Supplementary Materials

Supplementary Methods

Supplementary Figures 1-13

Supplementary Tables 1-2

References

Supplementary methods

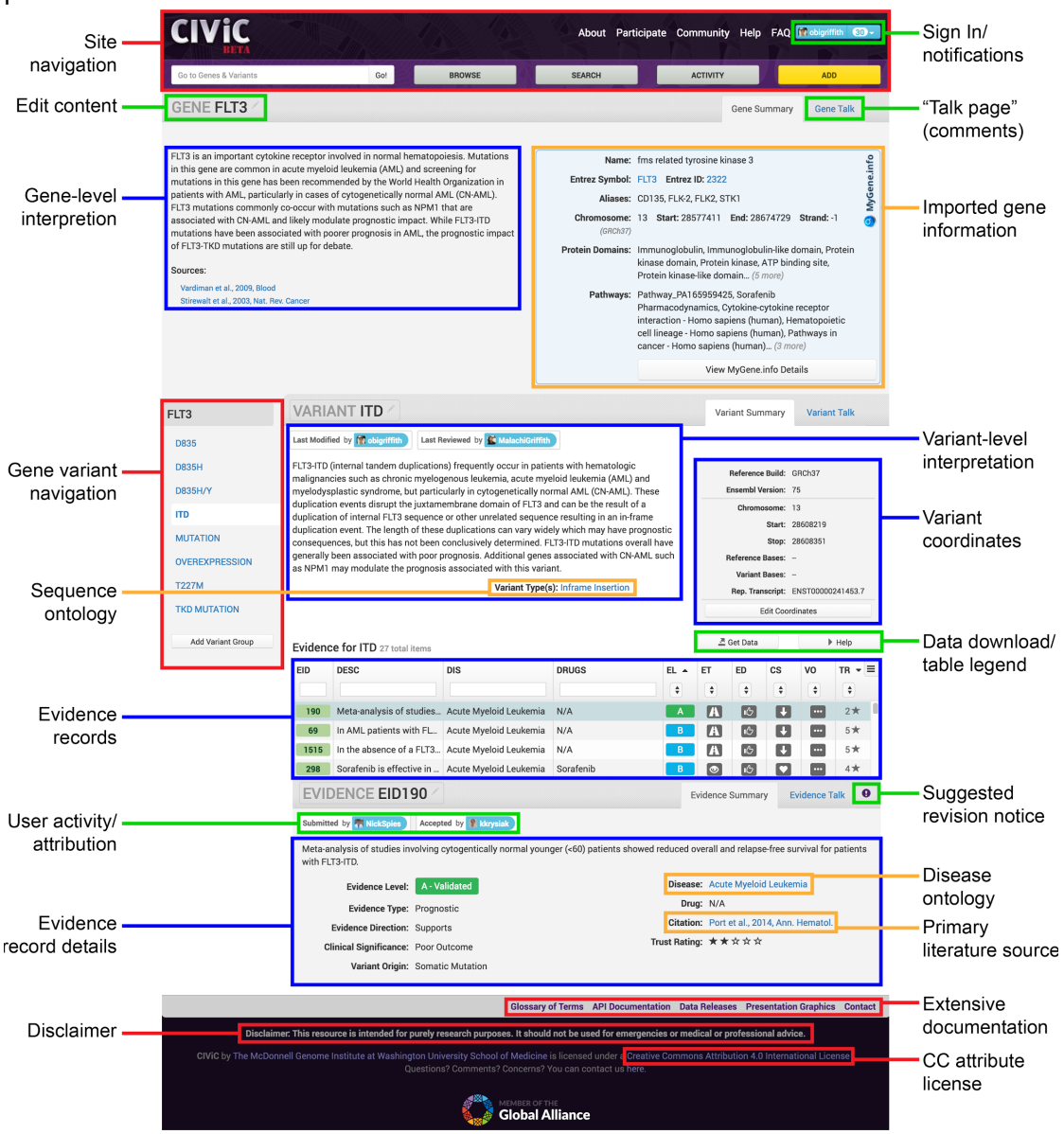
Implementation details

The CIViC source code and application are organized in a client-server model. The code is developed using a continuous integration and test-driven approach. The server side consists of a Ruby/Rails web application that interacts with a PostgreSQL relational database (**Supplementary Figure 5**). The server provides JSON API endpoints to the client. User authentication is managed by 'Oauth v2' and currently supports login with a user's existing ORCID, GitHub, or Google account. 'Code Climate' is used to evaluate code quality, 'Travis CI' for automated code testing, and 'Coveralls' to evaluate test coverage (currently 92%). The client side consists of an 'AngularJS' application that interacts with the CIViC server. It uses 'NPM' and 'Bower' for package management, and 'Gulp' to build the JavaScript application. Code changes are first pushed to a staging server for testing before being deployed to the public server using 'Puppet'. Current development efforts can be followed in the public GitHub pages at https://github.com/genome/civic-client (front end) and https://github.com/genome/civic-server (back end). Anyone is free to submit pull requests or issues (feature requests, bug reports, etc.) to these repositories. Using cutting edge methods and software development best-practices promotes integration with future end-user development and implementation tasks with incentive for developers to improve the underlying CIViC resource.

Supplementary Figures

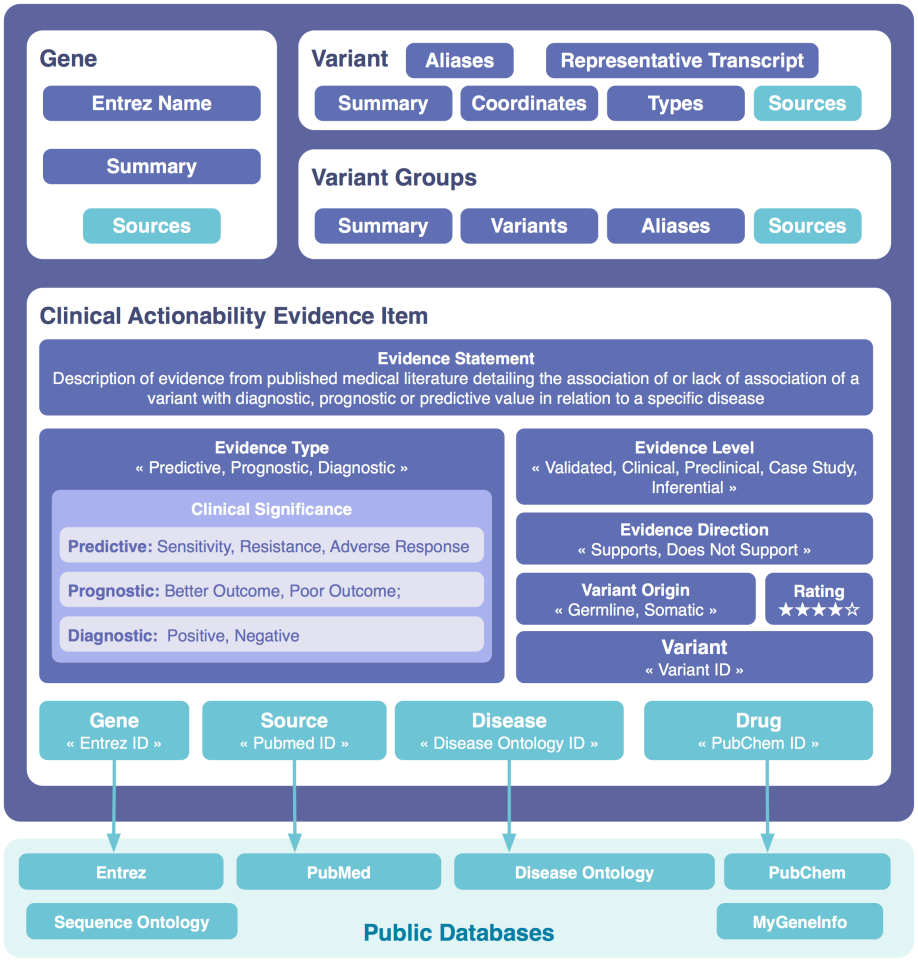
Supplementary Figure 1. CIViC interface overview

