Annotation Guidelines for ChemDisGene corpus

Introduction

We have generated a labeled corpus, called ChemDisGene, of pairwise interactions between chemicals, genes/gene products and diseases mentioned in scholarly papers (title and abstract). The corpus consists of (1) a large set of abstracts with relationships automatically derived from the Comparative Toxicogenomics Database (CTD) (Davis et al., 2020) using distant labeling and (2) a smaller corpus, manually annotated by domain experts. The labeled corpus is intended to be used as test data for machine-learning models.

This document outlines the annotation guidelines used by our domain experts for the manual labeling task, including the definitions of the relation classes and the curation rules. The guidelines evolved through several iterative annotation cycles, taking into account feedback and suggestions from the curators where rules needed to be more explicit.

Chemical, gene and disease entity mentions in titles and abstracts were detected by Pubtator Central (Wei et al., 2019) Pubtator annotations are linked to MeSH identifiers for diseases and chemical names and NCBI Gene identifiers for genes/gene products. The definitions for chemicals, genes/gene products and diseases used for the ChemDisGene corpus are therefore defined by the inclusion criteria for MeSH categories for Chemicals and Drugs and Diseases and NCBI Gene, respectively.

We included 10 relationship classes (see Annotation schema), some of which are further qualified by a ‘degree’, defining a total of 18 relation types. Some of the CTD chemical-disease interaction types that occur rarely were abstracted.

Pubtator Central has an F1 score for each entity type in the range 0.84–0.90 (Wei et al., 2019). The guidelines therefore include rules for entity mentions erroneously detected by Pubtator.

The guidelines then describe the curation steps:
(1) review of each relationship derived from CTD and approve or reject it based on the evidence provided in the title/abstract; and
(2) addition of all other relationships expressed in the title/abstract.

References

Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., Wiegers, J., Wiegers, T. C., and Mattingly, C. J. (2020). Comparative Toxicogenomics Database (CTD): update 2021. *Nucleic Acids Research*, 49(D1):D1138–D1143, 10.

Wei, C.-H., Allot, A., Leaman, R., and Lu, Z. (2019). PubTator Central: Automated concept annotation for biomedical full text articles. *Nucleic Acids Research*, 47(W1):W587–W593, 05.

Annotation schema: pairwise interaction types

For specific examples, see **Relationship Examples** below

Entities	That have this interaction/relationship	Relationship is defined as
Chemical - Disease	**Marker/mechanism**	A chemical that correlates with a disease (e.g., increased abundance in the brain of chemical X correlates with Alzheimer disease) or may play a role in the etiology of a disease (e.g., exposure to chemical X causes lung cancer).
Chemical - Disease	**Therapeutic**	A chemical that has a known or potential therapeutic role in a disease (e.g., chemical X is used to treat leukemia).
Chemical - Gene	Increases, decreases or affects **Expression**	Changes the expression of a gene product.
Chemical - Gene	Increases, decreases or affects **Activity**	Changes an elemental function of a molecule.
Chemical - Gene	Increases, decreases or affects **Metabolic processing**	Changes the biochemical alteration of a molecule's structure. **This does not include changes in expression, stability, folding, localization, splicing, or transport)** This includes: · acylation · alkylation · amination · carbamoylation · carboxylation · chemical synthesis · degradation (catabolism or breakdown) · cleavage (incl hydrolysis) · ethylation · glutathionylation · glycation · glycosylation · hydroxylation · lipidation · methylation · nitrosation · nucleotidylation

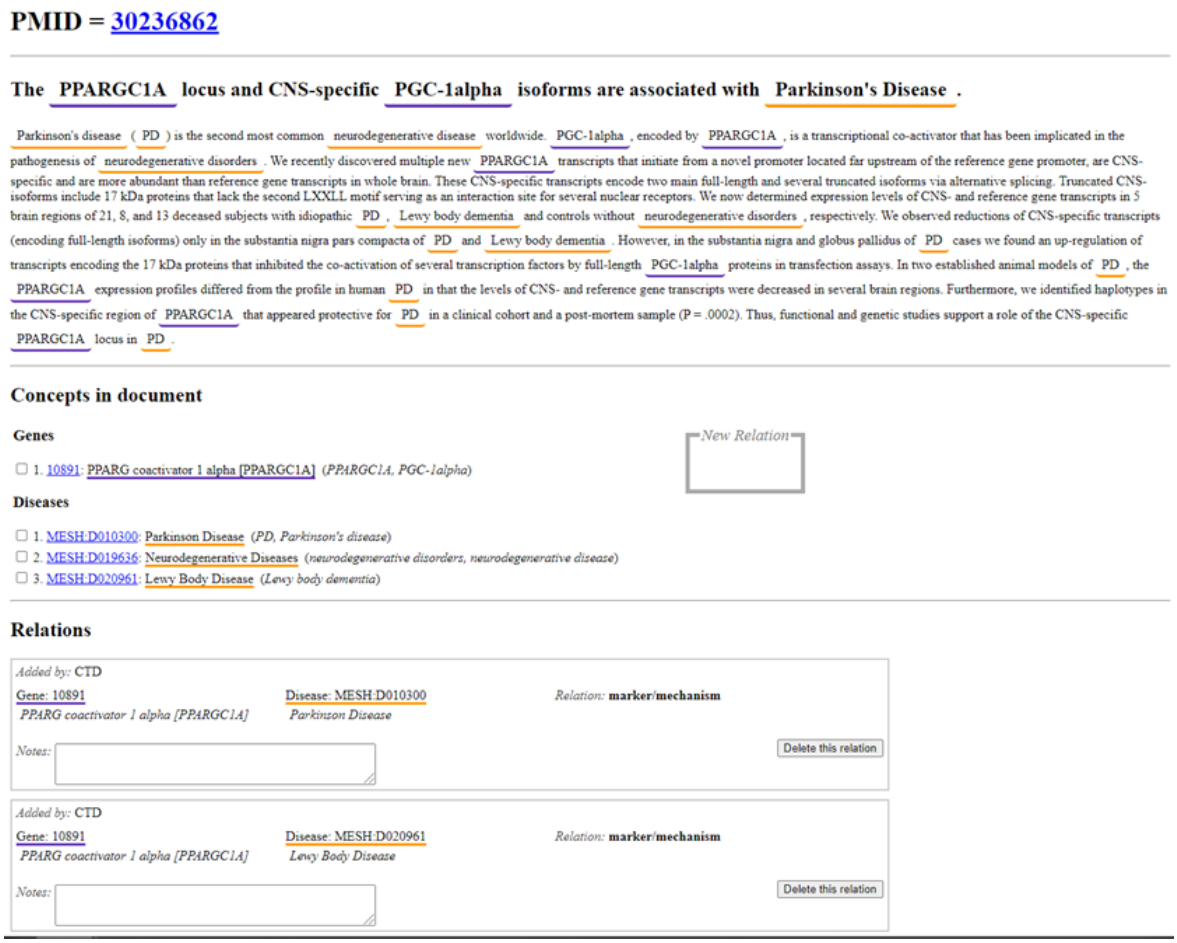
		· oxidation · phosphorylation · reduction · ribosylation · sulfation · sumoylation · ubiquitination (for more detailed definitions, see CTD glossary)
Chemical - Gene	Increases, decreases or affects **Transport**	Changes the movement of a molecule into or out of a cell. This includes: · secretion (the movement of a molecule out of a cell; by less specific means than export). · export (the movement of a molecule out of a cell; by more specific means than secretion). · uptake (the movement of a molecule into a cell; by less specific means than import). · import (the movement of a molecule into a cell; by more specific means than uptake).
Chemical - Gene	Affects **Localization**	Affects the part of the cell where a molecule resides
Chemical - Gene	Affects **Binding**	Affects the molecular interaction
Gene - Disease	**Marker/mechanism**	A gene that may be a biomarker of a disease (e.g., increased expression of gene X correlates with breast cancer) or play a role in the etiology of a disease (e.g., mutations in gene X causes liver cancer)..
Gene - Disease	**Therapeutic**	A gene that is or may be a therapeutic target in the treatment a disease (e.g., targeted reduction of gene X expression reduces susceptibility to emphysema).

