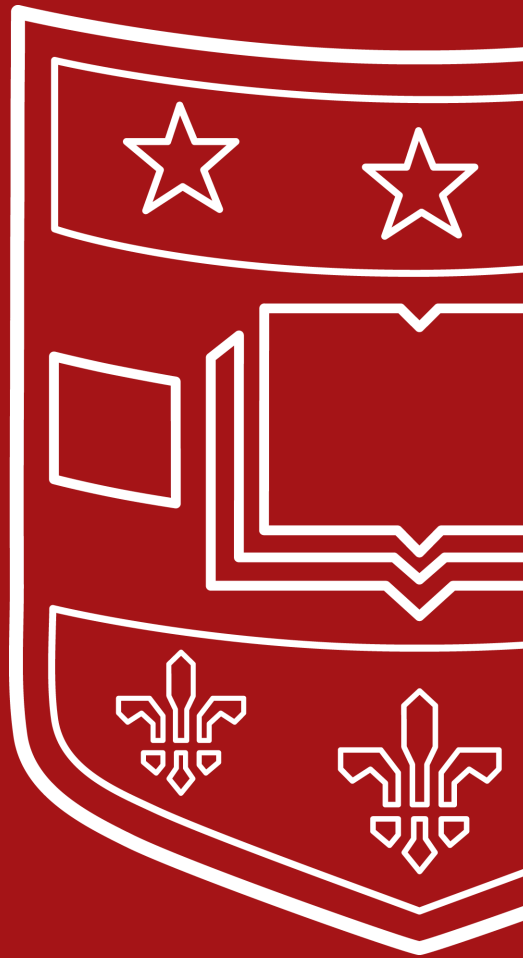
Knowledge standards drive clinical interpretation of genomic variants in cancers

Alex H Wagner, PhD
December, 2019 - UZH Irchel



Washington University in St. Louis

Introduction

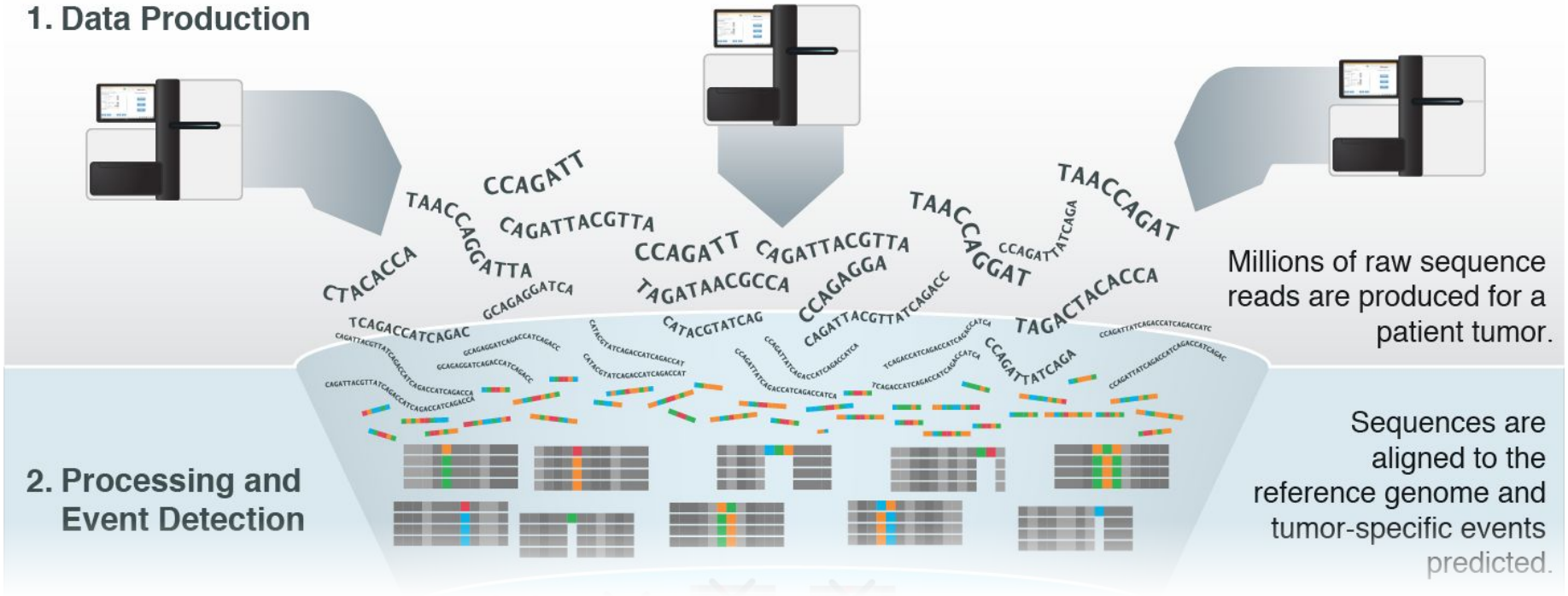
Precision Oncology

Tumor DNA

Clinical Genomic Variant Report

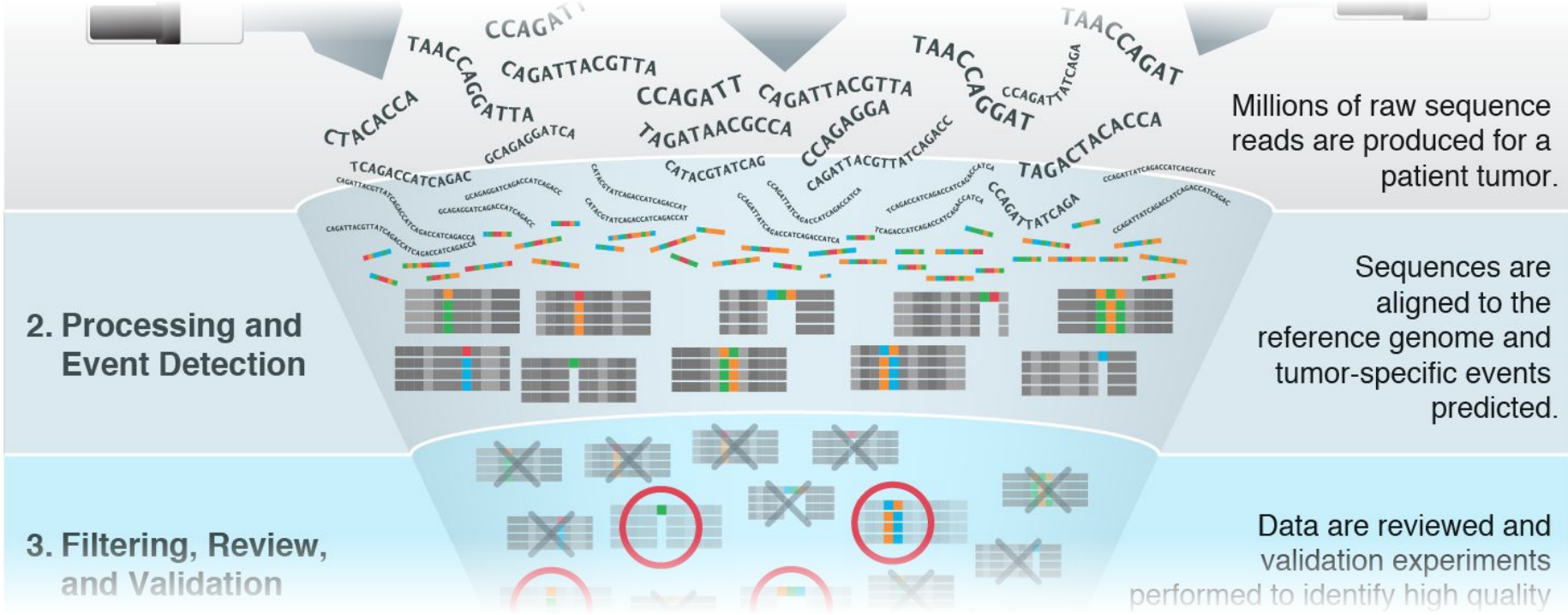
Patient Care

Precision Oncology Bottleneck



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

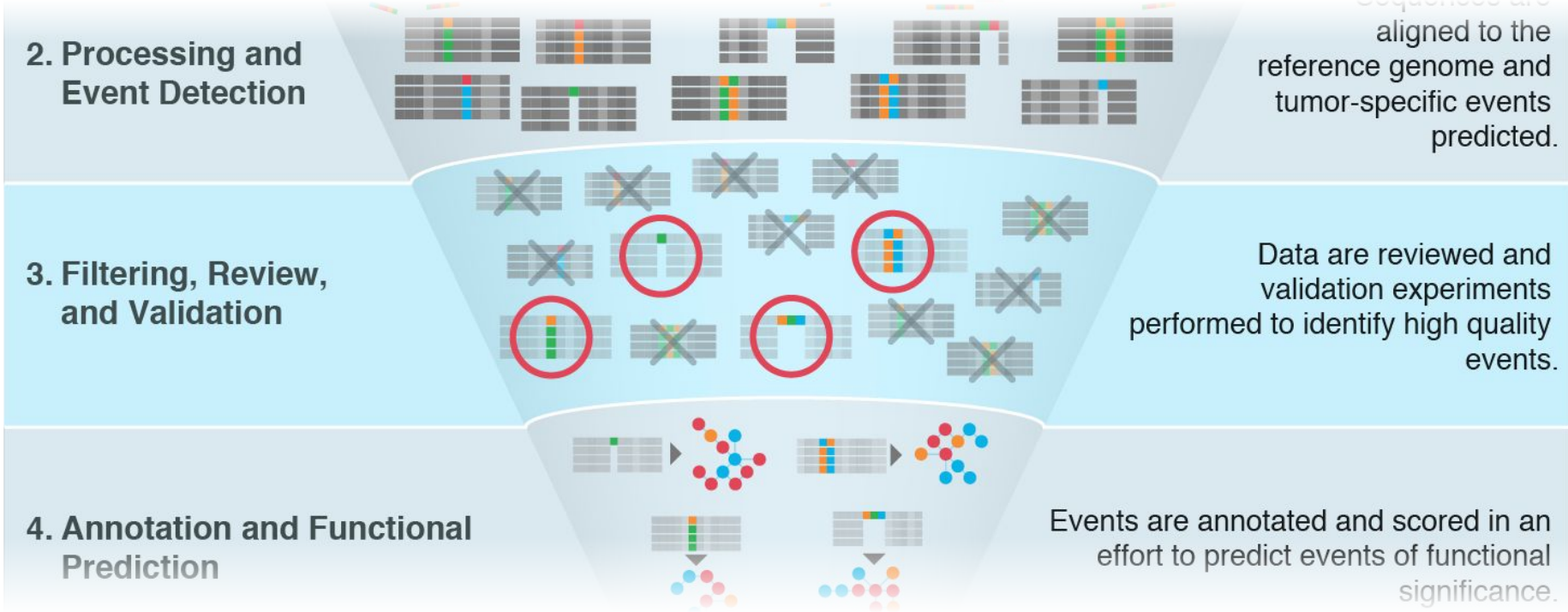
Precision Oncology Bottleneck



@handlerwagner

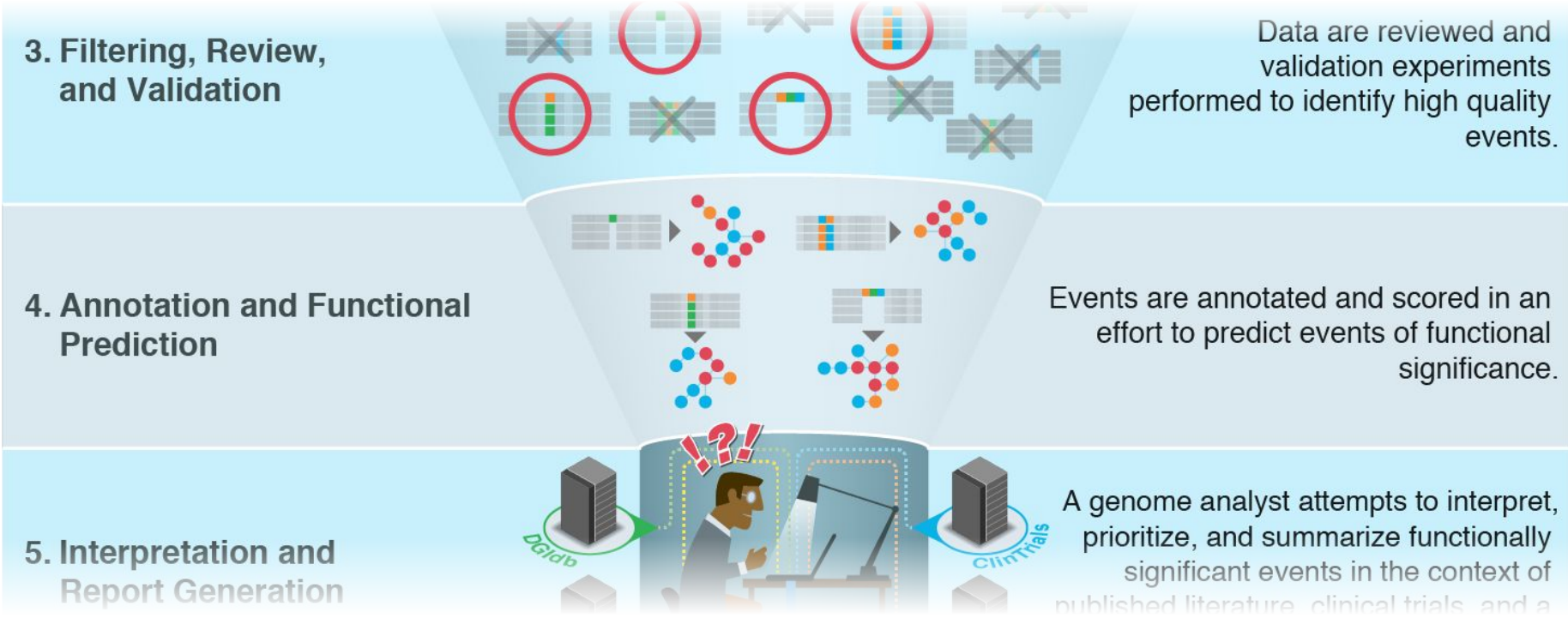
Good BM, Ainscough BJ, McMichael JF, Su AI†, Griffith OL†. 2014. Genome Biology. 15(8):438.

Precision Oncology Bottleneck



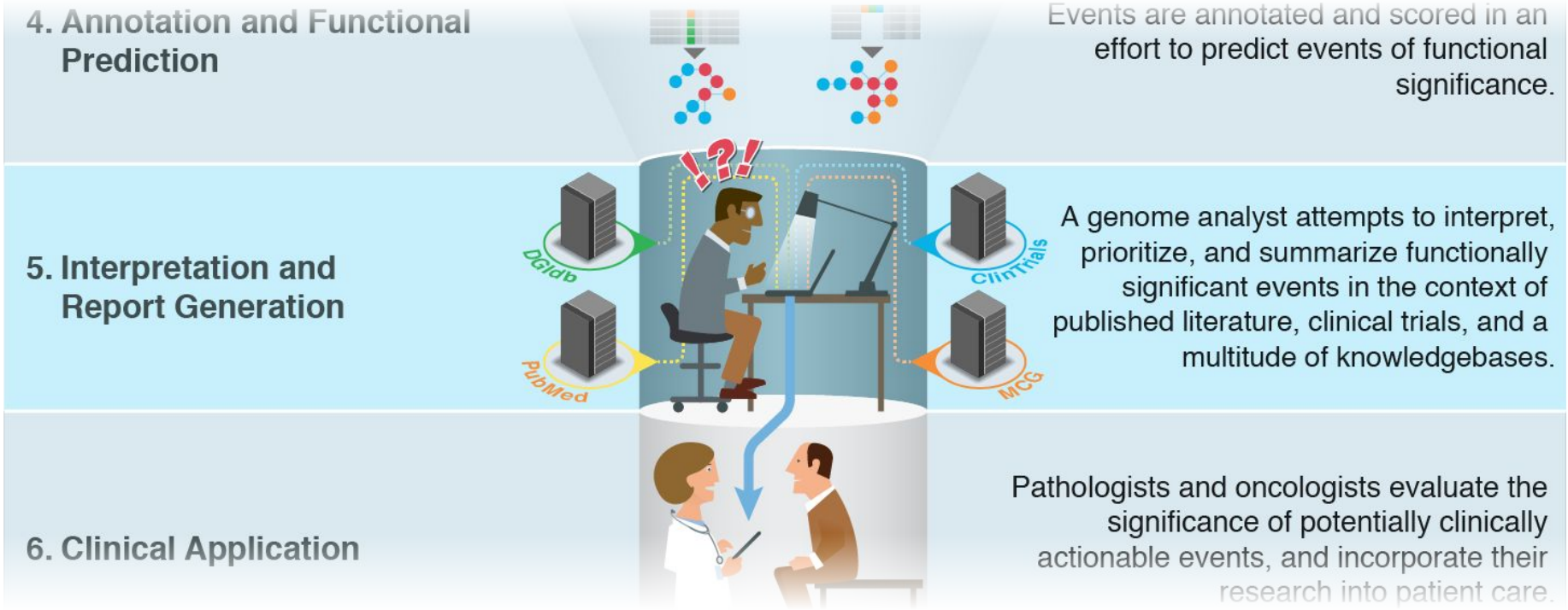
Good BM, Ainscough BJ, McMichael JF, Su AI†, Griffith OL†. 2014. Genome Biology. 15(8):438.

Precision Oncology Bottleneck



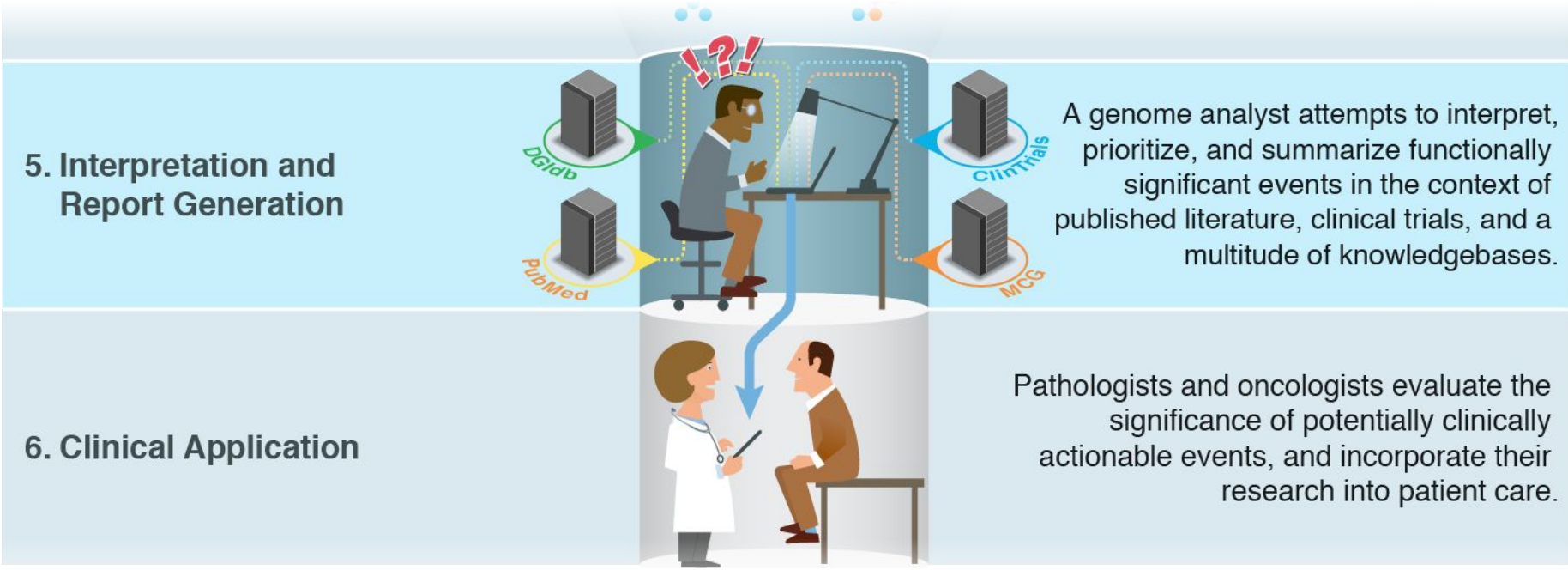
Good BM, Ainscough BJ, McMichael JF, Su AI†, Griffith OL†. 2014. Genome Biology. 15(8):438.

Precision Oncology Bottleneck



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

Precision Oncology Bottleneck



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

Precision Oncology Bottleneck

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

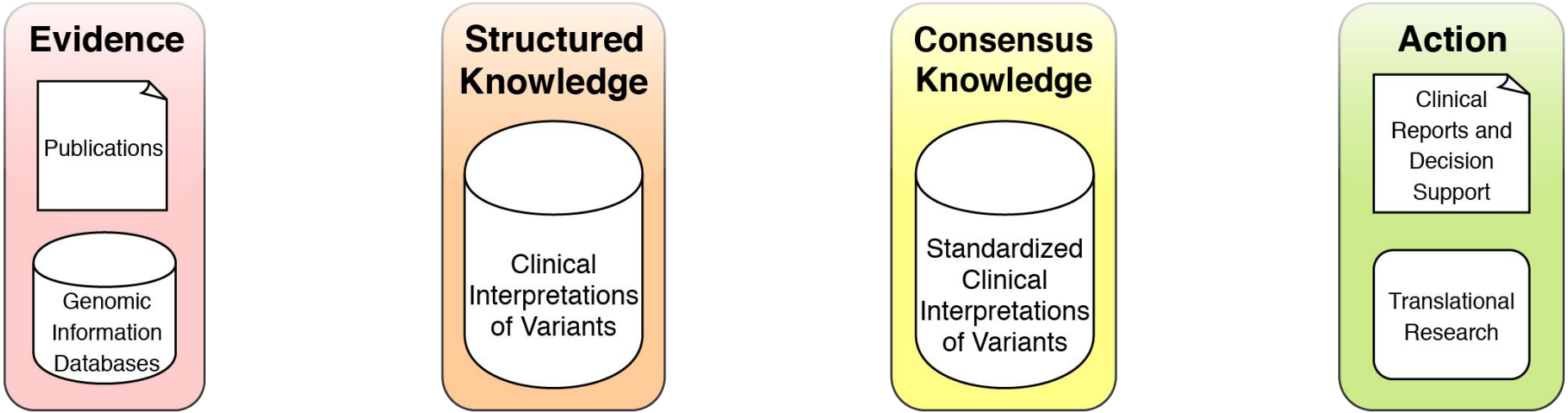
Precision Oncology Bottleneck

How do we alleviate this bottleneck?

