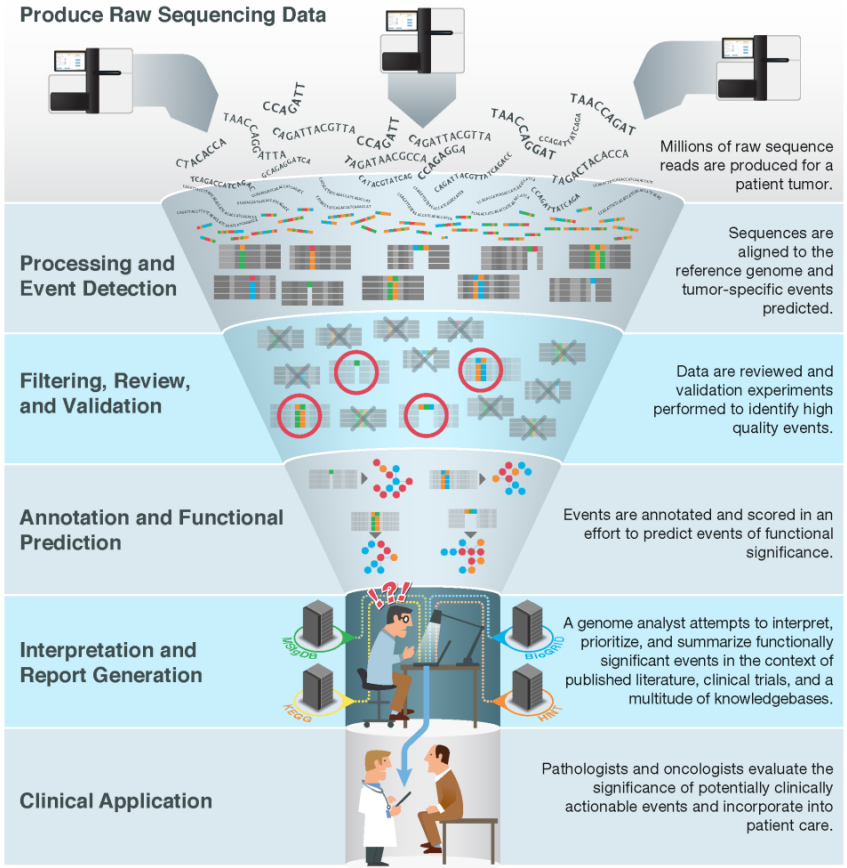
The Elizabeth H.
and James S. McDonnell III
McDonnell Genome Institute
at Washington University

Development of Informatics Resources for Interpretation of Clinically Actionable Variants in Cancer

ITCR 2017 Annual Meeting
Santa Cruz, CA
Wednesday, May 31, 2017

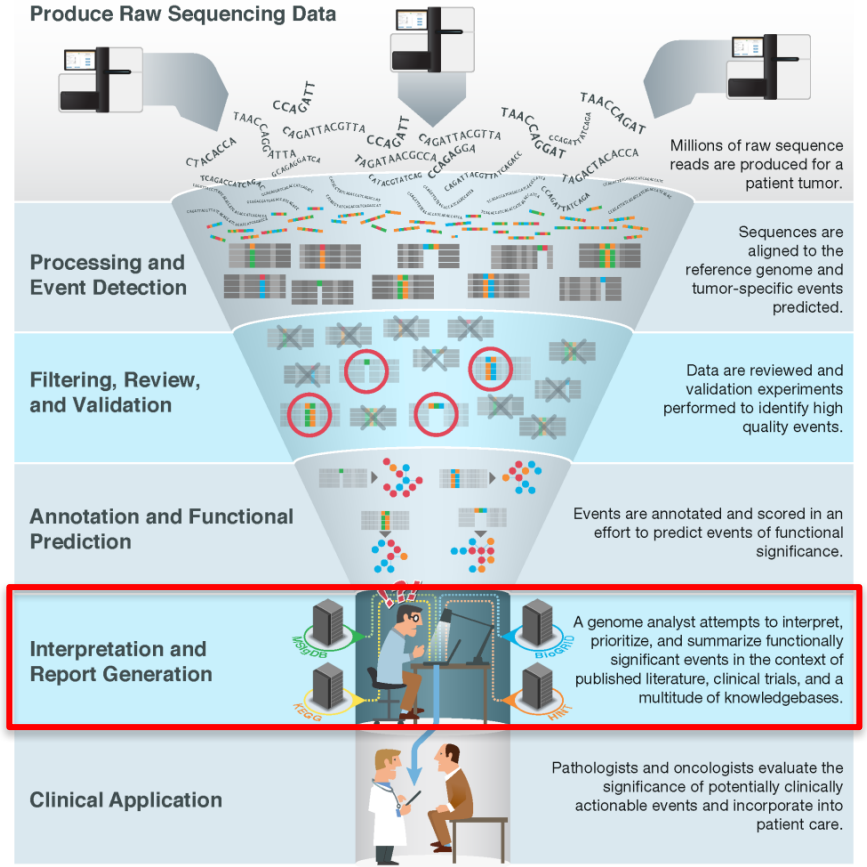
Obi Griffith <obigriffith@wustl.edu>

Background: High-throughput sequencing has been largely automated allowing rapid identification of somatic and germline variants in tumors



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

Problem: Clinical interpretation of genomic alterations remains a major bottleneck for realizing precision medicine



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

Clinical interpretations of variants are currently created in private academic silos or restricted-access commercial databases

Genomic Alterations

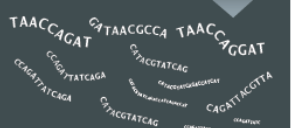
Gene Alteration	Interpretation
PIK3CA H1047R	Mutations in PIK3CA have been reported in 26% to 33% of breast cancer cases (COSMIC, Jun 2012 and Kalinsky et al., 2009; 19671852). Activating mutations in PIK3CA, such as the one seen here, may predict sensitivity to inhibitors of PI3 kinase or its downstream signaling pathway (the PI3K/Akt/mTOR pathway) (Huang et al., 2007; 18079394). The mTOR inhibitors temsirolimus and everolimus have been tested in several clinical trials in breast cancer, and have been approved by the FDA for use in other tumor types. Inhibitors of PI3K and Akt are currently in clinical trials in breast cancer, alone or in combination with other therapies. PIK3CA mutations may play a role in resistance to hormonal therapy in ER+ breast cancers (Miller et al., 2011; 22114931). Activating mutations in PIK3CA may also confer resistance to anti-Her2 therapies (Chakrabarty et al., 2010; 20581867, Kataoka et al., 2010; 19633047, Wang et al., 2011; 21676217); combined inhibition of Her2 and the PI3K pathway may be required in tumors with ERBB2 amplification and PIK3CA mutation, though this remains an area of active investigation.
CCND1 amplification	CCND1 amplification has been reported in approximately 10-15% of invasive breast cancers, more frequently in BRCA-negative cancers (Elsheikh et al., 2008; 17653856, Bane et al., 2011; 21327470). There are no approved therapies that directly target the protein product of CCND1 (Cyclin D1); however, CCND1 amplification may predict sensitivity to inhibitors of Cdk4 and Cdk6, which are currently under investigation in clinical trials. Overexpression of Cyclin D1 has also been associated with resistance to endocrine therapy in breast cancer (reviewed in Lange et al., 2011; 21613412; Musgrove and Sutherland, 2009; 19701242, Butt et al., 2005; 16113099).
CDH1 E167*	CDH1 mutations are present in approximately 17% of breast cancers, and more often in luminal type cancers (COSMIC, Jun 2012, Hollestelle et al., 2010; 19593635). Loss of the E-cadherin protein, which is encoded by the CDH1 gene, has been associated with poor prognosis in triple negative breast cancer (Kashiwagi et al., 2010; 20551954, Tang et al., 2011; 21519872). Presently, there are no targeted therapies to address loss of CDH1/E-cadherin.

- Interpretations have limited provenance and no mechanism for feedback
- This problem would be better addressed by an open public domain effort

We created CIViC to address this need - an open knowledgebase and curation system for clinical interpretation of variants in cancer

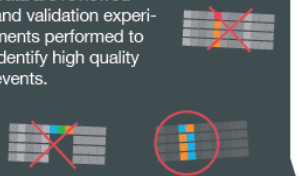
Data and Knowledge Production

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



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

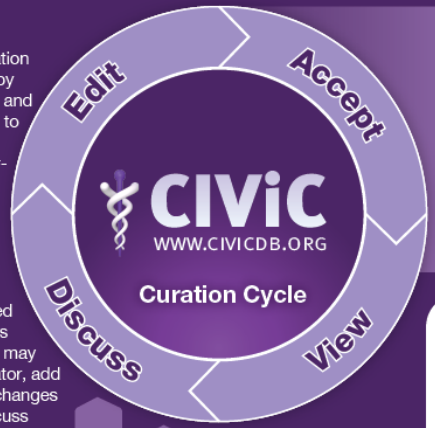
Data are reviewed and validation experiments performed to identify high quality events.



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

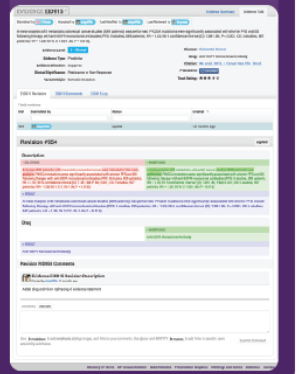
CIViC Curation

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.

