

# TP53 Mutations in Human Cancer: Database Reassessment and Prospects for the Next Decade

Bernard Leroy,<sup>1</sup> Martha Anderson,<sup>2</sup> and Thierry Soussi<sup>1,2\*</sup>

<sup>1</sup>Université Pierre et Marie Curie-Paris 6, Paris 75005, France; <sup>2</sup>Karolinska Institute Department of Oncology-Pathology Cancer Center Karolinska (CCK), Stockholm SE-171 76, Sweden

For the TP53 Special Issue

Received 16 January 2014; accepted revised manuscript 4 March 2014.

Published online 24 March 2014 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22552

**ABSTRACT:** More than 50% of human tumors carry TP53 gene mutations and in consequence more than 45,000 somatic and germline mutations have been gathered in the UMD TP53 database (<http://p53.fr>). Analyses of these mutations have been invaluable for bettering our knowledge on the structure–function relationships within the TP53 protein and the high degree of heterogeneity of the various TP53 mutants in human cancer. In this review, we discuss how with the release of the sequences of thousands of tumor genomes issued from high-throughput sequencing, the description of novel TP53 mutants is now reaching a plateau indicating that we are close to the full set of mutants that target the elusive tumor-suppressive activity of this protein. We performed an extensive and thorough analysis of the TP53 mutation database, focusing particularly on specific sets of mutations that were overlooked in the past because of their low frequencies, for example, synonymous mutations, splice mutations, or mutations-targeting residues subject to posttranslational modifications. We also discuss the evolution of the statistical methods used to differentiate TP53 passenger mutations and artifactual data from true mutations, a process vital to the release of an accurate TP53 mutation database that will in turn be an invaluable tool for both clinicians and researchers.

Hum Mutat 35:672–688, 2014. © 2014 Wiley Periodicals, Inc.

**KEY WORDS:** TP53; mutation; LSDB; cancer

## Introduction

“TP53 (MIM #191170) is the most frequently mutated gene in human cancer.” This sentence, found in the introductions of thousands of publications, can be traced back to 1990, 1 year after the description of the first TP53 mutations in lung and colorectal carcinoma [Baker et al., 1989; Takahashi et al., 1989]. Hundreds of novel cancer genes were subsequently identified, but none were capable of stealing this infamous crown from TP53. Even whole-genome sequencing of several thousand cancer genomes has not revealed any

new suitors for the throne and to this day TP53 mutations are still frequently found among the five most significantly mutated genes in the most common human cancers [Stratton, 2011; Kandoth et al., 2013].

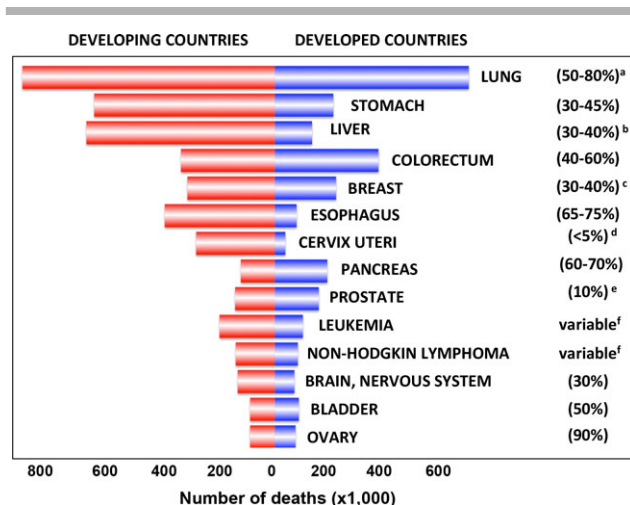
Nonetheless, the frequency of TP53 mutations is highly variable depending on the type of cancer (Fig. 1). Although nonsynonymous single-nucleotide variants (nsSNVs) are the most common TP53 alterations, it is worth noting that osteosarcoma is one of the few cancers that display a high frequency of TP53 gene deletion, an observation made in 1987 and confirmed by more recent whole-genome analysis, although the basis of this observation remains unknown [Masuda et al., 1987; Barretina et al., 2010]. The frequency of TP53 alterations appears to range from less than 5% in cervical carcinoma to 90% in ovarian carcinoma (Fig. 1), but these numbers must be taken with caution due to several factors.