Curation task

The goal is to label all chemical-gene, chemical-disease, gene-disease relationships that are clearly expressed in the title or abstract. The relation may be expressed anywhere in the text, including across several sentences, but it should be unambiguously implied by the text. Annotators must not use their prior knowledge to assign relationships.

Annotators should consider the final purpose of the project: if a user wants to find all papers mentioning the relationship “Chemical A is a drug for Disease X” [Mesh# <therapeutic> Mesh#] should this paper be detected? Would it be clear from the abstract why this paper shows in the result?

The task is facilitated by a user interface providing the document title and abstract, a list of concepts detected by Pubtator, and the Relations curated by CTD. Mentions of entities (Chemicals - Genes - Diseases) are highlighted and links to the corresponding MeSH ontology and NCBI Gene records are provided. Only relationships for which a correct “detected entity” is available should be labeled; only entities that are recognised by Pubtator can be labeled.
Example document:



PMID = 30236862

The PPARGC1A locus and CNS-specific PGC-1alpha isoforms are associated with Parkinson's Disease .

Parkinson's disease (PD) is the second most common neurodegenerative disease worldwide. PGC-1alpha , encoded by PPARGC1A , is a transcriptional co-activator that has been implicated in the pathogenesis of neurodegenerative disorders . We recently discovered multiple new PPARGC1A transcripts that initiate from a novel promoter located far upstream of the reference gene promoter, are CNS-specific and are more abundant than reference gene transcripts in whole brain. These CNS-specific transcripts encode two main full-length and several truncated isoforms via alternative splicing. Truncated CNS-isoforms include 17 kDa proteins that lack the second LXXLL motif serving as an interaction site for several nuclear receptors. We now determined expression levels of CNS- and reference gene transcripts in 5 brain regions of 21, 8, and 13 deceased subjects with idiopathic PD , Lewy body dementia and controls without neurodegenerative disorders , respectively. We observed reductions of CNS-specific transcripts (encoding full-length isoforms) only in the substantia nigra pars compacta of PD and Lewy body dementia . However, in the substantia nigra and globus pallidus of PD cases we found an up-regulation of transcripts encoding the 17 kDa proteins that inhibited the co-activation of several transcription factors by full-length PGC-1alpha proteins in transfection assays. In two established animal models of PD , the PPARGC1A expression profiles differed from the profile in human PD in that the levels of CNS- and reference gene transcripts were decreased in several brain regions. Furthermore, we identified haplotypes in the CNS-specific region of PPARGC1A that appeared protective for PD in a clinical cohort and a post-mortem sample (P = .0002). Thus, functional and genetic studies support a role of the CNS-specific PPARGC1A locus in PD .

Concepts in document

Genes

1. 10891: PPARG coactivator 1 alpha [PPARGC1A] (*PPARGC1A, PGC-1alpha*)

New Relation

Diseases

1. MESH:D010300: Parkinson Disease (*PD, Parkinson's disease*)
2. MESH:D019636: Neurodegenerative Diseases (*neurodegenerative disorders, neurodegenerative disease*)
3. MESH:D020961: Lewy Body Disease (*Lewy body dementia*)

Relations

Added by: CTD
Gene: 10891 — *PPARG coactivator 1 alpha [PPARGC1A]*
Disease: MESH:D010300 — *Parkinson Disease*
Relation: **marker/mechanism**
Notes:
Delete this relation