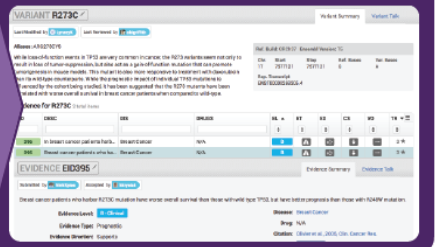


Add New Evidence

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.

www.civicdb.org

See paper published earlier this year for more details

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

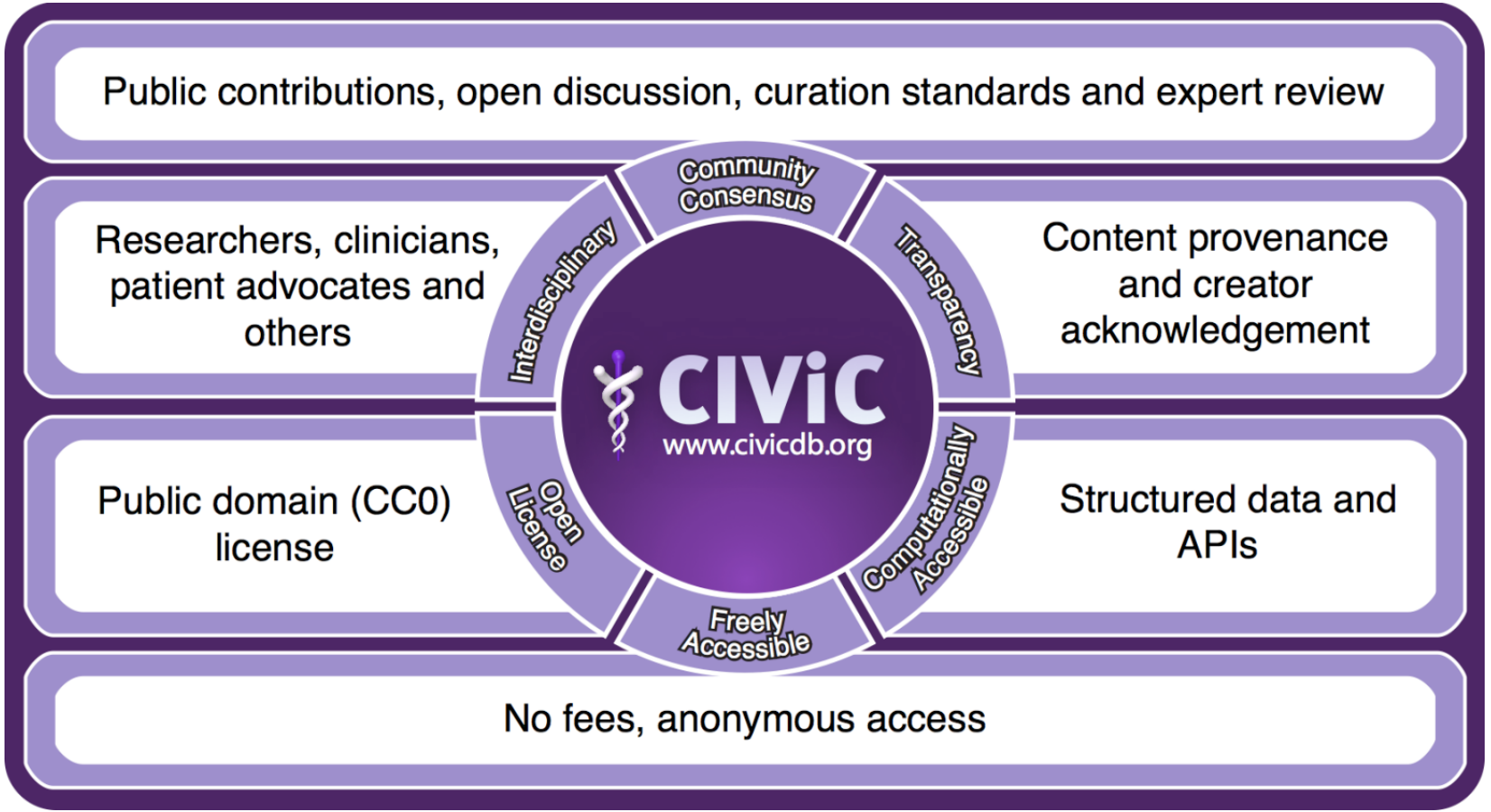
Malachi Griffith, Nicholas C Spies, Kilannin Krysiak, Joshua F McMichael, Adam C Coffman, Arpad M Danos, Benjamin J Ainscough, Cody A Ramirez, Damian T Rieke, Lynzey Kujan, Erica K Barnell, Alex H Wagner, Zachary L Skidmore, Amber Wollam, Connor J Liu, Martin R Jones, Rachel L Bilski, Robert Lesurf, Yan-Yang Feng, Nakul M Shah, Melika Bonakdar, Lee Trani, Matthew Matlock, Avinash Ramu, Katie M Campbell, Gregory C Spies, Aaron P Graubert, Karthik Gangavarapu, James M Eldred, David E Larson, Jason R Walker, Benjamin M Good, Chunlei Wu, Andrew I Su, Rodrigo Dienstmann, Adam A Margolin, David Tamborero, Nuria Lopez-Bigas, Steven J M Jones, Ron Bose, David H Spencer, Lukas D Wartman, Richard K Wilson, Elaine R Mardis & Obi L Griffith – Show fewer authors

