Human Mutation



MUT-TP53 2.0: A novel versatile matrix for statistical analysis of TP53 mutations in human cancer.

Journal:	Human Mutation
Manuscript ID:	humu-2010-0129.R1
Wiley - Manuscript type:	Informatics
Date Submitted by the Author:	n/a
Complete List of Authors:	Soussi, Thierry; Karolinska Institutet, Dept. of Oncology-Pathology Hamroun, Dalil; Institut Universitaire de Recherche Clinique, CNRS UPR 1142 Hjortsberg, Linn; Karolinska Institutet, Dept. of Oncology-Pathology Rubio Nevado, Jean Michel; Karolinska Institutet, Dept. of Oncology-Pathology Fournier, Jean Louis; Karolinska Institutet, Dept. of Oncology-Pathology Béroud, Christophe; INSERM U827, Laboratoire de Génétique Moléculaire et Chromosomique
Key Words:	database, curation, TP53, mutation, cancer



MUT-TP53 2.0: A novel versatile matrix for statistical analysis of TP53 mutations in human cancer.

Thierry Soussi,^{1,2*} Dalil Hamroun³, Linn Hjortsberg,¹ Jean Michel Rubio-Nevado,² Jean Louis Fournier², Christophe Béroud,³

Address for correspondence and reprints:

Thierry Soussi

thierry.soussi@ki.se

Keywords: TP53, mutation, database, curation, cancer

Databases

http://p53.free.fr http://www.umd.be:2072/

TP53 - OMIM: 191170; GDB: 120445; Genbank: X54156

Acknowledgements

This work is supported by Cancerföreningen i Stockholm, Cancerfonden and the Swedish Research Council (VR)

¹ Karolinska Institute Dept. of Oncology-Pathology Cancer Center Karolinska (CCK), SE-171 76 Stockholm, Sweden

² Université Pierre et Marie Curie-Paris 6, 75005 Paris, France

³ Laboratoire de Génétique Moléculaire et Chromosomique, Institut Universitaire de Recherche Clinique et CHU, CNRS UPR 1142, 641, avenue du Doyen Gaston Giraud, 34093 Montpellier Cedex 5

Abstract

Analysis of the literature reporting p53 mutations shows that 8% of report display typographical mistakes with a notable increase in recent years. These errors are sometimes isolated, but, in some cases, they concern several or even all mutations described in a single article. Furthermore, some works report unusual profile of p53 mutations whose accuracy is difficult to assess. In order to handle these problems we have developed MUT-TP53 2.0, an accurate and powerful tool that will automatically handle p53 mutations and generate tables ready for publication that will lower the risk of typographical errors. Furthermore, using functional and statistical information issued from the UMD p53 database, it's allow to assess the biological activity and the likelihood of every p53 mutant.

Introduction

Identification of novel genes associated with tumour development will provide new insight into cancer biology, and should also identify whether some of these mutated genes could be effective targets for anticancer drug development. For this purpose, partial and whole cancer genome sequencing has been initiated, but has led to the discovery of an unexpected landscape of in vivo somatic mutations with 10 to 20,000 base substitutions per genome [Stratton et al., 2009, Strausberg and Simpson, 2010]. The majority of these variations are somatic passenger mutations (or hitchhiking mutations) that have no active role in cancer progression and are co-selected by the driver mutations which are the true driving force for cell transformation [Chanock and Thomas, 2007]. Passenger mutations can be found in coding or non-coding regions of any genes and the distinction of these genes from driving mutations is a difficult but necessary task to obtain an accurate picture of the cancer genome. Several statistical approaches have been developed to resolve this problem such as comparing the observed to expected ratios of synonymous:non-synonymous variants. Alternatively, various bioinformatic methods are used to provide an indication of whether an amino acid substitution is likely to damage protein function on the basis of conservation through species or whether or not the amino acid change is conservative [Ng and Henikoff, 2001].

Reporting, storing, classifying and analyzing these mutations constitute a major challenge [Horaitis and Cotton, 2004]. For a long time, locus-specific databases have been developed for this purpose. Although each LSDB has been developed for a single gene, they are highly accurate and provide information that can be exploited for large-scale analysis. They often include structural, functional or evolutionary data that allow easy distinction between passenger and driving mutations. TP53 mutation (TP53; MIM# 191170) databases are a paradigm, as they constitute the largest collection of somatic mutations (30,000 mutations from 29,000 patients) for a single gene.

