

12/15/2021

## Annotation guidelines eMIND

This is the annotation guideline document for annotation of eMIND dataset for protein variation impacts related to protein level properties, in [AD/ADRDs](#)

- Biological Process
- Protein Activity/function
- Protein Interaction: includes complex formation
- Protein Aggregation
- Protein Abundance: abundance of a gene product or proteoform (e.g., increase levels of Abeta 42)
- Protein Modification: refers to post-translational modifications
- Protein Processing: common protein processing in AD/ADRDs includes TREM2 shedding, abeta processing (Abeta42/Abeta40) ratio, protein cleavage in general
- Protein Structure: amino acid changes affect the 3D structure of all or a protein region. Also to annotate impact on property of a region, motif, etc
- Protein localization: includes variants that impact cell localization, trafficking to compartment, level of secretion, or surface expression

## Some background on specific proteins

### Amyloid beta processing

For a review on Amyloid beta processing “Making the final cut: pathogenic amyloid- $\beta$  peptide generation by  $\gamma$ -secretase” Source:doi: 10.15698/cst2018.11.162. Particularly, Figure 1.

APP is first cleaved by beta-secretase in its ectodomain close to the extracellular/luminal membrane border thereby generating a C-terminal APP fragment (C99). Consecutive cleavages of C99 by gamma-secretase releases the APP intracellular domain (AICD) into the cytosol and 37-43 amino acid amyloid beta (Abeta) species into the extracellular space or lumen of secretory pathway organelles. Longer Abeta forms, e.g., Abeta42, are highly aggregation prone. An alternative cleavage of APP by alpha-secretase in the Abeta domain prevents the formation of Abeta.

### For annotation purposes:

- Levels of secretion of Abeta forms, although this reflects the processing we will annotate as impact on protein abundance (of proteoform) and indicate target proteoform in notes
- Ratio between proteoforms should be annotated to protein processing

## TREM2 processing

Recent studies suggest that sTREM2 can be generated by either alternative splicing or proteolytic cleavage of the full-length TREM2 protein. The longest TREM2 transcript consists of five exons, with exon 4 encoding a transmembrane domain. This isoform is anchored to the cell membrane and shed by ADAM10 and/or ADAM17, leading to the production of a soluble TREM2 and a C-terminal fragment. This can be further cleaved by gamma-secretase to generate a TREM2 intracellular domain (ICD). The shortest TREM2 transcript encodes a soluble form of TREM2 due to the lack of exon 4 which encodes the transmembrane domain of the receptor. Source: <https://doi.org/10.3389/fnagi.2019.00328>

### **For annotation purposes:**

- Shedding of TREM2 refers to cleavage, annotate as impact in processing
- Soluble TREM2 annotate as impact on localization

APOE Alleles (source:DOI: 10.5772/59809)

For annotation purposes we will treat APOE4 allele similar to a variant and annotate APOE4 as protein mutation but indicate in note "AD Allele". The APOE alleles differ in their interaction with Abeta, and APOE4 has been associated with AD risk (source:DOI: [10.1007/s12264-013-1422-z](https://doi.org/10.1007/s12264-013-1422-z))

## **Annotation task**

Find sentences with variants and impact, and annotate relevant entities (see below) and the impact relation. In addition, indicate the disease context for these. We only care about AD/ADRDs.

**Annotation tool:** TeamTat (<https://www.teamtat.org/>)

**Entities:** gene, disease, mutation, impact, sentence

Name	Color	Sample
Gene		sample annotated text in a sentence
ProteinMutation		sample annotated text in a sentence
Disease		sample annotated text in a sentence
DNAMutation		sample annotated text in a sentence
impact_protein_abundance		sample annotated text in a sentence
impact_protein_aggregation		sample annotated text in a sentence
impact_protein_function_activity		sample annotated text in a sentence
impact_protein_interaction		sample annotated text in a sentence
impact_protein_localization		sample annotated text in a sentence
impact_protein_process		sample annotated text in a sentence
impact_protein_ptm		sample annotated text in a sentence
impact_protein_processing		sample annotated text in a sentence
impact_protein_structure		sample annotated text in a sentence
sentence		sample annotated text in a sentence
impact_other		sample annotated text in a sentence

## Relation:

Has\_impact to relate the mutation and its impact(s).