The user-friendly CIViC interface is the primary point of contact with users whether they are consuming, editing or adding content. CIViC user-curated content (blue boxes) is visible without sign in and provides the bulk of visible content ordered from gene level (top) to variant level (middle), and finally individual evidence records (bottom). Curated content is enhanced by imported content and citations (orange boxes) that are linked directly to their original source. Website navigation and extensive documentation are highlighted with red boxes. Finally, a curator can interact (green boxes) with CIViC user-curated content by 1) suggesting changes (edit button) or adding content; 2) commenting on content or suggested revisions; 3) downloading content; or 4) viewing their activity, pending suggested changes, notifications or profile.



Supplementary Figure 2. The CIViC data model

Key elements of the CIViC data model are listed below. Briefly, CIViC aims to provide gene and variant level executive summaries of the clinical relevance of specific variants. Multiple structured evidence records are first created and then synthesized to produce these executive variant/gene summaries. Each evidence record is associated with a specific variant and gene. Each evidence record also corresponds to a single clinical assertion for a single cancer type from a single peer-reviewed publication. One publication can be used to generate multiple evidence records. The evidence record consists of a free-form, human readable statement and several structured elements. The statement is a summary of one to a few sentences written by a curator to summarize the clinical relevance of a variant according to evidence described in a particular publication. The curator attempts to concisely summarize the clinical assertion being made by the publication, as well as the nature of the evidence supporting that assertion and any caveats the reader should be aware of. The curator must also assign values for each structured element by evaluating details from the publication. These elements include evidence type, clinical significance, evidence direction, and others. Where possible, structured ontologies are used in the CIViC data model (e.g. the disease ontology for disease names). Dark blue boxes refer to primary CIViC entities and light blue boxes refer to external data.

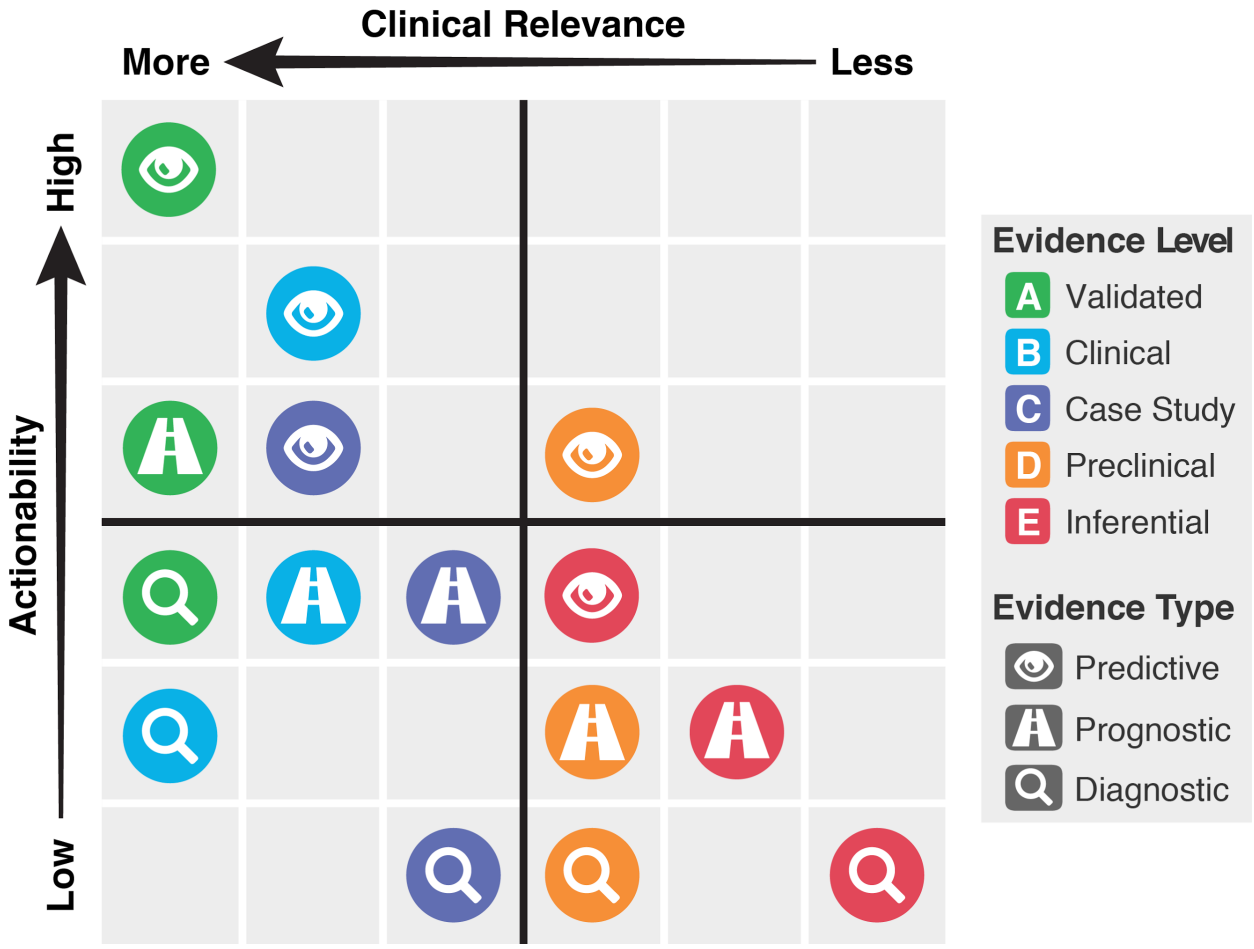


Supplementary Figure 3. Evidence level definitions and examples
Evidence levels defined in the CIViC data model are summarized below. Evidence levels are ordered A-E according to clinical utility (likelihood of relevance to a clinician reading a molecular report). A brief definition of each evidence level is provided along with an example obtained from www.civicdb.org. Updates to the CIViC data model (including to these evidence levels) will be maintained in the CIViC online documentation (https://civic.genome.wustl.edu/#/help/evidence). Additional examples of evidence records assigned to each evidence level can be obtained using the advanced search interface online: https://civic.genome.wustl.edu/#/search/evidence/.