Affiliations | Contributions | Corresponding authors

Nature Genetics **49**, 170–174 (2017) | doi:10.1038/ng.3774

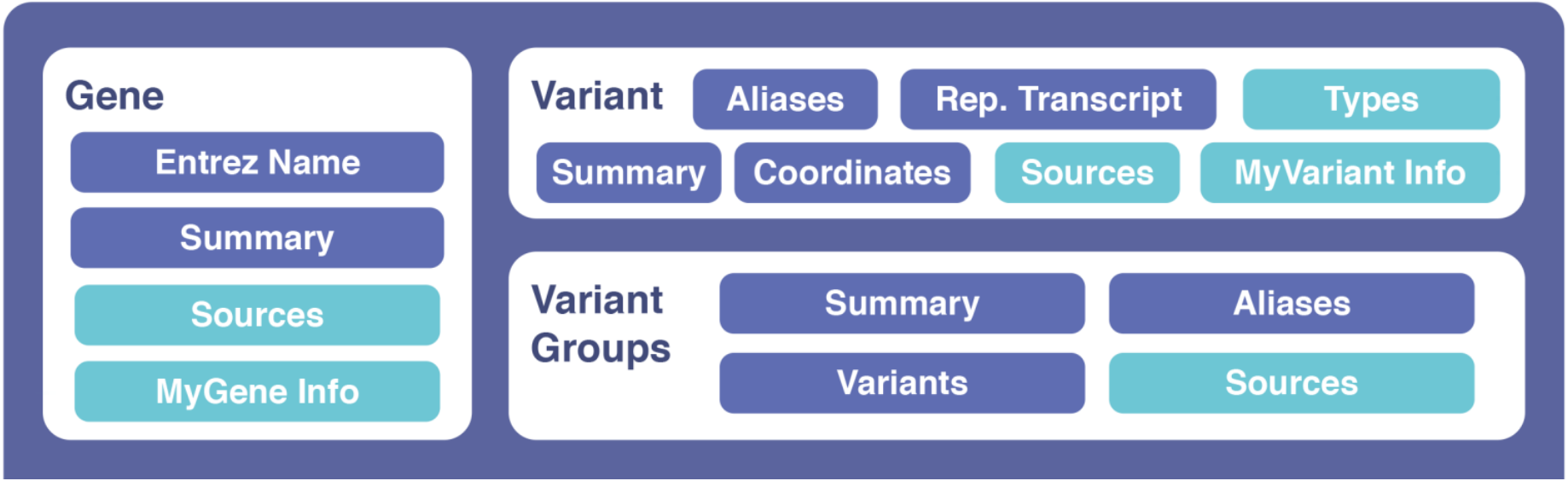
Published online 31 January 2017

CIViC Principles: Total commitment to standards, openness, transparency



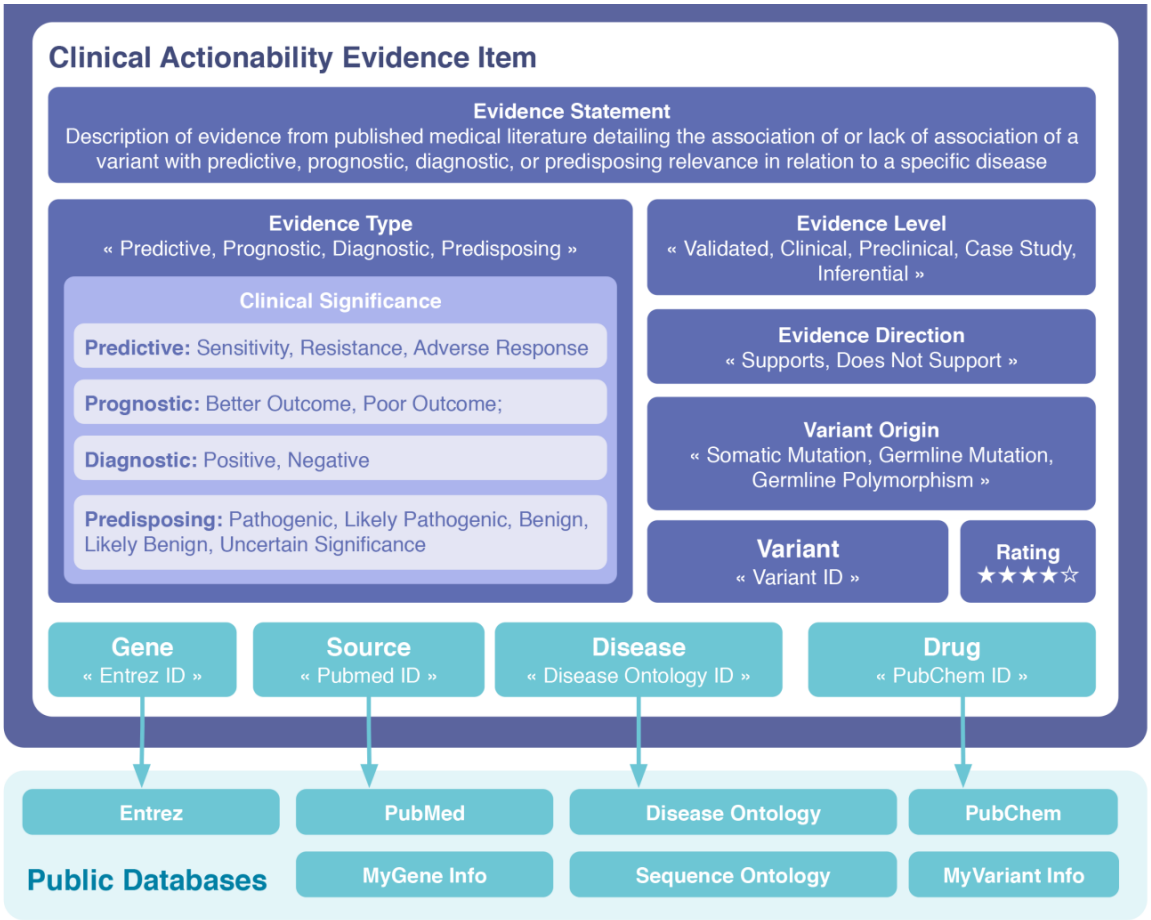
CIViC data model (gene -> variant -> evidence)

BRAF V600E predicts sensitivity to vemurafenib in melanoma



Variant = point mutations, insertions, deletions, translocations, RNA fusions, RNA expression, etc.
Somatic OR Germline

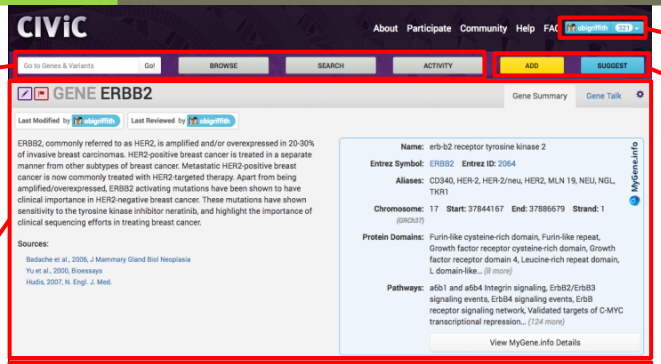
CIViC data model (gene -> variant -> evidence)



Evidence is categorized by:
Type (Drug response, prognostic, etc)
Level (Clinical, pre-clinical, etc)
Direction (supports or refutes)

Advanced search and browsing

Expert-curated knowledge



Optional login
Add/Suggest

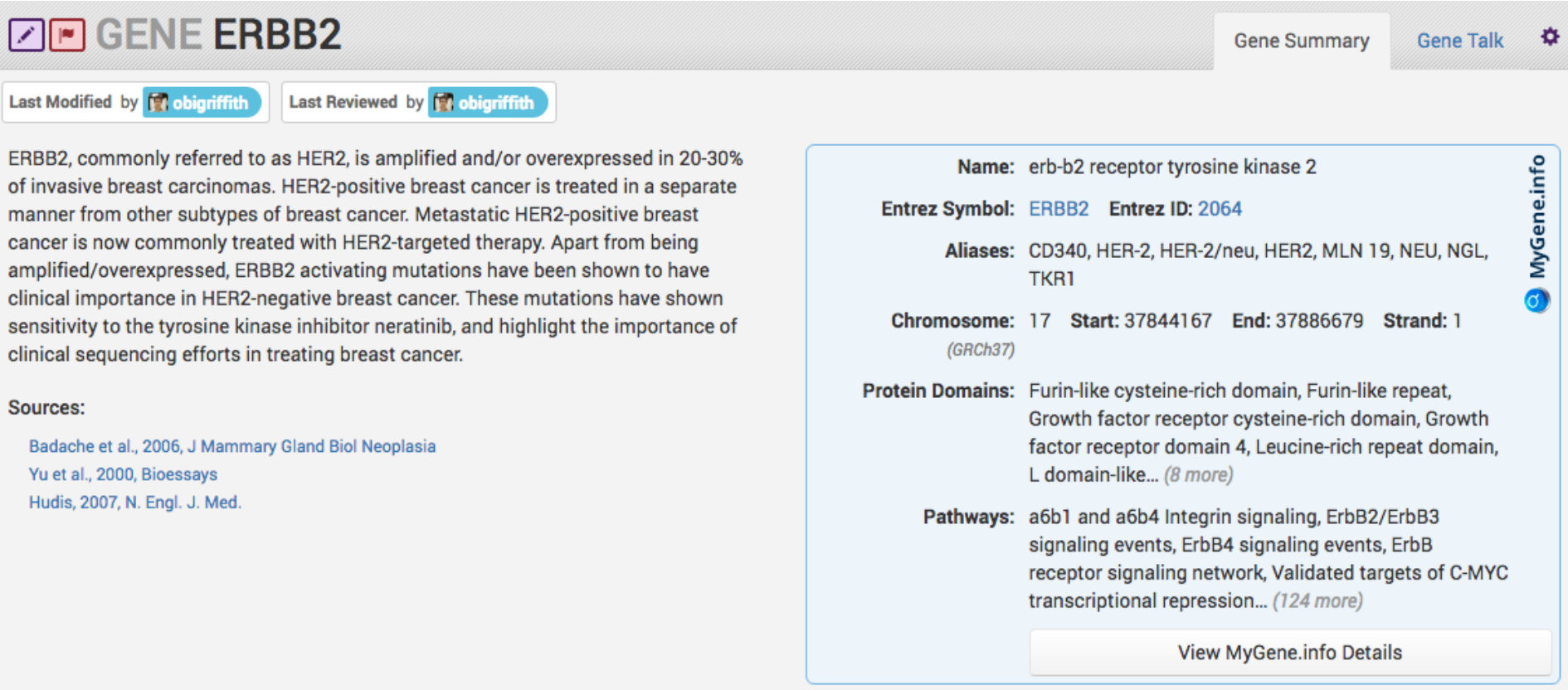
Gene summaries

Variant summaries and coordinates

Structured Evidence

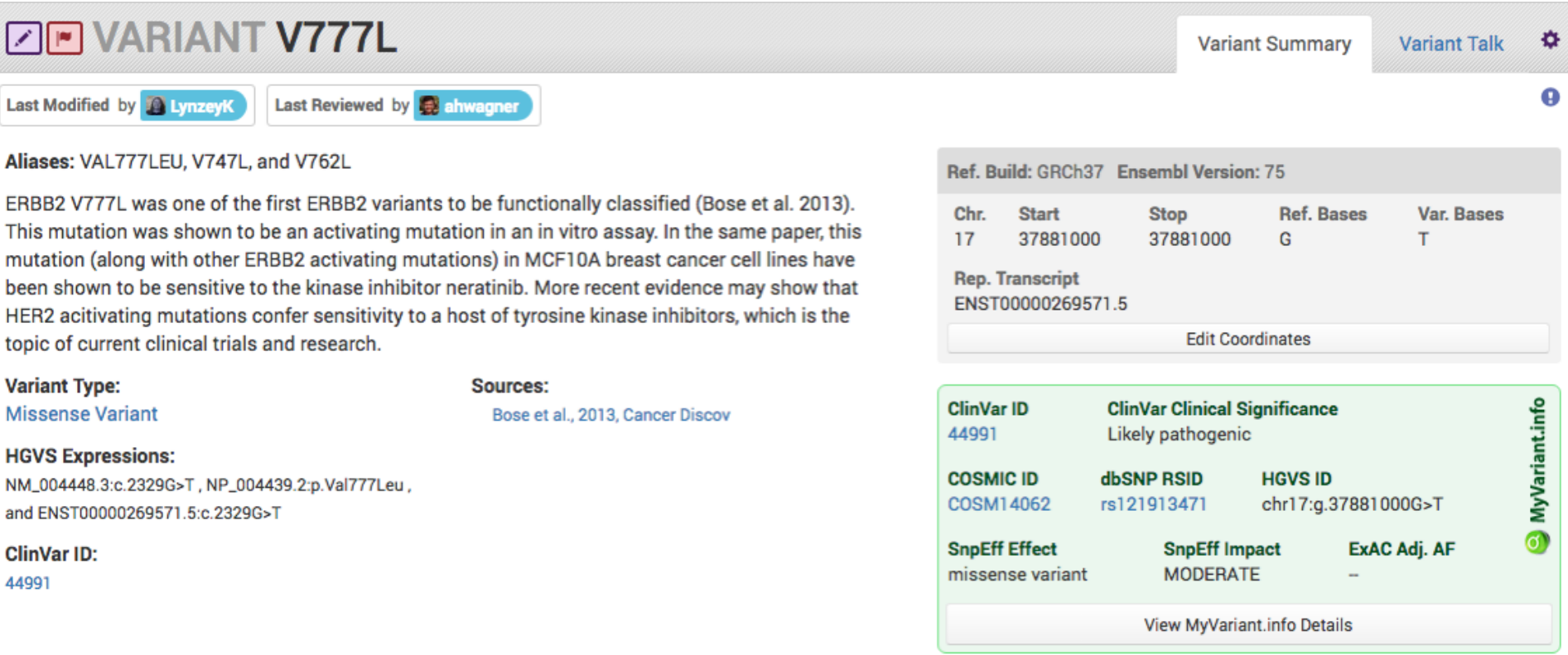
A quick tour of the CIViC site:
www.civicdb.org

