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 variant database.**

|  |  |
| --- | --- |
| **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/> |
| **SNP\_ID\*** | The SNP database now includes several pathogenic variants of the TP53 gene  <http://www.ncbi.nlm.nih.gov/snp>  \* 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.. |
| **HG18\_Variant** | Mutation nomenclature according to HGVS standards using the genomic sequence as reference  Reference sequence: NC\_000017.9 for genome build NCBI36/hg18 |
| **HG19\_Variant** | Mutation nomenclature according to HGVS standards using the genomic sequence as reference  Reference sequence: NC\_000017.10 for genome build NCBI37/hg19 |
| **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** | 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> |
| **Transcript t1 MN\_000546.5** | *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* |
| **Transcript t2 NM\_001126112.2** | *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* |
| **Transcript t3 NM\_001126114.2** | *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* |
| **Transcript t4 NM\_001126113.2** | *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* |
| **Transcript t5 NM\_001126115.1** | *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* |
| **Transcript t6 NM\_001126116.1** | *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* |
| **Transcript t7 NM\_001126117.1** | *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* |
| **Transcript t8 NM\_001126118.1** | *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* |
| **Transcript t1 LRG\_321** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence LRG\_321t1* |
| **Transcript t2 LRG\_321** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence LRG\_321t2* |
| **Transcript t3 LRG\_321** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence LRG\_321t3* |
| **Transcript t4 LRG\_321** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence LRG\_321t4* |
| **Transcript t5 LRG\_321** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence LRG\_321t5* |
| **Transcript t6 LRG\_321** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence LRG\_321t6* |
| **Transcript t7 LRG\_321** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence LRG\_321t7* |
| **Transcript t8 LRG\_321** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG (position 1 refers to the A of the start ATG): reference sequence LRG\_321t8* |
| **TP53\_alpha NP\_000537.3** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_000537.3* |
| **TP53\_beta NP\_001119586.1** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001119586.1* |
| **TP53\_gamma NP\_001119585.1** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001119585.1* |
| **Delta40\_TP53\_alpha NP\_001119590.1** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263690.1* |
| **Delta 40\_TP53\_beta NP\_001263625.1** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263625.1* |
| **Delta 40\_TP53\_gamma NP\_001263624.1** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263624.1* |
| **Delta 133\_TP53\_alpha NP\_001119587.1** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001119587.1* |
| **Delta 133\_TP53\_beta NP\_001119588.1** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001119588.1* |
| **Delta 133\_TP53\_gamma NP\_001119589.1** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001119589.1* |
| **Delta160\_TP53\_alpha NP\_001263626.1** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263626.1* |
| **Delta160\_TP53\_beta NP\_001263627.1** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263627.1* |
| **Delta160\_TP53\_gamma NP\_001263628.1** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263628.1* |
| **Protein p1 TP53\_alpha** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence LRG\_321p1* |
| **Protein p3 TP53\_beta** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence LRG\_321p3* |
| **Protein p4 TP53\_gamma** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence LRG\_321p4* |
| **Protein p8 Delta40\_TP53\_alpha** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence LRG\_321p8* |
| **Protein p9 Delta 40\_TP53\_beta** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence LRG\_321p9* |
| **Protein p10 Delta 40\_TP53\_gamma** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence LRG\_321p10* |
| **Protein p5 Delta 133\_TP53\_alpha** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence LRG\_321p5* |
| **Protein p6 Delta 133\_TP53\_beta** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence 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* |
| **Protein p11 Delta160\_TP53\_alpha** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263626.1* |
| **Protein p12 Delta160\_TP53\_beta** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263627.1* |
| **Protein p13 Delta160\_TP53\_gamma** | *Mutation nomenclature and coordinates are described according to HGVS, NCBI and LRG: reference sequence NP\_001263628.1* |
| **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 |