Level	Definition	Examples and further comments
A Validated association	Proven/consensus association in human medicine.	*"AML with mutated NPM1" is a provisional entity in WHO classification of acute myeloid leukemia (AML). This mutation should be tested for in clinical trials and is recommended for testing in patients with cytogentically normal AML.* Validated associations are often in routine clinical practice already or are the subject of major clinical trial efforts.
B Clinical evidence	Clinical trial or other primary patient data supports association.	*BRAF V600E is correlated with poor prognosis in papillary thyroid cancer in a study of 187 patients with PTC and other thyroid diseases.* The evidence should be supported by observations in multiple patients. Additional support from functional data is desirable but not required.
C Case study	Individual case reports from clinical journals.	*A single patient with FLT3 over-expression responded to the FLT3 inhibitor sunitinib.* The study may have involved a large number of patients, but the statement was supported by only a single patient. In some cases, observations from just a handful of patients (e.g. 2-3) or a single family may also be considered a case study/report.
D Preclinical evidence	In vivo or in vitro models support association.	*Experiments showed that AG1296 is effective in triggering apoptosis in cells with the FLT3 internal tandem duplication.* The study may have involved some patient data, but support for this statement was limited to in vivo or in vitro models (e.g. mouse studies, cell lines, molecular assays, etc.).
E Inferential association	Indirect evidence.	*CD33 and CD123 expression were significantly increased in patients with NPM1 mutation with FLT3-ITD, indicating these patients may respond to combined anti-CD33 and anti-CD123 therapy.* The assertion is at least one step removed from a direct association between a variant and clinical relevance.

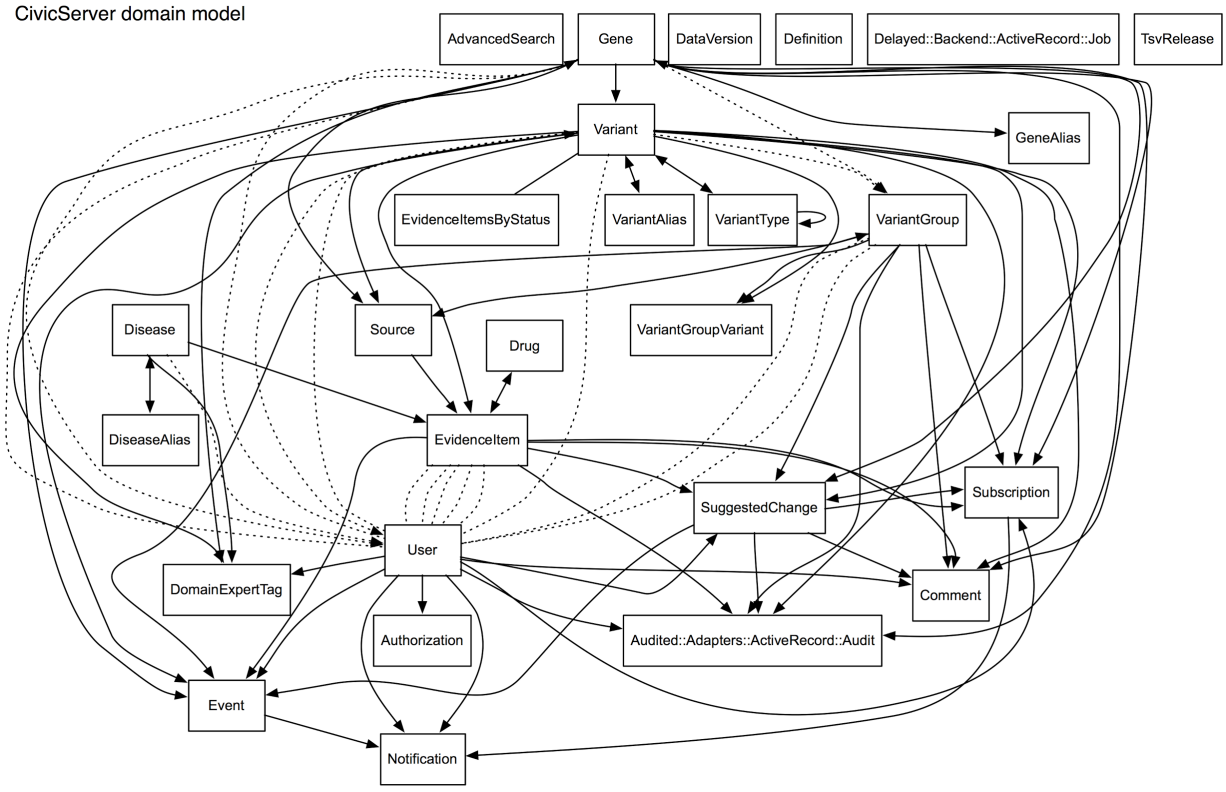
Supplementary Figure 4. CIViC evidence classes and their relative potential to influence clinical actions and understanding of disease

The following diagram attempts to order each combination of evidence level (A-E) and evidence type (predictive, prognostic, and diagnostic) according to their potential clinical relevance and actionability. 'Clinical relevance' refers to the contribution of the variant to clinical understanding of the disease and 'actionability' refers to the ability to identify a specific clinical action for a specific variant. In this assessment, validated predictive variants tend to be the most relevant and actionable, while inferential diagnostic are the least relevant. In general, higher evidence levels are more actionable and predictive assertions exceed prognostic and diagnostic evidence for clinical utility.



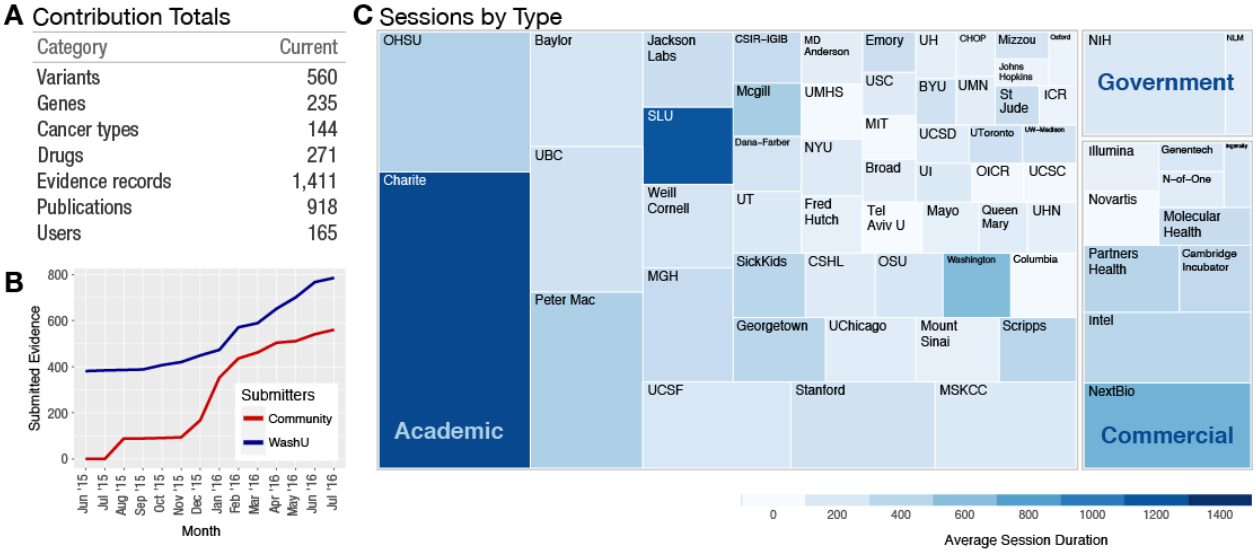
Supplementary Figure 5. CIViC database schema

