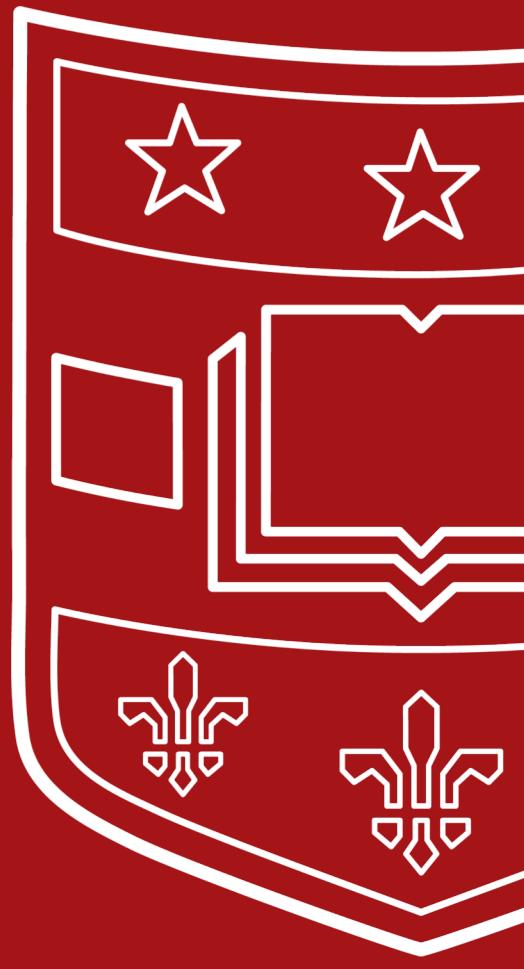


# Human Genome Variation

Clinical interpretation (I) and  
representation (II) of variants  
from human genomes

Alex H Wagner, PhD  
December, 2019 – UTH BIO390

 Washington University in St. Louis





# Part I – Clinical Interpretation of Variants



Swiss Institute  
of  
Bioinformatics

# Introduction to bioinformatics: Clinical Bioinformatics

Valérie Barbié, head of SIB Clinical Bioinformatics

Zürich, 26 November 2019

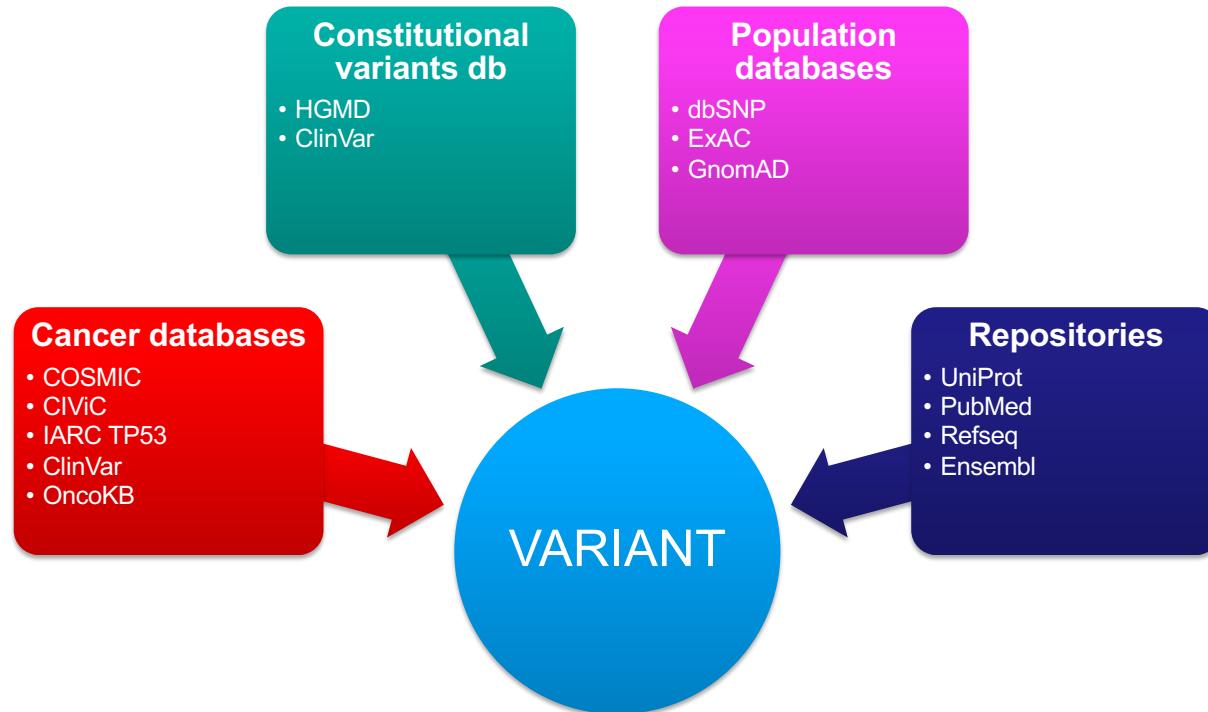


[www.sib.swiss](http://www.sib.swiss)

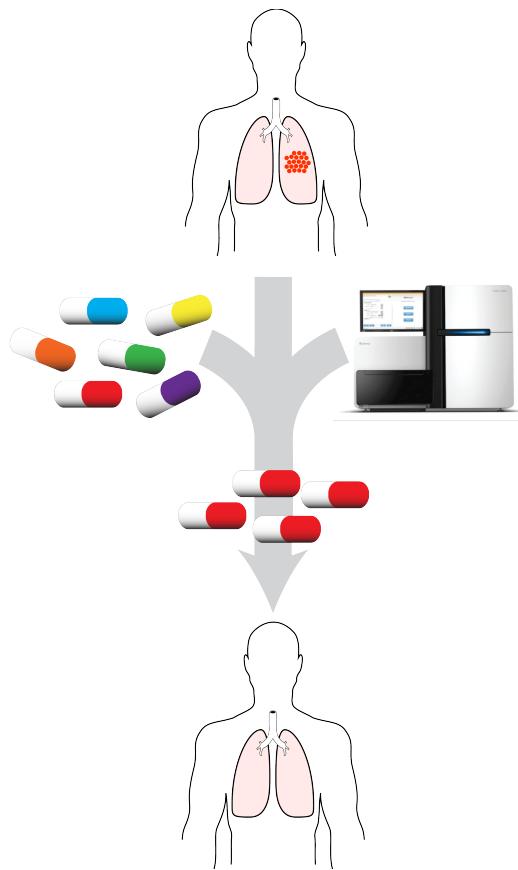


# Annotating a variant: knowledgebases

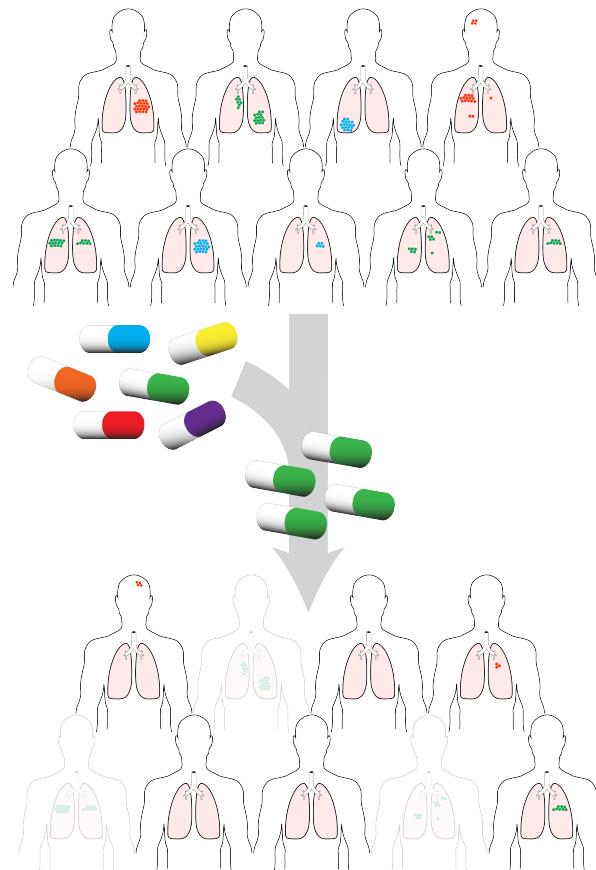
not exhaustive



## Precision medicine



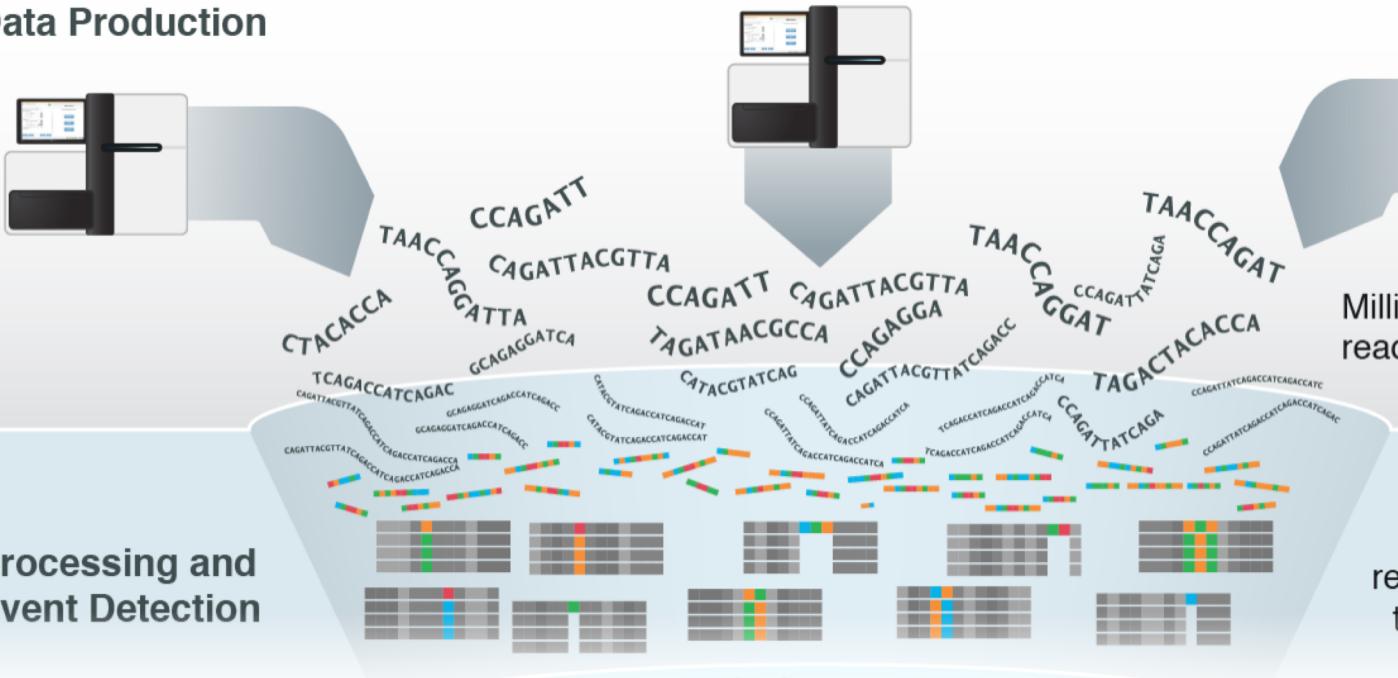
## Conventional therapies





# 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.

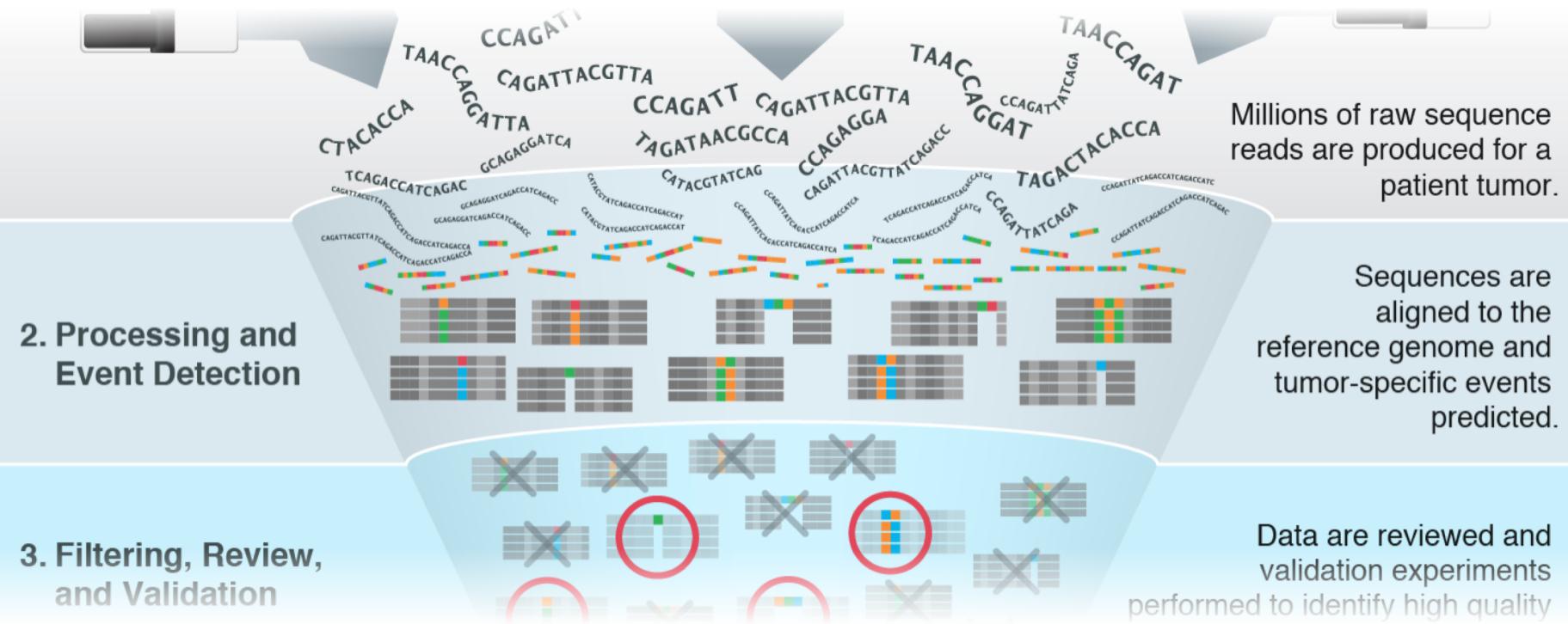


@handlerwagner

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



# 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



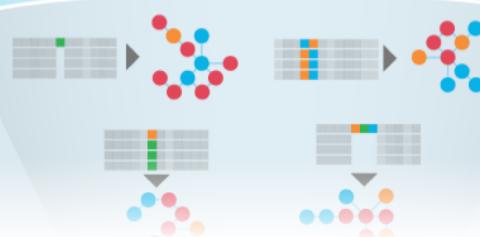
Sequencing data are aligned to the reference genome and tumor-specific events predicted.

