

Standardisation of pathogenicity classification for somatic alterations in solid tumours and haematologic malignancies

Florence Koeppel, Etienne Muller, Alexandre Harlé, Celine Guien, Pierre Sujobert, Olfa Trabelsi Grati, Olivier Kosmider, Laurent Miguet, Laurent Mauvieux, Anne Cayre, et al.

▶ **To cite this version:**

Florence Koeppel, Etienne Muller, Alexandre Harlé, Celine Guien, Pierre Sujobert, et al.. Standardisation of pathogenicity classification for somatic alterations in solid tumours and haematologic malignancies. European Journal of Cancer, 2021, 159, pp.1-15. 10.1016/j.ejca.2021.08.047 . hal-03408473

HAL Id: hal-03408473
https://hal.science/hal-03408473

Submitted on 11 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Standardisation of pathogenicity classification for somatic alterations in solid tumours and haematologic malignancies

Florence Koeppel [a], Etienne Muller [b,c], Alexandre Harlé [d], Céline Guien [e], Pierre Sujobert [f,g], Olfa Trabelsi Grati [h], Olivier Kosmider [i], Laurent Miguet [j], Laurent Mauvieux [j], Anne Cayre [k], David Salgado [e], Claude Preudhomme [l], Lucie Karayan-Tapon [m], Gaëlle Tachon [m], Florence Coulet [n], Alexandra Lespagnol [o], Christophe Beroud [e,p], Karen Leroy [q,l], Etienne Rouleau [r,*,l], Isabelle Soubeyran [s,l]

[a] *Gustave Roussy, Direction de la Recherche, Villejuif, F-94805, France*
[b] *Laboratoire de Biologie et Génétique du Cancer, Centre François Baclesse, Caen, 14000, France*
[c] *Inserm U1245, Normandie Univ, UNIROUEN, Normandy Centre for Genomic and Personalized Medicine, Rouen, 76031, France*
[d] *Université de Lorraine CNRS UMR 7039 CRAN, Service de Biopathologie, Institut de Cancérologie de Lorraine, Vandœuvre-lès-Nancy, F-54519, France*
[e] *Aix Marseille Univ, INSERM, MMG, Bioinformatics & Genetics, Marseille, France*
[f] *Hospices Civils de Lyon, Groupement Hospitalier Sud, Service d'hématologie biologique, Pierre-Bénite, France*
[g] *Cancer Research Center of Lyon, INSERM U1052 UMR CNRS 5286, Equipe labellisée Ligue Contre le Cancer, Université de Lyon, Lyon, France*
[h] *Unité de pharmacogénomique, Service de Génétique, Institut Curie, 26 rue d'Ulm, Paris, 75005, France*
[i] *AP-HP Centre, Hôpital Cochin, Service d'hématologie Biologique et Université de Paris, Paris-Descartes, France*
[j] *Laboratoire d'hématologie, CHRU Strasbourg, INSERM U1113, Avenue Molière, Strasbourg, 67100, France*
[k] *LBM OncoGenAuvergne, UF de Pathologie, Centre Jean Perrin, 58 Rue Montalembert, BP392, Clermont-Ferrand, 63011, France*
[l] *Center of Pathology, Laboratory of Hematology, University Hospital of Lille, Lille, France*
[m] *Université de Poitiers, INSERMU1084 et CHU de Poitiers, Laboratoire de Cancérologie Biologique, Poitiers, France*
[n] *Genetics Department, Assistance publique - Hôpitaux de Paris, Pitié Salpêtrière Hôpital, Paris, France*
[o] *CHU Pontchaillou - Laboratoire de Génétique Somatique des Cancers, Rennes, France*
[p] *AP-HM, Hôpital d'Enfants de la Timone, Département de Génétique Médicale et de Biologie Cellulaire, Marseille, France*
[q] *AP-HP Centre, Hôpital Européen Georges Pompidou, Service de Biochimie et Université de Paris, France*
[r] *Gustave Roussy, Département de biologie et pathologie médicales, Villejuif, F-94805, France*
[s] *Unité de Pathologie Moléculaire et Inserm U1218, Institut Bergonié, 229 cours de l'Argonne, Bordeaux, 33076, France*

Abstract ***Background:*** The difficulty in interpreting somatic alterations is correlated with the increase in sequencing panel size. To correctly guide the clinical management of patients with cancer, there needs to be accurate classification of pathogenicity followed by actionability assessment. Here, we describe a specific detailed workflow for the classification of the pathogenicity of somatic variants in cancer into five categories: benign, likely benign, unknown significance, likely pathogenic and pathogenic.
Methods: Classification is obtained by combining a set of eight relevant criteria in favour of either a pathogenic or a benign effect (pathogenic stand-alone, pathogenic very strong, pathogenic strong, pathogenic moderate, pathogenic supporting, benign supporting, benign strong and benign stand-alone).
Results: Our guide is concordant with the ACMG/AMP 2015 guidelines for germline variants. Interpretation of somatic variants requires considering specific criteria, such as the disease and therapeutic context, co-occurring genomic events in the tumour when available and the use of cancer-specific variant databases. In addition, the gene role in tumorigenesis (oncogene or tumour suppressor gene) also needs to be taken into consideration.
Conclusion: Our classification could contribute to homogenize best practices on somatic variant pathogenicity interpretation and improve interpretation consistency both within and between laboratories.

1. Introduction

Tumour sequencing is used in standard care and clinical trials to detect cancer genome alterations, and the accurate interpretation of any genomic alterations is key for the clinical management of patients. Some alterations have been well documented, and their interpretation is straightforward, but others are more challenging, either because they have been rarely observed in the considered cancer type or because they are in under-investigated areas of the genome. With the reduced cost and increased accessibility to sequencing technologies, the size of sequencing panels has increased over recent years. This resulted in finding an increased number of unknown variants and variants in under-investigated regions in the genome.

The underlying question is whether a patient with an unknown variant in an actionable gene should or should not be included into a precision oncology trial. Well-established pathogenic and benign variants are usually recognized easily, but there is a critical need to broaden the classification of a variant of interest from the 'unknown significance' category, to either the 'potentially benign' or the 'potentially pathogenic' category to support the inclusion of a patient into a trial. Unlike the detailed recommendations available for germline alterations in the context of rare genetic diseases [6,7], the current recommendations for somatic alterations are more granular [8–13]. They are based on a limited number of factors, are not detailed and may miss some elements required for adequate interpretation. To address this challenge, we have focused on optimising the pathogenicity assessment of uncommon single-nucleotide variations and small indels. It is important to note that a separate dedicated guide will be required for the interpretation of copy number alterations and fusion transcripts, some of which play a key role in haematologic malignancies in particular.

