

Purpose

The American Society of Hematology Somatic Working Group has integrated clinical classifications of genomic variants from multiple institutions to create a consensus somatic pathogenicity dataset. [Several institutions](#) contributed to this effort. This document is intended to assist downstream users in interpreting this dataset, by defining the classification system and the process for constructing these classifications for in our initial (v1) release. This document is versioned to coincide with the semantics and processes of the corresponding data releases.

Definitions

For this dataset, we define *Somatic Pathogenicity* of a variant as its role in driving tumorigenesis.

We use a three-class system for defining somatic pathogenicity:

- *Pathogenic variants* have been determined to drive tumorigenesis
- *Variants of unknown significance* (VUS) may have evidence supporting or refuting their effect on somatic pathogenicity, but the evidence is not strong enough to classify as pathogenic or benign
- *Benign variants* have been determined not to drive tumorigenesis

Input Classifications

As a consensus-building exercise, this procedure used classifications from 6 institutions. Below are details about the classification systems used as inputs to the procedure described in this document.

Institution 1

Resources utilized for variant interpretation are ClinVar, gnomAD, COSMIC, SIFT, Polyphen, Mutation assessor. Results submitted to the clinician are interpreted as disease related, VUS, or benign.

Institution 2

Annotation of variants was performed using Ion Reporter Software. Effects of mutations on the protein function were determined using prediction software (e.g. Sift, Polyphen 2 and Grantham). Association of variants with previously documented cancer specimens was determined using COSMIC (cancer.sanger.ac.uk/) and ClinVar (www.ncbi.nlm.nih.gov/clinvar/) databases. Results submitted to the clinician are interpreted as pathogenic, possible pathogenic, or unknown.

Institution 3

Resources utilized for variant interpretation are CIVIC, DoCM, COSMIC, ClinVar, gnomAD, IARC, FASMIC, oncoKB and in-silico prediction tools utilized are SIFT, LRT, Mutation Taster, Mutation Assessor, FATHMM, PROVEAN, Meta SVM, Meta LR, FATHMM MKL Coding. Results submitted to the clinician are according to the ACMG guidelines and provided as benign, likely benign, variant of unknown significance, likely pathogenic or pathogenic.

Institution 4

Resources utilized for variant interpretation are COSMIC, ClinVar, dbSNP, gnomAD (population freq.), dbNSFP (in-silico predictions) and [MLL Predictor](#) (Hutter et al, 2019, Blood Supplement_1). Results submitted to the clinician are interpreted as pathogenic, VUS, or benign.

Institution 5

Resources utilized for variant interpretation are gnomAD/ExAC, dbSNP, Variant Effect Predictor, COSMIC, oncoKB/cBioPortal, ClinVar and by an internally developed algorithm based upon literature review and experience. Results submitted to the clinician are interpreted as pathogenic (including likely pathogenic) and variant of unknown significance. Benign variants are not reported.

Institution 6

Resources [PeCanPIE](#) (Edmonson et al., 2019, Genome Res) utilized for variant interpretation are COSMIC, ClinVar, gnomAD/ExAC, IARC, ARUP, ASU:TERT,BIC, oncoKB, dbNSFP, St. Jude PCGP, LOVD. In-silico

prediction tools utilized are MutationAssessor and PolyPhen2, as well as SIFT, PolyPhen2, LRT, MutationAssessor, MutationTaster, FATHMM, CADD, and REVEL. Results submitted to the clinician are interpreted as somatic pathogenic, likely pathogenic, or VUS.

Integration Rules for Somatic Pathogenicity Classification

The somatic pathogenicity of a variant was determined by consensus agreement of the above ~~classification~~ interpretation systems. To do this, the following rules were applied:

1. Classifications were transformed to the [above-defined classification system](#) on a per-institution basis, in consultation with appropriate institutional representatives.
2. In cases where there were 3 or more classifications for a variant in agreement and no conflicting classification, we report the consensus classification.
3. In cases where there were fewer than 3 classifications for a variant and no conflicting classification, they were not included that those will be assessed for future releases
4. In cases where there were conflicting classifications, those were not included and we will have future releases where the methods by which conflicting interpretations are resolved will be documented.

Discrepant Interpretations

Contributors will be working through discrepant interpretations, and these will be submitted to the ASH website in the future. The methodologies (including bioinformatics pipelines) by which discrepant interpretations are resolved will be described in full detail.

Contributors

Rafael Bejar, MD, PhD: University of California, San Diego, CA; Ewalt: University of Colorado, Aurora, CO; Haferlach: MLL Munich Leukemia Laboratory, Munich, Germany; Kim: Brigham and Women's Hospital, Boston, Ma; Le Beau: University of Chicago, Chicago, IL; Ma: St. Jude Children's Research Hospital, Memphis, TN; Meggendorfer: MLL Munich Leukemia Laboratory, Munich, Germany; Nadarajah: MLL Munich Leukemia Laboratory, Munich, Germany; Shammo: Rush University, Chicago, IL; Ryan: American Society of hematology, Washington, DC; Siddon: Yale School of Medicine, New Haven, CT; Steensma: Dana-Farber Cancer Institute, Boston, MA; Wagner: Nationwide Children's Hospital, Columbus, OH; Walter: Washington University School of Medicine, St. Louis, MO; Zehir: Memorial Sloan Kettering Cancer Center, New York, NY; Zhang, St. Jude Children's Research Hospital, Memphis, TN.

± = gene list contributor; §= variant data contributor