## 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.



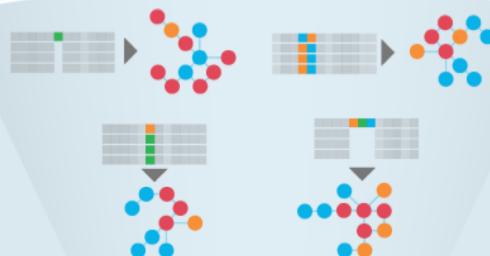
# Precision Oncology Bottleneck

## 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





# Precision Oncology Bottleneck



**How do we alleviate this bottleneck?**



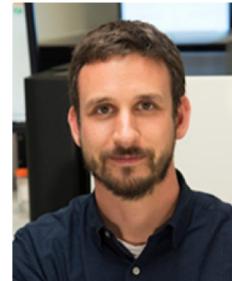
# CIViC

CLINICAL INTERPRETATIONS OF  
VARIANTS IN CANCER

[civicdb.org](http://civicdb.org)



Obi Griffith



Malachi Griffith

# CIViC principles emphasize collaborative, open sharing of interpretations

Public contributions, open discussion, curation standards and expert review

Researchers, clinicians,  
patient advocates and  
others

Public domain (CC0)  
license

No fees, anonymous access

Content provenance  
and creator  
acknowledgement

Structured data and  
APIs



# What is a variant of \*significance\* in cancer?

## A Prognostic of survival change

- *Biallelic CEBPA mutations are associated with improved overall survival in patients with acute myeloid leukemia*

## E Predictive of therapeutic response

- *BRAF V600E predicts sensitivity to vemurafenib*

## Q Diagnostic of tumor subtype

- *DNAJB1-PRKACA fusion differentiates fibrolamellar hepatocellular carcinoma from conventional HCC*

## ⚠ Predisposing for cancer development

- *Patients with the RUNX1 Y260\* mutation are associated with increased risk of developing acute myeloid leukemia*



# What is a variant of \*significance\* in cancer?



## Prognostic of survival change

- Biallelic *CEP1PA* mutations are associated with improved overall survival in patients



## Predictive

- *BRAF* V600E



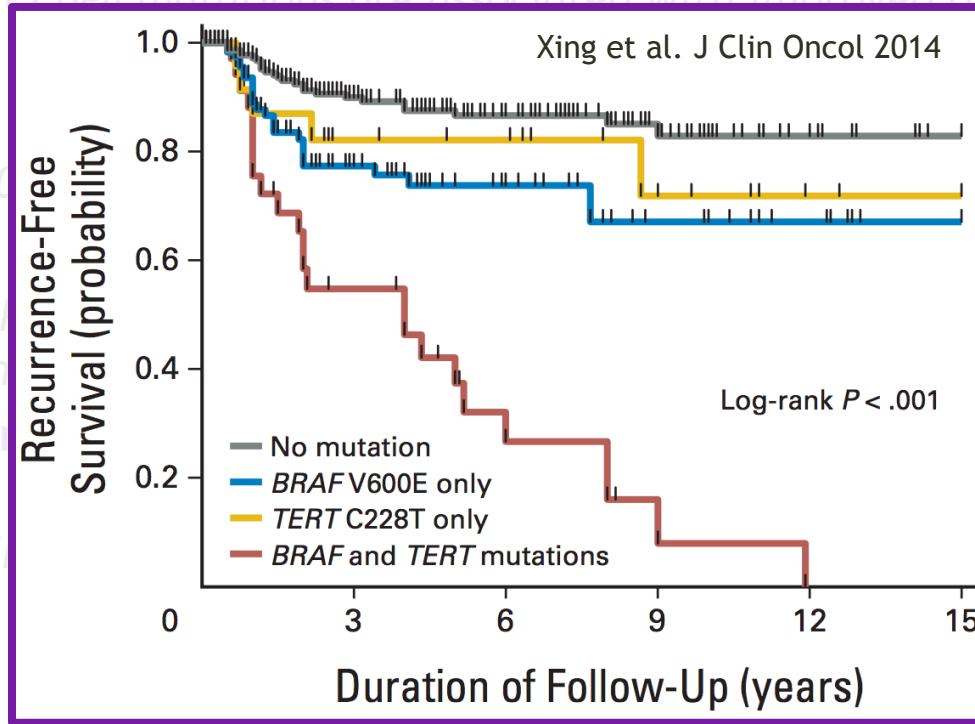
## Diagnostic

- *DNAJB1-L* from con

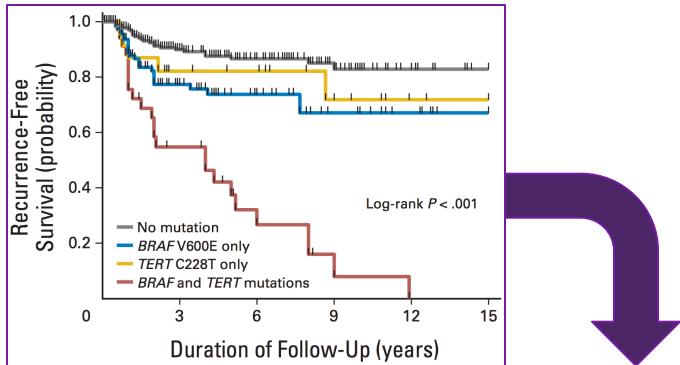


## Predisposition

- Patients developing



# CIViC provides a literature curation interface for the interpretation of clinically-relevant variants



## Structured Evidence Item

EVIDENCE EID656

Evidence Summary Evidence Talk

Submitted by gatoravi Last Modified by kkrysiak Last Reviewed by ahwagner Accepted by NickSpies

In patients with papillary thyroid cancer harboring both BRAF V600E and the TERT promotor mutation C228T (N=35), recurrence-free survival is worse than in patients harboring one of these mutations (N=159 BRAF, N=26 TERT promoter mutated) or no mutations in either gene (N=287)(P<0.001).

**Evidence Level:** B - Clinical      **Disease:** Papillary Thyroid Carcinoma

**Evidence Type:** Prognostic      **Associated Phenotype:** --

**Evidence Direction:** Supports      **Source:** Xing et al., 2014, J. Clin. Oncol.

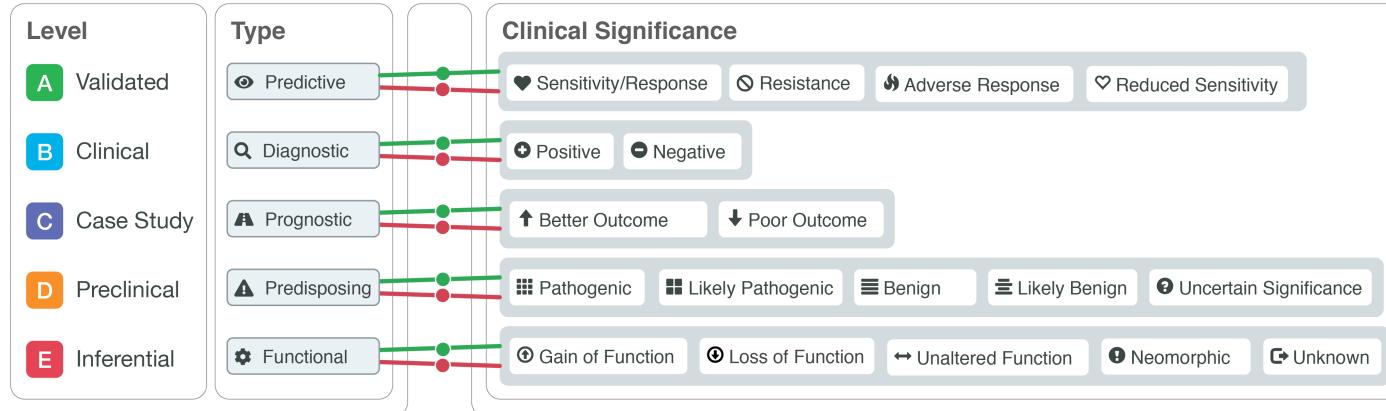
**Clinical Significance:** Poor Outcome      **PubMed ID:** [25024077](#)

**Variant Origin:** Somatic Mutation      **Clinical Trial:** --

**Trust Rating:** ★★★★★



# Curated evidence items are the foundational unit of CIViC



EVIDENCE EID656

[Evidence Summary](#) [Evidence Talk](#) [⚙️](#)

Submitted by gatoravi Last Modified by krysiak Last Reviewed by ahwagner Accepted by NickSpies

In patients with papillary thyroid cancer harboring both BRAF V600E and the TERT promotor mutation C228T (N=35), recurrence-free survival is worse than in patients harboring one of these mutations (N=159 BRAF, N=26 TERT promoter mutated) or no mutations in either gene (N=287)(P<0.001).

**Evidence Level:** B - Clinical

**Evidence Type:** Prognostic

**Evidence Direction:** Supports

**Clinical Significance:** Poor Outcome

**Variant Origin:** Somatic Mutation

**Disease:** Papillary Thyroid Carcinoma

**Associated Phenotype:** --

**Source:** Xing et al., 2014, J. Clin. Oncol.

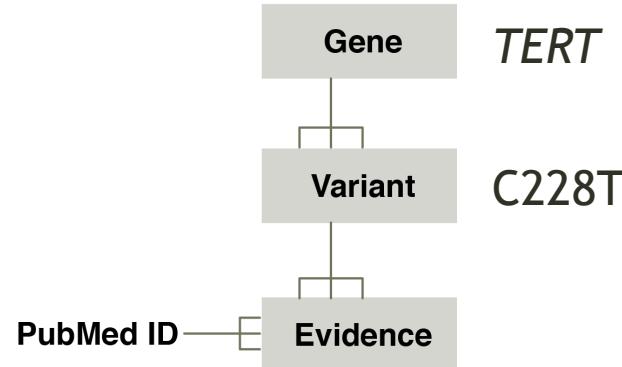
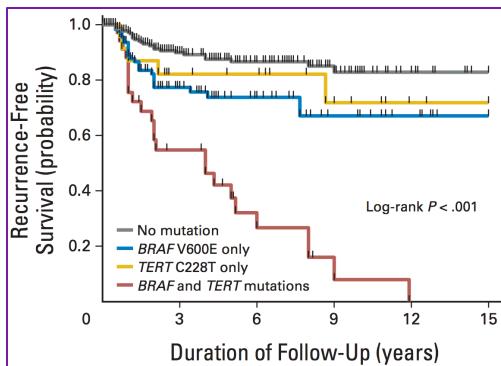
**PubMed ID:** [25024077](#)

**Clinical Trial:** --

**Trust Rating:** ★★★★★



# Curated evidence items are the foundational unit of CIViC



**EVIDENCE EID656**

Submitted by gatoravi Last Modified by kkrysiak Last Reviewed by ahwagner Accepted by NickSpies

In patients with papillary thyroid cancer harboring both BRAF V600E and the TERT promotor mutation C228T (N=35), recurrence-free survival is worse than in patients harboring one of these mutations (N=159 BRAF, N=26 TERT promoter mutated) or no mutations in either gene (N=287)(P<0.001).

**Evidence Level:** B - Clinical **Disease:** Papillary Thyroid Carcinoma

**Evidence Type:** Prognostic **Associated Phenotype:** -

**Evidence Direction:** Supports **Source:** Xing et al., 2014, J. Clin. Oncol.

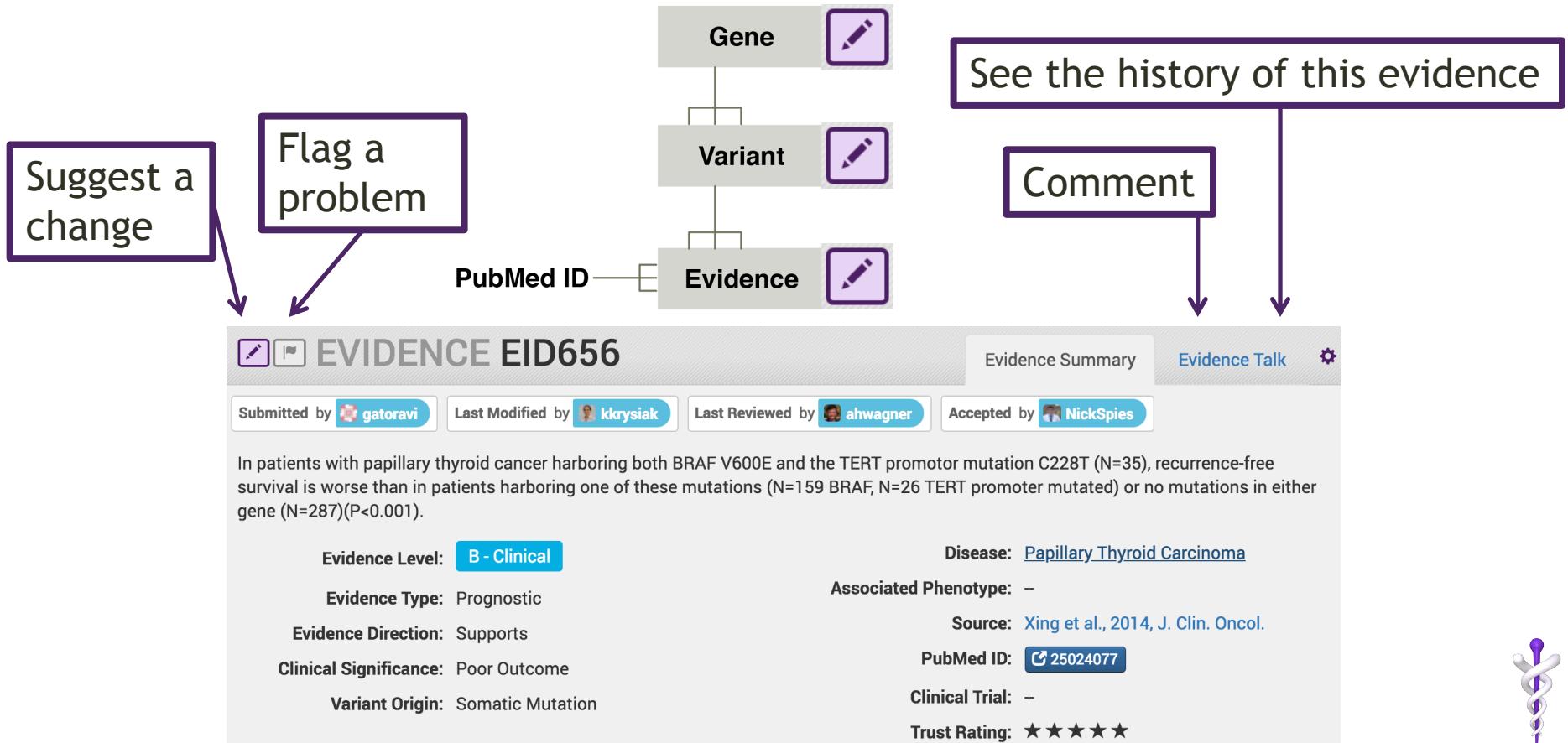
**Clinical Significance:** Poor Outcome **PubMed ID:** [25024077](#)

**Variant Origin:** Somatic Mutation **Clinical Trial:** -

**Trust Rating:** ★★★★★



# Curated evidence items are the foundational unit of CIViC



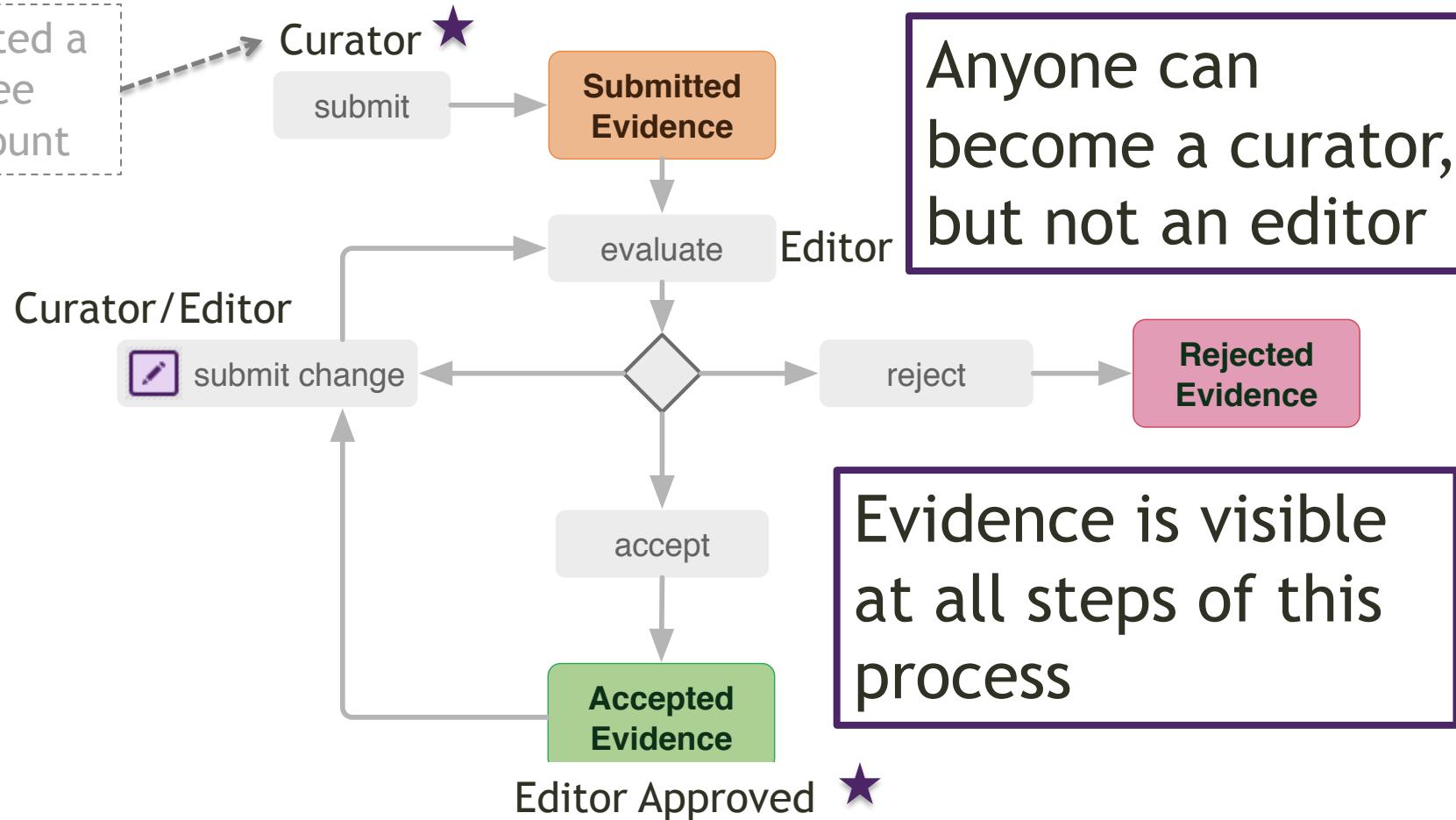
# Fundamental requirements of CIViC evidence