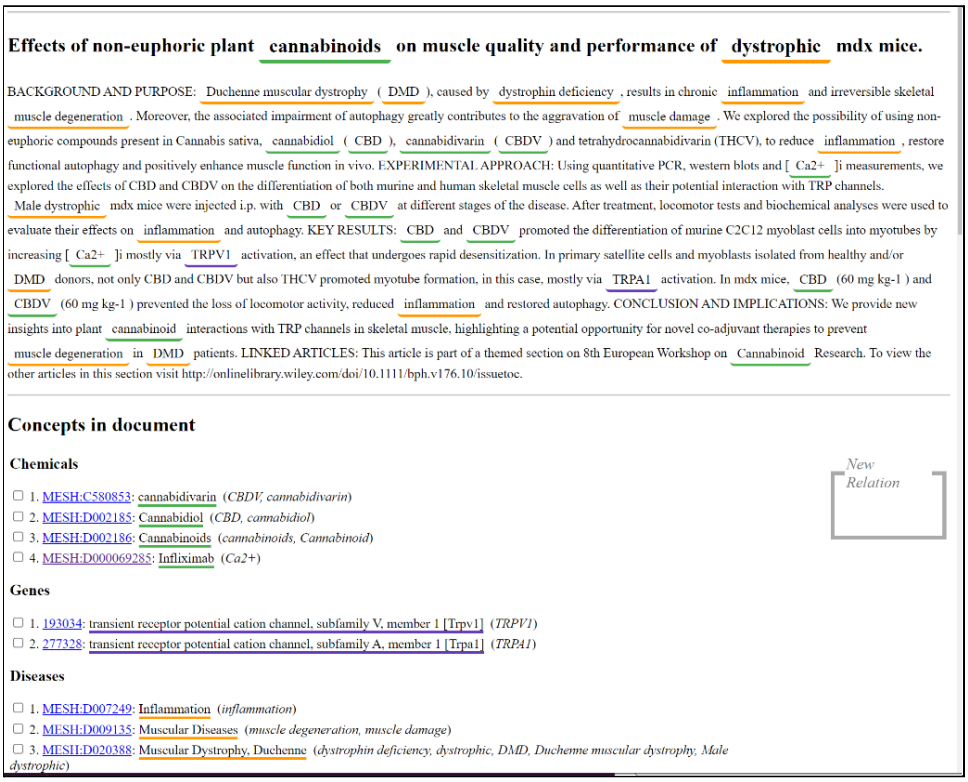
Added by: CTD
Gene: 10891 — *PPARG coactivator 1 alpha [PPARGC1A]*
Disease: MESH:D020961 — *Lewy Body Disease*
Relation: **marker/mechanism**
Notes:
Delete this relation

The task can be broken down into 3 steps:

1. Review the detected entities and identify the entities that can be included in an interaction pair.
2. Review the relationships already recorded (curated by CTD)
3. Add missing relationships

Step 1: Review the detected entities in “Concepts in document”

Annotators should consider: *Is the detected entity correct and mentioned anywhere in the abstract?*



Effects of non-euphoric plant cannabinoids on muscle quality and performance of dystrophic mdx mice.

BACKGROUND AND PURPOSE: Duchenne muscular dystrophy (DMD), caused by dystrophin deficiency , results in chronic inflammation and irreversible skeletal muscle degeneration . Moreover, the associated impairment of autophagy greatly contributes to the aggravation of muscle damage . We explored the possibility of using non-euphoric compounds present in Cannabis sativa, cannabidiol (CBD), cannabidivarin (CBDV) and tetrahydrocannabidivarin (THCV), to reduce inflammation , restore functional autophagy and positively enhance muscle function in vivo. EXPERIMENTAL APPROACH: Using quantitative PCR, western blots and [Ca2+]i measurements, we explored the effects of CBD and CBDV on the differentiation of both murine and human skeletal muscle cells as well as their potential interaction with TRP channels. Male dystrophic mdx mice were injected i.p. with CBD or CBDV at different stages of the disease. After treatment, locomotor tests and biochemical analyses were used to evaluate their effects on inflammation and autophagy. KEY RESULTS: CBD and CBDV promoted the differentiation of murine C2C12 myoblast cells into myotubes by increasing [Ca2+]i mostly via TRPV1 activation, an effect that undergoes rapid desensitization. In primary satellite cells and myoblasts isolated from healthy and/or DMD donors, not only CBD and CBDV but also THCV promoted myotube formation, in this case, mostly via TRPA1 activation. In mdx mice, CBD (60 mg kg-1) and CBDV (60 mg kg-1) prevented the loss of locomotor activity, reduced inflammation and restored autophagy. CONCLUSION AND IMPLICATIONS: We provide new insights into plant cannabinoid interactions with TRP channels in skeletal muscle, highlighting a potential opportunity for novel co-adjuvant therapies to prevent muscle degeneration in DMD patients. LINKED ARTICLES: This article is part of a themed section on 8th European Workshop on Cannabinoid Research. To view the other articles in this section visit http://onlinelibrary.wiley.com/doi/10.1111/bph.v176.10/issuetoc.

Concepts in document

Chemicals

- ☐ 1. MESH:C580853: cannabidivarin (*CBDV, cannabidivarin*)
- ☐ 2. MESH:D002185: Cannabidiol (*CBD, cannabidiol*)
- ☐ 3. MESH:D002186: Cannabinoids (*cannabinoids, Cannabinoid*)
- ☐ 4. MESH:D000069285: Infliximab (*Ca2+*)

New Relation

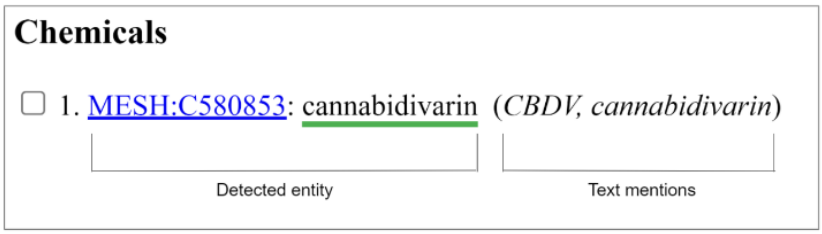
Genes

- ☐ 1. 193034: transient receptor potential cation channel, subfamily V, member 1 [Trpv1] (*TRPV1*)
- ☐ 2. 277328: transient receptor potential cation channel, subfamily A, member 1 [Trpa1] (*TRPA1*)

Diseases

- ☐ 1. MESH:D007249: Inflammation (*inflammation*)
- ☐ 2. MESH:D009135: Muscular Diseases (*muscle degeneration, muscle damage*)
- ☐ 3. MESH:D020388: Muscular Dystrophy, Duchenne (*dystrophin deficiency, dystrophic, DMD, Duchenne muscular dystrophy, Male dystrophic*)

Detected entity vs text mention:



The following points must be considered for this step:

- The entities are detected by the Pubtator annotator tool, which may introduce errors. Errors to look out for include:
-Acronym in the text linked to the incorrect entity
-Broad, potentially ambiguous terms (e.g. the text mention “developmental defect” might be incorrectly linked to the MeSH concept “Communication Disorder”)

- The underlined text in the abstract serves only as a guide for finding the mentions: the detected entity, if correct, can be used if it appears *anywhere* in the title or abstract.