First, most cancer types are heterogeneous entities comprising several subtypes [Ogino et al., 2012]. Diversities in cancer were initially defined via histological criteria and then confirmed by multiple molecular analyses that linked them to specific landscapes of genetic alteration. For example, two types of lung cancer have been identified: small cell lung cancer (SCLC) and non-SCLC (NSCLC). This latter is in turn divided into three main subtypes: adenocarcinoma (Adc), squamous cell carcinoma (SCC), and large cell carcinoma (LC) [Cooper et al., 2013]. The mutational profiles of the subtypes are different [CLCG Project, 2013]. KRAS mutation is frequent in Adc, but less so in the other subtypes (Supp. Fig. S1A). Similarly, the frequency of TP53 mutation varies across the types and subtypes, ranging from 50% in Adc and LC to 80% in SCLC and SCC (Supp. Fig. S1A). Mutational events are also distinct as the association of each subtype with smoking history is heterogeneous. Breast carcinoma too is a heterogeneous disease with multiple subtypes. Recent molecular profiling studies have identified four major subtypes: luminal A, luminal B, basal like, and HER2 [Weigelt et al., 2010]. The frequency of TP53 alteration in these subtypes ranges from 12% in luminal A to more than 80% in basal like (Supp. Fig. S1B) [Curtis et al., 2012]. Furthermore, for unknown reasons, basal-like tumors display a high frequency of TP53 nonsense and frameshift mutations compared with other breast cancer subtypes [Dumay et al., 2013]. In colon cancer, TP53 mutations occur less frequently in tumors with deficiencies in mismatch repair genes associated with high-microsatellite instability [Pugh et al., 2012].

A second factor influencing the frequency of TP53 mutations is the stage of development of the tumor. In prostate cancer, the frequency of TP53 mutations is low in primary tumors (10%–20%) but can reach 50% in metastatic tumors [Schlomm et al., 2008]. In leukemia and lymphoma, TP53 mutations are found more frequently at relapse, during blastic phase (CML) or in acute transformation (MDS) [Xu-Monette et al., 2012] and Malcikova et al.

Additional Supporting Information may be found in the online version of this article.

\*Correspondence to: Thierry Soussi, Karolinska Institute, Department of Oncology-Pathology, Cancer Center Karolinska, Stockholm SE-171 76, Sweden. E-mail: thierry.soussi@ki.se



**Figure 1.** Frequency of cancer deaths worldwide and relationship to the frequency of *TP53* mutations. **a:** The frequency of *TP53* mutations in lung cancer varies among the diverse subtypes (Supp. Fig. S1A). **b:** The frequency of *TP53* mutation in liver cancer can vary depending on the etiology of the tumor, for example, viral infection or exposure to food contaminants such as aflatoxin B1. **c:** The frequency of *TP53* mutations in breast cancer varies among the diverse subtypes (Supp. Fig. S1B). **d:** *TP53* mutation is negatively correlated to human papillomavirus infection in cervical cancer. **e:** *TP53* mutation in prostate cancer is more frequent in metastatic tumors. **f:** The frequency of *TP53* mutations in leukemia and lymphoma is highly heterogeneous among the various types, usually uncommon at diagnosis but can reach 50% at relapse or during blastic transformation. Cancer death numbers from GLOBOCAN 2008 (<http://globocan.iarc.fr>).

(2014) in this special issue. Further longitudinal analyses will be needed to assess the chronology of *TP53* mutation in various types of cancer and to understand how other genetic and/or epigenetic modifications cooperate with these mutations.