Curated gene summaries provide a high-level overview at gene level



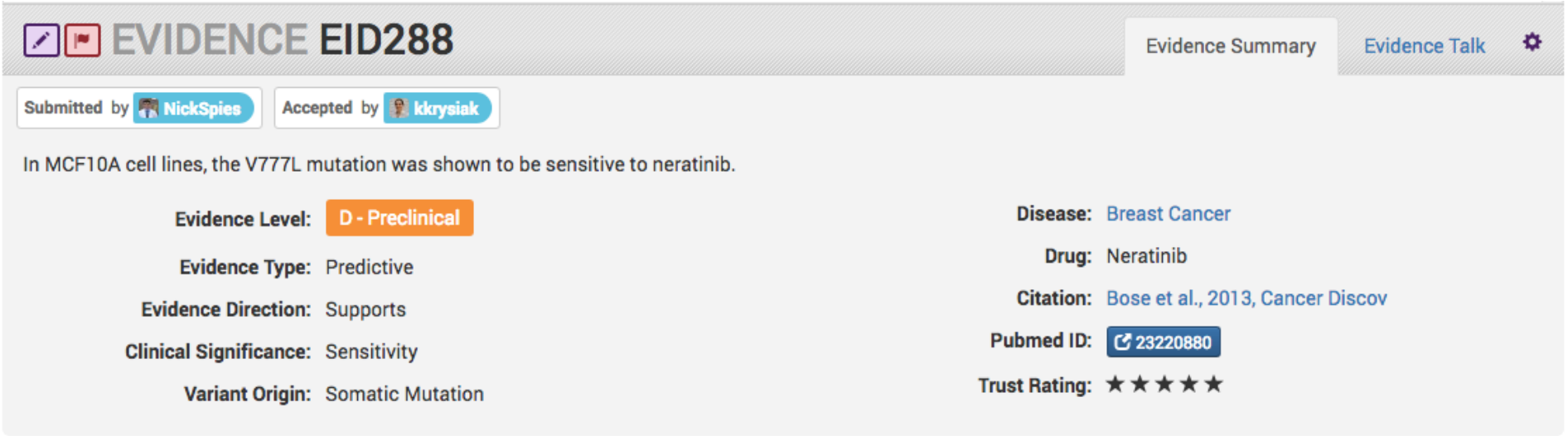
Supplemented with source citations and gene details from mygene.info (Entrez name, symbol, aliases, domains, pathways)

Curated variant summaries provide a high-level overview at variant level



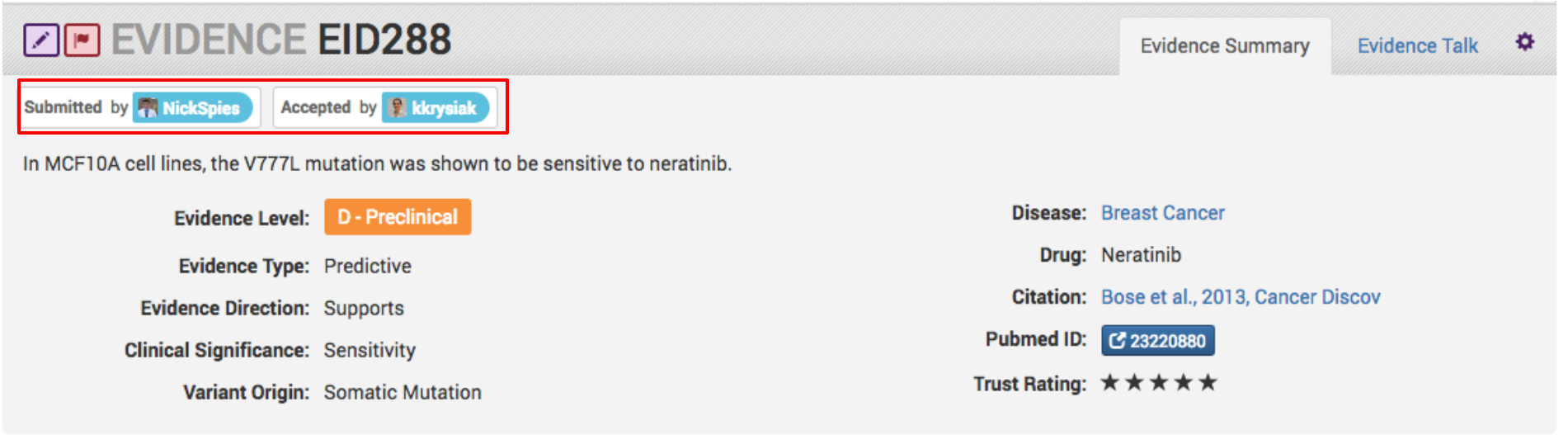
Supplemented with curated aliases, Sequence Ontology terms, sources, HGVS expressions, ClinVar links, coordinates and variant details from myvariant.info

Curated evidence statements are the foundational unit of CIViC

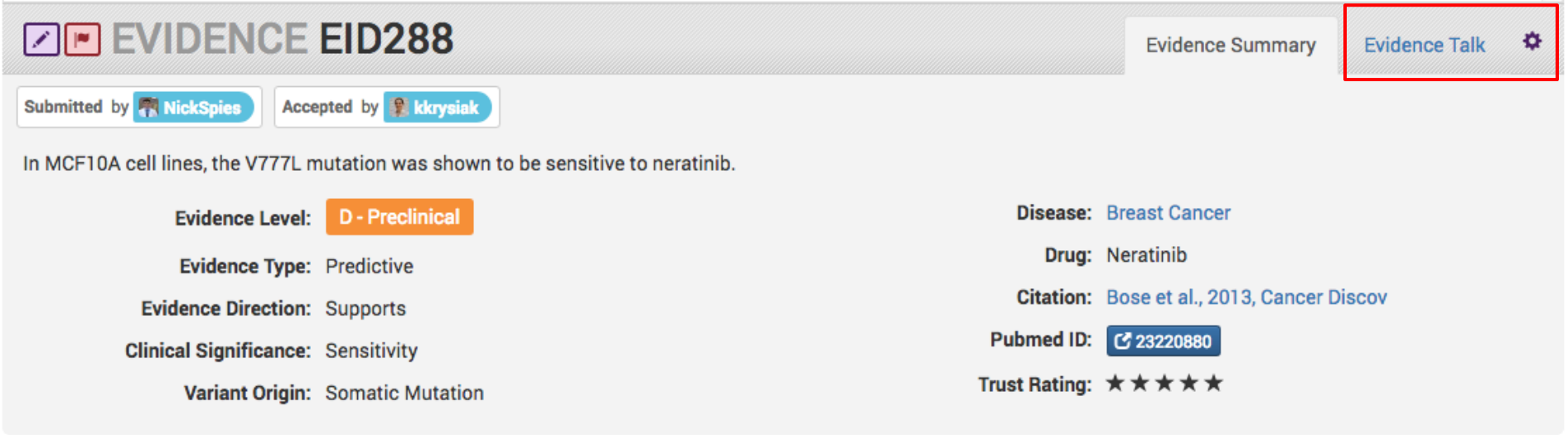


Evidence is curated with structured data (Level, Type, Direction, Variant Origin, and Clinical Significance), linked to Drugs, Disease Ontology, PubMed and Trust rating assigned with 5-star system

Curator activity is tracked for attribution, notifications and provenance



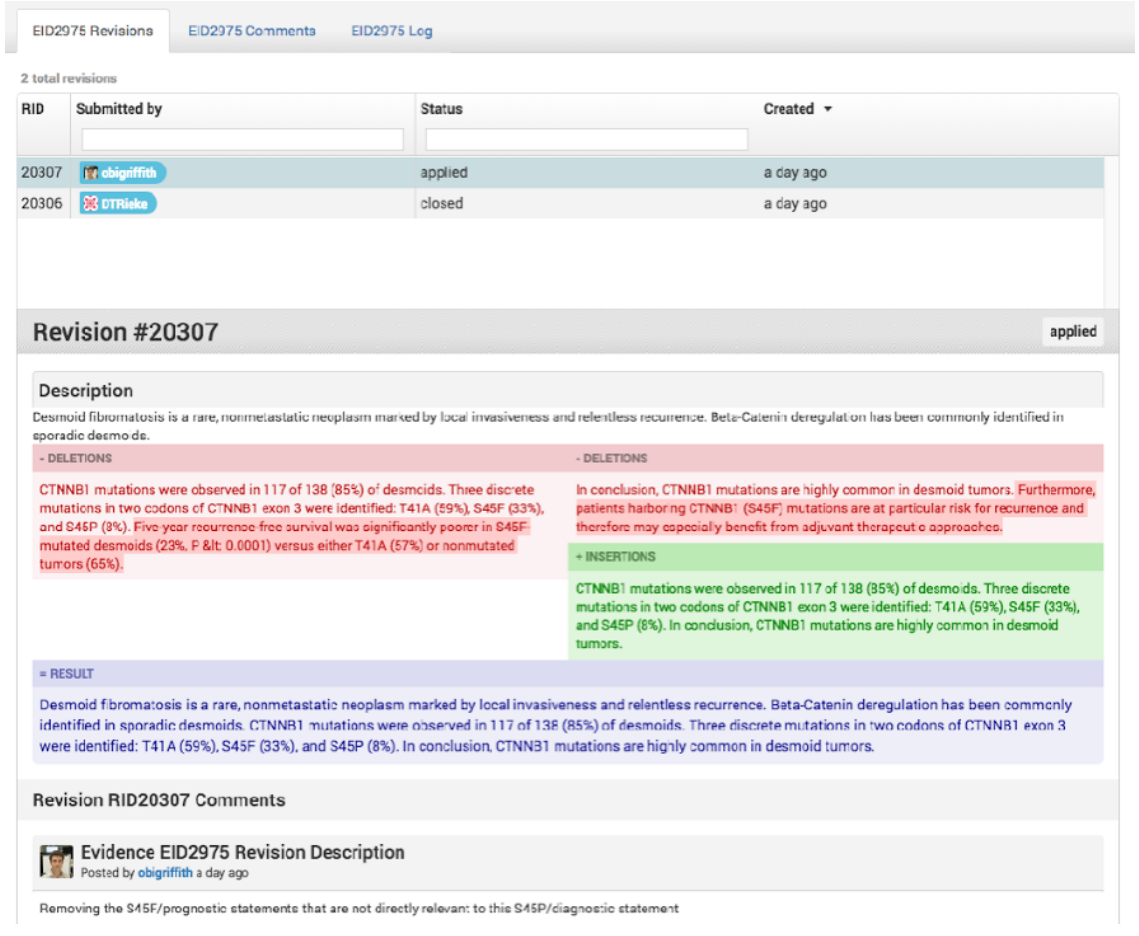
All curated entries have “Talk pages” and “Subscribe” option