A review of the literature reporting TP53 mutations shows that 8% of reports comprise typographical errors, with a marked increase over recent years (T Soussi, unpublished observations). These errors are sometimes isolated, but, in some cases, they concern several or even all mutations described in a single article. Furthermore, some papers report an unusual profile of p53 mutations. In 2006, we published a meta-analysis of 2,000 reports describing TP53 mutations and revealed that these dubious reports were associated with methodological bias for p53 analysis[Soussi et al., 2006]. To resolve these problems, we have developed, MUT-TP53, an accurate and powerful tool that automatically processes p53 mutations and generate tables ready for publication that will decrease the risk of typographical errors [Soussi et al., 2006]. Furthermore, using functional and statistical information derived from the UMD p53 database, this matrix could be used to assess the biological activity and likelihood of each TP53 mutant. Although this tool is used by numerous laboratories, reports of unusual patterns of p53 mutations are still published, leading to controversial discussions [Campbell et al., 2008, Roukos, 2008, Soussi Zander and Soussi, 2008, Zalcman et al., 2008]. In order to identify these

problems before publication, we present MUT-TP53 2.0, an extended version of our previous program comprising novel tools that allow authors, reviewers, editors and curators to: i) manage p53 mutation sequences based on the genetic code and wild-type p53 sequences; ii) check the frequency and activity of p53 mutations; iii) generate a p53 mutation table ready for publication that will lower the risk of typographical errors; iv) compare the profile of p53 inactivation with other publications and v) perform statistical analysis of p53 loss of function.

Results and Discussion

The latest version of the UMD p53 database (2010) contains 30,000 mutations from 29,200 patients (several patients have multiple mutations). As the p53 gene has several mutation hot spots, similar mutations are found in different patients and the true number of different p53 mutant is 2,300 including missense and non-sense mutations (whether or not they change the residue) and frameshift mutations. Among these mutations, only 1,439 mutant p53 have a single amino acid change.

These 1,439 mutants have been divided into 8 categories according to their frequency in the database (Figure 1A and 1B). The residual activity of mutants very frequently found in the TP53 database (categories 161+, 81-160, 41-80, and 21-40) is usually low with only a few mutants with higher activity for the 21-40 category (Figure 1A and 1B). These categories contain all of the hot spot mutants of the p53 gene and have been shown to be inactive by numerous studies. For categories 4-8, 2-3 or mutants found only once, the scatter is very heterogeneous, ranging from 0 to 160% compared to wt p53. Non-parametric statistical analysis using a Mann-Whitney test did not reveal any statistical difference between the four categories 161+, 81-160, 41-80, and 21-40. However, comparison of each of these categories with each of the low frequency categories showed a highly significant difference (p<0.0001, Figure 1A). As similar distribution and statistical differences was observed using TP53 data activity on other TP53 promoters or using different categories to classify TP53 mutation frequency (Supp. Figure S1). There is a clearly demonstrated inverse correlation between the frequency of TP53 mutants and their activity. Fifty five percent of the rare mutants found only once have an activity greater than 50% compared to wt TP53. This picture is similar to our previous analysis performed on 1,100 TP53 mutants [Soussi et al., 2005]. The presence of a few rare active p53 mutants in a large study is not a problem, as they can be rare passenger mutations. In contrast, several studies have described unusually high numbers of rare mutants that can represent 80% of all mutants. Some of these studies have been shown to be associated with methodologic biases and artefactual data. One of the goals of MUT-TP53 2.0 is to detect these rare p53 mutants in order to warn the user about an unusual mutation pattern.

In order to minimize typographical errors, another problem encountered in many publications, MUT-TP53 2.0 automates most of the data entry procedures and keeps the number of manual steps to a minimum. The matrix includes the sequence of the human wt TP53 cDNA and a genetic code for translation. The user only needs to enter 3 types of information: i) case number, ii) position of the mutation (at the codon level) and iii) mutant sequence. The matrix then automatically displays a table comprising all information concerning wt and mutant codon and amino acids (1- and 3-letter codes) (Figure 2). This information is linked to three types of comments and guidelines for the user about the likelihood of each mutation (Figure 2). Compared to version 1.0, MUT-TP53 2.0 contains numerous novel features to cover each single, double or triple nucleotide substitution at each codon of the p53 gene (25, 152 possibilities). Although the majority of p53 mutations are single substitutions, tandem mutations can be common in skin cancer, while triple substitutions are very rare. The nomenclature of the p53 mutants used in the matrix is based on the guideline described by Den Dunnen and Antonarakis [den Dunnen and Antonarakis, 2000]. A typical table generated by MUT-TP53 2.0 is shown in Figure 2 with various types of mutant. More examples are available in the Help document that can be downloaded with the program (http://p53.free.fr). In sample 1, mutant c.524G>A (p.R175H) is an inactive hot spot mutant described 1,233 times (frequency column) with a residual activity of 12% compared to wt p53 (activity column). Comment 1 concerns the