Variant interpretation corresponds to a pathogenicity classification (the biological impact of the alteration in the oncogenesis process) and an actionability assessment (the clinical impact in the standard of care of the patient [1]) and includes the theranostic application as well as diagnostic (for variant-driven sub-stratification of patients) and prognostic aspects. Interpretation of pathogenicity needs to be separate and evaluated before the interpretation of actionability to avoid potential confusion. When a variant is classified as pathogenic or probably pathogenic, the actionability can then be discussed in the clinical context, for example cancer histology and disease status (initial diagnosis, metastasis diagnosis, relapse, etc.) [2], and the correct level of actionability can be determined within this context. A pathogenic variant may be actionable in a certain cancer type but not actionable in another cancer type. Conversely, a polymorphism or a variant, which has been assessed to be neutral, cannot be actionable, even if some of these may modulate the therapeutic effect (e.g. *TP53* or *RASSF1A*). There are multiple existing actionability scales, such as the one provided by ESMO [3,4]. Synonymous variants should be classified with care because they are not necessarily benign; they may affect regulatory sequences or splicing and drive oncogenesis [5].

We have developed a detailed guide for the pathogenicity classification of somatic variants through a consensus

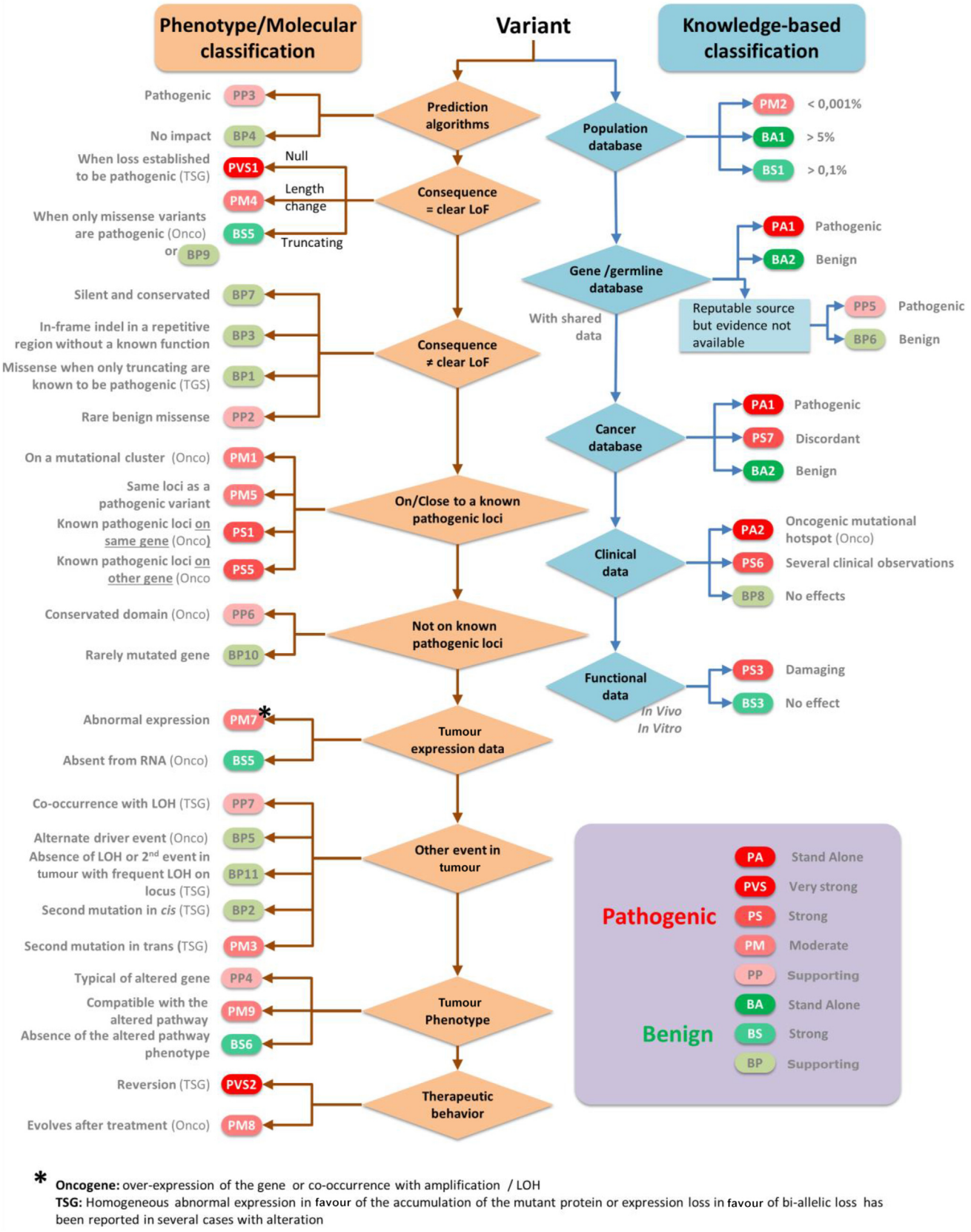


Fig. 1. Criteria determination workflow.

approach, involving thirteen centres across France. This network of laboratories is building a new database SoVaD, which shares alterations identified in French health care institutions and benefits from an expert curation (https://somaticdb.geneticsandbioinformatics.eu/). Relevant criteria were listed and assigned a score to establish the classification system. In addition, we illustrated our approach through examples of 'unknown' variants found in various solid and haematological malignancies (Supplementary Table 1). Some of these examples were taken from the Gen&Tiss external quality assessment program (2018 campaign) [14] and from the SoVad database. Supplementary Tables 2 and 3 highlight the correspondence with ACMG for pathogenic and benign criteria, respectively, and Fig. 1 illustrates the criterion determination workflow. Our aim is to provide a support for the decision-making, homogenise interpretation and limit variability both within and between laboratories.

2. General principles of the classification system

Pathogenicity classification requires combining sets of criteria specific to somatic variants and germline variants (Table 1). In concordance with the ACMG germline recommendations, we used a criterion identification code composed of a category (P or B depending on whether the criterion is in favour of a pathogenic or a benign effect, respectively), combined to a score (A/VS/S/M/P for stand-alone/very strong/strong/moderate/supporting, respectively, Table 2) and an incremental number. To avoid confusion, the incremental numbers of the ACMG germline criteria were used whenever suitable. For somatic-specific criteria, additional incremental numbers were created.