A simplified schema representing the CIViC data model below provides all table names of the CIViC relational database (running on PostgreSQL). Polymorphic associations are used to relate core domain objects such as evidence records, genes, and variants to the tables that power on-site workflows like moderation and discussion. This allows for a significant reduction in the total number of tables required at the expense of database enforced foreign key constraints. In lieu of traditional foreign keys, validations in the application's business logic are used to enforce data integrity. Solid lines in the diagram indicate direct relationships in the database implemented by a local foreign key (for example, a variant has an evidence item identifier in the variants table, and thus a direct relationship). Dotted lines indicate relationships that exist indirectly (the relationship goes through an intermediate event with some conditions attached to it). For a complete schema including all fields and foreign key relationships, refer to the CIViC backend code repository: https://github.com/genome/civic-server.



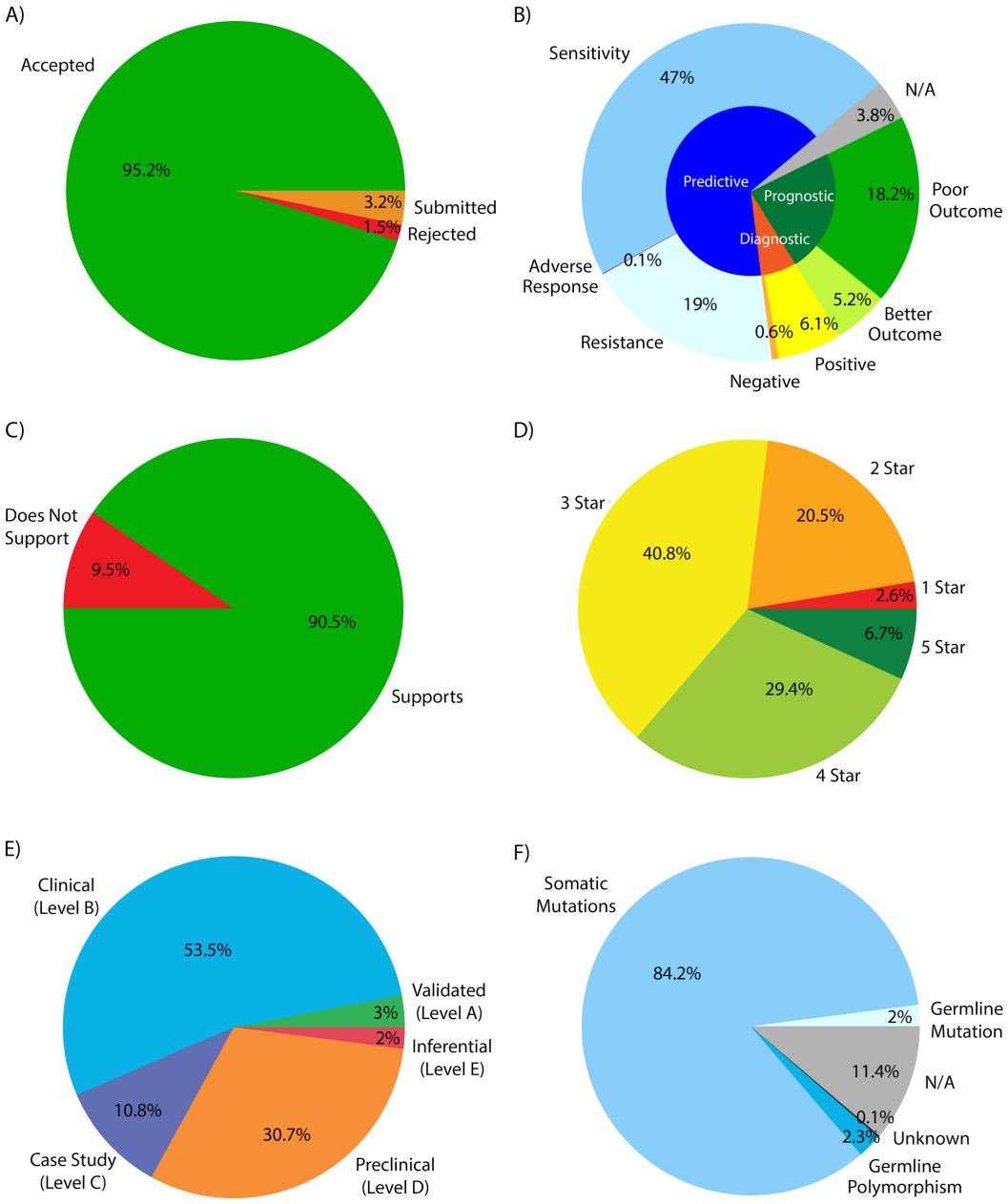
Supplementary Figure 6. Usage statistics and growth of content

A) CIViC content as of July 2016. B) Tracking of evidence statements within CIViC over time with respective contributions of internal (Washington University, 'WashU') and external (community) curation. C) Treemap with box size illustrating the relative number of visits (sessions) to the CIViC website www.civicdb.org from specific external organizations and colored by the average session duration (in seconds). Sessions from our own institute are excluded from this summary. D) Map illustrating the location where sessions originated. To date, the beta version of CIViC has achieved 26,295 visits from 12,350 unique visitors from 2,162 cities in 120 countries around the world.

