The TP53 UMD database

TP53 Mutation data

2017 Release R2

Read me: The mutation database

**Thierry Soussi and Bernard Leroy**

Leroy B, Anderson M, Soussi T. 2014. TP53 Mutations in Human Cancer: Database Reassessment and Prospects for the Next Decade. Hum Mutat 35: 672-688.

Soussi T. 2014. The TP53 gene network in a postgenomic era. Hum Mutat 35: 641-642.

Soussi T, Leroy B, Taschner PE. 2014. Recommendations for Analyzing and Reporting TP53 Gene Variants in the High-Throughput Sequencing Era. Hum Mutat 35: 766-778.

Soussi T. 2014. Locus-Specific Databases in Cancer: What Future in a Post-Genomic Era? The TP53 LSDB Paradigm. Hum Mutat 35: 643-653.

Leroy B, Girard L, Hollestelle A, Minna JD, Gazdar AF, Soussi T. 2014. Analysis of TP53 Mutation Status in Human Cancer Cell Lines: A Reassessment. Hum Mutat 35: 756-765.

Leroy, B., M. L. Ballinger, F. Baran-Marszak, G. L. Bond, A. Braithwaite, N. Concin, L. A. Donehower, W. S. El-Deiry, P. Fenaux, G. Gaidano, A. Langerød, E. Hellstrom-Lindberg, R. Iggo, J. Lehmann-Che, P. L. Mai, D. Malkin, U. M. Moll, J. N. Myers, K. E. Nichols, S. Pospisilova, P. Ashton-Prolla, D. Rossi, S. A. Savage, L. C. Strong, P. N. Tonin, R. Zeillinger, T. Zenz, J. F. Fraumeni, P. E. Taschner, P. Hainaut, and T. Soussi. 2017. Recommended Guidelines for Validation, Quality Control, and Reporting of TP53 Variants in Clinical Practice. *Cancer Res* 6: 1250-1260.

Read Me 1.00

TP53

**Important note: only the coding strand of the gene is used for the description of TP53 variants**

The UMD database comes in two files, the variant and the mutation database

The **mutation database** includes all patients carrying a TP53 mutation. Therefore, different patients expressing the same TP53 variant are included in this database.

The **variant database** includes each single TP53 variants found in the cases database.



Relationship between the variant and the mutation database (numbers can be slighly different in the files due to an update of the database)

**This read me file is specific for the mutation database.**

Mutation identifiers

|  |  |
| --- | --- |
| **cDNA\_Variant** | Mutation nomenclature according to HGVS standards using the coding sequence as reference (position 1 refers to the A of the start ATG): reference sequence [NM\_000546.5](http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NM_000546.5) |
| **UMD\_ID** | Unique mutation identifier used in the UMD database for each genomic variant |
| **COSMIC\_ID** | Mutation identifier used in COSMIC  <http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/> |
| **ATCC** | An official reference is available for several cell lines distributed by repository center  http://www.lgcstandards-atcc.org/ |
| **SNP\_ID\*** | The SNP database now includes several pathogenic variants of the TP53 gene  <http://www.ncbi.nlm.nih.gov/snp> |

\* **A note of caution:** Since 2011 (build 134), dbSNP started accepting submissions of germ line and somatic variations associated with various types of diseases and changed its name to “database of Short Genetic Variation” keeping the dbSNP acronym. Several frequent *TP53* variants (rs121912651, c.742C>T, p.Arg248Trp or rs11540652, c.743G>A, p.Arg248Gln) are included in dbSNP, but other hot spot variants are missing, whereas rare somatic variants **may be included**. This heterogeneity caused by biased dbSNP submissions is misleading, as it does not reflect the true occurrence and frequencies of *TP53* variants. Therefore, without further distinction, we can no longer assume that variants in dbSNP are associated with the lack of effect on disease and tumour characteristics.

Common SNPs such as rs1042522 (p.P72R), rs1800371 (p.P47S), rs1800372 (p.R213R) or rs1800370 (p.P36P) are not included in the database.

Mutation coordinates (Nucleotide)