Table 1
Comparison of criteria used for classification of germline versus somatic alterations.

Criteria common to germline variants	• Control population databases • Structural data • Scientific literature (functional data, clinical trials, case studies) • Clinical or genealogical criteria • The existence of associated alterations • Functional data
Criteria specific to somatic variants	• Cancer genomic data sets (e.g. TGCA, GENIE) • Information related to therapy (clinical outcomes under specific treatment) • Therapeutic information related to the conservation of similar protein domains (variant clustering) • Therapeutic context (cancer type, disease state) • Co-occurrence of an event in the tumour: - Mutational profiles—mutational signatures/chromosomal rearrangements - Loss of heterozygosity - Other deleterious mutation on the same gene - Other deleterious mutation on the same pathway - Additional data available on the tumour, such as IHC/RNA

We propose to classify pathogenicity into five classes because of combining the collected evidence (Table 3) in concordance with the ACMG system. We added a pathogenicity stand-alone class (PA) for criteria, which are sufficient to classify as class 5. Identification of several pathogenic very strong (PVS), pathogenic strong (PS) or pathogenic moderate (PM) criteria allows classifying either as class 5 or 4. Exclusive benign stand-alone (BA) or benign strong (BS) criteria exclusively are sufficient to classify as class 1. Combinations of benign criteria lead to class 2. In other cases or if there are enough criteria to classify as both benign and pathogenic simultaneously, then the classifications will be 'of unknown pathogenicity' in light of the knowledge available at the moment in time (class 3). The final interpretation of the variant should not remain the decision of a sole expert geneticist/molecular biologist, but needs to involve at some point a collegial discussion. Both local and shared databases also play a key role. The variant interpretation reflects the current knowledge and may evolve with relevant clinical, biological evidence collection. The unknown variant remains in class 3, until the point when there is enough available evidence (in databases or others) to classify it as likely pathogenic or likely benign. The relevance of the final interpretation is expected to be improved by following guidelines, which rationalise the classification process and help evidence gathering and decision-making in a more reliable manner.

3. Criteria and resources used for the classification

These criteria for pathogenicity classification of somatic variants are based on available databases, bioinformatics tools and published data. The pathogenicity and actionability classification of an unknown variant can be determined by examining the context of the tumour in which the alteration is encountered and the behaviour of the alteration during treatment (when information is available).

3.1. Criteria based on available databases

The first step in the interpretation workflow is to consult available databases. Epidemiological data are key for the interpretation of somatic variants. An expert geneticist or molecular biologist can use local laboratory variant databases in combination with single-nucleotide polymorphism (SNP) databases or disease variant databases, but the source of the information in any database

Table 2
Classes of gene activity, levels of evidence and pathogenicity.

Gene activity	• Oncogene • Tumour suppressor gene
Variant evidence weight	• **PA/PVS/PS/PM/PP**: stand-alone/very strong/strong/moderate/supporting elements for pathogenicity. • **BA/BS/BP**: stand-alone/strong/supporting elements for a benign impact.
Variant pathogenicity class	• **Class 1**: Benign • **Class 2**: Likely benign • **Class 3**: Unknown significance • **Class 4**: Likely pathogenic • **Class 5**: Pathogenic

needs to be validated by experts in the field (national/international collaborative database, curation methodology, publication in a scientific peer-reviewed journal). The database should also grant access to all the data required for classification. Reputable databases are curated based on a group of nationally or internationally approved experts, as defined by the Clinical Genome Resource ClinGen Program (https://clinicalgenome.org/).

3.1.1. Criteria based on SNP databases

Minor allele frequencies (MAFs) in the general population, sourced from SNP databases, are useful to classify variants. An MAF over 0.1% in gnomAD, Exome Sequencing Project or 1000 Genomes Project is an indication of the absence of pathogenicity, at least for cancer genes (BS1, Table 4). It can even be considered as a standalone criterion if the MAF is over 5% in gnomAD, Exome Sequencing Project or 1000 Genomes Project (BA1). Absence of or very low frequency from the SNP databases (MAF <0.001%) is compatible with pathogenicity (PM2). These data should be considered with caution because part of the variants featuring in these databases could actually be sequencing artefacts (e.g. *ASXL1* dupG, with MAF >3%). Moreover, the size of the database (number of individuals and number of alleles) limits the lower detectable MAF. Thresholds are given as an indication and may vary depending on the database size. Frequency in SNP databases should also be used cautiously for insertions/deletions (indels) and delins (particularly large ones) because these are poorly detected in the general population by high-throughput sequencing.

The allele frequency also needs to be considered in each of the subpopulations in addition to the global population. These databases are international and therefore not representative of specific subpopulations (regional or otherwise). Several pathogenic variations are overrepresented in particular populations because of the founder effect, such as in Ashkenazi Jewish and Finnish populations. Only allelic frequency data from cohorts of at least 100 individuals should be considered.

The absence of the alteration in a database should ideally be associated with verification of the sequencing depth at the location of the altered base in the shared data and evaluation of the quality of sequencing, but this information is often not available. The latter can be approximated by the number of genotypes available at the genomic position considered or at the adjacent positions (variable denominator for calculating allelic frequencies). Note that there is data redundancy between some of the databases.

3.1.2. Criteria based on Gene-specific germline variant databases and disease variant databases

Germline sources gather interpretative data on genes or germline variants known to be associated with a specific disease (Supplementary Table 4). The information is provided by scientific literature or laboratories, some of which is reviewed by expert committees. The source may use shared data to classify the alteration as pathogenic PA1 (Table 4) or benign using disease variant databases (BA2). The database may classify the variant without providing shared data, as pathogenic (PP5) or benign (BP6).

Co-segregation of the alteration with a strong phenotype can classify the alteration. Indeed, some phenotypes are highly specific for a gene, for example, mutations in the tumour suppressor gene *APC* with profuse polyposis in colon cancer.

3.1.3. Criteria based on cancer-specific variant databases

Cancer-specific variant databases indicate whether the alteration has previously been observed in different tumours and provide a classification of the variant (Table 5). The presence of several cases with the variant

Table 3
Pathogenicity classification based on the relative weight of the criteria.