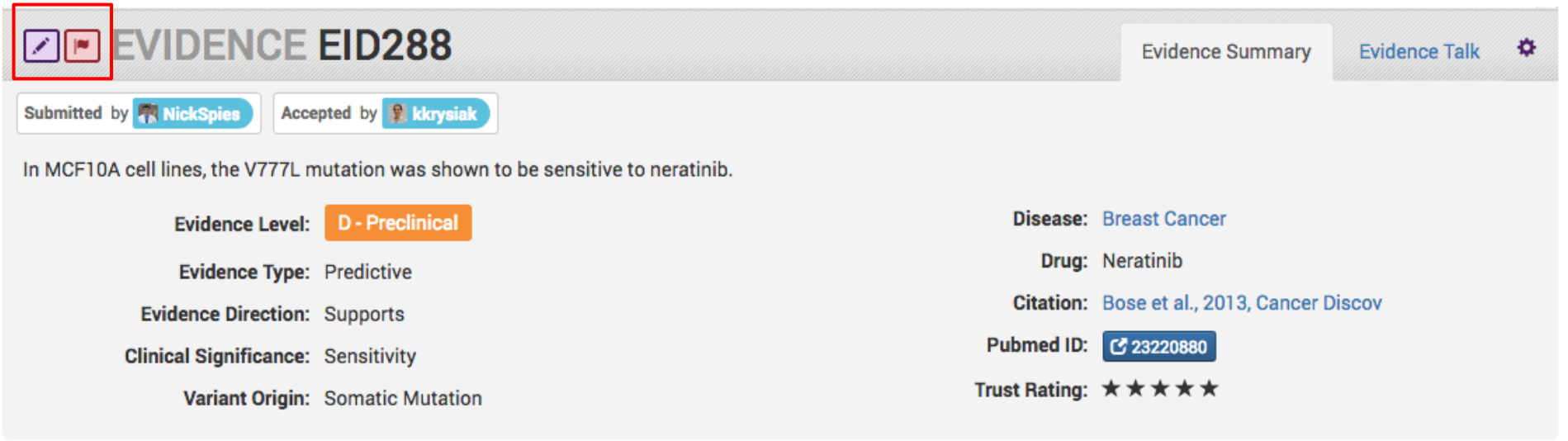
Talk pages (genes, variants, evidence) track the complete history of comments and revisions

Curators and editors have to option to “Follow” any entry (gene, variant, evidence) to receive notifications of proposed changes or additions

Revisions are tracked with detailed GitHub style diffs and comments

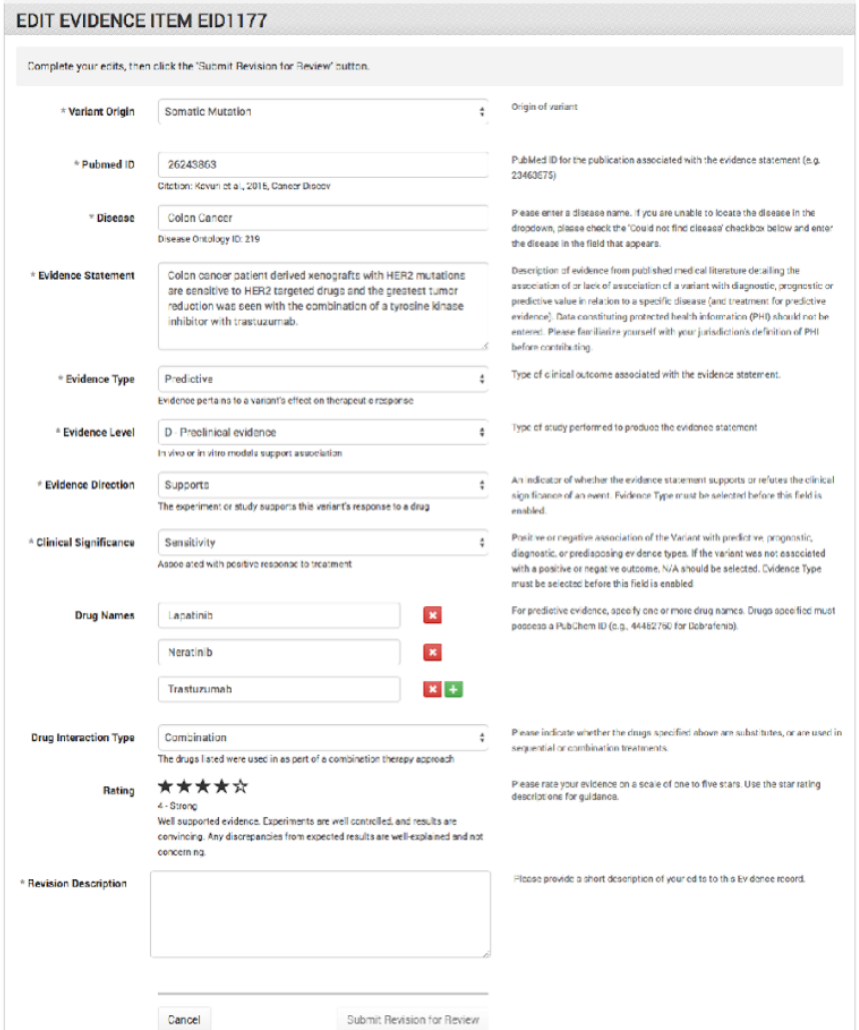


All curated entries (genes, variants, evidence) can be “Flagged” or Revised



All curated entries (genes, variants, evidence) can be “Flagged” for problems or revisions proposed

Flagging allows for very quick/easy marking of content needing immediate review or caution with use



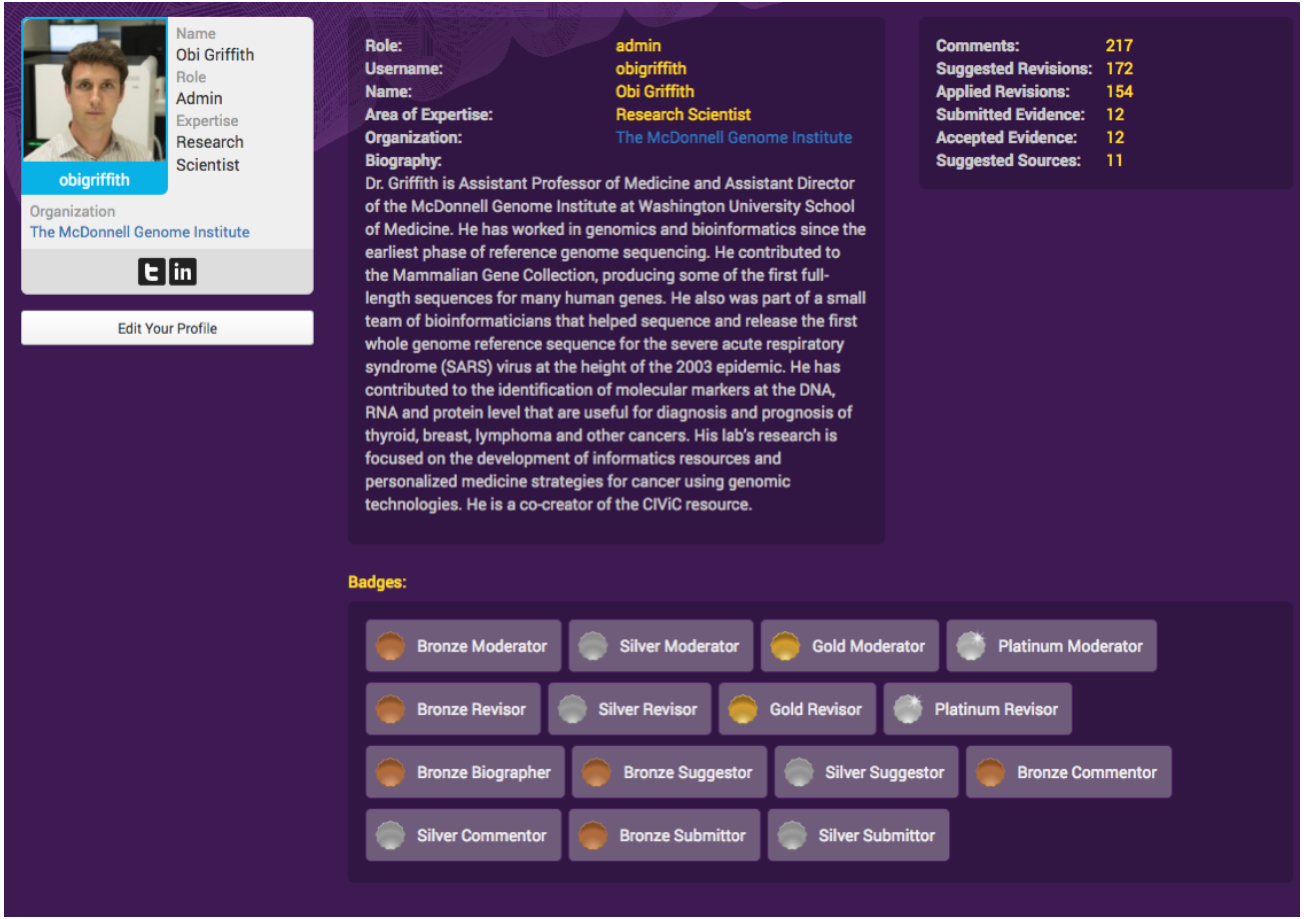
Revisions (or new submissions) are made through a sophisticated data entry page