The third factor that strongly modulates *TP53* mutation frequency is exogenous features such as viral or bacterial infection [Levine, 2009; Moody and Laimins, 2010; Schetter and Harris, 2012]. Most human viruses impair *TP53* activity, thus constraining the cell to proficient viral DNA replication, a mechanism that was, on another note, at the root of the discovery of *TP53* [Kress et al., 1979; Lane and Crawford, 1979; Linzer and Levine, 1979]. In cervical cancer, the human papillomavirus E6 protein targets the *TP53* protein to the proteasomal degradation pathway [Scheffner et al., 1990]. This interaction is the most significant association between a viral protein and its cellular target. In liver cancer associated with hepatitis B virus (HBV) infection, this interplay between viral proteins and *TP53* is less clear as many HBV-positive hepatocellular carcinomas display *TP53* mutation [Hussain et al., 2007]. Furthermore, recent evidence suggests that bacterial infection may trigger the *TP53* pathway and activate specific *TP53* isoforms [Terrier et al., 2013].

Finally, the highly heterogeneous geographical distribution of carcinogen exposure shapes *TP53* mutation both qualitatively and quantitatively. Thus, patient cohorts originating from different areas of the globe will have different landscapes of mutation, in *TP53* and other regions of the genome [Ozturk, 1991; Hartmann et al., 1997].

## The *TP53* Network: A Target in Human Cancer

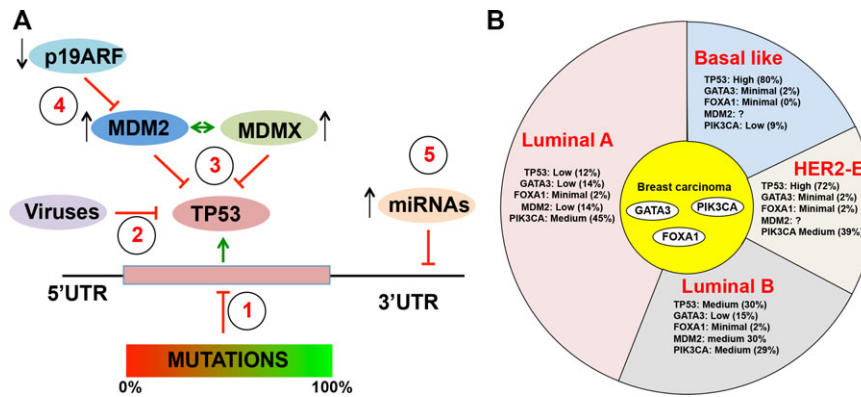
The *TP53* protein acts as a central hub that receives, integrates, and transmits multiple signals, generated during various stress events,

to ensure cell and tissue homeostasis. Although the most important activity of *TP53* is to act as a direct transcription activator for several hundred genes, it is also able to act as a transcription repressor. Moreover, it has several transcription-independent activities that make the investigation of this protein very complex. Discussing all the aspects of the various signaling pathways regulated by *TP53* is beyond the scope of this article. However, several recent reviews are available [Aylon and Oren, 2011; Levine et al., 2011; Berkers et al., 2013].