- If there is any ambiguity, annotators must check the MESH (or NCBI Gene) entry, including alternative names, to decide whether this concept is indeed mentioned in the abstract.

- For gene entities, annotators must consider which species the abstract is about. If an experimental model was used, it may be necessary to look up the species. Detected gene entities that are for another species must not be used (e.g. the human ortholog entity in NCBI Gene for a relationship should not be used if the statements in the abstract are based on work carried out in a mouse cell line).

What to do when the detected entity does not unambiguously match the text mention - different cases

Case 1: Detected entity is mentioned in the text but not underlined where the relationships are stated

Example:

PMID: 30825423
CHEMICAL Carbon Tetrachloride is correctly linked once to text mention “Carbon Tetrachloride“. But the relationship statements in the text use the acronym (text mention “CCl4”), which is incorrectly linked to a wrong GENE entity.
The Carbon Tetrachloride entity can be used for relationships instead of the incorrect CCl4.

Case 2: Detected entity is obviously incorrect

Examples:

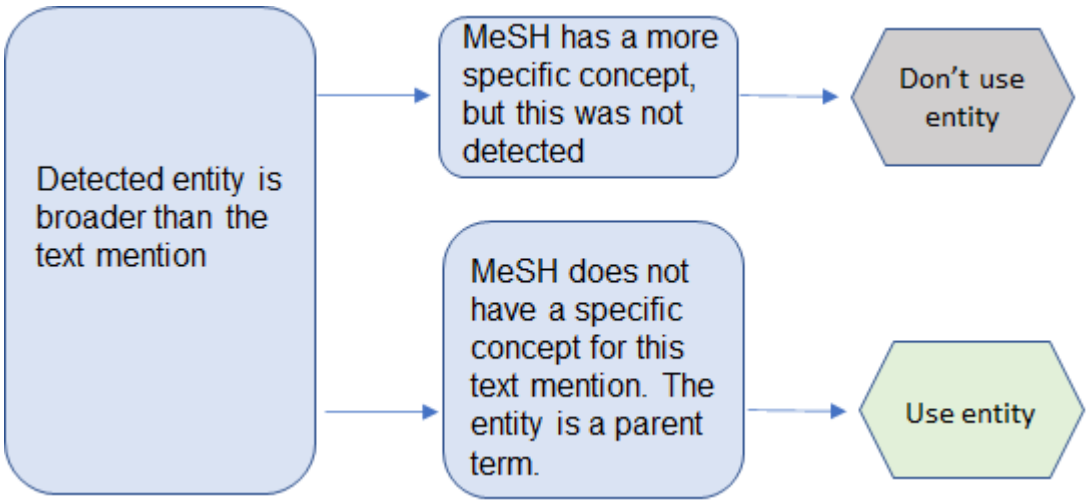
1) Wrong type of entity (e.g. chemical appears in gene category)

2) Incorrect detection of acronym:
PMID: 30773377
GENE hexokinase 2 [HK2] linked to mention “HK-2” (a cell line)

3) No biological sense:
PMID: 30615409
DISEASE Intellectual Disability linked to mention “retarded” gonadal development

Incorrect entities must not be used and relationship cannot be labeled.

Case 3: Detected entity is broader than the text mention



Examples:

MeSH does not have a specific concept, the entity is a parent term:
PMID: 30414920
CHEMICAL Silicon Dioxide linked to mention "Min-U-Sil"

PMID: 30414920
DISEASE Spinal Diseases linked to mention "Spinal deformities", which does not exist as a MeSH concept

If the detected entity is the most specific entity available in the ontology, it can be used in the relations, even if the text is talking about a more precise biological entity.

Case 4: Detected entity is more specific than the text mention

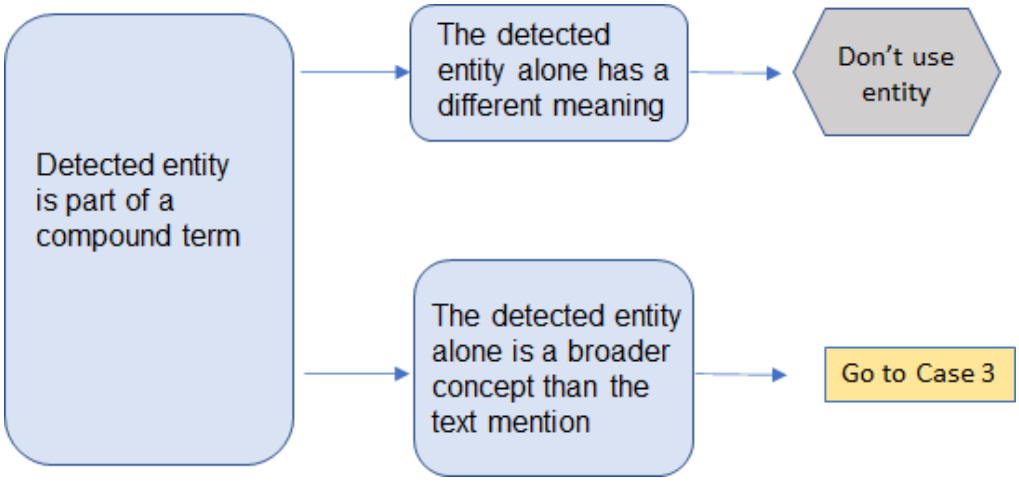
Examples:
PMID: 30391378
GENE ATP binding cassette subfamily B member 6 (Langereis blood group) [ABCB6] linked to text mention "ATP-binding cassette"

PMID: 30517881
DISEASE Drug-Related Side Effects and Adverse Reactions linked to text mention "toxicity"

This can be used if toxicity is linked to a drug but not generically for other types of toxicity.

If the detected entity more specific than the text mentions, it can be used in the relations.