Class 1—Benign	1 BA
	≥2 BS
Class 2—Likely benign	1 BS + 1 BP
	≥2 BP
Class 3—Unknown significance	Other criteria are unmet Criteria for benign and pathogenic are contradictory
Class 4—Likely pathogenic	1 PVS + 1 PM
	1 PS + 1–2 PM
	1 PS + ≥2 PP
	≥3 PM
	2 PM + ≥2 PP
	1 PM + ≥4 PP
Class 5—Pathogenic	1 PA
	1 PVS + ≥1 PS
	1 PVS + ≥2 PM/PP
	≥2 PS
	1 PS + ≥3 PM
	1 PS + 2 PM + ≥2 PP
	1 PS + 1 PM + ≥4 PP

Table 4
Criteria based on databases.

	Benign		Pathogenic			
	Strong = BS and stand-alone = BA	Supporting = BP	Supporting = PP	Moderate = PM	Strong = PS	Very strong = PVS and stand-alone = PA
SNP databases	• **BA1**: MAF >5% in gnomAD, Exome Sequencing Project or 1000 Genomes Project • **BS1**: MAF >0.1% (and below or equal to 5%) in gnomAD, Exome Sequencing Project or 1000 Genomes Project			• **PM2**: Absent from population database or MAF <0.001%		
Disease variant databases	• **BA2**: Classified as benign by a reputable germline source, with shared data	**BP6**: A reputable source recently reports the variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation	• **PP5**: A reputable source recently re ported the variant as pathogenic, but the evidence is not avail able to the labora tory to perform an independent evaluation			• **PA1**: Classi fied as pathogenic by a reputable germline source
Cancer-specific variant databases	• **BA2**: Classified as benign by a reputable cancer source			• **PM1** Onco: Located in a mutational clus ter with pathogenic alterations Identification of the same alteration in several tumours	• **PS7**: Classified as both pathogenic and likely pathogenic by various reputable cancer sources	• **PA1**: Classi fied as patho genic or actionable by several repu table cancer sources

SNP, single-nucleotide polymorphism; MAF, minor allele frequency.
Oncogene-specific criteria are in dark red.

Table 5
Examples of reputable databases.

Type	Example
SNP databases	• gnomAD (Genome Aggregation Database) • 1000 Genomes Project • ESP (NHLBI Exome Sequencing Project)
Gene-specific germline variant databases	• BRCAshare, BIC, BRCA exchange (*BRCA1, BRCA2*) • IARC (TP53) • UMD (Universal Mutation Database)
Disease-specific variant databases	• InSiGHT (Gastrointestinal hereditary tumours: *MLH1*, *MSH2*, *MSH6*, *PMS2*) • GGC (HBOC syndrome, Lynch syndrome, HDGC syndrome …)
Disease-non-specific variant databases	• HGMD (Human Gene Mutation Database) • ClinVar • LOVD (Leiden Open Variation Database) • OMIM (Online Mendelian Inheritance in Man)
Cancer-specific variant databases	• OncoKB [1] • DoCM (Database of Curated Mutations) [19] • CIViC (Clinical Interpretation of Variants in Cancer) [20] • CGI (Cancer Genome Interpreter) [21]

SNP, single-nucleotide polymorphism.

is in favour of pathogenicity if the variant is absent in any SNP database (PM1, Table 4). This holds true especially for oncogenes, which often carry hotspot alterations. A new type of database was recently implemented to explore the harmonised interpretations from several other bases [13]. These sources contain theranostic, prognostic or diagnostic information related to alterations, genes or signalling pathways. Some of these bases benefit from expert curation and classify the alteration based on evidence. An alteration may be classified consistently in several reputable databases interrogated as benign (BA2) or pathogenic (PA1). If classification is inconsistent, a conclusion can only be drawn if classification varies from likely pathogenic to pathogenic in the databases (PS7).

Cancer-specific databases should be used cautiously. Some data records, in particular the oldest ones, do not benefit from a control population or are based on a patient population of insufficient size and/or with insufficiently varied ethnicity to overcome potential biases. In addition, databases with case reports integrate alterations coming from articles, which are sometimes old, and their classification does not necessarily take into account all elements required for a standardised interpretation. Thus, some alterations in these databases are misclassified. These databases should be used as tools to retrieve publications reporting the alteration, and the biologist should interpret the alteration based on the elements available in the literature and taking into account the herein recommendations.

Colocalisation of alterations in a mutation cluster can be an element in favour of pathogenicity especially in oncogenes (PM1). Many clusters of mutations have been described in tumours, thanks to large sequencing programs, such as The Cancer Genome Atlas [15] and the International Cancer Genome Consortium [16]. This criterion may be used for recurrent hotspots and three-dimensional (3D) clustered hotspots, such as a mutated amino acid identified as a recurrent hotspot (statistically significant) and a 3D clustered hotspot in a population-scale cohort of tumour samples of various cancers [17,18].

3.1.4. Use of local databases of patients with cancer

Local databases of patient results at each health establishment are very helpful to refine interpretation. The goal is to homogenise interpretation for all patients with the same alteration. Alteration recurrence supports suspicion of an artefact or a polymorphism. It is strongly recommended to submit the classified alterations to international databases. The portal SoVaD also offers a user-friendly implementation of the classification workflow described herein to facilitate its adoption.

3.2. Criteria based on functional data and scientific literature

Classification is straightforward for mutational hotspots in oncogenes with approved targeted drugs (PA2, Table 6). For other alterations, it is necessary to review the literature with well-established *in vivo* or *in vitro* functional studies (BS3/PS3). Well-established *in vivo* or *in vitro* functional studies that show a deleterious impact of the alteration on the gene or its product strongly are an indicator of a pathogenic effect (PS3). Clinical cases that show a clinical response to a drug validated for the targeted gene are considered as PS6. Case reports that show no effect with a drug validated on the targeted genes are rather in favour of a benign effect (BP8). Yet, functional studies should be interpreted taking into account the cell or animal model, as well as the cancer type used to perform them, because there are tissue- and model-specific context biases.

3.3. Criteria based on structural and prediction data

3.3.1. Criteria based on structural information

3.3.1.1. In-frame missense alterations. In oncogenes, a nucleotide change resulting in the same amino acid change as a previously established pathogenic variant is strongly in favour of pathogenicity (PS1). Information

Table 6
Criteria related to functional data and scientific literature.

