OncoKB Curation Standard Operating Protocol

Version 1.1
March 2020
OncoKB.org

Chapter 1: Introduction

OncoKB is a Precision Oncology Knowledgebase that contains information about the biological effects and treatment implications of specific cancer genes and their somatic alterations. OncoKB is developed and maintained by the Knowledge Systems group in the Marie Josée and Henry R. Kravis Center for Molecular Oncology at the Memorial Sloan Kettering Cancer Center (MSK).

In OncoKB, genes are classified as either oncogenes or tumor suppressors based on the curated evidence. Alterations included in OncoKB are genetic changes that arise as a result of DNA-level variants in cancer: non-synonymous mutations, translocations, rearrangements / fusions, copy number amplifications and deletions. This document uses “Alterations”, “Mutations” and “Variants” interchangeably. All alterations in OncoKB are classified according to 1) their oncogenic effect and 2) their biological effect, based on the curated evidence (discussed in Chapter 2). The oncogenic effect of an alteration is an evidence-based assertion that classifies whether the mutation is oncogenic, likely oncogenic, neutral or inconclusive. The biological effect of an alteration is an evidence-based assertion that classifies whether the mutation is gain-of-function, loss-of-function, neutral or inconclusive.

If a cancer alteration in OncoKB is associated with clinical implications, these implications are also curated in OncoKB (discussed in Chapter 2). Alterations with clinical implications are further assigned a Therapeutic (Chakravarty et al., 2017), Diagnostic and/or Prognostic level of evidence. Each Level of Evidence assignment in OncoKB defines the strength of the evidence that supports the alteration as being a diagnostic, prognostic or therapeutic biomarker.

OncoKB Oversight and Governance

Oversight and governance of OncoKB is under the purview of the Lead Scientist and the Clinical Genomics Annotation Committee (CGAC). The Lead Scientist and CGAC are responsible for establishing standards and oversight of all processes in the scope of OncoKB. CGAC provides expertise in cancer variant interpretation, and, in particular, the assignment of the OncoKB Levels of Evidence to specific alterations. CGAC consists of “Core” members and “Extended” members. Core CGAC members guide OncoKB development, are at the forefront of clinical management and research and have translational cancer biology expertise in their respective major disease entities. Extended members are selected physicians and scientists who represent the broader MSK clinical leadership across departments and services, including service chiefs, physicians with clinical expertise in their fields, and scientists with specific gene or pathway expertise. Core members, in addition to responding to requests regarding clinical consensus, also maintain an active and responsive dialogue with the Lead Scientist, providing insight or updates regarding genomic biomarker-based clinical data.

OncoKB Staff

The OncoKB staff consists of the following:

1. *The OncoKB Lead Scientist* creates and maintains general oversight and governance procedures for the OncoKB staff including the development, approval, and coordination of all variant assessment activities. The Lead Scientist also liaises between the variant curation processes and their oversight and governance by CGAC. The OncoKB Lead Scientist does not have any relevant conflicts of interest.
2. *The Scientific Content Management Team (SCMT)*, which consists of the following: 1) OncoKB Scientists: Two Ph.D-level scientists with translational cancer biology expertise that provide day-to-day guidance and management of the OncoKB Curators regarding appropriate curation, editorial and scientific content review; 2) Lead Software Engineer: Executes database governance and data preservation as well as feature development and maintenance of the OncoKB Curation Platform (curation platform); 3) Lead OncoKB Data Curator: Liaises between the Lead Software Engineer and OncoKB Scientists to ensure seamless data maintenance, updates and access, and is responsible for database operations. No member of the SCMT has any relevant conflicts of interest.
3. *OncoKB Curators* include pre-doctoral graduate students, postdoctoral fellows and clinical fellows. They assess and curate alterations, their biological effects, and associated treatment implications in cancer in

compliance with the procedures described by the OncoKB SOP. OncoKB Curators are specifically trained in evaluating evidence from various sources and entering the appropriate information into the curation platform.

OncoKB Data Sources

Four primary data sources are used to identify and curate cancer variants and their biological and clinical therapeutic implications (**Fig. 1**):

1. Public cancer variant databases of alterations identified in tumor sequencing studies, e.g., cBioPortal and COSMIC (Catalogue of Somatic Mutations in Cancer).
2. Statistically significant and recurrent variants identified based on 24,592 sequenced tumors using methods described in Chang et al., 2018.
3. Disease-specific treatment guidelines such as those provided by the National Cancer Compendium Network (NCCN) and proceedings of major scientific and/or clinical conferences such as the American Society of Clinical Oncology (ASCO) and the American Association of Cancer Research (AACR).
4. General scientific literature, accessed through PubMed.

The external databases that we use as reference for curation are: 1) IARC TP53 (https://p53.iarc.fr/) 2) BRCA Exchange (https://brcaexchange.org/), 3) Cancer Hotspots (www.cancerhotspots.org). These databases are NOT used as primary curation sources. Rather, they are used for variant candidate selection by downloading the comprehensive list of alterations in each database and comparing them to the mutations curated in OncoKB. Post candidacy, each variant is independently curated using the processes specified in Protocols #2 and #3, and undergo necessary re-evaluation as specified in Chapter 5 sections IX and X. Thus far, we have candidacy selected from the IARC and BRCA Exchange (at the time, known as BIC) databases once in August 2015. Since then, manual review of publications with BRCA and TP53 variants has been our primary process of curation. For cancerhotspots.org two publications in 2016 and 2018 provided a variant candidate list which we reviewed per Protocol #2 and #3. Variants that had supporting scientific literature were classified as “Oncogenic” per Protocol #2 and variants which were considered hotspots based purely on statistical recurrence per Chang et al., 2018 were considered “Likely Oncogenic” per Protocol #2. The Cancer Hotspots website has a static list based on the 2018 publication and has not been updated since.

OncoKB Access

Data from OncoKB is used in four ways (**Fig. 1**):

1. OncoKB data is publicly available for personal and research purposes through an interactive website at www.oncokb.org. Usage terms of OncoKB are specified at https://www.oncokb.org/terms (Fig. 19).
2. The curated data is also available programmatically through the OncoKB application program interface (API). The different ways to access OncoKB data are documented at www.oncokb.org/DataAccess (Fig. 17).
3. The cBioPortal for Cancer Genomics (https://www.cbioportal.org) uses the OncoKB API for annotating cancer variants in its database.
4. OncoKB data is used to annotate the patient reports of the results from MSK-IMPACT, a targeted tumor sequencing test available to MSK patients.

Additionally, this document, a version-controlled OncoKB Curation Protocol v1.1 describing all processes and protocols involved in the maintenance, of OncoKB is publicly available on our website.

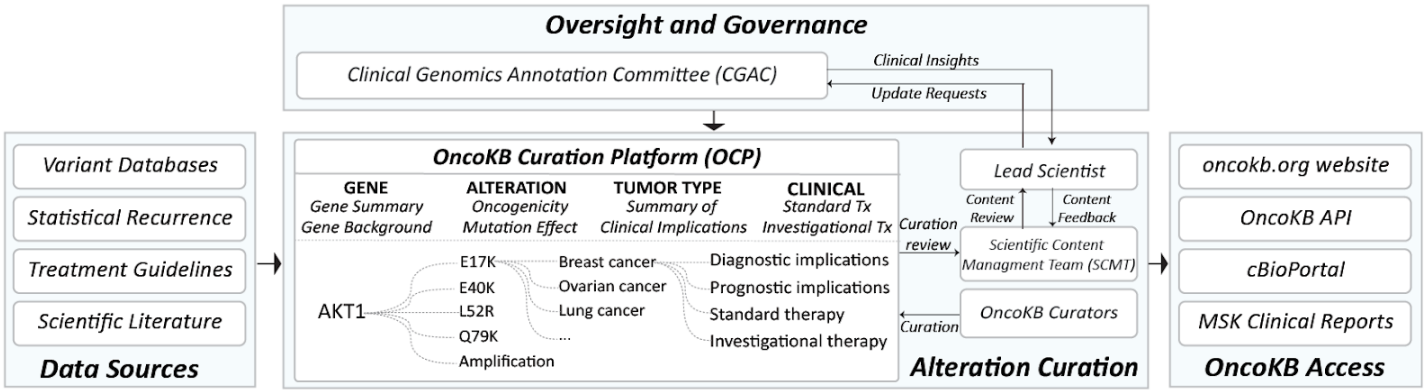


Figure 1: Summary of OncoKB processes. The schematic shows a summary of the data sources, knowledgebase architecture and processes that compose the OncoKB workflow.

Conflicts of Interest

Evidence-based assertions of the oncogenic and biological effect of an alteration (as described in **Protocols #2 and #3**) are not considered to be subject to conflicts of interest (COI). The evidence used to support specific assertions of oncogenic and biological effects is displayed on the website and link to the appropriate references in PubMed or to the scientific abstract website. Variant assertions are re-analyzed and re-evaluated by the OncoKB team in specific review cycles (refer to Chapter 5, Section X and Table 1) and any new content or inconsistencies are corrected at that time. Additionally feedback regarding updated content or inconsistencies from users of OncoKB either through the website or via cBioPortal are addressed within 48 hours of receipt (refer to Chapter 2, Section II.C and Chapter 7, Sections II.L.11 and V.B.6).

A subset of alterations in OncoKB are considered biomarkers that are predictive of response to certain drugs and asserted an OncoKB level of evidence in accordance with Protocol #4. Some of these drugs are FDA-approved and the biomarker is a consideration in standard care. In these cases, the biomarker is associated with either Level of Evidence 1 or 2 (refer to Chapter 5 and **Fig. 7**) and are not subject to COI. However, some of these drugs are either 1) FDA-approved, but the biomarker is in an off-label setting or 2) not FDA-approved and instead are being tested in clinical trials, and for these, COI may arise. In both of the latter scenarios, the biomarkers and drugs are considered investigational and are associated with a Level of Evidence, 3A, 3B or 4 (refer to Chapter 5 and **Fig. 7**).

To address and resolve potential COI, any new level assignments or changes to an existing level have to be approved unanimously by all CGAC members and there are at minimum 3 affirmative verifications from CGAC (please refer to Chapter 5, "Updating Level of Evidence Assertions of Clinically Actionable Variants", p25). The affirmative verifications from CGAC that must be received in order for a proposed change to the levels of evidence to be entered into OncoKB are the following:

1. From the Director of the Center for Molecular Oncology, Dr. David Solit
2. From a Disease Management Team (DMT) Chief in the indication of the proposed level of evidence change
3. A miscellaneous member of CGAC

Members of CGAC who may have COI with respect to the introduction or change of the levels of evidence assigned to a specific variant are allowed to provide advice and information regarding the assertion, but are excluded from the 3 CGAC member verification committee. Additionally, moving forward, for each change or introduction of a new level of evidence, the "News" announcement in the www.oncokb.org website will now include the names of the CGAC members that affirmatively verified the change, and the names of any CGAC members who may have a specific COI regarding the change or new leveled association.

Financial conflicts of interest for all OncoKB personnel including CGAC are disclosed publicly on the OncoKB website, www.oncokb.org/team (Fig. 22) and reported in publications or in conferences as appropriate. In the event of a conflict of interest arising for a specific CGAC member with regards to a Level of Evidence assignment, he or she is asked to recuse themselves from the consensus request. In the event that consensus cannot be immediately

reached, the Lead Scientist is responsible for mediating between conflicting advice to resolve any discrepancy. Should consensus not be reached, the proposed change in the Level of Evidence is rejected.

Additionally to capture any newly arising COIs, biannually the Lead Scientist will send out an email with the complete list of variants with a level of evidence assertion and request CGAC members to declare any conflicts of interest specific to this list. This will be published biannually on the OncoKB website.

External Advisory Board

To further mitigate issues of conflicts of interest (COI), we have convened an External Advisory Board (EAB), which consists of four leaders in the clinical oncology and genomics community: Dr. Victor Velculescu from Johns Hopkins University, Dr. Lillian Siu from Princess Margaret Hospital, Dr. Eliezer Van Allen from the Dana Farber Cancer Center and Dr. Alexander Lazar from MD Anderson. As part of the OncoKB EAB, these members have agreed to meet once a year via WebEx to review summarized OncoKB content, comment on any notable process or content changes based on the FDA-approval and clinical trial landscape, assess productivity of the OncoKB team and advise on improvements to the OncoKB infrastructure, process or content as necessary. Furthermore they will help mitigate and resolve any COI issues among members of CGAC that may arise.

Chapter 2: OncoKB Concepts

I. Concepts in OncoKB

To curate the clinical implications associated with an alteration in OncoKB in a structured way, each clinical implication must be associated with a specific gene, one or multiple alterations, and one or multiple tumor types. The following is the nested organization of key concepts for each gene in OncoKB (**Fig. 2**):

Gene

1. Summary
2. Background
3. Alteration
    i. Mutation Effect
    ii. Tumor type
        Clinical Implications
        1. Diagnostic Implications
        2. Prognostic Implications
        3. Therapeutic Implications
        4. Standard Sensitivity
        5. Standard Resistance
        6. Investigational Sensitivity
        7. Investigational Resistance

4. Variants of Unknown Significance

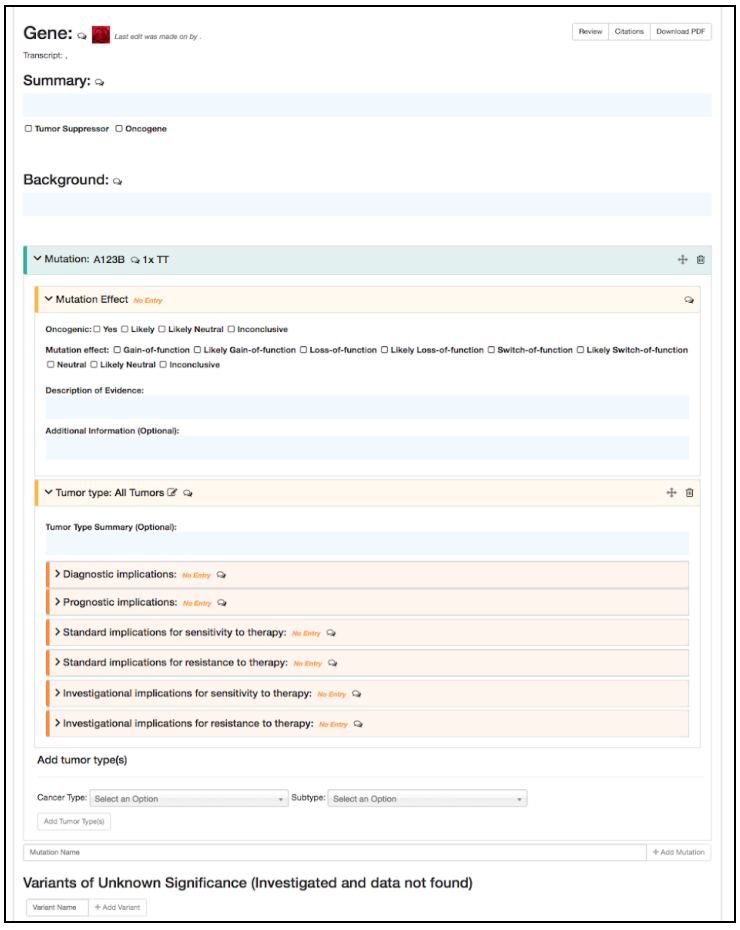


Figure 2: OncoKB is hierarchically organized by its key concepts. Any clinical implication, including drugs that show activity in tumors carrying a specific mutation, is always nested under a specific Mutation and Tumor type within a gene.

II. The OncoKB Curation Platform

Variant information is entered into the OncoKB curation platform, a custom web-based application that allows manual curation and review of variant information. All information entered into the curation platform are structured in a hierarchy of gene, alteration, tumor type and clinical implications. The latter include diagnostic, prognostic, and therapeutic implications. The OncoKB Lead Scientist requests periodic disease-specific content updates from individual CGAC members regarding genomic biomarker-based clinical data. The Lead Scientist also oversees and is responsible for all curation processes to ensure consistency and quality of variant curation and assertions by OncoKB Curators and curation review by SCMT. Addition of new or changes to the existing clinical implications in OncoKB may be prompted by new FDA approvals, FDA-breakthrough designations, and newly reported results of major clinical trials from clinical oncology conferences or publications, requiring clinical consensus among all members of CGAC. CGAC consensus feedback, clinical insights and recommendations are communicated to the Lead Scientist, then conveyed to the SCMT, and subsequently incorporated into OncoKB by the SCMT. All new content (including any updates, additions or deletions) that is entered into the OncoKB curation platform MUST go through a final review/quality control (QC) (refer to Chapter 5, Section IX) before it is finalized and released into public-facing OncoKB outputs (i.e., cBioPortal, oncokb.org and MSK-patient reports) (**Fig. 1**). This is implemented through the Review function on the OncoKB curation platform (**Fig. 3**). Additionally, to ensure that all variant assertions are accurate and the evidence supporting an assertion is up-to-date, a comprehensive reevaluation and reanalysis (refer to Chapter 5, Section X) of genes and their associated variants occurs in review cycles specified in **Table 1** using **Protocols #1-4**. The SCMT may execute the review themselves or assign specific gene(s) as needed for re-evaluation to curators.

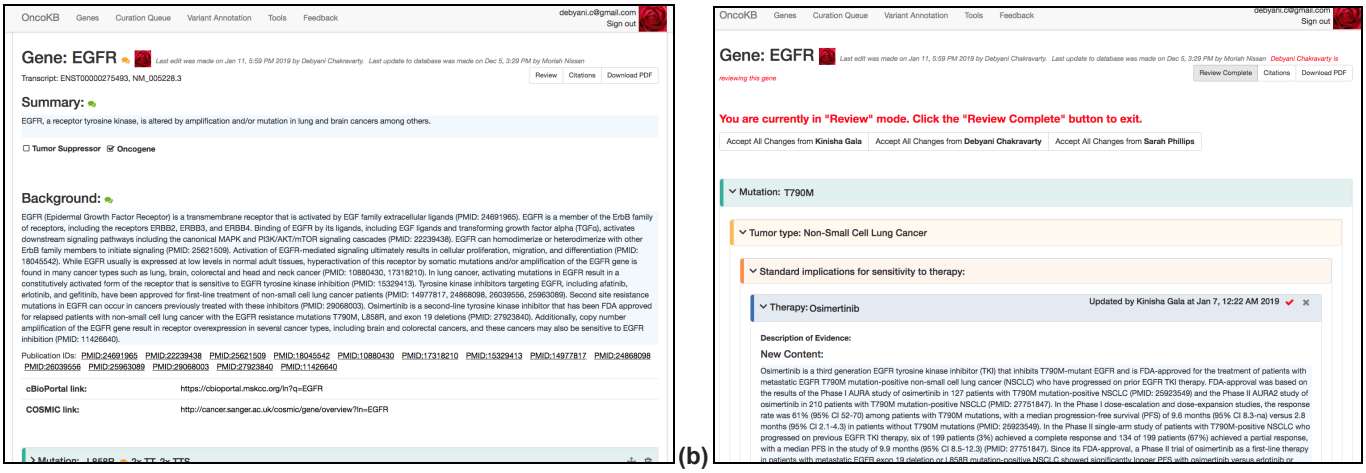


Figure 3: Curation Platform Review Interface. (a) The curation platform interface for curators. (b) The curation platform interface for the Lead Scientist and SCMT with administrative privileges including "Review interface" for reviewing and approving new content curated by OncoKB Curators.

The OncoKB Curation Interface Homepage is divided into the following pages:

A. Genes Homepage

The Genes page (**Fig. 4**) is displayed upon entering the OncoKB curation interface and is the main homepage of the curation interface. This page lists all genes (linking to its own Gene Curation Page) in the OncoKB curation system, along with the following information for each gene:

1. *Last modified:* Timestamp indicating when the Gene Curation Page was last modified
2. *Last modified by:* Name of the last user to edit the page
3. *Needs to be reviewed:* Indicates if there is new content in the Gene Curation Page that needs to be reviewed by the SCMT.
4. *Search Box*: Allows the user to search for their gene of interest.

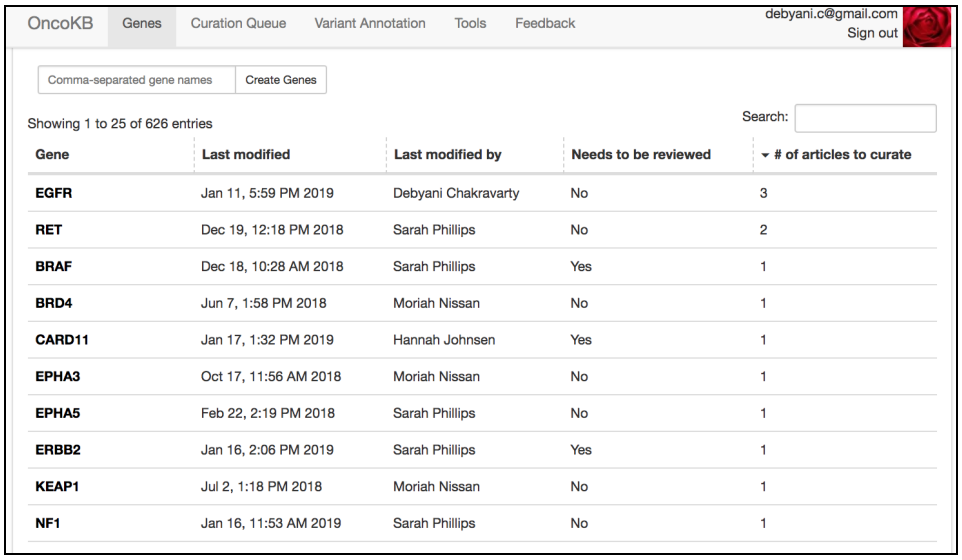


Figure 4: Gene homepage in the OncoKB Curation Platform. The Genes homepage lists all genes in the curation system.

B. Tools

The purpose of the Tools page is to provide data validation checks to the SCMT (**Fig. 5a**). This page is divided into several sections:

1. *Review History*: Allows the SCMT to visualize reviewed changes made to a specific Gene Page. Once a gene is specified, the following outputs are displayed:
   - **a.** Gene Name: The name of the queried gene
   - **b.** Reviewed By: The SCMT member who reviewed the data in question
   - **c.** Records: The specific section within the Gene Page that was reviewed by the SCMT member (i.e., Background, Mutation Effect), and the action taken (Added, Deleted or Updated)

   Each discrete piece of reviewed data within a Gene Page is displayed as its own entry.
2. *Query Reviewed Data*: Allows the SCMT to visualize the following outputs in a table format. These outputs are chosen from a drop-down list and can be downloaded as an XLS file by clicking the 'Download' button.
   - **a.** Oncogene/Tumor Suppressor: Lists all genes in OncoKB and their classification as an oncogene or tumor suppressor. The table also indicates whether the following alterations are curated for each gene: Truncating Mutations, Deletion, and Amplification.
   - **b.** Mutation Effect: Lists all alterations in OncoKB and provides the following information (extracted from the database) for each alteration: Associated Gene, Oncogenic Effect, Mutation Effect, Description of Mutation Effect, Citations.
   - **c.** Tumor Type Summary: Lists all Tumor Type Summaries in OncoKB and indicates the gene-alteration-tumor type combination for which they are associated.
   - **d.** Therapeutics: Lists all alterations associated with a Level of Evidence in OncoKB and provides the following information (extracted from the database) for each alteration: Associated Level of Evidence, Therapeutic, Therapeutic Description of Evidence.
3. *Additional Validation Checks*: SCMT can also query the following two validation questions:
   - **a.** Are truncating mutations curated for tumor suppressor genes?
     This query returns a list of genes in OncoKB that have Truncating Mutations curated as an alteration but are not marked as Tumor Suppressors.
   - **b.** Do all tumor suppressor genes have truncating mutations curated?
     This query returns a list of genes in OncoKB that are marked as Tumor Suppressors but do not have Truncating Mutations curated as an alteration. For some tumor suppressor genes, such as POLE, truncating mutations are purposely not curated as they lack evidence supporting their assertion as oncogenic. However, for the majority of tumor suppressors, truncating mutations are assumed to result in the loss-of-function of the protein and therefore considered oncogenic. Exceptions apply here as well, like in the case of BRCA2, where truncating mutations close to the C-terminus, such as K3326*, are known not to have an inactivating effect.

C. Feedback

The purpose of the Feedback page is to collate all user feedback received about specific OncoKB annotations from a feedback form within the cBioPortal. The feedback form in cBioPortal is also described in Chapter 4, Section III, B.4.f. In brief, the feedback form records the following user inputs (if applicable): gene, alteration, feedback, reference(s), user email address, and cBioPortal link. The Feedback page in the curation platform includes a "Complete" column, in which SCMT members can add the status of the response to the feedback, and a "Comments" column, in which SCMT members can add notes or comments regarding the feedback (**Fig. 5b**).