@handlerwagner

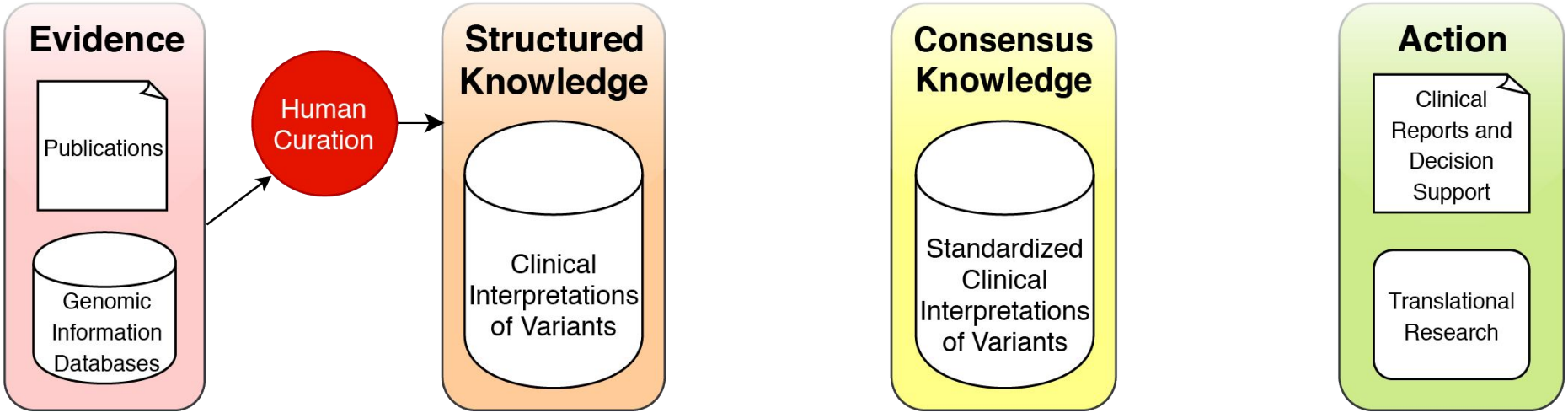
Good BM, Ainscough BJ, McMichael JF, Su AI†, Griffith OL†. 2014. Genome Biology. 15(8):438.

Roadmap from Evidence to Action



@handlerwagner

Roadmap from Evidence to Action



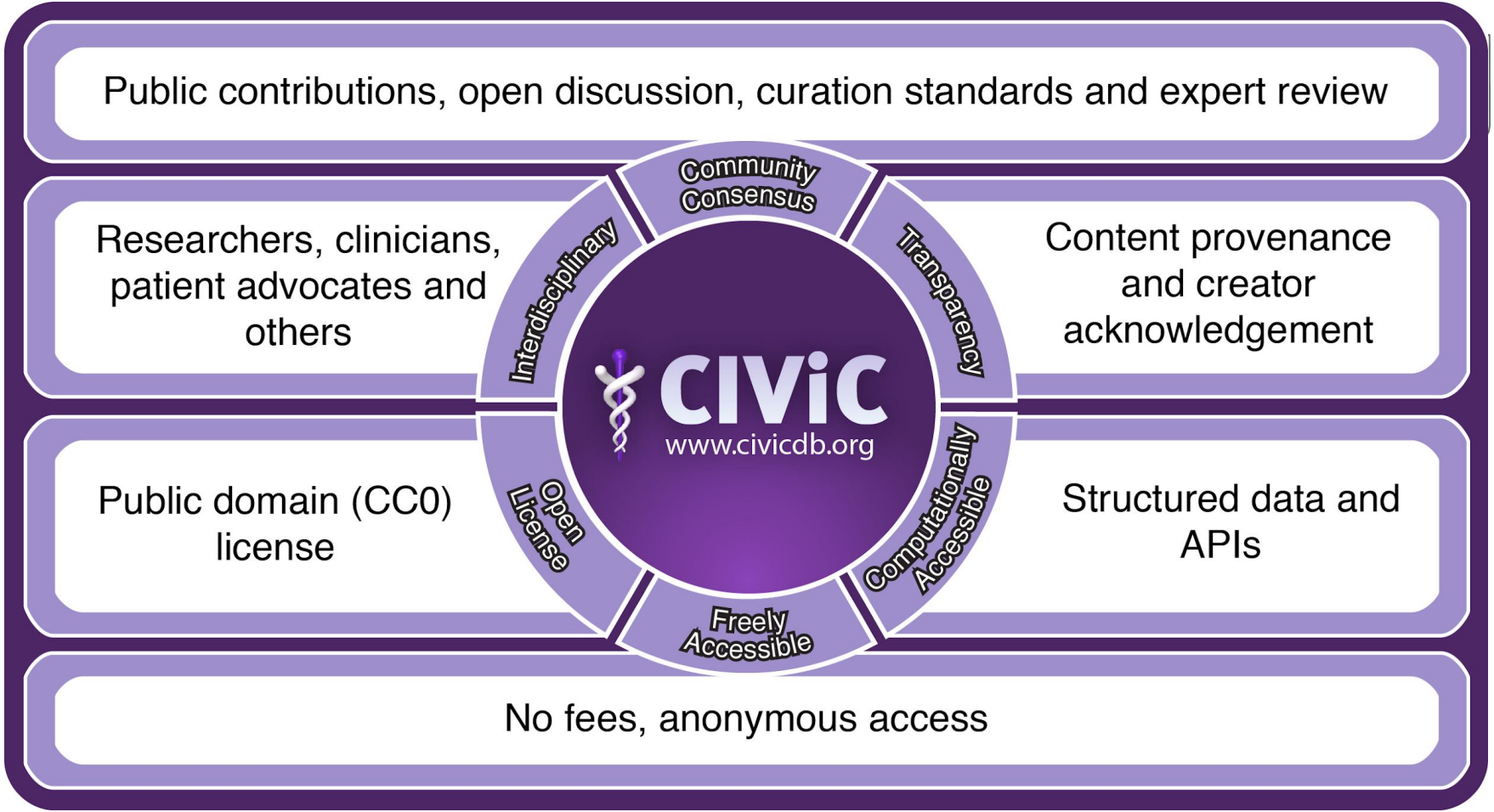
Alleviating the Curation Burden with CIViC

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



@handlerwagner

https://civicdb.org/graphics

Data and Knowledge Production

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

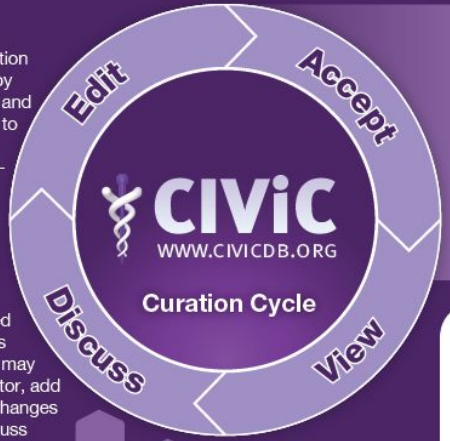
Sequences are aligned to the reference genome and tumor-specific events predicted.

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

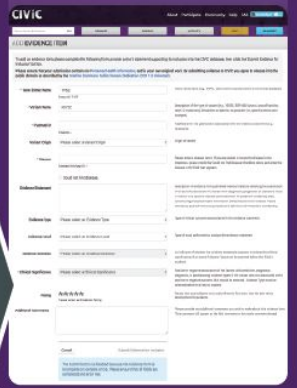
Events are annotated and scored in an effort to predict events of functional significance.

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.

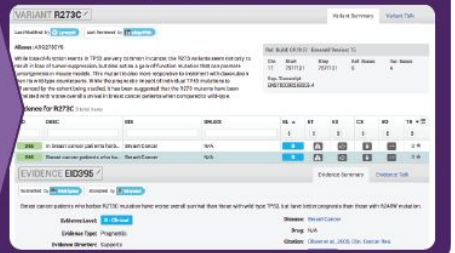


Add New Evidence



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.

@handlerwagner

https://civicdb.org/graphics

VARIANT V600E

Variant Summary | Variant Talk

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

HGVS Expressions:
NM_004333.4: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

Assertions:
AID7 | AID10 | AID20 | AID23

Show rejected:

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

HGVS ID
chr7:g.140453136A>T

SnpEff Effect
missense variant

SnpEff Impact
MODERATE

gnomAD Adj. AF
0

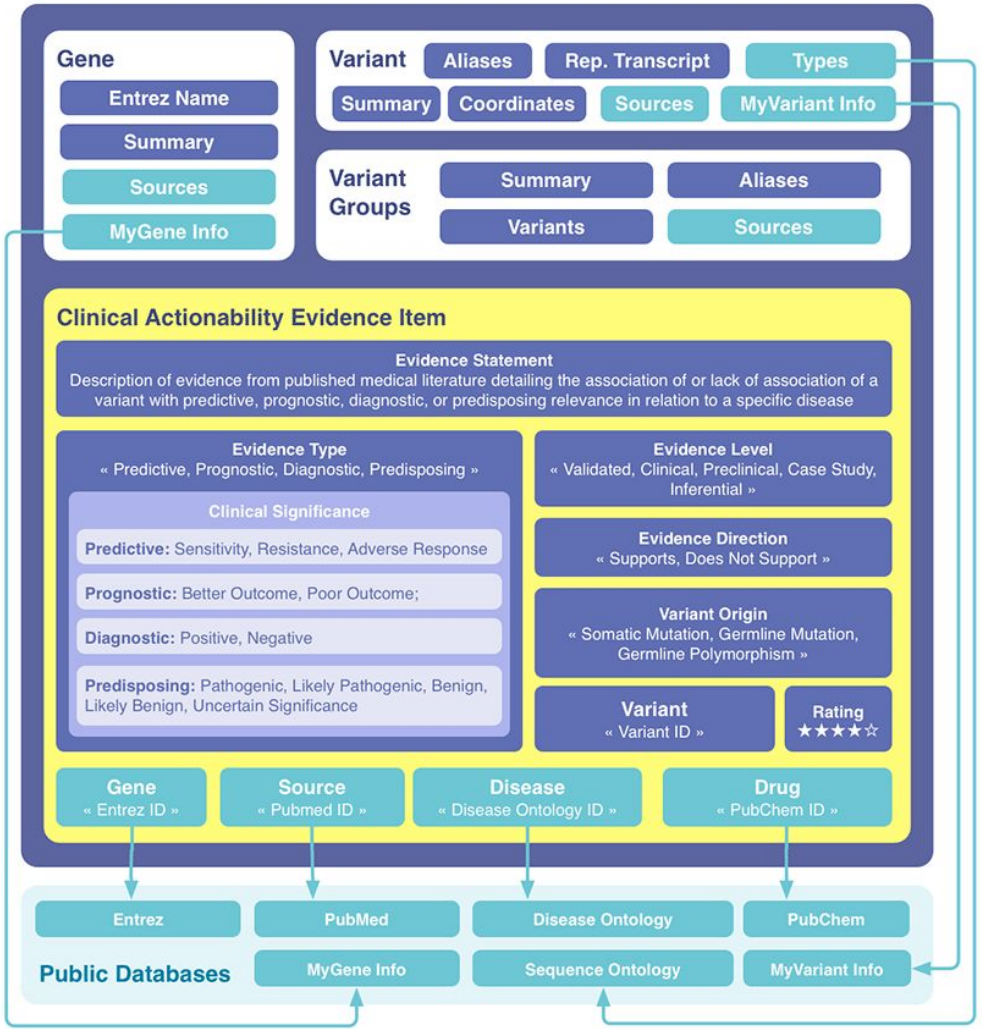
MyVariant.info

View MyVariant.info Details

Evidence for V600E 161 total items (showing 154)

Get Data | Help

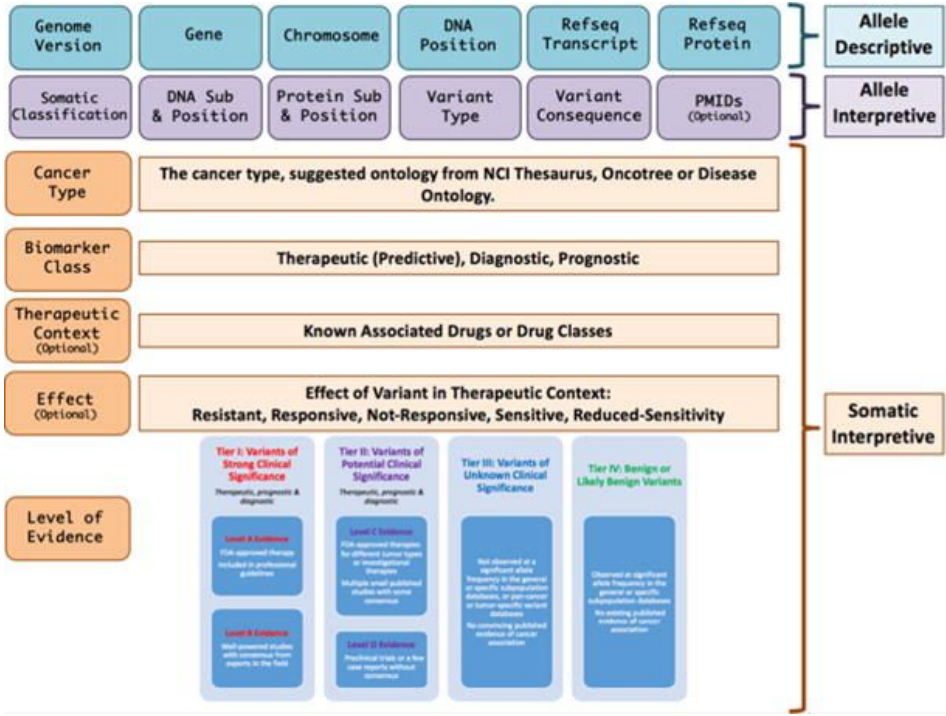
EID	DIS	DRUGS	DESC	EL	ET	ED	CS	VO	ER
7612	Colorectal Cancer	Binimetinib, Encorafenib, Cetuximab (Co...		A					5 ★
1409	Skin Melanoma	Vemurafenib		A					5 ★
3017	Lung Non-small Cell Carcinoma	Dabrafenib, Trametinib (Combination)		A					4 ★
102	Thyroid Gland Papillary Carcinoma	N/A		B					5 ★



@handlerwagner

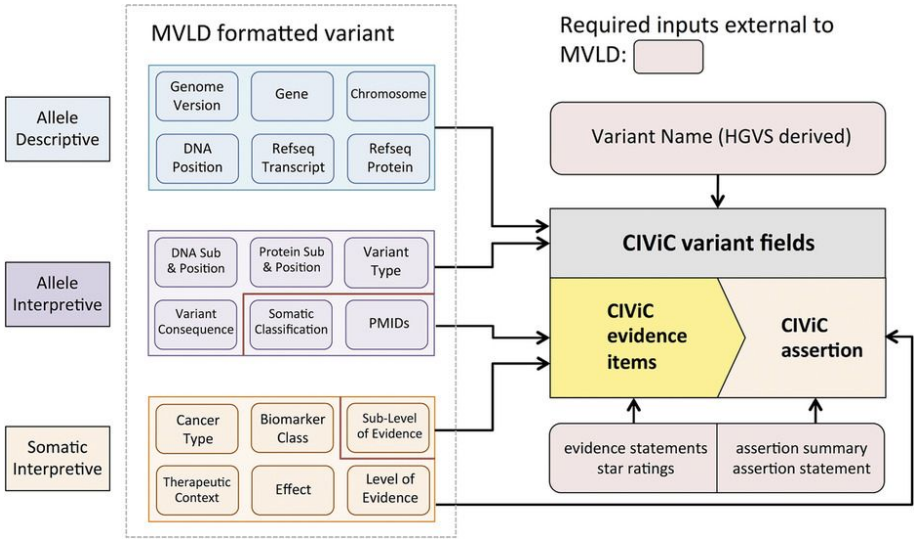
https://civicdb.org/graphics

ClinGen Somatic WG: MVLD Guidelines

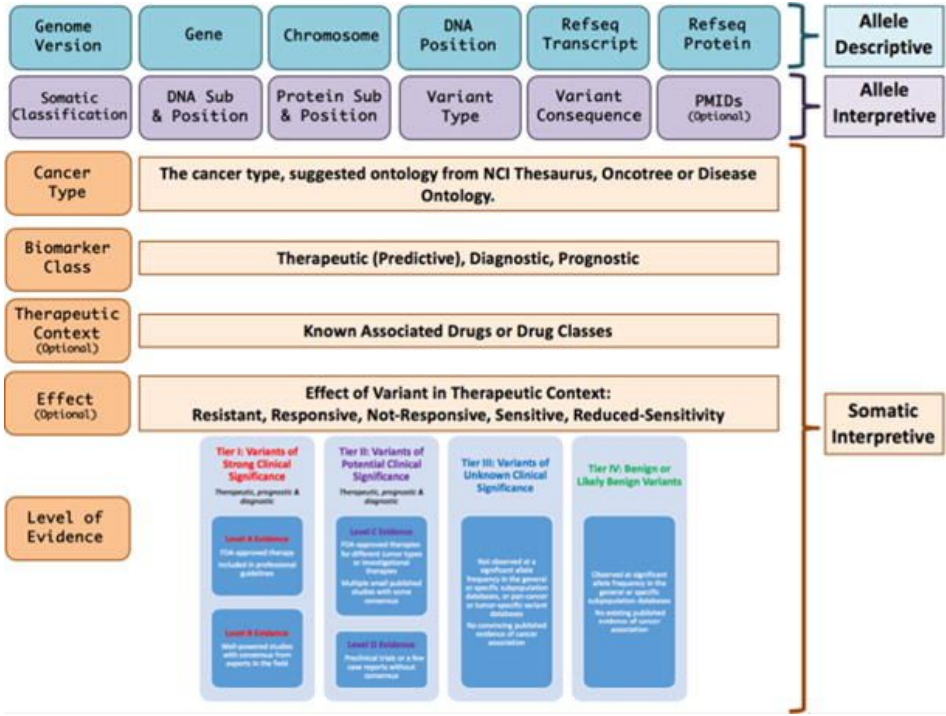


Madhavan S, et al. *Pac. Symp. Biocomput.* 2018

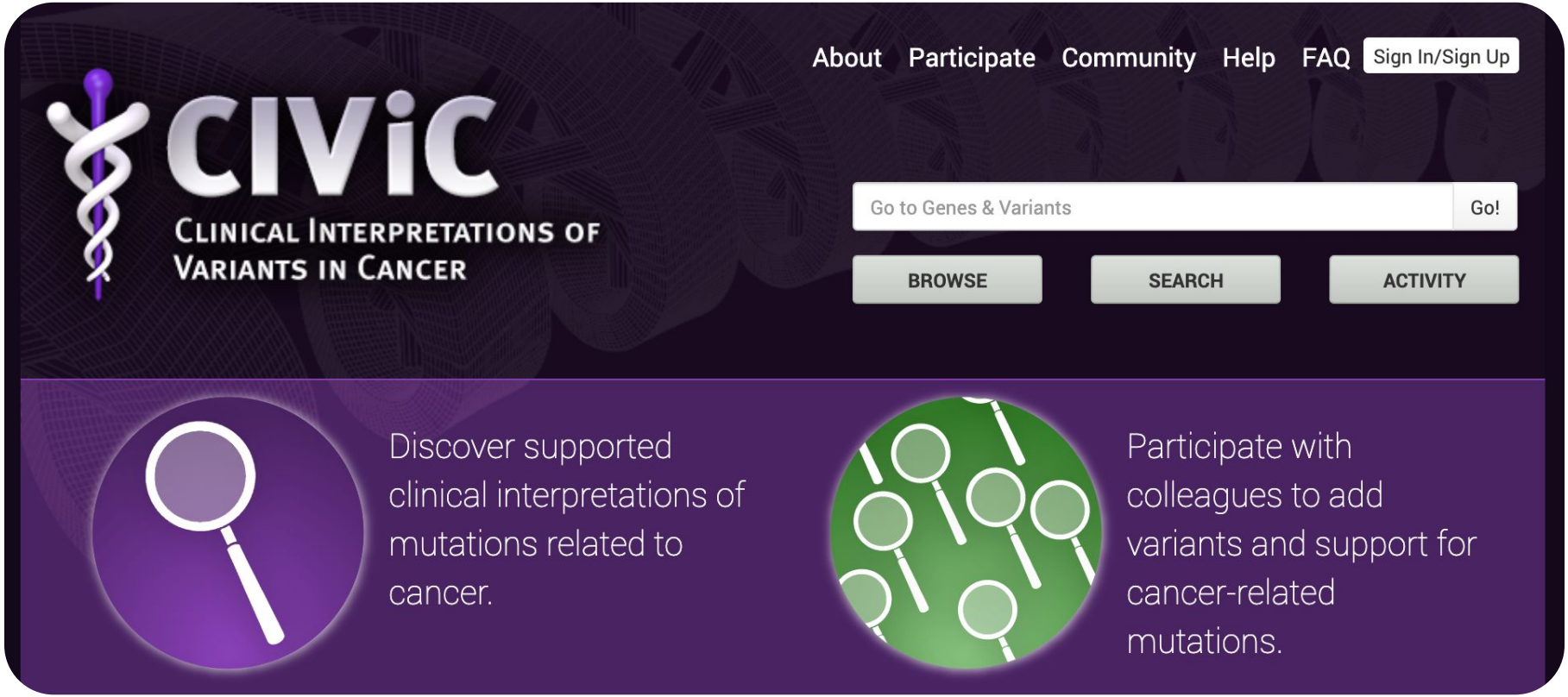
ClinGen Somatic WG: MVLD Guidelines



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



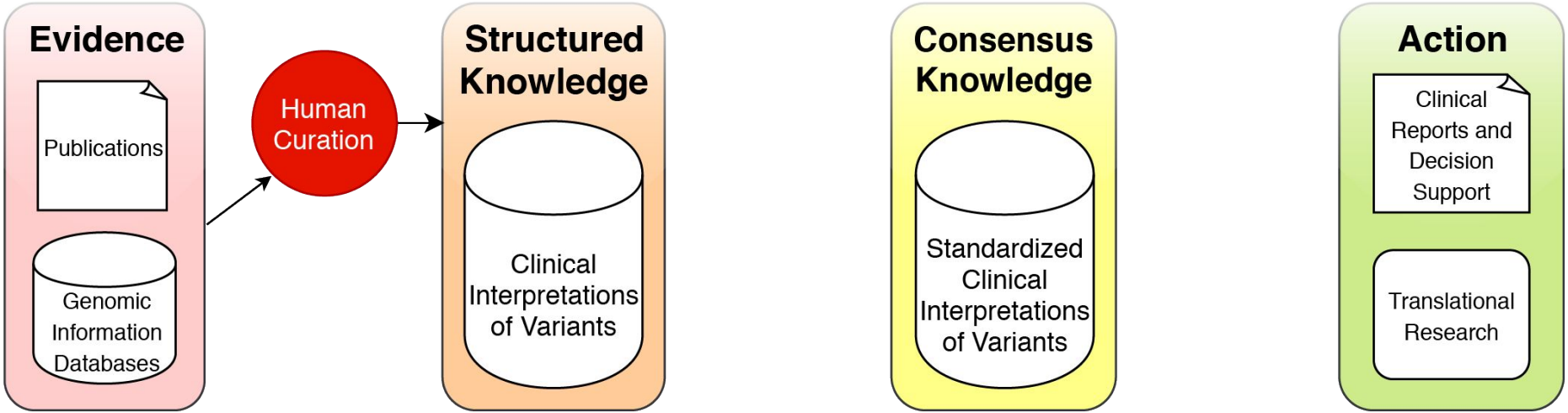
Madhavan S, et al. *Pac. Symp. Biocomput.* 2018



https://civicdb.org

@handlerwagner

Roadmap from Evidence to Action



Problem: CIViC is a Silo in a Diverse Knowledge Ecosystem

@handlerwagner

Variant Interpretation Knowledge is Siloed

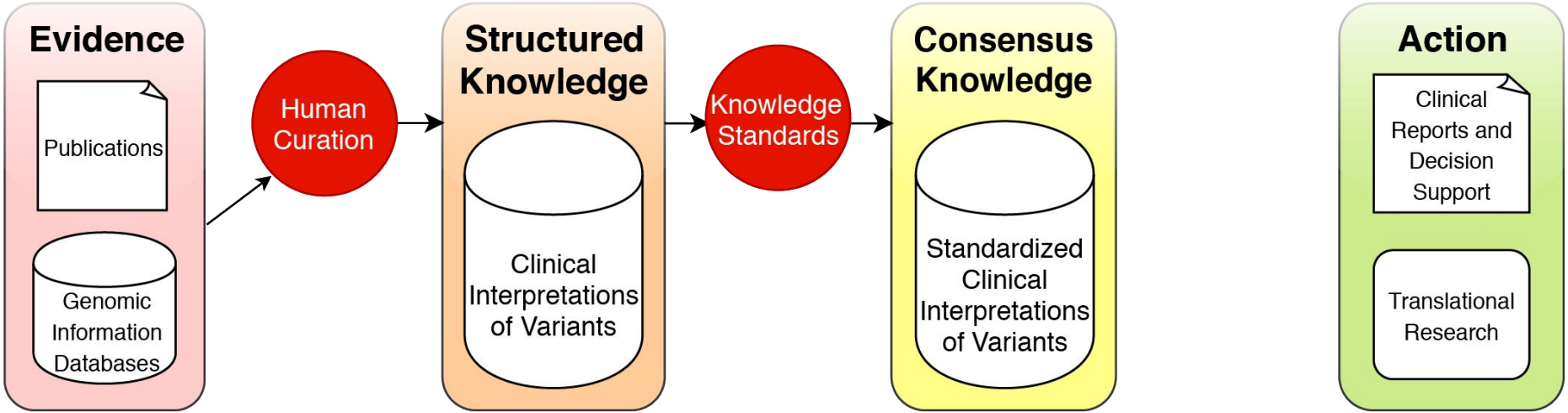
Established Interpretation Knowledgebases