|  |  |
| --- | --- |
| **HG19\_Start** | Mutation start coordinates using HG19 as reference |
| **HG19\_End** | Mutation end coordinates using HG19 as reference |
| **HG18\_Start** | Mutation start coordinates using HG18 as reference |
| **HG18\_End** | Mutation end coordinates using HG18 as reference |
| **HG38\_Start** | Mutation start coordinates using CRCh38 as reference |
| **HG38\_End** | Mutation end coordinates using CRCH38 as reference |
| **NG\_017013.2** | Nucleotide position using NCBI reference sequence NG\_0117013.2  <http://www.ncbi.nlm.nih.gov/nuccore/NG_017013.2>  This sequence is also the reference used by the Locus Reference Genomic  (<http://ftp.ebi.ac.uk/pub/databases/lrgex/LRG_321.xml>) |
| **HG19\_Variant** | Mutation nomenclature according to HGVS standards using the genomic sequence as reference  Reference sequence: NC\_000017.10 for genome build NCBI37/hg19 |
| **HG18\_Variant** | Mutation nomenclature according to HGVS standards using the genomic sequence as reference  Reference sequence: NC\_000017.9 for genome build NCBI36/hg18 |
| **HG38\_Variant** | Mutation nomenclature according to HGVS standards using the genomic sequence as reference  Reference sequence: NC\_000017.11 for genome build GRCh38.p2 |
| **NG\_017013.2\_variant** | Mutation nomenclature according to HGVS standards using the RefSeq Gene NG\_017013. sequence as reference  <http://www.ncbi.nlm.nih.gov/nuccore/NG_017013.2>  This sequence is also the reference used by the Locus Reference Genomic  (<http://ftp.ebi.ac.uk/pub/databases/lrgex/LRG_321.xml> |
| **Start\_cDNA** | Mutation start coordinate using the p53 cDNA as reference (position 1 refers to the A of the start ATG): reference sequence [NM\_000546.5](http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NM_000546.5) |
| **End\_cDNA** | Mutation end coordinate using the p53 cDNA as reference (position 1 refers to the A of the start ATG): reference sequence [NM\_000546.5](http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NM_000546.5) |
| **Exon:intron\_Start** | Location of the mutation start in the introns or exons of the TP53 gene. In most cases, Exon:intron\_Start and Exon:intron\_stop are similar.  A few large deletions encompass several exons and introns.  Although all intronic variations described in the literature have been included in the database, only mutations that target the canonical AG splice-acceptor site or GT splice-donor site (-1,-2, +1, +2) are considered to be pathogenic (see splice comment for more information). |
| **Exon:intron\_End** | Location of the mutation end in the introns or exons of the TP53 gene. In most cases, Exon:intron\_Start and Exon:intron\_stop are similar.  A few large deletions encompass several exons and introns.  Although all intronic variations described in the literature have been included in the database, only mutations that target the canonical AG splice-acceptor site or GT splice-donor site (-1,-2, +1, +2) are considered to be pathogenic (see splice comment for more information). |
| **Genome\_Base\_Coding** | Nucleotide at the start position of the mutation. |
| **cDNA\_Variant** | Mutation nomenclature according to HGVS standards using the coding sequence as reference (position 1 refers to the A of the start ATG): reference sequence [NM\_000546.5](http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NM_000546.5) |

Mutation coordinates (Protein P1, TP53 alpha, NP\_000537.3, 393 aa)

|  |  |
| --- | --- |
| **Codon** | **1-393:** Codon position using TP53 alpha (p1) as reference (NP\_000537.2)  **Splice**: mutations that target the canonical AG acceptor site or GT donor.  **Untranslated:** mutations that target other nucleotides (5'UTR; 3,UTR or Intron)  Large deletions with unknown boundaries are shown as "?".  **\***: stop codon  **nnn-beta or nnn-gamma**: codon position specific for isoforms beta and gamma |
| **WT AA\_1** | Wild-type amino acid: 1-letter nomenclature.  For intronic mutations, the intron number is displayed (**intron\_01 to intron 10, intron\_09\_beta and intron\_09\_gamma**).  Mutations that target the canonical AG splice-acceptor site or GT splice-donor site are displayed as i**ntron\_nn\_SA** or **intron\_nn\_SA**,where nn is the intron number. |
| **WT AA\_3** | Wild-type amino acid: 3-letter nomenclature.  For intronic mutations, the intron number is displayed (**intron\_01 to intron 10, intron\_09\_beta and intron\_09\_gamma**)  Mutations that target the canonical AG splice-acceptor site or GT splice-donor site are displayed as i**ntron\_nn\_SA** or **intron\_nn\_SA**,where nn is the intron number. |
| **Mutant AA\_1** | Mutant amino acid: 1-letter nomenclature |
| **Mutant AA\_3** | Mutant amino acid: 3-letter nomenclature |
| **Structure** | Structural motif of the TP53 protein according to the analysis described by Cho et al. (1994). |
| **Domain** | Domain of the TP53 protein   * **HCD I to V**: Highly Conserved Domain I to V * **DNA Binding**: DNA binding domain * **Negative regulation:** carboxy-terminus of the p53 protein associated with negative regulation of p53 DNA binding activity * **Transactivation TAD1:** transactivation domain 1 * **Transactivation TAD2:** transactivation domain 2 * **Proline Rich**: Proline-rich domain of the p53 protein * **NES:** Nuclear export signal of p53 * **NLS:** Nuclear localization signal of p53 * **Oligomerization:** Tetramerization domain of the p53 protein * Empty field: No specific domain available |
| **PTM** | Post-translational modifications   * Lys Acetylation * Lys Ubiquitination * Asp and Glu ADP Ribosylation * Ser or Thr Phosphorylation * Ser O-Linked Glycosylation * Cys Glutathionylation * Arg and Lys Methylation * Asn Isoapartyl methylation * Lys Neddylation * Tyr Nitrosylation * Lys Methylation |