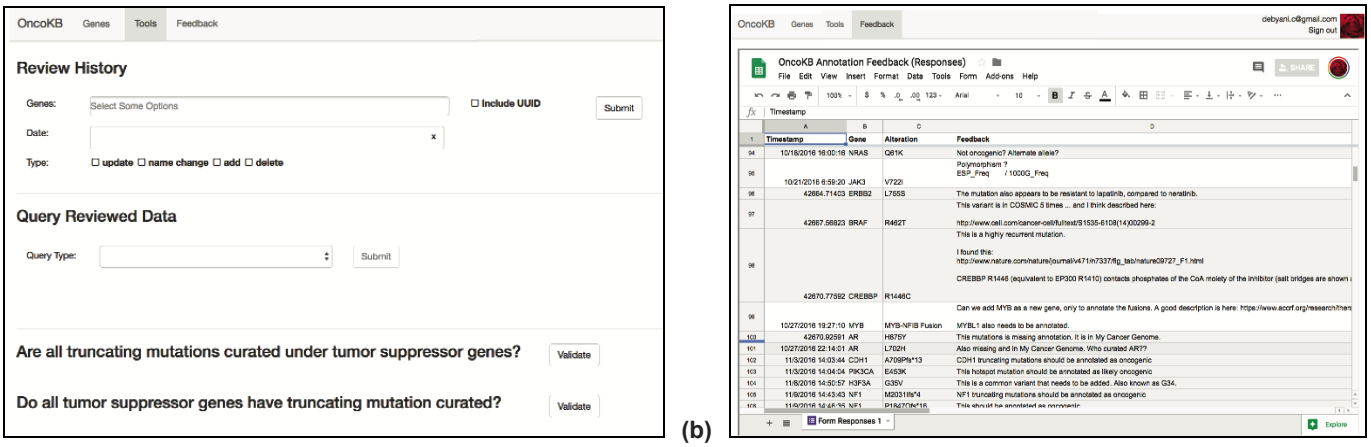


Figure 5: OncoKB Curation Platform Tools and Feedback Pages. (a) Includes ability to look up curation review history, query specific data and check the annotation of tumor suppressor genes. **(b)** All feedback received through cBioPortal is fed to a Google sheet that is accessible through the Curation Platform.

Chapter 3: Gene Curation

OncoKB uses the following standardizations for each gene:

- The HUGO gene symbols are used for gene names. We update to the latest HUGO symbols periodically.
- For each gene, a canonical transcript is selected for annotation. Both Ensemlbl and RefSeq transcript IDs are provided per gene.

The OncoKB Gene Curation Page contains the biological and clinical implications of each gene and its alterations. The Gene Curation Page contains the following sections (ordered by the hierarchy specified in the concept hierarchy section II):

I. Gene Summary

Provides a brief overview of the gene and its role in cancer. This section is free text and contains a 1-2 sentence summary. For the majority of genes, the summary is one sentence that describes the gene function and the cancer types in which it is most frequently altered, e.g., “EGFR, a receptor tyrosine kinase, is altered by amplification and/or mutation in lung and brain cancers among others.”

II. Gene Background

Provides a detailed overview of the biological function of the gene/protein in the normal cell, its role in cancer development and progression, and its clinical significance. The background section is free text and contains 6-10 sentences, although some genes with little published information may have shorter background sections. The background should contain sufficient detail to thoroughly explain the above-mentioned information but should not include minute details and extraneous information. The references used in this section should primarily come from high impact journals (i.e., New England Journal of Medicine, Journal of Clinical Oncology, Journal of Clinical Investigation, Cell, Cancer Discovery, Science, Nature, etc.).

III. Classifying a gene as an oncogene or tumor suppressor

Genes in OncoKB can be classified as oncogenes (e.g., BRAF), tumor suppressors (e.g., PTEN), both (e.g., NOTCH1), or neither (e.g., VTCN1). There are two checkboxes under the gene summary with which a curator may assign whether the gene is an oncogene and/or a tumor suppressor.

The following criteria is used to classify a gene and **Protocol #1** in the Appendix is used to assert oncogene or tumor suppressor for a gene:

A. Oncogene

In OncoKB, an oncogene is defined when a gene meets ≥1 criteria in Evidence I OR ≥1 criteria in Evidence II.
Evidence I. Any of the following features as demonstrated by the scientific literature in ≥1 studies:
(1) A cancer-inducing gene when activated by mutation OR
(2) A gene that can transform cells by increasing the selective growth advantage of the cell in which it resides as demonstrated by the scientific literature in ≥1 studies (Weinberg, p.G:20, 2014, Vogelstein et al., 2013).
Evidence II. A gene that, in tumor samples, has
(1) higher functional impact as defined by the PolyPhen2 Hum-Var prediction model and higher amplification frequency in comparison to those observed in neutral genes, AND
(2) lower loss-of-function mutations, splicing mutations and frequency of deletions and increased frequency of amplification compared to tumor suppressors (Davoli et al., 2013).

B. Tumor Suppressor

In OncoKB, a tumor suppressor is defined when a gene meets ≥1 criteria in Evidence I OR ≥1 criteria in Evidence II.

Evidence I. Any of the following features as demonstrated by the scientific literature in ≥1 studies:

(1) A gene whose partial or complete inactivation by mutation, occurring in either the germline or the genome of a somatic cell, leads to an increased likelihood of cancer development by increasing the selective growth advantage of the cell in which it resides OR
(2) A gene that is responsible for constraining cell proliferation OR
(3) A gatekeeper, a gene that operates to hinder cell multiplication or to further cell differentiation or cell death and in this way prevents the appearance of populations of neoplastic cells OR
(4) Mutated through protein-truncating alterations throughout their length (Weinberg, p.G:20, 2014, Vogelstein et al., 2013).

Evidence II. A gene that, in tumor samples, has
(1) higher frequencies of loss-of-function and splicing mutations, higher functional impact, and higher frequency of deletions compared to those found in neutral genes, AND
(2) higher frequencies of loss-of-function and splicing mutations, higher deletion frequency and lower amplification frequency compared to those found in oncogenes (Davoli et al., 2013).

C. Both

In some cases, a gene may have characteristics of both an oncogene and a tumor suppressor based on the tissue context in which the gene is altered and the criteria defined by OncoKB (refer to above). If a gene meets ≥1 criteria in Evidence I and/or ≥1 criteria in Evidence II classifying it as an oncogene AND meets ≥1 criteria in Evidence I and/or ≥1 criteria in Evidence II classifying it as a tumor suppressor, it is appropriate to check both the oncogene and tumor suppressor checkboxes in the curation platform.

D. Neither

If the gene does not meet the specific criteria for either an oncogene or a tumor suppressor, then both boxes may be left unchecked and the conclusion is that there is no clear evidence that the gene is an oncogene or tumor suppressor based on the criteria defined by OncoKB (refer to above).

Chapter 4: Alteration Curation

I. Nomenclature and Technical Rules for Alteration Curation

Specific nomenclature when curating alterations in OncoKB must be used to allow for seamless annotation of variants with its oncogenic and biological effects and clinical implications when using the OncoKB API.

A. General Curation Rules

1. Multiple mutations may be grouped together (comma separated) for curation of shared clinical implications and/or tumor type summaries. The oncogenic and mutation effect of each of the mutations should be curated separately.
2. Mutation ranges, which capture all amino acid substitutions in a specified amino acid range, can be used (e.g., TP53 102_292mis [TP53 DNA binding domain mutations], KIT C788_N828mut [KIT Exon 17 non-truncating mutations]). Mutation ranges must have an associated oncogenic effect, mutation effect, and description of evidence based on the available evidence. Clinical implications and/or tumor type summaries can also be curated under mutational ranges.
3. ***Alteration Codes*** – the following are codes that can be used for naming alterations in the OncoKB curation platform:
    a. mis = missense mutation - e.g., 102_292mis [DNA binding domain missense mutations]
    b. dup = duplication of a specified range - e.g., S501_A502dup
    c. del = in-frame deletion of a specified range - e.g., P551_E554del
    d. ins = in-frame insertion - e.g., W557_V559delinsC; e.g.T574insTQLPYD
    e. delins = in-frame alteration - whether it's in-frame insertion or deletion, will be interpreted by the number of amino acid changes. e.g., V600_K601delinsE = inframe deletion - e.g., R435_K436delinsKKR = in-frame insertion
    f. nontrunc = any non-truncating mutation - e.g., R449_E514 nontrunc
    g. fs = frameshift - e.g., N457Mfs*22
    h. _splice = splice mutations - e.g., X963_D1010splice or X963_splice
    i. trunc = truncating mutation - e.g., D286_L292trunc
    j. 1? = start lost - e.g., M1?
    k. * = stop gained - e.g., R2019*
4. ***Brackets and Parentheses in the Mutation Header***
    a. Square Brackets [ ] - used in the mutation header to rename a curated alteration. For example, to curate a specific insertion, amino acid positions are written in the mutation header to indicate the protein change (e.g., 729_761ins). However, for the purpose of displaying this alteration on the OncoKB website, the SCMT may want to refer to this alteration as "Exon 19 insertion". By using square brackets in the mutation header as follows: "729_761ins [Exon 19 insertion]", the OncoKB website will display the alteration as "Exon 19 insertion" instead of 729_761ins.
    b. Parentheses () - used in the mutation header to leave comments. Any text in () in the mutation header is for administrative purposes only and can only be viewed within the OncoKB curation interface. It will not affect the output of how a mutation is displayed on any output platform (cBioPortal, MSK-IMPACT Reports or OncoKB Website).

B. Missense Mutations

1. The naming convention for missense mutations is <ref_allele><position><tumor_allele> (e.g., V600E)
2. Every missense mutation needs to be separately curated with respect to its oncogenic and mutation effect.
3. Positional variants, which capture all amino acid substitutions at a given position, can be used for curation of shared clinical implications and/or tumor type summaries (e.g., KRAS G12, BRAF V600). Positional variants do not include curation of oncogenic effect or mutation effect, as this information should be captured under each allele-specific missense mutation for which there is functional data.

C. Truncating Mutations

"Truncating Mutations" can be curated as a specific alteration within a Gene Page. "Truncating Mutations" must have an associated oncogenic effect, mutation effect, and description of evidence.

1. Since "Truncating Mutations" captures all truncating alterations within the gene (some of which have not been functionally characterized), its oncogenic and mutation effect should be marked as "Likely Oncogenic " and "Likely Loss of Function" respectively.
2. Clinical implications and/or tumor type summaries can also be curated under "Truncating Mutations."
3. The oncogenic effect, mutation effect and clinical implications associated with "Truncating Mutations" can be limited by defining a range for the truncation (e.g., "CCND1 256_286trunc [C Terminal Truncating Mutations]"). Truncating mutations outside this range will not be associated with the designated oncogenic effect, mutation effect and clinical implication of those in the defined range.
4. "Truncating Mutations" include the following based on the Sequence Ontology:
    a. Stop_lost: A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript
    b. Start_lost: A codon variant that changes at least one base of the canonical start codon
    c. Stop_gained: A sequence variant where at least one base of a codon is changed, resulting in a premature stop codon and leading to a shortened transcript
    d. TFBS_ablation: A feature ablation where the deleted region includes a transcription factor binding site
    e. Feature_truncation: A sequence variant that causes the reduction of a genomic feature, with regard to the reference sequence
    f. Frameshift_variant: A sequence variant which causes a disruption of the translational reading frame, i.e., the number of nucleotides inserted or deleted is not a multiple of three
    g. Transcript_ablation: A feature ablation whereby the deleted region includes a transcript feature
    h. Splice_donor_variant: A splice variant that changes the 2 base region at the 5' end of an intron
    i. Splice_region_variant: A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron
    j. Stop_retained_variant: A sequence variant where at least one base in the terminator codon is changed, but the terminator remains
    k. Splice_acceptor_variant: A splice variant that changes the 2 base region at the 3' end of an intron
    l. Incomplete_terminal_codon_variant: A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed.

D. Fusions

"Fusions" can be curated as a specific gene alteration within a Gene Page, and include any fusion that involves the specified gene.

1. "Fusions" must have an associated oncogenic effect, mutation effect, and description of evidence.
2. Since "Fusions" captures all fusions within the gene (some of which have not been functionally characterized), its oncogenic and mutation effect should be marked as "Likely Oncogenic " and "Likely Gain of Function" respectively.
3. Clinical implications and/or tumor type summaries can also be curated under "Fusions."
4. Specific fusions, in which both fusion partners are specified, can be curated as separate alterations if there is functional evidence in the literature describing their oncogenic and/or mutation effect (e.g., "EML4-ALK fusion"). The oncogenic effect, mutation effect, and clinical implications of the specific fusion alteration will be prioritized over those of the "Fusions" alteration.
5. Although a specific fusion names two gene partners, the alteration is only curated in one Gene Page - the gene that is the main driver (or hypothesized to be the main driver) of the fusion oncoprotein (e.g., BCR-ABL1 is curated in the ABL1 Gene Page).

E. Copy Number Aberrations

"Amplification" and "Deletion" can be curated as specific gene alterations within a Gene Page if appropriate functional data exists:

1. "Amplification" and "Deletion" must have an associated oncogenic effect, mutation effect, and description of evidence.

2. Prognostic implications, clinical implications and/or tumor type summaries can also be curated under “Amplification” and “Deletion.”

F. In-frame Deletions or Insertions

In-frame deletions or insertions can be curated as a specific gene alteration within a Gene Page (refer to section IV.E.1).

1. “del” = in-frame deletion (e.g., P551_E554del, P191del)
2. “ins” = in-frame insertion (e.g., T574insTQLPYD)
3. “delins” = a specified in-frame alteration. Whether the alteration is an in-frame deletion or in-frame insertion is determined by the specified number of amino acid changes. For example:
    a. *V600_K601delinsE* is an in-frame deletion because the number of amino acids deleted (2) is greater than the number of amino acids inserted (1).
    b. *R435_K436delinsKKR* is an in-frame insertion because the number of amino acids inserted (3) is greater than the number of amino acids deleted (1).
4. Each curated alteration must have an associated oncogenic effect, mutation effect, and description of evidence.
5. Clinical implications and/or tumor type summaries can also be curated under an in-frame deletion or insertion.

G. Oncogenic Mutations

“Oncogenic Mutations” can be curated as a specific gene alteration within a Gene Page.

1. “Oncogenic Mutations” is used when there is tumor-specific information that applies to ALL functional (oncogenic/likely oncogenic) alterations within a Gene Page. The tumor-specific information will automatically get linked to all mutations in the Gene Page that have the ” Yes” or “Likely” boxes checked next to the Oncogenic label.
2. “Oncogenic Mutations” does not include curation of oncogenic effect, mutation effect, and description of evidence, as this information should be captured under each individual variant in the Gene Page for which “Oncogenic Mutations” applies.
3. If a gene has “Amplification” curated as “Oncogenic” or “Likely Oncogenic”, this alteration will NOT be associated with the tumor-type specific information under “Oncogenic Mutations.”

H. Tumor Suppressors and Oncogenes

For genes marked as Tumor Suppressors:

1. The alteration “Truncating Mutations” should be curated.
2. The alteration “Deletion” may be curated, but this is dependent on the data available in the literature.
3. For Oncogenes: Truncating Mutations in oncogenes are often nonfunctional/not oncogenic. However, there are some examples in which they are functional including the genes CCND1 and CALRX. In these cases, truncating mutations in the protein are often activating via loss of C-terminal negative regulatory domains and in these cases, truncating mutations are restricted to a specific range.

I. Hard-coded Alteration Names

Alterations that do not follow the above nomenclature are not supported unless they are hard coded. Examples of such alterations include:

1. FLT3: internal tandem duplication
2. EGFR: vIII
3. EGFR: Kinase domain duplication
4. EGFR: C-terminal domain

J. Hotspot Mutations

Mutational hotspots are defined as mutant residues arising more frequently than expected in the absence of selection based on the analysis by Chang et al., 2018. In this analysis 24,592 cancers including 10,336 prospectively sequenced patients with advanced disease were analyzed, and the authors identified 1,165 statistically significant missense or in-frame insertion or deletion hotspot mutations, of which 80% arose in 1 in 1,000 or fewer patients.

1. If there is functional data in the literature describing the oncogenic and/or mutation effect of an allele-specific hotspot, the hotspot should be curated as an individual variant within the appropriate Gene Page.
2. Curated hotspots must have an associated oncogenic effect, mutation effect, and description of evidence based on the available evidence.
3. If no allele-specific variants are curated for a hotspot (including if variants are only located in the VUS section of the Gene Page), the hotspot's oncogenic effect will be automatically designated as "predicted oncogenic" in any output platform (cBioPortal, MSK-IMPACT Reports or OncoKB website).

II. Evidence-based Alteration Curation

Alterations included in OncoKB are genetic changes that arise as a result of DNA-level variants in cancer: non-synonymous mutations, translocations, rearrangements / fusions, copy number amplifications and deletions. This document uses "alterations", "mutations" and "variants" interchangeably. OncoKB describes alterations by their effect on the protein and not at the DNA level. All alterations in OncoKB are classified according to 1) their oncogenic effect and 2) their biological effect, based on the curated evidence (**Fig. 6**).

The oncogenic and biological effects of a mutation are curated based on the properties of transformed cells described in the second edition of "The Biology of Cancer" by Robert Weinberg and the hallmarks of cancer described by Douglas Hanahan and Robert Weinberg in their manuscript "Hallmarks of cancer: the next generation." published in Cell in 2011 (Hanahan and Weinberg, 2011).

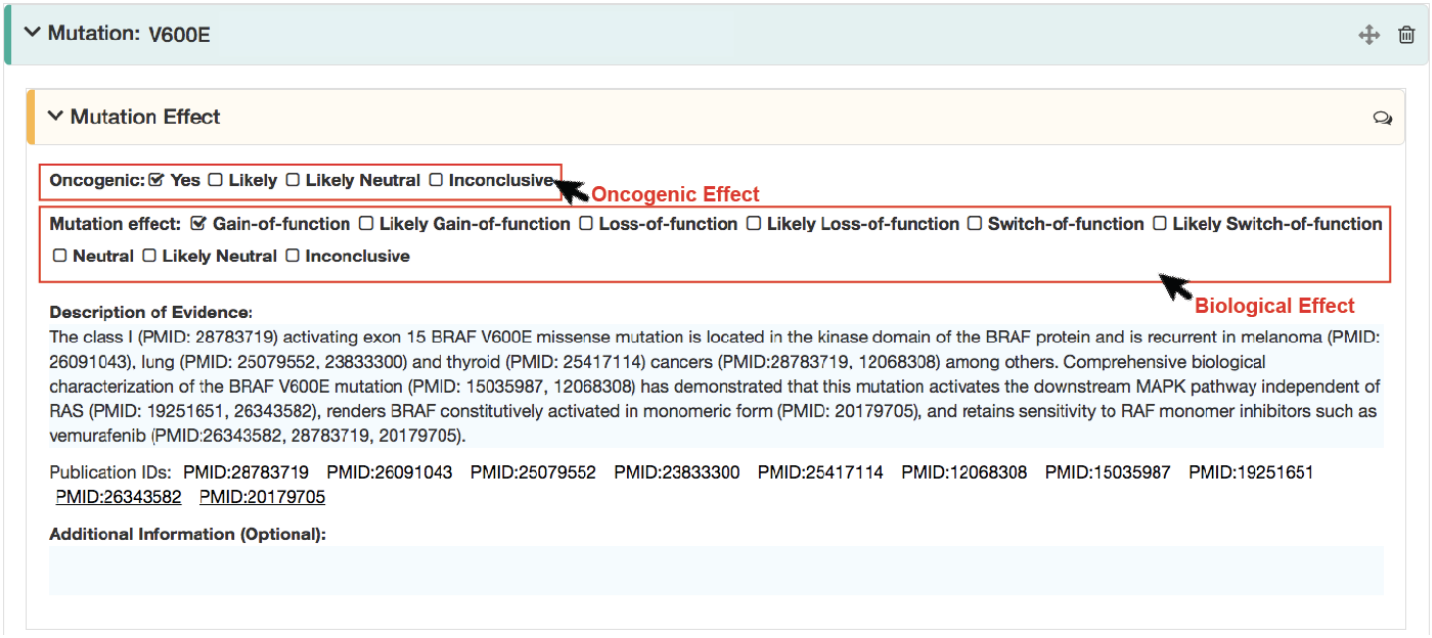


Figure 6: Curation of the Oncogenic and Biological effects of an alteration in OncoKB. An alteration is described by two assertions: 1) The Oncogenic Effect of the mutation and 2) The Biological Effect of the mutation. *Every variant in OncoKB must be curated with both of these assertions or placed in the Variants of Unknown Significance section of the curation platform. Otherwise entry of the variant is not allowed into the OncoKB database. *MSI-H and TMB are curated "alterations" in OncoKB that do not require an oncogenic and biological effect.

III. Defining the oncogenic effect of an alteration

In OncoKB, "oncogenic" is defined as "referring to the ability to induce or cause cancer" as described in the second edition of The Biology of Cancer by Robert Weinberg (2014). OncoKB distinguishes between five possible evidence-based assertions to describe the oncogenic effect conferred by the alteration when it is present in cells.

The following criteria is used to assert whether an alteration may be oncogenic, likely oncogenic, likely neutral or inconclusive and **Protocol #2** in the Appendix is used to determine this:

A. Oncogenic

Strong evidence shows that the alteration is established in the literature as promoting cell proliferation or other hallmark of cancer as defined by Douglas Hanahan and Robert Weinberg (Hanahan and Weinberg, 2011).

1. Compelling experimental data (e.g., genetically engineered mouse data with the mutation) in one or more studies directly demonstrating that the alteration is oncogenic and is associated with at least one hallmark of cancer as defined by Hanahan and Weinberg
2. The alteration is a known hotspot (Chang et al., 2018) AND there is at least one experimental study suggesting the alteration is oncogenic.
3. The alteration has been identified in a patient who responded to a targeted inhibitor, AND at least one experimental study provides strong evidence that the alteration is oncogenic.
4. The alteration is classified as either known gain/loss/switch-of-function AND there is at least one experimental study suggesting the alteration is oncogenic.

B. Likely Oncogenic (more permissive)

Evidence suggests the alteration likely promotes cell proliferation or other hallmark of cancer as defined by Douglas Hanahan and Robert Weinberg (Hanahan and Weinberg, 2011).

1. Representative experimental lines of data (e.g., downstream activation/inactivation of a signaling target/a hit in a high-throughput screen) in one or more studies pointing to possible oncogenic function or mutation associated with known germline syndrome.
2. At least one experimental study provides reasonable evidence suggesting the alteration is oncogenic.
3. The alteration is a known hotspot (Chang et al., 2018), AND there are no known functional studies describing the oncogenic potential of the alteration.
4. The alteration is classified as either known gain/loss/switch-of-function or likely gain/loss/switch-of-function AND there are no known functional studies describing the oncogenic potential of the alteration.

C. Likely Neutral

Evidence suggests the alteration does not alter protein activity or does not confer growth or survival advantage when expressed in cells.

1. The mutation effect of the alteration is neutral or likely neutral.
2. At least one experimental study provides reasonable evidence suggesting the alteration is likely neutral.

D. Inconclusive

There is conflicting and/or weak data describing the oncogenic effect of the mutant alteration

1. Conflicting data exists as to the oncogenic effect of the alteration.
2. Data is limited to “weak” experimental data describing the oncogenic effect of the alteration (small, under-powered experimental studies in one or multiple publications).
3. Data is limited to studies demonstrating either patient and/or in vitro sensitivity/resistance to a targeted drug.
4. Data is limited to in silico studies that predict the oncogenic effect of the alteration.

IV. Defining the biological effect of an alteration

In OncoKB, the Biological Effect is defined as the biological effect of a mutation/alteration on the protein function that gives rise to changes in the biological properties of cells expressing the mutant/altered protein compared to cells expressing the wildtype protein.

- Transformed cells are characterized by the following properties (Weinberg, p.82, Table 3.2, 2014):
    - Altered morphology (rounded shape, refractile in phase-contrast microscope)
    - Loss of contact inhibition (ability to grow over one another)
    - Anchorage independence (ability to grown without attachment to solid substrate)
    - Ability to proliferate indefinitely
    - Reduced requirement of mitogenic growth factors
    - High saturation density (ability to accumulate large numbers of cells in culture dish)
    - Inability to halt proliferation in response to deprivation of growth factors
    - Increase transport of glucose

    - Tumorigenicity (ability to form tumors in vivo following injection into appropriate host animals)
- The hallmarks of cancer comprise the biological capabilities acquired during the multistep development of human tumors. Mutations when expressed in cells may exhibit any one of these hallmarks of cancer in cells expressing the altered protein. Published experimental measurements of any of one these hallmarks of cancer may be taken as evidence that the mutation is oncogenic:
    - Sustaining proliferative signaling
    - Evading growth suppressors
    - Resisting cell death
    - Enabling replicative immortality
    - Inducing angiogenesis
    - Activating invasion and metastasis
    - Genome instability and mutation
    - Tumor-promoting inflammation
    - Deregulated cellular energetics
    - Evading immune destruction

OncoKB distinguishes between five possible evidence-based assertions to describe the biological effect conferred by the alteration when it is present in cells. An alteration is asserted as known or likely gain-, loss-, or switch-of-function, neutral, likely neutral, or inconclusive based on the following criteria using **Protocol #3**.

A. Known Gain/Loss/Switch-of-function

1. **Gain-of-function**: Strong evidence-based data demonstrating that the alteration increases the function of the protein, specifically:
    a. The alteration is associated with increased function of the protein
    b. Increased gene dosage
    c. Increased/ectopic mRNA expression
    d. Increased/constitutive protein activity
    e. Dominant negative
    f. Structural protein
    g. Toxic protein

2. **Loss-of-function**: Strong evidence-based data demonstrating that the alteration decreases the function of the protein, specifically:
    a. The alteration is associated with decreased function of the protein
    b. Haploinsufficiency

3. **Neutral:** Strong evidence-based data demonstrating that the function of the protein is unchanged by the alteration, specifically:
    a. The function of the protein is unchanged by the alteration
    b. There is no difference in measurable cell attributes expressing either the wildtype or mutant form of the gene.

