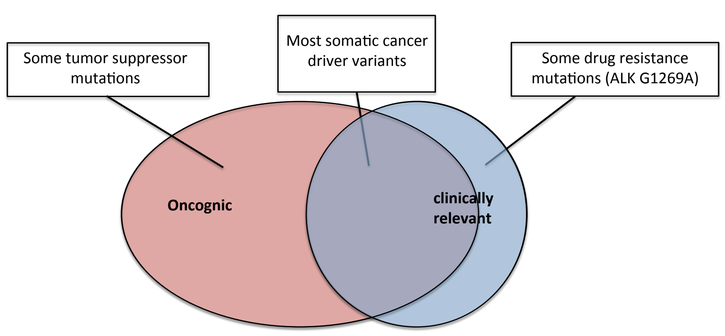
Integration of Biological Information into the CIViC Evidence Item

A. Danos, K. Krysiak, A. Wagner, S. Kiwala, J. McMichael, A. Coffman, E. Barnell, M. Griffith, O. Griffith

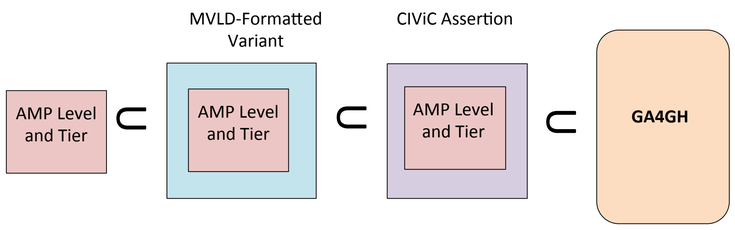
Overview

Oncogenic and clinically relevant variants may not always overlap



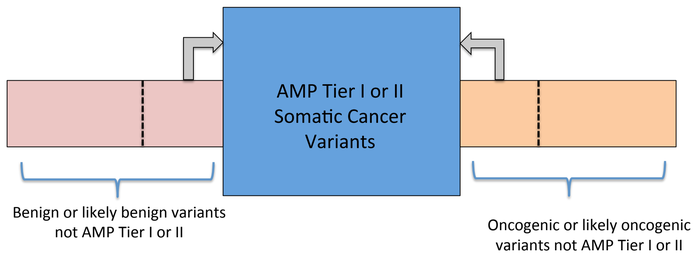
Somatic variants which play an active role in cancer can be broadly classified into variants with oncogenic activity, and variants with clinically relevant information directly associated. Variants with oncogenic properties may not always directly be associated with clinical information.

Somatic clinical cancer variant representation format inclusions



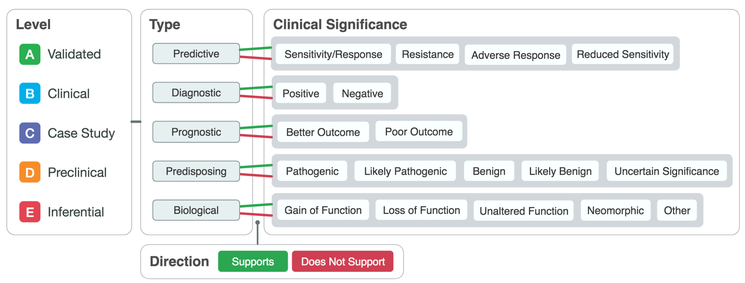
The AMP Tier and Level (Li *et. al*) has been broadly adopted in somatic cancer variant data formats and clinical information representations, including MVLD and CIViC, where the AMP Tier and Level is an organizing principle for the somatic assertion.

A subset of non-AMP Tier I or II oncogenic or benign variants can induce clinically relevant categorization of these variants



For a subset of variants with no clinical information directly attached to them, oncogenic or benign calls can induce clinical categorization of these variants and may also admit an AMP Tier and Level.

CIViC Evidence Items, including proposed Biological Evidence Item



CIViC Evidence Items (EIDs) associate literature derived evidence with a clinical significance type, creating a positive or negative statement is relation to that clinical significance. Evidence Items are labeled with evidence level based on what kind of evidence was presented in the particular study used to EID creation.

Abstract

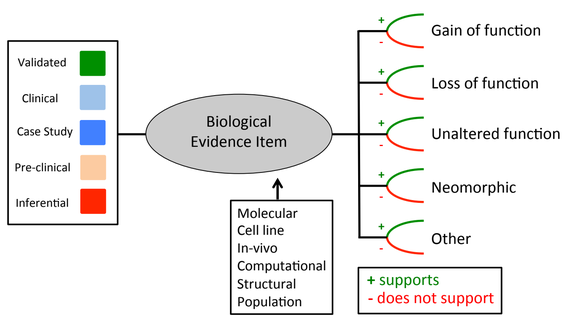
With the growing availability of clinical cancer genome sequencing coupled with growing knowledge on the vast array of cancer variants, there is great need for data resources focusing on the subset of clinically important cancer variants. The Clinical Interpretations of Variants in Cancer (**www.civicdb.org**) database is designed to meet this need. As construction of large datasets requires massive amounts of curation effort, some databases maintain this curated knowledge behind a paywall, presumably using fees for this curation. In contrast, CIViC's approach is to crowdsource and expertly moderate curation effort, which allows it to be a completely free, open-access public resource, able to quickly respond to new directions in the field as the active user base grows. To date CIViC has built a community of over 100 contributors and over 16,000 clinical and research users worldwide.