| **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). |
| **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) |
| **Genome base coding** | Nucleotide at the start position of the mutation. |
| **Mutant\_Allele** | Mutant 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. |
| **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 |
| **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\_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. |
| **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 |
| **Substitution\_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 |
| **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 and all transversions. |
| **Mutational\_Event** | Mutational events  **G>C**: (G to C base substitution)  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.** |
| **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:** Intronic tandem mutation that occurs across a splice site (+1/+2 or -1/-2) |
| **Variant\_Classification** | Translational effect of the mutation  (Missense, Nonsense, Synonymous, Nonstop, In\_frame\_Del, Inframe\_Ins, Frameshift\_Del or Frameshift\_Ins).  These 8 items are identical to the entries used in MAF file  Four novel items are used in the TP53 mutation database: In\_frame\_Del\_Complex, Inframe\_Ins\_Complex, Frameshift\_Del\_Complex or Frameshift\_Ins\_Complex) for mutations that span one or more than one exon-intron site. |
| **Variant\_Type** | Variant type as defined in MAF file  **SNV:** Single Nucleotide Variant  **DNP:** Change in two consecutive bases (dinucleotide variant)  **TNP:** Change in three consecutive bases (tri-nucleotide variant)  **ONP:** Change in four or more consecutive bases (oligo-nucleotide variant)  **INS:** Insertion  **DEL:** Deletion |
| **Mutation type** | **B:** Single nucleotide variant  **D**: Deletion  **I**: insertion  **ID**: complex event that involves an insertion and a deletion. |
| **Variant\_comment** | Specific comment concerning the consequences of the mutation. |
| **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 |
| **Structure** | Structural motif of the TP53 protein according to the analysis described by Cho et al. (1994). |
| **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 |
| **Records\_Number** | Number of occurrences of the **cDNA\_variant** in the database. |
| **Leukaemia\_Lymphoma\_Freq** | Frequency of the variant (cDNA\_nomenclature) in haematological malignancies  e.g.: for variant c.524G>A, the entry will be 2.91 (216/7,403)   * Frequency of c.524G>A in haematological malignancies: 2.91 % * c.524G>A in haematological malignancies: 216 * Total number of haematological malignancies in the database: 7,403 |
| **Solid\_Tumour\_Freq** | Frequency of the variant (cDNA\_nomenclature) in solid tumours  e.g.: for variant c.524G>A, the entry will be 4.40 (3,087/70,153)   * Frequency of c.524G>A in Solid tumours: 4.40 % * c.524G>A in Solid tumours: 3,087 * Total number of Solid tumours in the database: 70,153 |
| **Tumour\_Freq** | 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 4.33 (3,148/72,829)   * Frequency of c.524G>A in tumours: 4.33 % * c.524G>A in tumours: 3,148 * Total number of tumours in the database: 72,829 |
| **Cell\_line\_Freq** | 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 3.77 (146/3,864)   * Frequency of c.524G>A in tumours: 3.77 % * c.524G>A in tumours: 146 * Total number of cell lines in the database: 3,864 |
| **Somatic\_Freq** | Frequency of the variant (cDNA\_nomenclature) found as a somatic event  e.g.: for variant c.524G>A, the entry will be 4.13 (3,346/78,639)   * Frequency of c.524G>A in tumours: 4.26 % * c.524G>A in tumours: 3,346 * Total number of cell lines in the database: 78,639 |
| **Germline\_Freq** | Frequency of the variant (cDNA\_nomenclature) found as a germline event  e.g.: for variant c.524G>A, the entry will be 4.47 (52/1,169)   * Frequency of c.524G>A in tumours: 4.47 % * c.524G>A in tumours: 52 * Total number of germline in the database: 1,169   Note of caution 1: 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.  Note of caution 2: the germline Brazil mutation p.R337H has been shown to be a founder mutation and has only been included once in the database. |
| **Mutant activities (info)** | 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\_Act** | Residual transcriptional activity of mutant p53 on the WAF1 promoter (raw data from kato et al.) |
| **MDM2\_Act** | Residual transcriptional activity of mutant p53 on the MDM2 promoter (raw data from kato et al.) |
| **BAX\_Act** | Residual transcriptional activity of mutant p53 on the BAX promoter (raw data from kato et al.) |
| **14\_3\_3\_\_Act** | Residual transcriptional activity of mutant p53 on the 14-3-3-s promoter (raw data from kato et al.) |
| **AIP\_Act** | Residual transcriptional activity of mutant p53 on the AIP promoter (raw data from kato et al.) |
| **GADD45\_Act** | Residual transcriptional activity of mutant p53 on the GADD45 promoter (raw data from kato et al.) |
| **NOXA\_Act** | Residual transcriptional activity of mutant p53 on the NOXA promoter (raw data from kato et al.) |