4. **Switch-of-function**: Strong evidence-based data demonstrating that the alteration causes the protein to acquire a new function, specifically:
    a. The alteration is associated with a novel function of the protein
    b. New protein
    c. Altered substrate specificity

5. **Rules for classifying an alteration with a known function**
    a) Compelling experimental data in one or more studies directly establishing the function of the mutation.
    b) Multiple lines of data in one or more studies including but not limited to experimental data and statistical recurrence that together provide strong evidence establishing the function of the mutation.
    c) The alteration is a known hotspot (Chang et al., 2018) AND at least one experimental study provides strong evidence that the alteration confers gain-, loss-, or switch-of-function.

d) The alteration has been identified in a patient who responded to a targeted inhibitor AND at least one experimental study provides strong evidence that the alteration confers gain-, loss-, or switch-of-function.
e) Strong evidence-based data demonstrating that there is no difference in measurable cell attributes expressing either the wildtype or mutant form of the gene (Neutral)

B. Likely Gain/Loss/Switch-of-function

1. **Likely Gain-of-function**: Probable, possible, and/or evidence-based data suggesting that the alteration likely increases the protein function

2. **Likely Loss-of-function**: Probable, possible, and/or evidence-based data suggesting that the alteration likely decreases the protein function

3. **Likely Switch-of-function:** Probable, possible, and/or evidence-based data suggesting that the alteration likely causes the protein to acquire a new function

4. **Rules for classifying an alteration with a probable function**
   a) A single or multiple experimental studies from one publication including but not limited to experimental data or statistical recurrence establishing the function of the mutation
   b) The alteration is a known hotspot (Chang et al., 2018), and there are no known functional studies describing the mutation effect of the alteration.
   c) While conflicting evidence may exist, there is a reasonable assumption based on the data suggesting the alteration confers gain-, loss-, or switch-of or neutral function.
   d) The alteration has been identified in a patient who responded to a targeted inhibitor AND at least one experimental study provides limited evidence that the alteration confers gain-, loss-, or switch-of-function
   e) Probable, possible, and/or evidence-based data suggesting that there is no difference in measurable cell attributes expressing either the wildtype or mutant form of the gene (Likely neutral).

C. Inconclusive

There is conflicting and/or weak data describing the mutation effect of the alteration:

1. Conflicting data exists as to the mutational effect of the alteration.
2. Data is limited to "weak" experimental data describing the mutational effect of the alteration (small, under-powered experimental studies in one or multiple publications).
3. Data is limited to studies demonstrating patient and/or in vitro sensitivity/resistance to a drug.
4. Data is limited to in silico studies that predict the mutation effect of the alteration.

V. Tumor Type Curation

Tumor Type

Below each alteration in the curation interface, the user must choose one or multiple Tumor Type(s) for the purpose of curating alteration- and tumor type-specific clinical implications, if any. OncoKB uses OncoTree (http://oncotree.mskcc.org) to manage the vocabulary of tumor types. Currently OncoTree version 2019_12_01 is being used. The user may choose a main cancer type and/or subtype from the dropdown list. In addition to the Oncotree nodes, the dropdown list also contains the following categories:

A. *All Solid Tumors*: Includes all solid tumors within the Oncotree
B. *All Liquid Tumors*: Includes all liquid tumors (from the myeloid and lymphoid branches) within Oncotree
C. *All Tumors*: Includes all solid and liquid tumors within the Oncotree
D. *Other Tumors*: This tumor classification is a special case and is only utilized for the purpose of incorporating Tumor Type Summaries.

Chapter 5: Curation of Tumor Type-Specific Clinical Implications

A subset of alterations in OncoKB are considered biomarkers that are predictive of response to certain drugs. Some of these drugs are FDA-approved and the biomarker is a consideration in standard care. Alternatively, some of these drugs are either 1) FDA-approved, but the biomarker is in an off-label setting or 2) not FDA-approved and instead are being tested in clinical trials. In both of the latter scenarios, the biomarkers and drugs are considered investigational.

The original Levels of Evidence system was developed by OncoKB to rank the therapeutic implications associated with an alteration found in a patient tumor sample by the relative weight of the evidence (Chakravarty et al., 2017). On December 20, 2019, the Levels of Evidence were refined and simplified to be consistent with the Joint Consensus Recommendation by AMP, ASCO and CAP and the ESMO Scale for Clinical Actionability of molecular Targets (ESCAT) and to reflect the clinical data that demonstrates patients with investigational predictive biomarkers for a specific tumor type based on compelling clinical evidence (Level 3A) are more likely to experience clinical benefit compared to patients with predictive biomarkers that are considered standard care in a different tumor type (previously Level 2B, combined into Level 3B) (**Fig. 7**).

For example, an alteration that is recognized by the FDA to be predictive of response to an FDA-approved drug would have a higher Level of Evidence (Level 1) compared to an alteration that has been shown in preclinical studies to be sensitizing to an investigational drug that is being tested in a clinical trial (Level 4). Accordingly, the highest levels of evidence, Levels 1 and 2 refer to the standard implications for sensitivity to an FDA-approved drug. Additionally, Level R1 refers to the standard implications for resistance to an FDA-approved drug. Levels 3A, 3B and 4 refer to the investigational implications for sensitivity to either an FDA-approved or investigational drug (in the off-label setting, Level 3B) or an investigational drug (Levels 3A and 4). Level R2 includes investigational implications for resistance to either an FDA-approved or investigational drug. **Since the FDA does not endorse off-label use of drugs, the scope of FDA-recognition sought for the clinical implications of OncoKB is restricted for Level 1 (FDA-recognized variants that are biomarkers predictive of response to FDA-approved drugs), Level 2 (NCCN-listed variants that are biomarkers predictive of response to FDA-approved drugs) and Level 3 (Compelling clinical evidence supports the biomarker as being predictive of response to a drug in this indication)** variants only **(Refer to Appendix IV an V, Protocols #4A and #4B).** Each of these different sets of clinical implications are described in greater detail in Sections IV to VII below.

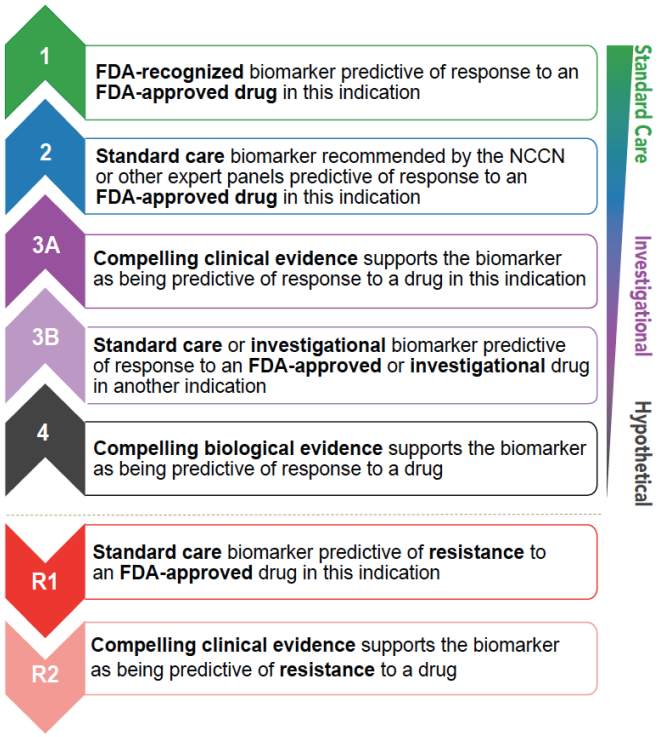


Figure 7: OncoKB (Therapeutic) Levels of Evidence.

Similarly, to rank the diagnostic and prognostic implications of an alteration found in a specific tumor type, the OncoKB Diagnostic and Prognostic Levels of Evidence schema were developed (refer to **Figs. 9 and 10**).

Updating Level of Evidence Assertions of Clinically Actionable Variants

CGAC members are responsible for advising the OncoKB team and entering into consensus regarding the assignment of a level of evidence to a biomarker. Requests for advice and consensus from CGAC occur in the form of periodic emails from the Lead Scientist to all CGAC members and are typically prompted by new FDA-approvals, FDA-breakthrough designations, or newly reported results of major clinical trials from clinical oncology conferences or publications.

Consensus emails have the following structure:

1. A statement describing the reason for a proposed new assignation of a level of evidence to an alteration or for changing the current level of evidence for a specific alteration and consequent change to OncoKB data.
2. A summary of the clinical data supporting the proposed assignation of a Level of Evidence to a specific alteration.
3. A sample Clinical Summary that includes the new OncoKB statement that is prompted by the new clinical data.
4. A request for feedback regarding the change to OncoKB data, in the form of a response within 5 business days of receipt of the request.

In order for a proposed change in the level of evidence to be approved, there are at minimum 3 affirmative verifications that must be received from CGAC, specifically the following CGAC members:
1) From the Director of the Center for Molecular Oncology, Dr. David Solit
2) From a Disease Management Team (DMT) Chief in the indication of the proposed level of evidence change
3) A miscellaneous member of CGAC

After review by 3 CGAC members the change in the level of evidence is further reviewed by a SCMT member and the OncoKB Lead Scientist following the process outlined in Chapter 5, Section IX. “Data Review” before it is finalized and released into public-facing OncoKB outputs (i.e., cBioPortal, oncokb.org and MSK patient reports).

Once a change is approved, it is entered into the OncoKB database, the outputs of which will be seen in the Clinical Summaries in the website, the cBioPortal and the MSK-IMPACT reports (refer to Chapter 1, “OncoKB Access”).

In the event that consensus cannot be immediately reached, the Lead Scientist is responsible for mediating between conflicting advice to resolve any discrepancy. Should consensus not be reached, the proposed change in the Level of Evidence is rejected.

Members of CGAC who may have COI with respect to the introduction or change of the levels of evidence assigned to a specific variant are allowed to provide advice and information regarding the assertion, but are excluded from the 3 CGAC member verification committee. Additionally, moving forward, for each change or introduction of a new level of evidence, the “News” announcement at www.oncokb.org will now include the names of the CGAC members that affirmatively verified the change, in addition to the names of any CGAC members who have a specific COI regarding the change or new leveled association.

The clinical implications of an alteration may be curated in one or more of seven sections (summarized in **Fig. 8**):

1. Tumor Type Summary
2. Diagnostic Implications
3. Prognostic Implications
4. Standard Implications for Sensitivity to Therapy
5. Standard Implications for Resistance to Therapy
6. Investigational Implications for Sensitivity to Therapy
7. Investigational Implications for Resistance to Therapy

Figure 8. The Clinical Implications of the mutation. If a mutation has a clinical implication, it is described within the context of the tumor type in which the clinical implication is relevant. If a mutation has a diagnostic clinical implication, it must be associated with a Diagnostic Level of Evidence. Similarly, if the tumor type-specific clinical implication is prognostic or therapeutic, it must be associated with a Prognostic or Therapeutic Level of Evidence respectively.

I. Clinical Summary

The clinical implications of an alteration is summarized in 1-2 sentences. These sentences describe the therapeutic, diagnostic and/or prognostic implications for alterations with a level of evidence. This section is free text codes may be used for curating tumor type summary in order to include patient's variant and tumor type in the sentence, since they may be different from the curated data, e.g., V600E in patient will be matched to V600.

A. [[variant]]: "gene" "mutation" mutant "tumor type" - e.g., BRAF V600E mutant melanoma
B. [[tumor type]]: "tumor type" - e.g., melanoma
C. [[gene]] - Adds the "gene" name - e.g., BRAF
D. [[mutation]] - Adds the "mutation" name - e.g., V600E
E. [[mutation]] [[mutant]] - Adds: "mutation" name and "mutant" - e.g., V600E mutant

II. Diagnostic Implications

The purpose of this section is to curate alterations which have tumor type specific diagnostic implications.

A. Level of Evidence

This section includes a drop-down list that allows a curator to choose the appropriate diagnostic Level of Evidence associated with the alteration in a specific tumor type. The drop-down list includes the following choices (**Fig. 9**):

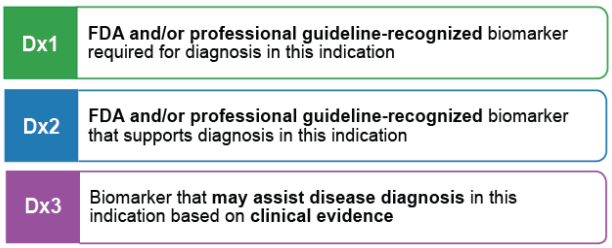


Figure 9: OncoKB Diagnostic Levels of Evidence Schema.

1. *Dx1* defined as "*FDA and/or professional guideline-recognized biomarker required for diagnosis in this indication.*"
2. *Dx2* defined as "*FDA and/or professional guideline-recognized biomarker that supports diagnosis in this indication.*"

3. *Dx3* defined as "*Biomarker that may assist disease diagnosis in this indication based on clinical evidence.*"

B. Description of Evidence

This section is free text and contains 4-6 sentences and describes an overview and results from clinical studies describing the prevalence of the gene-alteration in the specified disease including the cohort size, the genetic criteria for patient selection, and the total number and percent of patients with the specified gene-alteration.

C. Additional Information (Optional)

Provides the curator with space to add additional information and/or references to information that may be of relevance to the SCMT, but may not necessarily be included in the final output. The information in this section will only be accessible from the OncoKB curation interface and therefore will not be displayed on other platforms (ie. OncoKB website, cBioPortal, MSK-IMPACT Reports). This section is free text.

III. Prognostic Implications

The purpose of this section is to curate alterations which have tumor type specific prognostic implications.

A. Level of Evidence

This section includes a drop-down list that allows the user to choose the appropriate prognostic Level of Evidence associated with the alteration in a specific tumor type The drop-down list includes the following choices (**Fig. 10**):

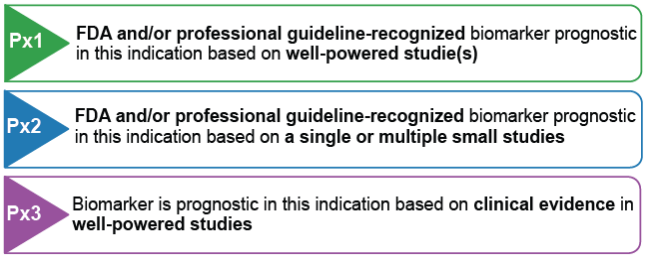


Figure 10: OncoKB Prognostic Levels of Evidence Schema.

1. *Px1* defined as "*FDA and/or professional guideline-recognized biomarker prognostic in this indication based on well-powered studies.*"
2. *Px2* defined as "*FDA and/or professional guideline-recognized biomarker prognostic in this indication based on a single or multiple small studies.*"
3. *Px3* defined as "*Biomarker is prognostic in this indication based on clinical evidence in well-powered studies.*"

B. Description of Evidence

An overview and results from clinical studies describing the prognostic implications of the gene-alteration in the specified disease including the cohort size, the genetic criteria for patient selection, the percent of patients with and without the specified gene-alteration, and the endpoints used to predict clinical benefit or harm (e.g., overall survival (OS), progression-free survival (PFS), disease-free survival (DFS) and associated p-values). This section is free text and contains 4-6 sentences.

C. Additional Information (Optional)

Provides the curator with space to add additional information and/or references to information that may be of relevance to the SCMT, but may not necessarily be included in the final output. The information in this section will only be accessible from the OncoKB curation interface and therefore will not be displayed on other platforms (ie. OncoKB website, cBioPortal, MSK-IMPACT Reports). This section is free text.

IV. Standard Implications for Sensitivity to Therapy

The standard therapeutic implications for sensitivity of alterations that are FDA- or NCCN- recognized as biomarkers predictive of response to FDA-approved therapies in specific tumor types are curated in this section (refer to **Fig. 7**). Here, a curator can enter the name of the standard sensitivity therapy in the "Therapy:" box. Therapies are chosen from a drop-down list linked to https://clinicaltrialsapi.cancer.gov/#/Interventions which provides standardized nomenclature for drugs. Once a therapy is entered, the following sections become available for curation:

A. Highest Level of Evidence

This section includes a drop-down list that allows the user to choose the appropriate standard Level of Evidence. The drop-down list includes the following choices (refer to **Fig. 7**):

1. *Level 1* defined as "*FDA-recognized biomarker predictive of response to an FDA-approved drug in this indication.*"
2. *Level 2* defined as "*Standard care (NCCN or other expert panels) biomarker predictive of response to an FDA-approved drug in this indication.*"

B. Level of Evidence in Other Tumor Types

Alterations that are Level 1 or 2 in a specified tumor type may or may not be considered Level 3B or Level 4 in other solid or other liquid tumor types. Whether to propagate a Level 1 or 2 indication to Level 3B or Level 4 in other solid and/or other liquid tumors is at the discretion of the SCMT and Lead Scientist and is based on the scientific literature.

This section includes two drop-down lists (one for solid tumors and one for liquid tumors) that allows the user to decide if the investigational therapy evidence should be propagated to Level 3B, Level 4 or No Level in i) other solid tumor types or 2) other liquid tumor types. The drop-down lists includes the following choices: (refer to **Fig. 7**):

1. *Level 3B* defined as "*Standard care or investigational biomarker predictive of response to an FDA-approved or investigational drug in another indication.*"
2. *Level 4:* defined as "*Compelling biological evidence supports the biomarker as being predictive of response to a drug*."
3. *No Level:* The curated standard therapeutic evidence will not be propagated to other tumor types. Therefore, if the gene-alteration combination is found in a tumor-type other than the one specified, it will not receive a Level of Evidence.

Level 3B evidences are not curated directly into OncoKB, but can be propagated from Level 1, 2, or 3A evidences to all other solid tumors or all other liquid tumors when the SCMT member specifically chooses to do so based on the scientific evidence and discussion with the Lead Scientist. Whether or not to propagate these associations involve a discussion with CGAC, as outlined above (refer to Chapter 5, "Updating Level of Evidence Assertions of Clinically Actionable Variants").

Level 1, 2 and 3A associations in solid tumors propagate to Level 3B in other solid tumors unless there is negative or conflicting evidence, in which case the association would propagate to Level 4 or No Level in other solid tumors in accordance with the evidence. Level 1, 2 and 3A associations in solid tumors do not propagate to liquid tumors unless there is specific scientific evidence to support the association as Level 3B or Level 4 in liquid tumors. Level 1, 2 and 3A associations in liquid tumors do not propagate to other solid or other liquid tumors unless there is specific scientific evidence to support the association as Level 3B or Level 4 in these tumor types.

C. Description of Evidence

This section is 4 to 6 sentences, consisting of free text that describes the following:

1. The therapy and its targets.
2. Overview and results from clinical studies testing the drug in patient populations including the cohort size, the genetic criteria for patient selection, and the results of the study (e.g., response rates and statistical analysis).

3. Description and results from studies testing the therapy in in vitro and/or in vivo models, if relevant. For Level 1 and 2 therapies, the curated studies reflect those referenced by the FDA and/or NCCN Compendium.

D. Additional Information (Optional)

Provides the curator with space to add additional information and/or references to information that may be of relevance to the SCMT, but may not necessarily be included in the final output. The information in this section will only be accessible from the OncoKB curation interface and therefore will not be displayed on other platforms (i.e., OncoKB website, cBioPortal, MSK-IMPACT Reports). This section is free text.

E. Updating Level of Evidence 1

The SCMT closely monitors all new FDA drug approvals in the Hematology/Oncology (Cancer) Approvals and Safety Notifications via updates received directly from the FDA by email from fda@info.fda.gov. When the FDA announces a new drug approval the SCMT immediately reviews and flags the FDA drug label specified genetic alteration as a potential OncoKB Level 1 alteration. Subsequently, the Lead Scientist sends a consensus email to CGAC seeking at minimum 3 affirmative verifications regarding the new level of evidence assignment (refer to Chapter 5, "Updating Level of Evidence Assertions of Clinically Actionable Variants", pg 25). Five business days after the consensus email is sent (during which time CGAC responses are received), if the level of evidence assignment has been approved, the data is entered into the OncoKB curation platform. Once entered, the data must go through one final round of review/quality control (QC) by the Lead Scientist or member of the SCMT who did not directly enter the data into the OncoKB curation platform (refer to Chapter 5, Section IX).

F. Updating Level of Evidence 2

Quarterly, the SCMT carefully reviews the NCCN Guidelines for Treatment of Cancer by Site (https://www.nccn.org/professionals/physician_gls/default.aspx#site). Guidelines that have been updated since the last review period are assessed, and alterations associated with an NCCN recommendation at category 2A or higher are flagged by the SCMT as potential OncoKB Level of Evidence 2. Upon notification by the SCMT, the Lead Scientist sends a consensus email to CGAC seeking affirmative verification regarding the new level of evidence assignment (refer to Chapter 5, "Updating Level of Evidence Assertions of Clinically Actionable Variants", pg 25). Five business days after the consensus email is sent (during which time CGAC responses are received), if the level of evidence assignment has been approved, the data is entered into the OncoKB curation platform. Once entered, the data must go through one final round of review/QC by the Lead Scientist or member of the SCMT who did not directly enter the data into the OncoKB curation platform (refer to Chapter 5, Section IX).

V. Standard Implications for Resistance to Therapy

The standard therapeutic implications for resistance of alterations that are NCCN- recognized as biomarkers predictive of resistance to FDA-approved therapies in specific tumor types are curated in this section (refer to **Fig. 7**). Here, a curator can enter the name of the standard resistance therapy in the "Therapy:" box. Therapies are chosen from a drop-down list linked to https://clinicaltrialsapi.cancer.gov/#/Interventions which provides standardized nomenclature for drugs. Once a therapy is entered, the following sections become available for curation:

A. Level R1

The highest and only standard level of resistance, Level R1. It is defined as "*Standard care biomarker predictive of resistance to an FDA-approved drug in this indication.*"

B. Description of Evidence

This section is free text and contains 4-6 sentences that describes the following:

1. The drug and its genetic targets.
2. Overview and results from clinical studies and/or case studies documenting resistance to the therapy.
3. The following information should be documented from clinical studies: cohort size, the genetic criteria for patient selection, and the clinical results of the study (e.g., response rates and statistical analysis).

4. Description and results from studies documenting resistance to the therapy in in vitro and/or in vivo models, if relevant. For Level R1 therapies, the curated studies reflect those referenced by the NCCN Compendium.

C. Additional Information (Optional)

Provides the curator with space to add additional information and/or references to information that may be of relevance to the SCMT, but may not necessarily be included in the final output. The information in this section will only be accessible from the OncoKB curation interface and therefore will not be displayed on other platforms (ie. OncoKB website, cBioPortal, MSK-IMPACT Reports). This section is free text.

VI. Investigational Implications for Sensitivity to Therapy

The investigational therapeutic implications for sensitivity of alterations for which there is published clinical (Level 3A) or preclinical (Level 4) data supporting the alteration as a predictive biomarker of response to an investigational therapy in specific tumor types are curated in this section (refer to **Fig. 7**). Here, a curator may enter the name of the investigational sensitivity therapy in the "Therapy:" box. Therapies are chosen from a drop-down list linked to https://clinicaltrialsapi.cancer.gov/#/Interventions, which provides standardized nomenclature for drugs. Once a therapy is entered, the following sections become available for curation:

A. Highest Level of Evidence

This section includes a drop-down list that allows the user to choose the appropriate investigational Level of Evidence. The drop-down list includes the following choices:

1. *Level 3A* defined as "*Compelling clinical evidence supports the biomarker as being predictive of response to a drug in this indication, but neither biomarker nor drug are standard care.*"
2. *Level 4 is defined* as "*Compelling biological evidence supports the biomarker as being predictive of response to a drug, but neither biomarker nor drug are standard care.*"

B. Level of Evidence in Other Tumor Types

Alterations that are Level 3A in a specified tumor type may or may not be considered Level 3B or Level 4 in other solid or other liquid tumor types. Whether to propagate a Level 3A indication to Level 3B or Level 4 in other solid and/or other liquid tumors is at the discretion of the SCMT and Lead Scientist and is based on the scientific literature.

This section includes two drop-down lists (one for solid tumors and one for liquid tumors) that allows the user to decide if the investigational therapy evidence should be propagated to Level 3B, Level 4 or No Level in i) other solid tumor types or 2) other liquid tumor types. The drop-down lists includes the following choices: (refer to **Fig. 7**):

1. *Level 3B* defined as "*Standard care or investigational biomarker predictive of response to an FDA-approved or investigational drug in another indication.*"
2. *Level 4:* defined as "*Compelling biological evidence supports the biomarker as being predictive of response to a drug*."
3. *No Level:* The curated standard therapeutic evidence will not be propagated to other tumor types. Therefore, if the gene-alteration combination is found in a tumor-type other than the one specified, it will not receive a Level of Evidence.