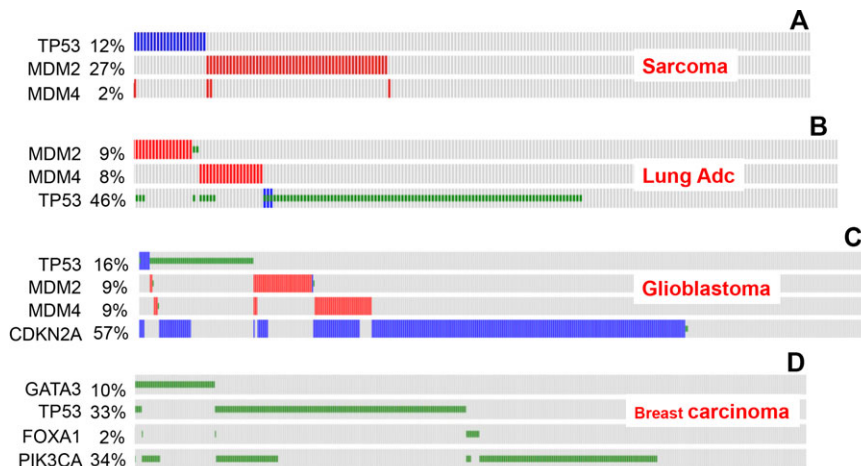
Although gene mutation is the major determinant of *TP53* inactivation, it is indisputable that the inactivation of genes closely associated with the *TP53* networks could substitute for inactivation of *TP53* itself. There should be significant heterogeneity among these related entities for several reasons. First, many of these pathways have significant tissue specificity, as illustrated by the specific requirements of particular types of cancer to target a specific set of driver mutations. The observation of gene deletions in sarcoma, and not *TP53* mutations per se, is one of the many arguments that comfort tissue specificity heterogeneity [Masuda et al., 1987; Barretina et al., 2010]. Second, *TP53* is associated with multiple networks that control the cellular response to several types of stress, and each one can be the driving force selected to impair *TP53* at different steps of neoplastic transformation. *TP53* has been shown to be an important player in angiogenesis, DNA damage response, oncogene activation, and aneuploidy propagation prevention to name just a few of its many roles [Vousden and Lane, 2007]. Whether the same set of *TP53* activities and partners must be targeted for each function of *TP53* is currently unknown. Finally, it is obvious that as a partner becomes more distant from *TP53* in the network, its association with *TP53* becomes weaker and the effect of its inactivation increasingly independent of *TP53*.

Via the sequencing of several thousand cancer genomes and the systematic identification of significant mutated genes in different types of cancer, the analysis of mutual exclusivity (or association) of different genomic alterations is possible [Ciriello et al., 2012]. Mutually exclusive genomic events identify genes that are active in common biological pathways where alteration of both genes would not add a selective advantage compared with alteration in one. Furthermore, these mutually exclusive events provide strong genetic evidence that the altered genes are functionally linked in a common biological pathway.

Figure 2A presents the most obvious branches of the *TP53* network that are impaired in human cancer. The MDM2 and MDM4 proteins are the prevailing targets since they act as direct negative regulators of *TP53*, downregulating its transcriptional activity (both MDM2 and MDM4) or targeting it to the ubiquitin-proteasome protein degradation pathway (MDM2 only) [Wade et al., 2013]. The *MDM2* gene is amplified in several tumors including breast carcinoma, glioblastoma, and sarcoma, whereas *MDM4* is amplified in melanoma, retinoblastoma, and breast carcinoma (Fig. 2B) [Oliner et al., 1993; Laurie et al., 2006; Wade et al., 2013]. Also, an abnormal accumulation of the mdm2 protein can occur in the absence of gene amplification, suggesting multiple mechanisms of deregulation and thus warranting different diagnostic procedures. In sarcoma, glioblastoma, or lung Adc, the distribution of *TP53*, *MDM2*, and *MDM4* alterations is highly exclusive, confirming that they target the same pathway in these types of cancer (Fig. 3A–C). We currently do not know whether the phenotype resulting from downregulation of *TP53* via the oncogenic activities of MDM2 in these tumors is the same as that resulting from an oncogenic mutant *TP53*. The *CDNK2A* gene encodes the 19/ARF protein that negatively regulates the interaction between *TP53* and MDM2 by targeting the latter in the nucleolus upon an oncogenic stress (Fig. 2A)



**Figure 2.** Heterogeneity of *TP53* inactivation in human cancer. **A1:** *TP53* mutations are found in as much as 50% of human cancers but their penetrance is highly heterogeneous, as reflected by the diversity of remaining transactivation activity, which ranges from 0% to 100%. 2: Various DNA viruses, such as SV40, HPV, or adenoviruses, encode proteins that target and impair the *TP53* protein. 3: MDM2 and MDM4 accumulation is found in numerous cancers, such as sarcoma or glioblastoma, that express wild-type *TP53*. 4: Inactivation of the *CDKN2A* locus that expresses p16 and p19ARF can be detected in various types of cancer. 5: Destabilization of *TP53* mRNA can occur either via dysregulation of miRNA expression or mutation of the 3'UTR of the RNA. **B:** Tissue specificity is an important component in the heterogeneity of *TP53* inactivation as shown by the specific, mutually exclusive pattern of *GATA3*, *TP53*, *PIK3CA*, and *FOXA1* alteration in breast cancer (see also Fig. 3D).