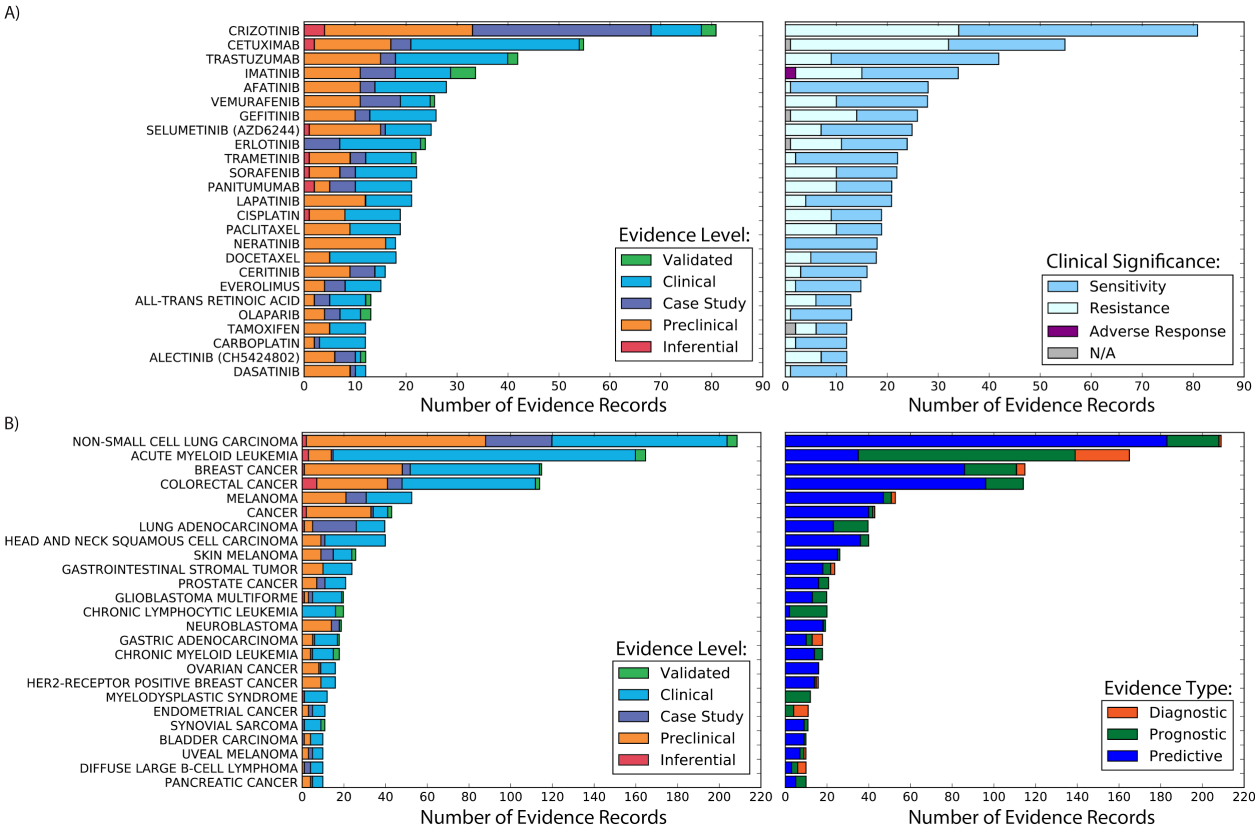


Supplementary Figure 7. Summary of current CIViC evidence records

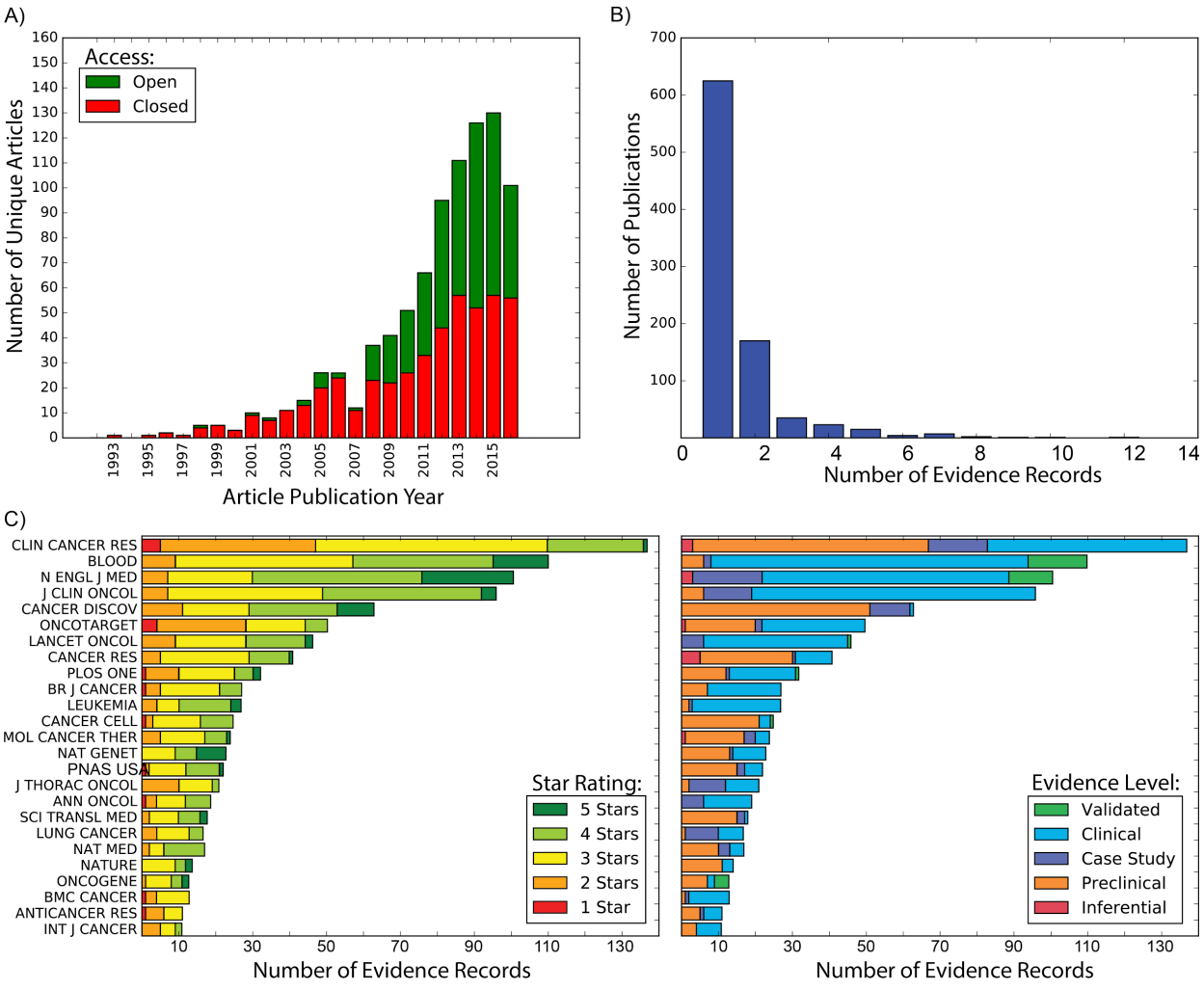
The following panels briefly summarize CIViC evidence records at the time of publication. A) Total publications used in 1,434 evidence records, broken down by review status of the evidence record. Panels B-F further summarize these evidence records after excluding those that had a 'rejected' status (leaving 1,411 submitted or accepted evidence records). B) Evidence records broken down by evidence type and clinical significance. C) Evidence records broken down by evidence direction. D) Evidence records broken down by evidence trust rating. E) Evidence records broken down by evidence level. F) Evidence records broken down by variant origin.