- Clinical relevance to cancer
- More than simple observation of a variant in a tumor
- Must not plagiarize the source material
- Published, traceable knowledge (PubMed ID or ASCO abstract)
  - Data available to support assertions
- Must not include personal health information
  - Published peer reviewed case reports are acceptable



# The CIViC evidence item life cycle is designed to promote quality, provenance, transparency and adaptation as knowledge evolves

Created a free account



# The goal of curating evidence is to document the clinical relevance of genes and variants, and ultimately to support final assertions

## CIViC

Go to Genes & Variants Go! BROWSE SEARCH ACTIVITY ADD ▾

### GENE ERBB2

Last Modified by obigriffith Last Reviewed by obigriffith

**Name:** erb-b2 receptor tyrosine kinase 2  
**Entrez Symbol:** ERBB2 Entrez ID: 2064  
**Aliases:** CD340, HER-2, HER-2/neu, HER2, MLN 19, NEU, NGL, TKR1  
**Chromosome:** 17 Start: 37844167 End: 37886679 Strand: 1 (GRCh37)

**Protein Domains:** Furin-like cysteine-rich domain, Furin-like repeat, Growth factor receptor cysteine-rich domain, Growth factor receptor domain 4, Leucine-rich repeat domain, L domain-like... (8 more)

**Pathways:** miR-targeted genes in muscle cell - TarBase, Leptin signaling pathway, Prolactin Signaling Pathway, Signaling Pathways in Glioblastoma, Integrated Pancreatic Cancer Pathway... (124 more)

[View MyGene.info Details](#)

### ERBB2 Variants & Variant Group

Show all: filter variants... Add Group

**AMPLIFICATION** D769H D769Y DEL 755-759 ERBB2 G776INSV\_G/C G309A G776L

G778\_P780DUP H878Y K753E K755S KINASE DOMAIN MUTATION L638S

L753E L755P L755S L755W L768S L865M M774DELINSLVL

M774INSAYVM MUTATION N857S NON-AMPLIFICATION OVEREXPRESSION P780INS

R678Q R896C S310F/Y SERUM LEVELS T798M T862A V773 V773A

V773L V777L V842I Y772\_A775DUP

### VARIANT V777L

Variant Summary Variant Talk

### VARIANT V777L

Last Modified by Lymzyell Last Reviewed by alhwagner Last Commented On by alhwagner

Allele Registry ID: CA135387

Ref. Build: GRCh37 Ensembl Version: 75

Chr.	Start	Stop	Ref. s	Var. Bases
17	37881000	37881000	G	T

Rep. Transcript ENST00000269571.5 Edit Coordinates

**Sources:**  
Bose et al., 2013, Cancer Discov

**HGVIS Expressions:**  
NM\_00448.3:c.232G>T, NP\_00449.2:p.Val777Leu, ENST00000269571.5:c.232G>T, and NC\_00017.10:g.37881000G>T

**ClinVar ID:** 44991 **ClinVar Clinical Significance:** Likely pathogenic

**COSMIC ID:** COSM14062 **dbSNP RSIID:** rs121913471 **HGVIS ID:** chr17:g.37881000G>T

**SnpEff Effect:** missense variant **SnpEff Impact:** MODERATE **gnomAD Adj. AF:** –

[View MyVariant.info Details](#)

### Evidence for V777L 4 total items

EID	DESC	DIS	DRUGS	EL	ET	ED	CS	VO	TR
5817	In a cohort of 205 Her2-recept...	Her2-receptor Positive Breast ...	Trastuzumab Emtansine, Tras...	C	🕒	⬆️	🚫	...	3 ★
288	In MCF10A cell lines, the V777...	Breast Cancer	Neratinib	D	🕒	⬆️	❤️	...	5 ★
1177	Colon cancer patient derived ...	Colon Cancer	Lapatinib, Neratinib, Trastuzu...	D	🕒	⬆️	❤️	...	4 ★
4452	bp] In an in vitro study, NCI-H...	Colorectal Cancer	Cetuximab	D	🕒	⬆️	🚫	...	...

[Get Data](#) [Help](#)

### EVIDENCE EID288

Submitted by NickSpies Accepted by kmyaski

In MCF10A cell lines, the V777L mutation was shown to be sensitive to neratinib.

**Evidence Level:** D - Preclinical

**Evidence Type:** Predictive

**Evidence Direction:** Supports

**Clinical Significance:** Sensitivity

**Variant Origin:** Somatic Mutation

**Drug:** Neratinib

**Disease:** Breast Cancer

**Associated Phenotype:** –

**Citation:** Bose et al., 2013, Cancer Discov

**PubMed ID:** 23220980

**Clinical Trial:** –

**Trust Rating:** ★★★★

[Evidence Summary](#) [Evidence Talk](#)

# Assertions are built from a collection of evidence items

- Each assertion aggregates multiple CIViC evidence items into a single clinical consensus
  - Specific variant, disease, evidence type, [drug]
- Underlying data is carefully organized and transparent, allowing for rapid updates and reassessment as new information emerges
- Once finalized, assertions are submitted to ClinVar



# Assertions - example

NSCLC with EGFR  
L858R mutation is  
sensitive to  
erlotinib and  
gefitinib

- NCCN guidelines
- FDA approval

...

**ASSERTION AID5**

Submitted by arpaddanos Last Modified by ebarnell Last Reviewed by arpaddanos Accepted by ebarnell

Gene: EGFR Variant: L858R Variant Allele Registry ID: CA126713

Disease: Non-small Cell Lung Carcinoma

Associated Phenotype: -

**Summary:** Non-small cell lung cancer with EGFR L858R mutation is sensitive to erlotinib or gefitinib.

**Description:** L858R is among the most common sensitizing EGFR mutations in NSCLC, and is assessed via DNA mutational analysis including Sanger sequencing and next generation sequencing methods. Tyrosine kinase inhibitors erlotinib and gefitinib are associated with improved progression free survival over chemotherapy in EGFR L858R patients. NCCN guidelines recommend (category 1) erlotinib and gefitinib for NSCLC with sensitizing EGFR mutations, along with afatinib and osimertinib.

ClinVar ID 16609	ClinVar Clinical Significance Pathogenic	
COSMIC ID COSM6224	dbSNP RSID rs121434568	HGVS ID chr7:g.55259515T>G
SnpEff Effect structural interaction variant	SnpEff Impact HIGH	gnomAD Adj. AF 
View MyVariant.info Details		

Assertion Type: Predictive  
Assertion Direction: Supports  
Clinical Significance: Sensitivity  
Drugs: Erlotinib and Gefitinib  
Drug Interaction Type: Substitutes  
AMP Category: Tier I - Level A  
NCCN Guideline: Non-Small Cell Lung Cancer (v3.2018)  
Regulatory Approval:   
FDA Companion Test:

Evidence Grid Evidence Cards

**Evidence Supporting AID5** 14 total items

EID	GENE	VARIANT	DESC	DIS	DRUGS	EL	ET	ED	CS	VO	TR
2994	EGFR	L858R	On May 14, 2013, the...	Non-small Cell Lung ...	Erlotinib	A					5 ★
2621	EGFR	L858R	In a phase 3 clinical t...	Non-small Cell Lung ...	Gefitinib	B					4 ★
1665	EGFR	L858R	90 NSCLC patients w...	Non-small Cell Lung ...	Gefitinib	B					3 ★
2634	EGFR	L858R	In a phase 3 clinical t...	Lung Adenocarcinoma	Gefitinib	B					3 ★
229	EGFR	L858R	There is no statistica...	Non-small Cell Lung ...	Erlotinib, Gefitinib (S...	B					3 ★
885	EGFR	L858R	A randomized phase ...	Non-small Cell Lung ...	Erlotinib	B					3 ★
2624	EGFR	L858R	Mutational profiling ...	Lung Adenocarcinoma	Gefitinib	C					1 ★

# RULES

**Gene and Variant must exist in CIViC**

**ASSERTION AID5**

Submitted by arpaddanos Last Modified by ebarnell Last Reviewed by arpaddanos Accepted by ebarnell

Gene: EGFR Variant: L858R Variant Allele Registry ID: CA126713

Disease: Non-small Cell Lung Carcinoma

Associated Phenotype: –

Summary: Non-small cell lung cancer with EGFR L858R mutation is sensitive to erlotinib or gefitinib.

Description: L858R is among the most common sensitizing EGFR mutations in NSCLC, and is assessed via DNA mutational analysis including Sanger sequencing and next generation sequencing methods. Tyrosine kinase inhibitors erlotinib and gefitinib are associated with improved progression free survival over chemotherapy in EGFR L858R patients. NCCN guidelines recommend (category 1) erlotinib and gefitinib for NSCLC with sensitizing EGFR mutations, along with afatinib and osimertinib.

ClinVar ID	ClinVar Clinical Significance	
16609	Pathogenic	
COSMIC ID COSM6224	dbSNP RSID rs121434568	HGVS ID chr7:g.55259515T>G
SnpEff Effect structural interaction variant	SnpEff Impact HIGH	gnomAD Adj. AF --
<a href="#">View MyVariant.info Details</a>		

MyVariant.info

Assertion Type: Predictive  
Assertion Direction: Supports  
Clinical Significance: Sensitivity  
Drugs: Erlotinib and Gefitinib  
Drug Interaction Type: Substitutes  
AMP Category: Tier I - Level A  
NCCN Guideline: Non-Small Cell Lung Cancer (v3.2018)  
Regulatory Approval: ✓  
FDA Companion Test: ✓

Evidence Grid Evidence Cards

**Evidence Supporting AID5** 14 total items

EID	GENE	VARIANT	DESC	DIS	DRUGS	EL	ET	ED	CS	VO	TR
2994	EGFR	L858R	On May 14, 2013, the...	Non-small Cell Lung ...	Erlotinib	A					5 ★
2621	EGFR	L858R	In a phase 3 clinical t...	Non-small Cell Lung ...	Gefitinib	B					4 ★
1665	EGFR	L858R	90 NSCLC patients w...	Non-small Cell Lung ...	Gefitinib	B					3 ★
2634	EGFR	L858R	In a phase 3 clinical t...	Lung Adenocarcinoma	Gefitinib	B					3 ★
229	EGFR	L858R	There is no statistica...	Non-small Cell Lung ...	Erlotinib, Gefitinib (S...	B					3 ★
885	EGFR	L858R	A randomized phase ...	Non-small Cell Lung ...	Erlotinib	B					3 ★
2624	EGFR	L858R	Mutational profiling ...	Lung Adenocarcinoma	Gefitinib	C					1 ★

# RULES

Gene and Variant must exist in CIViC

Assertion AID5

Submitted by arpaddanos Last Modified by ebarnell Last Reviewed by arpaddanos Accepted by ebarnell

Gene: EGFR Variant: L858R Variant Allele Registry ID: CA126713

Disease: Non-small Cell Lung Carcinoma

Associated Phenotype: –

Summary: Non-small cell lung cancer with EGFR L858R mutation is sensitive to erlotinib or gefitinib.

Description: L858R is among the most common sensitizing EGFR mutations in NSCLC, and is assessed via DNA mutational analysis including Sanger sequencing and next generation sequencing methods. Tyrosine kinase inhibitors erlotinib and gefitinib are associated with improved progression free survival over chemotherapy in EGFR L858R patients. NCCN guidelines recommend (category 1) erlotinib and gefitinib for NSCLC with sensitizing EGFR mutations, along with afatinib and osimertinib.

ClinVar ID	ClinVar Clinical Significance	
16609	Pathogenic	
COSMIC ID	dbSNP RSID	HGVS ID
COSM6224	rs121434568	chr7:g.55259515T>G
SnpEff Effect	SnpEff Impact	gnomAD Adj. AF
structural interaction variant	HIGH	–

[View MyVariant.info Details](#)

Assertion Type: Predictive  
Assertion Direction: Supports  
Clinical Significance: Sensitivity  
Drugs: Erlotinib and Gefitinib  
Drug Interaction Type: Substitutes  
AMP Category: Tier I - Level A  
NCCN Guideline: Non-Small Cell Lung Cancer (v3.2018)  
Regulatory Approval: ✓  
FDA Companion Test: ✓

Evidence Grid Evidence Cards

Evidence Supporting AID5 14 total items

EID	GENE	VARIANT	DESC	DIS	DRUGS	EL	ET	ED	CS	VO	TR
2994	EGFR	L858R	On May 14, 2013, the...	Non-small Cell Lung ...	Erlotinib	A	🕒	👉	❤️	...	5 ★
2621	EGFR	L858R	In a phase 3 clinical t...	Non-small Cell Lung ...	Gefitinib	B	🕒	👉	❤️	...	4 ★
1665	EGFR	L858R	90 NSCLC patients w...	Non-small Cell Lung ...	Gefitinib	B	🕒	👉	❤️	...	3 ★
2634	EGFR	L858R	In a phase 3 clinical t...	Lung Adenocarcinoma	Gefitinib	B	🕒	👉	❤️	...	3 ★
229	EGFR	L858R	There is no statistica...	Non-small Cell Lung ...	Erlotinib, Gefitinib (S...	B	🕒	👉	❤️	...	3 ★
885	EGFR	L858R	A randomized phase ...	Non-small Cell Lung ...	Erlotinib	B	🕒	👉	❤️	...	3 ★
2624	EGFR	L858R	Mutational profiling ...	Lung Adenocarcinoma	Gefitinib	C	🕒	👉	❤️	...	1 ★

Evidence must be fully curated

# RULES

## ASSERTION AID5

[Assertion Summary](#)[Assertion Talk](#)Submitted by  arpaddanosLast Modified by  ebarnellLast Reviewed by  arpaddanosAccepted by  ebarnell

Gene: EGFR Variant: L858R Variant Allele Registry ID: CA126713

Disease: Non-small Cell Lung Carcinoma

Associated Phenotype: -

**Summary:** Non-small cell lung cancer with EGFR L858R mutation is sensitive to erlotinib or gefitinib.

**Description:** L858R is among the most common sensitizing EGFR mutations in NSCLC, and is assessed via DNA mutational analysis including Sanger sequencing and next generation sequencing methods. Tyrosine kinase inhibitors erlotinib and gefitinib are associated with improved progression free survival over chemotherapy in EGFR L858R patients. NCCN guidelines recommend (category 1) erlotinib and gefitinib for NSCLC with sensitizing EGFR mutations, along with afatinib and osimertinib.

ClinVar ID	ClinVar Clinical Significance	
16609	Pathogenic	
COSMIC ID	dbSNP RSID	HGVs ID
COSM6224	rs121434568	chr7:g.55259515T>G
SnpEff Effect	SnpEff Impact	gnomAD Adj. AF
structural interaction variant	HIGH	—
<a href="#">View MyVariant.info Details</a>		

Assertion Type: Predictive

Assertion Direction: Supports

Clinical Significance: Sensitivity

Drugs: Erlotinib and Gefitinib

Drug Interaction Type: Substitutes

AMP Category: Tier I - Level A

NCCN Guideline: Non-Small Cell Lung Cancer (v3.2018)

Regulatory Approval: FDA Companion Test: 

Critical rationale / decisions are described

Evidence must be fully curated

[Evidence Grid](#) [Evidence Cards](#)

Evidence Supporting AID5 14 total items

EID	GENE	VARIANT	DESC	DIS	DRUGS	EL	ET	ED	CS	VO	TR	
2994	EGFR	L858R	On May 14, 2013, the...	Non-small Cell Lung ...	Erlotinib	A					5★	
2621	EGFR	L858R	In a phase 3 clinical t...	Non-small Cell Lung ...	Gefitinib	B					4★	
1665	EGFR	L858R	90 NSCLC patients w...	Non-small Cell Lung ...	Gefitinib	B					3★	
2634	EGFR	L858R	In a phase 3 clinical t...	Lung Adenocarcinoma	Gefitinib	B					3★	
229	EGFR	L858R	There is no statistica...	Non-small Cell Lung ...	Erlotinib, Gefitinib (S...	B					3★	
885	EGFR	L858R	A randomized phase ...	Non-small Cell Lung ...	Erlotinib	B					3★	
2624	EGFR	L858R	Mutational profiling ...	Lung Adenocarcinoma	Gefitinib	C					1★	

# RULES

Gene and Variant must exist in CIViC

Critical rationale / decisions are described

Evidence must be fully curated

**ASSERTION AID5**

Submitted by arpaddanos Last Modified by ebarnell Last Reviewed by arpaddanos Accepted by ebarnell

Gene: EGFR Variant: L858R Variant Allele Registry ID: CA126713

Disease: Non-small Cell Lung Carcinoma

Associated Phenotype: –

Summary: Non-small cell lung cancer with EGFR L858R mutation is sensitive to erlotinib or gefitinib.