The fundamental unit of knowledge in CIViC is the evidence item (EID). The EID associates a clinically relevant statement to a specific variant in a specific cancer type using evidence derived from peer-reviewed publications and linked to a PubMed ID. EIDs are used to build summary clinical assertion statements for a variant in a specific cancer context, to which an AMP Tier and Level is associated.

Here we introduce a new EID type called Biological Evidence. The CIViC EID currently consists of Predisposing, Predictive, Diagnostic or Prognostic evidence, and variants that are known to play an important role in cancer, but do not have directly actionable clinical information associated with them, are not admissible in CIViC. Upon user request for CIViC to be able to handle this type of evidence, and internal desire to expand the data model, Biological EIDs will be added to CIViC. This new evidence type will allow CIViC to gather information on important cancer driver mutations, and furthermore admit important new functionality enabling classification of AMP Tier III (VUS) and AMP Tier IV (known benign) variants. Additionally, Biological Evidence allows CIViC EIDs to incorporate PM1, PM4 and PM5 codes from the ACMG guidelines for assessment of variant pathogenicity at the assertion level. Current debates are underway regarding methods for assessment of somatic cancer variant oncogenicity, and integration of biological evidence will allow CIViC's evolving conception of the cancer variant to incorporate and help establish new standards for variant representation as they emerge.

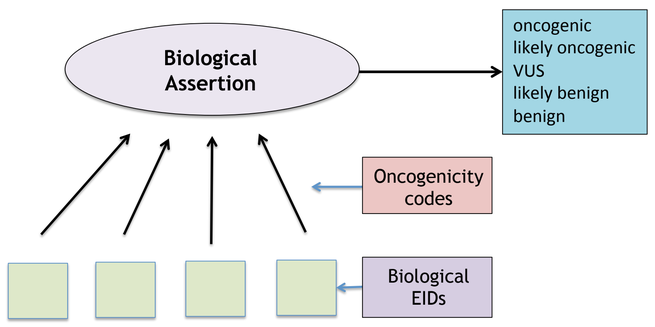
Biological Evidence and Assertion in CIViC

Proposed CIViC Biological Evidence Item



The CIViC Biological Evidence Item (EID) links literature derived biological, functional, structural and other evidence types not directly associated with a clinical cancer predictive, diagnostic or prognostic statement to a somatic cancer variant.

The CIViC Biological Assertion Summarizes Biological Evidence and Outputs an Oncogenicity Valuation



CIViC Evidence Items; each derived from a literature source with PMID (all evidence is for same variant, disease, and clinical statement type)

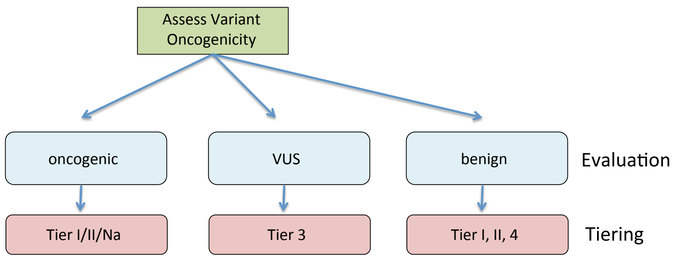
The CIViC Biological Assertion (AID) formally summarizes evidence from multiple Biological EIDs into one statement. In a fashion simmilar to ACMG Codes, (Richards *et. al.*) a proposed set of Oncogenicity codes will be used to guide the creation of assertion level valuations of oncogenicity.

Oncogenicity and AMP Tier

Possible workflow demonstrating oncogenic call influences AMP Tier: Example of protocol to follow (If approved)

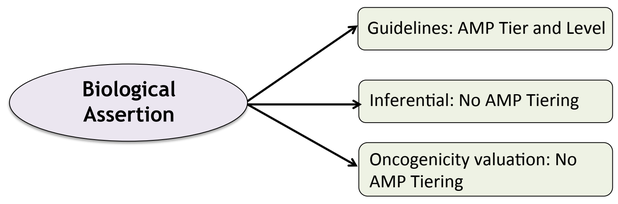
Example: BRCA1 mutation (both somatic and germline) approved for PARP inhibitor in breast cancer

- Check clinvar oncokb CIViC for oncogenicity interpretation using approved ACMG/AMP guidelines for germline/somatic variants.
- If interpretation is not found, perform interpretation using approved ACMG/AMP guidelines for germline/somatic variants (oncogenicity codes)
- If pathogenic/oncogenic it would be Tier I.