Case 5: Detected entity is part of a compound term

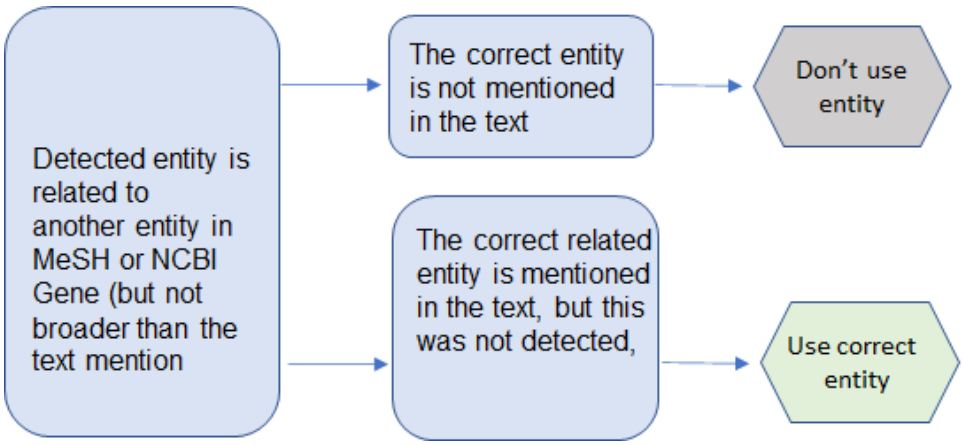


Example:

1) The detected entity alone has a different meaning than the compound text mention:
PMID: 30280449
CHEMICAL Indole linked to the text "indole-based synthetic cannabinoids"

If the detected entity was CHEMICAL Cannabinoid, this could be used as a broader entity, provided 'indole-based synthetic cannabinoids' does not exist as a concept in MeSH

Case 6: Detected entity is related (not broader or more specific) than the text mention



Example:

The correct entity is not mentioned in the text:
Detected GENE entity is for the human ortholog, but the abstract is about mouse; the human gene is not mentioned anywhere. The detected entity should not be used.

Step 2: Review the relationships already recorded (curated by CTD)

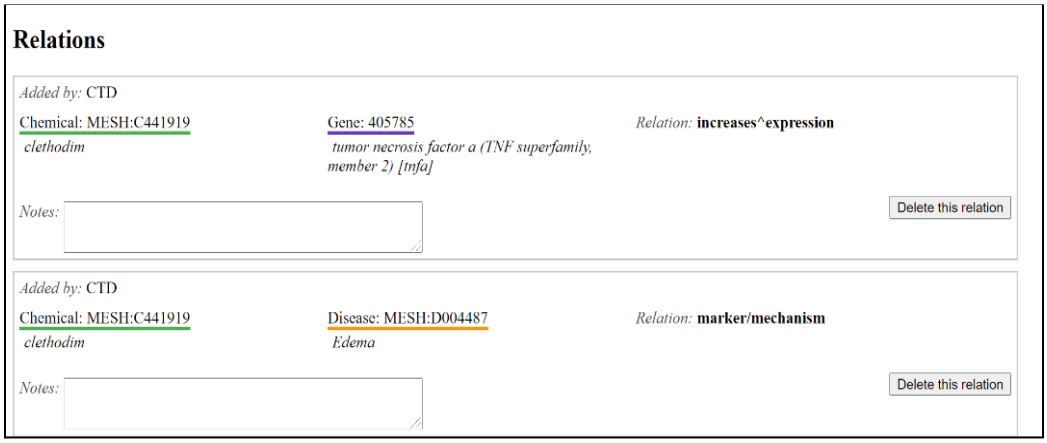
Annotators should consider: *Is the relationship recorded by CTD unambiguously stated in the abstract?*

The CTD relationships are listed in the RELATIONS section of the UI (below the detected entities/"concepts").

Relationships that are incorrect, or not supported by the title and abstract, must be removed using the "Delete this relation" button.

Annotators should bear in mind that CTD relationships may have been extracted from the full text. If the relationship is not stated in the abstract, it must be deleted.

(Points to consider when reviewing the relationships are outlined below.)

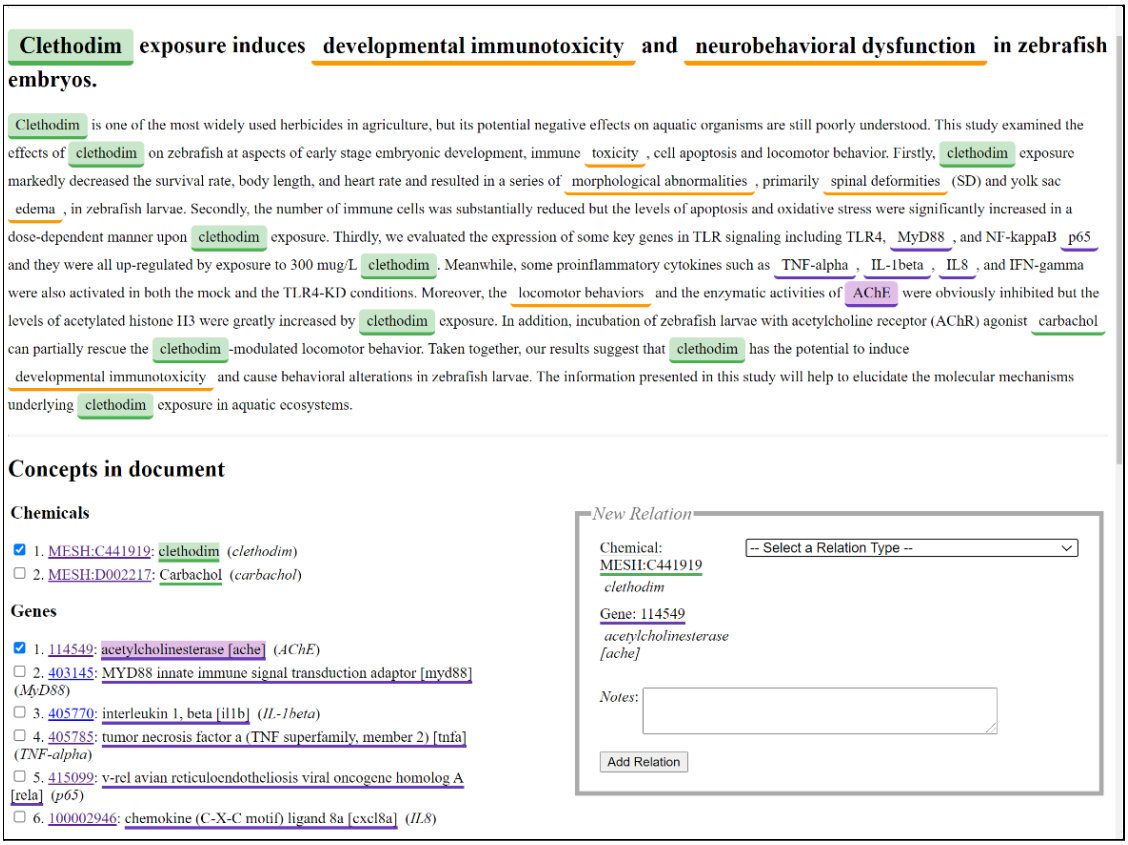


Step 3: Add missing relationships

Annotators should consider: *Are there any additional relationships mentioned in the abstract?*

The aim is to identify ALL possible relationships in a given abstract.