| **p53R2\_Act** | Residual transcriptional activity of mutant p53 on the p52R2 promoter (raw data from kato et al.) |
| **WAF1\_percent** | Residual transcriptional activity of mutant p53 on the WAF1 promoter (% compared to wild-type p53). |
| **MDM2\_percent** | Residual transcriptional activity of mutant p53 on the MDM2 promoter (% compared to wild-type p53). |
| **BAX\_percent** | Residual transcriptional activity of mutant p53 on the BAX promoter (% compared to wild-type p53). |
| **14\_3\_3\_\_percent** | Residual transcriptional activity of mutant p53 on the 14-3-3-s promoter (% compared to wild-type p53). |
| **AIP\_percent** | Residual transcriptional activity of mutant p53 on the AIP promoter (% compared to wild-type p53). |
| **GADD45\_percent** | Residual transcriptional activity of mutant p53 on the GADD45 promoter (% compared to wild-type p53). |
| **NOXA\_percent** | Residual transcriptional activity of mutant p53 on the NOXA promoter (% compared to wild-type p53). |
| **p53R2\_percent** | Residual transcriptional activity of mutant p53 on the p52R2 promoter (% compared to wild-type p53). |
| **Sift\_Prediction** | Predictive value using Sift |
| **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. |
| **PolyPhen-2 (info)** | **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/  **Polyphen-2 : prediction\_HUMDiv versus prediction\_HUMVar**  Two pairs of datasets were used to train and test PolyPhen-2 prediction models. The first pair, **HumDiv**, was compiled from all damaging alleles with known effects on the molecular function causing human Mendelian diseases, present in the UniProtKB database, together with differences between human proteins and their closely related mammalian homologs, assumed to be non-damaging. The second pair, **HumVar**, consisted of all human disease-causing mutations from UniProtKB, together with common human nsSNPs (MAF>1%) without annotated involvement in disease, which were treated as non-damaging.  The user can choose between HumDiv- and HumVar-trained PolyPhen-2 models. Diagnostics of Mendelian diseases requires distinguishing mutations with drastic effects from all the remaining human variation, including abundant mildly deleterious alleles. Thus, HumVar-trained model should be used for this task. In contrast, HumDiv-trained model should be used for evaluating rare alleles at loci potentially involved in complex phenotypes, dense mapping of regions identified by genome-wide association studies, and analysis of natural selection from sequence data, where even mildly deleterious alleles must be treated as damaging.  More info on the polyphen web site : <http://genetics.bwh.harvard.edu/pph2/dokuwiki/overview> |
| **Polyphen-2\_HumVar** | See above for more info |
| **Polyphen-2\_HumDiv** | See above for more info; this prediction is less accurate than HumVar for TP53 |
| **Mutassessor\_prediction:** | Functional impact of a variant : predicted functional (high, medium), predicted non-functional (low, neutral). |
| **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). |
| **Provean\_prediction** | Prediction - deleterious or neutral (using default cutoff at -2.5) |
| **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. T  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). |
| **Condel** | Predicted functional effect using Condel algorithm  <http://bg.upf.edu/fannsdb/> |
| **Condel\_Score** | 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 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). |
| **MutPred\_Splice\_Prediction\_Label** | See above and http://mutdb.org/mutpredsplice/about.htm |
| **MutPred\_Splice\_Confident\_Hypotheses** | See above and http://mutdb.org/mutpredsplice/about.htm |
| **Comment\_1\_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\_2\_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\_3\_Isoforms** | Number of TP53 isoforms targeted by the mutation |
| **Comment\_4\_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.  This prediction is shown for missense variant only and less accurate than the pathogenocity define in comment 8 |
| **Comment\_5\_Outliers** | Indicates whether or not the variant is associated with outlier publications.  Rare mutants only found in outlier studies should be considered to be suspicious. |
| **Comment\_6\_Splicing** | Indicates whether or not the mutation could impair TP53 splicing.  All TP53 gene substitutions have been analysed 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. |
| **Comment\_7\_Sequence** | Indicates the presence of homopolymeric tracts at the position of the mutation. |
| **Comment\_10\_population** | Population data and frequency of the SNP in various databases. |
| **Pathogenicity** | We used this specific standard terminology for TP53 variants: ‘pathogenic’, ‘likely pathogenic’, ‘uncertain significance’ (VUS), ‘likely benign’.  Specific algorithm were used to define TP53 variant pathogenicty (T Soussi et al. manuscript in preparation). |
| **Final comment** | Comment summary. |