frequency of the mutant in the database and comment 2 concerns the loss of function of this mutant (see material and methods). Comment 3 presents final advice concerning whether or not the user can trust this mutant. This final comment is based on mutant loss of activity and frequency (see material and methods). In sample 2, the mutation is less frequent but associated with loss of p53 activity. In sample 3, c.215C>G (p.P72P), comment 3 warns the user that this not a mutation but an exonic polymorphism of the human p53 gene. Although this polymorphism is well known, other polymorphisms, such as P36P and R213R, are less known and are often described as mutations. All exonic SNPs of the p53 gene have been included in MUT-P53 v2.0. Although a somatic mutation similar to a natural polymorphism cannot be formally excluded. they are hardly unlikely and difficult to detect if matched normal tissue is not analyzed. Sample 4 is a simple warning in the case of typographical errors where wt and mutant codon are identical. The mutation in example 5 (c.375G>A, p.T125T) is also a good example of a common mistake. For a long time, this mutation was (and still is) described as a neutral somatic mutation, as it does not change the amino acid. Nonetheless, this mutation localized in the last codon of exon 4, has been shown to lead to aberrant splicing and p53 inactivation, a feature now described in comment 2 [Varley et al., 2001]. Exonic mutations leading to aberrant splicing are guite common but difficult to detect if only DNA bases analyses are performed. All codons in the vicinity of an exon (3' and 5') have been documented in MUT-TP53 V2.0 and the user is warned accordingly (sample 6, Figure 2). This feature has allowed the detection of splicing defects in several cell lines. Mutant c.1010G>A (p.R337H) in sample 7 is very unusual and was detected as a germline mutation in Brazilian families with children prone to develop adrenocortical tumors [Ribeiro et al., 2001]. A few tumors with a similar somatic mutation have also been described. Although classical transcriptional assays suggest that this mutant behaves like wt p53, structural studies show that this mutant displays an abnormal conformation at low pH [DiGiammarino et al., 2002]. Whether or not this abnormal conformation is associated with a defect in p53 function in vivo has not been clearly established, but, as shown in Figure 2, this mutant has been flagged. The mutant in sample 8 (c. [748C>T;749C>T], p.P250F), is a tandem mutation only described in skin cancer and which is extremely rare in internal tumors. This double-base substitution found at dipyrimidine sites is associated with UV exposure [Brash, 1997]. As shown in comment 3, a warning is displayed to the user concerning this feature, as a large number of tandem mutations would be very unlikely in internal tumors. Mutations 9 (c.[343C>A:345T>G]. p.H115K. two noncontiquous substitutions) 10 (c.[688A>T;689C>G;690C>G], p.T230W, 3 substitutions in the same codon) are extremely rare in the p53 mutation database and a warning is displayed for the user. This last feature is not specific to p53 mutations, as a recent release of cancer genome sequences shows that a single nucleotide substitution is the most frequent nucleotide substitution. The table shown in Figure 2 should help the user to pinpoint uncommon p53 mutants and perform verification to validate the mutation.

The second tool included in MUT-TP53 V2.0 is entirely new and allows the user to perform statistical comparison of his/her data set with other publications. Since 2003, the UMD p53 mutation database includes functional information about the majority of p53 missense mutants (see also material and methods). This quantitative data has been extremely useful to classify and analyze p53 mutations. The range of p53 loss of function of all p53 mutants for each publication can be displayed by calculating the mean and 95% confidence interval (CI) of the residual activity of mutant p53 (Figure 3). The analysis shows that, for more than 90% of publications, the mean activity was situated between -1 and -1.2 [Kato et al., 2003]. This value corresponds to a residual transcriptional activity of about 10% compared to wild-type p53. The small range of the 95% CI indicates that the majority of mutant p53 proteins behave in a similar way. Figure 3A shows typical results for lung, colorectal and breast carcinomas, three types of cancer that have been extensively analyzed for p53 mutations. For each type of cancer, the five publications reporting the highest number of p53 mutants are shown and one publication situated outside of the range of other studies (Figure 3) is also shown. The reason why these out-of-range studies display an unusual number of mutant p53 proteins that retain wild-type activity has already been extensively discussed [Soussi et al., 2006]. The two out-of-range

studies in lung and colorectal cancer have been clearly shown to be artefactual and they include unusual mutations in other genes as well. The study in breast cancer described an unusual number of clustered mutations at positions never previously described and a large number of mutations that do not change the amino acid sequence. This is not a real problem, as this highly controversial manuscript was published in NEJM and raised the important question of the existence of genetic alterations in stroma cells from breast cancer patients, a highly debated field [Campbell et al., 2008, Roukos, 2008, Soussi Zander and Soussi, 2008, Zalcman et al., 2008].

This statistical analysis is now available to users of MUT-TP53 V2.0. It is performed automatically and displayed in the form of a table and a graphic and can be compared to other publications in the database (Figure 3B and supplementary material).

It should be stressed that an out-of-range finding should not be considered to be definitive and formal proof of a dubious study, but indicate the need for careful review of the data. If confirmed, this finding should be discussed in the publication, as it may represent a novel finding on p53 loss of activity for a particular set of p53 mutations in a specific type of cancer. For example, the c.1010G>A (p.R337H) mutant discussed above does not display loss of transcriptional activity and could have been missed if structural studies were not performed. Its association with adrenocortical tumor also suggests a specific link of this mutant with a specific cellular environment.