Additional relations can be added from the drop down. For this, the entities in the relationship are selected; only the relations possible for these entity types are available in the drop down.



Clethodim exposure induces developmental immunotoxicity and neurobehavioral dysfunction in zebrafish embryos.

Clethodim is one of the most widely used herbicides in agriculture, but its potential negative effects on aquatic organisms are still poorly understood. This study examined the effects of clethodim on zebrafish at aspects of early stage embryonic development, immune toxicity, cell apoptosis and locomotor behavior. Firstly, clethodim exposure markedly decreased the survival rate, body length, and heart rate and resulted in a series of morphological abnormalities, primarily spinal deformities (SD) and yolk sac edema, in zebrafish larvae. Secondly, the number of immune cells was substantially reduced but the levels of apoptosis and oxidative stress were significantly increased in a dose-dependent manner upon clethodim exposure. Thirdly, we evaluated the expression of some key genes in TLR signaling including TLR4, MyD88, and NF-kappaB p65 and they were all up-regulated by exposure to 300 mug/L clethodim. Meanwhile, some proinflammatory cytokines such as TNF-alpha, IL-1beta, IL8, and IFN-gamma were also activated in both the mock and the TLR4-KD conditions. Moreover, the locomotor behaviors and the enzymatic activities of AChE were obviously inhibited but the levels of acetylated histone H3 were greatly increased by clethodim exposure. In addition, incubation of zebrafish larvae with acetylcholine receptor (AChR) agonist carbachol can partially rescue the clethodim-modulated locomotor behavior. Taken together, our results suggest that clethodim has the potential to induce developmental immunotoxicity and cause behavioral alterations in zebrafish larvae. The information presented in this study will help to elucidate the molecular mechanisms underlying clethodim exposure in aquatic ecosystems.

Concepts in document

Chemicals

1. MESH:C441919: clethodim (*clethodim*)
2. MESH:D002217: Carbachol (*carbachol*)

Genes

1. 114549: acetylcholinesterase [ache] (*AChE*)
2. 403145: MYD88 innate immune signal transduction adaptor [myd88] (*MyD88*)
3. 405770: interleukin 1, beta [il1b] (*IL-1beta*)
4. 405785: tumor necrosis factor a (TNF superfamily, member 2) [tnfa] (*TNF-alpha*)
5. 415099: v-rel avian reticuloendotheliosis viral oncogene homolog A [rela] (*p65*)
6. 100002946: chemokine (C-X-C motif) ligand 8a [cxcl8a] (*IL8*)

New Relation

Chemical: MESH:C441919 *clethodim*

-- Select a Relation Type --

Gene: 114549 *acetylcholinesterase [ache]*

Notes:

Add Relation

Guidelines: What to consider when accepting or adding relationships

1. **Only add relationships that are clear from the title or abstract.**
   This can be known relationships or new relationships discovered in the paper.
   Consider whether a search for this relationship should return this paper- it should be clear from the title and abstract that this paper supports the relationship.

2. **Don't record relationships that were just investigated (and hence mentioned) for this paper, or provided as motivation, but which remain unknown and hypothetical.**

Therapeutic relationships include chemicals that have a *potential* therapeutic role, and genes that *may* be a therapeutic target; these can be recorded if their potential for treatment of the specific disease is mentioned in the abstract, or a conclusion of the paper. However, do not record a hypothetical role in another disease:
e.g. “Gene A is a therapeutic target for treatment of Disease X; it may therefore have a potential role in treatment of Disease Z” (record a relationship between Gene A and Disease X; do not record a relationship between Gene A and Disease Z).

3. **Do record all established relationships mentioned**, even if no evidence is provided here in the paper. For example, record relationships that are simply cited as known. (Controversial relationships can be recorded as there is evidence, albeit disputed.)

   For this, bear in mind that CTD may have ignored relationships that are stated but already well established and in their database. If the relationship is known (e.g. mentioned in background), do record it.

4. **Record direct relationships that can be clearly and unambiguously concluded from the abstract**.
   Indirect relationships (“two hops”) should only be recorded if they are ‘linear’ and not too complicated to understand from the context:

   For example: PMID: 32525551
   “Cr(VI) [CHEMICAL]… led to increased production of acetyl-CoA [GENE] and elevation of histone acetylation. This, in turn, up-regulated the expression of … ATP citrate lyase [GENE]”
   Record
   Cr(VI) [CHEMICAL] <increases expression> acetyl-CoA [GENE] *and*
   Cr(VI) [CHEMICAL] <increases expression> ATP citrate lyase [GENE]

   For example: PMID: 30590302
   “BPAF [CHEMICAL] can activate PI3K/Akt [GENE] and Erk [GENE] signals *via GPER*”
   Record
   BPAF [CHEMICAL] <increases activation> PI3K/Akt [GENE]

   For example: PMID: 30597128
   “CyPPA [CHEMICAL] was determined to modulate glycogen synthase kinase-3beta (GSK3beta) [GENE] activity, thereby leading to a decrease in beta-catenin/MITF [GENE] expression”
   Record
   CyPPA [CHEMICAL] <affects activity> glycogen synthase kinase-3beta (GSK3beta) [GENE] *and*
   CyPPA [CHEMICAL] <decreases expression> beta-catenin/MITF [GENE]

5. **The relationship may be inferred from the context given in the abstract**, i.e. it may not be stated directly in a single sentence.
   For example: PMID 30705370
   “We have previously identified a panel of fusion genes in aggressive prostate cancers [DISEASE]. In this study, we showed that ... CCNH-C5orf30 [Gene] and TRMT11-GRIK2 gene fusions were found in breast cancer, colon cancer…”
   Record C5orf30 [Gene] <marker/mechanism> prostate cancers [DISEASE]

6. **Relationships between two entities that depend on the presence of a third entity can be recorded:**
   For example: PMID 32589349
   The combined treatment with chemical 1 and 2 upregulates gene X, but not the single treatment
   Record
   "Chemical 1 <increases expression> of Gene X" *and*
   "Chemical 2 <increases expression> of Gene X"

   For this, the following use cases can be considered:
   1) Researcher wants to know whether Chem 1 interacts with Gene X. Recording the relationship "Chemical 1 <increases expression> of Gene X" will lead the user to this paper, which then provides additional information.
   2) Researcher wants to know about all chemicals that upregulate gene X. The answer will then contain both chemical 1 and 2, and this paper will provide additional information that both have to be together.]