Mutation features

|  |  |
| --- | --- |
| **Base\_Change\_Size** | Size of the substitution |
| **Ins\_Size** | Size of the deletion |
| **Del\_Size** | Size of the insertion; the sequence of the insertion is available for a few cases. |
| **Mutant\_Allele** | Mutant (Alt) nucleotide  For deletion, this field is empty  For insertion, this field includes the inserted sequence except when this sequence is unknown and is therefore left empty. |
| **WT\_Codon** | Nucleotide sequence of the wild-type codon in which the mutation occurred.  Mutations that target the canonical AG splice-acceptor site or GT splice-donor site are displayed as i**ntron\_nn\_SA** or **intron\_nn\_SA**,where nn is the intron number.  For intronic mutations, the intron number is displayed (**intron\_01 to intron 10, intron\_09\_beta and intron\_09\_gamma**)  Large deletions with unknown boundaries are shown as “**?**” |
| **Mutant\_Codon** | **NNN:** Sequence of the mutated codon.  **Del**: exonic deletion  **Ins**: exonic insertion  **Indel**: complex event that involves an exonic insertion and a deletion.  In accordance with the new HGVS rules, all tandem mutations are now included in this category as del2ins2 events.  See the HGVS website for more information (http://www.hgvs.org/mutnomen/).  **Splice**: mutation that targets the canonical AG splice-acceptor site or GT splice-donor site.  **NR:** not relevant, mutations targeting intronic sequence, 5'UTR or 3'UTR. |
| **Mutational\_Event** | Mutational events  **G>C**: (G to C base substitutions)  All other single substitutions are described in a similar way.  **CC>TT**: mutation that changes two contiguous nucleotides.  **Insertion**  **Deletion**  **Indel:** complex event that involves an insertion and a deletion.  **Important note: only the coding strand of the TP53 gene is used for mutation description.** |
| **Type** | **Ts:** Transition (a pyrimidine (C or T) is substituted by another pyrimidine, or a purine (A or G) is substituted by another purine);  **Tv:** Transversion (a transversion mutation involves substitution of a pyrimidine by a purine, or vice versa);  **Td**: tandem mutation  **Fr:** Frameshift mutations (deletions / insertions)  **Inf:** In-frame deletions or insertions |
| **Mutation\_Type** | **SNV:** Single nucleotide variant  **D**: Deletion  **I**: insertion  **ID**: complex event that involves an insertion and a deletion. |
| **CpG** | **Yes:** transition (G to A or C to T base change) at a CpG dinucleotide;  **No:** transitions (G to A or C to T base change) at non-CpG sites; |

|  |  |
| --- | --- |
| **Py-Py\_Doublets** | Indicates whether or not a mutation targets a Py-Py doublet  **Py-Py doublet:** two adjacent pyrimidine residues (cytosine or thymine) that can be targeted by UV light.  **Yes, non-coding strand:** mutation located at a Py-Py doublet on the non-coding strand of the p53 gene.  **Yes, coding strand:** mutation located at a Py-Py doublet on the coding strand of the p53 gene.  **No:** mutation located outside a Py-Py doublet. |
| **Tandem\_Class** | The majority of tandem mutations are found in skin tumours.  Several types of tandem mutations can occur in the open reading frame of the TP53 gene (or any other genes).  These mutations are considered to be single mutational events linked to UV exposure  **T1:** two different codons are modified by the substitution.  e.g.: codons 247 and 248 of the TP53 gene: AA**C - C**GG -> AA**T- A**GG  c.741\_742delCCInsTA (p.[N247N; R248R]  In the majority of T1 tandem mutations, the first substitution does not change the amino acid residue and results in a synonymous change.  **T2**: only one codon is modified by the substitution.  e.g.: codon 331, C**AG** >C**CA**.  c.992\_993delAGInsCA (p.Q331P).  **T3:** the tandem mutation occurs across a splice site |
| **Variant\_Classification** | Translational effect of the mutation  (Missense, Nonsense, Synonymous, Nonstop, In\_frame\_Del, Inframe\_Ins, Frameshift\_Del or Frameshift\_Ins). |
| **Variant\_Comment** | Specific comment concerning the consequences of the mutation. |
| **Variant\_Type** | Variant type as defined in MAF file  **SNP**: Single Nucleotide Variant  **DNP**: Change in two consecutive bases (dinucleotide polymorphism)  **TNP**: Change in three consecutive bases (tri-nucleotide polymorphism)  **ONP**: Change in four or more consecutive bases (oligo-nucleotide polymorphism)  **INS**: Insertion  **DEL**: Deletion |