Dynamic form adjusts for evidence type

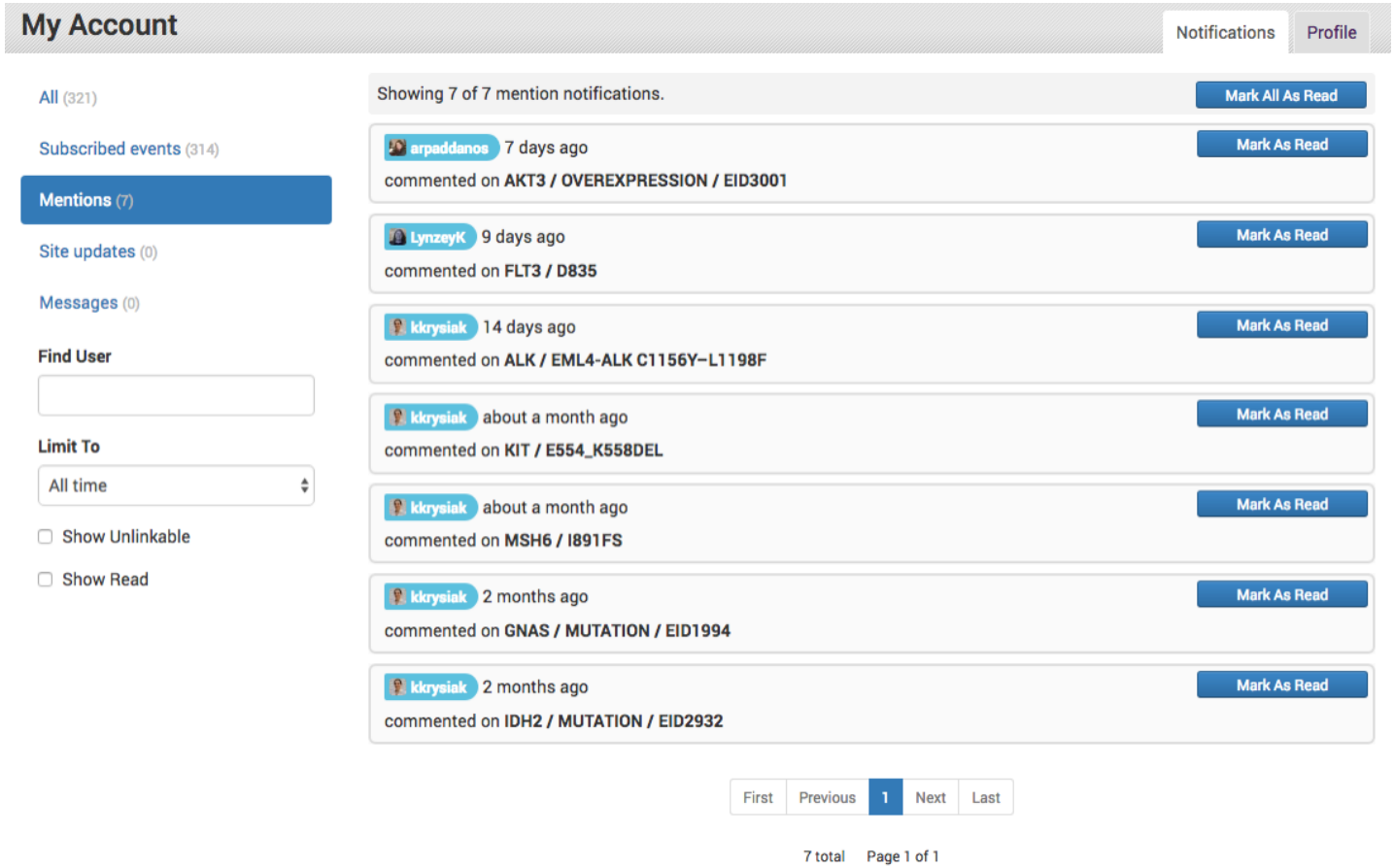
Live type-ahead suggestions and Ontology look-ups

Warnings for merge conflicts

Users can create detailed profiles, self-organize into curation groups, and compete for badges



Users receive notifications, site mentions, updates and messages



All data are available via TSV data dumps or API (json)

Data Releases

Disclaimer: TSV releases of CIViC data are provided at regular intervals for the convenience of those who require the use of a static file. For most users, we recomend utilizing our API which is documented here. Using the API will provide you with the richest metadata about CIViC entries as well as the most current versions of all evidence statements. In fact, the entire CIViC web frontend runs off the exact same API that is available for public use.

These TSV files do not contain user profile data, pending or rejected evidence items, discussion and commentary, or data provenance and revision history.

Date	Gene Summaries	Variant Summaries	Variant Group Summaries	Evidence Summaries
Nightly	GeneSummaries.tsv	VariantSummaries.tsv	VariantGroupSummaries.tsv	ClinicalEvidenceSummaries.tsv
01-may-2017	GeneSummaries.tsv	VariantSummaries.tsv	VariantGroupSummaries.tsv	ClinicalEvidenceSummaries.tsv
01-apr-2017	GeneSummaries.tsv	VariantSummaries.tsv	VariantGroupSummaries.tsv	ClinicalEvidenceSummaries.tsv
01-mar-2017	GeneSummaries.tsv	VariantSummaries.tsv	VariantGroupSummaries.tsv	ClinicalEvidenceSummaries.tsv

Get a list of evidence items

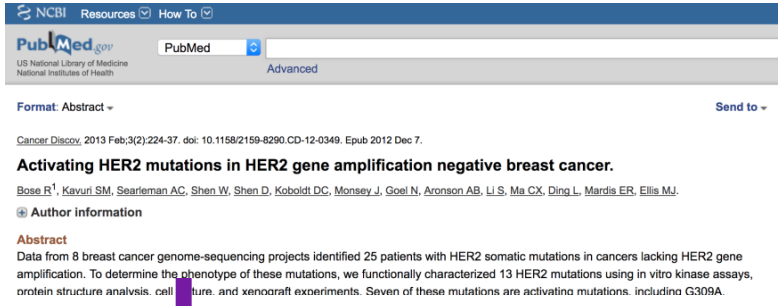
This endpoint returns a listing of evidence items in CIViC. This index style endpoint is *paginated* by default. You can use the `count` and `page` parameters or the `previous` and `next` links to iterate through all the evidence items.

HTTP Request

`GET https://civic.genome.wustl.edu/api/evidence_items`


```
{
  "_meta": {
    "current_page": 1,
    "per_page": 25,
    "total_pages": 69,
    "total_count": 1722,
    "links": {
      "next": "https://civic.genome.wustl.edu/api/evidence_items?count=25&pa
```

The CIViC curation cycle promotes quality, provenance and an up-to-date knowledgebase



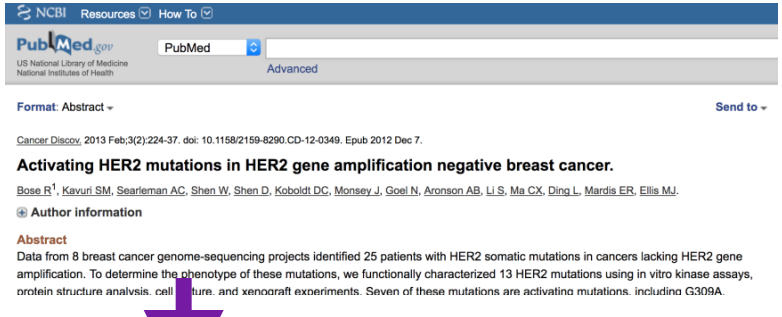
Evidence Record: Rating, Disease, Clinical Significance, Direction, Citation, Statement, Disease Ontology, Origin, Level, Drug, Type