- 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



VICC | Variant Interpretation for Cancer Consortium

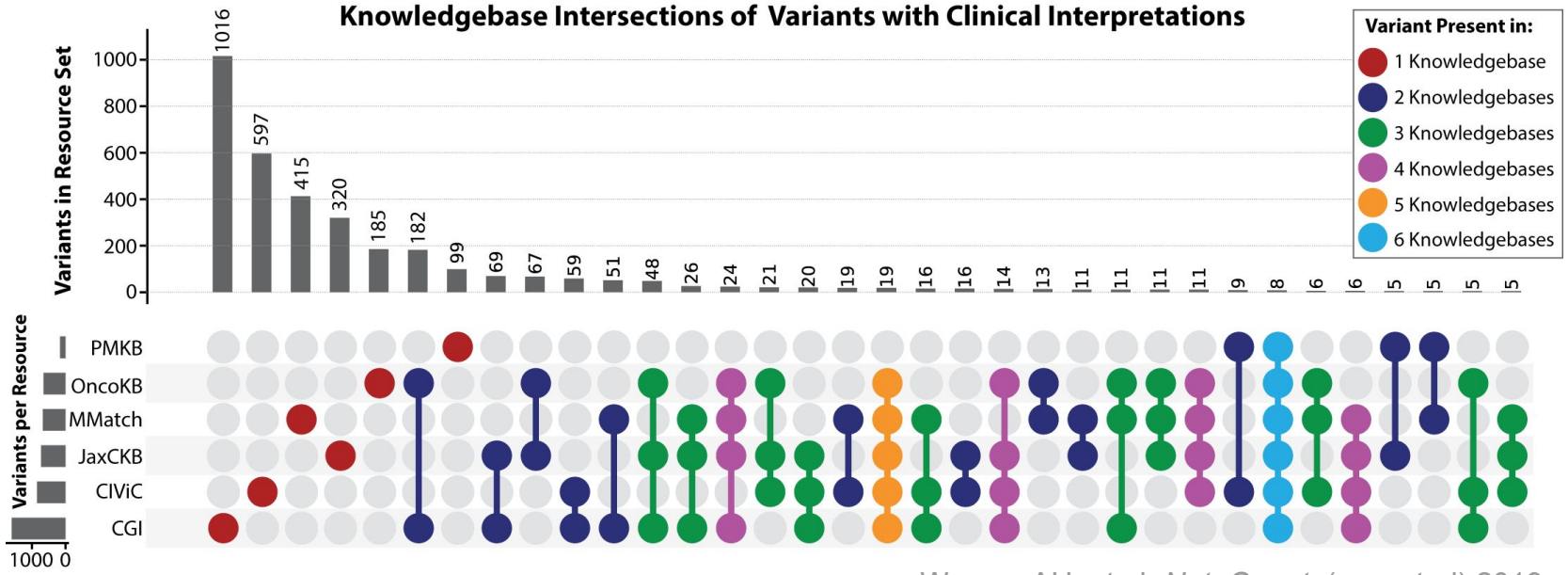
- Year Started: 2016
- Country: Global – USA, Barcelona, UK
- Institutions: WashU, MSKCC, DFCI, OHSU, IRB Barcelona, Cornell, Jackson Labs
  - –Consortium of Clinical Interpretations KBs: OncoKB, Jax-CKB, CIViC, PMKB, CGI, ...
- Mission:
  - Global integration of knowledgebases for clinical interpretation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

@handlerwagner

http://cancervariants.org/

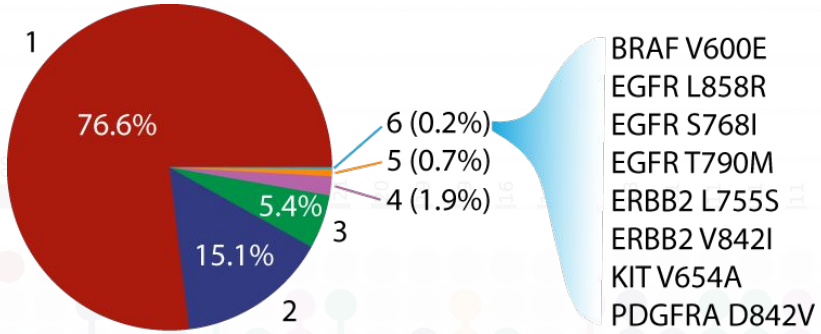
Variant Interpretations are Heterogeneous

Variant Representation Across Knowledgebases



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

Variant Representation Across Knowledgebases



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

Poor overlap of variants creates
inconsistent clinical interpretations

This can be addressed by
aggregating curated knowledge

Variant Interpretations are Structurally Disparate

@handlerwagner

Diversity in Structure of Variant Interpretations

	OncoKB	CIViC Clinical Interpretations of Variants in Cancer	PMKB
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.

@handlerwagner

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

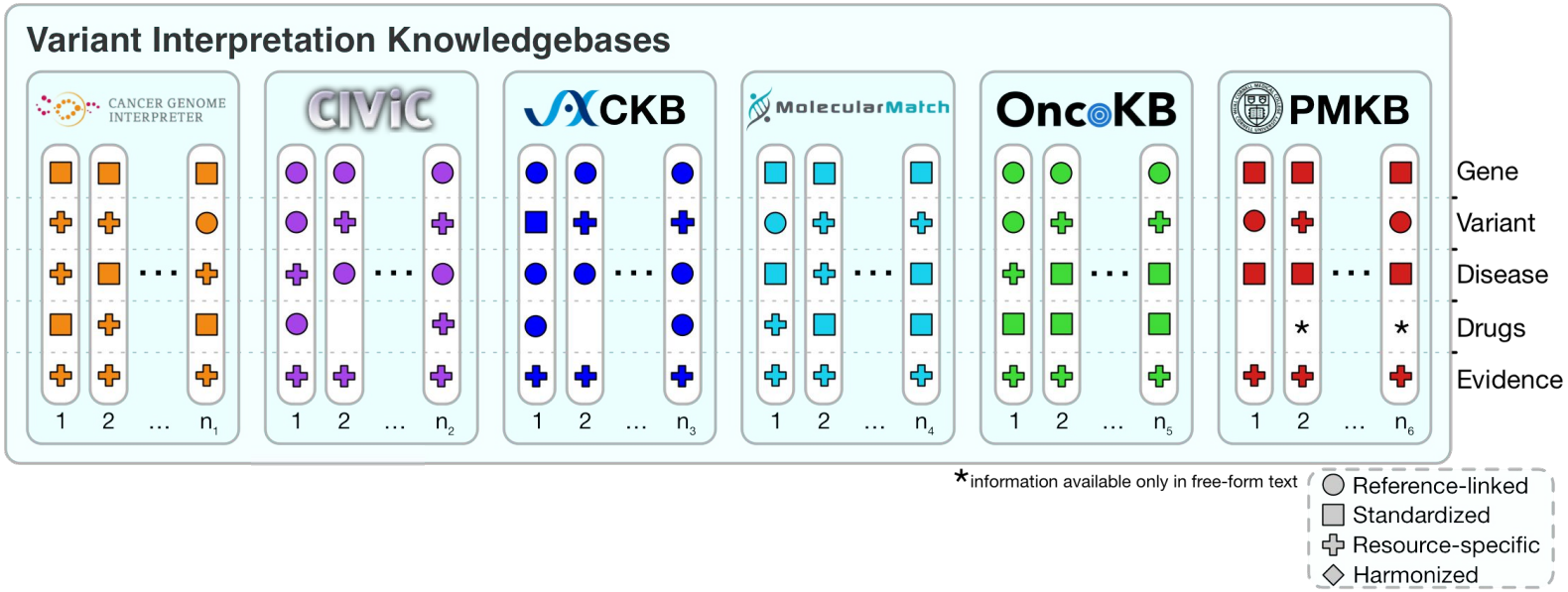
Poor overlap of variants creates
inconsistent clinical interpretations

We can aggregate, but...

Clinical variant interpretation
knowledgebases are not interoperable

This can be addressed by harmonizing interpretations

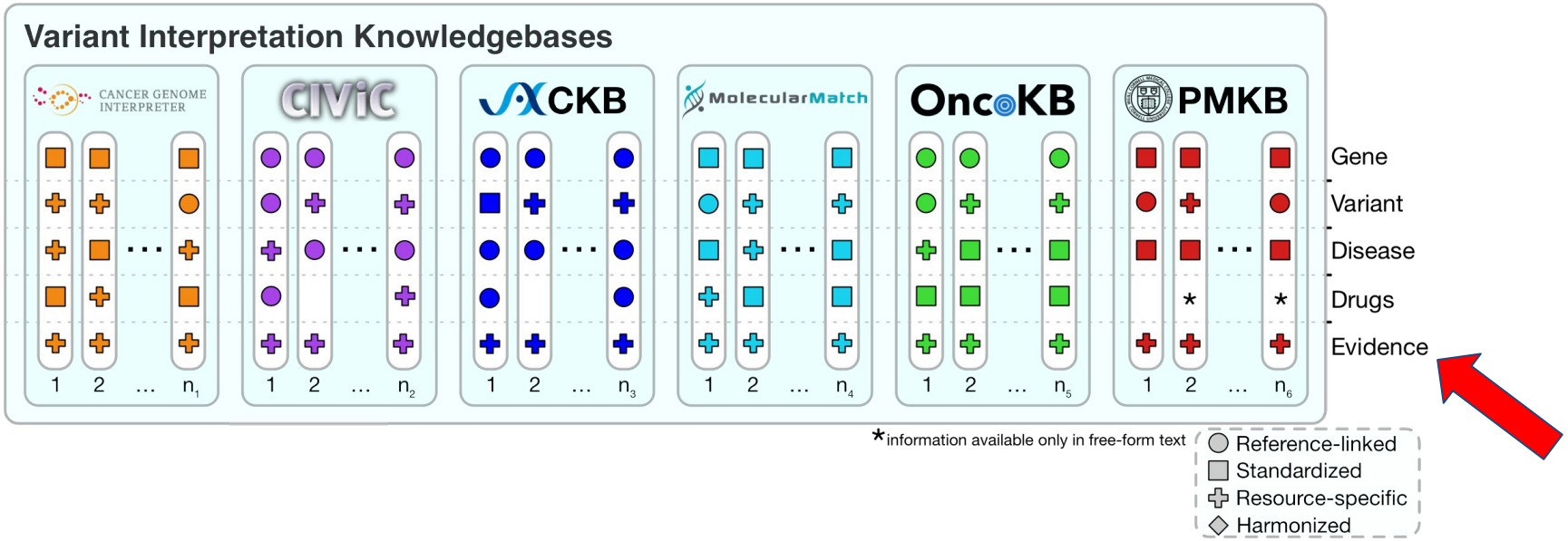
Structure of a Variant Interpretation



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

@handlerwagner

Structure of a Variant Interpretation



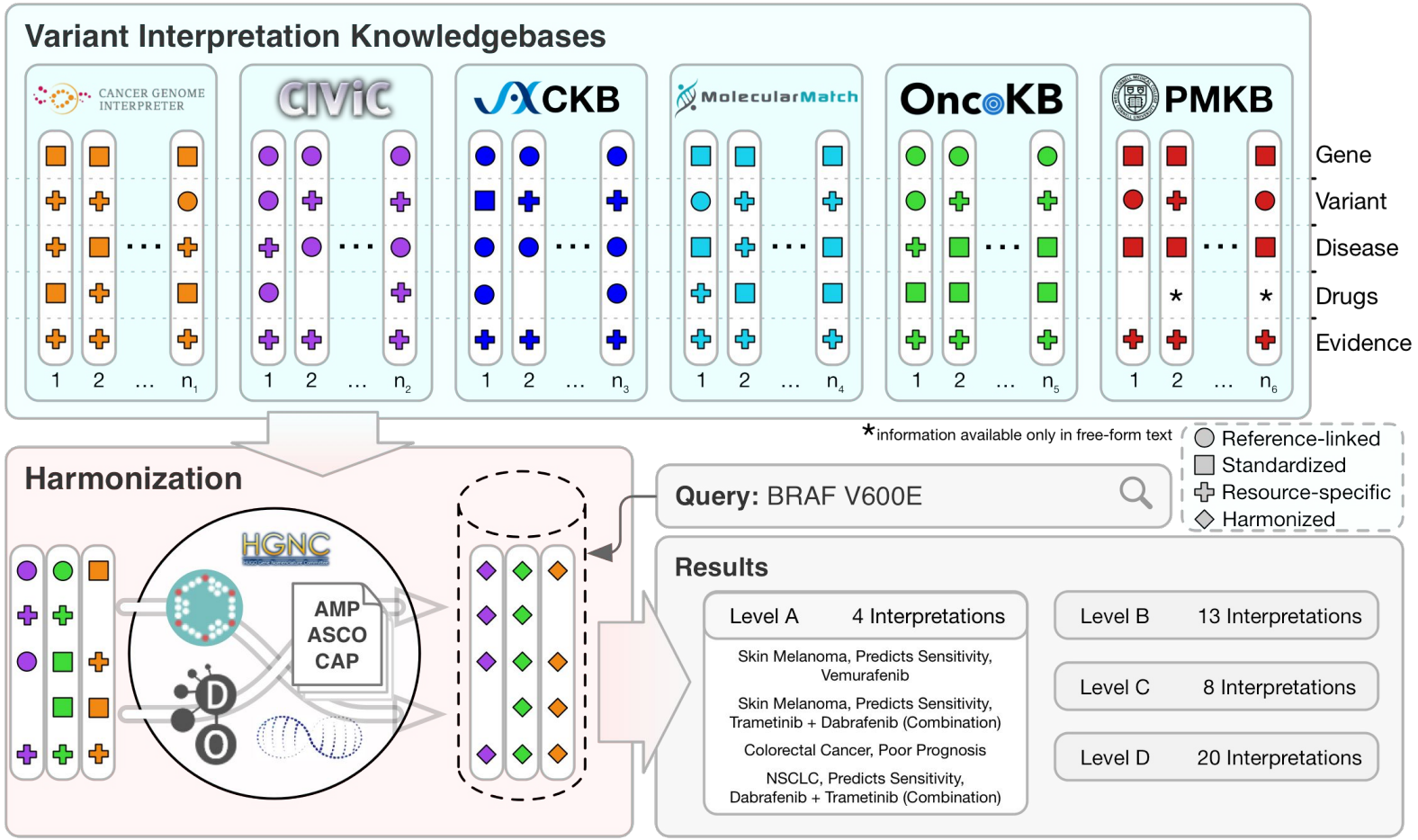
Wagner AH, et al. *Nat. Genet.* (accepted) 2019

@handlerwagner

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

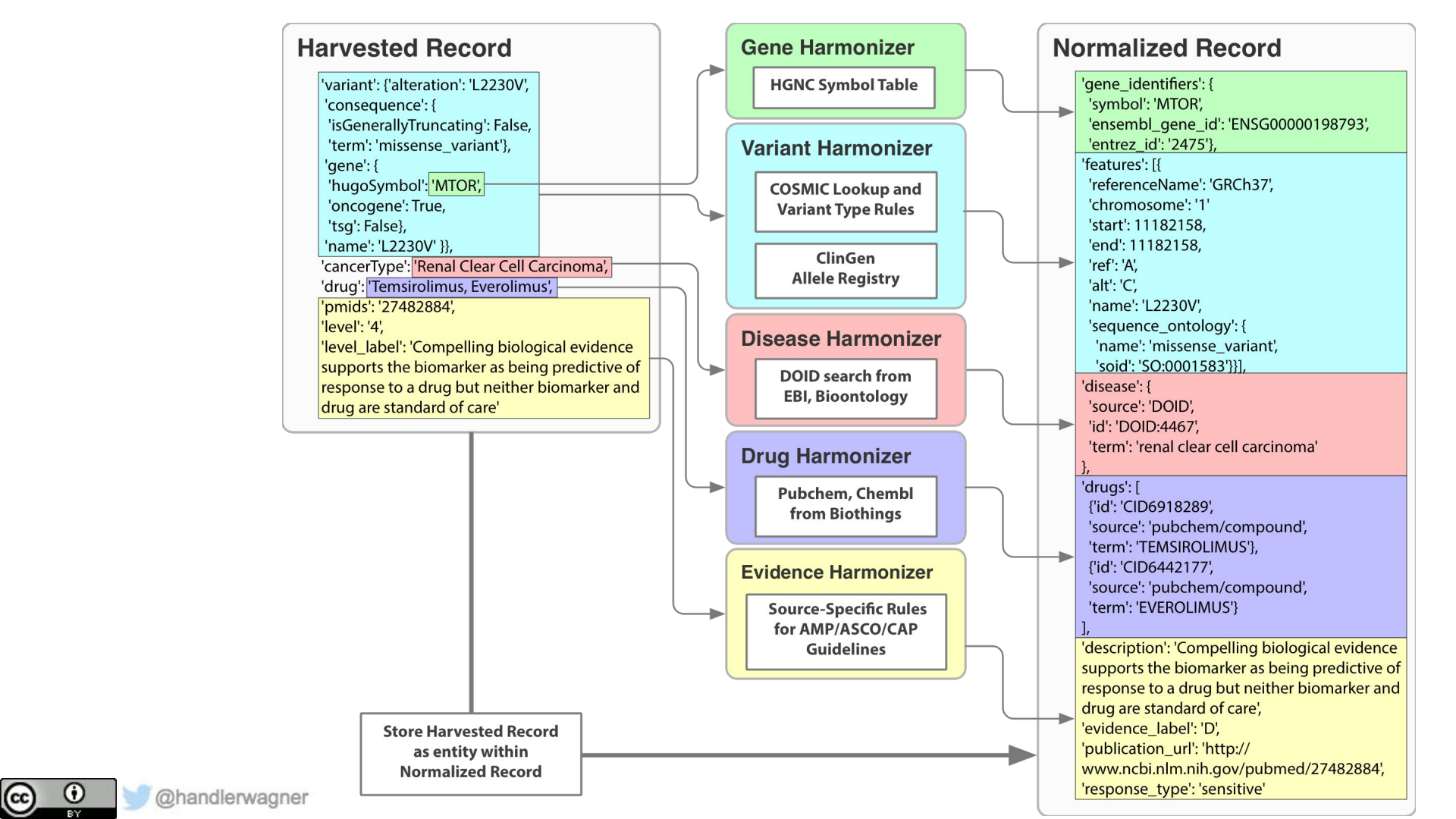
Wagner AH, et al. *Nat. Genet.* (accepted) 2019

@handlerwagner



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

@handlerwagner



Harmonization Increases Breadth of Interpretation

@handlerwagner

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

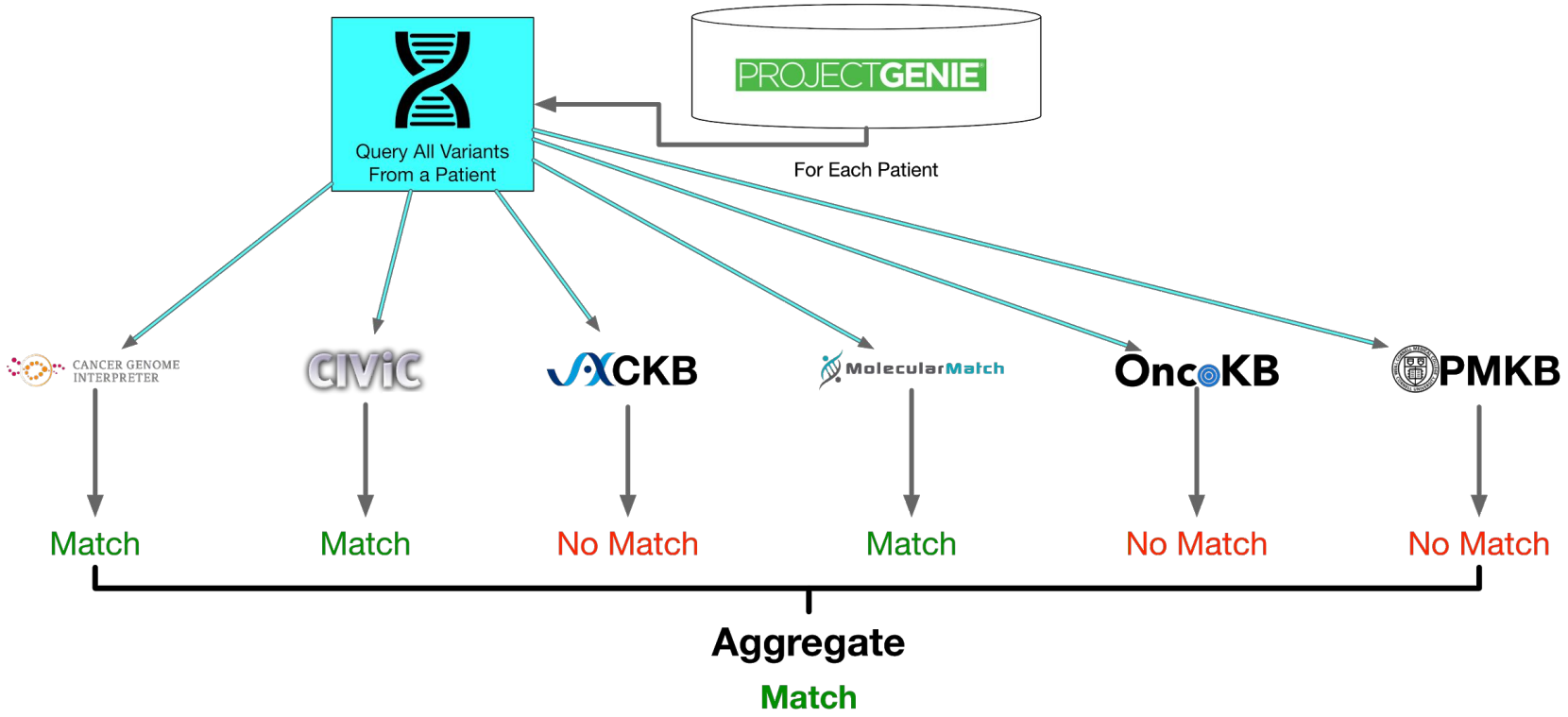
https://www.aacr.org/RESEARCH/RESEARCH/PAGES/AACR-PROJECT-GENIE.ASPX

