**SUPPLEMENTARY TEXT: Towards controlled terminology for reporting germline cancer susceptibility variants: an ENIGMA report.** Spurdle et al, J Medical Genetics.

**ENIGMA membership**

Eligibility for membership is broad: an ENIGMA member is currently defined as a researcher or research group (consortium) who is willing to work collaboratively towards classification of variants by contributing data from families and/or conducting statistical analysis or laboratory-based assays within a working group framework, namely clinical, analytical, functional, splicing and pathology working groups. The ENIGMA membership listing currently includes 309 individuals from 38 different countries. Most of these members (202) are from countries where research and clinical activities are conducted in a language other than English (see **Supp Figure 1**).



**Supplementary Figure 1:** Heatmap showing distribution of ENIGMA members across countries.

**Published recommendations on variant terminology**

**ACMG/AMP** [1]: “The terms “mutation” and “polymorphism,” however, which have been used widely, often lead to confusion because of incorrect assumptions of pathogenic and benign effects, respectively. Thus, it is recommended that both terms be replaced by the term “variant” with the following modifiers: (i) pathogenic, (ii) likely pathogenic, (iii) uncertain significance, (iv) likely benign, or (v) benign.”

**HVP** [2]: “The term “variant” should be used to describe all sequences changes irrespective of their contribution to phenotype. Mutation may be used in the correct sense of the word to describe the process by which variants arise. Use of the term polymorphism is deprecated.”

It is also the term used in publication to describe somatic alterations by the Variant Interpretation for Cancer Consortium (**VICC**) [3], although the terminology seems not to be transferred to the CIViC web-based resource for expert crowdsourcing of Clinical Interpretation of Variants in Cancer (<http://civicdb.org/>), a key output from this group.

**Clinical calibration, replication and validation**

Clinical calibration refers to the process where a specific type of information (e.g. bioinformatic score, protein function, splicing aberration, pathology information) is calibrated as a measure of variant pathogenicity against clinical predictors of variant pathogenicity (e.g. segregation data, family history profile, frequency in population controls). For example, Easton et al [4] provides an assessment of family history profiles of *BRCA1/2* pathogenic variant carriers versus non-carriers, according to variant location in specific motifs or domains, combined with bioinformatic prediction of missense effect. To avoid over-fitting, we recommend that initial calibration of a promising predictive measure (in particular functional assays) should be followed by a validation study using an independent set of variants, or at least an independent patient observational dataset. Further, as shown in **Supplementary Figure 2** below, and utilized in Drost *et al* [5], we suggest the following approach to describe replication and validation of a clinical calibration exercise. A functional assay method and its calibration can be considered to be replicated, *but not validated*, if at least one point estimated from the Training and Validation sets falls outside the other point estimate's 80% Confidence Interval (CI), but the point estimates are mutually within their 95% CI, which are wider. In this case, the method’s point estimates and confidence intervals will be refined by recalculation from the combined data set; in this scenario it may be considered appropriate to withhold the method from quantitative clinical use until it has met the stricter criterion for replication and validation against a later data set. A method and its calibration are considered invalidated if either or both point estimates fall outside the other point estimate’s 95% CI.

|  |  |
| --- | --- |
| **Supplementary Figure 2:** Depiction of confidence interval-based assay validation logic.Blue ball: point estimate.Blue line: 80% confidence limits.Yellow line: 95% confidence limits.1. Replicated **and** validated.
2. Replicated but **not** validated.
3. Invalidation.
 | abc |

**Definitions of key terms used to describe the clinical importance of sequence variation.**

Definitions, including health-related definitions where available, were sourced by online searches of the Oxford, Collins and Merriam-Webster dictionaries. Derivations were synthesized from all three sources. Considerations about translations of these terms into other languages was discussed by ENIGMA members whose primary language was not English.

***Pathogenic:***

Oxford: (of a bacterium, virus, or other microorganism) causing disease.

Collins: able to cause or produce disease (British), producing disease (American).

Merriam-Webster: Causing or capable of causing disease.

Derivation: From the Greek *pathos* (suffering) + *gen* (that which produces).

***Deleterious:***

Oxford: causing harm or damage

Collins: harmful, injurious, hurtful (English), harmful to health or well-being, injurious (American).

Merriam-Webster: damaging or harmful, or harmful often in a subtle or unexpected way (medical definition).

Derivation: via Medieval Latin from Greek *deleterios* (noxious/injurious/destructive).

***Benign***:

Oxford: (of a disease) not harmful in effect; (of a tumour) not malignant.

Collins: not threatening to life or health; not malignant (British), doing little or no harm, not malignant (American).

Merriam-Webster: of a mild type or character that does not threaten health or life (especially, not becoming cancerous); having no significant effect.