Description: L858R is among the most common sensitizing EGFR mutations in NSCLC, and is assessed via DNA mutational analysis including Sanger sequencing and next generation sequencing methods. Tyrosine kinase inhibitors erlotinib and gefitinib are associated with improved progression free survival over chemotherapy in EGFR L858R patients. NCCN guidelines recommend (category 1) erlotinib and gefitinib for NSCLC with sensitizing EGFR mutations, along with afatinib and osimertinib.

ClinVar ID 16609	ClinVar Clinical Significance Pathogenic		MyVariant.info
COSMIC ID COSM6224	dbSNP RSID rs121434568	HGVS ID chr7:g.55259515T>G	
SnpEff Effect structural interaction variant	SnpEff Impact HIGH	gnomAD Adj. AF --	

[View MyVariant.info Details](#)

**Evidence Grid** **Evidence Cards**

**Evidence Supporting AID5** 14 total items

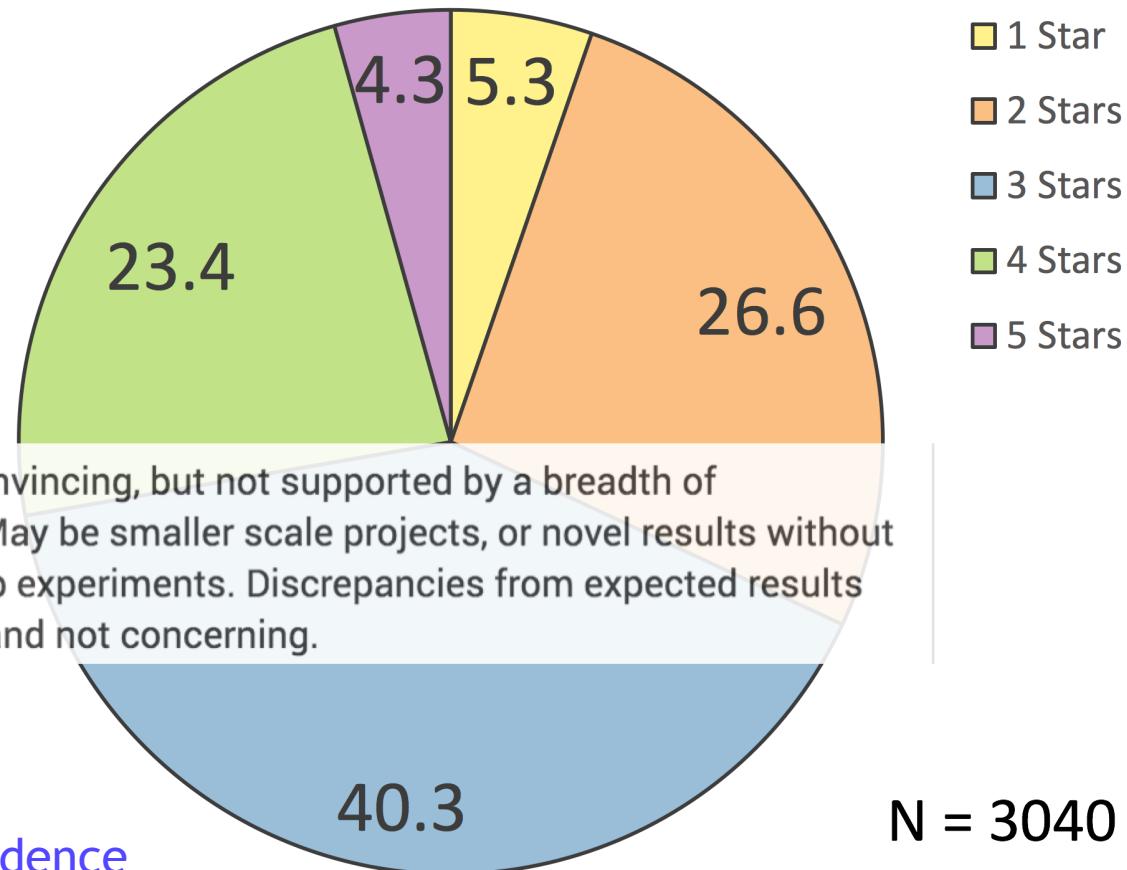
EID	GENE	VARIANT	DESC	DIS	DRUGS	EL	ET	ED	CS	VO	TR
2994	EGFR	L858R	On May 14, 2013, the...	Non-small Cell Lung ...	Erlotinib	A					5★
2621	EGFR	L858R	In a phase 3 clinical t...	Non-small Cell Lung ...	Gefitinib	B					4★
1665	EGFR	L858R	90 NSCLC patients w...	Non-small Cell Lung ...	Gefitinib	B					3★
2634	EGFR	L858R	In a phase 3 clinical t...	Lung Adenocarcinoma	Gefitinib	B					3★
229	EGFR	L858R	There is no statistica...	Non-small Cell Lung ...	Erlotinib, Gefitinib (S...	B					3★
885	EGFR	L858R	A randomized phase ...	Non-small Cell Lung ...	Erlotinib	B					3★
2624	EGFR	L858R	Mutational profiling ...	Lung Adenocarcinoma	Gefitinib	C					1★

Assertion Type: Predictive  
Assertion Direction: Supports  
Clinical Significance: Sensitivity  
Drugs: Erlotinib and Gefitinib  
Drug Interaction Type: Substitutes  
AMP Category: Tier I - Level A  
NCCN Guideline: Non-Small Cell Lung Cancer (v3.2018)  
Regulatory Approval:   
FDA Companion Test:

At least 1 Evidence item with 3+ stars

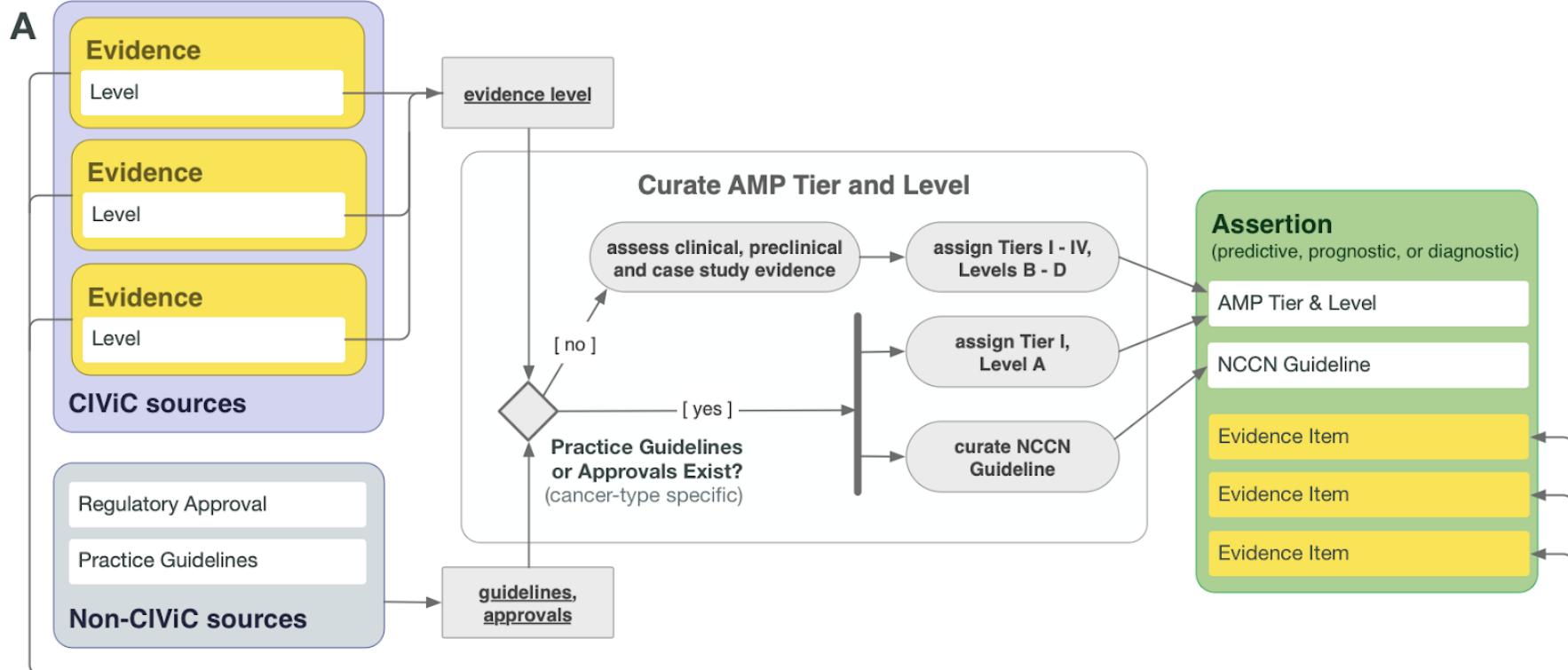
# Star ratings help guide Assertion creation

- Moderated evidence is appropriately distributed



<https://civicdb.org/statistics/evidence>

# Assertions summarize multiple lines of CIViC evidence, external evidence and assign overall AMP Tier/Level



# RULES

Assertion AID5

Assertion Summary

Assertion Talk

Submitted by arpaddanos

Last Modified by ebarnell

Last Reviewed by arpaddanos

Accepted by ebarnell

Gene: EGFR Variant: L858R Variant Allele Registry ID: CA126713

Disease: Non-small Cell Lung Carcinoma

Associated Phenotype: -

Summary: Non-small cell lung cancer with EGFR L858R mutation is sensitive to

Assertion Type: Predictive

Assertion Direction: Supports

Tier I: Variants of Strong Clinical Significance

Therapeutic, prognostic & diagnostic

Level A Evidence

FDA-approved therapy  
Included in professional guidelines

Level B Evidence

Well-powered studies with consensus from experts in the field

Tier II: Variants of Potential Clinical Significance

Therapeutic, prognostic & diagnostic

Level C Evidence

FDA-approved therapies for different tumor types or investigational therapies  
Multiple small published studies with some consensus

Level D Evidence

Preclinical trials or a few case reports without consensus

Tier III: Variants of Unknown Clinical Significance

Not observed at a significant allele frequency in the general or specific subpopulation databases, or pan-cancer or tumor-specific variant databases

No convincing published evidence of cancer association

Tier IV: Benign or Likely Benign Variants

Observed at significant allele frequency in the general or specific subpopulation databases  
No existing published evidence of cancer association

8)

Help

TR

▼

5★  
4★  
3★  
3★  
3★  
3★  
1★

# RULES

## ASSERTION AID5

[Assertion Summary](#)[Assertion Talk](#)Submitted by  arpaddanosLast Modified by  ebarnellLast Reviewed by  arpaddanosAccepted by  ebarnell

### Tier I: Variants of Strong Clinical Significance

Therapeutic, prognostic & diagnostic

#### Level A Evidence

FDA-approved therapy  
Included in professional guidelines

#### Level B Evidence

Well-powered studies with consensus from experts in the field

Gene: EGFR Variant: L858R Variant Allele Registry ID: CA126713

Disease: Non-small Cell Lung Carcinoma

Associated Phenotype: –

Summary: Non-small cell lung cancer with EGFR L858R mutation is sensitive to erlotinib or gefitinib.

Description: L858R is among the most common sensitizing EGFR mutations in NSCLC, and is assessed via DNA mutational analysis including Sanger sequencing and next generation sequencing methods. Tyrosine kinase inhibitors erlotinib and gefitinib are associated with improved progression free survival over chemotherapy in EGFR L858R patients. NCCN guidelines recommend (category 1) erlotinib and gefitinib for NSCLC with sensitizing EGFR mutations, along with afatinib and osimertinib.

ClinVar ID	ClinVar Clinical Significance	
16609	Pathogenic	
COSMIC ID	dbSNP RSID	HGVS ID
COSM6224	rs121434568	chr7:g.55259515T>G
SnpEff Effect	SnpEff Impact	gnomAD Adj. AF
structural interaction variant	HIGH	–
<a href="#">View MyVariant.info Details</a>		

[Evidence Grid](#)[Evidence Cards](#)

Evidence Supporting AID5 14 total items

EID	GENE	VARIANT	DESC	DIS	DRUGS	EL	ET	ED	CS	VO	TR	
2994	EGFR	L858R	On May 14, 2013, the...	Non-small Cell Lung ...	Erlotinib	A					5 ★	
2621	EGFR	L858R	In a phase 3 clinical t...	Non-small Cell Lung ...	Gefitinib	B					4 ★	
1665	EGFR	L858R	90 NSCLC patients w...	Non-small Cell Lung ...	Gefitinib	B					3 ★	
2634	EGFR	L858R	In a phase 3 clinical t...	Lung Adenocarcinoma	Gefitinib	B					3 ★	
229	EGFR	L858R	There is no statistica...	Non-small Cell Lung ...	Erlotinib, Gefitinib (S...	B					3 ★	
885	EGFR	L858R	A randomized phase ...	Non-small Cell Lung ...	Erlotinib	B					3 ★	
2624	EGFR	L858R	Mutational profiling ...	Lung Adenocarcinoma	Gefitinib	C					1 ★	

Assertion Type: Predictive

Assertion Direction: Supports

Clinical Significance: Sensitivity

Drugs: Erlotinib and Gefitinib

Drug Interaction Type: Substitutes

AMP Category: Tier I - Level A

NCCN Guideline: Non-Small Cell Lung Cancer (v3.2018)

Regulatory Approval:

FDA Companion Test:

At least 1 CIViC Level A or B Evidence item

# Creating evidence and assertions requires significant resources

**ASSERTION AID5**

Submitted by arpaddanos Last Modified by ebarnell Last Reviewed by arpaddanos Accepted by ebarnell

Gene: EGFR Variant: L858R Variant Allele Registry ID: CA126713  
Disease: Non-small Cell Lung Carcinoma  
Associated Phenotype: --  
Summary: Non-small cell lung cancer with EGFR L858R mutation is sensitive to erlotinib or gefitinib.  
Description: L858R is among the most common sensitizing EGFR mutations in NSCLC, and is assessed via DNA mutational analysis including Sanger sequencing and next generation sequencing methods. Tyrosine kinase inhibitors erlotinib and gefitinib are associated with improved progression free survival over chemotherapy in EGFR L858R patients. NCCN guidelines recommend (category 1) erlotinib and gefitinib for NSCLC with sensitizing EGFR mutations, along with afatinib and osimertinib.