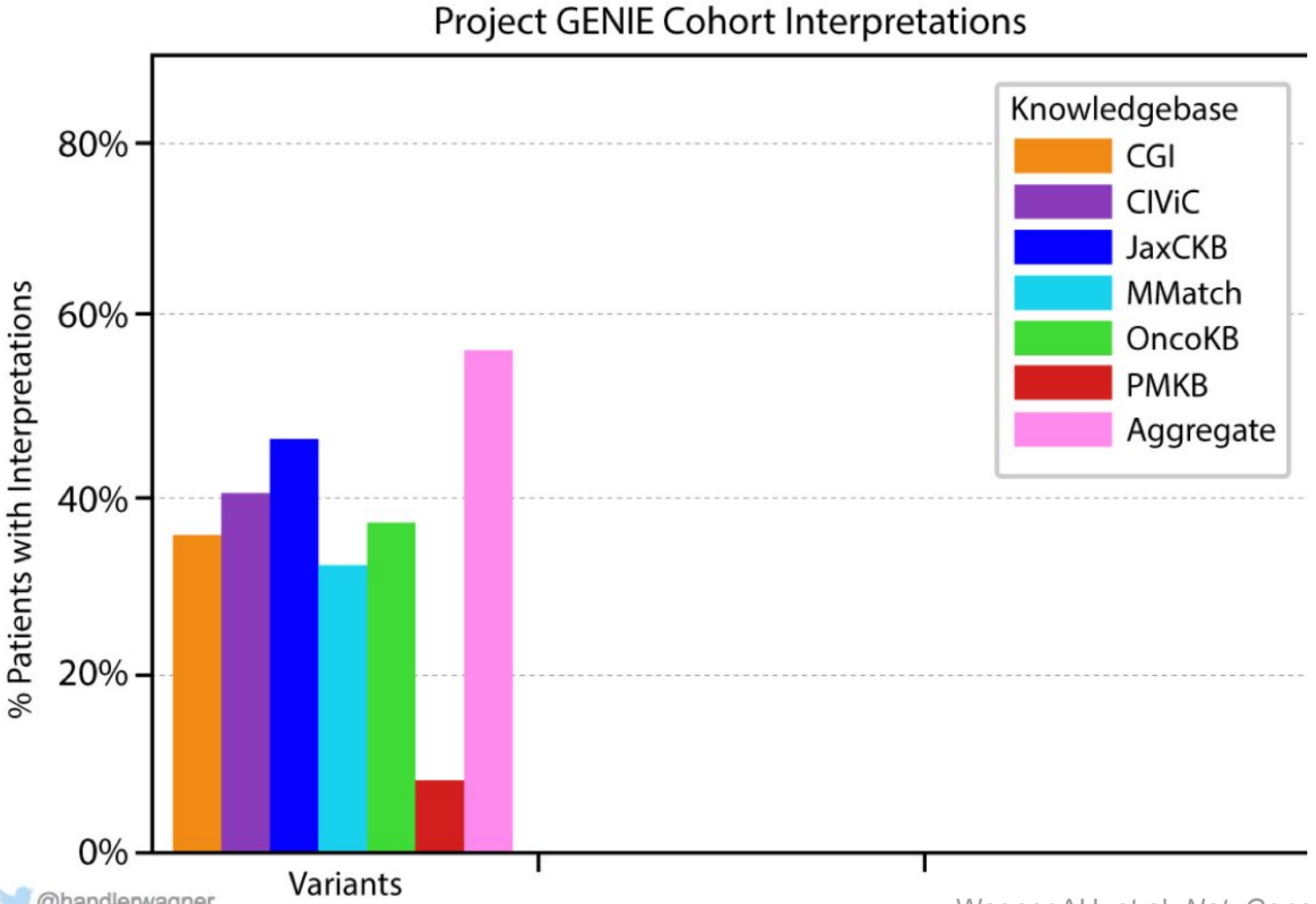
Project GENIE Cohort

- 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

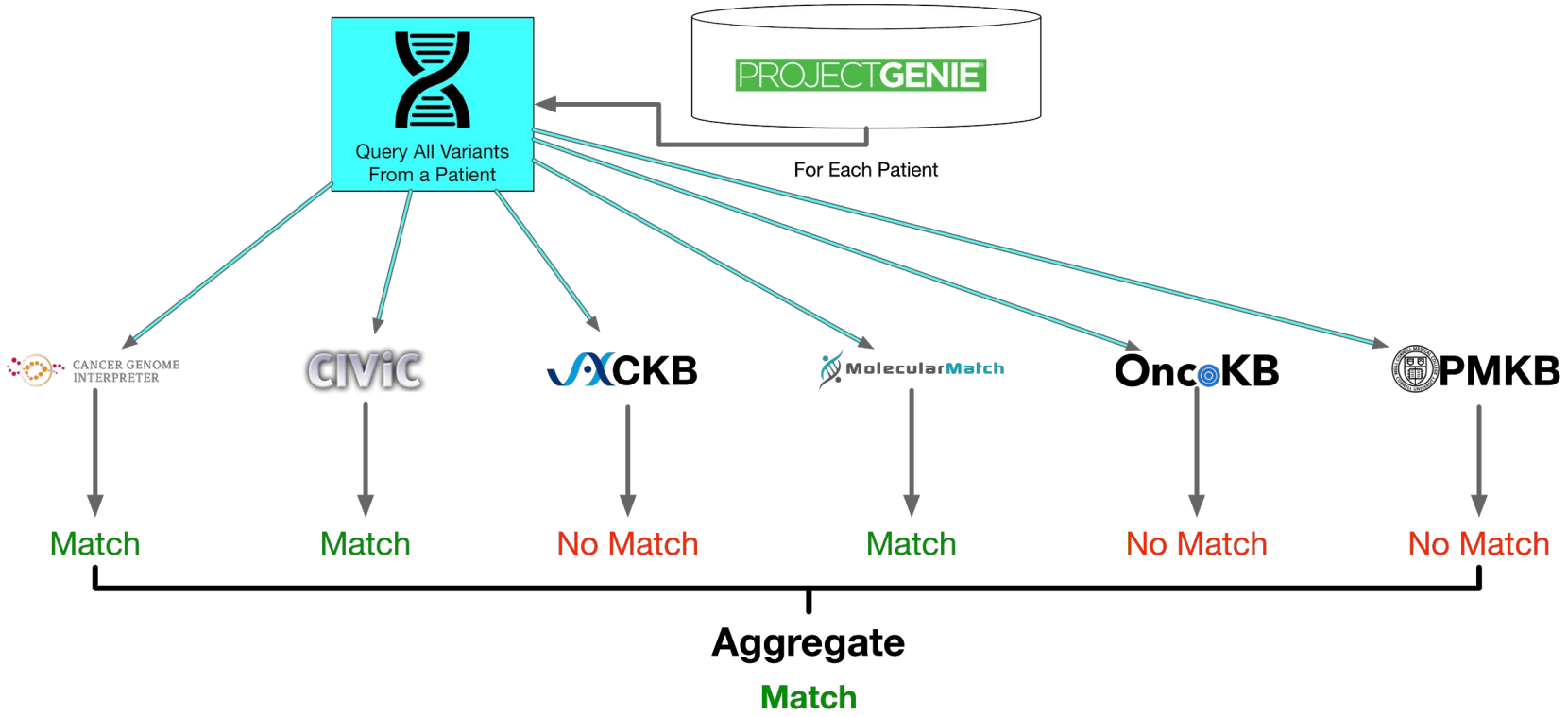




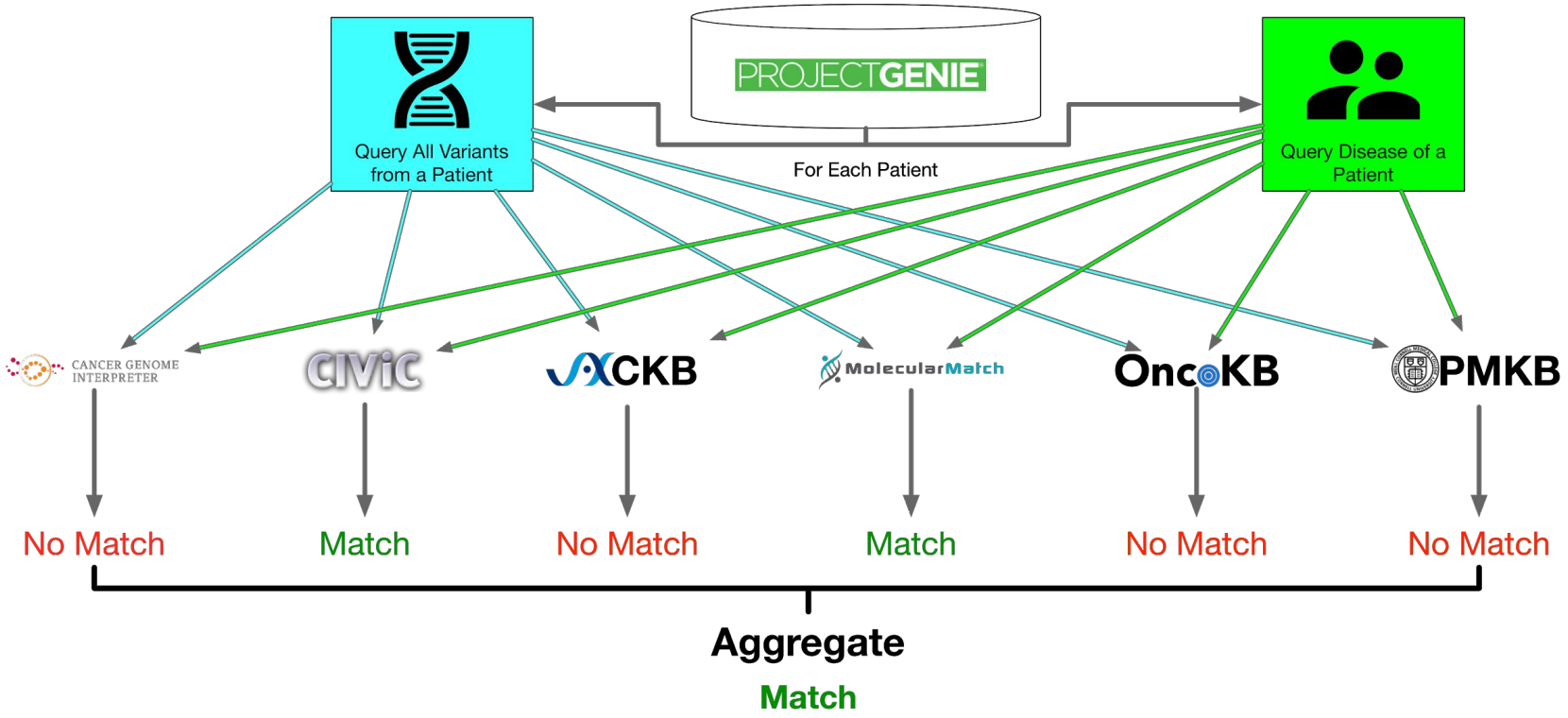
@handlerwagner

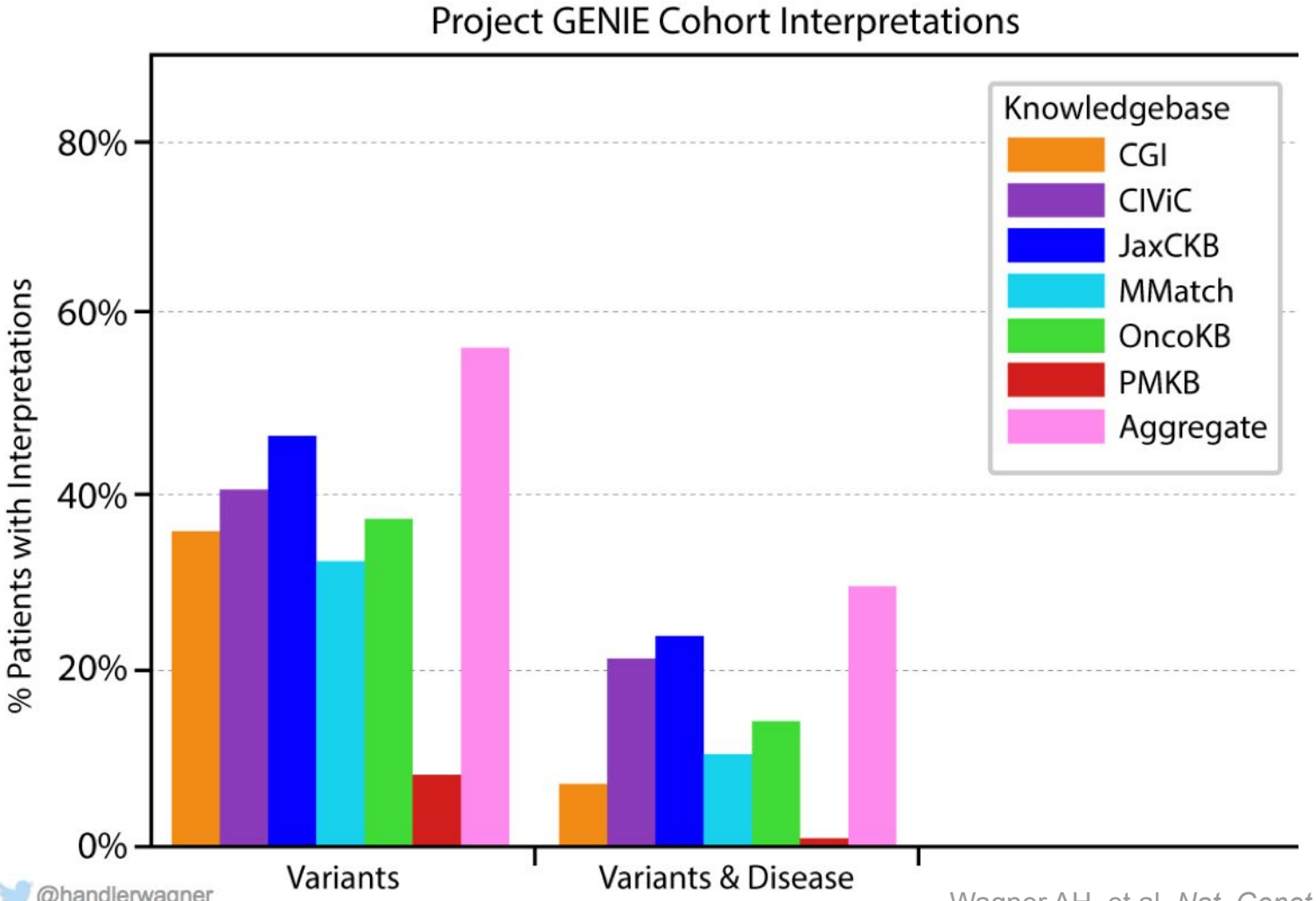
Wagner AH, et al. *Nat. Genet.* (accepted) 2019