Biochemical analyses have shown that mutant p53 proteins can be heterogeneous in terms of loss of DNA binding activity, transactivation or other activities [Soussi and Lozano, 2005]. The DNA binding site recognized by p53 is highly degenerated and wild-type or mutant p53 have variable affinities for the various biological sites [Resnick and Inga, 2003]. Mutant p53 proteins also exhibit a dominant negative effect via inactivation of the function of wild-type p53. This characteristic increases the significance of a single mutant p53 allele. While carcinogenesis requires the loss of both alleles of most tumor suppressor genes, mutation of one allele of p53 can result in total loss of function. While the dominant negative effect clearly occurs in cancer models, the mechanism by which it occurs has not been fully elucidated. Whether or not tumors expressing mutant p53 with this dominant negative characteristic are more aggressive is also still under investigation. MUT-TP53 V2.0 includes novel functional data derived from an unpublished database including more than 100,000 entries on mutant p53 in. Residual DNA binding activity (in vitro and in vivo), growth arrest, apoptosis, repressor and dominant negative activities can now be displayed (supplementary figures 2). Not only will this information be useful to define the degree of loss of activity of a mutant p53, but it should also help the user to define various classes of mutant p53 in order to perform more accurate clinical analysis. MUT-TP53 V2.0 with full documentation is available free of charge and can be downloaded from our website (htpp://p53.free.fr).

Sequencing of tumor DNA for the detection of p53 mutations (as well as other genes) and publication of this information appear to be currently moving in opposite directions. On the one hand, DNA sequencing is now available at very low cost and high throughput, allowing screening of a large number of patients and, on the other hand, the lack of space for publication now means that more than 50% of publications describing p53 mutations no longer provide the mutational data, leading to a decreased number of entries over recent years. An automated spreadsheet such as MUT-TP53 V2.0 can be used not only as a verification tool to generate accurate data, but also as a support that can be easily used by the UMD database software to import automatically data and more efficiently update the p53 mutation database.

Material and methods

MUT-TP53 V2.0 was developed using Microsoft *Excel*. The spreadsheet is available for both Windows and OS X platforms. Two different databases were used for development: First, the latest version of the UMD p53 mutation database (June 2010, 31,000 p53 mutations, http://p53.free.fr). It contains data on mutation frequency, mutation identity and transcriptional activity in yeast that are used to generate the result page in the spreadsheet (figure 2). The

second database is OPMA, a novel unpublished database that contains mutant p53 activity (T Soussi et al., manuscript in preparation). Briefly, OPMA has been developed by mining the literature focusing on mutant p53 loss of activity such as loss of transactivation in mammalian cells, dominant negative activity, in vitro and in vivo DNA binding, growth arrest and apoptosis. This data are used to generate the mutant page in the spreadsheet (Supplementary figure 2 and table 2).

The algorithm used for the development of MUT-TP53 V2.0 and the cut-offs used for the various comments are fully described in a supplementary document available with this manuscript. Mutant p53 activity has been described in detail in a previous report [Kato et al., 2003]. Briefly, 2,314 haploid yeast transformants containing p53 mutations and a GFP-reporter plasmid have been constructed. Mutant p53 activity was tested by measuring the fluorescent intensity of GFP that is controlled by the WAF1 promoter sequence of the plasmid after 3 days of growth at 37°C. The activity of the yeast without p53 or with wt p53 was -1.58 and 2.03, respectively. The activity of the majority of p53 mutants was situated between these two values.

An approach similar to that used for meta-analyses comparing clinical trials was used for data analysis and presentation of the statistical study [Soussi et al., 2006]. For each publication, the mean and 95% CI of p53 activity of each mutant were displayed graphically. The reference value corresponds to the mean and 95% CI of all studies for the specific cancer. Although the mean value of the entire database can be used as the reference value, the use of an individual reference value for each cancer type would more closely reflect the heterogeneous etiology and pattern of p53 mutations in various cancers. This procedure has been fully described in previous publications [Soussi et al., 2006].

Statistical analyses were performed with PRISM software (GraphPad Software Inc) on a Mac OS X platform.

MUT-TP53 V2.0 will be available for download at http://p53.free.fr.

For reviewers, it is available here: p53.free.fr/MUT-TP53_II.zip

Figure 1: Activity of mutant *TP53* according to their frequency in various subsets of the database. Mutant *TP53* were classified into 8 categories according to their frequencies. A: Box and whisker plots show the upper and lower quartiles and range (box), median value (horizontal line inside the box) and full range distribution (whisker line); analysis was performed for the 1,439 mutants found in tumors. P values listed above each bar refer to comparison to the 161+ category. The Mann–Whitney U test was used to evaluate statistical significance. Similar results were observed with the activity of seven other promoters regulated by TP53 (Supplementary figure 1). Changing the number of categories leads to similar results (Supplementary figure 1): B Each dot corresponds to the activity of a single mutant and its X position in the column is random. Black lines correspond to the mean value. N.S. Not significant; ***: p<0.0001. Black triangles corresponds to 0% (bottom) and 100% (top) TP53 activity.