Sample data information

|  |  |
| --- | --- |
| **Mutation\_Origin** | **Somatic**: Somatic mutations.  **Note of caution**:  In most cases, only tumour DNA is sequenced and alterations are compared to a reference sequence (not match DNA). It is likely that several of the mutations described in the database could be inherited.  **Germline**: inherited TP53 mutation.  Common SNPs such as rs1042522 (p.P72R), rs1800371 (p.P47S), rs1800372 (p.R213R) or rs1800370 (p.P36P) are not included in the database. Large sequencing projects such as the 1000 genomes project have revealed novel infrequent TP53 variants in the coding region of the gene. It is currently unknown whether or not these variants are pathogenic. Refer to the SNP section for more information. |
| **Disease** | Name of the disease as indicated in the publication. |
| **Sample\_pathology** | **Cancer:** all types of tumours.  **Premalignant diseases:** premalignant lesions, such as colorectal adenoma or prostatic intraepithelial neoplasia.  **Non-malignant disease**: non-malignant diseases such as gastritis or rheumatoid arthritis that have been shown to be associated with TP53 mutations. Statistical analysis showed that the pattern of these mutations is very different from those observed in malignant tumours; these diseases are therefore unsuitable for all cancer-related analyses.  **No disease:** cancer-free individuals with particular features (see the exposure and comments section for more information).  **Unknown:** sample of unknown origin |
| **Sample\_Origin** | Nature of the sample in which the mutation was identified.  **Adjacent tissue:** normal tissue (defined histologically) surrounding the tumour. Despite the strictest precautions, this tissue can contain a few tumour cells. Tissue from individuals without malignant disease is not included here.  **Adjacent tissue (stroma):** Stromal cells obtained by microdissection  **Cell line:** self-explanatory. For cell lines, when the status of the original tumour is also known, a second entry with the same name but with a “\_t” suffix is included in the database  **Circulating tumour cells**: DNA extracted from circulating tumour cells.  **Extra cellular DNA:** free DNA extracted from sputum, urine, plasma or serum.  **Normal tissue:** tissue from cancer-free individuals with particular features (see the exposure and comments section for more information). When available, the origin of the tissue is indicated.  **Pathological tissue:** sample from patients with non-malignant disease  **Peripheral blood lymphocytes:** self-explanatory.  **Tumour:** pathological sample from a cancer patient.  **Xenograft:** human tumour transplanted in mice. |

|  |  |
| --- | --- |
| **Name** | Name of the tumour/patient/cell line as indicated by the authors.  If the publication does not include a sample name, we have arbitrarily assigned a name, usually the first letters of the author’s last name, followed by the numbers in the series.  The same name or number can occur several times in a single study, as more than one mutation may be reported in some samples (see “complexity” column to find samples with multiple mutations).  For cell lines, we use the name given in the publication. Unfortunately, this can be confusing, as similar cell lines may be identified by multiple names (acronym or ATCC references, change/errors in the acronym). Every effort has been taken to ensure homogeneity of cell line names.  For cell lines available at the ATCC, the full name has been corrected when necessary |
| **Hereditary\_Syndrome** | Sample from patients with particular hereditary syndromes such as Fanconi anaemia, Li-Fraumeni syndrome or Xeroderma pigmentosum.  Several syndromes such as Li-Fraumeni or familial adrenocortical carcinoma in Brazil are associated with germline TP53 mutations  Other hereditary syndromes such as Familial adenomatous polyposis (FAP) or Xeroderma pigmentosum are associated with specific cancers with somatic TP53 mutations.  **Important note**: Most patients from Brazil with familial adrenocortical carcinoma carry the same germline mutations in the TP53 gene (p.R337H). This founder mutation has been shown to be very frequent in Southern Brazil, present in up to 0.3% of the population.  This germline mutation has been added only once in the UMD TP53 database to take into account the fact that it originates from a single event. |
| **Genetic\_Background** | Tumours can occur in patients with known germline mutations in other tumour genes that could shape TP53 mutations.  Only patients with clear genotype information have been included.  BRCA1: patients with a germline mutation in the BRCA1 gene  BRCA2: patients with a germline mutation in the BRCA2 gene  CHK2: patients with a germline mutation in the checkpoint kinase 2 gene associated with Li-Fraumeni syndrome  hMLH1: patients with germline mutations in the MLH1 gene associated with HNPCC.  hMSH2: patients with germline mutations in the MSH2 gene associated with HNPCC.  NBS1: patients with germline mutations in the NBS1 gene associated with Nijmegen breakage syndrome  NF1: patients with a germline mutation in the NF1 gene  PTEN; patients with a germline mutation in the Phosphatase and TENsin homologue gene associated with Cowden disease |
| **Internal** | **Internal tumours:** all internal tumours  **Skin tumour**: all skin tumours |
| **Solid** | **Solid tumour:** Non-hematological tumours  **Hematological malignancy:** all forms of leukemia and lymphomas |
| **Smoking** | Information on the patient’s smoking status  Yes, No, Unknown or Ex-smoker |
| **Aflatoxin** | Patient’s exposure to Aflatoxin B1  Yes, No or Unknown |
| **Radiations** | Patient’s exposure to radiation  Yes, No or Unknown |
| **Drinking** | Information on the patient’s drinking status  Yes, No or Unknown |
| **Asbestos** | Patient’s exposure to asbestos  Yes, No or Unknown |
| **Hepatitis\_B** | Detection of HBV in the tumour  Yes, No or Unknown |
| **Papilloma** | Detection of HPV in the tumour  Yes, No or Unknown |
| **Exposure** | Other exposure or viral infection such as Aristolochic acid, radon or HCV |
| **Complexity** | **SM**: Single mutational event in the tumour;  **DMU** (Double Mutation Unknown): Two p53 mutations in the same tumour, but their allelic distribution is unknown.  **DMD** (Double Mutation Different allele): Two p53 mutations in the same tumour on two different p53 alleles.  **DMS** (Double Mutation Same allele): Two p53 mutations in the same tumour on the same p53 allele.  **MM**(Multiple Mutation): More than two p53 mutations in the same tumour.  Tandem mutations are considered to be a single event in a tumour, as they arise from a unique alteration to two adjacent nucleotides. |
| **Unidentified\_Mutation** | Some mutations have been identified by pre-screening methodologies, but are not fully described and are indicated in this field. |