Interpretation Coverage of GENIE Cohort



Interpretation Coverage of GENIE Cohort

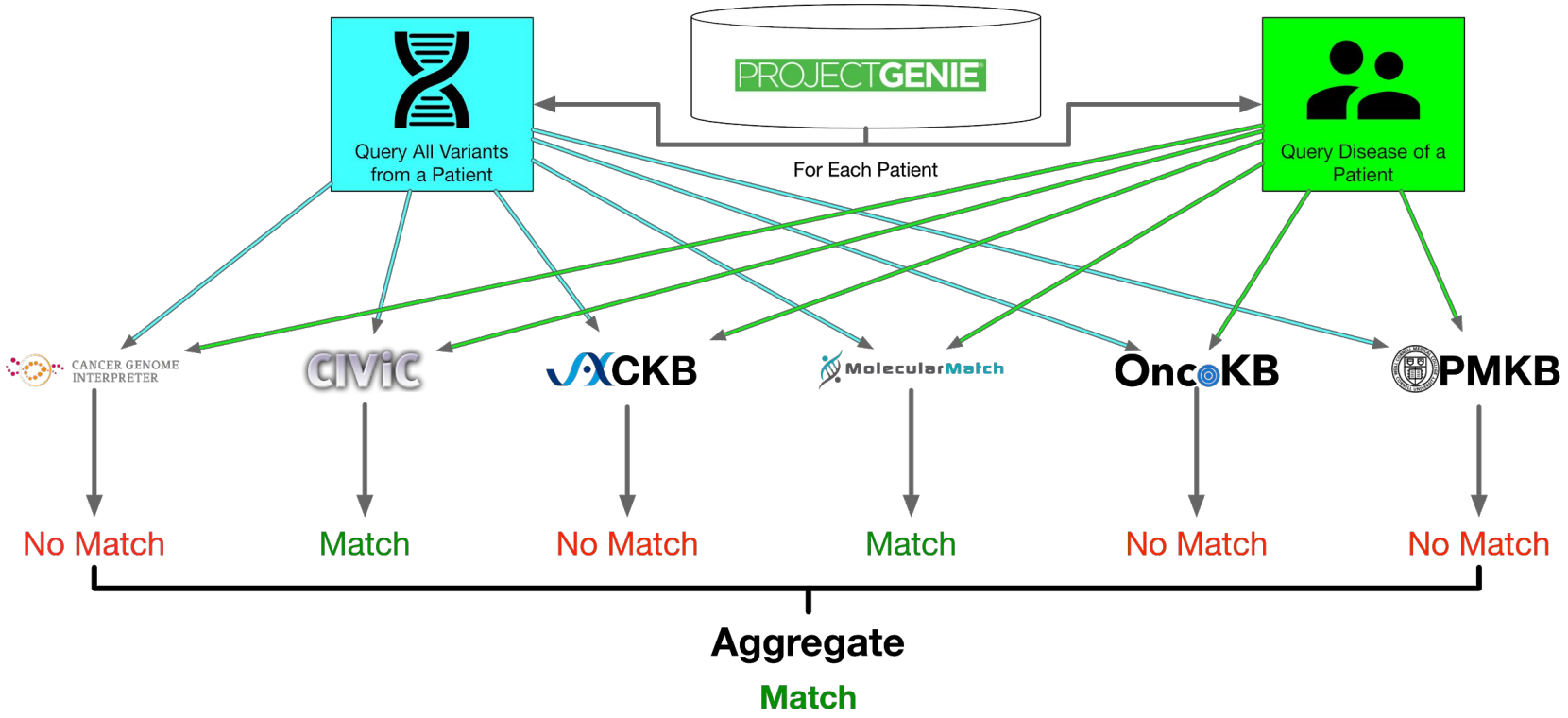




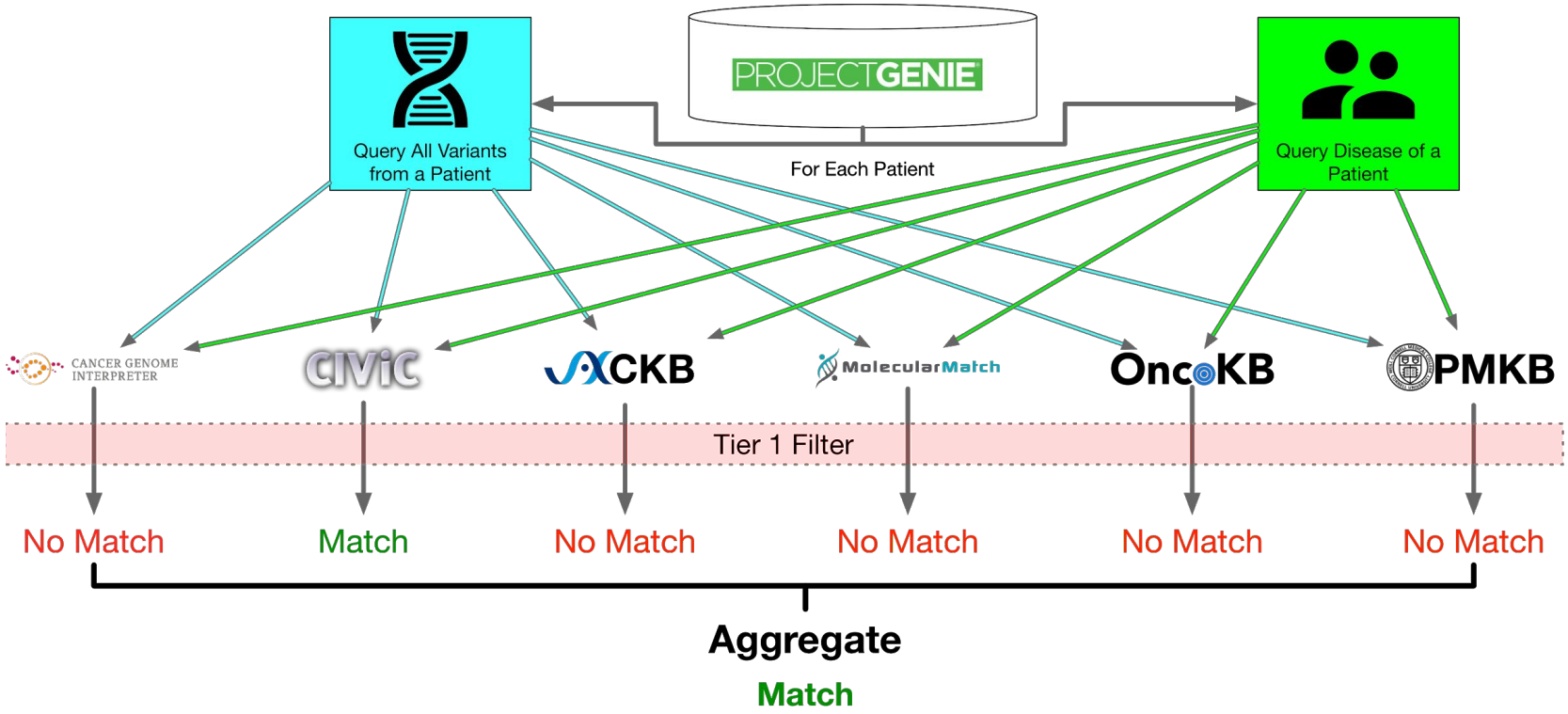
@handlerwagner

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

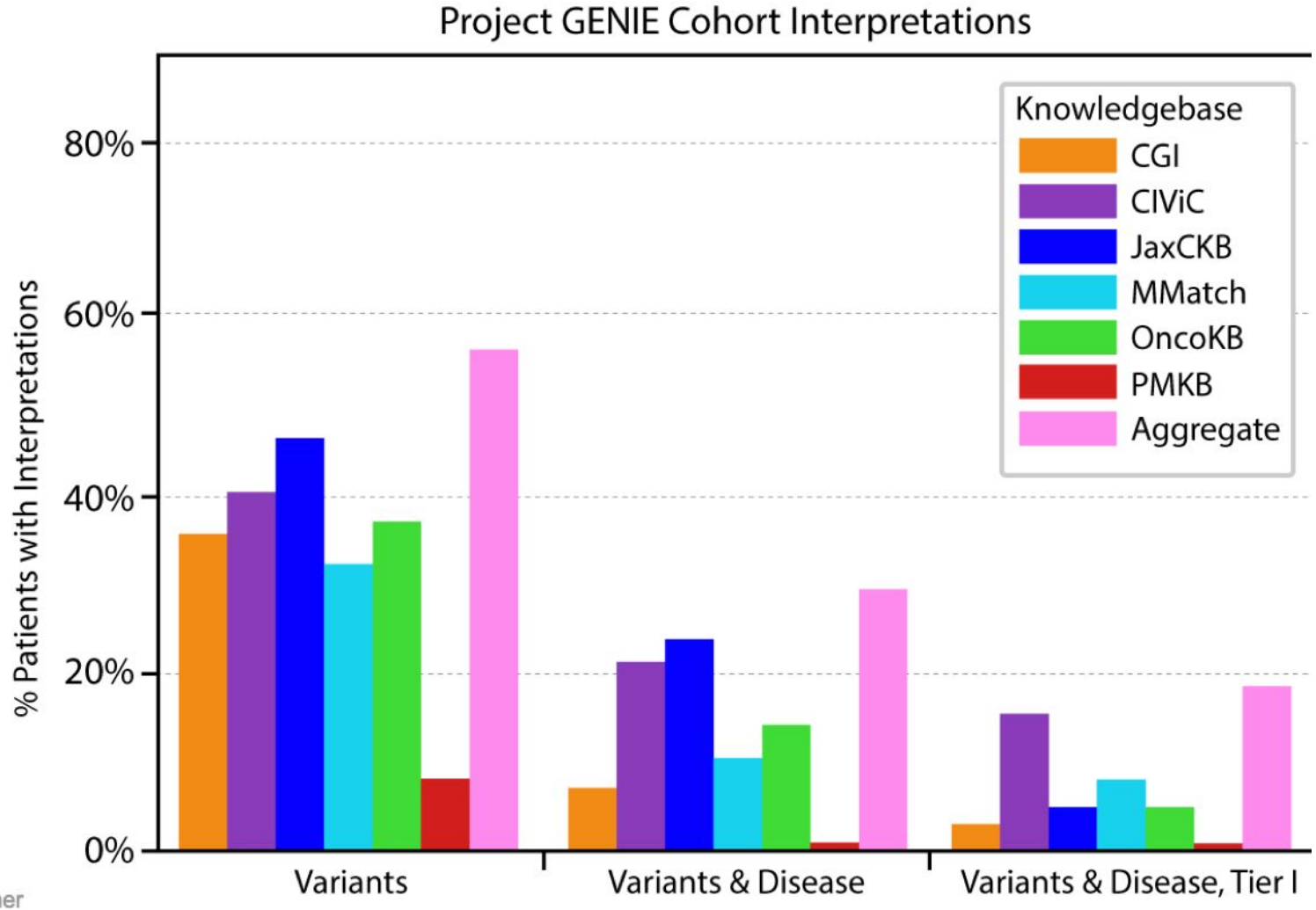
Interpretation Coverage of GENIE Cohort

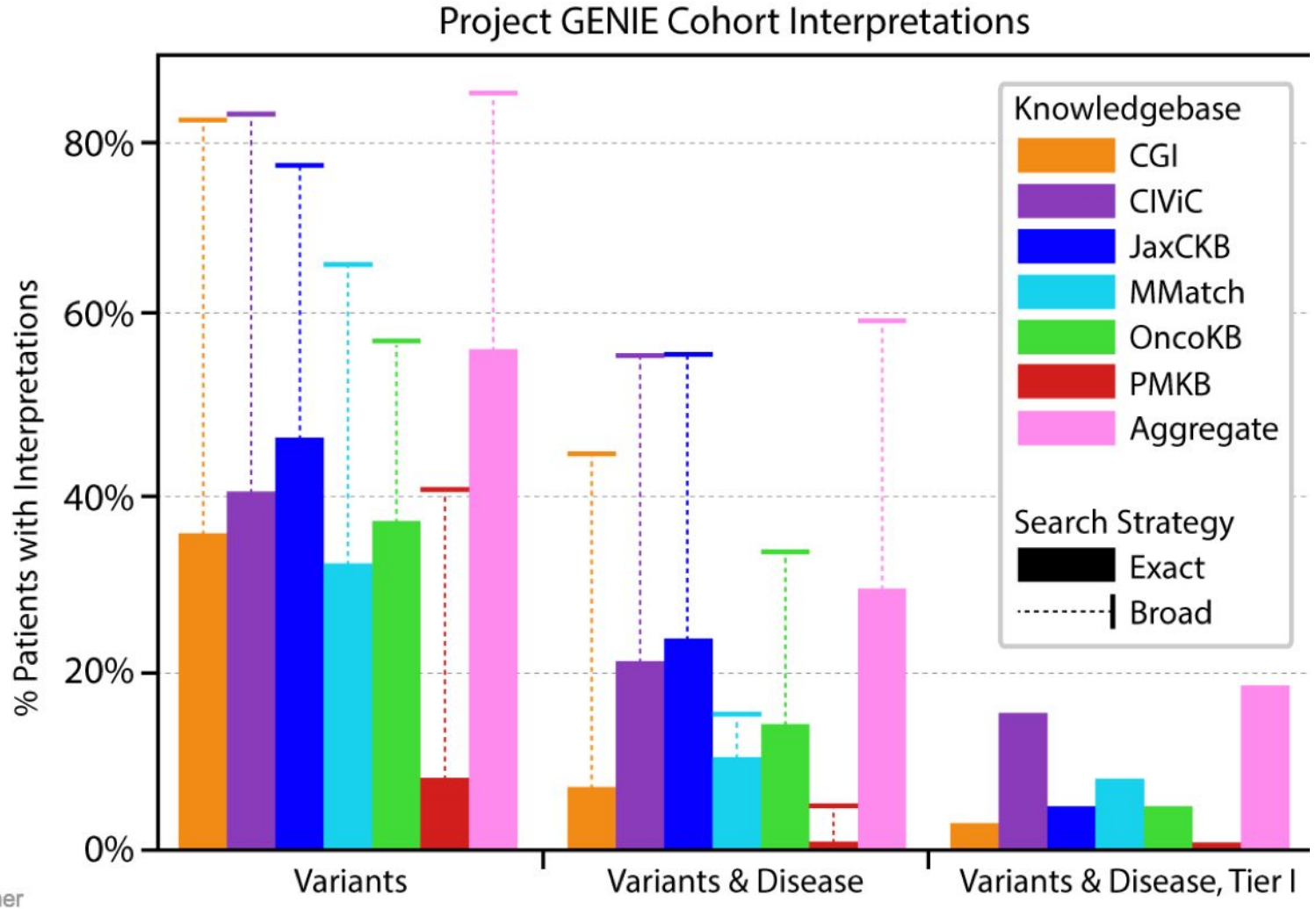


Interpretation Coverage of GENIE Cohort



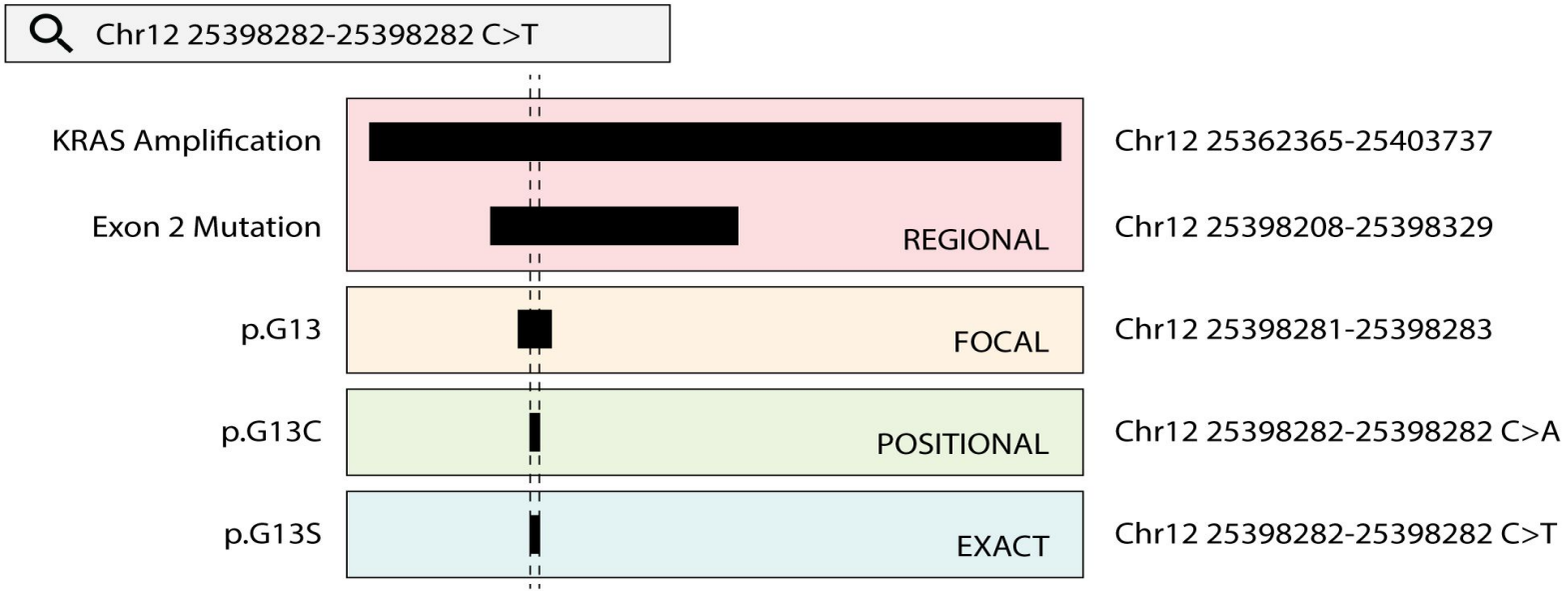
@handlerwagner





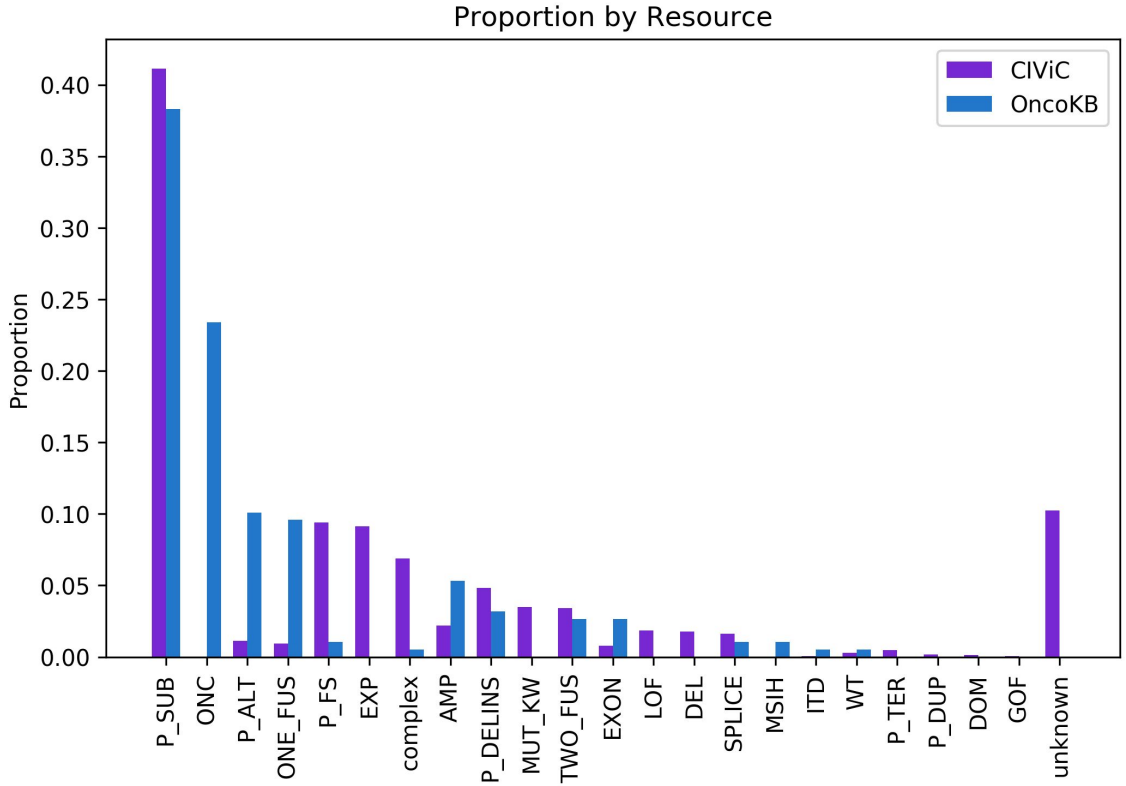
Meta-KB Search Strategies

a



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

@handlerwagner

Global Alliance
for Genomics & Health

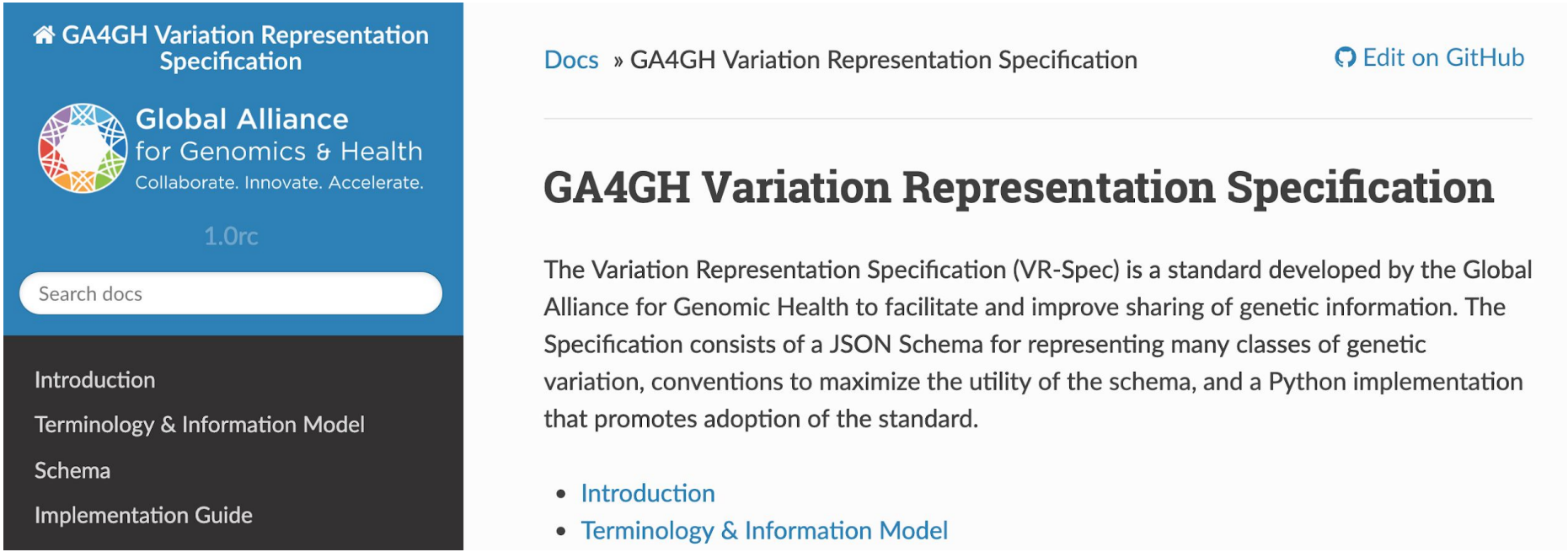
Collaborate. Innovate. Accelerate.

@handlerwagner

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/

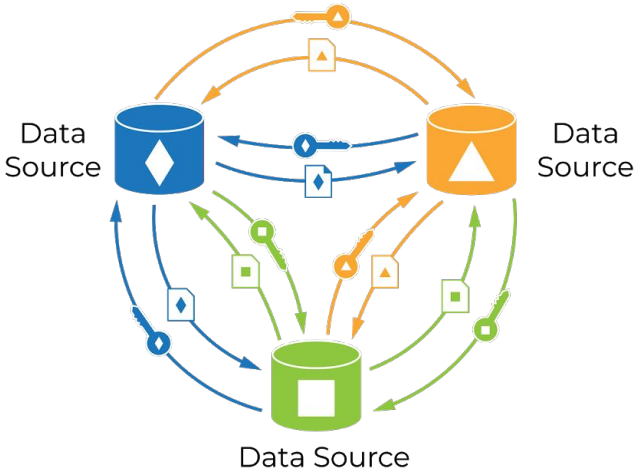


https://vr-spec.readthedocs.io

@handlerwagner

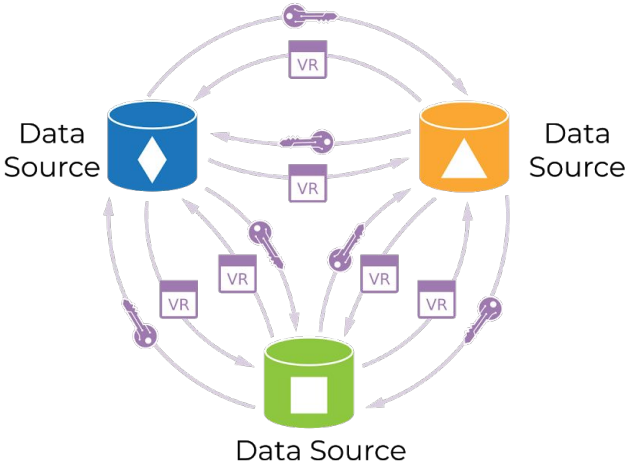
VR Specification Driving Interoperability

CURRENTLY...



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

WITH THE VR SPECIFICATION...

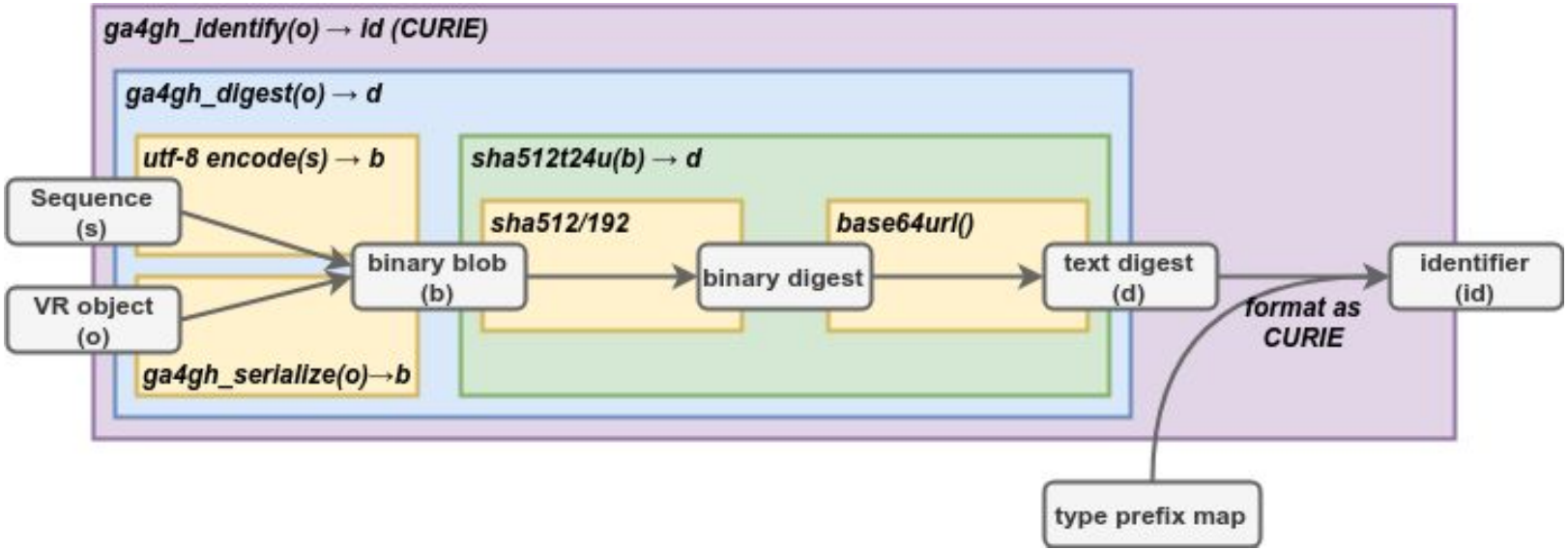


SYSTEMS USE A COMMON IDENTIFIER, COMPUTED FROM THE DATA ITSELF, AND A COMMON DATA FORMAT. ADDING A NEW SYSTEM IS MUCH EASIER.