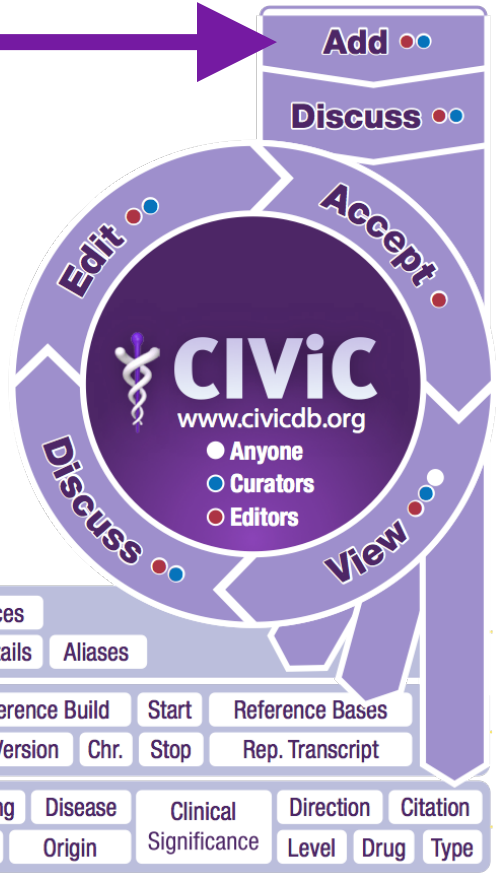
Gene Record: Sources, Summary, Imported Gene Details, Aliases

Variant Record: Reference Build, Start, Reference Bases, Summary, Type, Ensembl Version, Chr., Stop, Rep. Transcript

Evidence Record: Rating, Disease, Clinical Significance, Direction, Citation, Statement, Disease Ontology, Origin, Level, Drug, Type

The CIViC curation cycle promotes quality, provenance and an up-to-date knowledgebase

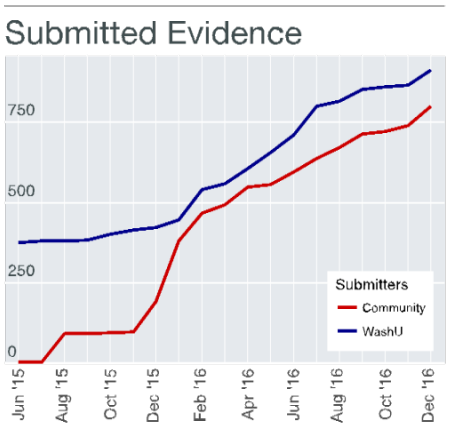


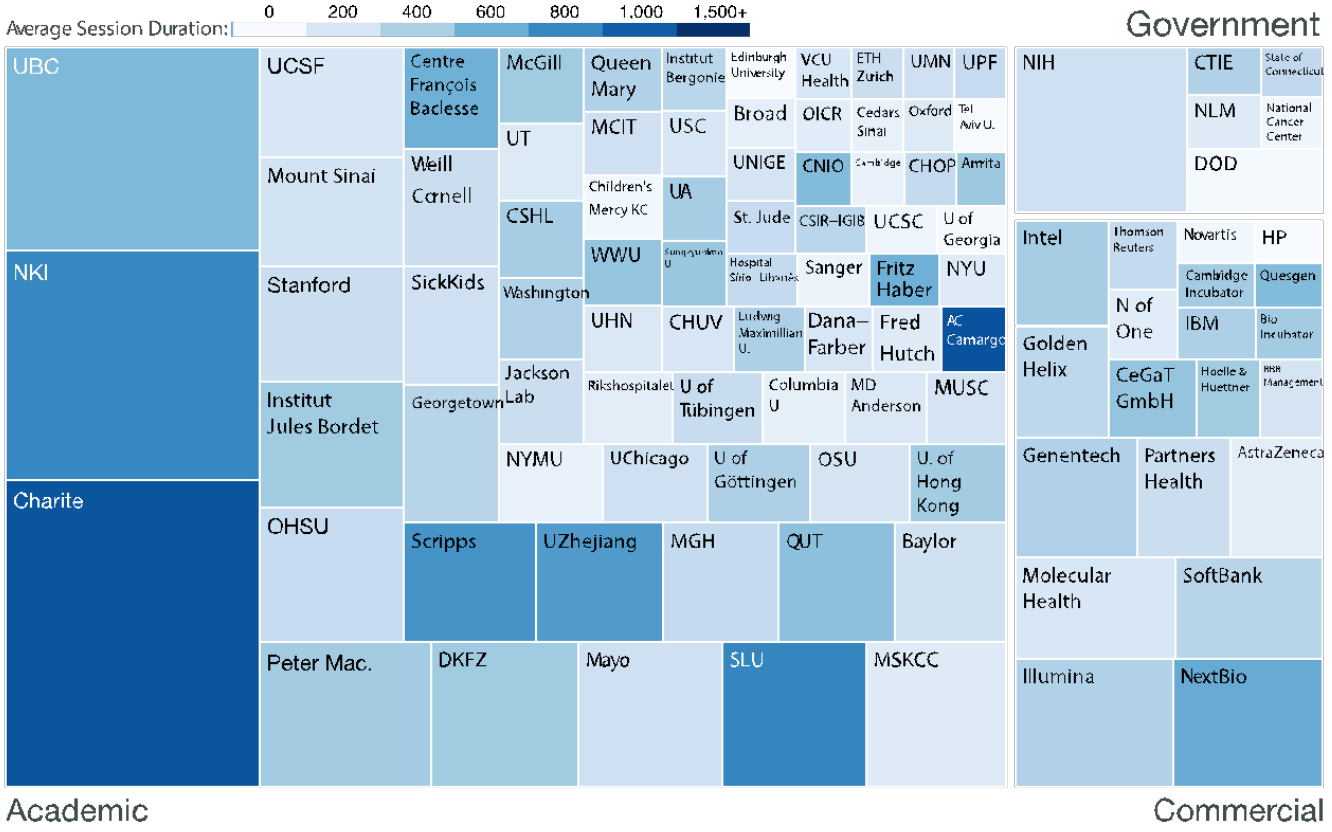


How is quality maintained?

- Content creation is completely transparent
- Anyone can become a curator but this role has limited powers
- All content must be reviewed by a site editor or domain expert
- Users can not accept their own contributions
- Problems can be identified by comment or flag

CIViC contributions and site visits demonstrate strong community involvement





2674 interpretations curated for 1012 variants, 308 genes, 188 cancer types, 1351 papers

1st Hackathon / Jamboree in December began the task of coordinating curation effort

- Improved coverage of group areas of interest
  - Germline variants
  - Resistance mutations
  - Clinically-relevant fusions
  - Predisposing variants
- UI improvements
  - Badge implementation for curator motivation
- Community efforts
  - WikiData incorporation of CIViC
  - CIViCmine implementation and rule definition

Dec 1, 2016 (1:30PM) -> Dec 2 (2:30PM)

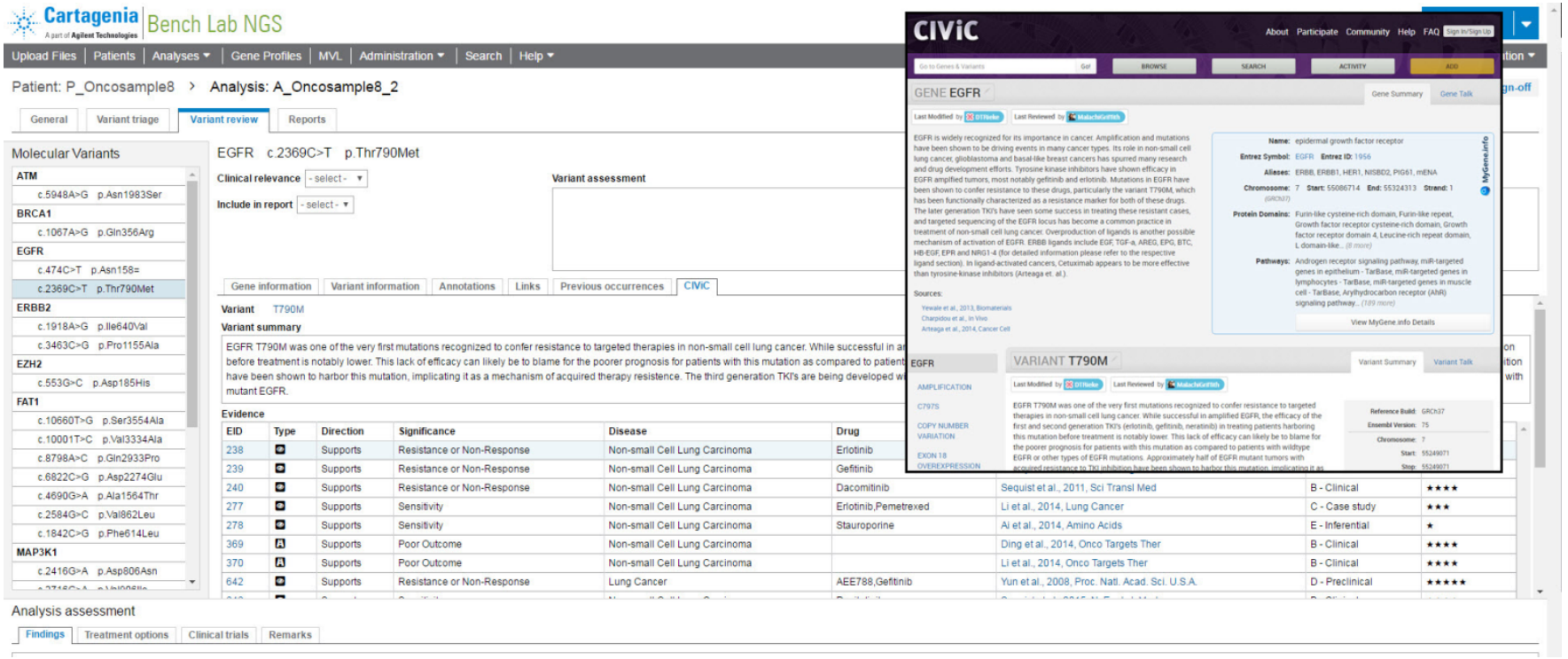
Database and Community Statistics

Genes		Variants		Evidence Items	
This Week:	14	This Week:	74	This Week:	90
This Month:	20	This Month:	87	This Month:	112
This Year:	201	This Year:	491	This Year:	1153
+14	Total: 280	+74	Total: 700	+88	Total: 1662