## Annotation guidelines

**Assign abstract as curatable if it has some relation and it is in the context of AD/ADRDs, otherwise select uncuratable.**

### 1-Only annotate abstracts

#### Entity annotations:

Abstracts are pre-tagged by Pubtator. You don't need to correct all these. Only need to correct the ones that are relevant to the annotation task. For example, if gene or disease are not captured correctly for the relation to be annotated, then this needs to be corrected (for example, if A beta 1 has been assigned to a gene that is not APP, and the right selection should have been A beta 1-42(43) instead )

**2-Highlight the mutation mention phrase (e.g., like "this mutation", "the variant") and add in the Note section the specific mutation following the [HVGs nomenclature](#).**

the variant

---

Location  
706

Entity Type  
ProteinMutation

Concept ID

Note  
p.Gly417Ala

Annotate each instance of this mention text

Update all mentions with the same concept ID (-- NO Concept ID assigned --)

Last updated by Cecilia at December 13, 2021 6:05 PM

Cancel

3-If the impact is on a different protein (so-called “target\_protein”), then add also info in the Note section of the impact type with name of the target\_protein

*“These results support our previous conclusions that the L435F and C410Y mutations cause loss of Presenilin function and gamma-secretase activity, including impaired Abeta production in the brain”*

These are mutations in Presenilin 1, then the impact “impaired Abeta production” should have Note Abeta (or APP).

impaired Abeta production

---

Location  
986

Entity Type  
impact\_protein\_abundance

+ Add New Entity Type

Concept ID

Note  
Abeta

Annotate each instance of this mention text

Update all mentions with the same concept ID (-- NO Concept ID assigned --)

Last updated by annotator2 at December 15, 2021 4:10 PM

✕ Delete + Add to Relation Cancel ✓ Update

4-For disease, these must be within the scope of AD/ADRDs. If only one disease is mentioned in abstract and it is tagged by Pubtator, then no need to do anything. If one disease is mentioned but not tagged (or the tag is incorrect), then annotate it as disease. If more than one disease, and only one involves the variant, then add a note on variant of impact sentence:

E.g., Variant X of gene Y affects structure and is associated with AD, while variant Z of gene Y affects function in Lewis body dementia.

Annotate variant X with Note disease: AD and select impact on structure

Annotate variant Z with Note disease: Lewis body dementia and select impact on function/activity

For example, there are 3 different diseases/disorders mentioned in the abstract below. In these cases all of them are under umbrella synucleinopathies. Since the abstract is pointing to the general disorder, when annotating the relation indicate in Note: disease:synucleinopathies

The effects of the novel **A53E alpha-synuclein** mutation on its **oligomerization** and **aggregation**

Lazaro, Diana F.,Dias, Mariana Castro.,Carija, Anita.,Navarro, Susanna.,Madaleno, Carolina Silva.,Tenreiro, Sandra.,Ventura, Salvador.,Outeiro, Tiago F.,Lazaro, Diana F. 2016. Vol. 4. -  
 Keyword: Alpha-synuclein Parkinson's disease Oligomerization Aggregation Neurodegeneration.  
[PMC 5148884](#)  
[PMID 27938414](#)

abstract 2 offset: 94 - 1118  
 alpha-synuclein (aSyn) is associated with both sporadic and familial forms of **Parkinson's disease (PD)**, the second most common **neurodegenerative disorder** after **Alzheimer's disease**. In particular, multiplications and point mutations in the gene encoding for aSyn cause familial forms of **PD**. Moreover, the accumulation of aSyn in Lewy Bodies and **Lewy neurites in disorders** such as **PD, dementia** with Lewy bodies, or multiple system **atrophy**, suggests aSyn misfolding and aggregation plays an important role in these disorders, collectively known as **synucleinopathies**. The exact function of aSyn remains unclear, but it is known to be associated with vesicles and membranes, and to have an impact on important cellular functions such as intracellular trafficking and protein degradation systems, leading to cellular pathologies that can be readily studied in cell-based models. Thus, understanding the molecular effects of aSyn point mutations may provide important insight into the molecular mechanisms underlying disease onset.

abstract 2 offset: 1120 - 1367  
 We investigated the effect of the recently identified **A53E aSyn** mutation. Combining in vitro studies with studies in cell models, we found that **this mutation reduces aSyn aggregation** and **increases proteasome activity, altering normal proteostasis**.

abstract 2 offset: 1368 - 1610  
 We observed that, in our experimental paradigms, the **A53E** mutation **affects specific steps of the aggregation process of aSyn** and **different cellular processes**, providing novel ideas about the molecular mechanisms involved in **synucleinopathies**.