@reecehart

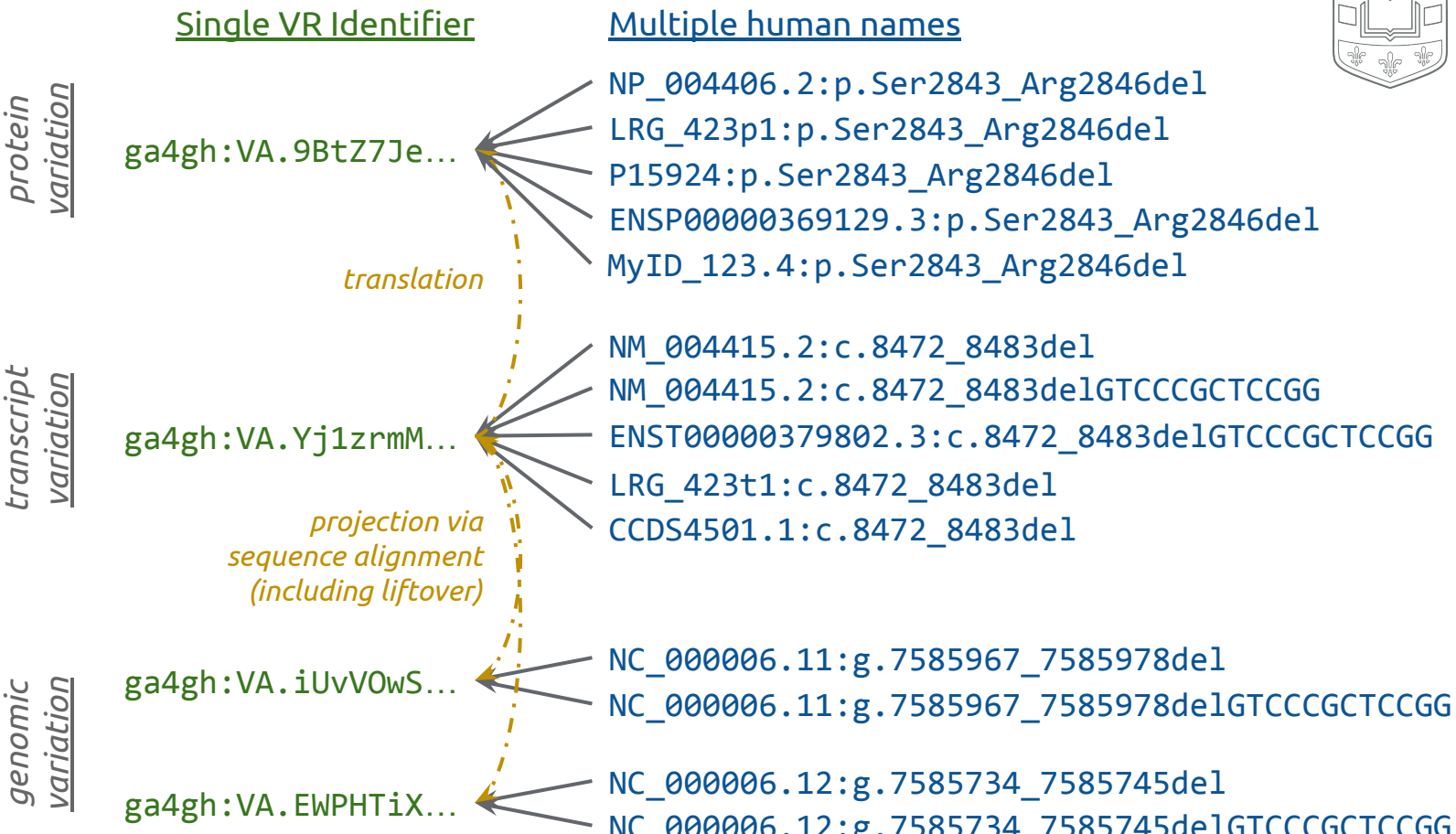
Rendering by Stephanie Li

VR Computed Identifiers



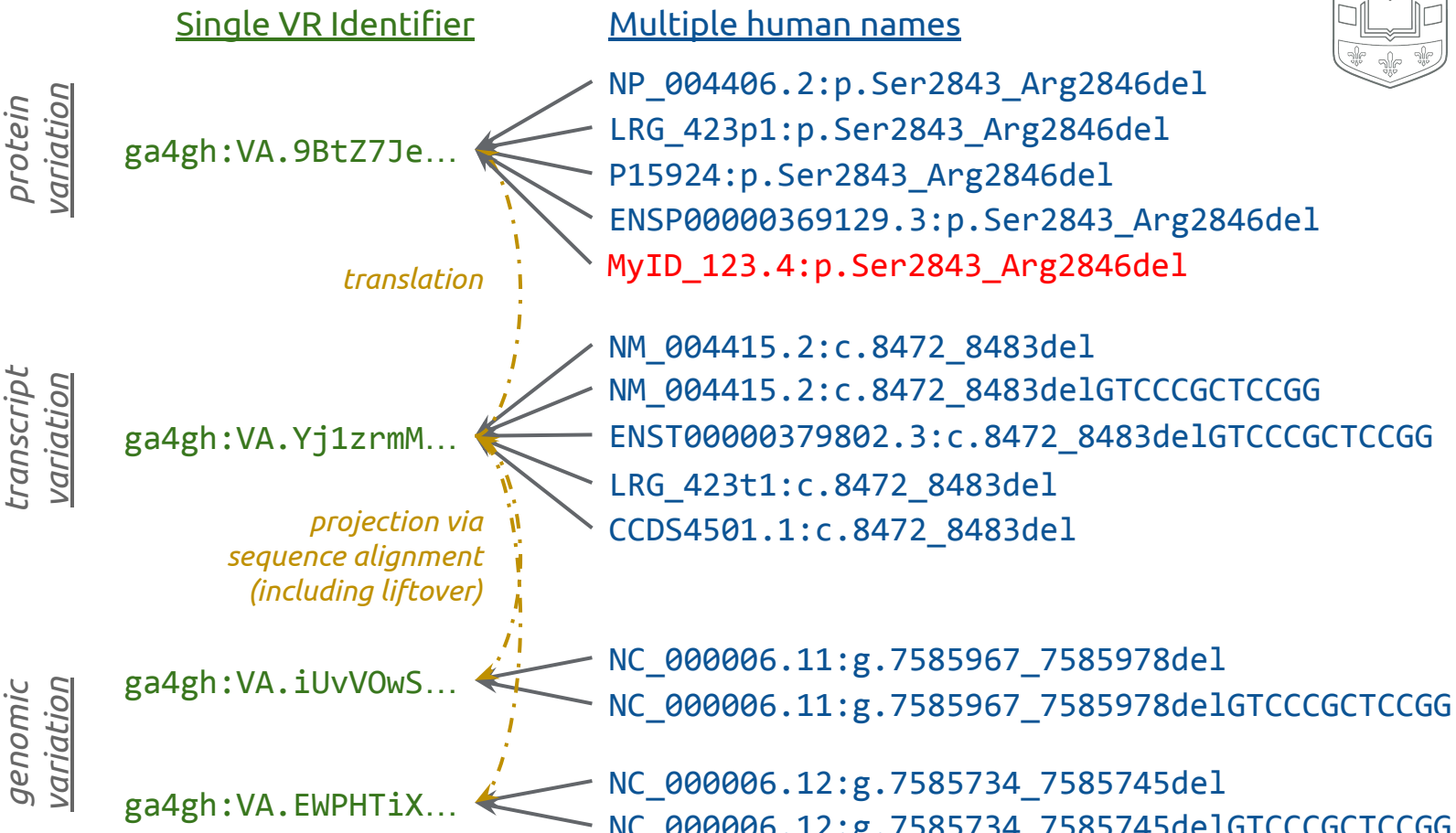
https://vr-spec.readthedocs.io/en/1.0/impl-guide/computed_identifiers.html

VR identifiers: precise, unique names for variation



@reecehart & @handlerwagner

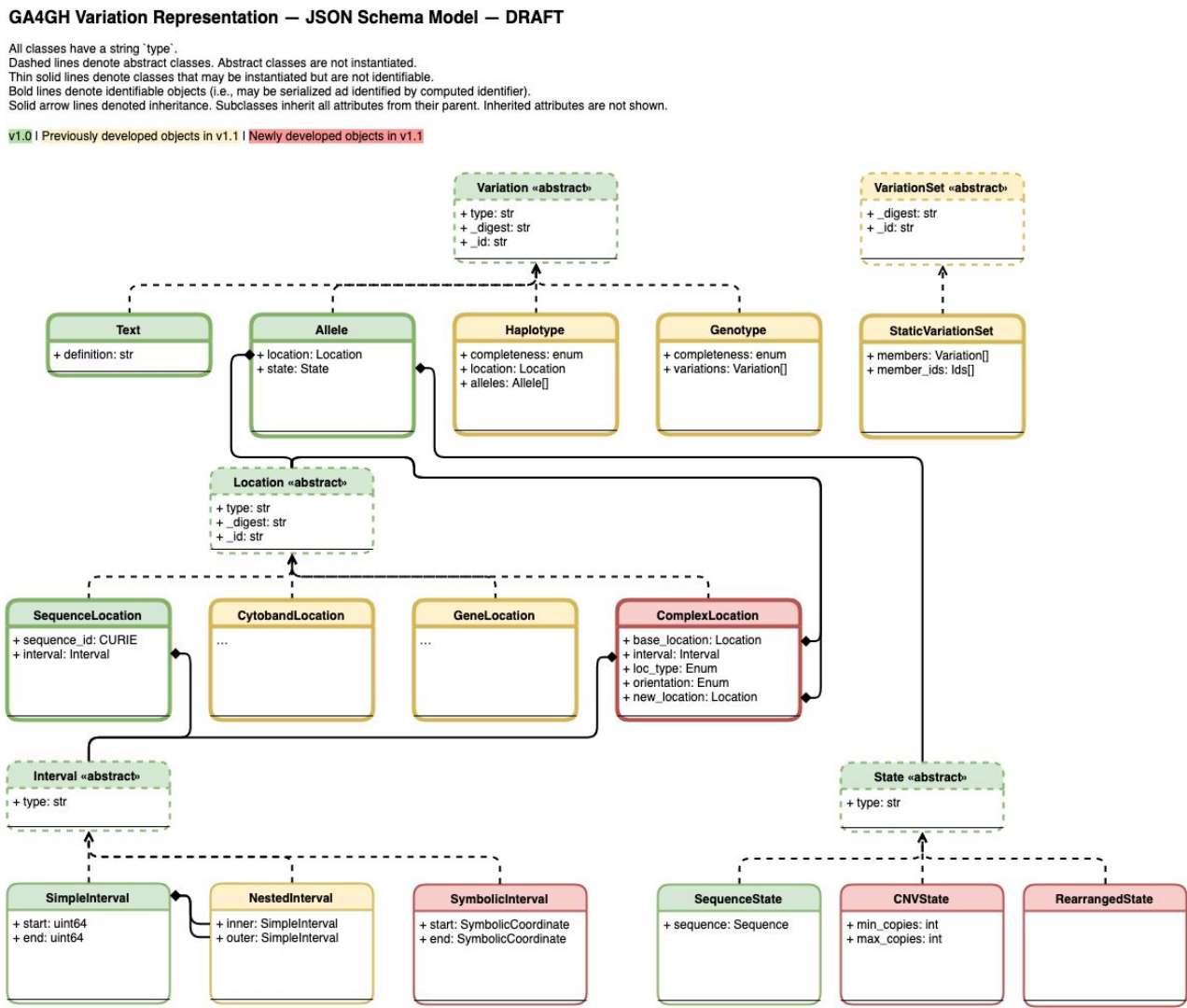
VR identifiers: precise, unique names for variation



@reecehart & @handlerwagner

Upcoming milestone goals:

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



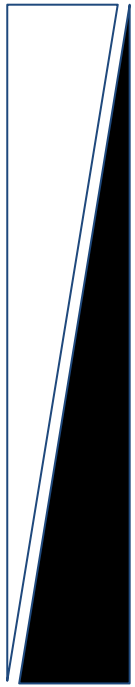
@handlerwagner

Fusion Specificity Needs Vary

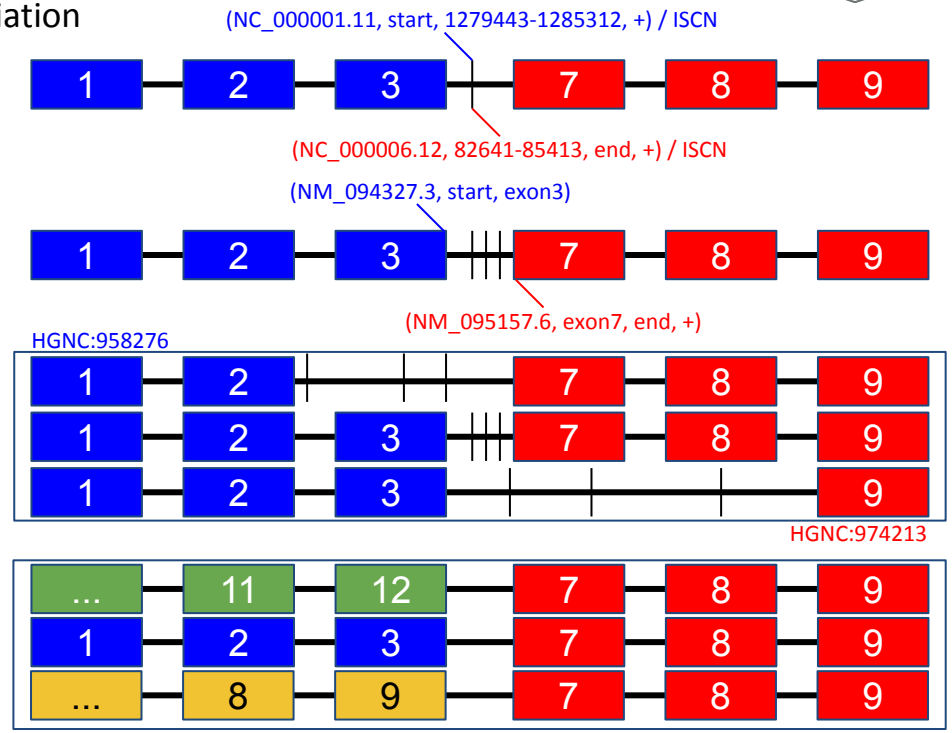
DNA Break Coordinates Specified

Exon-exon Junction(s) Specified

Gene(s) Specified



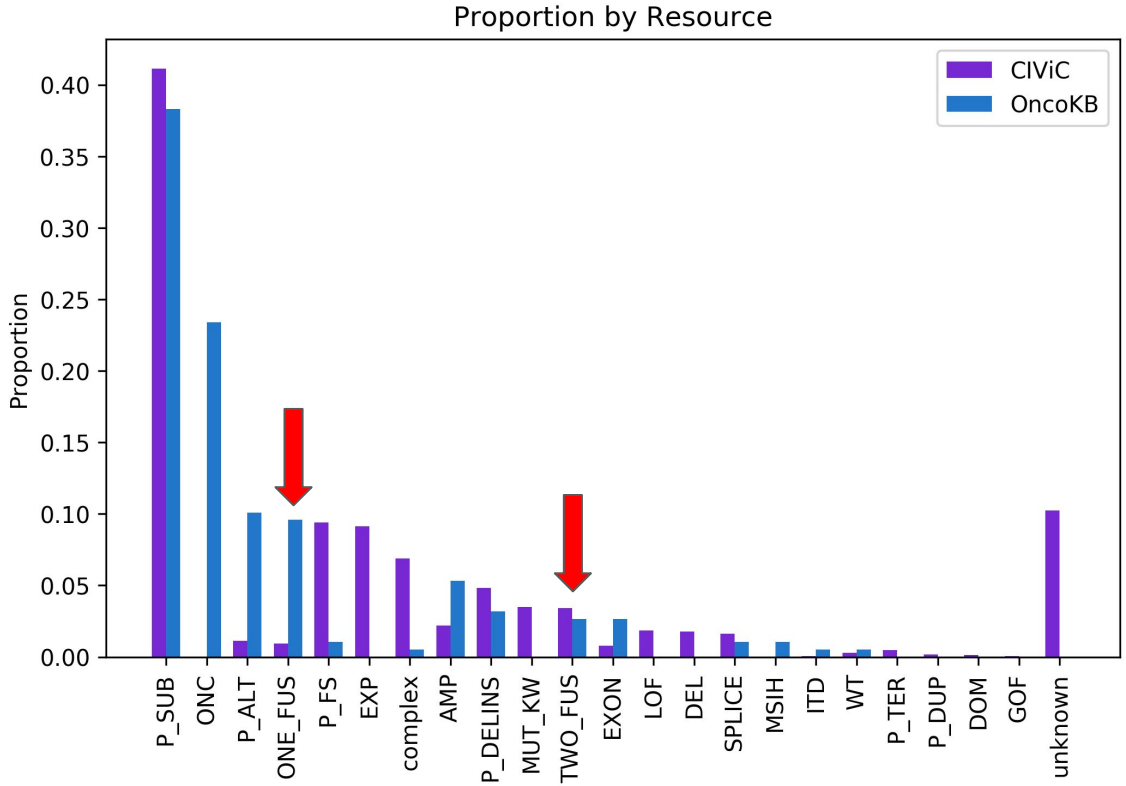
Discrete Variation



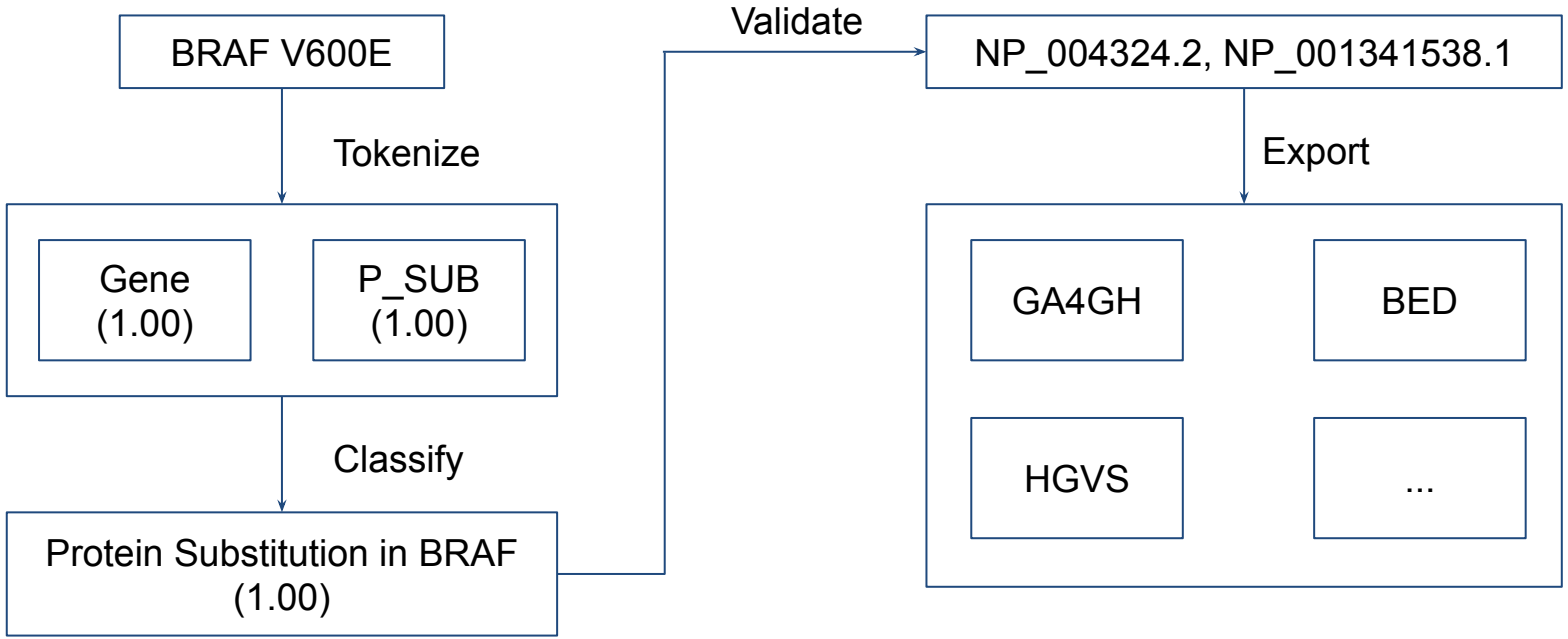
Categorical Variation

@handlerwagner

Fusions in the Somatic Variant Landscape



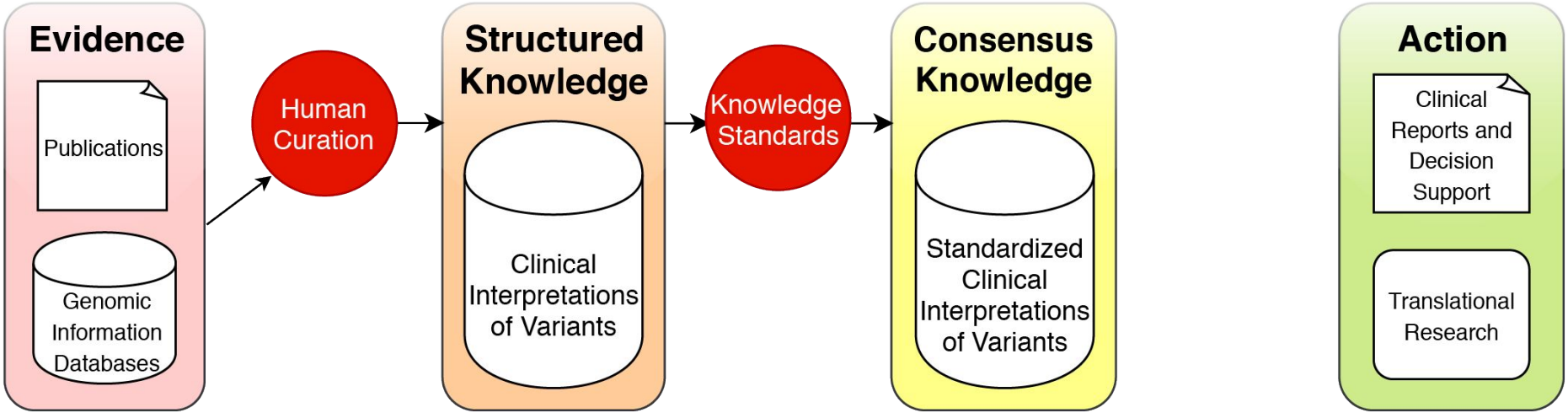
Lexing and Classification of Categorical Variants



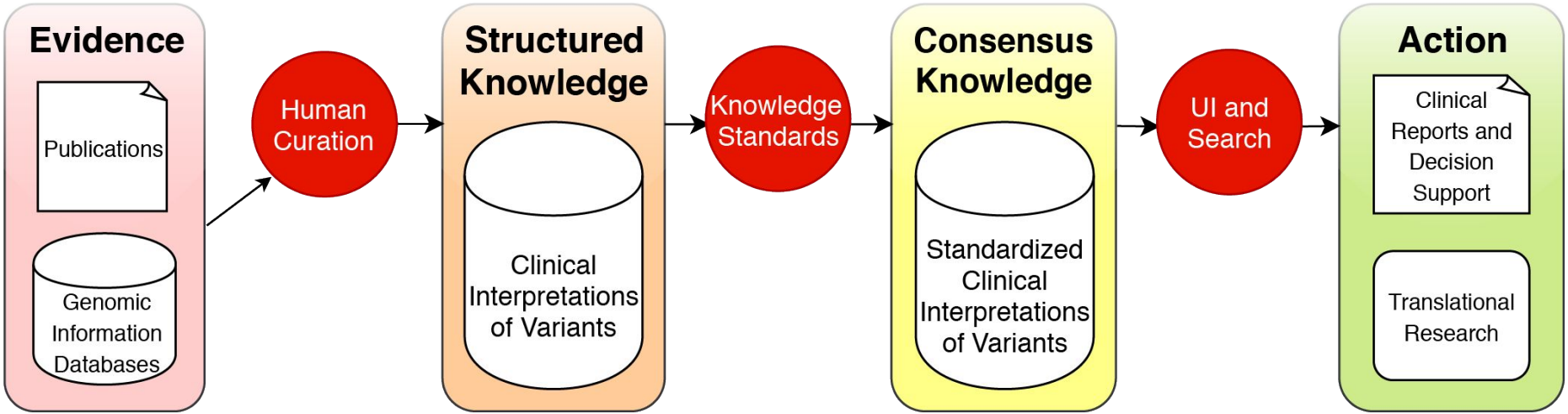
Improving Normalization of Drugs and Diseases

