Standard Operating Procedure of Syntax Determination

Determination of CDS syntax and amino acid syntax

1. Extract the gene, transcript version, mutation syntax, amino acid syntax, genomic position from literatures.
2. If all the information can be found in the literatures, validate the information with COSMIC and record the verified information in iCMDB.
3. If the genomic position is missing, identify the genomic position from COSMIC/NCBI based on the known information such as gene, transcript version, CDS syntax and amino acid syntax.
4. If CDS syntax or amino acid syntax is missing, check databases that report clinically relevant variations including COSMIC/Clinivar/NCCN/NCI to get the missing CDS syntax or amino acid syntax.
5. If the missing CDS syntax or amino acid syntax cannot be extracted in (4), convert the syntax based on the transcript sequences of CDS and protein, which are provided in Ensembl database.
6. Cross-validate the syntax with other clinically relevant variation databases including COSMIC/NCBI/NCCN/NCI
7. Record the CDS syntax and amino acid syntax in iCMDB

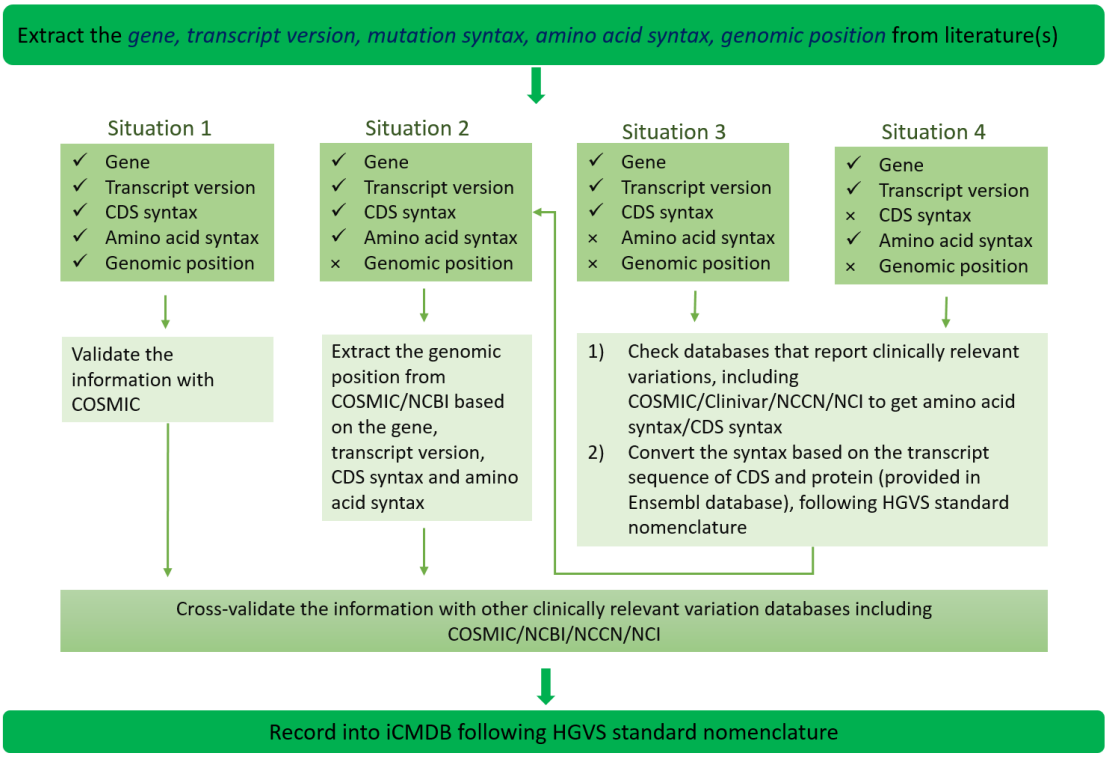

Figure. Flow chart of CDS/amino acid syntax determination

Determination of Exon

Record the transcript version and exon reported in literatures. If exon was not reported, use Ensembl biomart tool (http://asia.ensembl.org/biomart/martview/debbc2db068d0b880fdc8750de4e42c0) to identify the exon of mutation based on the transcript version used in literatures.