Level 3B evidences are not curated directly into OncoKB, but can be propagated from Level 1, 2, or 3A evidences to all other solid tumors or all other liquid tumors when the SCMT member specifically chooses to do so based on the scientific evidence and discussion with the Lead Scientist. Whether or not to propagate these associations involve a discussion with CGAC, as outlined above (refer to Chapter 5, "Updating Level of Evidence Assertions of Clinically Actionable Variants").

Level 1, 2 and 3A associations in solid tumors propagate to Level 3B in other solid tumors unless there is negative or conflicting evidence, in which case the association would propagate to Level 4 or No Level in other solid tumors in accordance with the evidence. Level 1, 2 and 3A associations in solid tumors do not propagate to liquid tumors unless there is specific scientific evidence to support the association as Level 3B or Level 4 in

liquid tumors. Level 1, 2 and 3A associations in liquid tumors do not propagate to other solid or other liquid tumors unless there is specific scientific evidence to support the association as Level 3B or Level 4 in these tumor types.

C. Description of Evidence

This section is free text and contains 4-6 sentences that describes the following:

1. The drug and its targets.
2. Overview and results from clinical studies and/or case studies testing the drug in patient populations including the cohort size, the genetic criteria for patient selection, and the results of the study (e.g., response rates and statistical analysis) (Level 3A only).
3. Description and results from studies testing the therapy in in vitro and/or in vivo models.

D. Additional Information (Optional)

Provides the curator with space to add additional information and/or references to information that may be of relevance to the SCMT, but may not necessarily be included in the final output. The information in this section will only be accessible from the OncoKB curation interface and therefore will not be displayed on other platforms (ie. OncoKB website, cBioPortal, MSK-IMPACT Reports). This section is free text.

E. Updating Investigational Levels of Evidence 3 and 4

Assertions of levels of evidence 3 or 4 to variants are incorporated from multiple different sources as described below:

1) Proceedings of major scientific and/or clinical conferences

Each year at least one member of the SCMT attends the following conferences: American Association for Cancer Research (ACCR), American Society of Clinical Oncology (ASCO), the European Society for Medical Oncology (ESMO) Congress, the American Society of Hematology (ASH) Annual Meeting, and the European Organisation for Research and Treatment of Cancer (EORTC)-NCI-AACR Symposium on Molecular Targets and Cancer Therapeutics, where information from oral presentations, posters and abstracts are assessed and flagged if the data could support a biomarker as being a leveled OncoKB alteration. Within two weeks following the conference, the data is compiled and analyzed in greater detail, and the SCMT notifies the Lead Scientist of any gene-biomarker-tumor type indications that might qualify for an OncoKB level of evidence (Sensitivity Levels 1-4 and Resistance Levels R1 and R2) based on the definitions outlined in Fig. 7.

Additionally, the SCMT reviews published highlights, abstracts and updates from various disease-specific conferences within one month following publication of the conference proceedings. These include but are not limited to: The San Antonio Breast Cancer Symposium, The World Conference on Lung Cancer, The AACR Special Conference on Melanoma, and The AACR Gastrointestinal Cancer Symposium. The SCMT notifies the Lead Scientist of any gene-biomarker-tumor type indications that might qualify for an OncoKB level of evidence.

2) The general scientific literature accessed through PubMed

The SCMT performs weekly literature reviews of high-impact journals including but not limited to: The New England Journal of Medicine, Cell, Cancer Cell, Cancer Discovery, JAMA, JAMA Oncology, Journal of Clinical Investigation, Nature, Lancet, Lancet Oncology, Cancer Research, Clinical Cancer Research, Journal of Clinical Oncology (JCO), JCO-Precision Medicine, Annals of Oncology, Lancet, Science, and Blood. Each week, the SCMT reviews the Table of Contents of newly published issues from these journals and flags articles to further assess. Every two weeks, a member of the SCMT team critically reviews the curated list of articles, and notifies the Lead Scientist of any gene-biomarker-tumor type indications that might qualify for an OncoKB level of evidence.

When critically assessing sources form 1 and 2 above, the SCMT specifically looks for new information on: 1) cancer genes, 2) cancer-associated alterations, 3) clinical trial results related to biomarker-specific patient responses and 4) biomarker-associated drug studies in the preclinical setting where the biomarker comprises an eligibility criteria in a currently open and recruiting clinical trial. 3 and 4 above comprise data related to potential Level 3 and Level 4 indications.

3) Recommendations from CGAC

Members of CGAC are in frequent contact with the Lead Scientist and can nominate gene-alteration-tumor type-drug associations for Level 3 or 4 status based on their knowledge and expertise in the field. As detailed in Chapter 1, "OncoKB Oversight and Governance", members of CGAC are at the forefront of clinical management and research and have translational cancer biology expertise in their respective major disease entities. Therefore, CGAC members have first-hand knowledge of new biomarker-tumor type-drug associations that may qualify for an OncoKB level of evidence, specifically those that may qualify as a Level 3A/3B or Level 4 association since qualification for these levels is based on clinical trial enrollment criteria, preclinical biomarker-drug studies, and results from case studies and larger clinical trials. If a CGAC member proposes a gene-biomarker-tumor type indication for an OncoKB level of evidence, the SCMT immediately reviews the data to determine the appropriate level classification (if any) and provides the Lead Scientist with the findings.

4) Recommendations from OncoKB users

There are various mechanisms for users to provide feedback to the OncoKB team (refer to Chapter 7, Section II.L.11 and Section V.B.6 and Fig. 34 and Fig. 40). If a user proposes a new or update to an OncoKB leveled association, the SCMT immediately reviews the data to determine the appropriate level classification (if any) and notifies the Lead Scientist with the findings.

Considering the various data sources outlined in 1-4 above, the SCMT team is continually analyzing and reviewing data that may qualify a gene-alteration-tumor-type-drug association as a Level 3A or Level 4 indication. A detailed SOP including granular rules for mapping variants to the OncoKB levels of evidence (including Levels 3A and 4) are outlined in **Protocol #4**. Once the SCMT flags a gene-biomarker-tumor type-drug indication for Level 3A or 4 status, the Lead Scientist sends a consensus email to CGAC seeking affirmative verification regarding the new level of evidence assignment (refer to Chapter 5, "Updating Level of Evidence Assertions of Clinically Actionable Variants", pg 25). Five business days after the consensus email is sent (during which time CGAC responses are received), if the level of evidence assignment has been approved, the data is entered into the OncoKB curation platform. Once entered, the data must go through one final round of review/QC by the Lead Scientist or member of the SCMT who did not directly enter the data into the OncoKB curation platform (refer to Chapter 5, Section IX).

VII. Investigational Implications for Resistance to Therapy

The investigational therapeutic implications for resistance of alterations are those for which there is compelling clinical data that supports that the alteration may serve as a biomarker predictive of resistance to FDA-approved or investigational therapies in specific tumor types are curated in this section (refer to **Fig. 7**). Here, a curator may enter the name of the investigational resistance therapy in the "Therapy:" box. Therapies are chosen from a drop-down list linked to https://clinicaltrialsapi.cancer.gov/#/Interventions which provides standardized nomenclature for drugs. Once a therapy is entered, the following sections become available for curation:

A. Level R2

The highest and only investigational level of resistance, Level R2. It is defined as "*Compelling clinical evidence supports the biomarker as being predictive of resistance to a drug.*"

B. Description of Evidence

This section is free text and contains 4-6 sentences that describes the following:

1. The drug and its targets.
2. Overview and results from clinical studies and/or case studies (if applicable) documenting resistance to the therapy.
3. The following information should be documented from clinical studies: cohort size, the genetic criteria for patient selection, and the clinical results of the study (e.g., response rates and statistical analysis).
4. Description and results from studies documenting resistance to the therapy in in vitro and/or in vivo models.

C. Additional Information (Optional)

Provides the curator with space to add additional information and/or references to information that may be of relevance to the SCMT, but may not necessarily be included in the final output. The information in this section will only be accessible from the OncoKB curation interface and therefore will not be displayed on other platforms (ie. OncoKB website, cBioPortal, MSK-IMPACT Reports). This section is free text.

VIII. Variants of Unknown Significance (VUS)

VUS are added to a unique section within the OncoKB Gene Curation Page called “Variants of Unknown Significance (Investigated and data not found)” (**Fig. 11**). Once a VUS is entered, it is linked to a timestamp displaying the date the VUS was last edited. If a VUS on the Gene Curation Page is investigated at a future date and still no data is found, the “Refresh” button can be clicked to update the timestamp associated with the VUS in question.

VUS are alterations for which limited or no information is publicly available and falls into one of three possible classes:

1. No data exists.
2. The variant has been identified within a tumor, but not functionally tested (in this case, the comment bubble for each variant lists the appropriate publications for SCMT reference).

A VUS on the Gene Curation Page entered:

1. Grey = Curated < 3 months prior to the current date.
2. Yellow = Curated 3 > 6 months prior to the current date.
3. Red = Curated > 6 months prior to the current date.

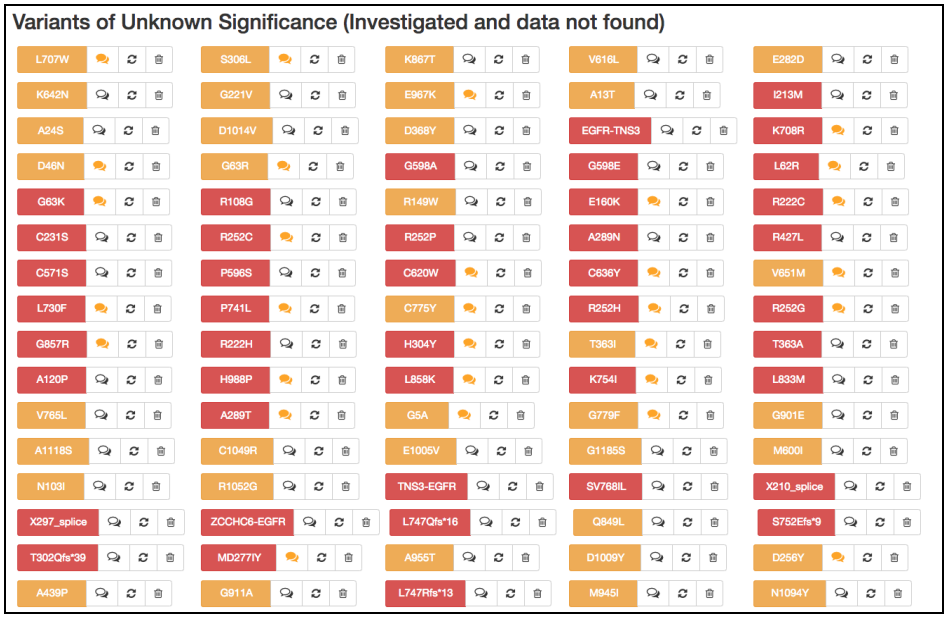


Fig. 11: Variants of Unknown Significance section in OncoKB Curation Platform.

IX. Data Review

All new content (including any updates, additions or deletions) that is entered into the OncoKB curation platform MUST go through a final review/QC before it is finalized and released into public-facing OncoKB outputs (i.e., cBioPortal, oncokb.org and MSK-patient reports). This is implemented through the Review function on the OncoKB curation platform. The OncoKB Curation Interface Homepage lists each gene and whether or not that gene has data to be reviewed. Each gene page on the curation platform has a ‘Review’ button that leads to the Review Page. The ‘Review’ button and Review page are only accessible to the SCMT and Lead Scientist of the OncoKB team. Data entries and deletions made on the gene page are NOT considered final (and therefore not released to OncoKB

public facing outputs) until they are reviewed and accepted on the Review Page by a member of the SCMT or the Lead Scientist who did NOT directly enter that change into the OncoKB curation platform.

The Review page records and stores all data entries and deletions that were made on the corresponding gene page. It provides the following information: 1) the location on the gene page where the data edit was made, 2) the exact text that was added, modified or deleted, 3) the name of the person who made the data entry or deletion, 4) the date and time the edit was made, and 5) a button to accept or reject the change. The Review page allows every discrete piece of information to be separately reviewed and accepted or rejected by the reviewer (**Fig. 3**). If a data edit or entry is high priority, the SCMT (or Lead Scientist) who entered the data immediately alerts another SCMT member (or the Lead Scientist) to review that change via Slack instant messaging. All questions and discussions about the data entry are carried out in real time via Slack. Once the new data is accepted or rejected, the reviewer documents this on the Slack channel and notes that the review process is complete. Data entries, edits or deletions that are not high priority are reviewed weekly by members of the SCMT.

X. Reanalysis and Reevaluation

A. Quality Control Procedures

Prior to each OncoKB data release, all reanalysis and reevaluation of OncoKB assertions and data is executed by the SCMT under the guidance of the Lead Scientist and occurs every 8 weeks. Each OncoKB data release is logged in the OncoKB GitHub data repository and accessible to registered users through the OncoKB website.

Reanalysis and reevaluation of potential data discrepancies are identified using the following four database queries:

a. Variants with conflicting/inconclusive assertions of oncogenic/biological effect
b. Variants without oncogenic or mutation effect assertions
c. Variants with oncogenic and mutation effect assertions but without curated Evidence (i.e., absence of PMIDs)
d. Comparison of all variants associated with a Level of Evidence between previous and about-to-be released website versions

B. Resolving Identified Errors

Any discrepancies and errors identified through these queries are re-curated using **Protocols #1-4**. They are reviewed using criteria detailed in Chapter 5, Section IX. Reanalysis and reevaluation is repeated until no errors arise in the current data release.

Once reanalysis and reevaluation is complete, a beta oncokb.org is created for final review. This website is carefully reviewed by the SCMT and Lead Scientist to ensure that there are no errors in the data output and all updates are properly displayed. Specifically, the SCMT reviews:

1. **The Homepage:** To check that the number of genes, alterations, tumor types and drugs are accurate, as well as the number of leveled genes (Levels 1-4 and R1/R2).
2. **The Actionable Genes Page**: To check that all updated levels of evidence are properly displayed on the table.
3. **News:** To ensure that the news is accurate and comprehensive and properly displayed.
4. **Gene and Variant Pages**: Gene and variant pages that have a new or updated level of evidence are reviewed to ensure data is accurate and properly displayed. Additionally, each member of the SCMT reviews 5 gene and 5 variant pages to ensure data is consistent with the curation platform and properly displayed.
5. **Additional Tabs:** One member of the SCMT is responsible for reviewing all additional website tabs (Cancer Genes, Data Access, About, Team, Terms) to ensure all information (previous and updated) is properly displayed.

If errors are identified or changes need to be made, these are implemented in the OncoKB curation platform following the rules outlined in Chapters 3-5, and reviewed according to Chapter 5, Section IX. The beta website is then updated and steps 1-5 above are repeated. This process continues until all errors are resolved and the data is considered finalized and ready for public release

In addition, to ensure that all variant assertions are accurate and the evidence supporting an assertion is up-to-date, comprehensive reevaluation and reanalysis of genes and their associated variants occur in review cycles specified in **Table 1**. The SCMT may execute the review themselves or assign specific gene(s) as needed for reevaluation to curators.

Table 1. OncoKB data as of 2/1/2019.

	Genes (%)	Variants (%)	Review Cycle
# with a Level of Evidence	81 (14)	161 (4)	Every 8 weeks
# with Variant Assertions	311 (54)	4220 (96)	~50 genes every 4 months (all genes evaluated in ~2 years)
# without Variant Assertions	187 (32)	N/A	All gene summary and backgrounds reviewed every 2 years
Total	579 (100)	4381 (100)	-

In the OncoKB curation platform, all variant assertions in the OncoKB website are associated with a Description of Evidence that has been curated by OncoKB curators and/or SCMT with links to the supporting evidence sources (e.g., PMIDs or Abstracts). Per specific review cycle, these descriptions of evidence for the set of genes being re-evaluated can be downloaded for review (**Fig. 12**). Should the SCMT find that the Description of Evidence or sources supporting a variant assertion is inaccurate, the SCMT, in consultation with the Lead Scientist, makes the appropriate changes.

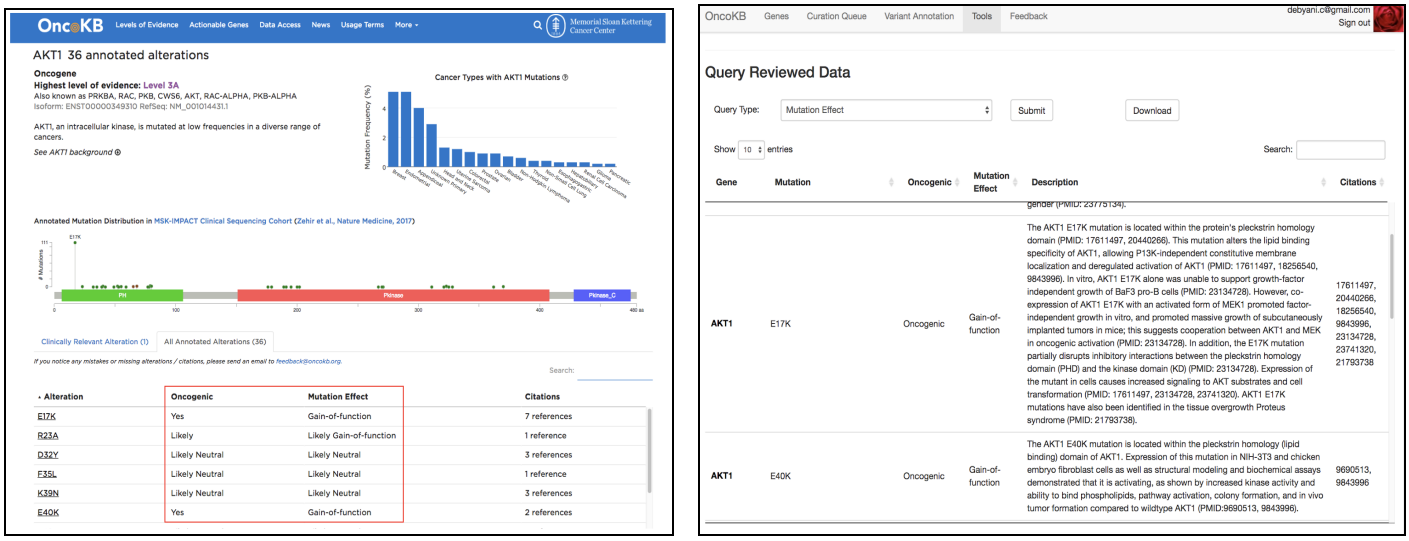


Figure 12: OncoKB Data Reanalysis and Reevaluation. Variant assertions in the OncoKB website (left panel, boxed in red) have curated descriptions of the evidence supporting the assertion in the curation platform. These are used by the Lead Scientist and SCMT to reevaluate assertions.

Chapter 6: Annotation of Variants in Patient Tumor Samples

With the curated content as the foundation, OncoKB has implemented tools for annotating variants detected in sequenced patient tumors (including a web application programming interface and an annotator tool, both described in Chapter 7). OncoKB annotates variants with assertions of its oncogenic and biological effects, and with its tumor type-specific clinical implications using automation based on specific rules described below. These rules are in place to simplify the curation process when possible, and provide annotations to variants for which there may not be specific functional data, but whose oncogenic and mutation effect can be inferred from other functionally validated variants or through its statistical recurrence in cancer.

I. Variant Annotation Process

In cBioPortal, OncoKB data is used to annotate alterations found in individual patient tumor samples. These annotations contain three brief statements:

1. ***Gene summary:*** One to two sentences detailing the functional role of the gene in a cell and in which tumor types it is frequently altered. *e.g., BRAF, an intracellular kinase, is frequently mutated in melanoma, thyroid and lung cancers among others.*
2. ***Oncogenic summary:*** An evidence-based assertion that defines the oncogenic effect of the alteration. Possible assertions include Oncogenic, Likely Oncogenic, Neutral, Likely Neutral, or Inconclusive. (refer to Chapter 4, Section III and **Protocol #2**) e.g., *The BRAF V600E mutation is known to be oncogenic.*
3. ***Clinical Summary:*** The clinical summary is one or two sentences summarizing the therapeutic implications of the queried alteration with a therapeutic level of evidence in a specific tumor type. *e.g The RAF-inhibitors encorafenib, dabrafenib and vemurafenib alone or in combination with the MEK-inhibitors binimetinib, trametinib and cobimetinib, respectively, are FDA-approved for the treatment of patients with BRAF V600E/K mutant melanoma.*

When an alteration in a patient tumor sample is queried, the clinical implications associated with all matched curated alterations and all matched curated tumor types are matched to the queried alteration and tumor type. However, only one oncogenic effect, mutation effect, and description of evidence can be associated with the queried alteration and tumor type. Therefore, to assign the specific oncogenic effect, biological effect, and description of evidence to a queried alteration, the process described in **Fig. 13** is used:

A. Match gene.
   Curated genes can be queried by HUGO symbols or Entrez Gene IDs.

B. Retrieve gene summary.
   The curated gene summary will be retrieved to annotate the queried variant.

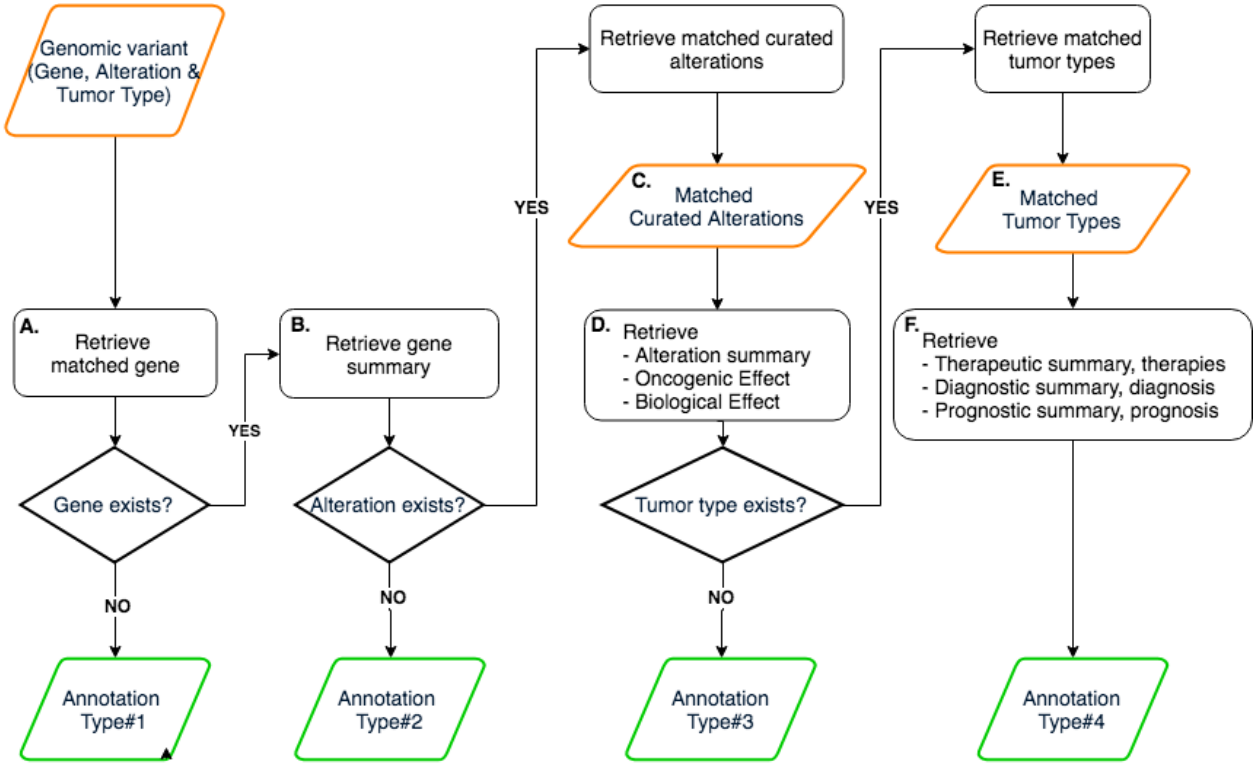
C. Match curated alterations.
   The process to match curated alterations is described in the Nomenclature and Rules section (Chapter 6, Section II).

D. Retrieve mutation summary, oncogenic and biological effects for the alteration.
   This is based on matched curated alterations (refer to Chapter 6, Section II).

E. Match curated tumor types.
   Refer to Section II. Nomenclature and Rules**.**

F. Retrieve tumor type summary and clinical implications.
   A tumor type summary will be generated for the queried variant. All clinical implications related to matched curated alterations and tumor types will be pulled. These implications will then be sorted by OncoKB level priorities (defined above). The resistance level implication has a higher priority than sensitivity levels if they are associated with the same therapy.

A) Summary of Variant Annotation Workflow



B) Sample Annotation Workflows

Genomic Variant #1 = PHF6 C242Y Acute Myeloid Leukemia
Annotation Type#1 = There is currently no information about this gene in OncoKB.