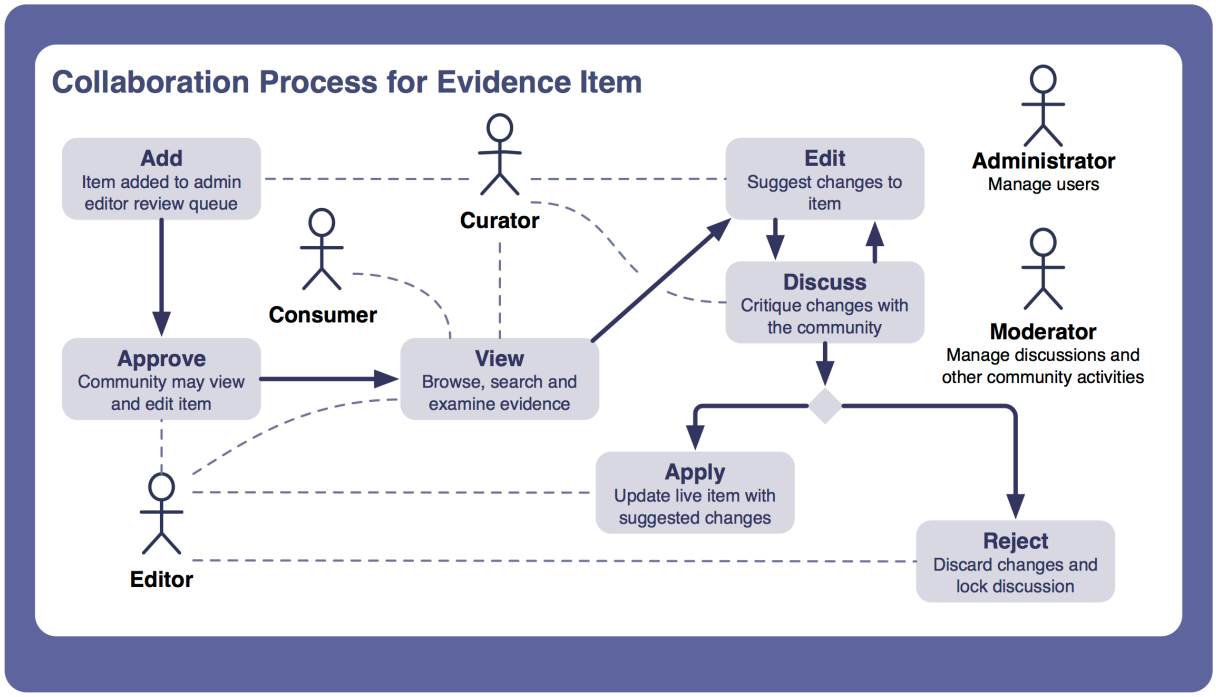
Supplementary Figure 8. Summary of the most curated drugs and diseases in CIViC
A summary of the drugs and diseases represented in CIViC evidence records ranked by the number of evidence records associated with each. A) The top 25 drugs were identified from 945 accepted or submitted evidence records of the predictive evidence type. The evidence records for these drugs are broken down by evidence level (left panel) and clinical significance (right panel). B) The top 25 cancer types (distinct disease ontology terms) were identified from all 1,411 accepted or submitted evidence records. The evidence records for these diseases are broken down by evidence level (left panel) and evidence type (right panel).



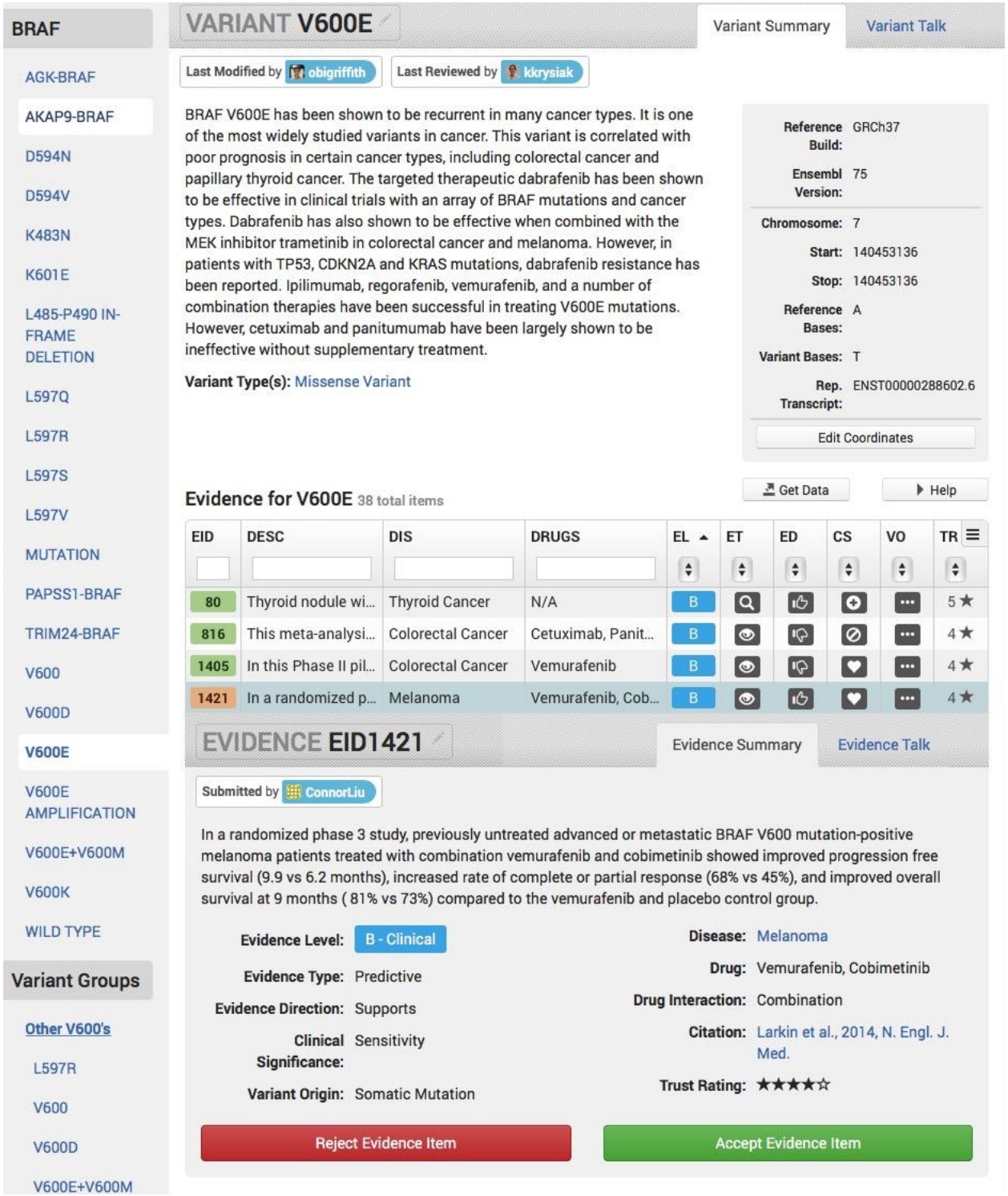
Supplementary Figure 9. CIViC evidence records summarized by literature sources
The published literature used to create all CIViC evidence records are summarized below. A total of 1,411 accepted or submitted evidence records were derived from 918 peer-reviewed publications. A) A histogram summarizing articles used in CIViC evidence records broken down by year of publication (and further divided according to their open versus closed access status). B) A histogram showing the distribution of number of evidence records obtained from single publications. Most publications yield only a single evidence record, but as many as 12 have been obtained. C) Evidence records obtained from the top 25 journals most commonly mined in CIViC are summarized and broken down by the evidence type of evidence records extracted from these journals on the left. The same evidence records from the top 25 journals broken down by evidence star rating are displayed on the right.