**Figure 3.** OncoPrint of genetic alterations for *TP53* and *TP53*-related genes in various cancers. The OncoPrint view provides an overview of genomic alterations in particular genes (horizontal rows) affecting particular individual samples in a large cohort (vertical columns). For each patient, alteration in a specific set of genes is immediately identified in a single column. Blue color: gene deletion; red color: gene amplification; green bars: mutations; gray color: no alteration. **A:** In sarcoma, the frequency of *TP53* mutations is very low as deletion is the major mechanism for *TP53* inactivation. *MDM2* amplification is found predominantly in tumors that express wild-type *TP53*. **B:** In lung Adc, *TP53* mutations are frequent but *MDM2*/*MDM4* alteration can be identified in 15% of wild-type tumors. **C:** In glioblastoma, *MDM2* or *MDM4* amplification is found in tumors that do not express mutant *TP53*. **D:** In breast carcinoma, alterations in *GATA3*, *TP53*, *FOXA1*, and *PIK3CA* are mutually exclusive, suggesting that they act in a common pathway. These mutual exclusivities to *TP53* mutations have been shown to be statistically significant. Data issued from TCGA portal analyses, <http://www.cbioportal.org/public-portal/index.do> (December 9, 2013) [Cerami et al., 2012; Gao et al., 2013].

[Li et al., 2011]. Deletion or methylation of this gene is frequently observed in human cancer, particularly in tumors with wild-type *TP53* (Fig. 3C). In breast carcinoma, *TP53*, *PIK3CA*, *GATA3*, and *FOXA1* mutations appear almost mutually exclusively relative to one another (Figs. 2B and 3D). *GATA3* mutations are more frequent in luminal-A type cancers presenting a low frequency of *TP53* mutations, suggesting that both genes act on a similar pathway, at least for this specific subtype [Curtis et al., 2012]. The *GATA3* transcription factor is an essential component for the differentiation of mammary stem/progenitor cells into luminal cells, but its relationship with the *TP53* pathway is still unclear [Zheng and Blobel, 2010]. The inverse correlation between *TP53* mutations and *PIK3CA* is also specific to breast cancer, as it is not observed in lung or colorectal carcinoma (Fig. 3D).

The *TP53* network is also under the direct or indirect control of a number of miRNAs having either positive or negative effects on *TP53* (Fig. 2A) [He et al., 2007; Wrighton, 2009]. Direct negative regulation can occur via the binding of specific miRNAs to the 3'UTR of the *TP53* mRNA (miR-125b, miR-504, and miR33 among others). Conversely, some miRNAs have a net positive effect because they target MDM2 or other negative regulators of *TP53*. It would be temptingly simple to state that either the upregulation of miRNAs associated with a negative effect or the downregulation of those associated with a positive effect will impair the *TP53* pathway. However, such a picture would be oversimplified and would not take into account the fact that most miRNAs have multiple targets and that the deregulation of any one of them may lead to a multitude of effects. In addition to *TP53*, miR-125b also directly represses other



target genes including *BBC3* (*PUMA*), *CCNC*, *CDC25A*, or *BAK1* [Hermeking, 2012]. Indeed, knowledge is lacking as to whether or not alterations in miRNA networks upstream or downstream of the TP53 pathway can substitute for TP53 mutations and this issue will be complex to assess, as several parameters (such as tissue specificity) are tightly associated with miRNA expression.