Genomic Variant #2 = BCL2 A131D Glioblastoma Multiforme
Annotation Type#2 = BCL2, an anti-apoptotic protein, is frequently altered in non-Hodgkin lymphomas. As of 01/03/2019, there was no available functional data about the BCL2 A131D mutation. However, it has been identified as a statistically significant hotspot and is predicted to be oncogenic (http://cancerhotspots.mskcc.org). There are no FDA-approved or NCCN-compendium listed treatments specifically for patients with BCL2 A131D mutant glioblastoma multiforme.
Genomic Variant #3 = ERBB2 G292R Melanoma
Annotation Type#3 = ERBB2, a receptor tyrosine kinase, is altered by mutation, amplification and/or overexpression in various cancer types, most frequently in breast, esophagogastric and endometrial cancers. The ERBB2 G292R mutation is likely oncogenic. While the anti-HER2 antibody ado-trastuzumab emtansine (T-DM1) is NCCN-compendium listed for the treatment of patients with ERBB2 mutant non-small cell lung cancer (NSCLC) and there is promising clinical data in patients with breast and NSCLC with known oncogenic ERBB2 alterations treated with the ERBB-targeted inhibitor neratinib, their clinical utility in patients with ERBB2 G292R mutant melanoma is unknown.
Genomic Variant #4 = BRAF V600E Melanoma
Annotation Type#4 = BRAF, an intracellular kinase, is frequently mutated in melanoma, thyroid and lung cancers among others. The BRAF V600E mutation is known to be oncogenic. The RAF-inhibitors encorafenib, dabrafenib and vemurafenib alone or in combination with the MEK-inhibitors binimetinib, trametinib and cobimetinib, respectively, are FDA-approved for the treatment of patients with BRAF V600E/K mutant melanoma.

Figure 13: Variant and Sample Annotation Workflow. ***A, Summary:*** To annotate variants found in patient tumor samples with its oncogenic and biological effects, and with tumor type-specific clinical implications OncoKB uses semi-automation summarized by this workflow. *A: Match gene*; Curated genes can be queried by HUGO symbols or Entrez Gene IDs. *B: Retrieve gene summary;* The curated gene summary will be retrieved to annotate the queried variant. *C: Match curated alterations*; The process to match curated alterations is described in the Nomenclature and Rules section. *D: Retrieve mutation summary, oncogenic and biological effects for the alteration;* This is based on matched curated alterations (refer to Chapter 6, Section II. Nomenclature and Rule). *E: Match curated*

tumor types; Refer to Section II. Nomenclature and Rules. *F; Retrieve tumor type summary and clinical implications.* A tumor type summary will be generated for the queried variant. All clinical implications related to matched curated alterations and tumor types will be pulled. These implications will then be sorted by OncoKB level priorities (defined above). The resistance level implication has a higher priority than sensitivity levels if they are associated with the same therapy. Orange rhombus = Input;Green rhombus = Output; Rectangle = Process; Diamond = Decision. ***B, Examples:*** Shown are sample annotations for the four different annotation types shown in part A of the figure.

II. Nomenclature and Rules Related to Annotation

A. OncoKB Cancer Gene List

OncoKB maintains a list of genes we consider as cancer genes based on their inclusion in various different sequencing panels, the Sanger Cancer Gene Census, or Vogelstein et al. (2013).

B. Curated Genes

Not every gene in the OncoKB Cancer Gene list has been curated by the team. We release new genes incrementally and refer to these genes as Curated Genes.

C. Matched Genes

OncoKB accepts gene HUGO symbols, Entrez gene IDs and gene aliases in the query to identify Curated Genes.

D. Matched Curated Alterations

When an alteration is queried in the OncoKB database, it may be associated with several alterations curated in the Gene Page and their associated annotations which include their oncogenic and biological effect, clinical implications and tumor type summary. The various curated alterations in OncoKB that match the queried alteration are referred to as Matched Curated Alterations.

1. *Overall Matching Logic*

Each queried alteration may be associated with one oncogenic effect and one biological effect. Therefore, the biological effect can be automatically associated with the queried alteration. The order of retrieving the information is the following:

a. Exact Match (single mutation header, e.g., V600E)
b. Exact Match (mutation in a string, e.g., V600E, V600K)
c. Positional Variant Match (e.g., V600)
d. Range Mutations (e.g., V600_K601delinsEQ)
e. Fusions
f. Deletion
g. Truncating Mutations
h. Oncogenic Mutations
i. Gain of Function Mutations
j. Loss of Function Mutations
k. Special Rules for Alterations

2. *Special Rules for Alterations*

a. ***Missense Mutations:*** If a specific missense mutation (e.g., BRAF V600E) is queried, it will be mapped to all curated mutations that reference the specific mutation position. This may include:
   i. the exact mutation match (V600E)
   ii. the exact mutation match in a list of mutations (V600E, V600K)
   iii. the positional variant match (V600)
   iv. a missense mutation range that includes the queried mutation (V600_K601mut)

b. ***In-frame Mutations***: OncoKB can curate in-frame mutations within an amino acid range. In-frame mutations will be mapped when the queried alteration position intersects within a curated range.

c. ***Oncogenic Mutations***: Any queried alteration that is annotated as “Oncogenic” or “Likey Oncogenic” in the OncoKB database, will be mapped to “Oncogenic Mutations”.

d. ***Fusions***: If a specific fusion is queried, it will be mapped to: 1) the specific fusion and 2) “Fusions” if curated.
e. ***Truncating Mutations***: If a truncating alteration is queried, it will be mapped to: 1) the specific truncating alteration and 2) “Truncating Mutations” if curated.
f. ***Duplications:*** For small tandem duplications (dups), the queried alteration must be an exact match to get mapped.
g. ***Deletion***: If a deletion event is queried, it will be mapped to: 1) “Deletion” and 2) “Truncating Mutations” if curated. If a deletion event is queried, and “Truncating Mutations” but not “Deletion” is curated.

E. Hotspots

Mutational hotspots are defined as mutant residues arising more frequently than expected in the absence of selection based on the analysis by Chang et al., 2018.

F. Matched Curated Tumor Types

Clinical implications are matched based on the patient’s tumor type. Queried tumor type will be associated with curated tumor types for the summary and clinical implication. As long as the curated tumor type is the same as or the parent node (based on OncoTree definition) of the query tumor type, it will be matched as a matched curated tumor type. We also include a few general tumor types (All Tumors, All Solid Tumors, ALl Liquid Tumors) and they will be mapped accordingly.

G. OncoKB Therapeutic Implication Levels of Evidence Priorities

Multiple therapeutic implications may be matched to a variant in a patient. When ranking them, we use the following order to keep the highest level of the implications.
Level R1 > Level 1 > Level 2 > Level R2 > Level 3A > Level 3B > Level 4

III. Annotation Summaries

A. Gene Summary

Gene summary will be retrieved as curated in the system, e.g., “BRAF, an intracellular kinase, is frequently mutated in melanoma, thyroid and lung cancers among others.”

B. Variant Biological Summary

The biological summary is one sentence that describes the oncogenic effect of the queried alteration. This sentence is programmatically generated based on the oncogenicity of the genetic alteration (refer to **Table 2**). The mutation summary is included in the variant-annotation endpoints of the OncoKB API.

Table 2. Example mutation summaries.

Mutation	If the alteration selected in cBioPortal is…	The sentence in the OncoKB card will be...
BRAF V600E	Oncogenic	The BRAF V600E mutation is known to be oncogenic.
BRAF T241P	Likely Oncogenic	The BRAF T241P mutation is likely oncogenic.
BRAF R509Q	Likely Neutral	The BRAF R509Q mutation is likely neutral.
BRAF Q201H	Inconclusive	There is conflicting and/or weak data describing the oncogenic function of the BRAF V600X mutation
BRAF	Variant of Unknown Significance (VUS)	As of 10/17/2018, there was no available functional

A762V	assessed by SCMT	data about the BRAF A762V alteration.
BRAF P318S	VUS not assessed by SCMT	The BRAF P318S mutation has not specifically been reviewed by the OncoKB team, and its oncogenic function is considered unknown.
ARID1A G2087V	Hotspot (VUS not assessed by SCMT)	The ARID1A G2087V mutation has been identified as a statistically significant hotspot and is predicted to be oncogenic.
[Gene] [Mutation]	Hotspot (VUS assessed by SCMT)	As of [date], there was no available functional data about the [gene] [mutation] mutation. However, it has been identified as a statistically significant hotspot and is predicted to be oncogenic (http://cancerhotspots.mskcc.org).
DAXX Duplicati on	Structural variant within a gene that has “Truncating Mutations” curated as likely oncogenic	This DAXX duplication may be a truncating alteration and is likely oncogenic.
BRAF Q201*	Truncating mutation in an oncogene	BRAF is considered an oncogene and truncating mutations in oncogenes are typically nonfunctional.

C. Clinical Summary

The clinical summary is one or two sentences summarizing the therapeutic implications of the queried alteration with a therapeutic level of evidence in a specific tumor type. For example, “The RAF-inhibitor dabrafenib in combination with the MEK1/2-inhibitor trametinib is FDA-approved for the treatment of patients with BRAF V600E mutant anaplastic thyroid cancer.”

When a specific alteration in a patient tumor sample is queried for annotation, multiple curated alterations may be matched and each matched curated alteration may have its own clinical summary. However, only one clinical summary will be associated with each specific alteration in a patient of a specific tumor type.

Therefore, in order to assign the clinical summary, the matched curated alterations are prioritized based on the order below (using BRAF V600E in a patient with Colorectal Cancer (CRC) as an example):

1. Clinical summary under the exact match alteration (V600E) for the tumor type in question (CRC) (so in the example of V600E in CRC, we will stop here because we have curated the alteration and tumor type specific clinical summary)
2. Clinical summary under the relevant positional variant (V600) for the tumor type in question (CRC)
3. Clinical summary under the exact match alteration (V600E) for “Other Tumor Type”
4. Clinical summary under the relevant positional variant (V600) for “Other Tumor Type”
5. Clinical summary under the highest priority relevant alteration (see above for prioritization of matched curated alterations) for the tumor type in question (CRC)
6. Search under the highest priority relevant alteration (refer to Chapter 6, Section II.D) for the other tumor type and use that summary (if present)
7. Continue steps 7-8 until all matched curated alterations have been evaluated for clinical summaries
8. If the queried alteration is associated with an “Oncogenic” or “Likely Oncogenic” mutation effect, search under “Oncogenic Mutations” for the tumor type in question (CRC)
9. If the queried alteration is associated with an “Oncogenic” or “Likely Oncogenic” mutation effect, search under “Oncogenic Mutations” for “Other Tumor Types”

D. Resistance Mutations

For alterations with an associated Level R1 or R2, the specified therapy (i.e., the therapy to which the alteration is considered a biomarker of resistance) will ONLY be associated with resistance (and NOT sensitivity).

Chapter 7: OncoKB Data Access

There are three ways that the public may access OncoKB data:

1. Through the OncoKB API
2. Through the publicly available website www.oncokb.org
3. Through cBioPortal

I. The OncoKB API

The OncoKB data can be accessed through a REST API (https://oncokb.org/api/v1/swagger-ui.html). The API is defined and organized using swagger annotation. MAF file annotation is also possible by using OncoKB Annotator (https://github.com/oncokb/oncokb-annotator) which is fully supported by using OncoKB REST APIs.

II. The OncoKB Website: www.oncokb.org

The OncoKB.org website (www.oncokb.org) was first released to the public at the American Society of Clinical Oncology (ASCO) Annual Meeting in 2016. This website provides to the clinical and scientific community worldwide the current and detailed annotation of the oncogenic effects and therapeutic implications of alterations observed in cancer.

As of 02/2020, the website has information about 5150 variants annotated in 671 genes across 48 tumor types, with therapeutic information for 88 drugs (**Fig. 14**).

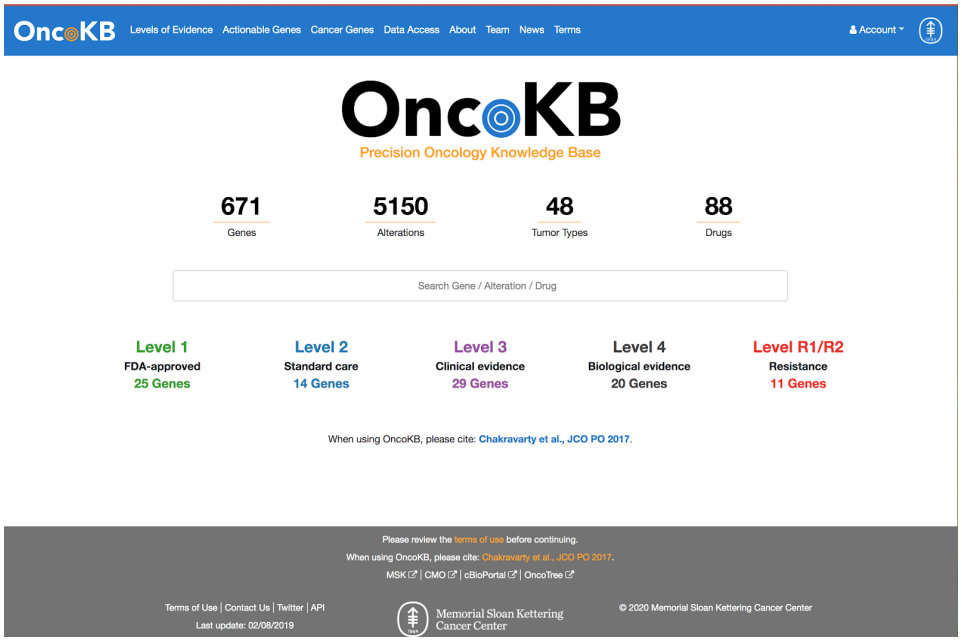


Figure 14: OncoKB.org Homepage.

The homepage of oncokb.org (**Fig. 14**) displays the following sections and functionalities:

A. Data Summary

The website shows the current number of genes (clickable), alterations, actionable tumor types and drugs curated in OncoKB. The “genes” number links to the OncoKB Cancer Gene List page. Below the search bar, the number of genes with alterations associated with a level of evidence are summarized. The number of genes below each Level of Evidence links to the Actionable Genes page.

B. Search Bar

Queries can be entered in the search box to lookup genes, aliases, EntrezID or gene-variant combinations in OncoKB. Upon entering a query, a drop-down menu will automatically appear listing possible gene and variant matches. Additionally, each suggested variant in the drop-down menu will be associated with an oncogenicity and (if relevant) the highest associated level of evidence. Clicking on a variant in the drop-down menu links to the variant page. Currently only one gene and/or one variant can be queried at a time.

C. Levels of Evidence

The Levels of Evidence page (**Fig. 15**) shows the hierarchy and definitions of the OncoKB Levels of evidence, as described in Chapter 5. This schematic can be downloaded in PDF or PPT format.

D. Actionable Genes

The Actionable Genes page (**Fig. 16**) lists all the gene-alteration-tumor type combinations that are associated with a level of evidence (Sensitivity Levels 1-4 and Resistance Levels R1-R2). The table is divided into five columns: Level, Gene, Alterations, Tumor Type and Drugs. Clicking on the entry under “Gene” will bring the user to the respective gene page. Clicking on the entry under “Alteration” will bring the user to the variant page. The user can customize the table by selecting 1 or more levels from the top of the page, thus only visualizing the data associated with the selected levels. The page also contains search bars for gene, tumor type and drug, thus allowing the user to customize the table with his/her desired search terms.

Figure 15: Levels of Evidence page in oncokb.org.

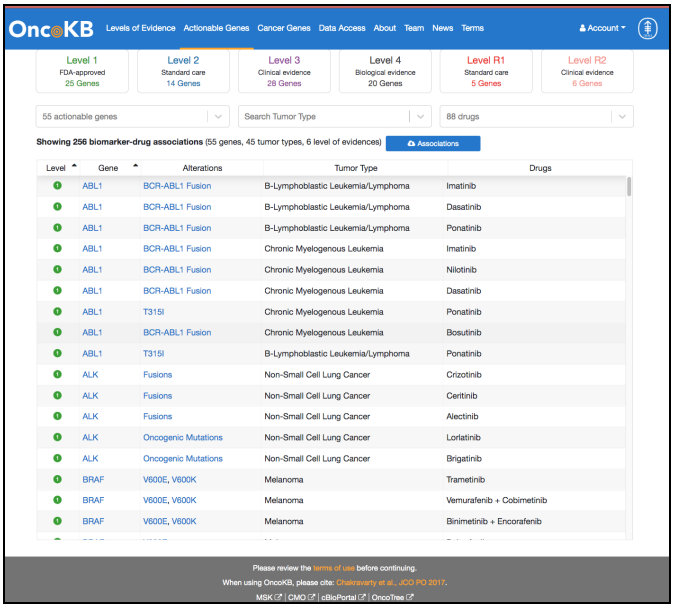


Figure 16: Actionable Genes page in oncokb.org.

E. Data Access

The OncoKB Data Access page (**Fig. 17**) allows the user to register for a license for the purpose of accessing OncoKB data via its web API (refer to Chapter 7, Section III for User Login and Registration details). Once registered and logged in, the user will have access to the following:

1. Annotating Files: The user can annotate data files (mutations, copy number alterations, fusions, clinical data) with the OncoKB Annotator.
2. Web API: The user can programmatically access the OncoKB data via its web API.

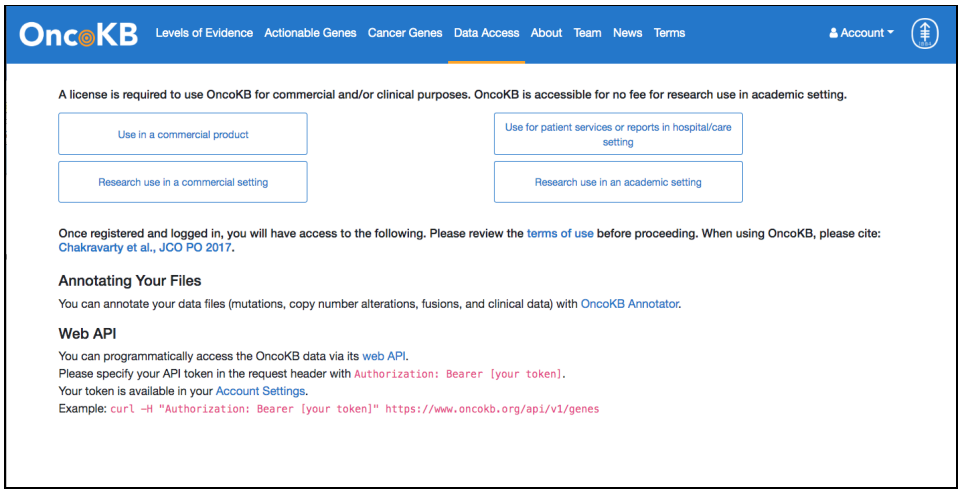


Figure 17: Data Access page in oncokb.org.

F. News

The News page (**Fig. 18**) contains: 1) details of any new data and/or updates added at each OncoKB version release, 2) the date of each release, and 3) a link to sign up to receive low-volume OncoKB email updates. Website updates are released approximately monthly.

Specifically highlighted in the news are:

1. Changes to actionable alterations, levels of evidence or therapeutics
2. Addition of new genes
3. Changes to any functions on the website
4. Additionally, moving forward, for each change or introduction of a new level of evidence, the news will now include the names of the CGAC members that affirmatively verified the change, in addition to the names of any CGAC members who have a specific COI regarding the change or new leveled association.

Not highlighted are:

1. Changes to mutation effect or oncogenic effect of alterations
2. Changes to citations
3. Addition or subtraction of alterations
4. Changes to a gene's designation as tumor suppressor or oncogene

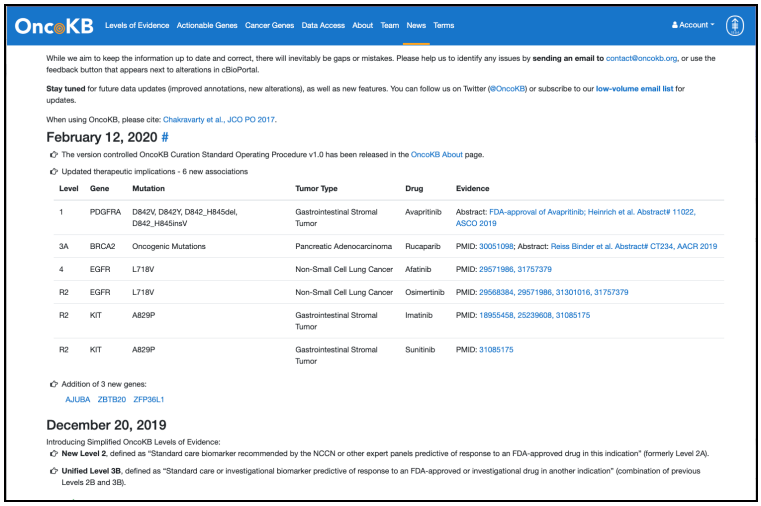


Figure 18: News page in oncokb.org.

G. Usage Terms

This page contains OncoKB licensing and data usage terms and guidelines (**Fig. 19**). The usage guidelines must be read and understood before using the data in OncoKB. Any additional inquiries about OncoKB usage terms may be directed to contact@oncokb.org.

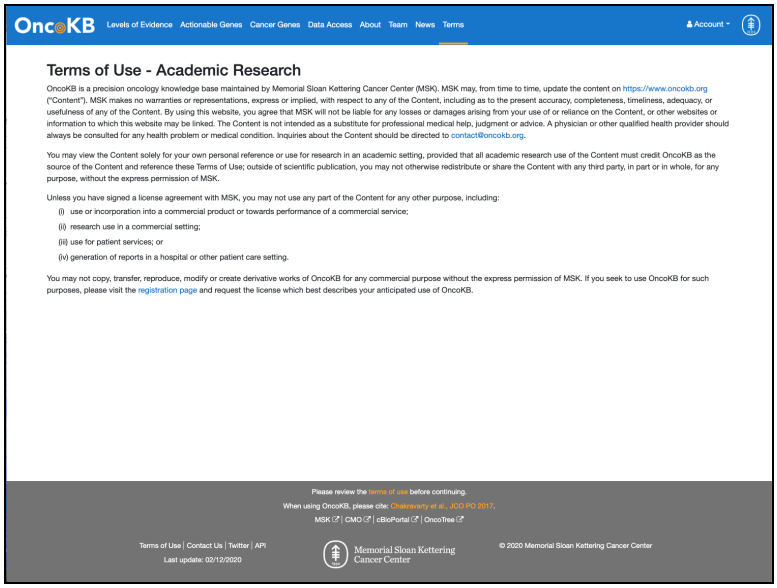


Figure 19: Usage Terms in oncokb.org.

H. OncoKB Cancer Gene List

The OncoKB Cancer Gene List page (**Fig. 20**) contains the genes considered by OncoKB to be cancer genes and indicates with a checkmark their inclusion in a specified resource, including:

1. MSK-IMPACT
2. MSK-IMPACT Heme
3. Foundation One
4. Foundation One Heme
5. Sanger Cancer Gene Census
6. Vogelstein et al., 2013.

Each gene is further classified as an Oncogene or Tumor Suppressor based on the criteria outlined in **Protocol #1** (refer to Chapter 3, Section III). The data on this page can be downloaded as a tab delimited file by clicking on the button in the upper right-hand corner of the page.

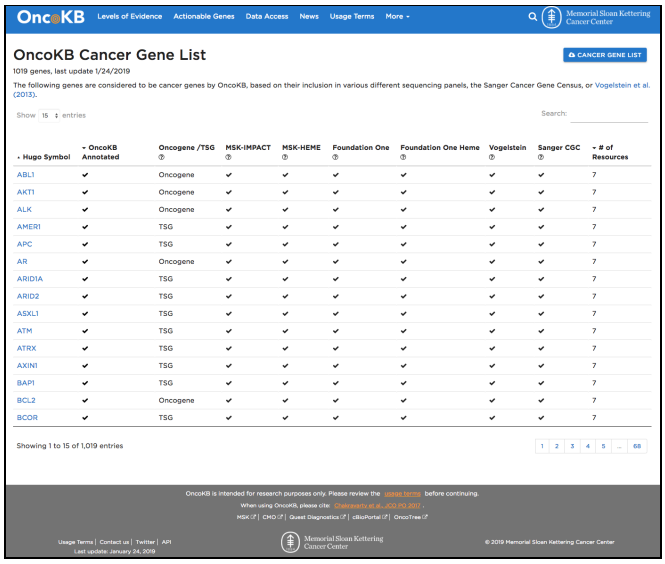


Figure 20: OncoKB Cancer Gene List in oncokb.org.

I. About OncoKB

The About page (**Fig. 21**) provides information about the history of OncoKB, and provides a schema delineating its oversight and governance, inputs, workflow and outputs. Additionally, a link to the first version of the OncoKB SOP titled *OncoKB Standard Operating Procedure v1.0* can be found here.

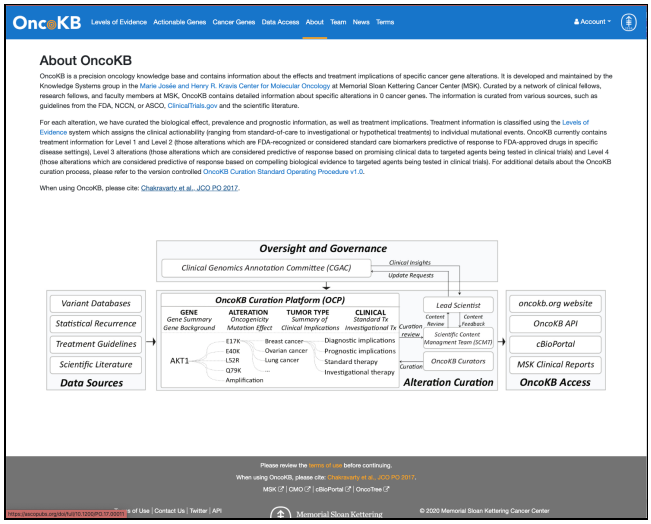


Figure 21: The About Page in oncokb.org.