Mutation assessment and pathogenicity

|  |  |
| --- | --- |
| **Records\_Number** | Number of occurrences of the mutant in the database. |
| **Leukaemia\_Stat** | Frequency of the variant (cDNA\_nomenclature) in hematological malignancies  e.g.: for variant c.524G>A, the entry will be 2.81 (159/5,642)  Frequency of c.524G>A in hematological malignancies: 2.81 %  c.524G>A in hematological malignancies: 159  Total number of hematological malignancies in the database: 5,642 |
| **Solid\_Stat** | Frequency of the variant (cDNA\_nomenclature) in solid tumours  e.g.: for variant c.524G>A, the entry will be 4.36 (2,260/51,782)  Frequency of c.524G>A in Solid tumours: 4.36 %  c.524G>A in Solid tumours: 2,260  Total number of Solid tumours in the database: 51,782 |
| **Tumour\_Stat** | Frequency of the variant (cDNA\_nomenclature) in tumours only (excluding variants from cell lines, germline and non-neoplastic diseases)  e.g.: for variant c.524G>A, the entry will be 4.18 (2,257/53,875)  Frequency of c.524G>A in tumours: 4.18 %  c.524G>A in tumours: 2,257  Total number of tumours in the database: 53,875 |
| **Cell\_line\_Stat** | Frequency of the variant (cDNA\_nomenclature) in cell lines only (excluding variants from tumours, germline and non-neoplastic diseases)  e.g.: for variant c.524G>A, the entry will be 3.83 (143/3,729)  Frequency of c.524G>A in tumours: 3.83 %  c.524G>A in tumours: 143  Total number of cell lines in the database: 3,729 |
| **Somatic\_Stat** | Frequency of the variant (cDNA\_nomenclature) found as a somatic event  e.g.: for variant c.524G>A, the entry will be 4.13 (2,439/59,003)  Frequency of c.524G>A in tumours: 4.13 %  c.524G>A in tumours: 2,439  Total number of cell lines in the database: 59,003 |
| **Germline\_Stat** | Frequency of the variant (cDNA\_nomenclature) found as a germline event  e.g.: for variant c.524G>A, the entry will be 5.40 (48/888)  Frequency of c.524G>A in tumours: 5.40 %  c.524G>A in tumours: 48  Total number of cell lines in the database: 888  Note of caution: for multiple variants, this frequency will be 0, as many of these variants (particularly those associated with carcinogen exposure) are only found as somatic events. A value of 0 should therefore not be used to predict mutation pathogenicity. |
| **Tumour\_ Repetition** | Several tumours display more than one mutation.  Tumour\_rep denotes the total number of mutations associated with this mutant in a single tumour.  e.g.:  1: only one mutation found in the tumour  14: 14 different mutations found in the tumour  All mutations found in a single tumour can be displayed by using the “tag-name” field that is unique for each tumour in the database. Values of three and more are highly suspicious. |
| **Publication\_Repetition** | Number of occurrences of this particular mutant in the publication.  A large number of non-hot spot mutants in a single publication is highly suspicious. |
| **Comment\_Prediction** | Several prediction algorithms have been used to predict TP53 pathogenicity (SIFT, Mutassessor, Provean, PolyPhen, see the corresponding rows for each individual analysis).  A prediction index has been deduced from the various analyses  Damaging  Probably damaging  Tolerated    Note of caution: for TP53 mutation, the sensitivity of the various algorithms is never higher than 80%  Other parameters such as frequency in the database or residual activity are more predictive for pathogenicity as the requirements between loss of function and selection as a driver mutation are different.  e.g.: c.69G>T (p.W23C)  this mutation targets a highly conserved residue of the TP53 protein localized in the binding domain for the mdm2 protein.  It is predicted to be highly deleterious by all of the currently available predictive algorithms, but transactivation activity for this mutant is not impaired.  This mutation has never been described in human cancer, as it is defective for mdm2 binding and will be counter-selected in human cancer as it is lethal. |
| **Comment\_Frequency** | Specific information related to the frequency of the mutation in the database.  Four categories have been defined:  i: This mutation is very frequent  ii) This mutation is frequent  iii: This mutation is not frequent  iv: This mutation is rare  see Leroy et al. TP53 Mutations in Human Cancer: Database Reassessment and Prospects for the Next Decade. Human Mutation (2014) 35, 672-688 |
| **Comment\_Outliers** | Indicates whether or not the mutation is associated with outlier publications.  Rare mutants only found in outlier studies should be considered to be suspicious. |