Figure 2: Output example: Ten mutant p53 showing the various features of MUT-MAT V2.0 were analyzed. Mutants in sample 1 (p.R175H), sample 2 (p.K132E), sample 5 (p.T125T), sample 6 (p.S260S), sample 7 (p.R337H) and sample 8 (p.N200P) have been described in the tumor database. In sample 3, the spreadsheet identifies the exonic SNP in exon 4 at codon 72. In sample 4, the matrix identifies a typographical error, as wt and mutant codon are identical. Mutants in sample 9 and 10 correspond to the substitution of two non-contiguous residues or 3 residues. These mutational events are very unusual in the p53 gene and are flagged accordingly. See text for more information. Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence

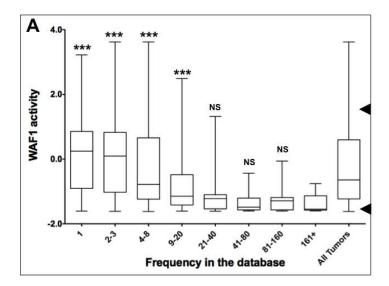
(GDB: 120445).

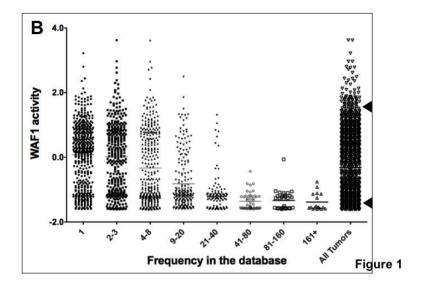
Figure 3. A: Meta-analysis of p53 loss of function. Dot and bars: mean and 95% CI of mean p53 activity as measured by transactivation with the WAF1 promoter. The y-axis corresponds to p53 transactivation activity, with a value of -1.58 for the negative control and a value of 2.03 for 100% of wt activity. The publication code is indicated on the x-axis: the first letter corresponds to the cancer type and the second number is an anonymous ID for the publication. Six publications were analyzed for lung (NSCLC), colorectal (CRC) and breast cancer. Five publications displayed a homogenous distribution of p53 loss of activity (green data), while one study reported out-of-range results (red data). **B: Output generated by the spreadsheet.** Data from the user are displayed at the far right of the graph (red arrow). Statistical analysis for various cancers is shown in the graph.

Publication or cancer type code is indicated on the x-axis: CRC: colorectal cancer; GLB: glioblastoma; HNSCC: head and neck squamous cell carcinoma; HCC: hepatocellular carcinoma; NSCLC: non-small cell lung carcinoma: SCLC: small cell lung carcinoma; Eso SCC: esophageal squamous cell carcinoma; Ast: astrocytoma. The first letters correspond to the cancer type and the second number is an anonymous ID for the publication. All cancer: analysis of the entire UMD p53 database

References

- Brash DE. 1997. Sunlight and the onset of skin cancer. Trends Genet 13:410-414.
- Campbell IG, Qiu W, Polyak K, Haviv I. 2008. Breast-cancer stromal cells with TP53 mutations. N Engl J Med 358:1634-5; author reply 1636.
- Chanock SJ, Thomas G. 2007. The devil is in the DNA. Nat Genet 39:283-284.
- den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15:7-12.
- DiGiammarino EL, Lee AS, Cadwell C, Zhang W, Bothner B, Ribeiro RC, Zambetti G, Kriwacki RW. 2002. A novel mechanism of tumorigenesis involving pH-dependent destabilization of a mutant p53 tetramer. Nat Struct Biol 9:12-16.
- Horaitis O, Cotton RG. 2004. The challenge of documenting mutation across the genome: the human genome variation society approach. Hum Mutat 23:447-452.
- 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.
- Ng PC, Henikoff S. 2001. Predicting deleterious amino acid substitutions. Genome Res 11:863-874.
- Resnick MA, Inga A. 2003. Functional mutants of the sequence-specific transcription factor p53 and implications for master genes of diversity. Proc Natl Acad Sci U S A 100:9934-9939.
- Ribeiro RC, Sandrini F, Figueiredo B, Zambetti GP, Michalkiewicz E, Lafferty AR, DeLacerda L, Rabin M, Cadwell C, Sampaio G, Cat I, Stratakis CA, Sandrini R. 2001. An inherited p53 mutation that contributes in a tissue-specific manner to pediatric adrenal cortical carcinoma. Proc Natl Acad Sci U S A 98:9330-9335.
- Roukos DH. 2008. Breast-cancer stromal cells with TP53 mutations. N Engl J Med 358:1636; author reply 1636.
- Soussi Zander C, Soussi T. 2008. Breast-cancer stromal cells with TP53 mutations. N Engl J Med 358:1635; author reply 1636.
- Soussi T, Asselain B, Hamroun D, Kato S, Ishioka C, Claustres M, Beroud C. 2006. Metaanalysis of the p53 mutation database for mutant p53 biological activity reveals a methodologic bias in mutation detection. Clin Cancer Res 12:62-69.
- Soussi T, Kato S, Levy PP, Ishioka C. 2005. Reassessment of the TP53 mutation database in human disease by data mining with a library of TP53 missense mutations. Hum Mutat 25:6-17.
- Soussi T, Lozano G. 2005. p53 mutation heterogeneity in cancer. Biochem Biophys Res Commun 331:834-842.
- Soussi T, Rubio-Nevado JM, Ishioka C. 2006. MUT-TP53: a versatile matrix for TP53 mutation verification and publication. Hum Mutat 27:1151-1154.
- Stratton MR, Campbell PJ, Futreal PA. 2009. The cancer genome. Nature 458:719-724.
- Strausberg RL, Simpson AJ. 2010. Whole-genome cancer analysis as an approach to deeper understanding of tumour biology. Br J Cancer 102:243-248.
- Varley JM, Attwooll C, White G, McGown G, Thorncroft M, Kelsey AM, Greaves M, Boyle J, Birch JM. 2001. Characterization of germline TP53 splicing mutations and their genetic and functional analysis. Oncogene 20:2647-2654.
- Zalcman G, Bergot E, Hainaut P. 2008. Breast-cancer stromal cells with TP53 mutations. N Engl J Med 358:1635-6; author reply 1636.