ClinVar ID <a href="#">16609</a>	ClinVar Clinical Significance Pathogenic	
COSMIC ID <a href="#">COSM6224</a>	dbSNP RSID <a href="#">rs121434568</a>	HGVS ID <a href="#">chr7:g.55259515T&gt;G</a>
SnpEff Effect structural interaction variant	SnpEff Impact HIGH	gnomAD Adj. AF --

[View MyVariant.info Details](#)

Assertion Type: Predictive  
Assertion Direction: Supports  
Clinical Significance: Sensitivity  
Drugs: Erlotinib and Gefitinib  
Drug Interaction Type: Substitutes  
AMP Category: Tier I - Level A  
NCCN Guideline: Non-Small Cell Lung Cancer (v3.2018)  
Regulatory Approval: ✓  
FDA Companion Test: ✓

Evidence Grid Evidence Cards

**Evidence Supporting AID5** 14 total items

EID	GENE	VARIANT	DESC	DIS	DRUGS	EL	ET	ED	CS	VO	TR	
2994	EGFR	L858R	On May 14, 2013, the...	Non-small Cell Lung ...	Erlotinib	A	🕒	🕒	♥	...	5	★
2621	EGFR	L858R	In a phase 3 clinical t...	Non-small Cell Lung ...	Gefitinib	B	🕒	🕒	♥	...	4	★
1665	EGFR	L858R	90 NSCLC patients w...	Non-small Cell Lung ...	Gefitinib	B	🕒	🕒	♥	...	3	★
2634	EGFR	L858R	In a phase 3 clinical t...	Lung Adenocarcinoma	Gefitinib	B	🕒	🕒	♥	...	3	★
229	EGFR	L858R	There is no statistica...	Non-small Cell Lung ...	Erlotinib, Gefitinib (S...	B	🕒	🕒	♥	...	3	★
885	EGFR	L858R	A randomized phase ...	Non-small Cell Lung ...	Erlotinib	B	🕒	🕒	♥	...	3	★
2624	EGFR	L858R	Mutational profiling ...	Lung Adenocarcinoma	Gefitinib	C	🕒	🕒	♥	...	1	★



# How do we scale up without compromising quality (ClinGen Somatic)?

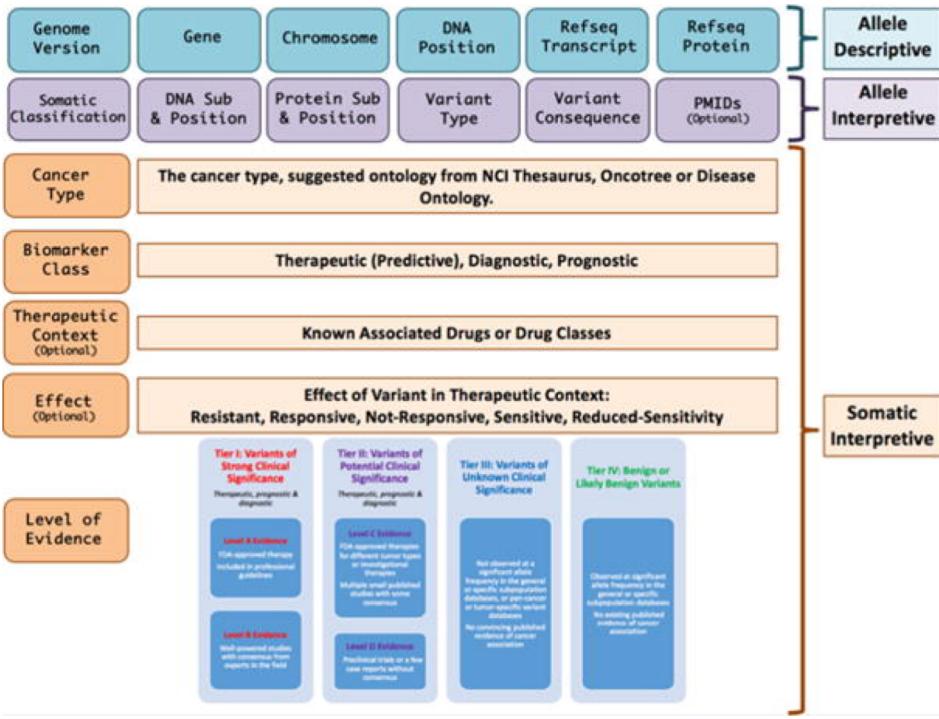
- Collaboration with the ClinGen Somatic Working Group is critical to achieving broad engagement by domain experts
  - Subha Madhavan and Shruti Rao
- Focused working groups forming (disease, gene, variant type, etc.)
- >40 members
- >250 evidence items submitted

The screenshot shows the CIViC website interface. At the top, there's a navigation bar with links for About, Participate, Community, Help, FAQ, and a user profile for 'MalachiGriffith'. Below the navigation is a search bar with placeholder 'Go! to Genes & Variants' and buttons for BROWSE, SEARCH, ACTIVITY, and ADD. The main content area is titled 'Organization Summary' and features a section for the 'Pediatric Cancer Task Force at ClinGen'. This section includes the ClinGen logo and a brief description: 'Pediatric Cancer Task Force within the ClinGen Somatic Work Group focuses on curation of clinically relevant somatic variants in pediatric cancers.' A 'Organization Website' button is also present. The 'Members' section displays eight profiles in a grid:

Member Profile	Name	Role	Expertise	Organization	ORCID ID
	Deb Irene	Curator	Research Scientist	Pediatric Cancer Task Force at ClinGen	-
	Shruti Rao	Curator	-	Pediatric Cancer Task Force at ClinGen	-
	Chimene Kesserwan	Curator	-	Pediatric Cancer Task Force at ClinGen	0000-0001-6043-2065
	Laura Corson	Curator	Research Scientist	Pediatric Cancer Task Force at ClinGen	-
	Gordana Racica	Curator	Clinical Scientist	Pediatric Cancer Task Force at ClinGen	-
	Angshumoy Roy	Curator	Clinical Scientist	Pediatric Cancer Task Force at ClinGen	0000-0001-7248-8576
	Kristen Lipscomb Sund	Curator	-	Pediatric Cancer Task Force at ClinGen	-
	Nishant Tiwari	Curator	Clinical Scientist	Pediatric Cancer Task Force at ClinGen	-



# ClinGen Somatic WG: MVLD Guidelines



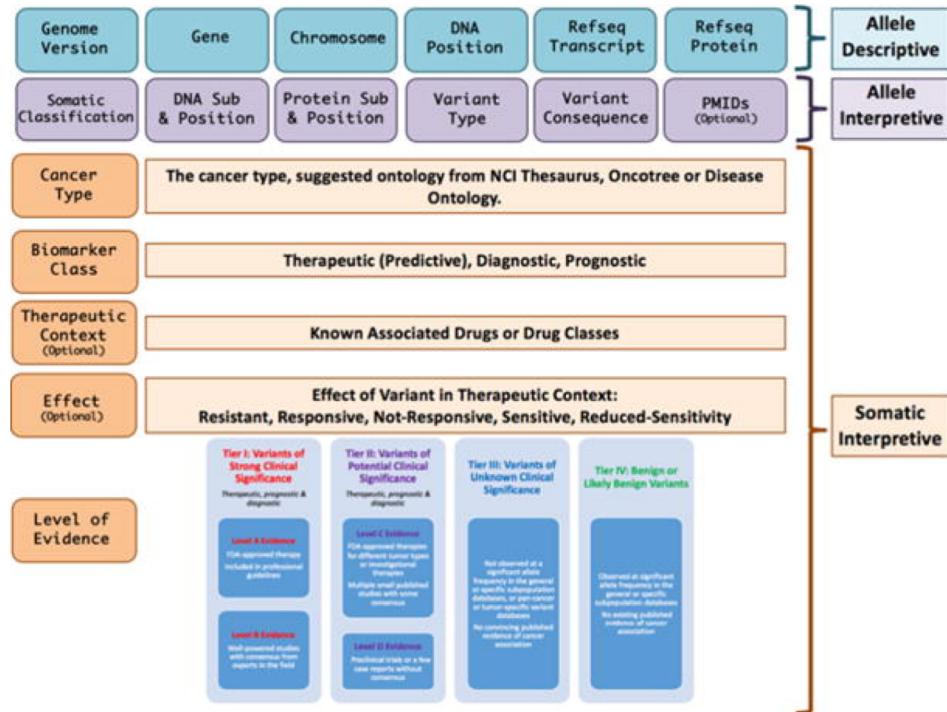
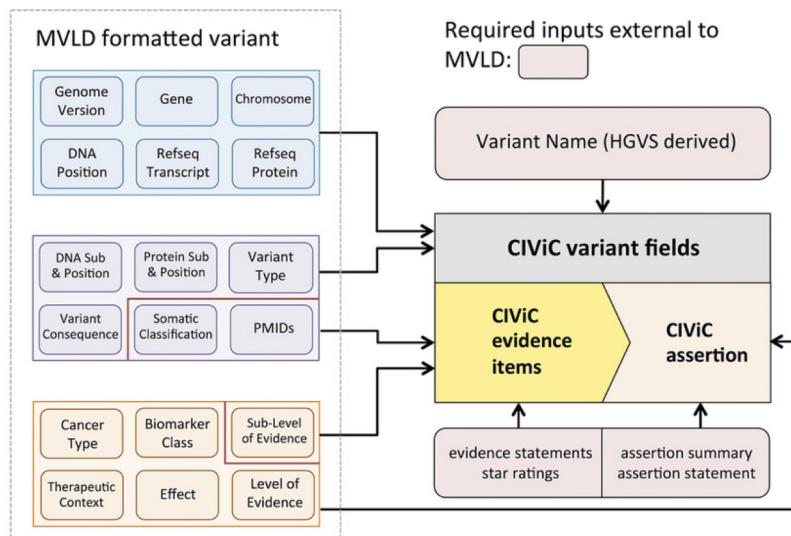
Madhavan S, et al. *Pac. Symp. Biocomput.* 2018

# ClinGen Somatic WG: MVLD Guidelines

Allele  
Descriptive

Allele  
Interpretive

Somatic  
Interpretive



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

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

# Creation of standards, SOPs, and training materials is needed

BMC Part of Springer Nature

Search

## Genome Medicine

Home About Articles Submission Guidelines

Correspondence | Open Access | Published: 29 November 2019

### Standard operating procedure for curation and clinical interpretation of variants in cancer

Arpad M. Danos, Kilannin Krysiak, Erica K. Barnell, Adam C. Coffman, Joshua F. McMichael, Susanna Kiwala, Nicholas C. Spies, Lana M. Sheta, Shahil P. Pema, Lynzey Kujan, Kaitlin A. Clark, Amber Z. Wollam, Shruti Rao, Deborah I. Ritter, Dmitriy Sonkin, Gordana Raca, Wan-Hsin Lin, Cameron J. Grisdale, Raymond H. Kim, Alex H. Wagner, Subha Madhavan, Malachi Griffith & Obi L. Griffith

[Genome Medicine](#) 11, Article number: 76 (2019) | [Cite this article](#)

380 Accesses | 4 Altmetric | [Metrics](#)

## Abstract

Manually curated variant knowledgebases and their associated knowledge models are serving an increasingly important role in distributing and interpreting variants in cancer. These knowledgebases vary in their level of public accessibility,

## Somatic Training Materials

Somatic Variant Training Materials Documents

Interested in Somatic Variant Curation? In order to get involved with our activities, please fill out our volunteer survey: <http://bit.ly/clingenvolunteersurvey>. For questions about existing materials or requests for new materials, contact us at [clingen@clinicalgenome.org](mailto:clingen@clinicalgenome.org).

Training Modules  Additional Supporting Materials

- 1 Recommended**  
The CIViC Knowledge Model and Standard Operating Procedures for Curation and Clinical Interpretation of Variants in Cancer  
→ The CIViC Knowledge Model and Standard Operating Procedures for Curation and Clinical Interpretation of Variants in Cancer
- 2 Recommended**  
CIViC - Getting Started  
This video covers: Description of CIViC and its goals, Navigating through CIViC's core pages, Browsing, searching, and consuming knowledgebase content.  
→ CIViC - Getting Started
- 3 Recommended**  
CIViC - Adding Evidence  
This video covers: Scanning a publication for curatable details, Signing into CIViC to Add Evidence, Walking through the Add Evidence form, Viewing the submitted evidence.  
→ Adding Evidence to CIViC
- 4 Recommended**  
CIViC - Editing Entities  
→ CIViC - Editing Entities

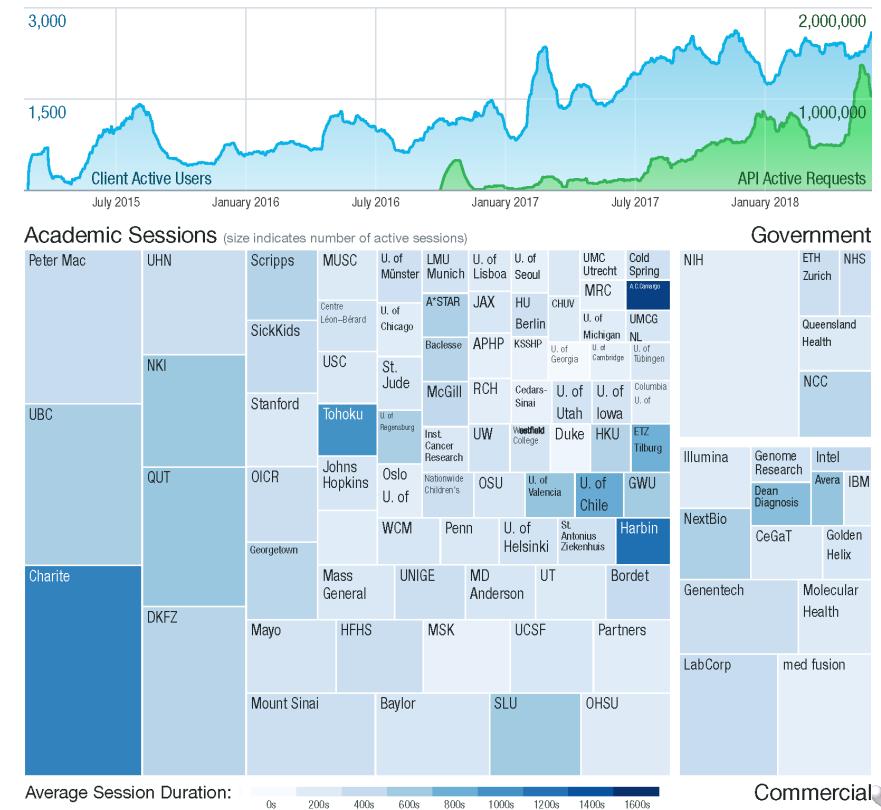
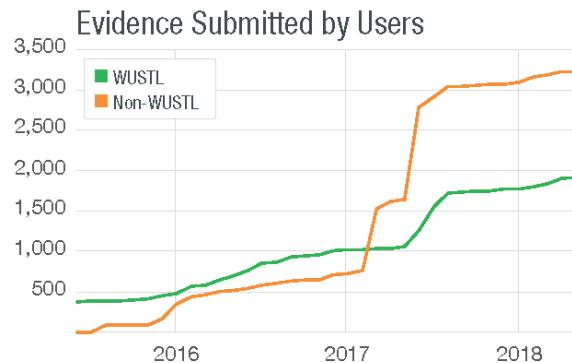
<https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-019-0687-x>

<https://clinicalgenome.org/curation-activities/somatic/training-materials/>

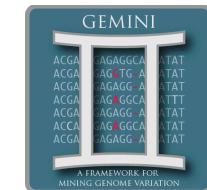


# How do we measure success?

- 6,468 evidence lines curated for 2,357 variants, 402 genes, 274 cancer types, 2,374 papers. 24 assertions to date.
- >3,000 users per month
- 191 contributors to date
- >750,000 API requests per month



# Open model promotes adoption by commercial and academic data clients

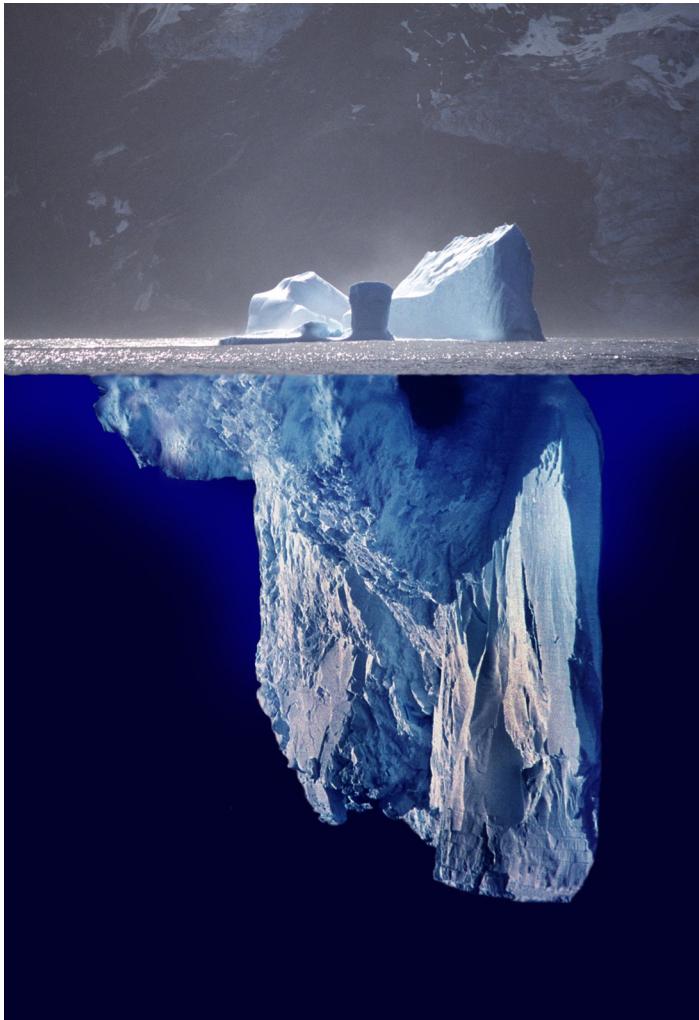


Alissa Clinical Informatics Platform  
Alissa Interpret



CIViC has >35 known data clients (<https://civicdb.org/about#data-clients>)