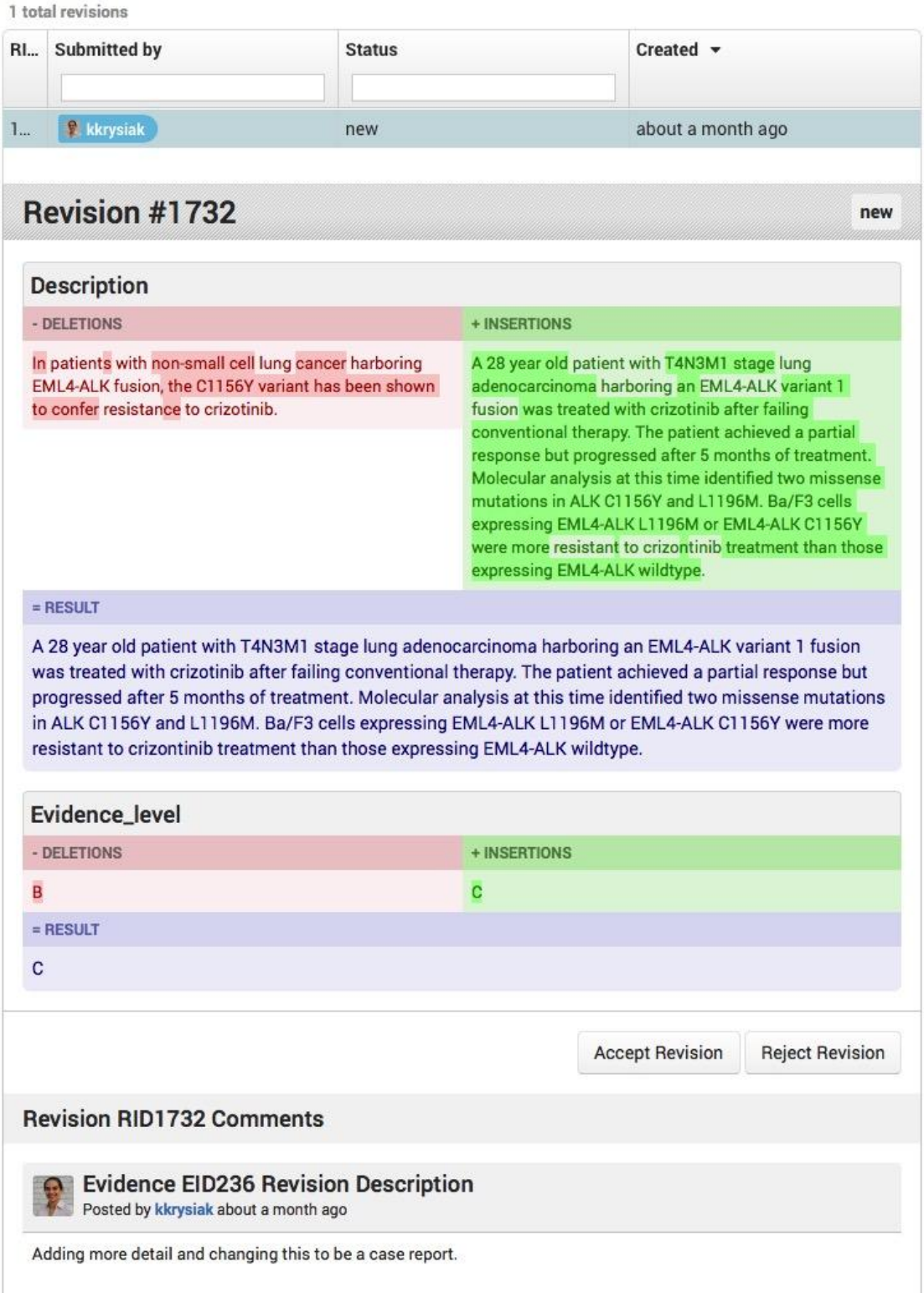
Supplementary Figure 10. The collaborative process and user roles in creating evidence CIViC consists of an online web resource whose target audience is an international community of cancer researchers, clinicians, and patient advocates. Participants in CIViC fall into various categories with increasing privileges or capabilities in the interface. The first category and most basic level of user is that of 'consumer'. Consumers may view, download and programmatically (via API) access all of the content of CIViC under the terms of the Creative Commons Public Domain Dedication license (CC0). No login is required to use CIViC. No requirement to login, fees, or other encumbrances will be introduced in future versions of CIViC. Consumers may not add, approve, edit, or discuss revisions of content in CIViC. The second category of users includes all those roles that do permit modification and discussion in the site: 'curators', 'editors', and 'administrators'. 'Curators' may add new evidence records describing clinical relevance of variants, add or improve variant/gene summaries, and discuss existing content. While comments/discussion are automatically accepted, additions and revisions to existing content are initially entered in a pending state and must be approved prior to acceptance in CIViC. Rejected content is not deleted and may be revived after further discussion and revision. Editors have the additional capability to approve or reject additions and revisions of content. However, an editor cannot approve their own submissions or revisions, meaning that all content in CIViC must be created in collaboration between at least two members of the community. Editors are selected by a committee of existing editors, based on direct knowledge of the editor's expertise or by promotion from curator after demonstrating extensive contributions to CIViC. Finally, administrators have the abilities of editors but may also change user roles and use advanced site management utilities (e.g. merging duplicate records).



Supplementary Figure 11. Screenshot of the editor view for a submitted evidence record
Every new evidence record and any revision of existing content in CIViC must be approved by at least one independent editor prior to acceptance. The following screenshot shows a new evidence record submitted by a curator that is awaiting review by an editor.