J. Team

The Team page (**Fig. 22a**) lists the names of the individuals involved in the creation, development and maintenance of OncoKB, including:

1. Design & Development Team (including members of the Lead Scientist, SCMT members and Leadership)
2. Current OncoKB Curators
3. Past Contributors to OncoKB
4. Clinical Genomics Annotation Committee

Note, financial conflicts of interest for all OncoKB personnel are disclosed publicly on the OncoKB website via linking to an online spreadsheet that lists all relevant relationships (**Fig. 22b**).

(a) (b)

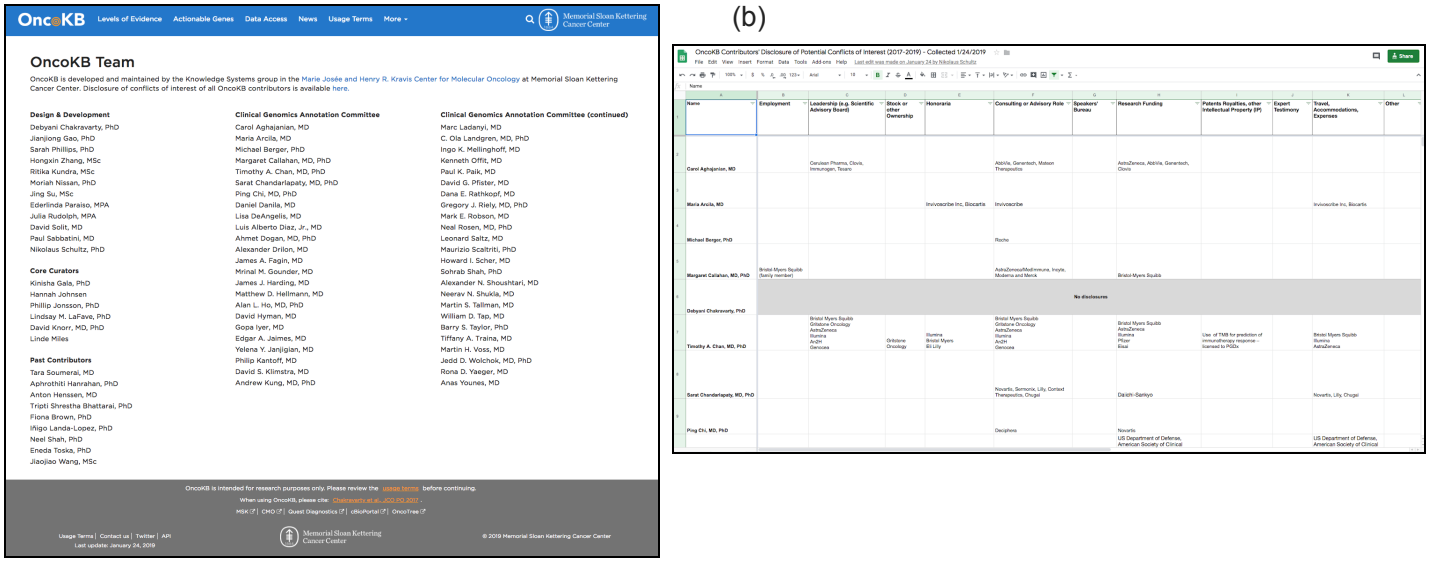


Figure 22: OncoKB Team List in oncokb.org. (a) All OncoKB personnel including past contributors are listed here. **(b)** The word "here" in the introduction statement "Disclosure of conflicts of interest of all OncoKB contributors is available here." links to a spreadsheet that lists the relevant financial conflicts of all OncoKB personnel.

K. Gene Pages

Gene-specific data in OncoKB can be found on individual gene pages (**Fig. 23**). Note: Not all genes in the OncoKB Cancer Genes List have gene pages in OncoKB. Gene pages include the following information:

(a) (b)

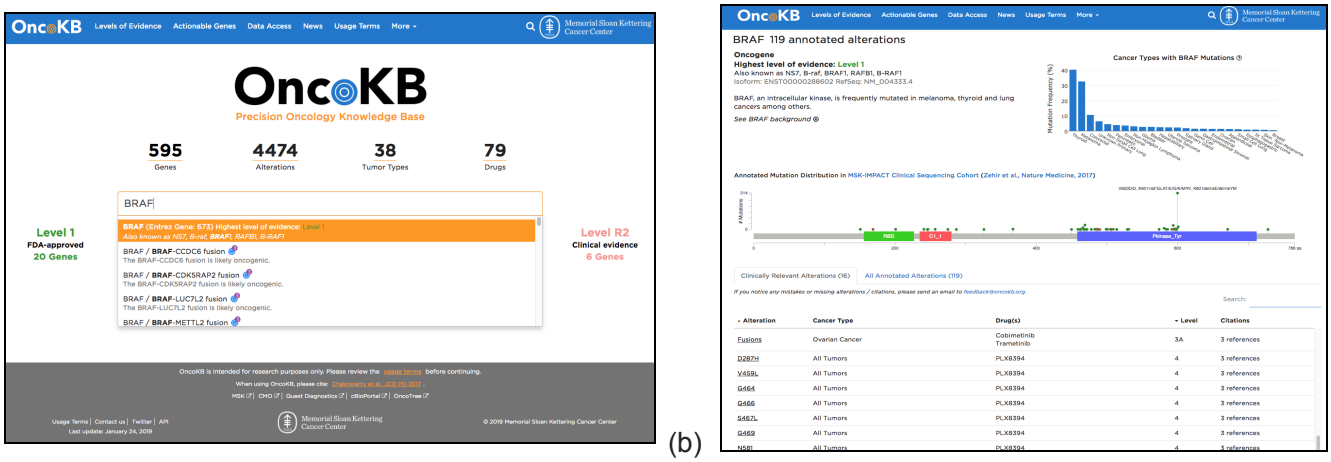


Figure 23: BRAF gene Page. (a) Searching for a specific gene will highlight all possible links and take you to the appropriate gene or alteration page. **(b)** BRAF Gene page shown here as an example.

1. **Gene summary:**
   The gene summary at the top of the gene page contains the following elements (**Fig. 24**)
   a. Gene name and its total number of annotated alterations in OncoKB
   b. Evidence-based classification of the gene as either oncogene and/or tumor suppressor
   c. The highest gene-associated Level of Evidence (if any)
   d. Gene-name aliases
   e. OncoKB utilized gene isoform and RefSeq ID

   Additionally, the gene summary has 1-2 sentences detailing the functional role of the gene in a cell and tumor types in which it is frequently altered.

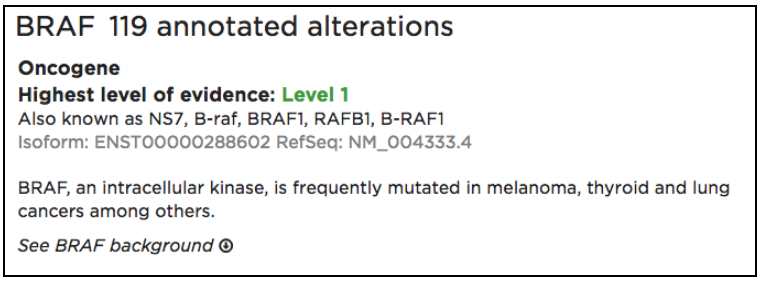


Figure 24: Gene Summary. BRAF shown as an example.

2. **Gene background:** Clicking the "See [Gene] background" below the Gene Summary expands the "Gene Background" text (refer to **Fig. 25**, example gene is BRAF), which describes the role of the gene-encoded protein in normal cells, its function in tumorigenesis, and its prevalence and mutation pattern in relevant tumor types. PMIDs in the gene background link out to the referenced paper abstract in PubMed in a new browser page.

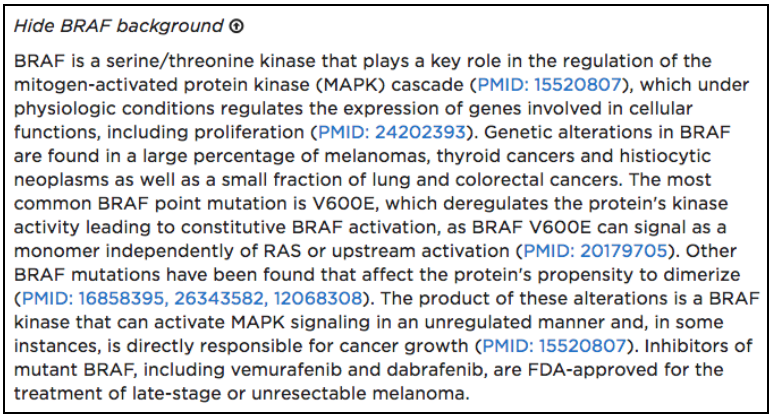


Figure 25: Gene Background. BRAF shown as an example.

3. **Gene-specific "Cancer Types" histogram: Fig. 26** shows the mutation frequency of the gene in different tumor types. The Y-axis shows the percent of samples that carry a mutation in the specific genes (including missense mutations, truncating mutations, and frameshift mutations) and the X-axis specifies tumor type. Data for this histogram is sourced from the ~10,000 tumor samples of the MSK-IMPACT Clinical Sequencing Cohort (Zehir et al., 2017) and does not account for copy number changes, chromosomal translocations or cancer types with fewer than 50 samples. Clicking on a bar in the histogram changes the data in the lollipop plot to reflect the selected tumor type.

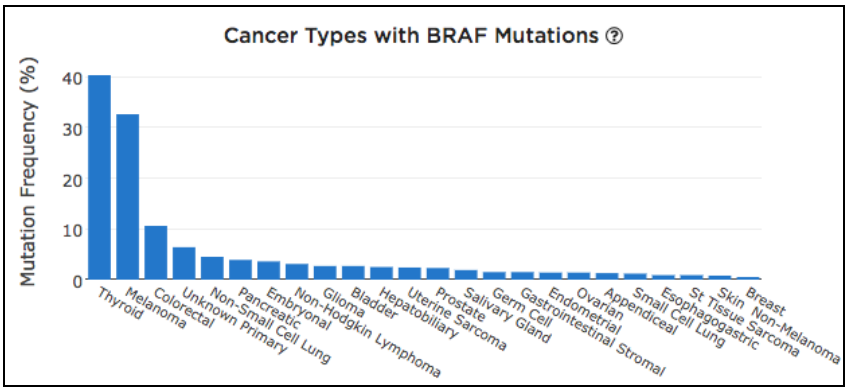


Figure 26: Gene-specific Cancer-Types histogram. BRAF shown as an example.

4. **Gene-specific lollipop plot:** The gene-specific lollipop plot is a schematic that displays the gene-encoded protein (**Fig. 27**). The X-axis of the plot is the amino acid position in the gene-encoded protein and the Y-axis of the plot is mutation count. On this schematic, the location of each mutation on

the protein is indicated by a “lollipop”, and the height of the lollipop signifies the mutational frequency of the mutant allele. Data for this histogram is sourced from the 10,000 tumor samples of the MSK-IMPACT Clinical Sequencing Cohort (Zehir et al., 2017). Clicking a specific mutation (or clicking a single lollipop) restricts the Alterations table to display oncogenic and actionability information (if any) associated with the selected mutation. Clicking on a tumor type in the “Cancer Types with [Gene] Mutations” histogram will restrict the displayed mutations in the lollipop plot to only those found in the selected tumor type. To undo the tumor type filter, the user can click “Current view shows filtered results. Click here to reset all filters”. The user can customize the plot and download it as a PDF or SVG file using the buttons that appear when the user hovers over the upper right of the lollipop plot.

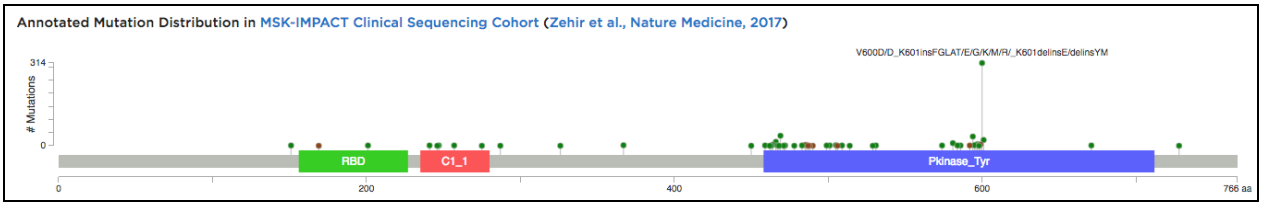


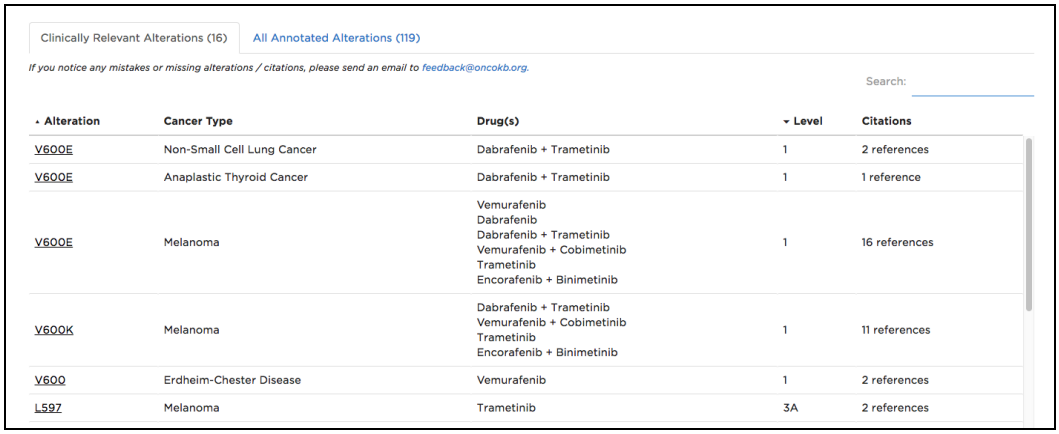
Figure 27: The gene-specific lollipop plot based on the published MSK clinical sequencing cohort in Zehir et al., 2017. BRAF shown as an example.

5. **Clinically Relevant and All Annotated Alterations tables:** Below the lollipop plot are two tabs, the *Clinically Relevant Alterations* and the *All Annotated Alterations* tables (**Fig. 28**). Both tables are searchable using the search bar indicated on the right-hand side of the table. By default, the Clinically Relevant Alterations table is shown. Each column in both tables is sortable. Clicking on the alteration brings you to the individual alteration page. Hovering over the citation column reveals a dialogue box that lists the title, citation and PMID of each source used to support the association. Clicking on either the title or the PMID will link out to the referenced paper abstract in PubMed in a new browser page.

   a. *Clinically Relevant Alterations (# of alterations):* Gene-specific alterations associated with a level of evidence indicating potential clinical actionability are shown in this tab (**Fig. 28a**) which lists:
      i. Clinically Relevant Alterations: Gene alteration considered clinically relevant
      ii. Tumor type in which the alteration is considered clinically relevant
      iii. Drug(s) associated with the clinical relevance of the alteration
      iv. Level of evidence for the alteration-tumor type-drug association
      v. Relevant citations

   b. *All Annotated Alterations (# of alterations):* All OncoKB curated Gene-specific alterations are shown in this tab (**Fig. 28b**):
      i. Gene alteration
      ii. Oncogenic status: Yes, Likely, Neutral, Likely Neutral or Inconclusive
      iii. Mutation Effect: Gain-, Loss-, Switch-of-function, Neutral, Likely Gain-, Likely Loss-, Likely switch-of-function, Likely Neutral, Inconclusive.
      iv. Citations: Citation number is listed with a mouse-over dialogue box that lists the title, citation and PMID of all references. Clicking on either the title or the PMID links out to the referenced paper abstract in PubMed in a new browser page.
      v. Clicking on the alteration links to the individual alteration page (refer to Chapter 7, Section II.L below).

(a)



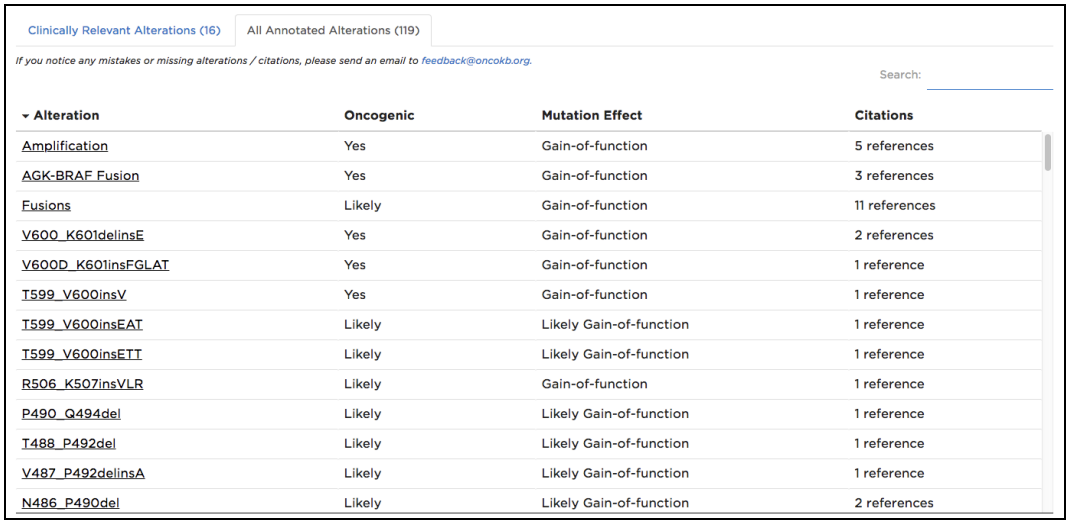
Clinically Relevant Alterations (16) | All Annotated Alterations (119)

If you notice any mistakes or missing alterations / citations, please send an email to feedback@oncokb.org.

Search:

Alteration	Cancer Type	Drug(s)	Level	Citations
V600E	Non-Small Cell Lung Cancer	Dabrafenib + Trametinib	1	2 references
V600E	Anaplastic Thyroid Cancer	Dabrafenib + Trametinib	1	1 reference
V600E	Melanoma	Vemurafenib Dabrafenib Dabrafenib + Trametinib Vemurafenib + Cobimetinib Trametinib Encorafenib + Binimetinib	1	16 references
V600K	Melanoma	Dabrafenib + Trametinib Vemurafenib + Cobimetinib Trametinib Encorafenib + Binimetinib	1	11 references
V600	Erdheim-Chester Disease	Vemurafenib	1	2 references
L597	Melanoma	Trametinib	3A	2 references

(b)



Clinically Relevant Alterations (16) | All Annotated Alterations (119)

If you notice any mistakes or missing alterations / citations, please send an email to feedback@oncokb.org.

Search:

Alteration	Oncogenic	Mutation Effect	Citations
Amplification	Yes	Gain-of-function	5 references
AGK-BRAF Fusion	Yes	Gain-of-function	3 references
Fusions	Likely	Gain-of-function	11 references
V600_K601delinsE	Yes	Gain-of-function	2 references
V600D_K601insFGLAT	Yes	Gain-of-function	1 reference
T599_V600insV	Yes	Gain-of-function	1 reference
T599_V600insEAT	Likely	Likely Gain-of-function	1 reference
T599_V600insETT	Likely	Likely Gain-of-function	1 reference
R506_K507insVLR	Likely	Gain-of-function	1 reference
P490_Q494del	Likely	Likely Gain-of-function	1 reference
T488_P492del	Likely	Likely Gain-of-function	1 reference
V487_P492delinsA	Likely	Likely Gain-of-function	1 reference
N486_P490del	Likely	Likely Gain-of-function	2 references

Figure 28: Clinically Relevant and All Annotated Alterations Tables. BRAF shown as an example. (a) The Clinically Relevant Alterations (with 16 total alterations) is selected by default (in black). The All Alterations table (with 119 total alterations) is clickable in blue. (b) When selected, the All Annotated Alterations tab (with 119 total alterations) is shown.

L. Alteration Pages

Similar to gene-specific data, alteration-specific data in OncoKB can be found on individual alteration pages. Typing the alteration into the homepage or OncoKB header search bars can access these pages (Fig. 29, BRAF V600E example shown). Alterations across all pages in oncokb.org also link to their respective Alteration pages. Note, not all alterations have alteration pages.

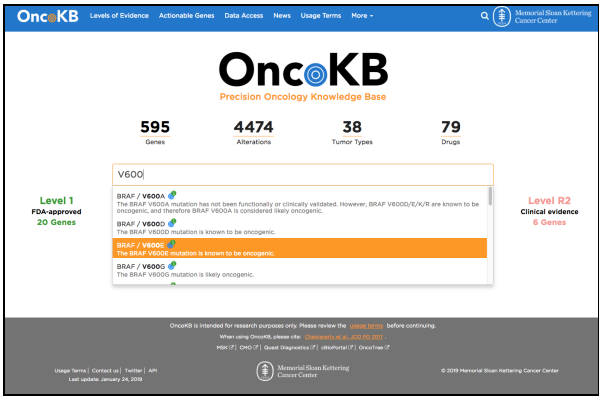


Figure 29: BRAF V600E Alteration Page. (a) Searching for a specific alteration will highlight all possible alterations to select from and take the user to the appropriate alteration page. (b) BRAF V600E Alteration page shown as an example.

Each Alteration page has the following information:

1. **Gene and alteration name.**

2. **Evidence-based classification of the oncogenic effect** of the alteration (refer to Chapter 4, Section III and **Protocol #2**). Possible classifications include Oncogenic, Likely Oncogenic, Neutral, Likely Neutral, Inconclusive
3. **Evidence-based classification of the biological effect** of the alteration (refer to Chapter 4, Section Iv and **Protocol #3**). Possible classifications include Gain-, Loss-, Switch-of-function, Neutral, Likely Gain-, Likely Loss-, Likely Switch-of-function, Likely Neutral, Inconclusive.
4. **Evidence-based classification of the clinical effect** of the alteration **and its highest alteration-associated therapeutic Level of Evidence (if any)**. Possible levels of evidence include the following (refer to Chapter 5):
   a. Therapeutic: Levels 1, 2, 3A, 3B, 4, R1 and R2
   b. Diagnostic: Levels Dx1, Dx2, Dx3
   c. Prognostic: Levels Px1, Px2, Px3
5. **Gene summary:** Refer to K.1 in this section and **Fig. 24 and 30**
6. **Alteration summary**: Summary of the evidence-based classification of the oncogenic effect of the alteration (refer to Chapter 4, Section III) is given in sentence form (highlighted in blue in **Fig. 30**).

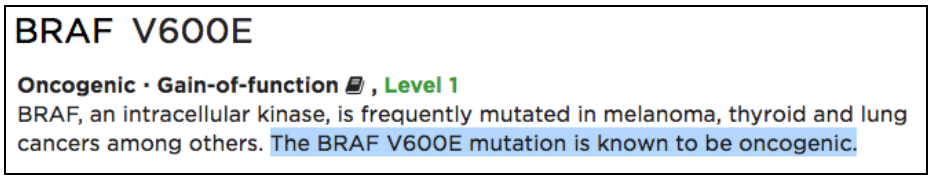


Figure 30: Alteration Summary. In addition to the gene summary the alteration summary is also shown in the Alteration page (highlighted in blue). BRAF V600E shown as an example.

7. **Additional gene information:** Information is described in Items L.1, **Fig. 24** and Item L.2, **Fig. 25** in this section. Briefly, whether the gene is an oncogene or tumor suppressor, the highest level of evidence associated (if any), the gene aliases and the gene background, with PMIDs that link directly to the reference.
8. **Alteration-specific "cancer types" histogram:** The Cancer Types histogram (**Fig. 31**) shows the frequency of the specific alteration in different tumor types. The Y-axis shows the percent of samples that carry the specific alteration and the X-axis specifies the tumor type. Data for this histogram is sourced from the ~10,000 tumor samples of the MSK-IMPACT Clinical Sequencing Cohort (Zehir et al., 2017). Alteration pages for copy number changes or chromosomal translocations do not have this histogram.

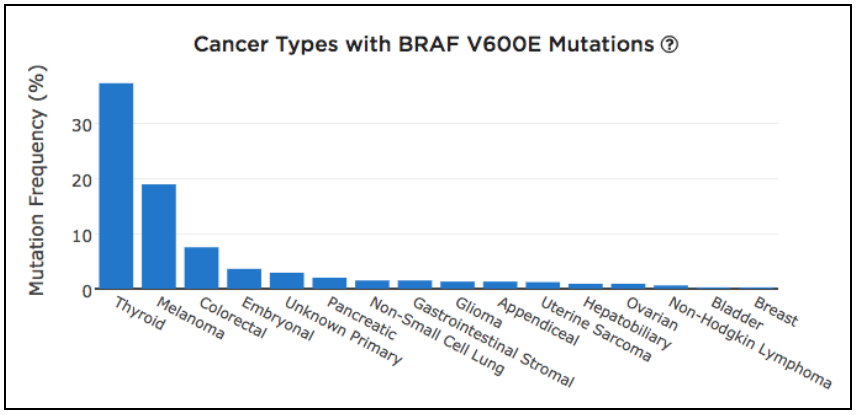


Figure 31: Alteration-specific Cancer Types histogram. BRAF V600E is shown as an example.