	Benign		Pathogenic			
	Strong = BS and stand-alone = BA	Supporting = BP	Supporting = PP	Moderate = PM	Strong = PS	Very strong = PVS and stand-alone = PA
Functional *in vitro*/ *in vivo* data	• **BS3**: Well-established *in vitro* or *in vivo* functional studies show no damaging effect on protein function or splicing				• **PS3**: Well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product	
Clinical data		• **BP8**: Case report without any effect of a targeted treatment			• **PS6**: Several clinical cases showing clinical response to a drug validated with the targeted gene	• **PA2** Onco: mutational hotspot with targeted and validated drugs.

Oncogene-specific criteria are in dark red.

related to structural domains can contribute to the classification of alterations in similar proteins (PS5), for example, the tyrosine kinase domain found in the epidermal growth factor receptor (EGFR) and ERBB2 receptors.

A missense alteration next to, or upon, a codon known from databases and/or functional studies to have a strong functional impact may be pathogenic (PM5, Table 7). In a tumour suppressor gene (TSG), in which only truncating alterations are known to be pathogenic, a missense might be benign (BP1). However, a predicted deleterious alteration is very likely to be pathogenic in a TSG in which loss of function (LOF) is a known pathogenic mechanism (PVS1). An in-frame indel or an alteration in a repeated region with no known function is likely to be benign (BP3), although there are some counter examples. Missense in conserved protein domains with functional impact reported in other proteins with similar domains might be pathogenic (PP6). Furthermore, if the missense alteration is within a gene in which missense alterations are frequently pathogenic, it is supporting pathogenicity (PP2).

When multiple nucleotide variants (MNVs) correspond to two in-frame alterations, spaced or distributed over a few nucleotides, the global impact of these alterations must be considered for interpretation.

Synonymous alteration without predicted impact on splicing and for which the nucleotide sequence is not highly conserved through vertebrates' species might be benign (BP7). It is important to consider the evolutionary depth of aligned species to ensure the most appropriate representation of known genomes. The degeneracy of the genetic code, involving the third base of the codon, induces a conservation which is often very weak at these positions. Thus, observed conservation at these positions could be considered more significant, suggesting the existence of a possible regulatory sequence at this level (e.g. exonic splicing enhancer (ESE) or silencer (ESS)). Conservation data are available from phyloP or PhastCons score databases and are used in the algorithms of certain predictors (e.g. SIFT and PolyPhen).

3.3.1.2. Alterations affecting the length of the protein. For oncogenes, an alteration which is truncating or induces a premature stop codon might be benign (BP9). This criterion is only poorly in favour of a benign effect as some deletion can be activating effect such as the activating PEST truncating mutations in *NOTCH1* in leukaemia. For the TSG in which LOF is the known pathogenic mechanism, a predicted deleterious effect (non-sense, frameshift, canonical splicing site: −2, −1, +1, +2, codon of translation initiation, out-of-frame deletion of one or more exons, in-frame deletion of one or more exons carrying a functional domain) is a very strong element in favour of pathogenicity (PVS1).

Table 7
Criteria related to structural and prediction data.

	Benign		Pathogenic			
	Strong = BS and stand-alone = BA	Supporting = BP	Supporting = PP	Moderate = PM	Strong = PS	Very strong = PVS and stand-alone = PA
Structural and prediction data/ presumed structural and functional impact		• **BP1** TSG: Missense variant in a gene for which only trun cating variants are known to be pathogenic • **BP3**: In-frame deletions/ insertions in a repetitive region without a known function • **BP4**: A missense variant with mul tiple lines of computational evidence suggests no impact on the gene or gene product. • **BP7**: A synony mous (silent) variant for which splicing prediction algo rithms predict no impact • **BP9** Onco: A truncating alter ation or an alter ation inducing a premature stop codon	• **PP2**: Missense alteration in a gene with a low rate of benign missense alterations • **PP3**: Multiple lines of computational evidence pre dict a delete rious effect on the gene or gene product • **PP6** Onco: Missense in conserved protein do mains with functional impact re ported in other proteins with similar domains	• **PM4**: Protein length changes as a result of in-frame deletions/ insertions in a non-repeat region or stop-loss variants • **PM5**: Novel missense change at an amino acid residue where a different missense change deter mined to be pathogenic has been seen before	• **PS1** Onco: Same amino acid change as a pre viously estab lished pathogenic variant regardless of nucleotide change • **PS5** Onco: Same amino acid change as an established path ogenic alteration in the same domain but from another gene	• **PVS1** TSG: Pre sumed deleterious alteration in a gene where LOF is a known path ogenic mechanism Deletion of a functional domain for which the loss is estab lished to be path ogenic (structural variation/splicing)

LOF, loss of function.
Criteria apply to tumour suppressor genes only (TSG, in blue), to oncogenes only (Onco, in dark red) or to both (in black)

Table 8
Classification criteria related to the context of the tumour.

	Benign		Pathogenic			
	Strong = BS and stand-alone = BA	Supporting = BP	Supporting = PP	Moderate = PM	Strong = PS	Very strong = PVS and stand-alone = PA
Context of the tumour (tumour type, CNA, IHC, RNAseq …)	• **BS5** Onco: Absence of the alteration in the RNA. Truncating alteration in the gene where only missense variants are pathogenic • **BS6**: Absence of the tumour phenotype related to the pathway alteration (e.g. absence of high TMB for *POLE*/*POLD1* or absence of MSI for MMR genes)	• **BP2** TSG: Found in a case with a deleterious mutation in *cis* on the same gene • **BP5** Onco: Found in a case with another oncogenic alteration in a gene on the same pathway (use with caution) • **BP5** Onco: Found with an alternate driver event, which is usually mutually exclusive (use with caution) • **BP10**: Genes rarely mutated on this type of tumours • **BP11** TSG: Absence of LOH in a tumour with frequent LOH on this locus	• **PP4**: Described in cases with a typical tumour phenotype, e.g. *APC* and colorectal cancer • **PP7** TSG: Co-occurrence with LOH	• **PM3** TSG: Alteration detected in *trans* • PM7 Onco: Over expression of the gene (ex *ERBB2*) or co-occurrence with amplification/ LOH • **PM7** TSG: Homogeneous abnormal expression in favour of the accumulation of the mutant protein, or expression loss in favour of bi-allelic loss has been reported in several cases with the alteration		

- **PM9**: Presence of a tumour phenotype compatible with the alteration of the pathway (e.g. TMB, MSI, HRD score)

CNA, copy number alteration; IHC, immunohistochemical staining; TMB, tumour mutational burden; TSG, tumour suppressor gene; MSI, microsatellite instability; LOH, loss of heterozygosity; HRD, homologous recombination deficiency.
Criteria apply to tumour suppressor genes only (TSG, in blue), to oncogenes only (Onco, in dark red) or to both (in black).

