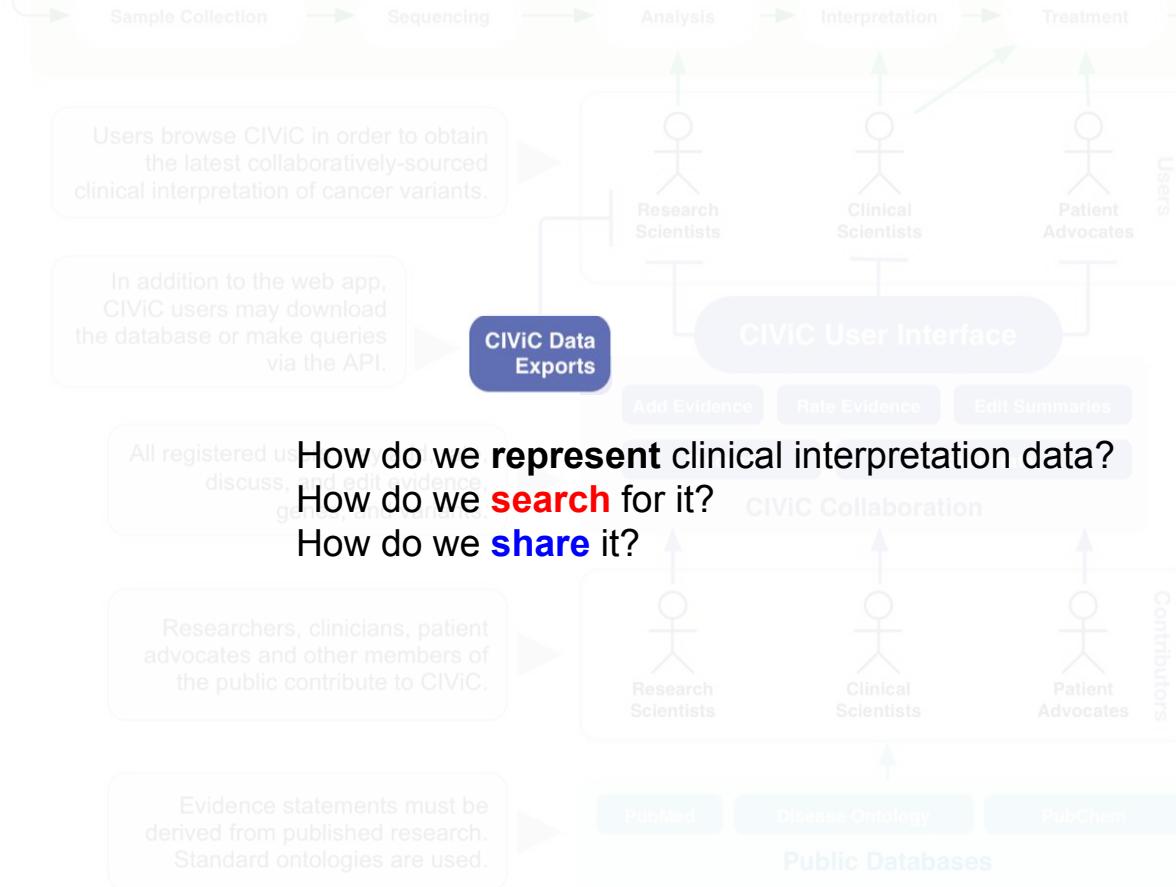
Research Scientists Clinical Scientists Patient Advocates
Contributors

Published Disease Ontology PubChem

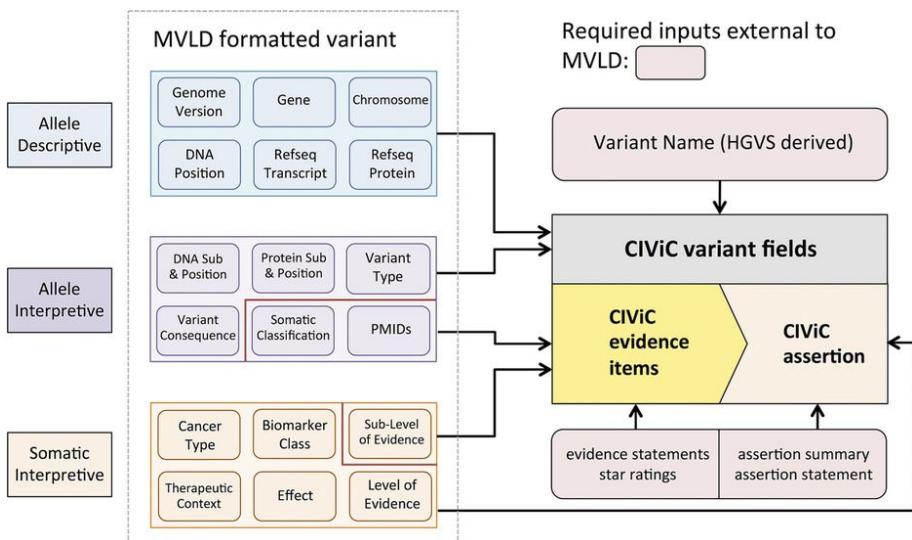
Public Databases



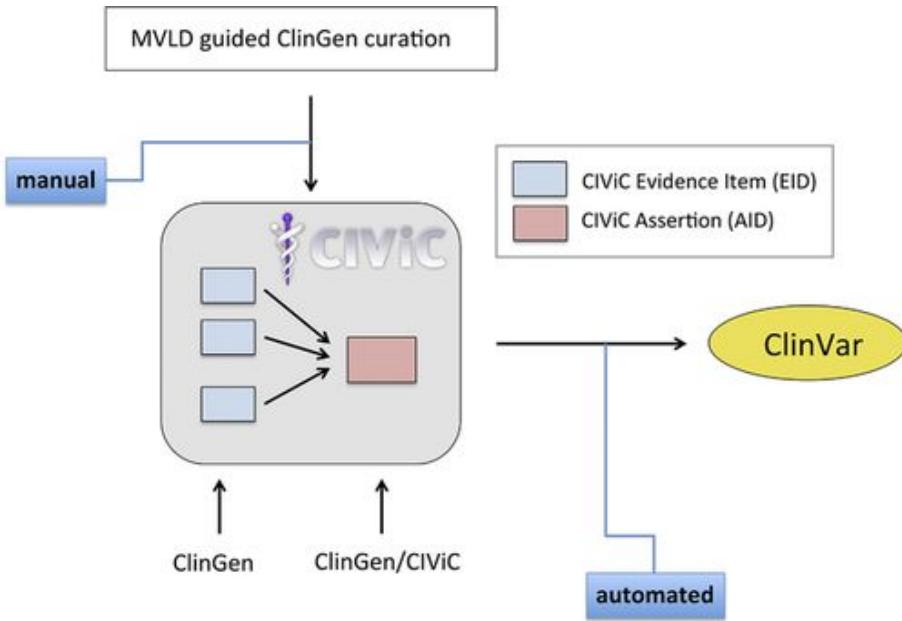
Precision Cancer Treatment Cycle



ClinGen Somatic WG: MVLD Guidelines



Danos AM, et al. *Hum. Mut.* 2018



Danos AM, et al. *Hum. Mut.* 2018