Derivation: Middle English - from Old French *benigne*, from Latin *benignus*, (probably) from *bene* (well) + -*genus* (-born). Alternative derivation (Collins) - from Old French *benigne*, from Latin *benignus*, from *bene* (well) + *gignere* (to produce).

***Mutation***:

Oxford: the process or an instance of change or alteration; a genetic change which, when transmitted to offspring, gives rise to heritable variations (Australian), action or process of mutating, changing of the structure of a gene – resulting in a variant form that may be transmitted to subsequent generations; a distinct form resulting from genetic mutation (English).

Collins: the act or process of mutating, change, alteration; a change or alteration; a change in the chromosomes or genes of a cell, which when in gametes the structure and development of the resultant offspring may be affected (English). a changing or being changed; a change as in form, nature, qualities; a sudden variation in some inheritable characteristic in a germ cell of an individual animal or plant, as distinguished from a variation resulting from generations of gradual change; an individual resulting from such variation; an abrupt and relatively permanent change in somatic cells that is transmitted only to daughter cell and can be inherited only in plants that reproduce asexually (American).

Merriam-Webster: a significant and basic alteration; a relatively permanent change in hereditary material that involves either a change in chromosome structure or number or a change in the nucleotide sequence of a gene's codons (as in frameshift or missense errors) and that occurs either in germ cells or in somatic cells but with only those in germ cells being capable of perpetuation by sexual reproduction; an individual, strain, or trait resulting from mutation

Derivation: Middle English from Latin *mutatio*, from *mutare* (change).

***Significance:***

Oxford: the quality of being worthy of attention, important.

Collins: consequence or importance.

Merriam-Webster: quality of being important, the quality of having notable worth or influence.

Derivation: Middle English via Old French *significance,* from Latin *significantia*, from *significare* (indicate, portend).

**Annotation based on variant sequence/position and (predicted) effect on gene/protein product**

Variant annotation, often used in the context of next generation sequencing bioinformatic pipelines, is the process of assigning various descriptors or data points relevant for a given sequence variant. The purpose is to assess variously: the quality of a sequence variant call; the location of a variant in relation to functional genomic regions (exons, introns, splice sites, regulatory sites); bioinformatically predicted effect on mRNA transcript/encoded protein, nucleotide protein or evolutionary conservation; and multiple other possible features used to interpret the clinical or research importance of genetic variation. As an example, Variant Effect Predictor (VEP, [6]) is applied commonly in sequencing pipelines, and uses sequence ontology (SO) terms [7] to describe a genomic variant by type of sequence alteration, genomic features altered by the change, and the predicted impact of the alteration. Pipelines may also include cross-reference to identify the variant (or variant position) in internal or external datasets that provide information used for curation of disease gene variants against recognized variant classification criteria. For example: variant frequency in outbred population groups such as gnomAD (http://gnomad.broadinstitute.org/)[8]; presence and pathogenicity assertion in ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) [9], presence in the COSMIC repository of variation identified in tumors (Catalogue of Somatic Mutations in Cancer; https://cancer.sanger.ac.uk/cosmic)[10].

**References**

1. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;**17**(5):405-24 doi: 10.1038/gim.2015.30[published Online First: Epub Date]|.

2. Project) HHV. HVP Guideline HVP/GL/003-01/EN, 2016.

3. Griffith M, Spies NC, Krysiak K, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet 2017;**49**(2):170-74 doi: 10.1038/ng.3774[published Online First: Epub Date]|.

4. Easton DF, Deffenbaugh AM, Pruss D, et al. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. Am J Hum Genet 2007;**81**(5):873-83 doi: 10.1086/521032[published Online First: Epub Date]|.

5. Drost M, Tiersma Y, Thompson BA, et al. A functional assay-based procedure to classify mismatch repair gene variants in Lynch Syndrome. Genetics in Medicine 2018;**in press**

6. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biol 2016;**17**(1):122 doi: 10.1186/s13059-016-0974-4[published Online First: Epub Date]|.

7. Cunningham F, Moore B, Ruiz-Schultz N, et al. Improving the Sequence Ontology terminology for genomic variant annotation. J Biomed Semantics 2015;**6**:32 doi: 10.1186/s13326-015-0030-4[published Online First: Epub Date]|.

8. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;**536**(7616):285-91 doi: 10.1038/nature19057[published Online First: Epub Date]|.

9. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 2018;**46**(D1):D1062-D67 doi: 10.1093/nar/gkx1153[published Online First: Epub Date]|.

10. Forbes SA, Beare D, Boutselakis H, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res 2017;**45**(D1):D777-D83 doi: 10.1093/nar/gkw1121[published Online First: Epub Date]|.