| **Comment\_Activity** | Specific information related to the residual activity of this TP53 mutant in the database based on the overall transcriptional activity (TA) on 8 different promoters as measured by Kato et al. For each mutant, the median of the 8 promoter-specific activities (expressed as percent of the wild-type protein) has been calculated.  For **missense variants,** five categories have been defined:  No activity: median <=20  Partial activity: median >20 and <=75  Fully active: median >75 and <=140  Hyper active: median >140  No data: this mutant has not been tested  For **nonsense** **variants**, one category has been used  The activity of truncated p53 is assumed to be nil  For **frameshift** variants, two categories have been used:  The consequence of this in-frame mutation is unknown (In-frame of 15 bp or less).  The activity of truncated p53 is assumed to be nil (out-of-frame insertion and deletion, in-frame mutation >18 bp or mutation across an intron:exon junction).  For **synonymous** variants, two categories have been used:  This synonymous mutation is known to impair TP53 splicing.  Synonymous mutation with unknown consequences.  For mutations that target the canonical AG **splice-acceptor site** or GT **splice-donor site**:  Splicing defect: impaired TP53 activity  Activity for each individual promoter is also available (see the various rows in the database: WAF, MDM2, BAX, 14-3-3-s, AIP, GADD45, NOXA and P53R2). |
| **Comment\_Splicing** | Indicates whether or not a mutation could impair TP53 splicing.  **For exonic mutation:**  All TP53 gene substitutions have been analyzed by using mutpred\_splice  M. Mort *et al.*, *Genome Biol* **15**, R19 (2014) courtesy of M. Mort  A MutPred Splice general score probability cutoff of ≥0.70 was used to indicate a predicted SAV.  For mutations close to an exon, a cutoff of ≥0.60 was used.  Raw data for mutpred\_splice are also available in this table.  **For intronic mutation**  Although all intronic variations described in the literature have been included in the database, only mutations that target the canonical AG splice-acceptor site or GT splice-donor site (-1,-2, +1, +2) are considered to be pathogenic. |

|  |  |
| --- | --- |
| **Sift\_Score** | Predicted functional effect using SIFT algorithm  http://sift.jcvi.org/  SIFT (Sorting Intolerant From Tolerant) prediction is based on the degree of conservation of amino acid residues in sequence alignments derived from closely related sequences.  Ranges from 0 to 1. The amino acid substitution is predicted to be damaging when the score is <= 0.05, and tolerated when the score is > 0.05. |
| **Sift\_Prediction** | Predictive value using Sift |
| **Provean\_Score** | Predicted functional effect using Provean algorithm  http://provean.jcvi.org/index.php  **PROVEAN** (**Pro**tein **V**ariation **E**ffect **An**alyzer) is a software tool which predicts whether an amino acid substitution or indel has an impact on the biological function of a protein.  PROVEAN introduces a **delta alignment score** based on the reference and variant versions of a protein query sequence with respect to sequence homologues collected from the NCBI NR protein database through BLAST. For maximum separation of deleterious and neutral variants for all 4 classes of human protein variants, the **default score cutoff** is currently set at -2.5 for binary classification (i.e. deleterious vs neutral). |
| **Provean\_prediction** | Prediction - deleterious or neutral (using default cutoff at -2.5) |
| **Mutassessor\_score:** | Predicted functional effect using Mutassessor algorithm  http://mutationassessor.org/  B. Reva, Y. Antipin, C. Sander, *Nucleic Acids Res* **39**, e118 (2011).  Functional impact combined score  the **default score cutoff** is currently set at -1.938 for classification (i.e. High or medium vs low or neutral). |
| **Mutassessor\_prediction:** | Functional impact of a variant : predicted functional (high, medium), predicted non-functional (low, neutral). |
| **Polyphen:** | Predicted functional effect using Polyphen algorithm  **PolyPhen-2** is an automatic tool for prediction of the possible impact of an amino acid substitution on the structure and function of a human protein. This prediction is based on a number of features comprising the sequence, phylogenetic and structural information characterizing the substitution.  http://genetics.bwh.harvard.edu/pph2/  Qualitative ternary classification appraised at 5%/10% (HumDiv) or 10%/20% (HumVar) FPR thresholds ("benign", "possibly damaging", "probably damaging") |