9. **Alteration-specific lollipop plot schematic:** The alteration-specific lollipop plot shows the position of the alteration in the gene-encoded protein and the tumor-type-specific mutational count of the specific mutant allele (as indicated by the height of the lollipop) (**Fig. 32**). Similar to the lollipop plot in Section K.4 and **Fig. 27** of this section, the X-axis of this schematic is the amino acid position in the

gene-encoded protein and the Y-axis of the plot is mutation count. Data for this histogram is sourced from the ~10,000 tumor samples of the MSK-IMPACT Clinical Sequencing Cohort (Zehir et al., 2017).

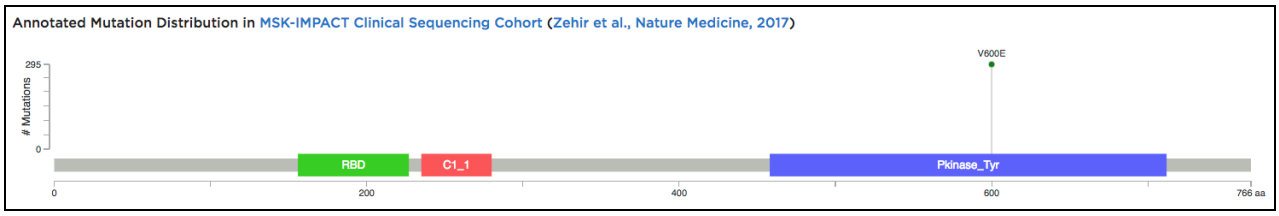


Figure 32: The alteration-specific lollipop plot based on the published MSK clinical sequencing cohort in Zehir et al., 2017. BRAF is shown as an example.

10. **Alteration-specific table:** Alterations with no associated level of evidence will not have an alteration-specific table. For clinically relevant alterations associated with a level of evidence indicating potential clinical actionability, an alteration-specific table (**Fig. 33**) will list the following:
    a. Gene alteration considered clinically relevant
    b. Cancer type in which the alteration is considered clinically relevant
    c. Drug(s) associated with the alterations clinical relevance
    d. Level of evidence for the alteration-tumor-type-drug association
    e. Relevant citations

    Each column in the table is sortable. Clicking on the alteration brings you to the individual alteration page. Hovering over the citation column reveals a dialogue box that lists the title, citation and PMID of each source used in support of the association. Clicking on either the title or the PMID will link out to the referenced paper abstract in PubMed in a new browser page.

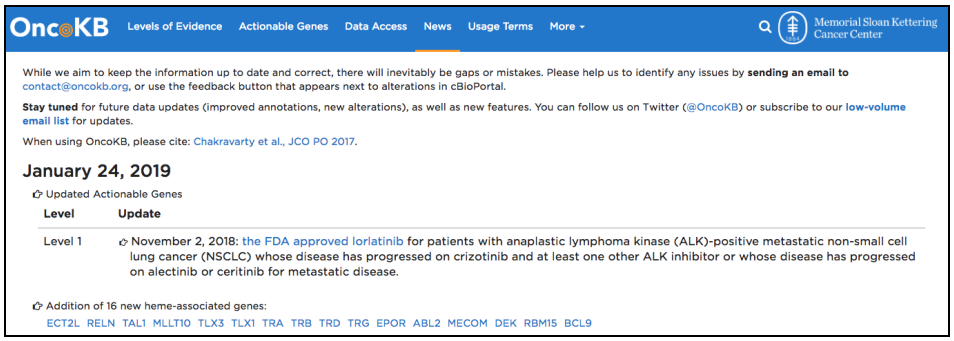
Search:

Alteration	Cancer Type	Drug(s)	Level	Citations
V600E	Non-Small Cell Lung Cancer	Dabrafenib + Trametinib	1	2 references
V600E	Melanoma	Vemurafenib Dabrafenib Dabrafenib + Trametinib Vemurafenib + Cobimetinib Trametinib Encorafenib + Binimetinib	1	16 references
V600E	Anaplastic Thyroid Cancer	Dabrafenib + Trametinib	1	1 reference
V600E	Colorectal Cancer	Encorafenib + Binimetinib + Cetuximab	3A	2 references
V600K	Melanoma	Dabrafenib + Trametinib Vemurafenib + Cobimetinib Trametinib Encorafenib + Binimetinib	1	11 references
V600	Erdheim-Chester Disease	Vemurafenib	1	2 references
V600	Colorectal Cancer	Panitumumab + Dabrafenib + Trametinib	3A	8 references

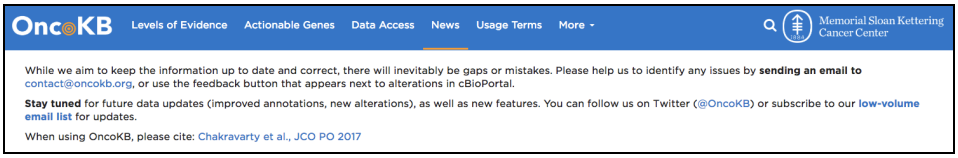
Figure 33: Alteration-specific tables. Alteration-specific tables are only available for those alterations associated with a level of evidence. The alteration specified as well as leveled therapeutic evidence related to the specified alteration are displayed. BRAF V600E is shown as an example.

11. **Feedback through OncoKB.org**: Assertion feedback by OncoKB users is an important feature of the knowledgebase. There are two web-based mechanisms through which users may provide feedback on OncoKB content: 1) the OncoKB website, and 2) via the cBioPortal for Cancer Genomics. Any feedback, comments or questions may also be sent via email to contact@oncokb.org, which is provided in multiple places within the OncoKB website (**Fig. 34**). Emails sent to contact@oncokb.org are received by the Lead Scientist and all SCMT members and answered within 48 hours.

(a)



(b)



(c)

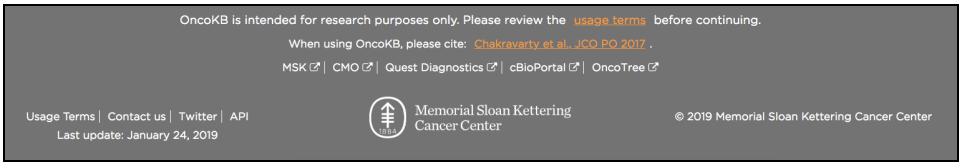


Figure 34: Feedback through Oncokb.org. Users of Oncokb.org may provide feedback on the website by clicking the email link for contact@oncokb.org (a) In the News section, (b) In the Usage Terms section, or by clicking “Contact Us” in (c) the OncoKB webpage footer.

III. User Login and Registration

OncoKB public website has released the User Login/Registration module to streamline user management and provide enhanced data protection. While all users can view gene/variant information on the website, API services are only available to approved registered users.

A. License Types

There are four types of licenses that a user may choose from when registering for an account at https://www.oncokb.org/account/register (**Table 3**).

Table 3. OncoKB licenses types that users may choose from when registering for an OncoKB account.

License Type	Description
Academic License	Research use in an academic setting.
Hospital License	Use for patient services or reports in hospital/care settings.
Research in Commercial License	Research use in a commercial setting.
Commercial License	Use in a commercial product.

B. User Registration Form

Once a user selects the license type, they will be prompted to complete a registration form and agree with OncoKB Terms of Use (**Fig. 35**). When the registration is complete, the system automatically sends the user an email with a verification link that must be clicked to complete registration.

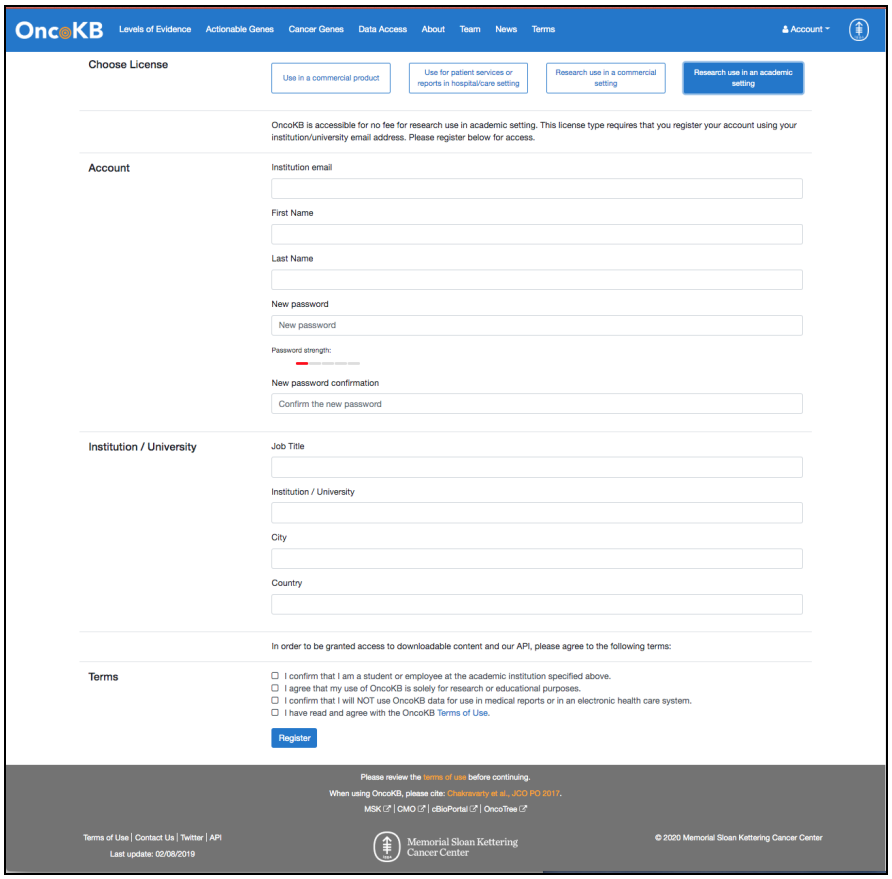


Figure 35: OncoKB User Registration Form. OncoKB users who want to gain access to the API must register by completing the above form and agreeing to the Terms of Use.

C. License Request Review

The OncoKB Team is immediately notified about every license request via a private SLACK channel (**Fig. 36**), allowing requests to be processed in real time. Users who register with an MSK email (@mskcc.org) are automatically approved by the system. Academic license requests are verified and approved by members of the OncoKB team. For academic licenses, users are required to use their institutional email. All hospital and commercial license requests are logged and forwarded to the MSK Office of Development for further review and contract negotiation.

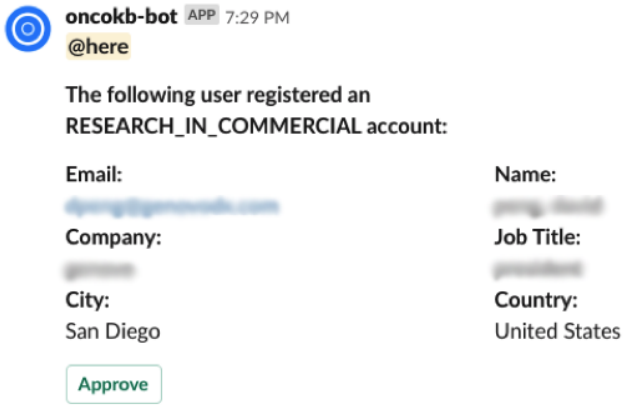


Figure 36: SLACK notification for an OncoKB Research in Commercial License

D. User Login

The OncoKB public website stores variant data (Variant Database) and user data (User Database) in separate MySQL databases. When a user logs in to the public website using their username and password, their credentials are sent to the system and verified in the User Database. Once the user passes the authentication, they are allowed to access OncoKB data by API services.

E. API Services & Token

The OncoKB API services are protected and only available to registered/approved users. The OncoKB website automatically creates a token for all approved users, allowing them to programmatically access the OncoKB data via its web API and token https://www.oncokb.org/swagger-ui/index.html. If a user tries to access the OncoKB API without a token, the system will return “Not authorized user” error and that user will not be granted access to the API.

IV. Data and Website Security

A. Data Security

Oncokb.org uses token-based authentication enabled by Spring Security layer to protect the data. For each registered and approved user, the OncoKB website will automatically create a token and store it in the database. Once the token is generated, it cannot be altered by the user. When the user successfully logs in using his/her credentials following authentication, his/her token will be returned. Once a user is logged in, each subsequent request will include the token, allowing the user to access routes, services, and resources that are permitted with that token. With this system in place no one can access OncoKB data without an assigned token. Importantly, OncoKB APIs provide read-only data. Therefore, no one can modify OncoKB related data through either the website or OncoKB APIs. Additionally, the public database that stores data for oncokb.org is backed up daily and can quickly be recovered if needed. For the purpose of curating data (data which once reviewed will be displayed on the OncoKB public website, oncokb.org), there is a separate OncoKB curation website that is deployed in an internal server and protected under MSK firewall.

B. Website Security

OncoKB has mechanisms in place to prevent cyber attacks as well as a procedure to follow in case of an on-going attack.

OncoKB’s attack surface is kept small through a variety of mechanisms at different levels of the stack. The REST API is a microservice written in Java using the Spring framework, which has built-in protection against several forms of attacks. Similarly for the frontend, which uses React. Both of these components of the stack are open source and hosted on GitHub. GitHub provides automatic detection of vulnerabilities in dependencies for both Java and JavaScript. The app runs inside a Docker container that uses an official Java Docker image. It has read-only access to the MySQL database that contains the variant information. The containers run in a dedicated namespace on a Kubernetes cluster. All these are preventative measures to help decrease the attack surface and prevent escalation of the attack in case the container is compromised.

To be able to detect an on-going attack, team members can utilize a variety of dashboards to monitor the logs of the web service, HTTP requests and database queries, and gain insight into activity on the Kubernetes cluster or Amazon Web Services. In case an attack is detected the following procedure can be followed by several members of the team:

1. If the web service itself is not compromised. Determine the IP address(es) of the attack by inspecting the logs of Nginx and block it.
2. If the container running OncoKB is compromised, do (1) and restart all the containers.
3. If the cluster is compromised. Create a new cluster, limit access to only your own IP and update DNS records to point to the new cluster. Remove the old cluster.
4. If AWS or Google Domains is compromised follow the SOP of those services to regain ownership of the account.

V. OncoKB Content Accessible through cBioPortal

The OncoKB knowledgebase is integrated into cBioPortal (cbioportal.org) through annotation of mutation effect, oncogenic effect and level of evidence of alterations visualized on the platform.

A. OncoKB icons in cBioPortal

OncoKB icons are coded and used in cBioPortal to communicate the oncogenic and biological effect and actionability of a given variant. The following are the rules of the icons used in the cBioPortal:

1. cBioPortal uses the OncoKB symbols to signify information known about the variant.
2. In addition to specifying the oncogenic effect, the portal icon will display the highest levels of evidence for the given variant and the tumor type.
3. “Predicted variants” are mutations that are mutational hotspots in cancer Chang et al., 2018 but that are not specifically curated in OncoKB.

B. OncoKB Cards in cBioPortal

OncoKB information is displayed in cBioPortal in OncoKB cards that appear when the user hovers over the OncoKB icon that is next to an alteration in the mutation table in the “mutations” tab of a gene query or in the Patient View of a sample in the Mutations tab.

The card is divided into the following sections:

1. **Header:** The header lists the gene, alteration, and tumor type of the respective sample
2. **Clinical Implications:** The clinical implications tab (**Fig. 37**) describes the oncogenicity of the alteration. This section is clickable and changes the information in the “description” space directly below.
3. **Description:** By default, the information displayed in the description section is the “clinical implications” information. The “clinical implications” information includes the:
   a. Gene summary
   b. Mutation summary
   c. Tumor type summary
   d. Clinical actionability table: The information in this table includes:
      i. Level of evidence icon: if the user hovers over the icon, the definition of the level is displayed. While the OncoKB icon on the “mutations” tab displays the highest level of evidence for the alteration, the OncoKB Card lists all levels of evidence associated with the alteration.
      ii. Alteration associated with the level of evidence
      iii. Drugs associated with the level of evidence
      iv. Tumor type associated with the level of evidence
      v. Citation icon: Upon mouse-over, this icon shows sources associated with each leveled evidence.

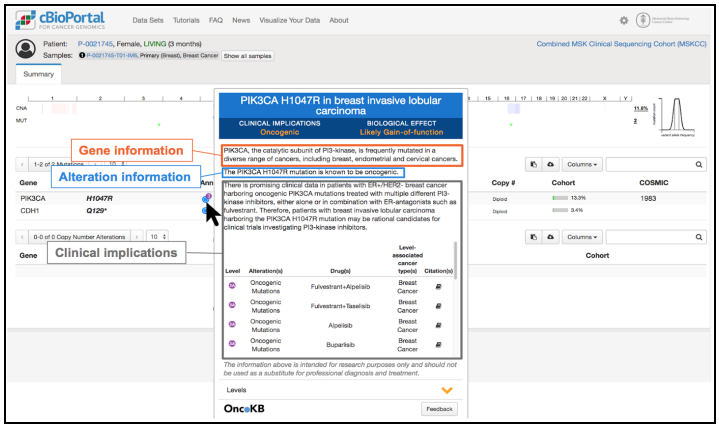


Figure 37: The OncoKB Card in cBioPortal, Clinical Implications: Hovering over the OncoKB card in the patient view or mutations tab in cBioPortal will display the OncoKB card. Gene-specific information is outlined orange, alteration-specific information is outlined in blue and clinical implications (if relevant for the specified tumor-type) is displayed in grey.

4. **Biological Effect:** The biological effect tab (**Fig. 38)** describes the biological effect of the alteration (whether the alteration is gain-of-function, loss-of-function, neutral, etc.). Clicking on the biological effect tab in the gene card will switch views to display the biological effect of the alteration. In this section of the OncoKB card, an evidence-based classification of the biological effect of the alteration is provided and the list of references supporting this classification.

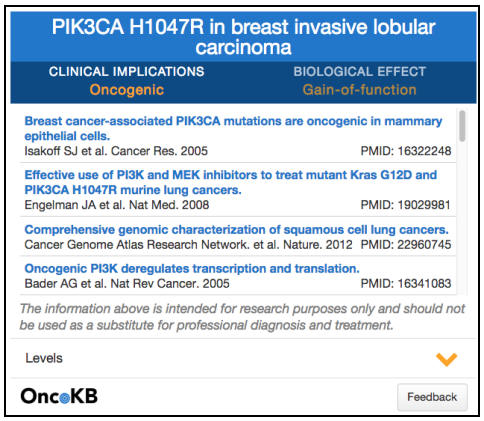
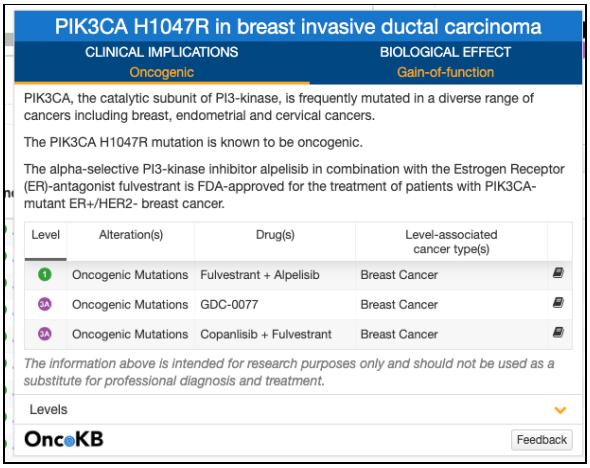


Figure 38: The OncoKB Card in cBioPortal, Biological Effect: Clicking on the biological effect tab in the OncoKB gene card shows a list of references that support the assertion of the biological effect shown in dark blue (example shown here; the PIK3CA H1047 mutation [found in breast invasive lobular carcinoma] is Gain-of-function) and link out to the respective PubMed Abstract page.

5. **Levels:** Levels in the OncoKB card (**Fig. 39**) refers to the Levels of Evidence that support the mutation being predictive of response to the targeted therapies. Clicking on the down arrow next to "Levels" reveals a drop down description of all the OncoKB levels of evidence (both sensitivity and resistance).

(a)



(b)

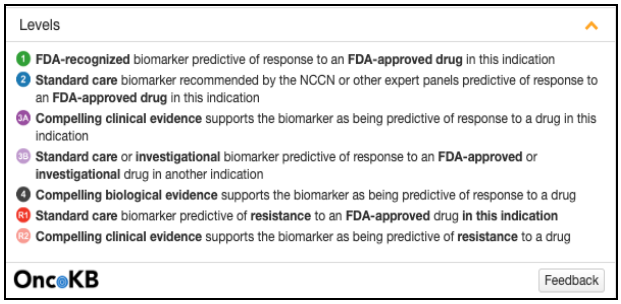


Figure 39: Levels in the OncoKB card: Clicking on the yellow arrowhead in the OncoKB card displays a glossary of the definition of the Levels of evidence.

6. **OncoKB website and feedback:** Clicking on the OncoKB logo will bring the user to the OncoKB.org website. Clicking on "Feedback" (**Fig. 40a**) results in a pop-up comment card (**Fig. 40b**) that allows the user to provide feedback about the gene-alteration combination directly to the OncoKB team via Google forms. In the "OncoKB Annotation Feedback" pop-up form, information about the gene and alteration, the email address used to log into the portal, and the web address of the specific portal instance will be pre-populated in the feedback form. Users may then enter specific feedback and associated references in the Feedback and References fields before submitting the feedback. Submission of feedback by a cBioPortal user will auto-populate in a Google spreadsheet with all the information entered above. Changes to this Google Sheet will trigger an automatic email sent to the Lead Scientist and SCMT alerting them of user feedback via cBioPortal and will be answered within 48 hours. Upon completion of any necessary deliverables as suggested by the feedback (either curation or software related), the appropriate OncoKB staff member fills in the "Complete" column and adds their initials as well as any comments related to the feedback item (**Fig. 40c**). The Feedback Page collates all cBioPortal

user feedback related to OncoKB assertions and is a log of OncoKB development based on cBioPortal user-feedback.

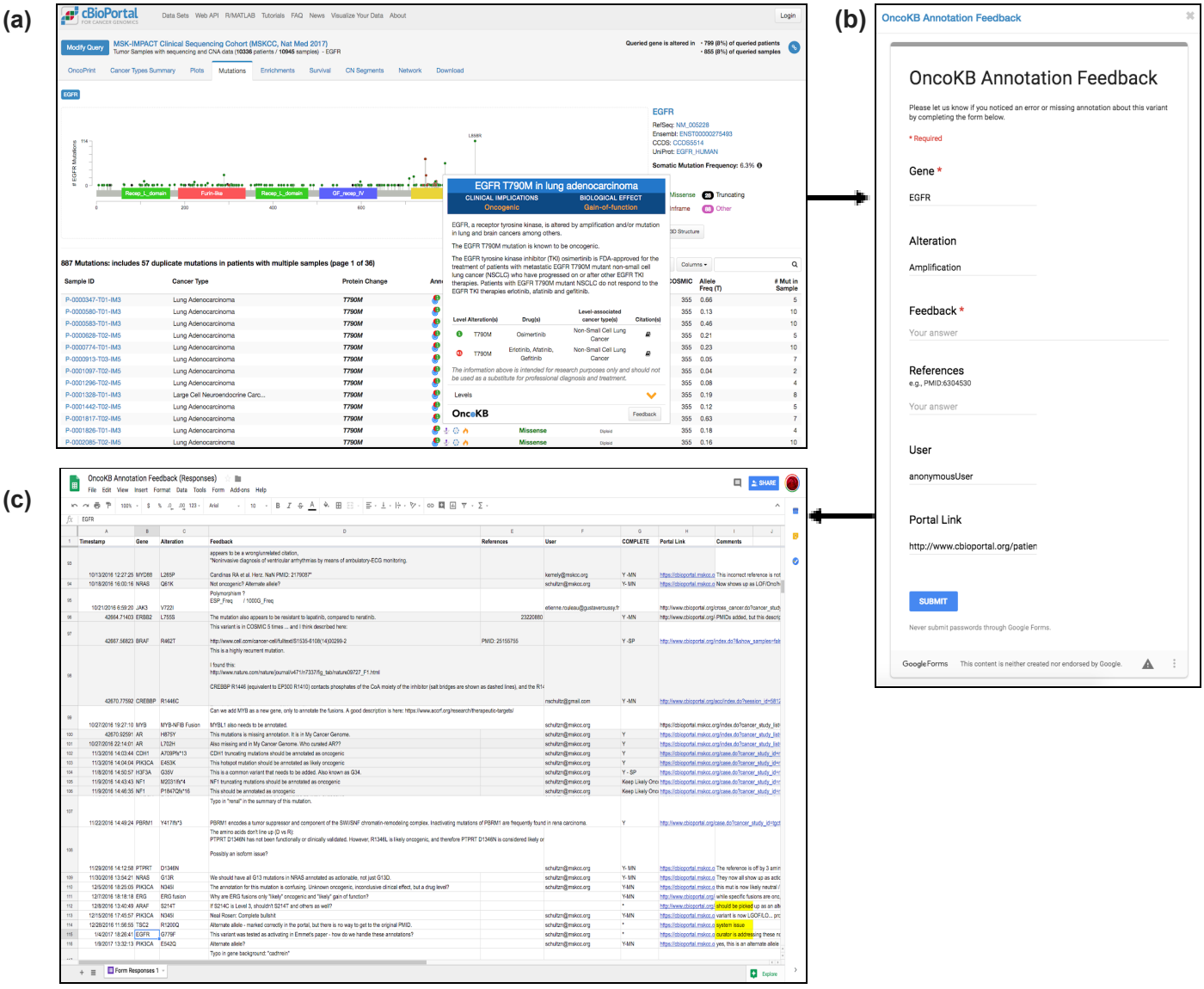


Figure 40: OncoKB Feedback through cBioPortal. On cBioPortal, if hovering over the OncoKB icon, a pop up with OncoKB information appears (a), clicking on the OncoKB icon in the pop-up will take users to the OncoKB homepage, clicking on the "Feedback" button in cBioPortal results in a pop-up comment card (b) that allows the user to provide feedback about the OncoKB annotation on the specific variant. User feedback is auto-populated into a google spreadsheet (c) which the OncoKB SCMT accesses and uses to answer user questions with a 48-hour turnaround period.