### Relation #R11

Note

Relation Type

+ Add New Relation Type

Nodes You can change the order by dragging

Type	Concept	Text	Offset	Role
A	ProteinMutation	this mutation	1264	<input type="text"/>
A	impact_protein_aggregation	reduces aSyn aggre...	1278	<input type="text"/>
A	impact_protein_function_activity	increases proteaso...	1307	<input type="text"/>
A	impact_protein_process	altering normal prot...	1338	<input type="text"/>

+ Add reference node

Last updated by annotator 1 at December 16, 2021 9:19 AM

5-Select all relevant impact sentences at the end, and tag it as entity: sentence. Make sure you indicate as new annotation and do not erase any existing annotation (unless is intentional)

We investigated the effect of the recently identified **A53E aSyn** mutation. Combining in vitro studies with studies in cell models, we found that **this mutation reduces aSyn aggregation** and **increases proteasome activity, altering normal proteostasis**

In gray is the sentence, where other annotated entities are highlighted in their respective colors.

## 6-Relation annotation

1). We are annotating has\_impact relation between a mutation and an impact type

When more than one impact in the same sentence for a mutation, create separate relations for each of them.

The screenshot displays a text annotation interface. On the left, a snippet of a scientific abstract is shown with various terms highlighted in different colors (blue, red, yellow, green). On the right, a 'Relations' panel is visible, listing 11 relations (R3 to R11) between mutations and their impacts. Each relation is represented as a table row with columns for ID, Type, and Nodes.

Annotations: Xia, Dan., Kelleher, Raymond J., Shen, Jie. 2016. Vol. 90. Issue 2. 417 - 422. PMC 4840410 PMID 27100200

abstract\_title\_1 2 offset: 87 - 94

SUMMARY

abstract 2 offset: 95 - 1114

We recently reported that homozygous Presenilin-1 (Psen1) knockin (KI) mice carrying the familial Alzheimer's disease (FAD) mutation L435F or C410Y recapitulate the phenotypes of Psen1-/- mice. Production and steady-state levels of Abeta40 and Abeta42 are undetectable in KI/KI brains and reduced in KI/+ brains, though the Abeta42/Abeta40 ratio is slightly increased in KI/+ brains. Moreover, the FAD mutation impairs synaptic function, learning and memory, and age-dependent neuronal survival in the adult brain. Here we extend our analysis of the effects of the L435F and C410Y mutations to the generation of Abeta43. Similar to Abeta40 and Abeta42, production of Abeta43 is undetectable in KI/KI brains and reduced in KI/+ brains. These results support our previous conclusions that the L435F and C410Y mutations cause loss of Presenilin function and gamma-secretase activity, including impaired Abeta production in the brain. This Matters Arising Response paper addresses the Veugelen et al. Matters Arising paper.

title\_1 2 offset: 1115 - 1127

ID	Type	Nodes
R3	has_impact	L435F generation of Abeta43
R5	has_impact	L435F loss of Presenilin function
R6	has_impact	L435F gamma-secretase activity
R7	has_impact	L435F impaired Abeta production
R4	has_impact	C410Y generation of Abeta43
R8	has_impact	C410Y loss of Presenilin function
R9	has_impact	C410Y gamma-secretase activity
R10	has_impact	C410Y impaired Abeta production
R13	has_impact	the FAD mutation impairs synaptic function
R15	has_impact	the FAD mutation learning and memory
R16	has_impact	the FAD mutation age-dependent neuronal survival in the adult brain
R11	has_impact	Loss of Abeta43 production Presenilin-1 mutations

2) Investigational sentences (those stating what authors intend to do, not the result of it): only need to annotate such sentences with entity type "sentence", no need to annotate the relation(s) in them; Add note "investigation"

3) Negation sentences: We will still add the relation between the mutation and impact but need to add NOTE "negation";

4) Unique relation: if multiple sentences provide the same relation annotations (in terms of uniqueness in the relation), then highlight all the impact sentences, annotate the unique relation out of only one sentence of them.