- 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

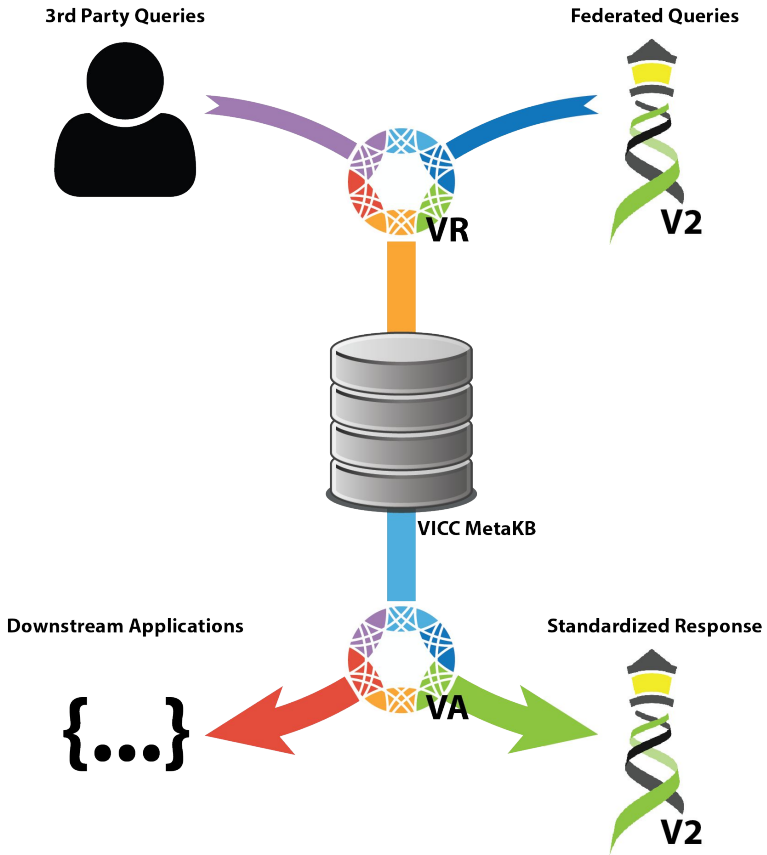
Public Search Tool

Documentation:
docs.cancervariants.org

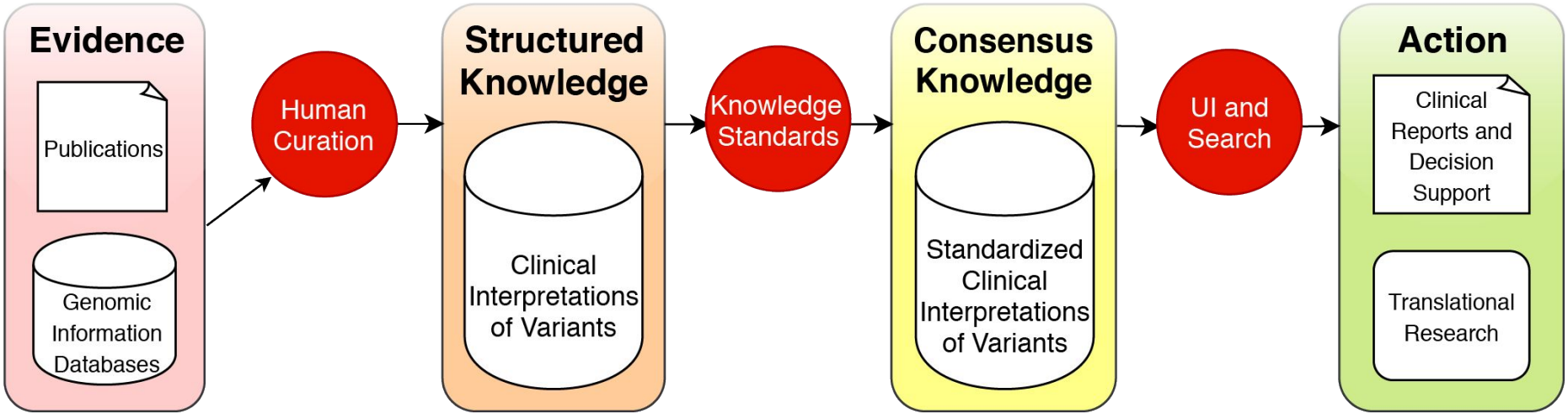
Web Application:
search.cancervariants.org

@handlerwagner

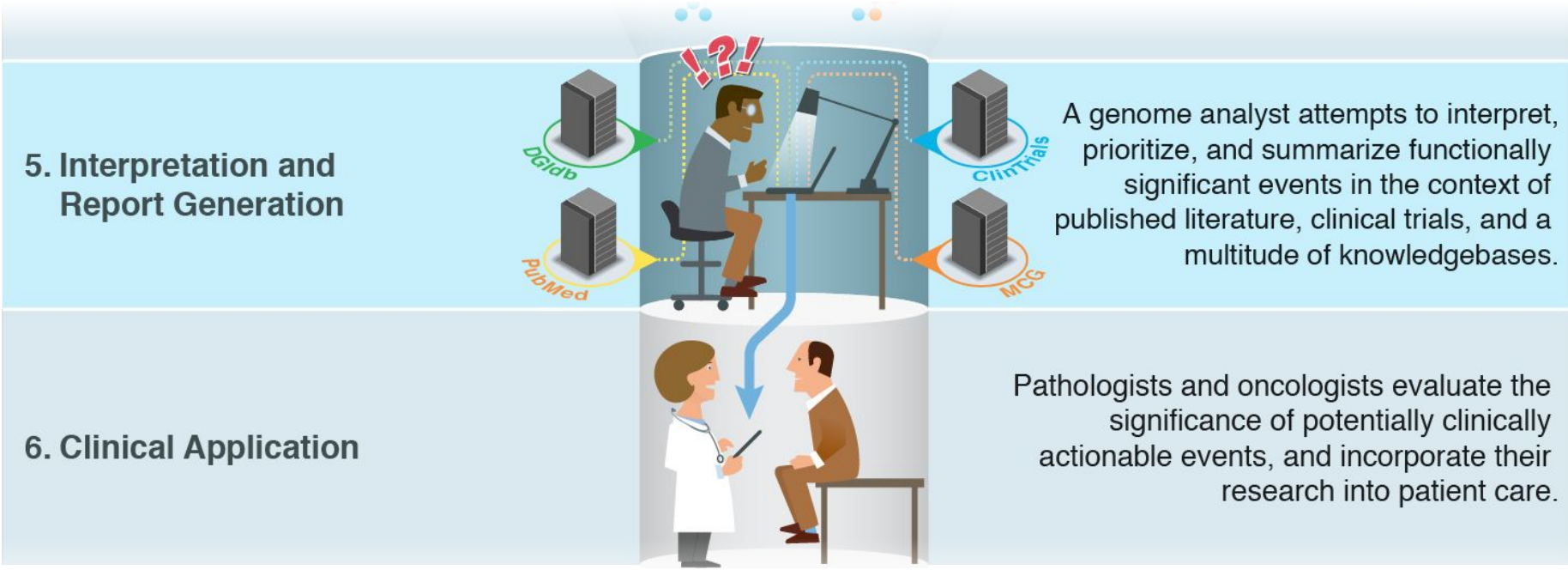
VR, VA, and Beacon v2



Roadmap from Evidence to Action

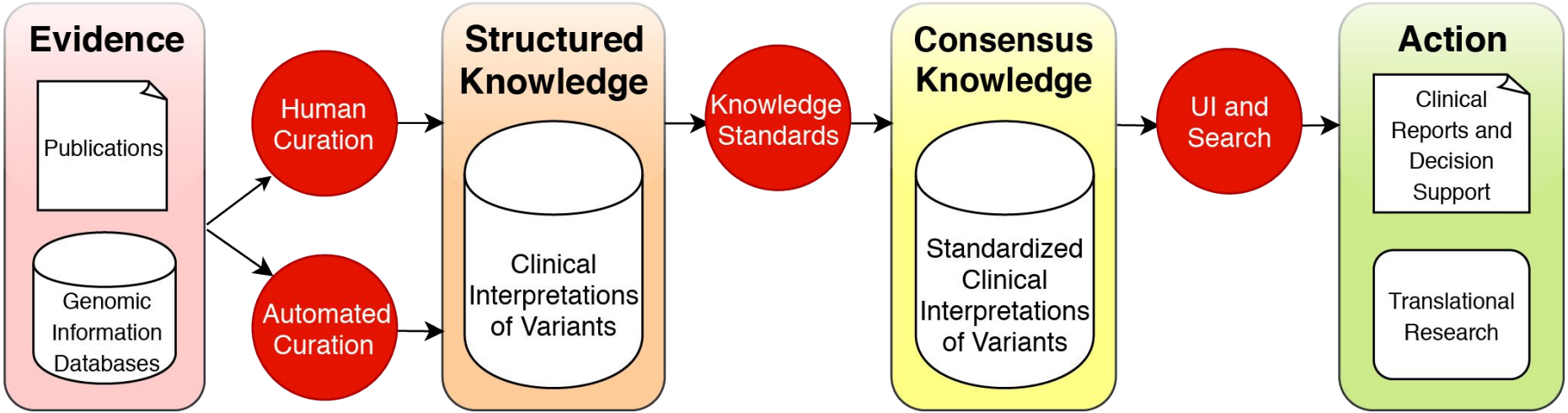


Precision Oncology Bottleneck



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

Roadmap from Evidence to Action



@handlerwagner

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:

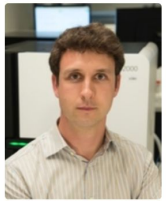
Specification: https://vr-spec.readthedocs.org

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

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

GA4GH GKS Work Stream

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

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, Jianjiong 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 Micheel, 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 Zaranek, Zhenyu Zhang.

F32 CA206247

NIH NATIONAL CANCER INSTITUTE

K99 HG010157

NIH National Human Genome Research Institute

Thank You

Questions?

Email: awagner24@wustl.edu

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

Resources

CIViC: civicdb.org

CIViCpy: civicpy.org

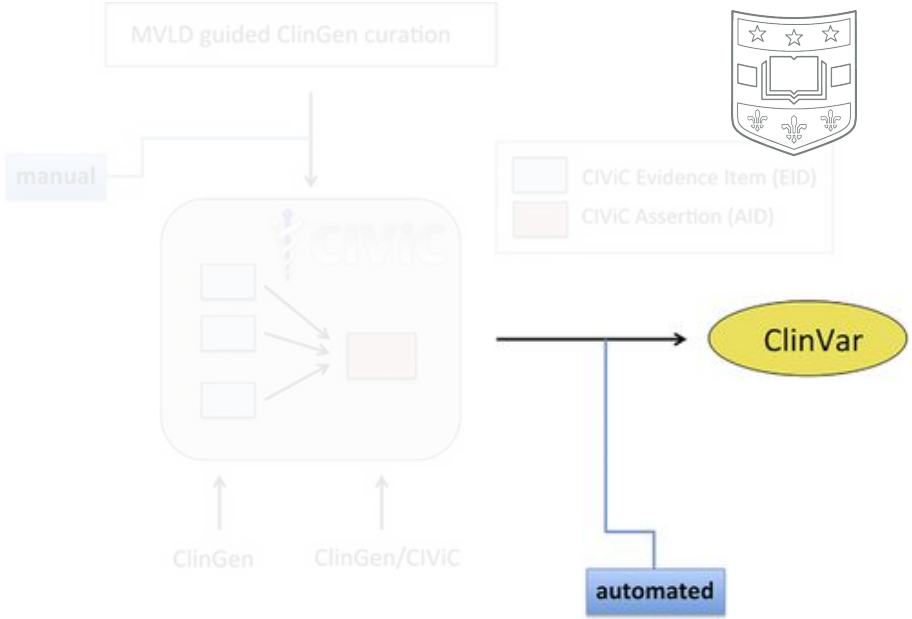
VICC: cancervariants.org

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

Remaining Content

ClinGen Somatic WG: MVLD Guidelines

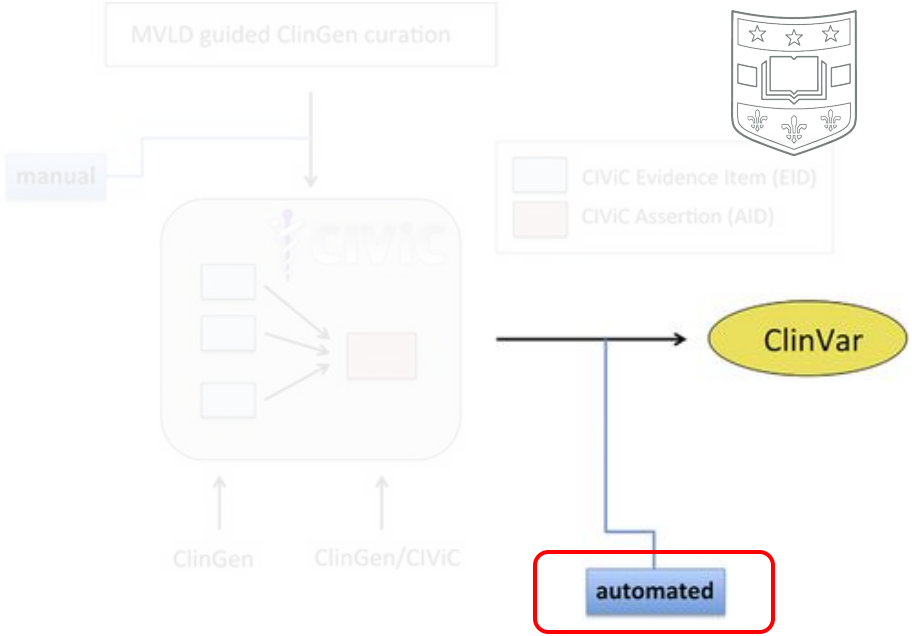
Danos AM, et al. *Hum. Mut.* 2018



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

ClinGen Somatic WG: MVLD Guidelines

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

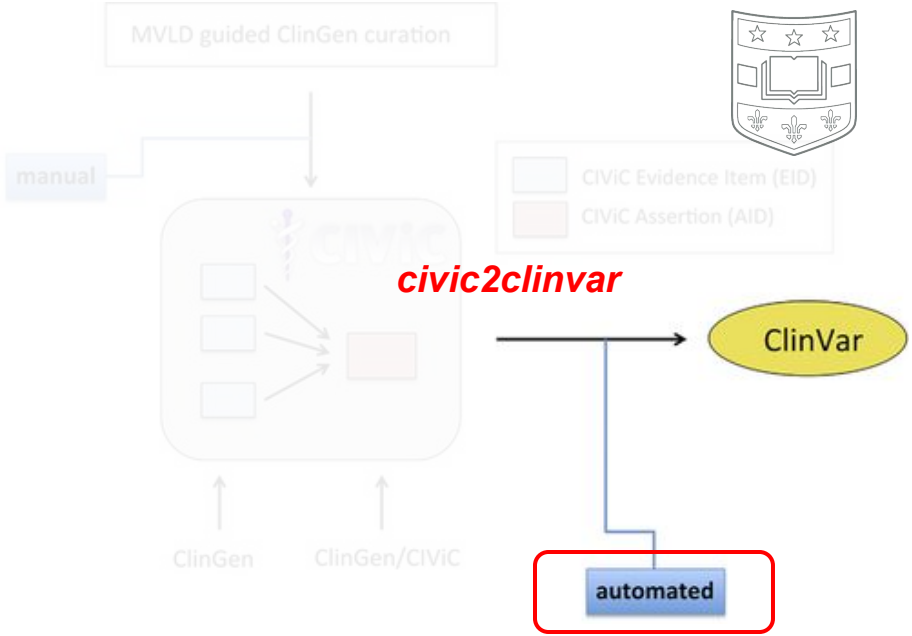


Danos AM, et al. *Hum. Mut.* 2018

@handlerwagner

ClinGen Somatic WG: MVLD Guidelines

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



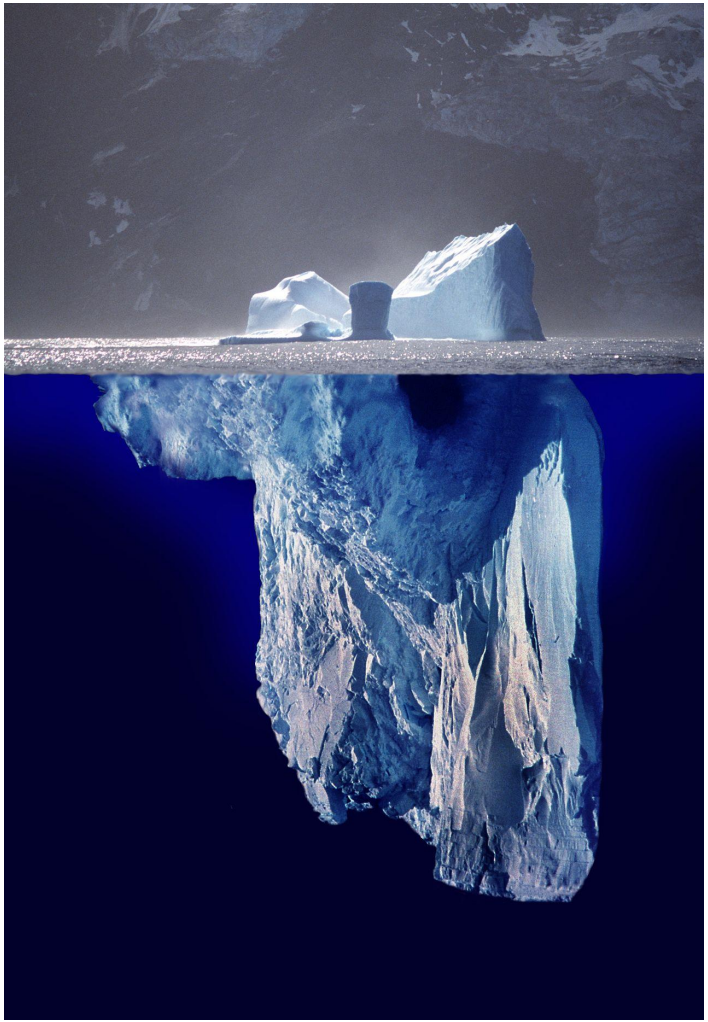
Danos AM, et al. *Hum. Mut.* 2018

@handlerwagner

civic2clinvar

https://commons.wikimedia.org/wiki/File:Iceberg.jpg

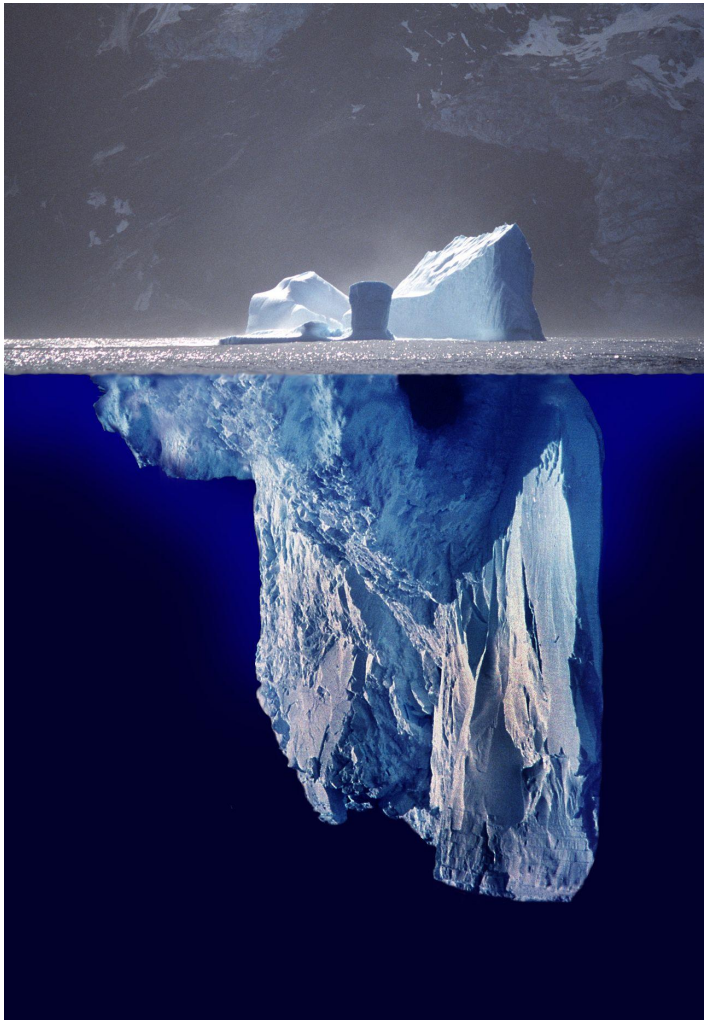
@handlerwagner



https://commons.wikimedia.org/wiki/File:Iceberg.jpg

civic2clinvar

@handlerwagner



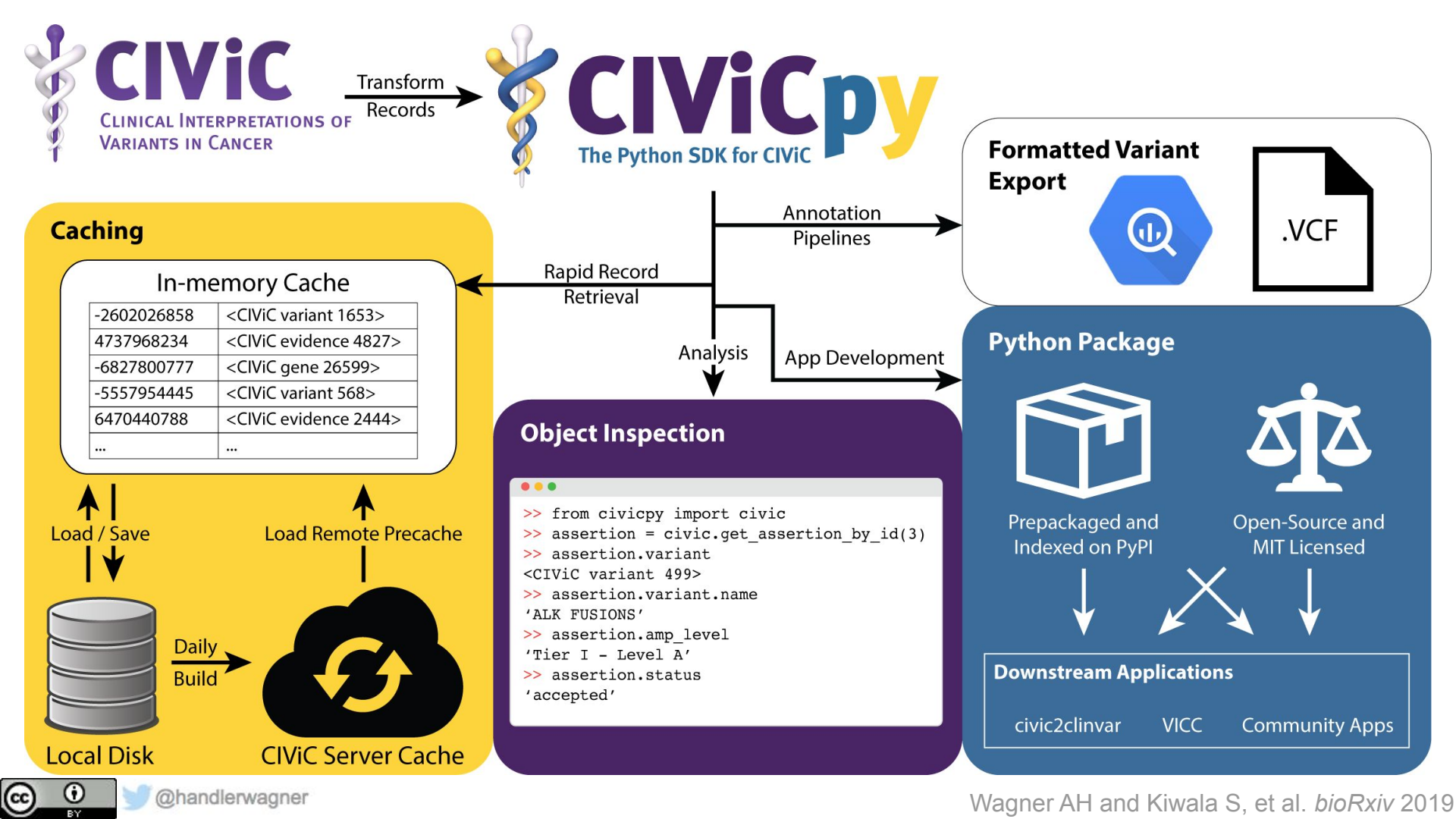
https://commons.wikimedia.org/wiki/File:Iceberg.jpg