APPENDIX

Appendix I. OncoKB icons in cBioPortal.
For each oncogenic effect, the most common biological effects assigned to OncoKB variants are shown.

OncoKB Icon	Oncogenic Effect	Biological Effect
	Oncogenic	Gain-of-Function (GOF) / Likely GOF Loss-of-Function (LOF) / Likely LOF Switch-of-Function (SOF) / Likely SOF
	Likely Oncogenic	Likely GOF Likely LOF Likely SOF
	Likely Neutral	Neutral Likely Neutral
	Inconclusive	Inconclusive
	SCMT reviewed Variant of Unknown Significance (VUS)	SCMT reviewed VUS
	Unknown (SCMT non-reviewed VUS)	Unknown (SCMT non-reviewed VUS)

Appendix II. OncoKB Levels of Evidence and their icons in cBioPortal.
Variants with clinical implications are given a specific OncoKB icon in cBioPortal as described here.

Level of Evidence (per Chakravarty et al., 2017)	OncoKB Icon in cBioPortal
1 **FDA-recognized** biomarker predictive of response to an **FDA-approved drug** in this indication	
2 **Standard care** biomarker recommended by the NCCN or other expert panels predictive of response to an **FDA-approved drug** in this indication	
3A **Compelling clinical evidence** supports the biomarker as being predictive of response to a drug in this indication	
3B **Standard care** or **investigational** biomarker predictive of response to an **FDA-approved** or **investigational** drug in another indication	
4 **Compelling biological evidence** supports the biomarker as being predictive of response to a drug	
R1 **Standard care** biomarker predictive of **resistance** to an **FDA-approved** drug in this indication	
R2 **Compelling clinical evidence** supports the biomarker as being predictive of **resistance** to a drug	

Appendix III. Protocol #1: Assertion of gene function.

Assertion of OG or TSG or Both requires at least 1 of criteria from Evidence I or II. If the evidence is weak and/conflicting, or there is insufficient evidence to classify a gene as an OG or TS, that gene will not be labeled as an OG or TS.

Evidence	**ASSERTIONS**		
	Oncogene (OG)	**Tumor Suppressor (TSG)**	**Both**
I. Weinberg, p.G:20, 2014 Vogelstein et al., 2013	***RULE OG-1*** Any of the following features as demonstrated by the scientific literature in ≥1 studies. (1) A cancer-inducing gene when activated by mutation OR (2) A gene that can transform cells by increasing the selective growth advantage of the cell in which it resides as demonstrated by the scientific literature in ≥1 studies.	***RULE TSG-1*** Any of the following features as demonstrated by the scientific literature in ≥1 studies. (1) A gene whose partial or complete inactivation by mutation, occurring in either the germline or the genome of a somatic cell, leads to an increased likelihood of cancer development by increasing the selective growth advantage of the cell in which it resides OR (2) A gene that is responsible for constraining cell proliferation OR (3) A gatekeeper, a gene that operates to hinder cell multiplication or to further cell differentiation or cell death and in this way prevents the appearance of populations of neoplastic cells 4) Mutated through protein-truncating alterations throughout their length	***RULE TSGOG-1*** Meets at least one of the criteria for both OG and TSG
II. Davoli et al., 2013	***RULE OG-2*** A gene that, in tumor samples, has i) higher functional impact as defined by the PolyPhen2 Hum-Var prediction model and higher amplification frequency in comparison to those observed in neutral genes, AND ii) lower loss-of-function mutations, splicing mutations and frequency of deletions and increased frequency of amplification compared to tumor suppressors	***RULE TSG-2*** A gene that, in tumor samples, has i) higher frequencies of loss-of-function and splicing mutations, higher functional impact, and higher frequency of deletions compared to those found in neutral genes, AND ii) higher frequencies of loss-of-function and splicing mutations, higher deletion frequency and lower amplification frequency compared to those found in oncogenes	***RULE TSGOG-2*** Meets OG AND TSG criteria

Appendix IV. Protocol #2: Assertion of the oncogenic effect of a somatic alteration.

Assertion of the oncogenic effect of an alteration (A-E) requires at least 1 of criteria from the corresponding Evidence

Assertion	Definition	Criteria	Evidence (the alteration meets any of the following criteria)
A. Oncogenic	Strong evidence shows that the alteration is established in the literature as promoting cell proliferation or other hallmark of cancer as defined by Douglas Hanahan and Robert Weinberg (Hanahan and Weinberg, 2011).	1	Compelling experimental data (e.g,. genetically engineered mouse data with the mutation) in one or more studies directly demonstrating that the alteration is oncogenic and is associated with at least one hallmark of cancer as defined by Hanahan and Weinberg
		2	The alteration is a known hotspot (Chang et al., 2018) AND there is at least one experimental study suggesting the alteration is oncogenic.
		3	The alteration has been identified in a patient who responded to a targeted inhibitor, AND at least one experimental study provides strong evidence that the alteration is oncogenic.
		4	The alteration is classified as either known gain/loss/switch-of-function AND there is at least one experimental study suggesting the alteration is oncogenic.
B. Likely Oncogenic	Evidence suggests the alteration likely promotes cell proliferation or other hallmarks of cancer as defined by Douglas Hanahan and Robert Weinberg (Hanahan and Weinberg, 2011). This criteria is more permissive than Criteria 1.	1	Representative experimental lines of data (e.g., downstream activation/inactivation of a signaling target/a hit in a high-throughput screen) in one or more studies pointing to possible oncogenic function or mutation associated with known germline syndrome.
		2	At least one experimental study provides reasonable evidence suggesting the alteration is oncogenic.
		3	The alteration is a known hotspot (Chang et al., 2018) AND there are no known functional studies describing the oncogenic potential of the alteration.
C. Likely neutral	Evidence suggests the alteration does not alter protein activity or does not confer growth or survival advantage when expressed in cells.	1	The mutation effect of the alteration is neutral or likely neutral.
		2	At least one experimental study provides reasonable evidence suggesting the alteration is likely neutral.
D. Inconclusive	There is conflicting and/or weak data describing the oncogenic effect of the mutant alteration	1	Conflicting data exists as to the oncogenic effect of the alteration.
		2	Data is limited to “weak” experimental data describing the oncogenic effect of the alteration (small, under-powered experimental studies in one or multiple publications).
		3	Data is limited to studies demonstrating either patient and/or in vitro sensitivity/resistance to a targeted drug.
		4	Data is limited to in silico studies that predict the oncogenic effect of the alteration.

Appendix V. Protocol #3: Assertion of the biological effect of a somatic alteration.

Assertion of the biological effect of an alteration requires at least 1 of criteria from Assertion Type I (only 1 Assertion Type I (A, B, C, D or E) can be chosen for each variant) and at least 1 criteria from Assertion Type II (only 1 Assertion Type II can be chosen for each variant A or B).

ASSERTION TYPE I **Choose from A, B, C, D or E;** ***Based on any of the following criteria in each**	**AND**	**ASSERTION TYPE II** **If Type I = A / B / C / D choose from A or B;** ***Based on any of the criteria in each**	**AND**	**FINAL ASSERTION**
A: Gain of function* 1. The alteration is associated with Increased function of the protein 2. Increased gene dosage 3. Increased/ectopic mRNA expression 4. Increased/constitutive protein activity 5. Dominant negative 6. Structural protein 7. Toxic protein		**A: Known function*** 1. Compelling experimental data in one or more studies directly establishing the function of the mutation. 2. Multiple lines of data in one or more studies including but not limited to experimental data and statistical recurrence that together provide strong evidence establishing the function of the mutation. 3. The alteration is a known hotspot (Chang et al., 2018) AND at least one experimental study provides strong evidence that the alteration confers gain-, loss-, or switch-of or neutral function. 4. Rescue experiment provides evidence that the alteration is neutral. (Neutral) 5. The alteration has been identified in a patient who responded to a targeted inhibitor AND at least one experimental study provides strong evidence that the alteration confers gain-, loss-, or switch-of or neutral function. 6. Strong evidence-based data demonstrating that there is no difference in measurable cell attributes expressing either the wildtype or mutant form of the gene (Neutral).		**IA.IIA** Known Gain of function
				IB.IIA Known Loss of function
B: Loss of function* 1. The alteration is associated with decreased function of the protein 2. Haploinsufficiency				**IC.IIA** Known Switch of function
C: Switch of function* 1. The alteration is associated with a novel function of the protein 2. New protein 3. Altered substrate specificity				**ID.IIA** Known Neutral function
D: Neutral function* 1. The function of the protein is unchanged by the alteration 2. There is no difference in measurable cell attributes expressing either the wildtype or mutant form of the gene.		**B: Likely function*** 1. A single or multiple experimental studies from one publication including but not limited to experimental data or statistical recurrence establishing the function of the mutation 2. The alteration is a known hotspot (Chang et al., 2018), and there are no known functional studies describing the mutation effect of the alteration. 3. While conflicting evidence may exist, there is a reasonable assumption based on the data suggesting the alteration confers gain-, loss-, or switch-of or neutral function. 4. The alteration has been identified in a patient who responded to a targeted inhibitor AND at least one experimental study provides limited evidence that the alteration confers gain-, loss-, or switch-of-function. 5. Probable, possible, and/or evidence-based data suggesting that there is no difference in measurable cell attributes expressing either the wildtype or mutant form of the gene (Likely neutral).		**IA.IIB** Likely Gain of function
E: Inconclusive function* 1. Conflicting data exists as to the mutational effect of the alteration. 2. Data is limited to “weak” experimental data describing the mutational effect of the alteration (small, under-powered experimental studies in one or multiple publications). 3. Data is limited to studies demonstrating patient and/or in vitro sensitivity/resistance to a drug. 4. Data is limited to in silico studies that predict the mutation effect of the alteration.				**IB.IIB** Likely Loss of function
				IC.IIB Likely Switch of function
				ID.IIB Likely Neutral function
				E Inconclusive

Appendix VI. Protocol #4A: Detailed criteria for assertion of an OncoKB level of evidence of an alteration.

The following protocol outlines [I.] Treatment Guidelines (A. FDA drug labels or B. Disease-specific NCCN guidelines) and supporting [II.] Scientific Evidence (C. Clinical data or D. Preclinical data) required to assert an OncoKB level of evidence to an alteration.

<table>
<tr><td colspan="2">OncoKB Levels of Evidence</td><td>1</td><td>R1</td><td>2</td><td>3A</td><td>3B</td><td>4</td><td>R2</td></tr>
<tr><td colspan="2">DATA SOURCE TYPE</td><td colspan="7">CRITERIA</td></tr>
<tr><td rowspan="4">I. Treatment Guidelines</td><td rowspan="2">A. FDA-drug labels</td><td>1. Variant must be specified in the FDA-drug label as a FDA-recognized biomarker of response.</td><td colspan="6"></td></tr>
<tr><td colspan="3">2. Must be an FDA-approved drug.</td><td>3. FDA-approved drug OR drug being tested via enrollment in a clinical trial with compelling clinical data.</td><td>4. FDA-approved OR drug is being tested via enrollment in a clinical trial with compelling clinical data in another indication</td><td>5. FDA-approved drug OR drug being tested via enrollment in a clinical trial with compelling clinical data OR drug that has recently been tested via enrollment in a clinical trial but the data is not yet mature to assess for level 3A status.</td></tr>
<tr><td rowspan="3">B. Disease-specific NCCN guidelines</td><td colspan="3">1. Variant is described as predictive biomarker of response (or resistance for R1) to an FDA- approved targeted therapy at NCCN Level 2A or higher. (This is often, but not always the case for Level 2)</td><td></td><td>4. Variant is described as predictive biomarker at NCCN Level 2A or higher in another tumor type.</td><td></td></tr>
<tr><td colspan="2"></td><td colspan="2">2. If the variant is FDA-recognized as a germline biomarker predictive of response to an FDA-approved targeted therapy AND there is clinical data demonstrating patient response to the same targeted therapy in the somatic setting.</td><td colspan="3" rowspan="2"></td></tr>
<tr><td colspan="2"></td><td colspan="2">3. If the targeted therapy is FDA-approved in an</td></tr>
</table>

			indication where the predictive variant biomarker is pathognomonic to the indication, the variant is considered Level 2 or 3A based on the available clinical data.			
II. Scientific Evidence	**C. Clinical data**	**1.** Prospective randomized/non-randomized clinical trials in a specific tumor type with survival endpoints.		**7.** Criteria C1, C2, C3, C4 C5 or C6 **in another tumor type.**		**10.** Prospective randomized clinical trials in a specific tumor type with tumor **resistance** data but no survival endpoints.
		2. Prospective randomized/non-randomized clinical trial in a specific tumor type with tumor response data.				
		3. Basket clinical trials with tumor response data.				**11.** Retrospective clinical study with tumor **resistance** data in a specific tumor type comparing variant positive vs. negative cohorts.
			4. Retrospective clinical study with tumor response data in a specific tumor type comparing variant positive vs. negative cohorts.			
			5. Clinical case series (n ≥ 3 pts) demonstrating **response** associated with variants in specific tumor type with supporting preclinical data.		**8.** One or 2 clinical case study(s) in a tumor type with supporting preclinical data.	**12.** Clinical case series (n≥3 pts) demonstrating **resistance** associated with variants in specific tumor type with supporting preclinical data.
			6. Multiple single clinical case studies in specific tumor type with supporting		**9.** Multiple case reports (n≥3) for the variant in a specific tumor type but absence of supporting preclinical data.	

			preclinical data (n ≥3 pts).			
II. Scientific Literature	**D. Preclinical data**	**1.** Preclinical studies connecting the variant to response to a targeted therapeutic using *in vivo* or *in vitro* model systems.			**3.** May or may not have supporting preclinical data.	**4.** Preclinical studies connecting the variant to **resistance** to a targeted therapeutic using *in vivo* or *in vitro* model systems.
			2. Eligibility criteria in an ongoing clinical trial or in a trial that has recently closed but for which the survival outcomes or tumor response data is still not mature.			

Appendix VII. Protocol #4B: Required criteria for assertion of an OncoKB level of evidence of an alteration.

The following protocol outlines the required logic to assign an alteration an OncoKB level of evidence.

OncoKB Levels of Evidence		1	R1	2	3A	3B	4	R2
DEFINITIONS		FDA-recognized biomarker predictive of response to an FDA-approved drug in this indication **(A1 AND A2 AND B1 AND C1/2/3 AND D1)**	Standard care biomarker predictive of resistance to an FDA-approved drug in this indication **(A2 AND B1 AND C1/2/3 AND D1)**	Standard care biomarker recommended by the NCCN or other expert panels predictive of response to an FDA-approved drug in this indication **(A2 AND B1/2/3 AND C1/2/3 AND D1)**	Compelling clinical evidence supports the biomarker as being predictive of response to a drug in this indication **(A3 AND B2/3 or C1/2/3/4/5/6 AND D1/2)**	Standard care or investigational biomarker predictive of response to an FDA-approved or investigational drug in another indication **(A4 OR B4 AND C7 AND D1/2)**	Compelling biological evidence supports the biomarker as being predictive of response to a drug **(A5 AND C8/9 AND D2/3)**	Compelling clinical evidence supports the biomarker as being predictive of resistance to a drug **(A5 AND C10/11/12 AND D4)**
REQUISITE CRITERIA	**A. FDA-drug labels**	A1 ***AND*** A2	A2	A2	A3	A4	A5	A5
	B. Disease specific NCCN guidelines	***AND*** B1		***AND*** B1, B2 ***OR*** B3	***AND EITHER*** B2 B3, C1, C2, C3, C4, C5 **OR** C6	***OR*** B4	NA	
	C. Clinical data	***AND EITHER*** C1, C2 ***OR*** C3				***AND*** C7	***AND EITHER*** C8 ***OR*** C9	***AND EITHER*** C10*, C11 ****OR*** C12*
	D. Preclinical data	***AND*** D1			***AND EITHER*** D1 ***OR*** D2		***AND EITHER*** D2 ***OR*** D3	***AND*** D4

Appendix VIII. Mapping the OncoKB levels of evidence to the FDA levels of evidence.

Below are the rules for mapping variants with an OncoKB Level of Evidence (Level 1-3A and Level R1 and R2) to the FDA Levels of Evidence. OncoKB leveled variants do not map to FDA Level of Evidence 1 since there are no corresponding CDx tests.

DEFINITION OF ONCOKB LEVEL OF EVIDENCE	ONCOKB LEVEL OF EVIDENCE	FDA LEVEL OF EVIDENCE	DEFINITION OF FDA LEVEL OF EVIDENCE
Does not map to an OncoKB Level of Evidence		**1**	**Companion diagnostics (CDx)** are tests that provide information that is essential for the safe and effective use of a corresponding therapeutic product, such as a drug. Tumor profiling NGS tests may include CDx claims that are prescriptive for a specific therapeutic product, such as the **Table 1** claims listed in the intended use for the Oncomine Dx Target Test and FoundationOne CDx. Such claims are supported by analytical validity of the test for each specific biomarker and a clinical study establishing either the link between the result of that test and patient outcomes or clinical concordance to a previously approved CDx.
FDA-recognized biomarker predictive of response to an FDA-approved drug in this indication with analytical validity based on the mutation itself	**1**	**2**	**Cancer Mutations with Evidence of Clinical Significance** Tests for biomarkers described as cancer mutations with evidence of clinical significance enable health care professionals to use information about their patients' tumors in accordance with the clinical evidence, such as clinical evidence presented in professional guidelines, as appropriate. Such claims are supported by a demonstration of analytical validity (either on the mutation itself or via a representative approach, when appropriate) and clinical validity (typically based on publicly available clinical evidence, such as professional guidelines and/or peer-reviewed publications).
Standard care biomarker recommended by the NCCN or other expert panels predictive of response to an FDA-approved drug in this indication with analytical validity based on the mutation itself	**2**		
Standard care biomarker predictive of resistance to an FDA-approved drug in this indication with analytical validity based on the mutation itself	**R1**		
FDA-recognized or standard care biomarker supported by analytical validity via a representative approach	**1 or 2**	**3**	**Cancer Mutations with Potential Clinical Significance** Mutations not considered biomarkers in Level 1 or Level 2 can be described as cancer mutations with potential clinical significance. These mutations may be informational or used to direct patients towards clinical trials for which they may be eligible. Such claims are supported by analytical validation, principally through a representative approach, when appropriate, and clinical or mechanistic rationale for inclusion in the panel. Such rationales would include peer-reviewed publications or in vitro pre-clinical models.
Compelling clinical evidence supports the biomarker as being predictive of response to a drug in this indication	**3A**		
Standard care or investigational biomarker predictive of response to an FDA-approved or investigational drug in another indication	**3B**	***Does not map to an FDA Level of Evidence***	

REFERENCES

1. Chakravarty, D, Gao, JJ, Phillips, SM, Kundra, R, Zhang, H, Wang, J, Rudolph, JE, Yaeger, R, Soumerai, T, Nissan, MH, Matthew T. Chang, Sarat Chandarlapaty, Tiffany A. Traina, Paul K. Paik, Alan L. Ho, Feras M. Hantash, Andrew Grupe, Shrujal S. Baxi, Margaret K. Callahan, Alexandra Snyder, Ping Chi, Daniel C. Danila, Mrinal Gounder, James J. Harding, Matthew D. Hellmann, Gopa Iyer, Yelena Y. Janjigian, Thomas Kaley, Douglas A. Levine, Maeve Lowery, Antonio Omuro, Michael A. Postow, Dana Rathkopf, Alexander N. Shoushtari, Neerav Shukla, Martin H. Voss, Ederlinda Paraiso, Ahmet Zehir, Michael F. Berger, Barry S. Taylor, Leonard B. Saltz, Gregory J. Riely, Marc Ladanyi, David M. Hyman, José Baselga, Paul Sabbatini, David B. Solit, and Nikolaus Schultz. **OncoKB: A Precision Oncology Knowledge Base** *JCO Precision Oncology* (2017) 1:1, 1-16

2. Chang MT, Bhattarai TS, Schram AM, Bielski CM, Donoghue MTA, Jonsson P, Chakravarty D, Phillips S, Kandoth C, Penson A, Gorelick A, Shamu T, Patel S, Harris C, Gao J, Sumer SO, Kundra R, Razavi P, Li BT, Reales DN, Socci ND, Jayakumaran G, Zehir A, Benayed R, Arcila ME, Chandarlapaty S, Ladanyi M, Schultz N, Baselga J, Berger MF, Rosen N, Solit DB, Hyman DM, Taylor BS. **Accelerating Discovery of Functional Mutant Alleles in Cancer.** *Cancer Discovery.* (2018) Feb;8(2):174-183. doi: 10.1158/2159-8290.CD-17-0321. Epub 2017 Dec 15. PMID: 29247016

3. Wu K, Hinson SR, Ohashi A, Farrugia D, Wendt P, Tavtigian SV, Deffenbaugh A, Goldgar D, Couch FJ. **Functional evaluation and cancer risk assessment of BRCA2 unclassified variants.** *Cancer Research* (2005) Jan 15;65(2):417-26.PMID: 15695382

4. Chang MT, Asthana S, Gao SP, Lee BH, Chapman JS, Kandoth C, Gao J, Socci ND, Solit DB, Olshen AB, Schultz N, Taylor BS. **Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity.** *Nature Biotechnology* (2016) Feb;34(2):155-63. doi: 10.1038/nbt.3391 Epub 2015 Nov 30. PMID: 26619011

5. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, Srinivasan P, Gao J, Chakravarty D, Devlin SM, Hellmann MD, Barron DA, Schram AM, Hameed M, Dogan S, Ross DS, Hechtman JF, DeLair DF, Yao J, Mandelker DL, Cheng DT, Chandramohan R, Mohanty AS, Ptashkin RN, Jayakumaran G, Prasad M, Syed MH, Rema AB, Liu ZY, Nafa K, Borsu L, Sadowska J, Casanova J, Bacares R, Kiecka IJ, Razumova A, Son JB, Stewart L, Baldi T, Mullaney KA, Al-Ahmadie H, Vakiani E, Abeshouse AA, Penson AV, Jonsson P, Camacho N, Chang MT, Won HH, Gross BE, Kundra R, Heins ZJ, Chen HW, Phillips S, Zhang H, Wang J, Ochoa A, Wills J, Eubank M, Thomas SB, Gardos SM, Reales DN, Galle J, Durany R, Cambria R, Abida W, Cercek A, Feldman DR, Gounder MM, Hakimi AA, Harding JJ, Iyer G, Janjigian YY, Jordan EJ, Kelly CM, Lowery MA, Morris LGT, Omuro AM, Raj N, Razavi P, Shoushtari AN, Shukla N, Soumerai TE, Varghese AM, Yaeger R, Coleman J, Bochner B, Riely GJ, Saltz LB, Scher HI, Sabbatini PJ, Robson ME, Klimstra DS, Taylor BS, Baselga J, Schultz N, Hyman DM, Arcila ME, Solit DB, Ladanyi M, Berger MF. **Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients.** *Nature Medicine.* (2017) Jun;23(6):703-713. doi: 10.1038/nm.4333. Epub 2017 May 8. PMID: 28481359

6. Weinberg RA. **The Biology of Cancer.** Garland Science; New York: 2007. p.G:20, G28 and 82.

7. Vogelstein B1, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. **Cancer genome landscape**. *Science.* (2013) Mar 29;339(6127):1546-58. doi: 10.1126/science.1235122. PMID: 23539594

8. Davoli T, Xu AW, Mengwasser KE, Sack LM, Yoon JC, Park PJ, Elledge SJ. **Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome.** *Cell.* (2013) Nov 7;155(4):948-62. doi: 10.1016/j.cell.2013.10.011. Epub 2013 Oct 31. PMID: 24183448

9. National Cancer Institute [Internet]. Bethesda, MD: National Cancer Institute (US); [cited 2014 Dec 04]. NCI Dictionaries; [cited 2014 Dec 04]. Available from:http://www.cancer.gov/dictionary

10. Hanahan D, Weinberg RA. **Hallmarks of cancer: the next generation.** *Cell.* (2011) Mar 4;144(5):646-74. doi: 10.1016/j.cell.2011.02.013. Review. PMID: 21376230

11. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. **The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.** *Cancer Discovery* (2012) May;2(5):401-4. doi: 10.1158/2159-8290.CD-12-0095. PMID: 22588877

12. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. **Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.** *Science Signaling*. (2013) Apr 2;6(269):pl1. doi: 10.1126/scisignal.2004088.PMID: 23550210