7. **Do not use your prior knowledge; look for evidence for the relationship in the text.**

8. **Relationships implicit in certain terms can also be recorded.**
   For example, the text “bisphenol P showed estrogen receptor antagonistic activities“ can be recorded as < bisphenol P [chemical] decreases activity of estrogen receptor [gene] >

9. **If two opposite relationships are described in the same abstract, both should be recorded.**
   For example: PMID 30703377
   “TGHQ [CHEMICAL] -induced post-translational stabilization of Nrf2 in HK-2 cells resulted in the expected upregulation of HO1 and NQO1 [GENE] mRNA, TGHQ actually decreased Nrf2 mRNA in HL-60 cells, with a concomitant decrease in NQO1 [GENE] mRNA”
   Record
   TGHQ [CHEMICAL] increases^expression NQO1 [GENE] AND
   TGHQ [CHEMICAL] decreases^expression NQO1 [GENE]

10. **Do not record negative relationships**
    For example, do not record “treatment with chemical x did not result in…”

11. **Use degree “affects” when direction is not clear**
    For interactions “expression”, “activity”, “metabolic processing” and “transport” use the degree “affects” when the context does not make it clear whether the effect is an up-or down-regulation.

Relation examples

Chemical - Disease: Marker/mechanism

A chemical that correlates with a disease (e.g., increased abundance in the brain of chemical X correlates with Alzheimer disease) or may play a role in the etiology of a disease (e.g., exposure to chemical X causes lung cancer).

Examples	Text extract	PMID
MESH:D008942 MESH:D064420 Chemical^Disease:marker/mechanism	cytotoxicity [disease] of mitoxantrone [chemical]	30610963
MESH:C441919 MESH:D004487 Chemical^Disease:marker/mechanism	clethodim [chemical] exposure ... resulted in ... yolk sac edema [disease]	30517881
MESH:D001564 MESH:D056486 Chemical^Disease:marker/mechanism	Benzo(alpha)pyrene (BaP) [chemical] possesses a forceful hepatotoxicity [disease]	30726812
MESH:D002251 MESH:D058186 Chemical^Disease:marker/mechanism	acute kidney injury [disease] caused by CCl4 [chemical]	30825423
MESH:D005978 MESH:D003924 Chemical^Disease:marker/mechanism	Diabetic [disease] patients had markedly ... decreased levels of total GSH [chemical] in plasma	32715377

Chemical - Disease: Therapeutic

A chemical that has a known or potential therapeutic role in a disease (e.g., chemical X is used to treat leukemia).

Examples	Text extract	PMID

MESH:D005905 MESH:D003920 Chemical^Disease:therapeutic	the diabetes [disease] drug glyburide [chemical]	30610963
MESH:C070840 MESH:D009369 Chemical^Disease:therapeutic	licochalcone A [chemical] is a promising therapeutic agent for the treatment of human nasopharyngeal cancer [disease] cells	30983163
MESH:D024482 MESH:D058866 Chemical^Disease:therapeutic	combination of vitamin K2 [chemical] and PTH increased differentiation of osteoblast and had a synergistic effect on bone formation in osteoporotic calvarial bone defect [disease]	30639440
MESH:C000618475 MESH:D009369 Chemical^Disease:therapeutic	5-oxo-hexahydroquinoline [chemical] derivatives ... represent promising agents to overcome MDR in cancer [disease] cells	30391378

Chemical - Gene: Affects/Increases/Decreases Expression

Changes the expression of a gene product.
“Affects” can be used if direction not clear. However, if both increase AND decrease are reported in the same abstract (e.g. under different conditions), then both should be recorded as separate relationships.
Obvious relationships mention “expression”, “mRNA or protein levels”
Consider the context as “up/downregulation” may refer to activation.

Examples	**Text extract**	**PMID**
MESH:C087670 30390 Chemical^Gene:affects^expression	AZO [chemical] at 20.0 mug/L inhibited growth... These effects were associated with altered expression of cyp19a [gene]	30615409
MESH:C460579 3553 Chemical^Gene:affects^expression	LPS-induced IL-1beta [gene] responses .... as shown by inhibition with zYVAD-fmk [chemical]	30414920
MESH:C078765 4780 Chemical^Gene:decreases^expression	TGHQ [chemical] actually decreased Nrf2 [gene] mRNA in HL-60 cells	30703377
MESH:C552889 1499 Chemical^Gene:decreases^expression	CyPPA [chemical] was determined to modulate glycogen synthase kinase-3beta (GSK3beta) activity, thereby leading to a decrease in beta-catenin/MITF [gene] expression. "	30597128
MESH:C078765 1728 Chemical^Gene:increases^expression	TGHQ [chemical] -induced post-translational stabilization of Nrf2 in HK-2 cells resulted in the expected upregulation of HO1 and NQO1 [gene] mRNA	30703377
MESH:C441919 403145 Chemical^Gene:increases^expression	... TLR4, MyD88 [gene], and NF-kappaB p65 and they were all up-regulated by exposure to 300 mug/L clethodim [chemical]	30517881

MESH:C583074 2099 Chemical^Gene:increases^expression	BPAF [chemical] ... significantly enhances the protein expression of estrogen receptor alpha [gene]	30590302

Chemical - Gene: Affects/Increases/Decreases Activity

Changes an elemental function of a molecule
As above, “affects” can be used if direction not clear. However, if both increase AND decrease are reported in the same abstract (e.g. under different conditions), then both should be recorded as separate relationships.
Consider the context and whether the up/downregulation of a gene is in fact a change in function.
Obvious relationships can be recorded where chemicals are denoted “agonist or antagonist”, with the notion that:
An agonist is a chemical that binds to a receptor and activates the receptor to produce a biological response; an antagonist blocks the action of the agonist; an inverse agonist causes an action opposite to that of the agonist and decreases the activity of the receptor below the basal level.