*civic2clinvar*



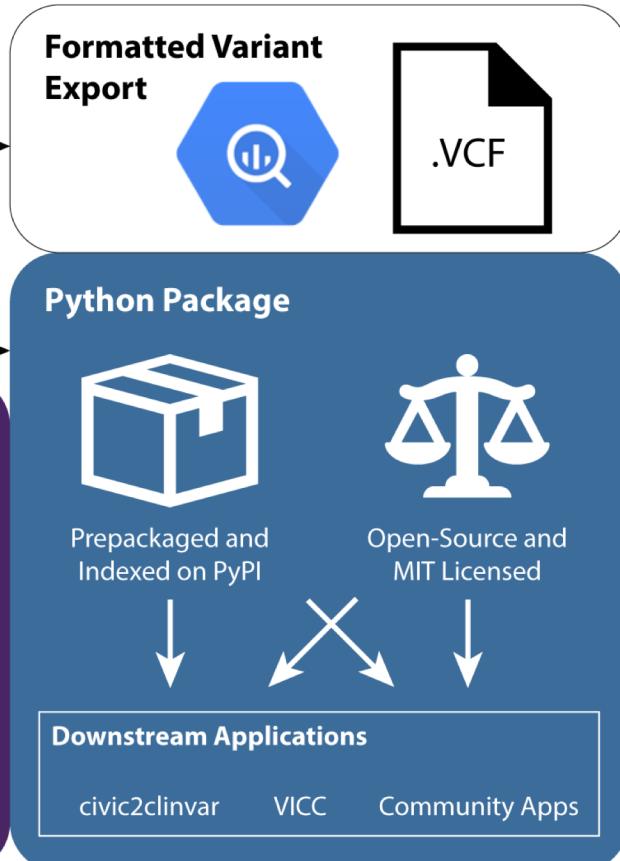
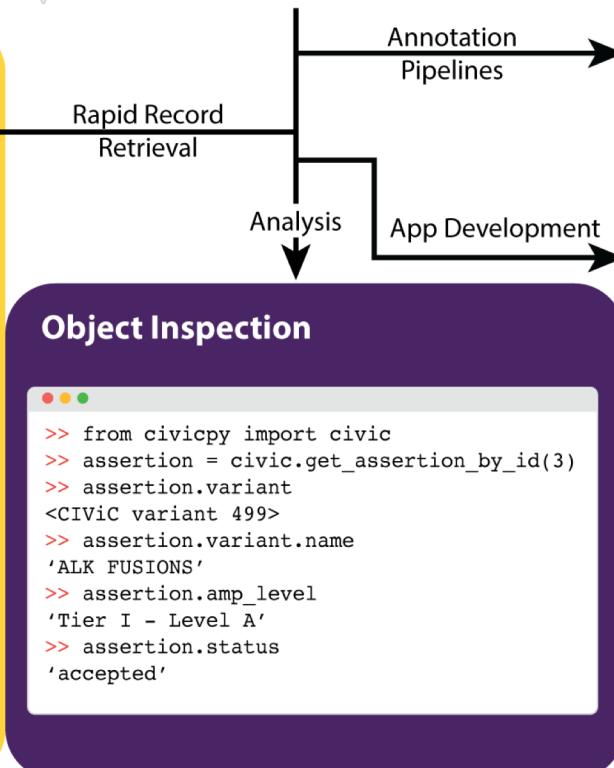
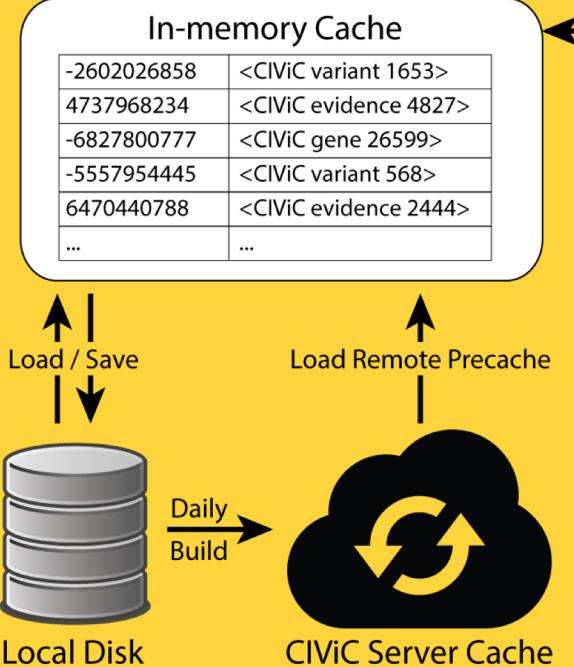
**CIViCpy**  
The Python SDK for CIViC



@handlerwagner



## Caching





### 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

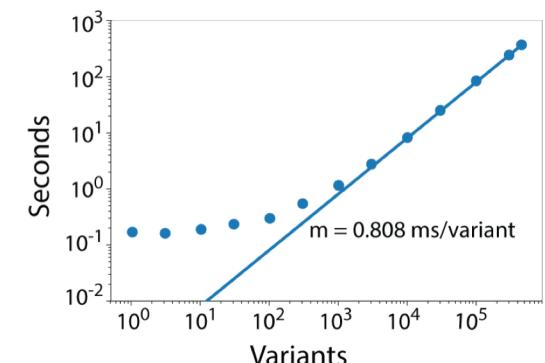
Record	Query	Any	QE	RE	Exact
Record	Query	✓	✓	✓	✗

###### Allele Match

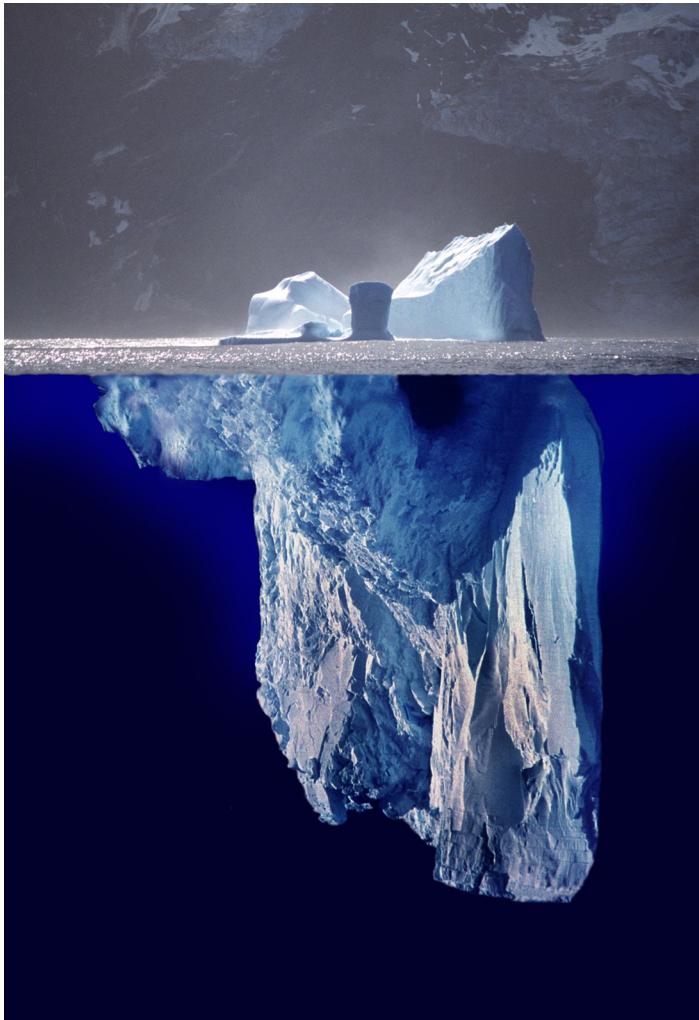
Record	Query	Any	QE	RE	Exact
Record	Query	✓	✓	✓	✓

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

### ClViCpy Search Performance



@handlerwagner



# CIViCpy

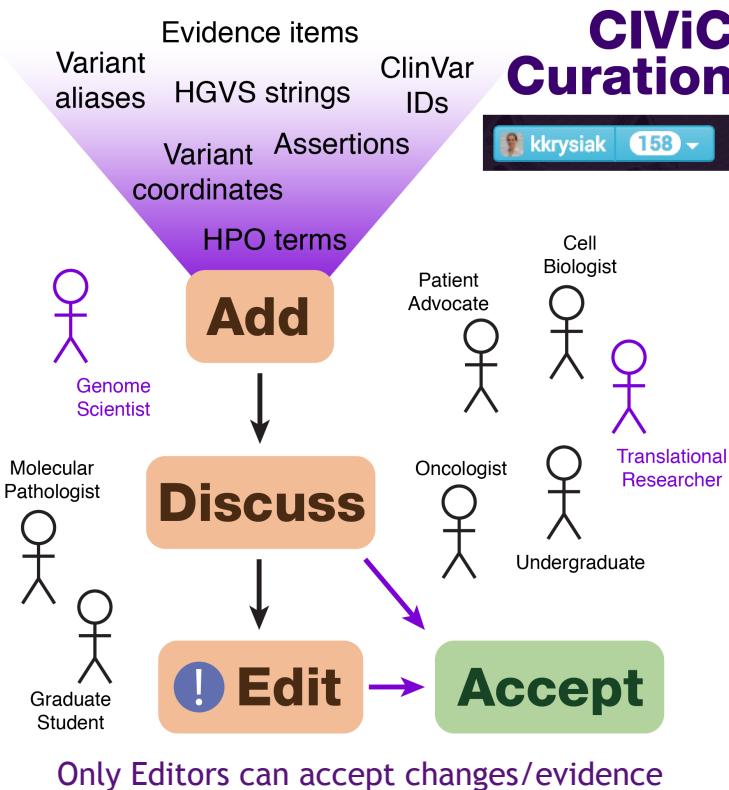
The Python SDK for CIViC

[civicpy.org](https://civicpy.org)



@handlerwagner

# CIViC's open access and crowd-sourcing model has advantages and disadvantages



- Advantages
- Scalable
- Promotes transparent consensus building (with provenance)
- Diverse community (remote, asynchronous, global contributions)
- Disadvantages
- Quality (perceived and real)
- Difficult to focus curation effort
- Funding



# Adoption of existing standards

HUGO  
(genes)

Disease Ontology  
(cancer types)

HPO  
(phenotypes)

Assertion AID5

Submitted by arpaddano Last Modified by ebamell Last Reviewed by arpaddano Accepted by ebamell

Gene: EGFR Variant: L858R Variant Allele Registry ID: CA126713

Disease: Non-small Cell Lung Carcinoma

Associated Phenotype: -

Summary: Non-small cell lung cancer with EGFR L858R mutation is sensitive to erlotinib or gefitinib.

Description: L858R is among the most common sensitizing EGFR mutations in NSCLC, and is assessed via DNA mutational analysis including Sanger sequencing and next generation sequencing methods. Tyrosine kinase inhibitors erlotinib and gefitinib are associated with improved progression free survival over chemotherapy in EGFR L858R patients. NCCN guidelines recommend (category 1) erlotinib and gefitinib for NSCLC with sensitizing EGFR mutations, along with afatinib and osimertinib.

ClinVar ID 16609 ClinVar Clinical Significance Pathogenic MyVariant.info

COSMIC ID COSM6224 dbSNP RSID rs121434568 HGVS ID chr7:g.55259515T>G

SnpEff Effect structural interaction variant SnpEff Impact HIGH gnomAD Adj. AF --

Assertion Type: Predictive Assertion Direction: Supports Clinical Significance: Sensitivity Drugs: Erlotinib and Gefitinib Drug Interaction Type: Substitutes AMP Category: Tier I - Level A NCCN Guideline: Non-Small Cell Lung Cancer (v3.2018) Regulatory Approval: FDA Companion Test:

ClinGen Allele Registry (variants)

NCIT (drugs)

AMP or ACMG tiering (clinical significance)

Evidence Grid Evidence Cards

Evidence Supporting AID5 14 total items

EID	GENE	VARIANT	DESC	DIS	DRUGS	EL	ET	ED	CS	VO	TR	...
2994	EGFR	L858R	On May 14, 2013, the...	Non-small Cell Lung ...	Erlotinib	A	🕒	ⓘ	❤️	...	5	★
2621	EGFR	L858R	In a phase 3 clinical t...	Non-small Cell Lung ...	Gefitinib	B	🕒	ⓘ	❤️	...	4	★
1665	EGFR	L858R	90 NSCLC patients w...	Non-small Cell Lung ...	Gefitinib	B	🕒	ⓘ	❤️	...	3	★
2634	EGFR	L858R	In a phase 3 clinical t...	Lung Adenocarcinoma	Gefitinib	B	🕒	ⓘ	❤️	...	3	★
229	EGFR	L858R	There is no statistica...	Non-small Cell Lung ...	Erlotinib, Gefitinib (S...	B	🕒	ⓘ	❤️	...	3	★
885	EGFR	L858R	A randomized phase ...	Non-small Cell Lung ...	Erlotinib	B	🕒	ⓘ	❤️	...	3	★
2624	EGFR	L858R	Mutational profiling ...	Lung Adenocarcinoma	Gefitinib	C	🕒	ⓘ	❤️	...	1	★





## Problem: CIViC is a Silo in a Diverse Knowledge Ecosystem



# Variant interpretation knowledge remains siloed

Established Interpretation Knowledgebases (somatic)

- [CIViC \(WashU\)](#)
- [OncoKB \(MSKCC\)](#)
- [JAX-Clinical Knowledgebase \(Jackson lab\)](#)
- [Cancer Genome Interpreter \(Barcelona\)](#)
- [MyCancerGenome \(Vanderbilt\)](#)
- [PMKB \(Cornell\)](#)
- [KnowledgeBase for Precision Oncology \(MD Anderson\)](#)
- [CanDL \(Ohio State\)](#)
- [COSMIC \(Sanger\)](#)

Germline

- [PharmGKB](#)
- [ClinVar](#)
- [ClinGen Evidence Repository](#)

Additionally...

- Many ad hoc “databases” (academic centers / hospitals)
- Many industry examples

# Collaborative standards and interoperability work is needed: Variant Interpretation for Cancer Consortium (VICC) - GA4GH Driver Project



...

[cancervariants.org](http://cancervariants.org)



**Global Alliance**  
for Genomics & Health

Variant Interpretation for Cancer

- Gene
- Variant
- Cancer subtype
- Clinical implication: drug sensitivity, drug resistance, adverse response, diagnostic, or prognostic
- Source (e.g., PubMed identifier)
- Curation group

[ga4gh.org](http://ga4gh.org)

## Goals/Principles:

- Global integration of clinical cancer variant interpretation
- Standards and guidelines
- Open content for sharing
- Facilitate cross-knowledgebase queries

[search.cancervariants.org](http://search.cancervariants.org)