The criterion is of moderate impact if the truncation is located towards the end of the protein sequence and does not affect the functional domain of the protein (PM4). Most of the short transcripts are not translated because of non-sense–mediated decay [22]. Assigning an LOF effect to truncating alterations located after or within 50 nucleotides of the last intron–exon junction should be performed with caution because of the absence of non-sense–mediated mRNA decay and conservation of a large part of the encoded protein in such a case. Moreover, physiological splicing may occur, resulting in multiple transcripts.

For alterations affecting translation initiation codons, alternative translation initiation codons may be present. For stop codon loss–type alterations, the presence of potential alternative stop codons should be considered (ORFfinder analysis…).

3.3.2. Criteria based on bioinformatics predictions

For missense alterations, it is necessary to use several algorithms to predict the potential alteration of the protein function (Supplementary Table 4). Because of the lack of specificity of most of the bioinformatics predictions, we need to be cautious in situations where there is no predicted effect, to avoid reporting a false negative. This information must be interpreted with regard to the other elements available.

To obtain the best sensitivity and specificity [23], it is recommended to combine tools. Recently, scores have been developed to combine the predictions of a given number of other prediction tools (CADD [24], Condel [25], KGGseq [26] …). They have the advantage of allowing predictions on both non-exonic and exonic alterations. However, like all bioinformatics prediction tools, their performance is dependent on the sets of alterations used for training (machine learning).

Predicting the potential effect of an alteration on splicing is complex and requires additional supporting data. Splice site alterations of any nature might alter splicing, including non-sense, missense and synonymous alterations. The currently available tools include MaxEntScan [27] and NNSPLICE [28], GeneSplicer [29], SpliceSiteFinder, Human Splicing Finder [30] and combinatorial tool SPiCE [31]. Some tools like Mutation Taster or UMD-Predictor [32] incorporate both the prediction of an effect on the amino acid and an effect on splicing. Nevertheless, it is preferable to use tools dedicated to prediction of splicing effects. Current tools predicting splice alterations are efficient for creation of cryptic sites and the abolition of splice sites, but require improvement for element enhancer or silencer splicing. In case of a predicted deleterious splice effect, analysis of the tumour's transcripts or *ex vivo* splicing tests can be performed to objectify this potential effect and possibly provide a strong element in favour of pathogenicity.

Multiple lines of computational evidence are required to suggest a deleterious effect (PP3) on the gene or gene

Table 9
Classification criteria related to the therapeutic behaviour.

	Benign		Pathogenic			
	Strong = BS and stand-alone = BA	Supporting = BP	Supporting = PP	Moderate = PM	Strong = PS	Very strong = PVS and stand-alone = PA
Therapeutic behaviour				• PM8 Onco: Occurrence or increase after a targeted treatment • PM8 Onco: Primary resistance to a targeted treatment on several cases with the same alteration		• PVS2 TSG: A reversion on the variant after a targeted treatment has been observed

Criteria apply to tumour suppressor genes only (TSG, in blue), to oncogenes only (Onco, in dark red) or to both (in black).

product or no impact (BP4). It is important to consider that all bioinformatics prediction tools are purely speculative, and *in vivo* experiments are needed to confirm the impact of the variant. To ensure homogeneity of the interpretation of alterations, it would be interesting to agree on the prediction tools to query. In addition, when predictions are discordant, a conclusion on pathogenicity can hardly be drawn and the prediction criteria cannot be taken into account. Yet, the level of redundancy between the tools should be estimated to avoid a detrimental effect on the final result.

3.4. Criteria based on the tumour context

3.4.1. Tumour phenotype

Mutations in certain genes are associated with particular tumour types, for example, *APC* in colon cancer. Therefore, tumour type needs to be considered in the assessment of pathogenicity (PP4, Table 8). Similarly, if the alteration is in a gene which is rarely mutated in the considered tumour type, it adds weight towards a benign effect of the variant (BP10, e.g. *EGFR* and melanoma), but might also be indicative of an erroneous diagnosis.

Mutations in certain pathways are associated with a specific tumour phenotype, such as high tumour mutational burden, microsatellite instability or high homologous recombination deficiency (HRD) score. Such associations are therefore in favour of pathogenicity when observed (PM9) or of a benign effect when absent (BS6).

3.4.2. Co-occurrence of alteration on the same gene or in the same pathway

Mutual exclusivity of specific driver alterations is frequently observed before treatment, meaning that primary alterations targeting similar biological pathways are redundant, with one alteration being sufficient to deregulate the affected process. Mutually exclusive alterations often affect the same pathways. Mutual exclusivity needs to be estimated with a sufficient number of cases to draw a conclusion. In most situations, co-occurrence is very rare in a primary tumour. In the TSG, detection of a deleterious mutation in *cis* on the same gene favours a benign effect (BP2). Co-occurrence in *cis* is used for the *BRCA1* variant classification [33,34]. In oncogenes, co-occurrence of another oncogenic mutation in a gene from the same pathway favours a benign effect, as well as co-occurrence of an alternate driver event reported to be usually mutually exclusive (BP5).

These points should nonetheless be considered very cautiously, as true co-occurrence does exist within the same gene or pathway (e.g. co-occurrence in the PIK3CA or JAK-STAT signalling pathways or co-occurrence of Thr790Met and Cys797Ser in EGFR) and is often considered synergistic. It may mean that alteration of only one of the two genes would be insufficient, whereas cells with alterations in both progress to malignancy. An example is the combination of PIK3CA and PTEN mutations [35]. Co-occurrence of alterations on the same gene or pathway may also be due to parallel or convergent evolution of branched mutations [36]. Not all drivers are mutually exclusive; each tumour most often bears several oncogenic activating and inactivating events. The presence of another oncogenic driver event therefore depends on the context (acquired resistance alteration, sub-clonal evolution) and can only be used in the absence of any hypermutability. An example of this is the co-occurrence of up to three mutations in the EGFR gene, resulting in drug resistance.

