

Exercises for Systematics

HGNC and VGNC

Find alias names and symbols.

NLM1 is a gene symbol whoes gene model is [At4g19030.1](#). It is annotated in the Arabidopsis thaliana (Mouse-ear cress) and encoding [A0A1P8B5K6 · A0A1P8B5K6_ARATH](#) (NOD26-like major intrinsic protein 1).

Therefore, I would analyze *ADA* in the following assignments.

alias names and symbols for gene *ADA*:

- [ADA12](#)

Is there a mouse homologue?

Yes: [Ada](#)

MANE

Find MANE transcript for your gene.

Search term on NIH nucleotide: `(ADA[Gene Name] AND srcdb_refseq[PROP] AND MANE Select[Keyword] AND Homo sapiens[Organism])` and on ensembl. Or just copy from HGNC entry:

- NM_000022.4
- ENST00000372874.9

RefSeq

How many transcripts are there and how do they differ?

There are four transcript variants. They are different in their sequence span on a genome assembly.

UniProt

How many variants are there in SwissProt and of which types?

Download variants report json file from UniProt and count id and consequenceType:

Total:	595
missense	526
frameshift	33
stop gained	28
-	5

Total:	595
stop lost	2
inframe deletion	1

for the id entry, there seems to be different system. Could this be redundant? i.e., COSMO id and rs id actually are the same allele. Therefore, which entry should be taken to count the variants? what is the "-" in the consequenceType entry?

EC

What is the EC number for the protein and what does it mean?

[EC 3.5.4.4: adenosine deaminase](#)

HGVS

How are frameshift variations systematically named?

Frameshift