|  |  |
| --- | --- |
| **Condel:** | Predicted functional effect using Condel algorithm  <http://bg.upf.edu/fannsdb/> |
| **MutPred\_Splice\_General\_Score** | http://mutdb.org/mutpredsplice/about.htm  The MutPred Splice outputs are:  1, **General Score,** which is the probability that the **exonic** variant disrupts splicing. We use a general score >=0.6 to identify a variant which disrupts splicing.  e.g. general score >=0.6 labels a variant as a Splice Affecting Variant (SAV)  e.g. general score <0.6 labels a variant as a Splice Neutral Variant (SNV)  2, Additional supporting evidence is provided by a **confident hypothesis** about the splicing mechanism disrupted.  Practical advice  MutPred Splice can be used to prioritise your dataset into three partitions:  1, High Confident calls of splicing variants - predicted SAV (general score >=0.6) where a confident hypothesis is available.  2, Confident calls of splicing variants - predicted SAV (general score >=0.6) where a confident hypothesis not available.  3, Not predicted to disrupt splicing (SNV) (general score <0.6). |
| **Prediction\_Label** | See above and http://mutdb.org/mutpredsplice/about.htm |
| **Confident\_Hypotheses** | See above and http://mutdb.org/mutpredsplice/about.htm |

Mutation coordinates (all transcripts t1 to t8 and all protein isoforms p1 to 12)

|  |  |
| --- | --- |
| *Mutation nomenclature and coordinates used below are described according to HGVS, NCBI and LRG* | |
| Transcript t1 | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence NM\_000546.5 or LRG\_321t1* |
| Transcript t2 | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence NM\_001126112.2 or LRG\_321t2* |
| Transcript t3 | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence NM\_001126114.2 or LRG\_321t3* |
| Transcript t4 | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence NM\_001126113.2 or LRG\_321t4* |
| Transcript t5 | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence NM\_001126115.1 or LRG\_321t5* |
| Transcript t6 | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence NM\_001126116.1 or LRG\_321t6* |
| Transcript t7 | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence NM\_001126117.1 or LRG\_321t7* |
| Transcript t8 | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence NM\_001126118.1 or LRG\_321t8* |
| Protein p1 (TP53\_alpha) | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_000537.3 or LRG\_321p1* |
| Protein p3 (TP53\_beta) | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001119586.1 or LRG\_321p3* |
| Protein p4 (TP53\_gamma) | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001119585.1 or LRG\_321p4* |
| Protein p8 (Delta\_40\_TP53) | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263690.1 or LRG\_321p8* |
| Protein p9 (Delta\_40\_TP53\_beta) | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263625.1 or LRG\_321p9* |
| Protein p10 (Delta\_40\_TP53\_gamma) | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263624.1 or LRG\_321p10* |
| Protein p5 (Delta\_133\_TP53\_alpha) | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001119587.1 or LRG\_321p5* |
| Protein p6 (Delta\_133\_TP53\_beta) | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001119588.1 or LRG\_321p6* |
| Protein p7 (Delta\_133\_TP53\_gamma) | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001119589.1 or LRG\_321p7* |
| Protein p11 (Delta\_160\_TP53\_alpha) | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263626.1 or LRG\_321p11* |
| Protein p12 (Delta\_160\_TP53\_beta) | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263627.1 or LRG\_321p12* |
| Protein p13 (Delta\_160\_TP53\_gamma) | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263628.1 or LRG\_321p13* |