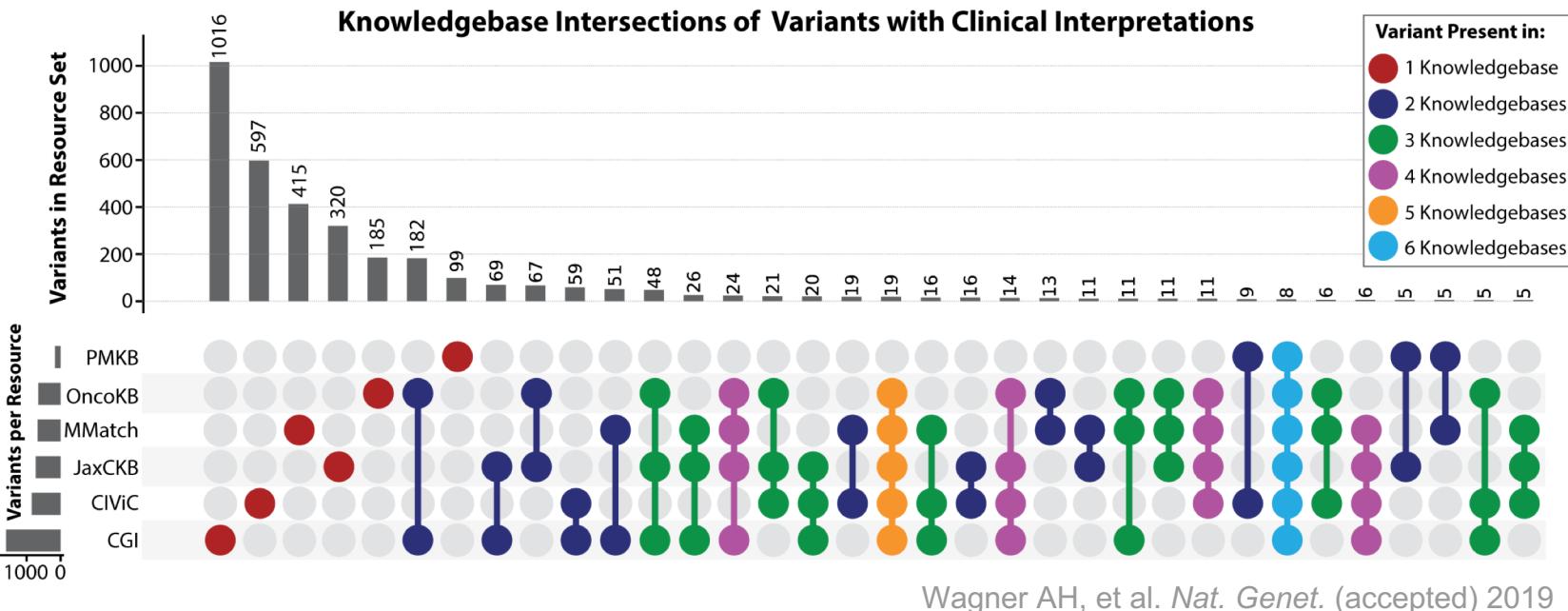




# Variant Interpretations are **Heterogeneous**



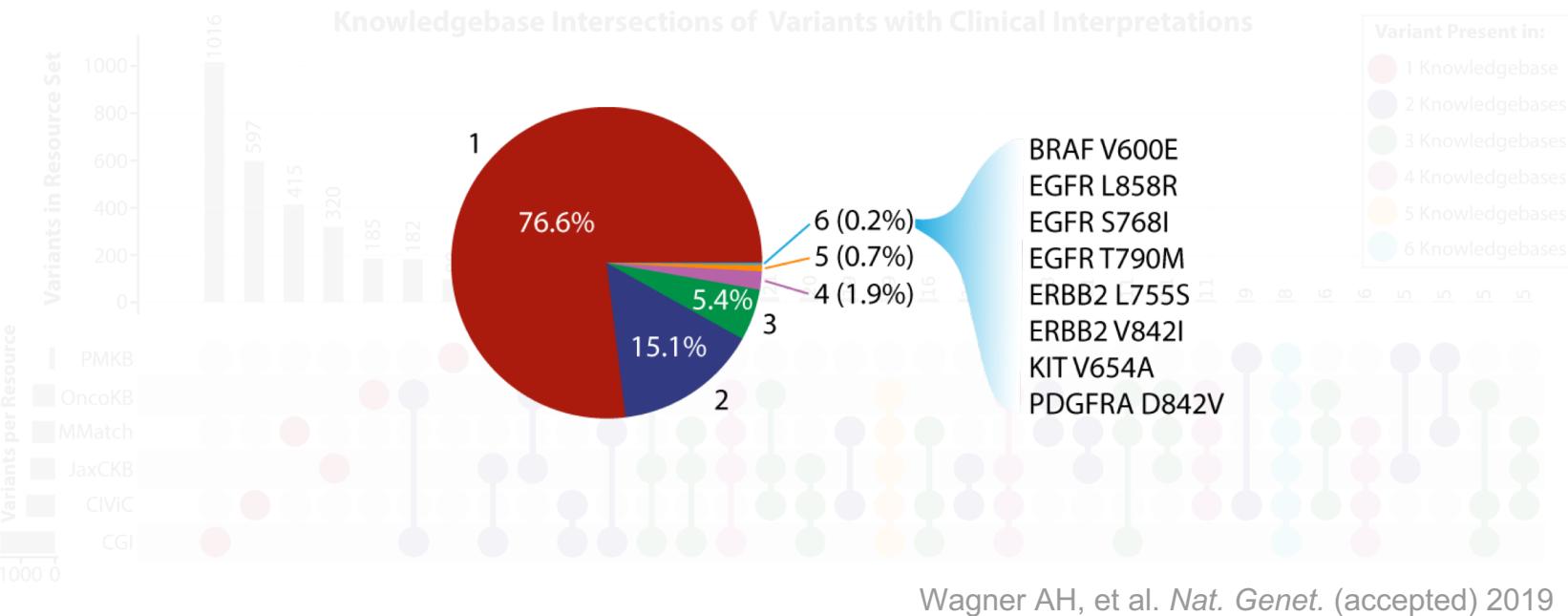
# Variant Representation Across Knowledgebases



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...

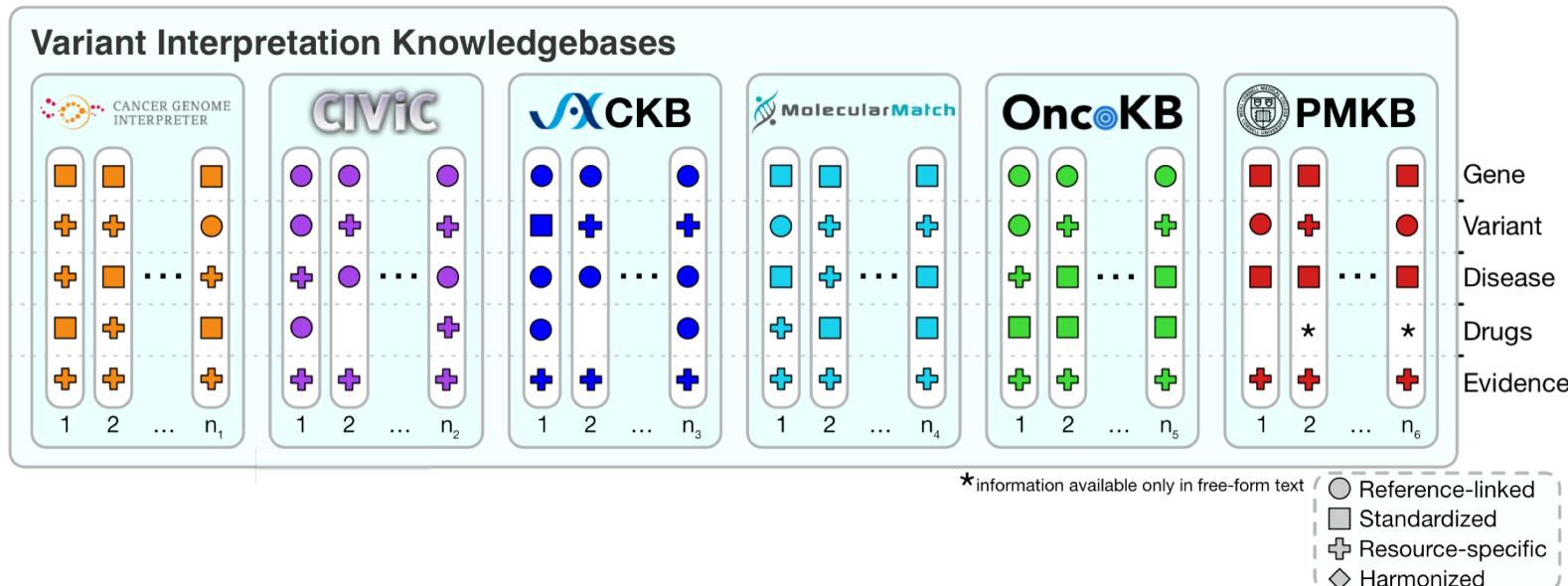
Clinical variant interpretation  
knowledgebases are not interoperable

This can be addressed by harmonizing interpretations





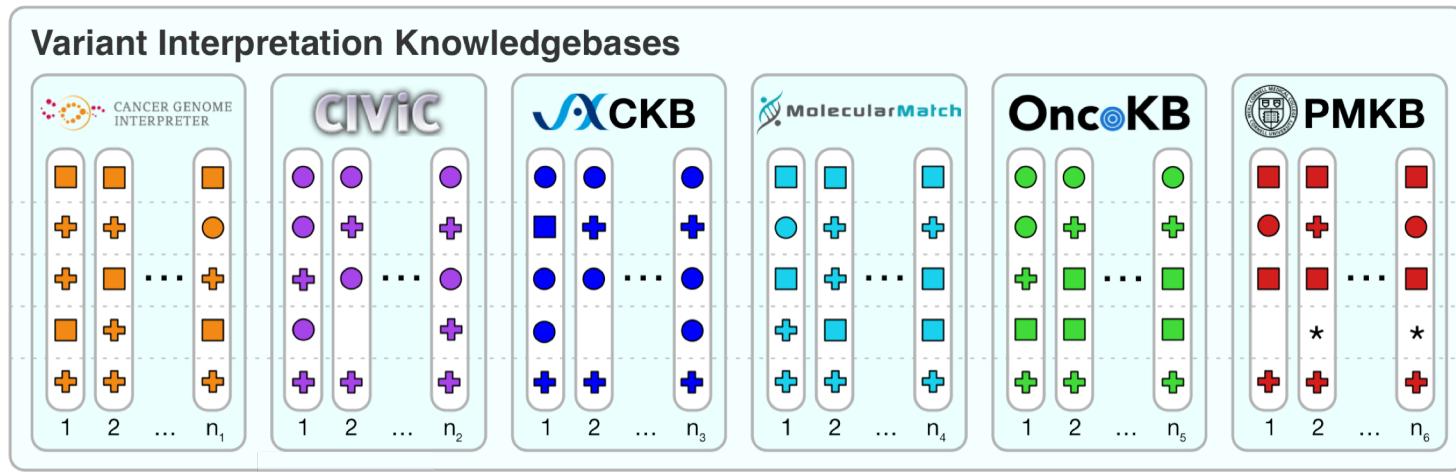
# Structure of a Variant Interpretation



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	



## 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'
},
```

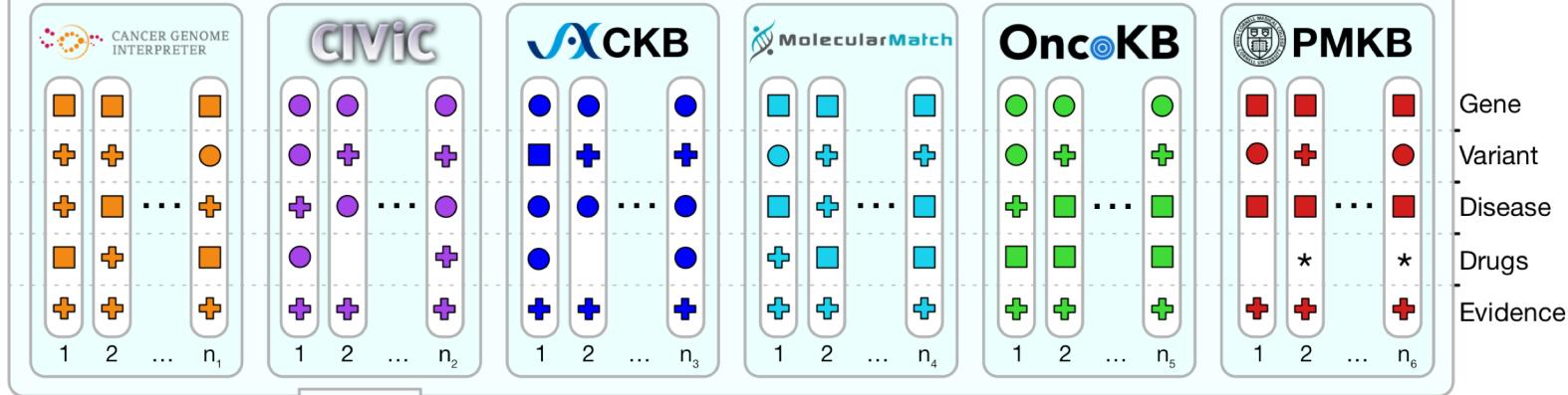
```
'drugs': [
```

```
  {'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

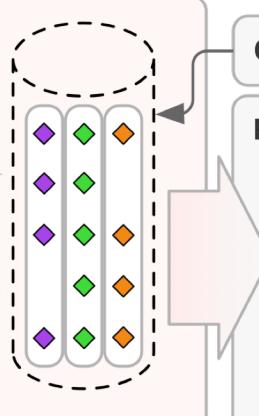
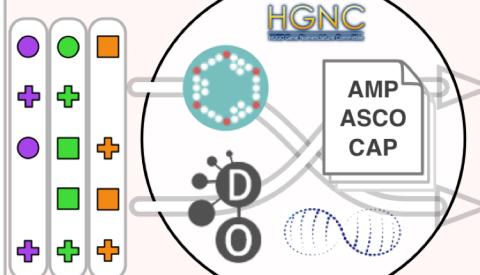
# Variant Interpretation Knowledgebases



\*information available only in free-form text

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

## Harmonization



Query: BRAF V600E

## Results

Level A 4 Interpretations

Skin Melanoma, Predicts Sensitivity, Vemurafenib  
Skin Melanoma, Predicts Sensitivity, Trametinib + Dabrafenib (Combination)  
Colorectal Cancer, Poor Prognosis  
NSCLC, Predicts Sensitivity, Dabrafenib + Trametinib (Combination)

Level B 13 Interpretations

Level C 8 Interpretations

Level D 20 Interpretations



## Part II: An **Expressive** and **Computable** specification for Variation Representation (VR-spec / VR)



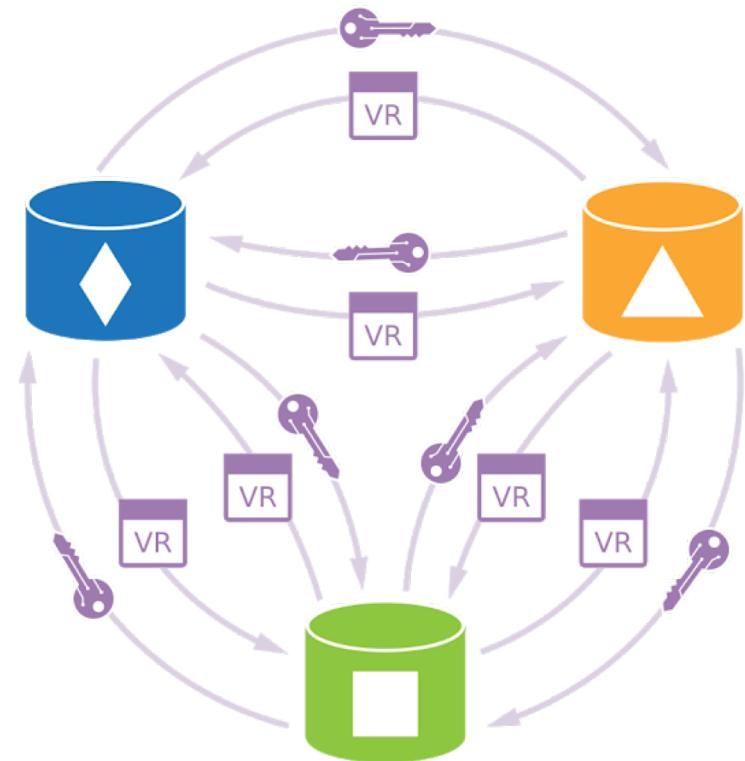
Reece Hart

# Overview

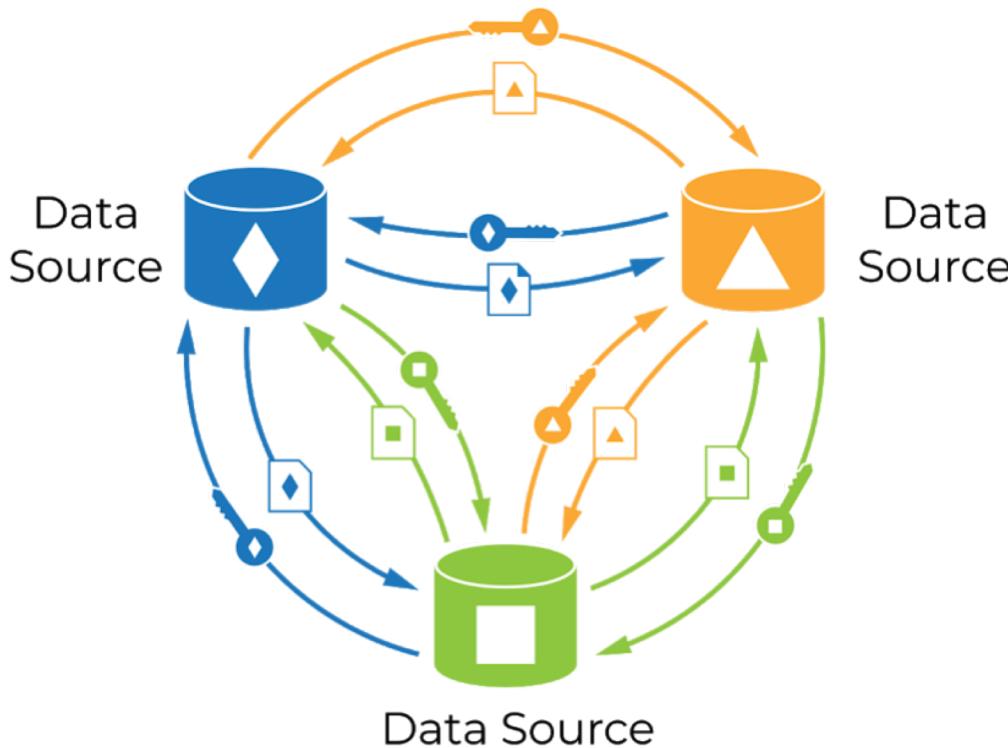
VR aims to standardize the definition, sharing, and identification of biological variation within and between systems.