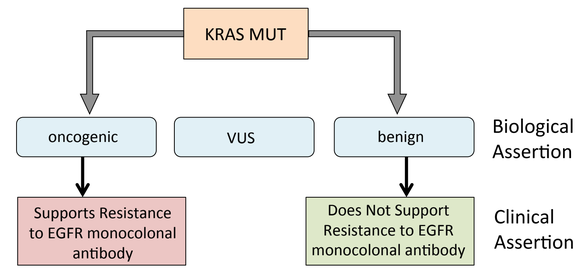
A proposed general workflow and specific example with BRCA1 for assessing mutations with no direct clinical evidence associated. Oncogenic valuation made from BRCA1 variants in the context of breast cancer.

Biological Assertion



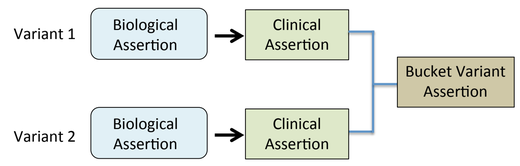
Here we propose that CIViC Biological Assertions can be envisioned to have three principal implementations. When strictly adhering to guidelines, Biological Assertions may admit an AMP Tier and Level and be clinical grade assertions. Another use case is Inferential, where biological features of the variant may be similar to those of other variants with defined clinical utility. This use case will not admit an AMP Tier, but is similar to a recently proposed tiering by ESMO where it would be Tier III-B. Finally, Biological Assertions may be used as stand-alone valuations of a given variant's oncogenicity status.

Oncogenicity Assertions may relate to Clinical Assertions in CIViC



Given a KRAS mutation without direct clinical evidence associated to it, if an oncogenicity call of benign or likely benign is possible, then this Biological Assertion in CIViC can feed into a Clinical Assertion for this variant. The possibility for Clinical Assertion resides in the fact that NCCN Guidelines recommend FOLFIRI + cetuximab/panitumumab for colorectal cancer only in cases of wt KRAS/NRAS. This induces a CIViC Assertion: *Benign KRAS Mut does not support resistance to EGFR antibodies cetuximab/panitumumab*. Since the assertion is backed by NCCN guidelines, we would propose a Tier I Level A AMP rating for this CIViC assertion.

Oncogenicity Assertions point to Bucket Variant Assertions



Multiple variants from the same gene may generate Biological Assertions, which in turn generate Clinical Assertions which point to an overall Clinical Assertion for a bucket variant to which the individual variants belong.

Data Types

The current ACMG/AMP germline guidelines are based on

- population data
- Computational data
- Functional data
- Segregation data
- De novo data (de novo mutations)
- Allelic data (coexist with non-pathogenic variant)

In addition: Recent publication, Walsh *et. al.*: somatic hotspot -> ACMG code

These data types are also those used to create the biological variant, and thus the oncogenicity codes (as currently proposed) follow the ACMG codes in a close fashion.

Acknowledgements and Citations

We would like to acknowledge Dmitriy Sonkin, PhD for providing assistance and sharing these ideas, which were developed in collaboration with the ClinGen Somatic Working Group.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., ... & Voelkerding, K. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine, 17(5), 405.

Li, M. M., Datto, M., Duncavage, E. J., Kulkarni, S., Lindeman, N. I., Roy, S., ... & Nikiforova, M. N. (2017). Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. The Journal of molecular diagnostics, 19(1), 4-23.

Walsh, M. F., Ritter, D. I., Kesserwan, C., Sonkin, D., Chakravarty, D., Chao, E., ... & Kulkarni, S. (2018). Integrating somatic variant data and biomarkers for germline variant classification in cancer predisposition genes. Human Mutation, 39(11), 1542-1552.

Oncogenicity Codes – Initial Proposal

Category Very strong
OVS1_somatic null mutation (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a bona fide tumor suppressor gene.

Caveats:

- Use caution interpreting LOF mutations at the extreme 3' end of a gene.
- Use caution with splice mutations that are predicted to lead to exon skipping but leave the remainder of the protein intact.
- Use caution in the presence of multiple transcripts.

Oncogenic Driver evidence as OVS2?

Category Strong
OS1_somatic Same amino acid change as a previously established oncogenic mutation (by appropriate expert group) regardless of nucleotide change Example: Val→Leu caused by either G>C or G>T in the same codon Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level.
OS2_somatic Well-established in vitro or in vivo functional studies supportive of an oncogenic effect of the gene or gene product Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established.
OS3_somatic Located in one of the hotspots in cancerhotspots.org with at least 50 samples with mutation at amino acid position and the particular amino acid change count in cancerhotspots.org is at least 10. (Use caution with hotspots driven by truncating mutations.) This rule cannot be used if OS1_somatic is applicable.