civic2clinvar

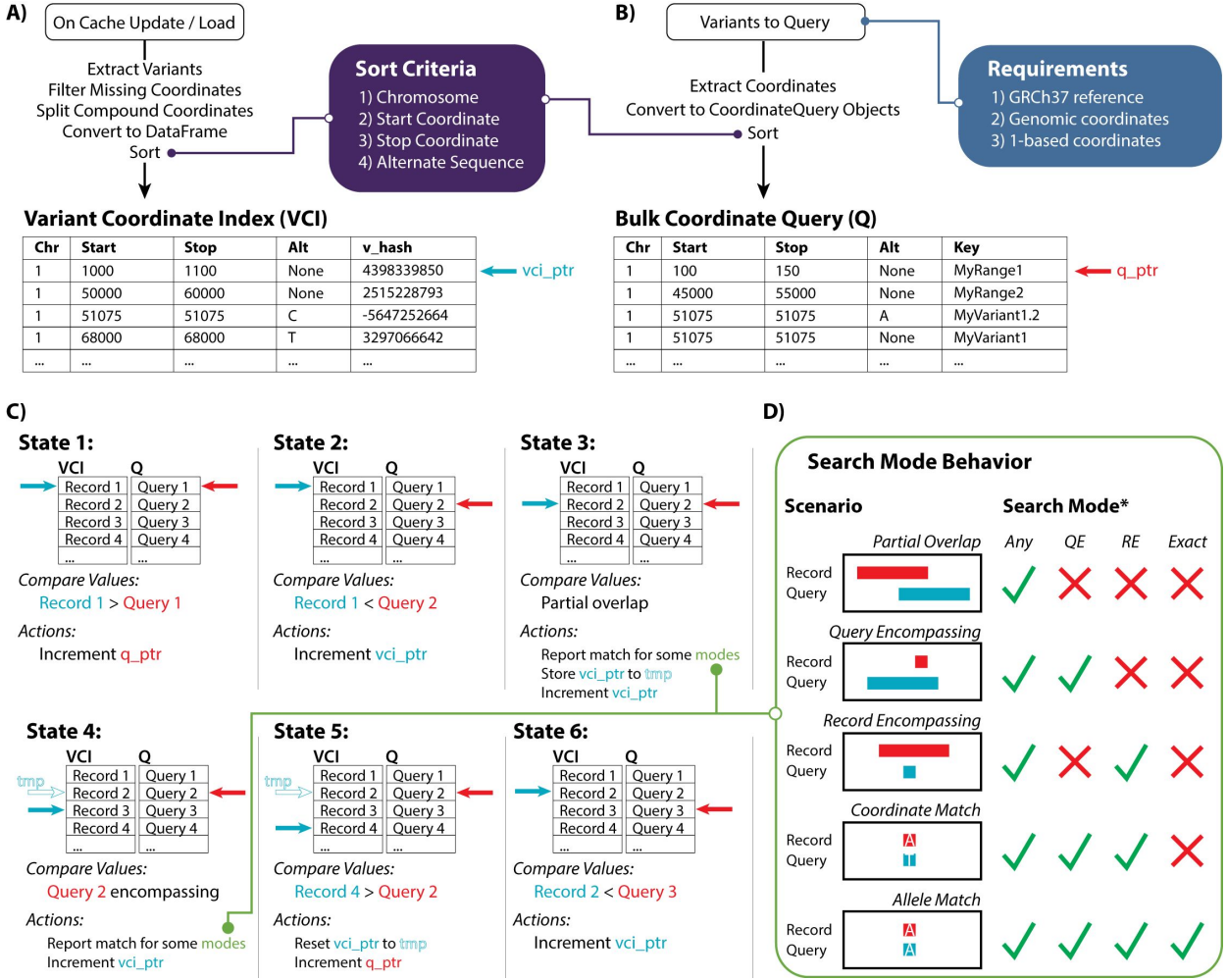
CIViCpy

The Python SDK for CIViC

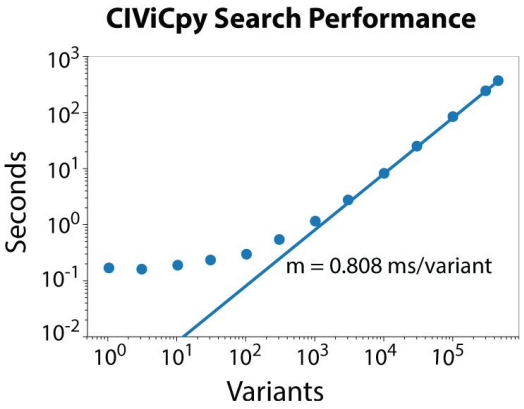
@handlerwagner



Wagner AH and Kiwala S, et al. *bioRxiv* 2019

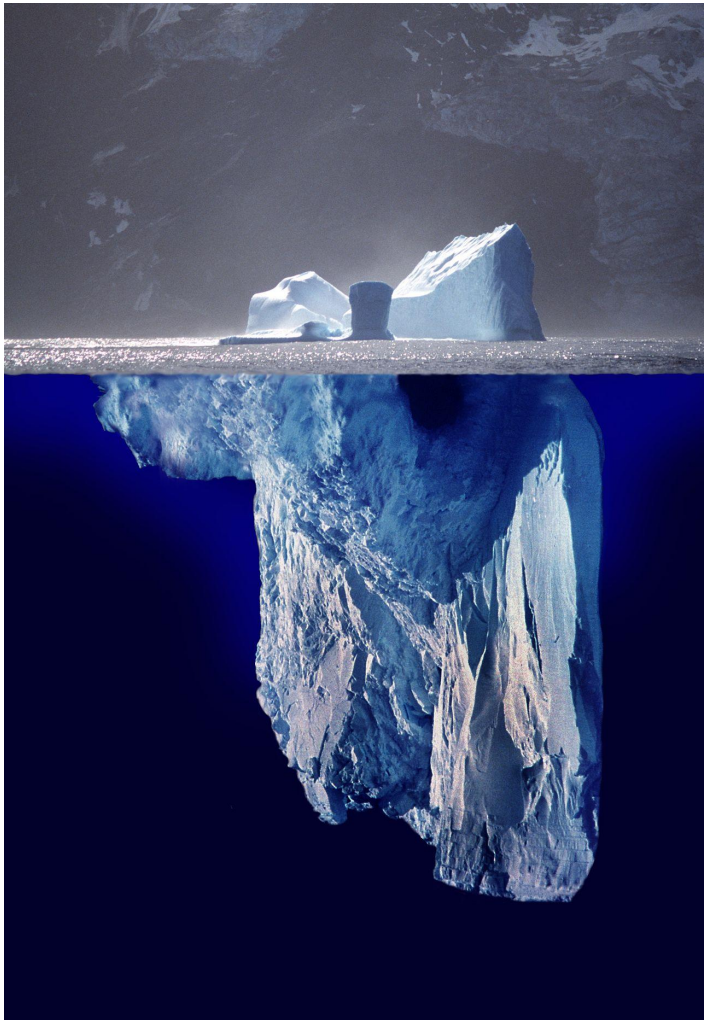


Searches **>1,200** variants/second



@handlerwagner

Wagner AH and Kiwala S, et al. *bioRxiv* 2019

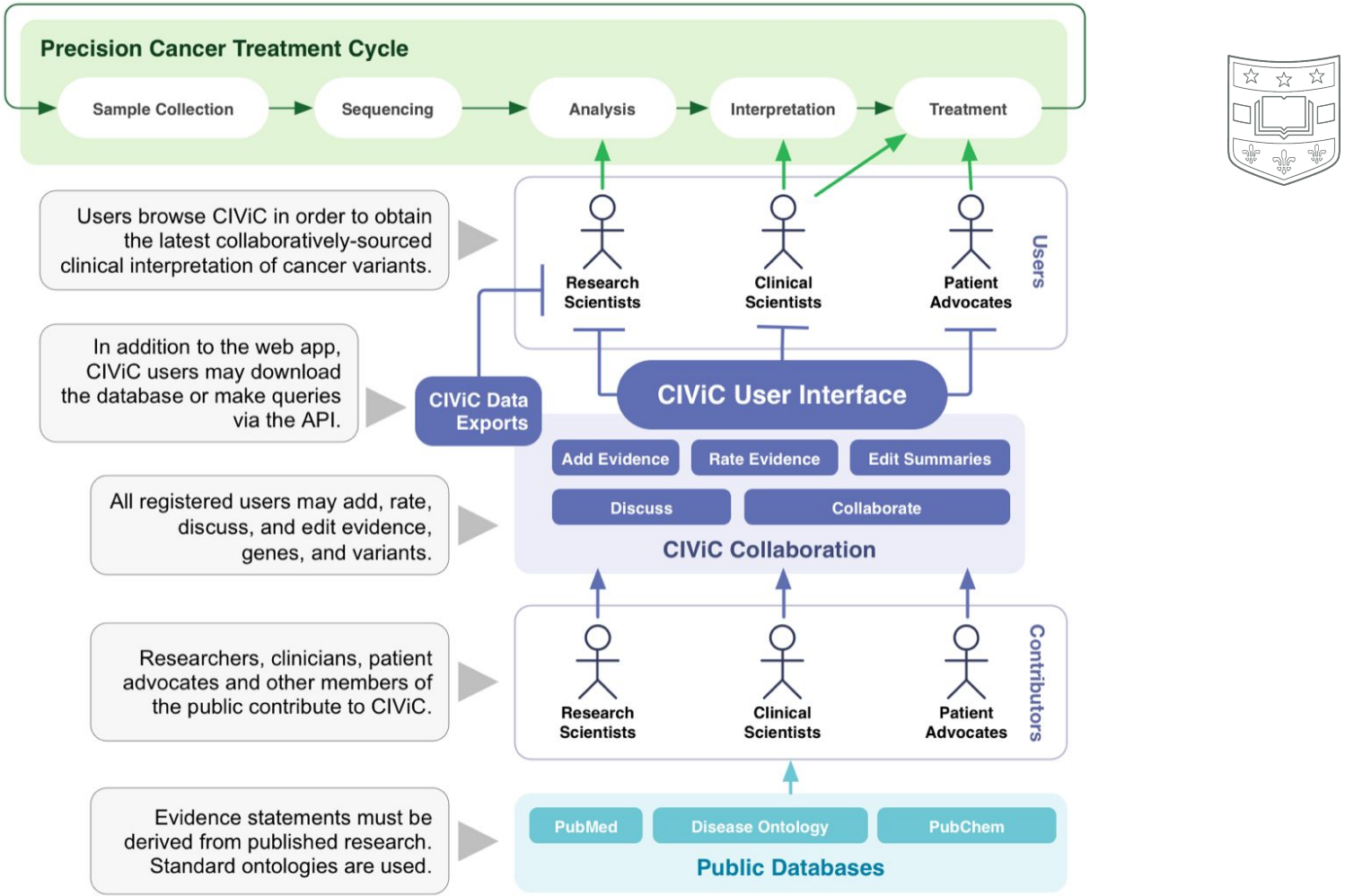


https://commons.wikimedia.org/wiki/File:Iceberg.jpg

CIViCpy

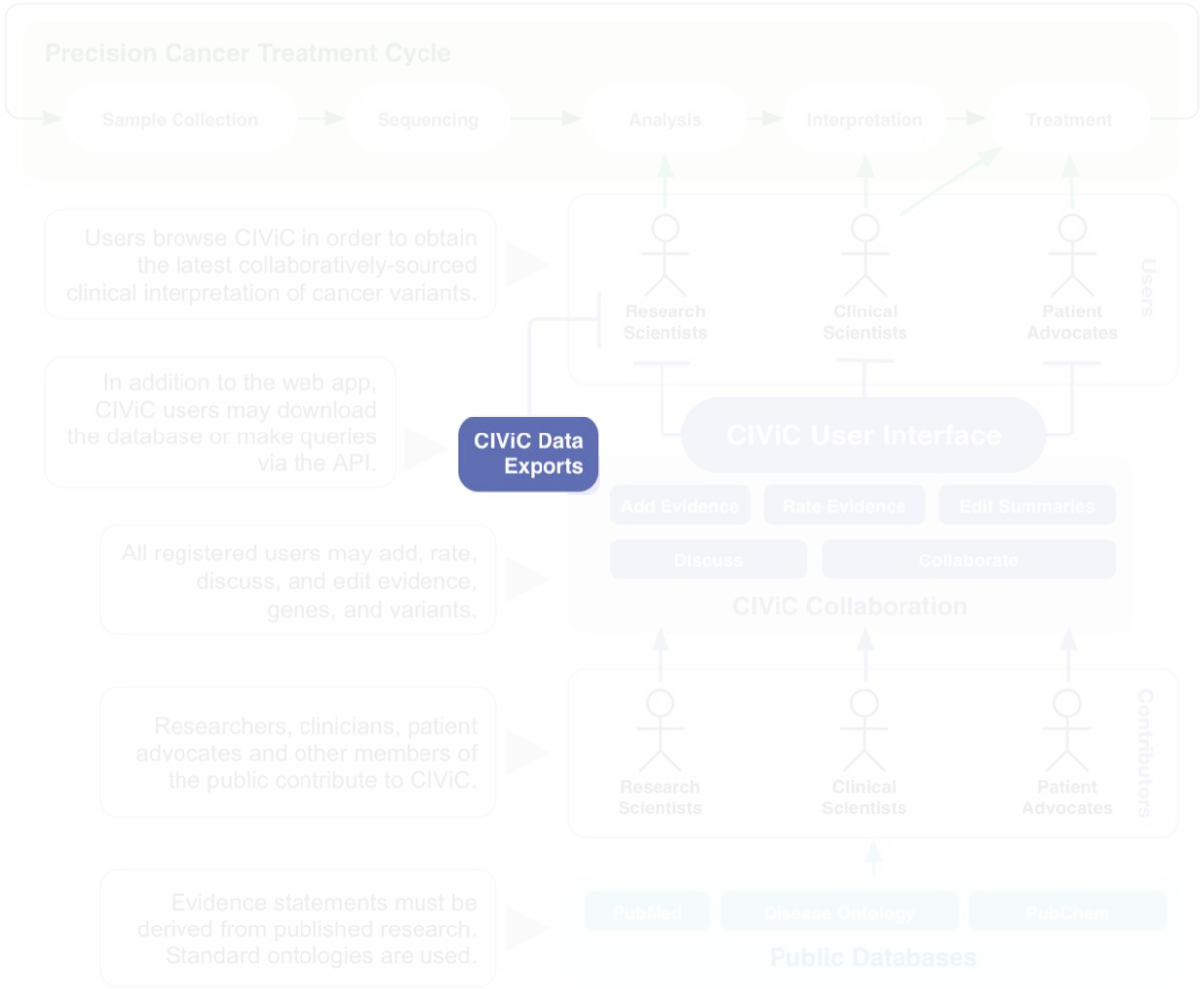
The Python SDK for CIViC

civicpy.org



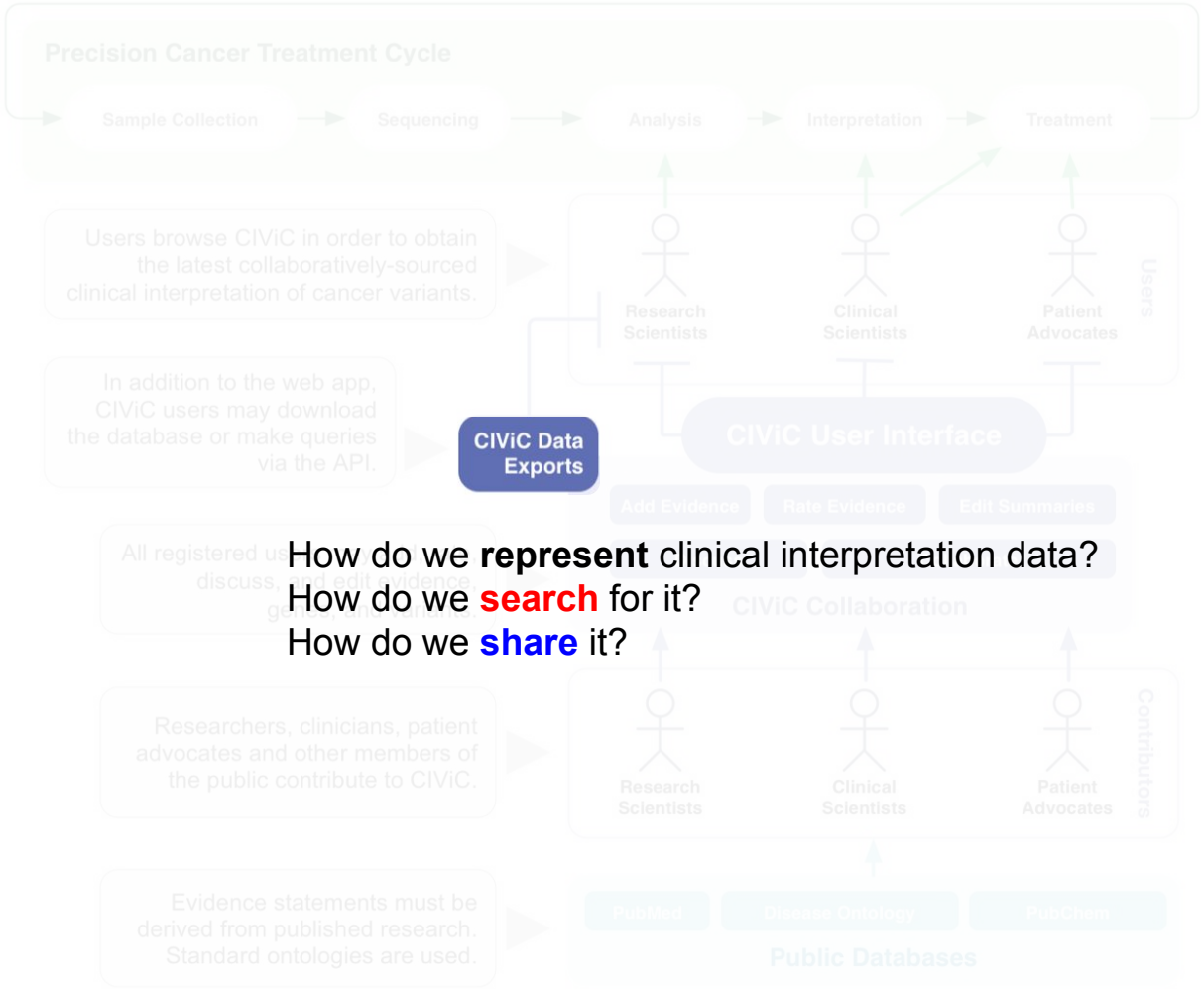
@handlerwagner

https://civicdb.org/graphics



@handlerwagner

https://civicdb.org/graphics

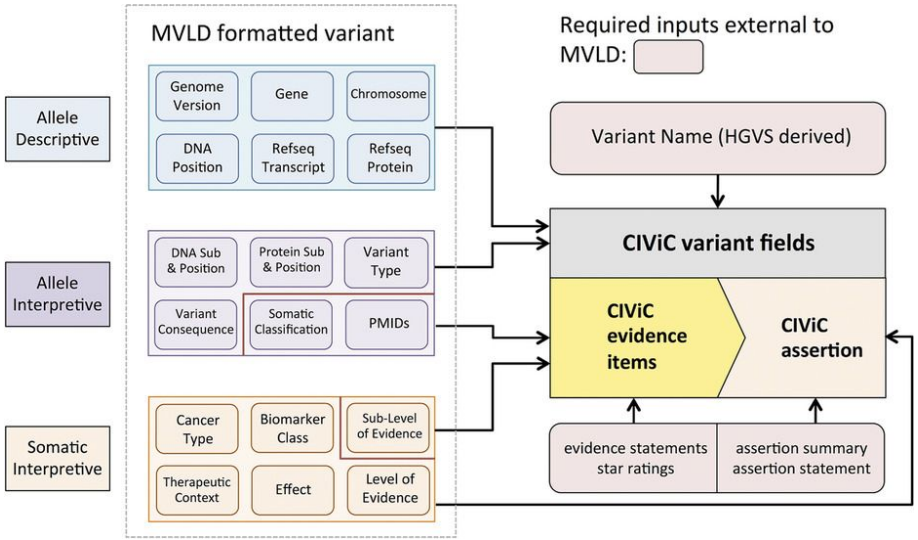


How do we **represent** clinical interpretation data?
How do we **search** for it?
How do we **share** it?

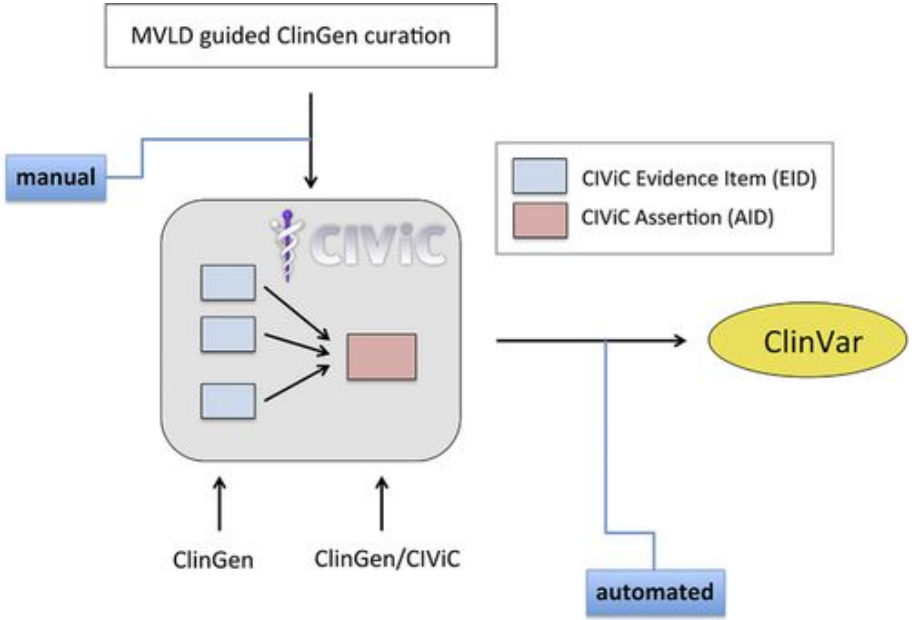
@handlerwagner

https://civicdb.org/graphics

ClinGen Somatic WG: MVLD Guidelines



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



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