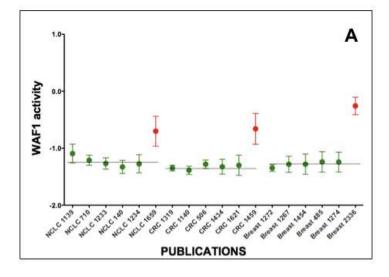


264x382mm (72 x 72 DPI)

Spie	Pos	WT	MUT	p.Mutant (1)	c.Mutent	Frequency	Activity	Comments 1	Commerce 2	Comments 3
1	175	CGC	CAC	p.R175H	c.524G>A	1233	12,41	This is a hot spot mutant	This mutant is inactive	No Problem
2	132	AAG	GAG	p.K132E	c.394A>G	28	0,56	This mutant is frequent	This mutant is inactive	No Problem
3	72	ccc	cgc	p.P72R	c.215C>G	0	×		1 222 1	This is not a mutation I This change has been reported as a natural polymorphism o the p53 gene
4	273	CGT	CGT	p.R273R	***	4	×	***	***	Wt and Mutant codon are identical : this is not a mutation !!
5	125	ACG	ACA	p.T125T	c.375G>A	14	×	This mutant is not frequent	This mutation does not change the amino acid residue but it is known to lead to abberant splicing.	No Problem
6	260	тсс	TCA	p.S260S	c.780C>A	:0:3	×	This mutant is not frequent	This mutation does not change the amino acid residue. It can change splicing, translation or RNA stability.	This codon is at the vicinity of en exon extremity, Check RNA if possible.
7.	337	ccc	CAC	p.R337H	c.1010G>A	87	×	This mutant is very rare	The stability of this mutant is highly sensitive to pH	Germline alteration R337H is associated with children from southern Brazil with adrenocortical tumours
8	250	ccc	TTC	p.P250F	c.[748C>T;749C>T]	6	ND	This mutant is not frequent	The remaining activity for this mutant p53 is unknown	Unusual mutational event except in skin cancer: check the trace or the autoradiogram
9	115	CAT	AAG	p.H115K	E.[343C>A;345T>G]	0	ND	This mutant has never been described so far	The remaining activity for this mutant p53 is unknown	This is an unusual mutational event: check the trace or the autoradiogram
10	230	ACC	TGG	p.T230W	c.[688A>T;689C>G;690C>G]	0	ND	This mutant has never been described so far	The remaining activity for this mutant p53 is unknown	This is a very unusual mutational event: check the trace or the autoradiogram

Figure 2

275x190mm (72 x 72 DPI)



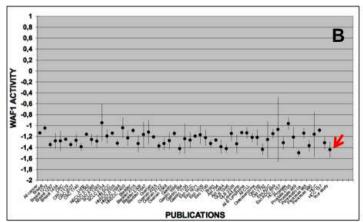


Figure 3

190x275mm (72 x 72 DPI)