Category Supporting
OP1_somatic Missense mutation in a gene that has a low rate of benign missense variation and in which missense mutations are known to be oncogenic.
OP2_somatic Multiple lines of computational evidence support an oncogenic effect of the gene or gene product. (conservation, evolutionary, splicing impact, etc.) Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. Can be used only once in any evaluation of a mutation.
OP3_somatic Mutation in a gene in malignancy with a single genetic etiology.
OP4_somatic Located in one of the hotspots in cancerhotspots.org and the particular amino acid change count in cancerhotspots.org is below 10. (Use caution with hotspots driven by truncating mutations.)

Category Moderate
OM1_somatic Located in a well-established mutational hot spot and/or critical and well-established functional domain. (e.g., active site of an enzyme.)
OM2_somatic Absent from controls (or at extremely low frequency) in 1000 Genomes Project, or Exome Aggregation Consortium.
Caveat: Population data for insertions/deletions may be poorly called by next-generation sequencing.
OM3_somatic Protein length changes as a result of in-frame deletions/insertions in known oncogene or tumor suppressor gene or stop-loss mutations in known tumor suppressor gene.
OM4_somatic Novel missense change at an amino acid residue where a different missense change determined to be oncogenic (by appropriate expert group) has been seen before Example: Arg156His is oncogenic; now you observe Arg156Cys.
Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level.
OM5_somatic Located in one of the hotspots in cancerhotspots.org with less than 50 samples with mutation at amino acid position and the particular amino acid change count in cancerhotspots.org is at least 10. (Use caution with hotspots driven by truncating mutations.) This rule cannot be used if OM1_somatic or OM4_somatic is applicable.

Evidence of benign impact
Category Stand-alone
BA1_somatic Allele frequency is >5% in 1000 Genomes Project, or Exome Aggregation Consortium.
Category Strong
BS1_somatic Well-established in vitro or in vivo functional studies show no oncogenic effects.
BS2_somatic Allele frequency is >1% in 1000 Genomes Project, or Exome Aggregation Consortium.
Category Supporting
BP1_somatic Multiple lines of computational evidence suggest no impact on gene or gene product. (conservation, evolutionary, splicing impact, etc.)
Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. Can be used only once in any evaluation of a mutation.
BP2_somatic A synonymous (silent) mutation for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved.

Oncogenic Rules
(i) 1 Very strong (OVS1_somatic) AND
(a) ≥1 Strong (OS1_somatic – OS3_somatic) OR
(b) ≥2 Moderate (OM1_somatic – OM5_somatic) OR
(c) 1 Moderate (OM1_somatic – OM5_somatic) and 1 supporting (OP1_somatic – OP4_somatic) OR
(d) ≥2 Supporting (OP1_somatic – OP4_somatic)
(ii) ≥2 Strong (OS1_somatic – OS3_somatic) OR
(iii) 1 Strong (OS1_somatic – OS3_somatic) AND
(a) ≥3 Moderate (OM1_somatic – OM5_somatic) OR
(b) 2 Moderate (OM1_somatic – OM5_somatic) AND ≥2 Supporting (OP1_somatic – OP4_somatic) OR
(c) 1 Moderate (OM1_somatic – OM5_somatic) AND ≥3 supporting (OP1_somatic – OP4_somatic)

Likely oncogenic
(i) 1 Very strong (OVS1_somatic) AND 1 moderate (OM1_somatic – OM5_somatic) OR
(ii) 1 Strong (OS1_somatic – OS3_somatic) AND 1–2 moderate (OM1_somatic – OM5_somatic) OR
(iii) 1 Strong (OS1_somatic – OS3_somatic) AND ≥2 supporting (OP1_somatic – OP4_somatic) OR
(iv) ≥3 Moderate (OM1_somatic – OM5_somatic) OR
(v) 2 Moderate (OM1_somatic – OM5_somatic) AND ≥2 supporting (OP1_somatic – OP4_somatic) OR
(vi) 1 Moderate (OM1_somatic – OM5_somatic) AND ≥3 supporting (OP1_somatic – OP4_somatic)
Benign
(i) 1 Stand-alone (BA1_somatic) OR
(ii) 2 Strong (BS1_somatic – BS2_somatic)
Likely benign
(i) 1 Strong (BS1_somatic – BS2_somatic) and 1 supporting (BP1_somatic – BP2_somatic) OR
(ii) ≥2 Supporting (BP1_somatic – BP2_somatic)
Uncertain significance
(i) Other criteria shown above are not met OR
(ii) the criteria for benign and oncogenic are contradictory