Mutant activity

|  |  |
| --- | --- |
| **Mutant activities (mutant)** | Data for WAF, MDM2, BAX, 14-3-3-s, AIP, GADD45, NOXA and P53R2 are taken from the publication by Kato et al. (Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R, Ishioka C (2003) Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. Proc Natl Acad Sci U S A 100: 8424-8429).  Transactivation was tested using a yeast assay. The residual transcriptional activity of mutant p53 is always compared to wild-type p53 for the same promoter (%).  **Syn**: mutation that does not change the amino acid: however, some of these mutations can change splicing or RNA stability.  **Fr**: Frameshift mutations. No activity data are available, but it is generally assumed that no p53 is produced.  **Tr:** Terminating mutation: No activity data are available, but it is generally assumed that no p53 is produced.  **ND**: No data available for this mutant.  **Splice:** splice mutation. No activity data are available, but it is generally assumed that no p53 is produced.  see **Comment frequency** for a final assessment |
| **WAF1** | Residual transcriptional activity of mutant p53 on the WAF1 promoter (% compared to wild-type p53). |
| **MDM2** | Residual transcriptional activity of mutant p53 on the MDM2 promoter (% compared to wild-type p53). |
| **BAX** | Residual transcriptional activity of mutant p53 on the BAX promoter (% compared to wild-type p53). |
| **14\_3\_3\_s** | Residual transcriptional activity of mutant p53 on the 14-3-3- promoter (% compared to wild-type p53). |
| **AIP** | Residual transcriptional activity of mutant p53 on the AIP promoter (% compared to wild-type p53). |
| **GADD45** | Residual transcriptional activity of mutant p53 on the GADD45 promoter (% compared to wild-type p53). |
| **NOXA** | Residual transcriptional activity of mutant p53 on the NOXA promoter (% compared to wild-type p53). |
| **p53R2** | Residual transcriptional activity of mutant p53 on the p52R2 promoter (% compared to wild-type p53). |

References

|  |  |
| --- | --- |
| **Reference ID** | Reference identification number |
| **Authors** | Authors |
| **Year** | Year |
| **Title** | Title |
| **Journal** | Journal |
| **Volume** | Volume |
| **Page(s)** | Page(s) |
| **Medline** | Medline |
| **Act\_Outliers** | For each publication, the mean and 95% or 99% confidence interval (CI) of the residual activity of TP53 mutants have been calculated.  **IN:** no specific problem  **95**: indicates that the set of TP53 mutants described in this publication is statistically different from the distribution of all reports and displays significant activity according to the 95% CI.  **99:** indicates that the set of TP53 mutants described in this publication is statistically different from the distribution of all reports and displays significant activity according to the 99% CI.  **NA**: not available, as only publications with more than 6 mutations have been analyzed.  Soussi T, Asselain B, Hamroun D, Kato S, Ishioka C, Claustres M, Beroud C (2006) Meta-analysis of the p53 mutation database for mutant p53 biological activity reveals a methodologic bias in mutation detection. Clin Cancer Res 12:62-69 |
| **PCA\_Outliers** | For publications with more than 5 mutations, multicriteria analysis was performed using data-driven methodologies. A score was used to filter for high confidence reports within the database. Publications that deviated from the median by >2 SD were tagged as outliers.  The value of the SD for each publication is displayed in “PCA\_Score”.  **In:** no problem SD<2.  **Out**: Outlier publications SD=>2.  **NA:** not available, as only publications with more than 6 mutations have been analyzed.  See comment\_8\_Publication for more info on each publication.  Edlund K, Larsson O, Ameur A, Bunikis I, Gyllensten U, Leroy B, Sundström M, Micke M, Botling J, Soussi T (2012) Data-driven unbiased curation of the TP53 tumor suppressor gene mutation database and validation by ultradeep sequencing of human tumors. Proc Natl Acad Sci U S A 2012 Jun 12;109(24):9551-6. |
| **PCA\_Score** | For publications with more than 5 mutations, multicriteria analysis was performed using data-driven methodologies. A score was used to filter for high confidence reports within the database. Publications that deviated from the median by >2 SD were tagged as outliers.  The value of the SD for each publication is displayed in this row.  Publications with values **greater than 2** should be considered to be suspicious.  Publications with values **greater than 5** are probably artefactual.  See comment\_8\_Publication for more info on each publication.  Edlund K, Larsson O, Ameur A, Bunikis I, Gyllensten U, Leroy B, Sundström M, Micke M, Botling J, Soussi T (2012) Data-driven unbiased curation of the TP53 tumor suppressor gene mutation database and validation by ultradeep sequencing of human tumors. Proc Natl Acad Sci U S A 2012 Jun 12;109(24):9551-6. |
| **Comment\_7\_Sequence** | Specific information about homopolymeric tracts in the TP53 gene |
| **Commrnt\_8\_Publication** | Specific information regarding this mutation and the publication. |
| **Comment\_9\_SNP** | Specific comments regarding the specificity of each SNP including novel SNP detected in new sequencing projects |
| **General Comment** | Specific information regarding this mutation and/or the publication. This field also includes important information regarding the status of this mutation in various cell lines. |
| **Pathogenicity** | We used this specific standard terminology for TP53 variants: ‘pathogenic’, ‘likely pathogenic’, ‘uncertain significance’ (VUS), ‘likely benign’.  A new TP53 specific algorithm was used to define TP53 variant pathogenicty (T Soussi et al. manuscript in preparation). |