Examples	**Text extract**	**PMID**
MESH:D019319 1499 Chemical^Gene:affects^activity	effects of OA [chemical] on the ... amount of transcriptionally active beta-catenin [gene]	31115591
MESH:C552889 2932 Chemical^Gene:affects^activity	CyPPA [chemical] was determined to modulate glycogen synthase kinase-3beta (GSK3beta) [gene] activity	30597128
MESH:D019319 1499 Chemical^Gene:increases^activity	significant induction of the canonical Wnt/beta-catenin [gene] -signaling pathway by OA [chemical]	31115591
MESH:C583074 5594 Chemical^Gene:increases^activity	BPAF [chemical] can activate PI3K/Akt and Erk [gene] signals"	30590302
MESH:C006780 9970 Chemical^Gene:decreases^activity	BPA [chemical] and five its analogues acted as CAR [gene] inverse agonists	30582956
MESH:C003135 2099 Chemical^Gene:increases^activity	phoxim [chemical], altrenogest and nandrolone were determined to be ER [gene] agonists	30742918
MESH:D015029 2099 Chemical^Gene:decreases^activity	zeranol [chemical] was found to exhibit weak ER [gene] antagonistic activity	30742918

Chemical - Gene: Affects/Increases/Decreases Metabolic processing

Changes the biochemical alteration of a molecule's structure.
This **does not include** changes in expression, stability, folding, localization, splicing, or transport)
This includes acylation, carboxylation, chemical synthesis, degradation (catabolism or breakdown), cleavage (incl hydrolysis), methylation, oxidation, phosphorylation, reduction. **etc**
Note that the metabolic process should be explicitly mentioned; e.g. do not infer from the mention of a signaling molecule that is typically phosphorylated (background knowledge) that phosphorylation has

occurred.

Examples	Text extract	PMID
MESH:D005576 1385 Chemical^Gene:increases^metabolic_processing	forskolin [chemical]-induced phosphorylation of the cAMP response element binding protein (CREB) [gene]	30610963
MESH:C583074 207 Chemical^Gene:increases^metabolic_processing	BPAF [chemical] … increases phosphorylation levels of protein kinase B (Akt) [gene]	30590302
MESH:C070840 142 Chemical^Gene:increases^metabolic_processing	licochalcone A [chemical] ... by the upregulation of ... cleaved-poly ADP-ribose polymerase [gene]	30983163
MESH:D008095 1576 Chemical^Gene:increases^metabolic_processing	cytochrome P450 3A (CYP3A) [gene] -mediated oxidative metabolism of lithocholic acid [chemical]	31102695
MESH:C078814 1385 Chemical^Gene:decreases^metabolic_processing	AEA [chemical] completely inhibited forskolin-induced phosphorylation of the cAMP response element binding protein (CREB) [gene]	30610963
MESH:D019319 1499 Chemical^Gene:affects^metabolic_processing	effects of OA [chemical] on the phosphorylation state, cellular redistribution as well as on the amount of transcriptionally active beta-catenin [gene]	31115591
MESH:C078765 4780 Chemical^Gene:decreases^metabolic_processing	TGHQ [chemical] -induced decreases in Nrf2 [gene] phosphorylation	30703377
MESH:D019319 1499 Chemical^Gene:decreases^metabolic_processing	effects of OA on the phosphorylation state …of beta-catenin" + "confirmed by...hypophosphorylated beta-catenin"	31115591

Chemical - Gene: Affects/Increases/Decreases Transport

Changes the movement of a molecule into or out of a cell. This includes secretion, export, uptake, import

Examples	Text extract	PMID
MESH:D008070 7124 Chemical^Gene:increases^transport	LPS- [chemical] and AFH-induced release of tumor necrosis factor alpha (TNF-alpha) [gene]	30414920

Chemical - Gene: Affects Localization

Affects the part of the cell where a molecule resides

Examples	Text extract	PMID
MESH:D019319 1499 Chemical^Gene:affects^localization	nuclear translocation of beta-catenin [gene] mediated by non-cytotoxic OA [chemical] concentrations	31115591

Chemical - Gene: Binding

Affects the molecular interaction

Examples	Text extract	PMID
MESH:C078814 1385 Chemical^Gene:affects^binding	AEA [chemical] also decreased p-CREB [gene] binding to the BCRP promoter	30610963
MESH:C006780 2104 Chemical^Gene:affects^binding	interaction between the ERRgamma [gene] ligand-binding domain (LBD) and compounds of the bisphenol [chemical]	31127318
MESH:D063388 1268 Chemical^Gene:affects^binding	It [CB1] [gene] binds several compounds in its orthosteric site, including the endocannabinoids [chemical]	30322873

Gene - Disease: Marker/mechanism

A gene that may be a biomarker of a disease (e.g., increased expression of gene X correlates with breast cancer) or play a role in the etiology of a disease (e.g., mutations in gene X causes liver cancer).

Examples	Text extract	PMID
54822 MESH:D002583 Gene^Disease:marker/mechanism	miR-543/TRPM7 [gene] axis mediated CC [disease] progression	30710498
3845 MESH:D021441 Gene^Disease:marker/mechanism	KRAS [gene] is one of the most frequently mutated proto-oncogenes in pancreatic ductal adenocarcinoma (PDAC) [disease]	30654191
1268 MESH:D009765 Gene^Disease:marker/mechanism	this receptor [CB1] [gene] is implicated in several maladies, such as obesity [disease]	30322873

Gene - Disease: Therapeutic

A gene that is or may be a **therapeutic target** in the treatment a disease (e.g. targeted reduction of gene X expression reduces susceptibility to emphysema).

Examples	Text extract	PMID
100126316 MESH:D021441 Gene^Disease:therapeutic	miR-873 [gene] nanoparticles inhibited KRAS expression and tumor growth in PDAC [disease]	30654191
100126316 MESH:D001943 Gene^Disease:therapeutic	miR-873 [gene] -based gene therapy may be a therapeutic strategy in PDAC and TNBC [disease]	30654191
2729 MESH:D003924 Gene^Disease:therapeutic	Genetic variants in glutamate cysteine ligase [gene] confer protection against type 2 diabetes [disease]	32715377