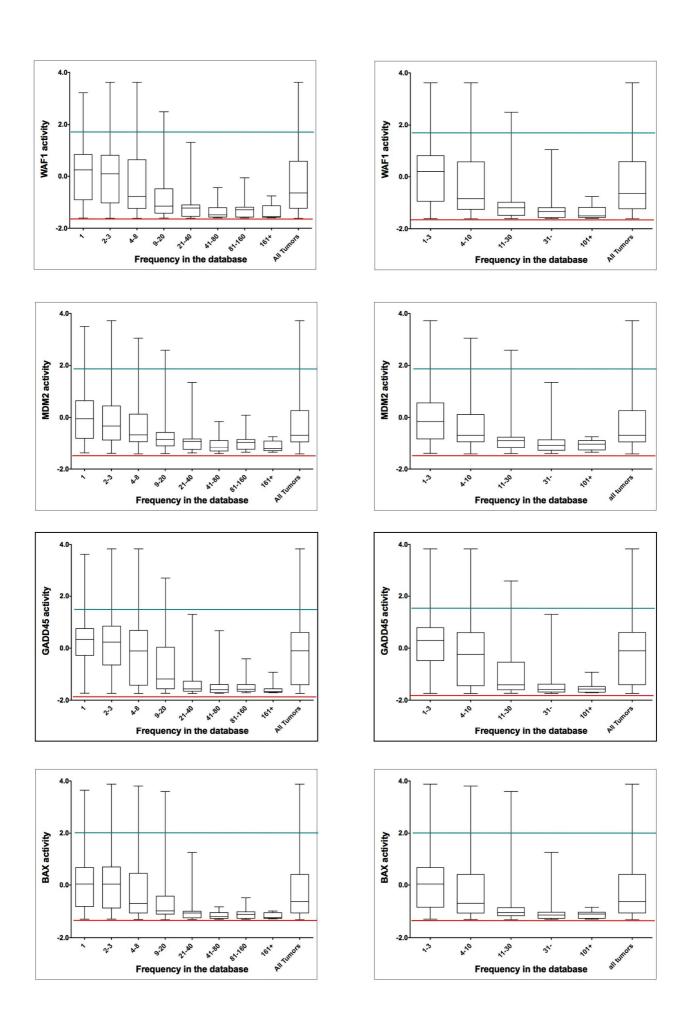
Supp. Figure S1: Activity of mutant *TP53* according to their frequency in various subsets of the database. Mutant *TP53* were classified into 8 (left) or 5 (right) categories according to their frequencies in the database. Box and whisker plots show the upper and lower quartiles and range (box), median value (horizontal line inside the box) and full range distribution (whisker line); analysis was performed for the 1,439 mutants found in tumors. P values listed above each bar refer to comparison to the 161+ category. The Mann–Whitney U test was used to evaluate statistical significance. Similar results were observed with the activity of WAF1 (Growth arrest), GADD45 (DNA repair), MDM2 (p53 reglation) or BAX (apoptosis). No difference are observed in the number of cateogries are changed. N.S. Not significant; ***: p<0.0001. Red and green bars correspond to 0 and 100% of p53 activity respectively

Supp. Figure S2: algorithm used for the development of MUT-TP53 V2.0

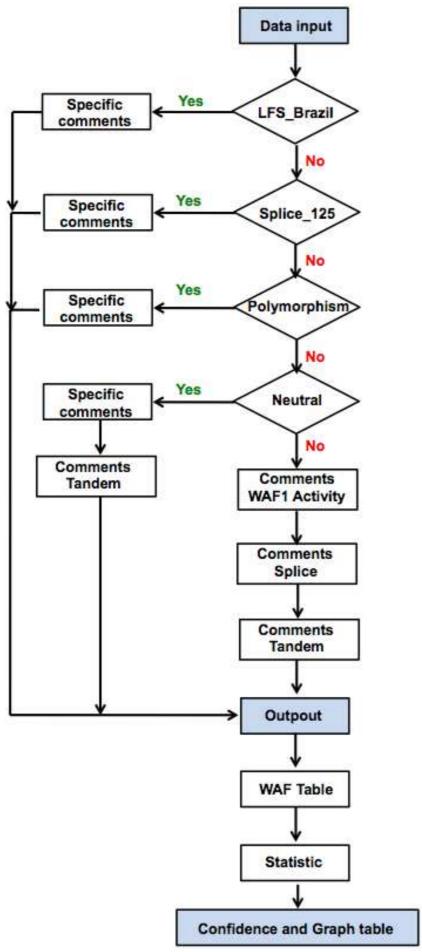
Supp. Figure S3: mutant p53 loss of activities. The OPMA database contains mutant p53 loss of function for different p53 activities such as transactivation (in yeast or in mammalian cells), dominant negative activity, repression of transcription and DNA binding (*in vitro* and *in vivo* when available, T Soussi et al., manuscript in preparation, see also Supp. Table S2). A subset of this information for each mutant p53 at position 175 is shown in the figure. For mutant p.R175H, a hot spot mutant found in human cancer, most studies in the literature show that this mutant is devoid of any transactivation activity, whereas only a few show opposite results. For DNA binding activity, the trend is similar. Mutant p.R175C, an infrequent mutant, displays an opposite behavior, as most studies found that this mutant does not display any loss of activity.

Supp. Table S1: cut-off used for data analysis. For comment 1, associated with the frequency of mutant p53 in the database, 5 categories have been defined and a specific comment has been specifically assigned to each category. For comment 2, associated with the residual activity of mutant p53, 5 categories have been defined. We assumed that nonsense mutations had no activity. Comment 3, related to the confidence of the mutation, is dependent on both comments 1 and 2.

Supp. table S2: : data included in the OPMA database.