## VR 1.0 Aims:

- An extensible schema for variation
- Minimization of representational ambiguity
- Shared identifiers for variation
- Example Implementation

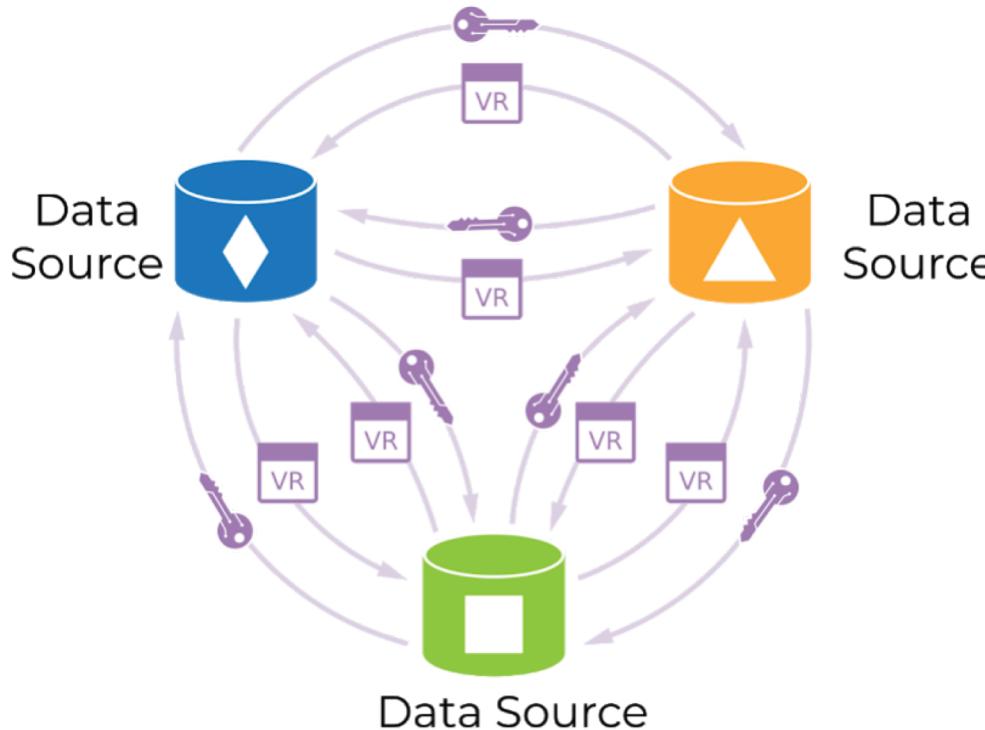


# CURRENTLY...



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.

# Presentation ≠ Representation

NM\_080588.2:c.139\_140insC  
ENST00000367279:c.139\_140insC  
NM\_080588.2:c.139dup  
ENST00000367279:c.139dup



```
{  
  "sequence_id": "ga4gh:SQ.0a1b2c3d",  
  "location": (139,139),  
  "state": "C"  
}
```

*Multiple human names*  
due to choice of accession, normalization,  
ins/dup, etc.



Structured representations  
Clear conventions  
= VR

*One conceptual variant*

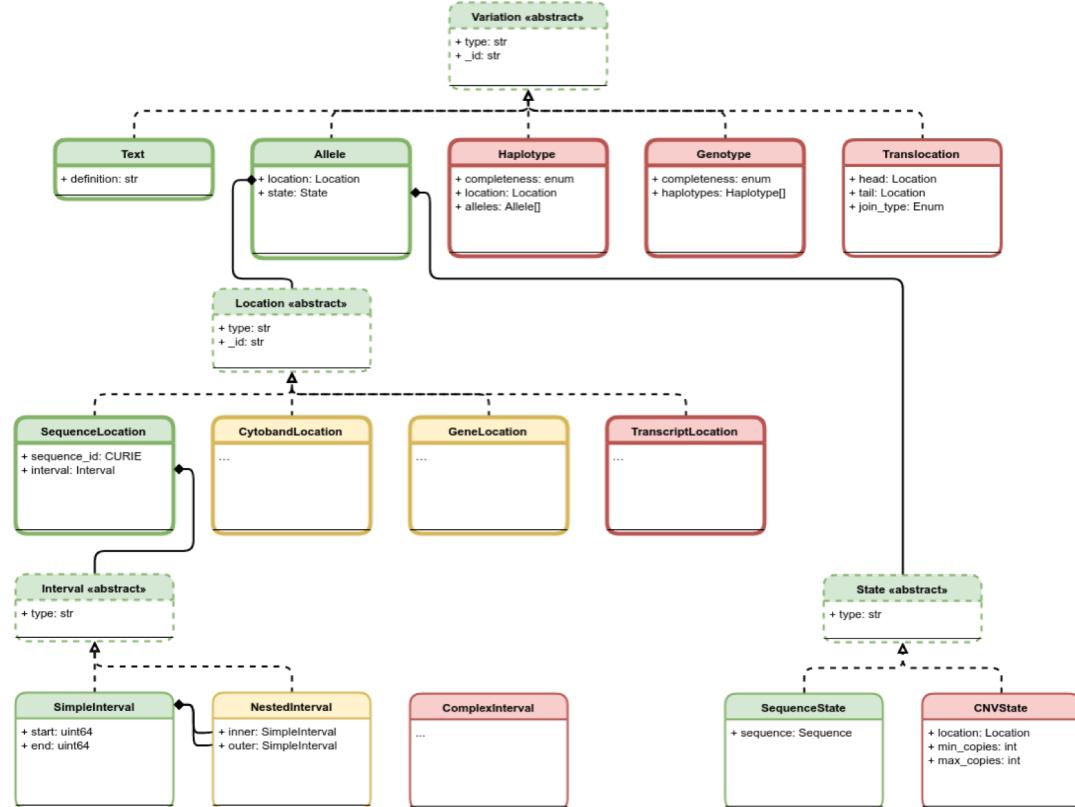
# VR Schema is Extensible

1.0 Released

Near-term

Planning

- Alternative location types
  - Genes
  - Cytobands
- Variation Sets
- Structural Variations
  - CNVs
  - Translocations
  - Fusions
- Haplotypes
- Genotypes



# Example: NC\_000013.11:g.32936732G>C

Boxes indicate schema classes and composition.

Computed identifiers are shown, but spec permits other formats.

The goal is a unambiguous computational representation (not human readability).

```
"ga4gh:VA.n9ax-9x6g0C00Et73VMYqCBfqfxG1XUH": {
    "location": {
        "interval": {
            "end": 32936732,
            "start": 32936731,
            "type": "SimpleInterval"
        },
        "sequence_id": "ga4gh:SQ._0wi-qoDrvram155UmSC-zA5ZK4fpLT",
        "type": "SequenceLocation"
    },
    "state": {
        "sequence": "C",
        "type": "SequenceState"
    },
    "type": "Allele"
}
```

<https://vr-spec.readthedocs.io/en/1.0rc/impl-guide/example.html>

# Fully-Justified Normalization Captures Region of Shuffling Ambiguity

*Example: Variant replaces CA with CAGCA in TCAGCAGCT at bases 5-6  
actual location of variation is ambiguous due to the sequence context*

$TCAG \left[ \frac{CA}{CAGCA} \right] GCT$

left shuffle ↲  
(à la VCF)

$T \left[ \overline{CA} \right] CAGCAGCT$



*regions of location ambiguity*

↘ right shuffle  
(à la HGVS)

$TCAGCAGC \left[ \overline{AGC} \right] T$

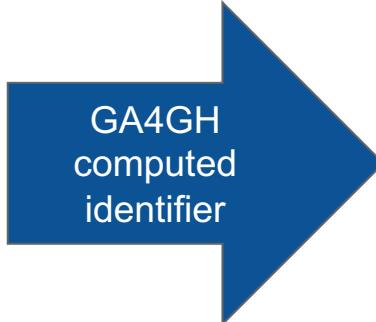


The mission of the VR is to  
*standardize the representation of all classes  
of biological variation*  
to enable  
*accurate curation and  
reliable computed interpretation.*



# Computed identifiers are unique and precise

```
{  
  "location": {  
    "interval": {  
      "end": 32936732,  
      "start": 32936731,  
      "type": "SimpleInterval"  
    },  
    "sequence_id": "refseq:NC_000013.11",  
    "type": "SequenceLocation"  
  },  
  "state": {  
    "sequence": "C",  
    "type": "SequenceState"  
  },  
  "type": "Allele"  
}
```



GA4GH  
computed  
identifier

ga4gh:VA.EgHPXXhU...

# Computed Identifier Algorithm



Same data ⇒ same identifier



- Obviates central registration services
- Simplifies id generation in private environments
- Facilitates data sharing by distributed groups
- Works for any serializable type

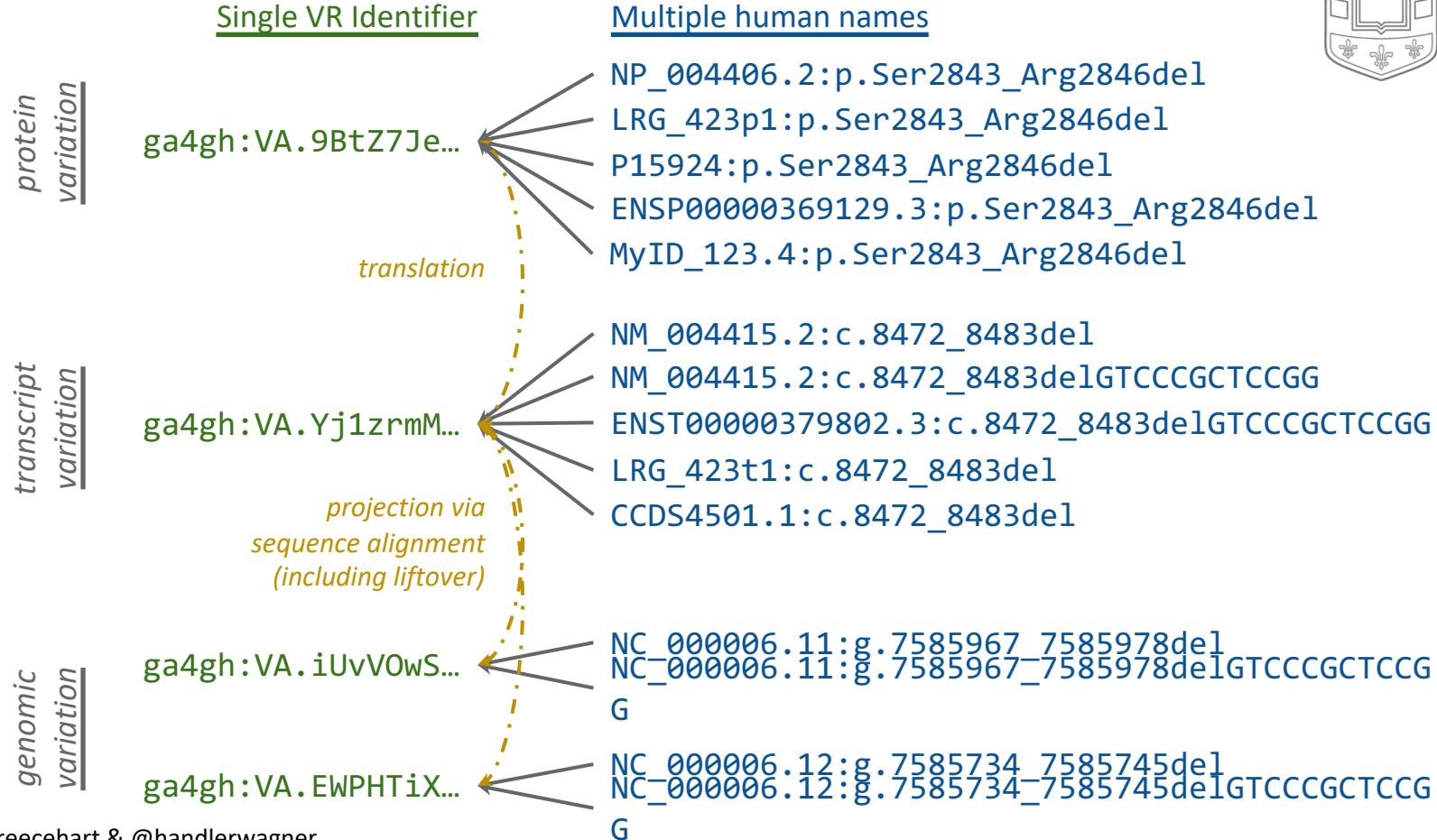
Identifier format:

"ga4gh:" *type* ":" *digest*

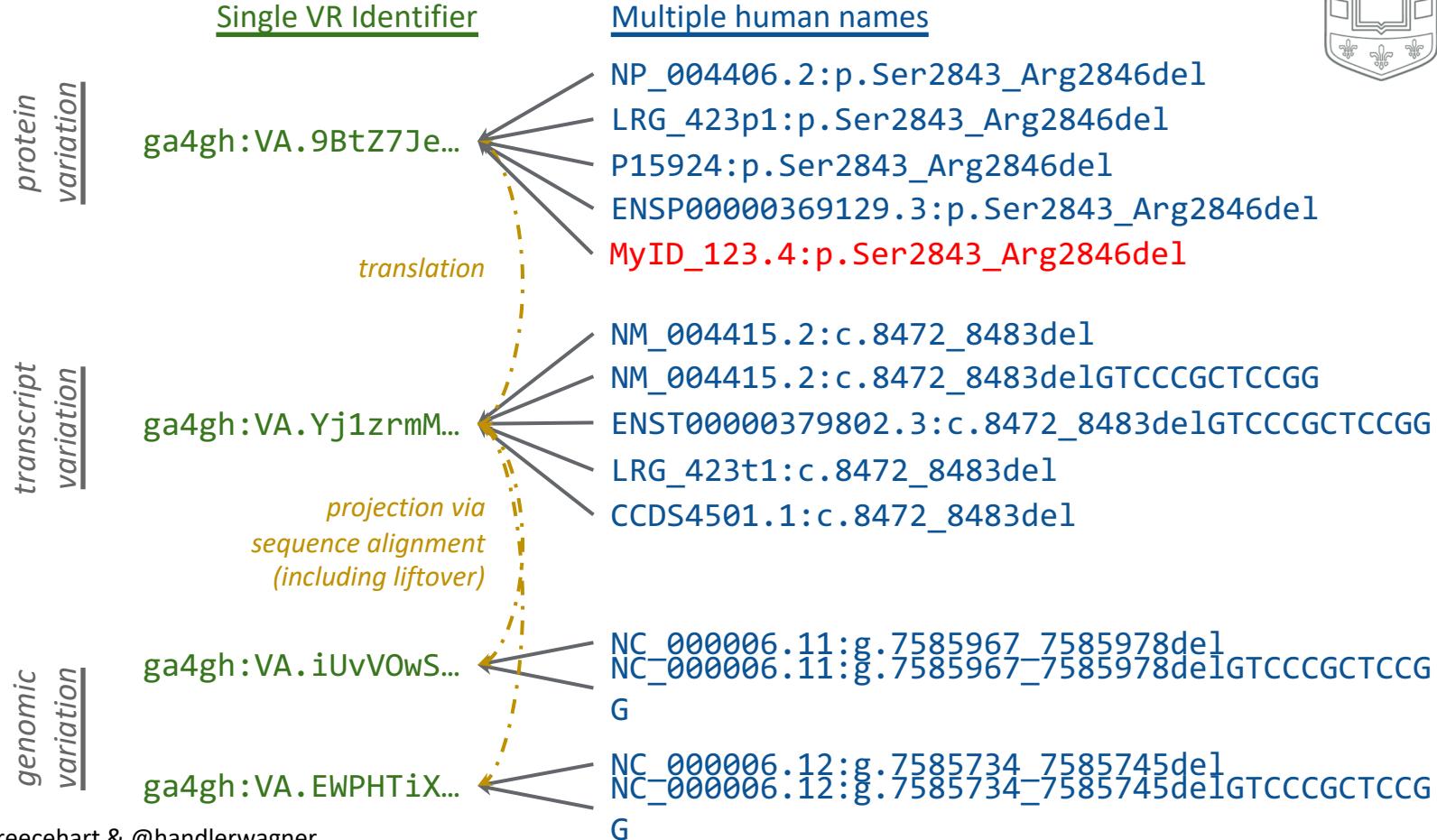
ga4gh:**V**A.zA5ZK4fpLT...

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

# VR identifiers: precise, unique names for variation



# VR identifiers: precise, unique names for variation



# VR In a Nutshell

- VR provides ...
  - a *terminology document*
  - a *logical model*
  - a concrete (JSON Schema) specification
  - a digest-based method for generating globally unique identifiers
- for ...
  - Sequences, Locations, Alleles, Haplotypes, and Genotypes
    - NGS "readthrough" variation now
    - CNVs, SVs, fuzzy locations later
- in order to ...
  - Create an extensible *computational* representation of variation on all sequence types



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>