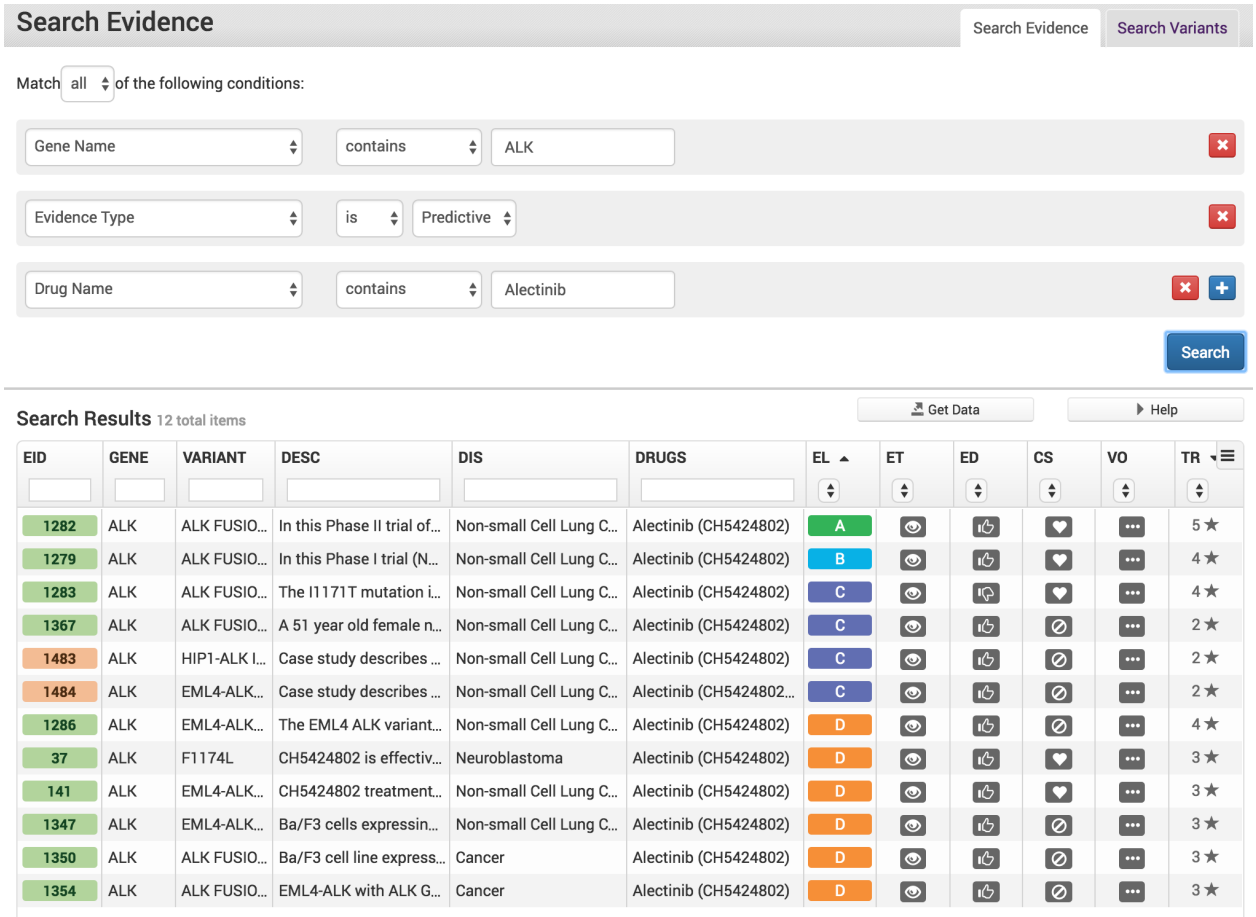


Supplementary Figure 12. Screenshot of the editor view for a pending revision
After proposing a revision to existing content, a contributor is presented with a summary of the fields they are proposing to modify. An independent editor must approve these revisions before they are displayed in the canonical CIViC results (the web interface and API).



Supplementary Figure 13. Screenshot of a complex evidence query

CIViC has an advanced search interface that currently supports complex queries for evidence records and variants. An arbitrary number of query conditions can be set and the query can be configured to match any one, or all of these conditions. Evidence records can be queried by sixteen variables including disease, variant name, publication ID, evidence type, evidence level, trust rating, curator name, etc. In the following screenshot, the advanced search interface is being used to retrieve all evidence records that correspond to variants involving the gene *ALK,* where the evidence type is 'Predictive', and the drug involved is alectenib. From this query, 12 evidence records are returned and sorted according to their quality level (evidence level, and trust rating). The standard CIViC evidence datagrid is used to display a summary of the 12 evidence records including: evidence identifier (EID), gene name, variant name, evidence statement (DESC), cancer type (DIS), drugs, evidence level (EL), evidence type (ET), evidence direction (ED), clinical significance (CS), variant origin (VO), and evidence trust rating (TR). The 'Help' button provides a comprehensive legend of all abbreviations, symbols, and colors used to encode information in the evidence record summary. Clicking any row will take the user to the comprehensive display for that evidence record. Every advanced search generates a unique URL that can be used generate an updated result for a complex query at a later time, or easily share the result with a colleague.



Supplementary Tables

Supplementary Table 1. Related resources

This table compares CIViC to other resources with regard to their curation model, ability to view content without registering, existence of a public API, ability to download bulk data, open licensing of the code and content, and various technical features.

This table can be downloaded as a spreadsheet from the journal's website. Alternatively, a live version that will be updated as these resources develop can be found here: https://goo.gl/5WAZmd

Supplementary Table 2. Literature covered by CIViC compared to related resources

At time of publication, CIViC contained curated evidence records obtained from 895 peer-reviewed publications. A summary of the overlap between these publications and those curated by related resources is provided below. Refer to **Supplementary Table 1** for extensive details of each related resource.

This table can be downloaded as a spreadsheet from the journal's website.

	Cancer Genome Interpreter (CGI)	CanDL (CDL)[1]	Gene Drug Knowledge Database (GDKD)[2]	OncoKb (OKB)	Precision Medicine Knowledge base (PMKB)	Jackson Knowledge base (JKB)[3]	My Cancer Genome (MCG)[4]
Total unique publications	530	126	409	3,700	560	787	840
Percentage of publications in this resource found in CIViC	21.9%	24.6%	26.9%	6.8%	6.6%	8.6%	14.9%
Percentage of publications in CIViC found in this resource	13.0%	3.4%	12.3%	1.6%	4.1%	7.6%	14.0%
Total overlapping publications with CIViC	116	30	110	61	37	68	125
Maximum overlapping publications with any other resource	293 (55.3%) (GDKD)	38 (30.2%) (MCG)	293 (71.6%) (CGI)	91 (2.5%) (PMKB)	91 (16.3%) (OKB)	73 (9.3%) (MCG)	125 (14.9%) (CIViC)

References

1. Damodaran, S. *et al.* Cancer Driver Log (CanDL): Catalog of Potentially Actionable Cancer Mutations. *J Mol Diagn* **17**, 554-9 (2015).
2. Dienstmann, R., Jang, I.S., Bot, B., Friend, S. & Guinney, J. Database of genomic biomarkers for cancer drugs and clinical targetability in solid tumors. *Cancer Discov* **5**, 118-23 (2015).
3. Patterson, S.E. *et al.* The clinical trial landscape in oncology and connectivity of somatic mutational profiles to targeted therapies. *Hum Genomics* **10**, 4 (2016).
4. Yeh, P. *et al.* DNA-Mutation Inventory to Refine and Enhance Cancer Treatment (DIRECT): a catalog of clinically relevant cancer mutations to enable genome-directed anticancer therapy. *Clin Cancer Res* **19**, 1894-901 (2013).