Supp. Figure S1



Supp. Figure S2

Mutant	Freq	Transactivation	Dominant negative activity for transactivation	Repressor activity	In vitro DNA binding	In vivo DNA bindin
•	•	₹ Too few Data	No Data	No Data	↑ Too few Data	No Data
R175A	0	1 Wt Wild type	Too few Data	Too few Data	1 Mut No Data	No Data
R175C	29	35 Wt / 5 Mut Mutant	1 Wt Too few Data	5 Wt Too few Data	No Data	No Data
R175D	1	7 Mut	1 Mut	3 Wt / 2 Mut		and the same of the
R175F	0	Mutant 12 Mut	Too few Data 5 Mut	Too few Data 2 Mut	No Data	No Data
R175G	28	Mutant 15 Wt / 27 Mut	No Data	Too few Data 1 Mut	No Data	No Data
R175H	1279	Mutant 2 Wt / 265 Mut	Mutant 6 Wt / 36 Mut	Mutant 2 Wt / 5 Mut	Mutant 15 Mut	Too few Data 5 Mut
R175I	0	Wild type 4 Wt / 2 Mut	No Data	Too few Data 1 Wt / 1 Mut	No Data	No Data
175K	0	discrepency	No Data	Too few Data	No Data	No Data
11/5K	U	3 Wt / 3 Mut Wild type	Wild type	1 Wt / 1 Mut Too few Data	No Data	No Data
R175L	27	85 Wt / 30 Mut Wild type	11 Wt / 1 Mut Too few Data	3 Wt / 2 Mut Too few Data	No Data	No Data
175N	0	5 Wt / 3 Mut	1 Wt / 1 Mut	5 Wt	NO Data	No Data
175P	8	Wild type	Too few Data	Mutant	Too few Data 1 Wt / 2 Mut	No Data
TYSP	•	40 Wt / 27 Mut Wild type	3 Wt / 2 Mut Too few Data	3 Wt / 5 Mut Too few Data	No Data	No Data
175Q	5	4 Wt / 3 Mut	1 Wt	1 Wt / 4 Mut	N. B. L.	
1755	10	Wild type 27 Wt / 6 Mut	No Data	Too few Data 1 Wt / 1 Mut	No Data	No Data
R175T		Wild type	No Data	Too few Data	No Data	No Data
20000000	1	4 Wt / 2 Mut Mutant	Too few Data	2 Wt Too few Data	No Data	No Data
175W	0	7 Mut Mutant	1 Mut Too few Data	3 Wt / 2 Mut Too few Data	No Data	No Data
	_		1 Mut	3 Wt / 2 Mut	No Data	No Data
upp.	• Figur	re S3	THUC	J M. / 2 Mac		
				J W. / 2 Mac		
				J m. / 2 mac		
				J m. / I m.		
			T MU			
			T MU			
			T MU			
			T Put			
			T Put			

Comment 1

200 and more	20-199	6-19	1-5	Not described
This is a hot spot	This mutant is frequent	This mutant is not	This mutant is very	This mutant has
mutant		frequent	rare	never been
				described so far

Comment 2

Non sense mutation	Activity 0-25%	Activity 25-50%	Activity >50%	No data
The activity of truncated p53	This mutant is inactive	There is partial activity for this	This mutant does not display	Remaining activity of this
is assumed to be null.		mutant	a significant loss of activity	mutant is unknown

Comment 3

	200 and more	20-199	6-19	1-5	Not described
No data available	No Problem	No Problem	CHECK CAREFULLY	CHECK CAREFULLY	CHECK CAREFULLY
			BEFORE PUBLICATION	BEFORE PUBLICATION	BEFORE PUBLICATION
Non sense mutation	No Problem	No Problem	No Problem	CHECK CAREFULLY	CHECK CAREFULLY
				BEFORE PUBLICATION	BEFORE PUBLICATION
Activity 0-25%	No Problem	No Problem	No Problem	CHECK CAREFULLY	CHECK CAREFULLY
				BEFORE PUBLICATION	BEFORE PUBLICATION
Activity 25 to 50%	The remaining activity of this p53		CHECK CAREFULLY	CHECK CAREFULLY	CHECK CAREFULLY
	mutant raises some concerns	*	BEFORE PUBLICATION	BEFORE PUBLICATION	BEFORE PUBLICATION
	about its association with				
	transformation				
Activity > 50%	The remaining activity of this p53	CHECK CAREFULLY	CHECK CAREFULLY	CHECK CAREFULLY	CHECK CAREFULLY
	mutant raises some concerns	BEFORE	BEFORE PUBLICATION	BEFORE PUBLICATION	BEFORE PUBLICATION
	about its association with	PUBLICATION			
	transformation				

Supp. Table S1

Supplementary table 2: Structure of the OPMA database.

	Number of entries	Number of publications
Transactivation	78,000	180
DNA binding	845	66
Apoptosis	404	45
Growth arrest	385	68
Dominant negative activity	29	1617
Thermosensitivity	24,392	35
Mutant specific transactivation	657	62

Supp. Table S2