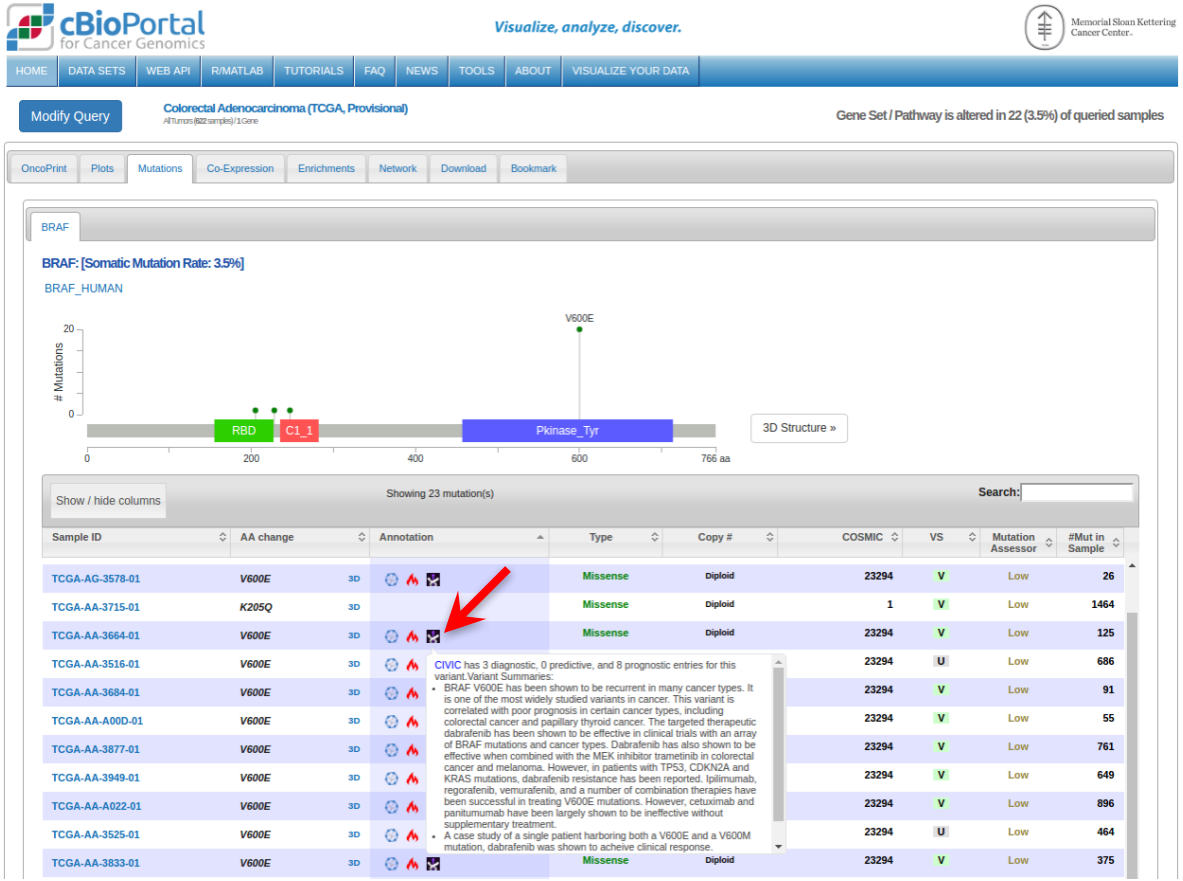
Drugs		Diseases		Publications	
This Week:	8	This Week:	50	This Week:	38
This Month:	14	This Month:	52	This Month:	58
This Year:	193	This Year:	132	This Year:	765
+7	Total: 289	+49	Total: 208	+36	Total: 1066

Revisions		Contributors		Comments	
This Week:	47	This Week:	23	This Week:	154
This Month:	233	This Month:	30	This Month:	453
This Year:	2471	This Year:	51	This Year:	5167
+7	Total: 2843	+11	Total: 57	+129	Total: 5965

CIViC is now integrated into Agilent's Cartagenia Bench



Also coming to cBioPortal: www.cbioportal.org/beta/



Many more:
BioGPS
DoCM
DGIdb
SolveBio
TGex
UCSC Browser
Varsome
WikiData

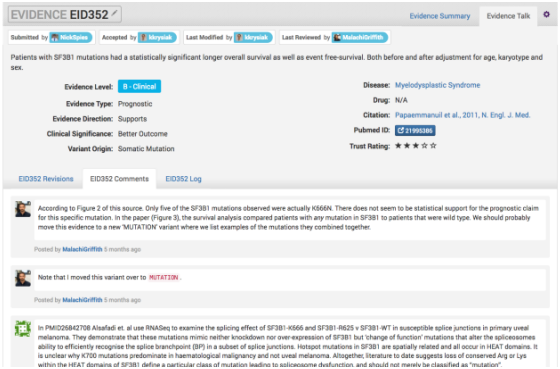
AACR CIViC announcement: Illumina has made a major knowledge donation to CIViC

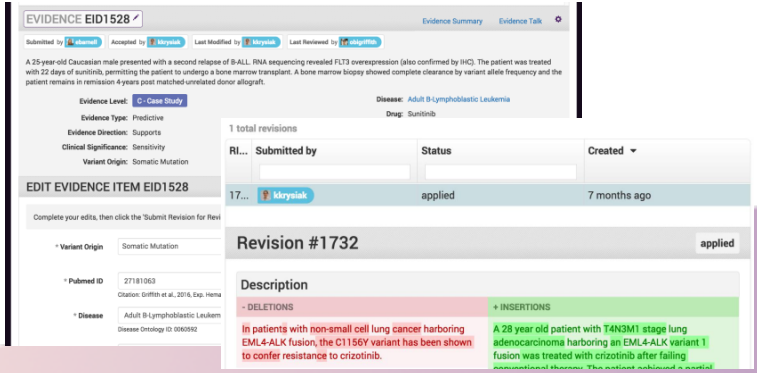
Watch for another press release at ASCO and visit www.civicdb.org to learn more…

How can you participate? Visit www.civicdb.org to join as a curator

LOW CURATION BURDEN

HIGH CURATION BURDEN



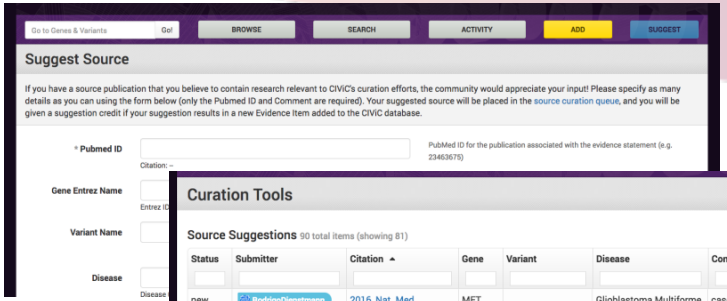


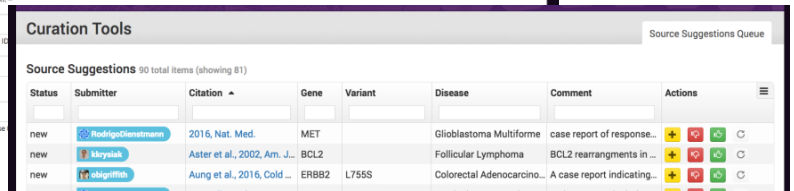
Discuss or flag for review

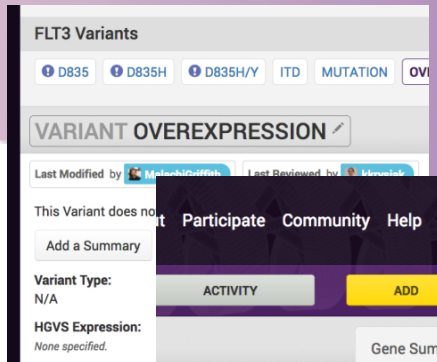
Suggest publications

Suggest revisions

Add new content







Conclusion and future direction

- Conclusion:
  - A high-quality knowledgebase of expert-curated variant interpretations
  - A sophisticated framework/platform for curation
  - A model for community-driven open science

- Future directions:
  - ClinGen collaboration – engage more expert curators
  - Support complex genotypes
  - Support conference proceedings
  - APTRC supplement to create CIViC-based NGS panel (Salipante)

Acknowledgments

The CIViC Community

McDonnell Genome Institute

Griffith Lab (and CIViC team)

Obi Griffith
Malachi Griffith
Benjamin Ainscough
Erica Barnell
Katie Campbell
Adam Coffman
Kelsy Cotto
Arpad Danos
Yan-Yang Feng
Felicia Gomez
Jasreet Hundal
Kilannin Krysiak
Lynzey Kujan
Jason Kunisaki
Josh McMichael
Cody Ramirez
Zachary Skidmore
Nick Spies
Lee Trani
Alex Wagner
Jason Walker

Funding National Cancer Institute (ITCR U01)

Questions?