3.4.3. Loss of heterozygosity and additional available data

The effects of the alteration at the RNA or protein level are additional valuable elements. It should always be interpreted considering all candidate pathogenic

variants in the gene because it will not indicate which variant is related to the observed effect. Tumour tissue analyses, such as immunohistochemical staining, transcriptome analysis (RNA), copy number alterations or methylation status, contribute useful elements for the classification. For example, overexpression of an oncogene like *ERBB2* is in favour of a pathogenic effect of a co-occurring *ERBB2* variant (PM7). Loss of heterozygosity (LOH) can be a factor of pathogenicity (PP7 for the TSG, PM7 for oncogenes). For oncogenes, the absence of the alteration at the protein and/or at the RNA level, if such information is available, is indicative of a likely benign effect (BS5). Some tumours show a high level of chromosomal instability as ovarian cancer, which may appear as LOH at the gene level, without being informative on the pathogenicity of the TSG variant. In that latter case, the absence of LOH may be considered as supporting benign (BP11) classification (e.g. LOH of the *BRCA1* or *BRCA2* wild-type allele in ovarian cancer). It is recommended to consider several tumours with the same variant before reaching a conclusion. For the TSG, the detection of the alteration in *trans* of a validated deleterious mutation is moderately in favour of pathogenicity (PM3). Moreover, integrated analysis of DNA and RNA sequencing can document an allelic imbalance or correlate overexpression with amplification or gene fusion.

3.4.4. Mutational signature

The presence of a tumour phenotype compatible with the alteration of the pathway is a useful criterion, such as an HR gene variant and high HRD score [37] or a *POLE* variant with a high mutational burden [38] (PM9 if the phenotype is observed and BS6 if the phenotype is absent). General chromosomal aberrations may also be a sign of the pathogenicity of the identified alteration, if it is localised to the alteration but not genome doubling with hyperploidy [39].

3.5. Criteria based on the variant behaviour after patient treatment

Little information is available on the variant behaviour after patient treatment because such sequencing data are rarely collected, although they may bring precious information for the variants' pathogenicity classification. If several cases with the same alteration respond to a therapy targeting the gene or pathway, the alteration is strongly likely to be pathogenic (PS3, Table 9). During targeted treatment, if an alteration arises on the targeted gene or pathway, the alteration could play a role in acquired resistance. For oncogenes, it is the onset of new variants after targeted treatment, for example resistance mutations in *ALK*, *ROS1*, *MET* are clearly functional variants, which usually prevent the binding of the drug. For oncogenes, occurrence or increase after targeted treatment or primary resistance to a targeted treatment in several cases with the alteration is in favour of a pathogenic effect (PM8).

For the TSG, the appearance of a reversion variant on the locus with a variant after a targeted treatment strongly suggests a benign effect of the combination of the two variants. The reversion variant itself, when present alone, may also be pathogenic, especially if it resets the correct open reading frame, for example *BRCA* reversions can be considered as functional variants. Also for the TSG, the occurrence of reversion on the locus with the alteration after a targeted treatment is very strongly in favour of an initial pathogenic variant (PVS2), for example the variant in *BRCA1* with a reversion which deletes this specific variant. The absence of response to a targeted treatment in a case report is weakly in favour of a benign effect (BP8), but it might just be that the tested drug was not appropriate.

4. Conclusion

Many relevant sources of information are available to support somatic variant interpretation, but as described, they should be used cautiously because the user should be aware of the criteria used for classification and potential discrepancies in classification between sources. Sharing results and expertise is highly valuable, and networking initiatives should be facilitated and encouraged. A web portal SoVad was implemented for sequencing platforms to share variants and communicate on somatic variant pathogenicity classification. It offers user-friendly implementation of the proposed classification workflow to facilitate its adoption. Other support systems are offering similar approaches, such as Franklin, and the scientific and medical value and friendly use of each of these systems need to be evaluated in large collaborative studies. This study was based on work by several French institutions in the SoVaD database. This is an important step to coordinate the information on rare somatic variants at a national level, and all the discussion should be extended to a broader international consortium. Similar guides would also be needed for the interpretation of somatic copy number alterations and fusion transcripts.

We expect the work presented here to help increase the quality of the genomic information used in patient management and the reproducibility of the classification among molecular tumour boards. We intend to validate interlaboratory consistency of the classification using an existing EQA program. Furthermore, the genetic and clinical data collected through genomic-based standard care and clinical trials could be used to fuel *de novo* AI

systems supporting both variant interpretation and the clinical decision-making process.

Author contributions section

Florence Koeppel: Writing - Original Draft, Project administration. Etienne Muller: Investigation, Writing - Review & Editing. Alexandre Harlé: Investigation, Writing - Review & Editing. Céline Guien: Software, Writing - Review & Editing. Pierre Sujobert: Writing - Review & Editing. Olfa Trabelsi Grati: Writing - Review & Editing. Olivier Kosmider: Writing - Review & Editing. Laurent Miguet: Writing - Review & Editing. Laurent Mauvieux: Writing - Review & Editing. Anne Cayre: Writing - Review & Editing. David Salgado: Software, Writing - Review & Editing.. Claude Preudhomme: Writing - Review & Editing. Lucie Karayan-Tapon: Writing - Review & Editing. Gaëlle Tachon PharmD: Writing - Review & Editing. Florence Coulet: Writing - Review & Editing. Alexandra Lespagnol: Writing - Review & Editing. Christophe Beroud: Software, Conceptualisation, Supervision, Writing - Review & Editing. Karen Leroy: Conceptualisation, Investigation, Writing - Original Draft. Etienne Rouleau: Conceptualisation, Supervision, Investigation, Funding acquisition. Isabelle Soubeyran: Conceptualisation, Investigation, Writing - Original Draft.

Funding

This research benefited from a funding by the French National Cancer Institute (INCa), grant number 2017-189.

Conflict of interest statement

The authors declare no conflict of interest.

Acknowledgements

Grateful acknowledgement is given to John Rowell for the proofreading of the manuscript. The SoVAD database and network benefit from the funding from National Cancer Institute (INCa - 2017-189 (NGS 2017)).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejca.2021.08.047.

References

[1] Chakravarty D, Gao J, Phillips S, Kundra R, Zhang H, Wang J, et al. OncoKB: a precision oncology knowledge base. JCO Precis Oncol 2017;1:1–16.

[2] Hyman DM, Puzanov I, Subbiah V, Faris JE, Chau I, Blay JY, et al. Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. N Engl J Med 2015;373:726–36.

[3] Condorelli R, Mosele F, Verret B, Bachelot T, Bedard PL, Cortes J, et al. Genomic alterations in breast cancer: level of evidence for actionability according to ESMO Scale for Clinical Actionability of molecular Targets (ESCAT). Ann Oncol 2019;30:365–73.

[4] Mateo J, Chakravarty D, Dienstmann R, Jezdic S, Gonzalez-Perez A, Lopez-Bigas N, et al. A framework to rank genomic alterations as targets for cancer precision medicine: the ESMO Scale for Clinical Actionability of molecular Targets (ESCAT). Ann Oncol 2018;29:1895–902.

[5] Supek F, Miñana B, Valcárcel J, Gabaldón T, Lehner B. Synonymous mutations frequently act as driver mutations in human cancers. Cell 2014;156:1324–35.

[6] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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;17:405–24.

[7] Amendola LM, Jarvik GP, Leo MC, McLaughlin HM, Akkari Y, Amaral MD, et al. Erratum: performance of ACMG-AMP variant-interpretation guidelines among nine laboratories in the clinical sequencing exploratory research Consortium. Am J Hum Genet 2016;98(6):1067–76. https://doi.org/10.1016/j.ajhg.2016.03 (S0002929716300593) Am J Hum Genet, 2016, 99:247.

[8] Van Allen EM, Wagle N, Levy MA. Clinical analysis and interpretation of cancer genome data. J Clin Oncol 2013;31:1825–33.

[9] Sukhai MA, Craddock KJ, Thomas M, Hansen AR, Zhang T, Siu L, et al. A classification system for clinical relevance of somatic variants identified in molecular profiling of cancer. Genet Med 2016;18:128–36.

[10] Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer. J Mol Diagnostics 2017;19:4–23.

[11] Ritter DI, Roychowdhury S, Roy A, Rao S, Landrum MJ, Sonkin D, et al. Somatic cancer variant curation and harmonization through consensus minimum variant level data. Genome Med 2016;8.

[12] Froyen G, Mercier M Le, Lierman E, Vandepoele K, Nollet F, Boone E, et al. Standardization of somatic variant classifications in solid and haematological tumours by a two-level approach of biological and clinical classes: an initiative of the belgian compermed expert panel. Cancers (Basel) 2019;11.

[13] Wagner AH, Walsh B, Mayfield G, Tamborero D, Sonkin D, Krysiak K, et al. A harmonized meta-knowledgebase of clinical interpretations of somatic genomic variants in cancer. Nat Genet 2020;52:448–57.

[14] Rouleau E, Leroy K, Dufraing K, Harlé A, Lacroix L, Vanwelden K, et al. First national external quality assessment for the interpretation of somatic variants: assessment of 25 variants in colorectal, lung, ovarian cancers and melanoma in France. Ann Oncol 2019;30:v30.

[15] Wang Z, Jensen MA, Zenklusen JC. A practical guide to the cancer genome Atlas (TCGA). Methods Mol Biol 2016;1418:111–41.

[16] Hudson TJ, Anderson W, Aretz A, Barker AD, Bell C, Bernabé RR, et al. International network of cancer genome projects. Nature 2010;464:993–8.

[17] Chang MT, Asthana S, Gao SP, Lee BH, Chapman JS, Kandoth C, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nat Biotechnol 2016;34:155–63.

[18] Gao J, Sumer SO, Zhang H, Solit DB, Taylor BS, Schultz N, et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. Genome Med 2017;9:4.

[19] Ainscough BJ, Griffith M, Coffman AC, Wagner AH, Kunisaki J, Choudhary MNK, et al. DoCM: a database of curated mutations in cancer. Nat Methods 2016;13:806–7.
[20] Griffith M, Spies NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet 2017;49:170–4.
[21] Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med 2018;10:25.
[22] Brogna S, Wen J. Nonsense-mediated mRNA decay (NMD) mechanisms. Nat Struct Mol Biol 2009. https://doi.org/10.1038/nsmb.1550.
[23] Suybeng V, Koeppel F, Harlé A, Rouleau E. Comparison of pathogenicity prediction tools on somatic variants. J Mol Diagnostics 2020. https://doi.org/10.1016/j.jmoldx.2020.08.007.
[24] Jain P, O'Roak BJ, Cooper GM, Witten DM, Shendure J, Kircher M. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310–5.
[25] González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet 2011;88:440–9.
[26] Li MX, Gui HS, Kwan JSH, Bao SY, Sham PC. A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. Nucleic Acids Res 2012;40:e53.
[27] Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 2004;11:377–94.
[28] Reese MG. Improved splice site detection in Genie. J Comput Biol 1997;4:311–23.
[29] Pertea M. GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Res 2001;29:1185–90.
[30] Desmet FO, Hamroun D, Lalande M, Collod-Bëroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009. https://doi.org/10.1093/nar/gkp215.
[31] Leman R, Gaildrat P, Gac GL, Ka C, Fichou Y, Audrezet MP, et al. Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined in silico/in vitro studies: an international collaborative effort. Nucleic Acids Res 2018;46:7913–23.
[32] Salgado D, Desvignes JP, Rai G, Blanchard A, Miltgen M, Pinard A, et al. UMD-predictor: a high-throughput sequencing compliant system for pathogenicity prediction of any human cDNA substitution. Hum Mutat 2016;37:439–46.
[33] Gowen LC, Johnson BL, Latour AM, Sulik KK, Koller BH. Brca1 deficiency results in early embryonic lethality characterized by neuroepithelial abnormalities. Nat Genet 1996;12.191–104.
[34] Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J Med Genet 2006;43:295–305.
[35] Pearson HB, Li J, Meniel VS, Fennell CM, Waring P, Montgomery KG, et al. Identifi cation of pik3ca mutation as a genetic driver of prostate cancer that cooperates with pten loss to accelerate progression and castration-resistant growth. Canc Discov 2018;8:764–79.
[36] Venkatesan S, Birkbak NJ, Swanton C. Constraints in cancer evolution. Biochem Soc Trans 2017;45:1–13.
[37] Timms KM, Abkevich V, Hughes E, Neff C, Reid J, Morris B, et al. Association of BRCA1/2 defects with genomic scores predictive of DNA damage repair deficiency among breast cancer subtypes. Breast Cancer Res 2014;16:475.
[38] Hino H, Shiomi A, Kusuhara M, Kagawa H, Yamakawa Y, Hatakeyama K, et al. Clinicopathological and mutational analyses of colorectal cancer with mutations in the POLE gene. Cancer Med 2019;8:4587–97.
[39] Bielski CM, Zehir A, Penson AV, Donoghue MTA, Chatila W, Armenia J, et al. Genome doubling shapes the evolution and prognosis of advanced cancers. Nat Genet 2018;50:1189–95.