## Trends in TP53 Reports

Before the release of cancer genome high-throughput sequencing studies, mutation reports were highly stable because modifications to published data were very infrequent. Despite the obvious errors due to sequencing artifacts, incorrect data analysis or badly managed typography, authors almost never changed or retracted their works, and many dubious mutation reports still pollute the scientific literature. In contrast, mutations obtained via NGS are directly stored in large databases after being processed. Raw data from tumor DNA sequencing must go through a pipeline of different algorithms that successively remove background noise associated with the methodology, then neutral variants stored in the dbSNP at the NCI (<http://www.ncbi.nlm.nih.gov/SNP/>) before finally labeling a modification as a specific tumoral variant [Chin et al., 2011]. Thereafter, the greatest challenge is distinguishing driver mutations from random passenger mutations. The former confer a growth advantage and, by definition, reside in the subset of genes known as cancer genes, whereas the latter are distributed randomly in the genome at positions that do not lead to phenotypic change.

While updating the most recent release of the TP53 mutation database, we noticed many fluctuations in the various datasets of TP53 mutations associated with specific studies. For example, TP53 data from the TCGA study on ovarian cancer have been modified several times since their publication in 2010 [Bell et al., 2011]. Some modifications were due to improvements in the pipeline used to identify significant mutated genes and the use of novel algorithms to predict mutational effects on protein structures and functions [Lawrence et al., 2013]. The significance of other modifications was more obscure. For example, we noticed that all synonymous single-nucleotide variants (sSNVs) had been removed from the cbiportal database (<http://www.cbiportal.org/public-portal/>) but not from others. Similar observations were made for other sets of data. Although sSNVs are often associated with passenger mutations, they can also have drastic effects on splicing or RNA stability. In the case of TP53, several sSNVs have been clearly identified as detrimental for TP53 splicing (see the section *Silent Mutations*).

This fluctuation and heterogeneity in the various data sets will make it difficult for curators to carefully update mutation databases and be among the many reasons for the disappearance of locus-specific databases in their current format [Soussi, 2014].

The trend of TP53 mutation description since 1989 is shown in Figure 4 (see Soussi in this issue for a more in-depth discussion on mutation evolution [Soussi, 2014]). This graph, updated from its original publication in 2011, shows a renewal of trends due to the release of more than 4,000 mutations by the various tumor genome sequencing projects and the widespread use of NGS platforms [Soussi, 2011a]. Before this rebound, the numbers for publications and mutations published each year had culminated at about 200 and 2,400, respectively, before beginning to decline at about the turn of the century and through to 2009 (Fig. 4A and B, blue column). As discussed in the accompanying article, this was largely due to space considerations in scientific journals [Soussi, 2014]. After the nadir of 2009, the results of large studies using NGS started becoming available and led to a powerful rebound of

new data in the early 2010s (Fig. 4A and B, red column). Since the TP53 gene is the most frequently mutated gene in human cancer, we can expect that the large sequencing projects still underway will continue to generate great amounts of TP53 mutation data.

The current issue of the TP53 database contains 45,000 mutations with 1,540 variants issued from single-nucleotide substitutions and 2,000 frameshift mutants issued from small insertions or deletions. The trend in the number of variants being published for the first time, TP53 variant novelty, is shown in Figure 4C and D. During the first years, this trend was high because the pool of potential inactive TP53 variants was large. Variant novelty peaked in the years 1993-95 then began decreasing regularly as studies describing TP53 variants became increasingly redundant (Fig. 4C).

It is interesting to note that this trend is highly specific for SNVs as the rate of frameshift mutation novelties has remained stable over the past 20 years and is furthermore similar across conventional Sanger sequencing and NGS studies (Fig. 4C and D). The novel TP53 variants detected between 2010 and 2013 are mostly localized in exons 2, 3, 4, and 10, which were not commonly analyzed in previous studies.

TP53 SNV novelty is reaching a plateau, suggesting that we may now have the full set of variants selected for a loss of tumor-suppressor function (Fig. 4D). With this set of variants in hand, it is now possible to paint an accurate landscape of the TP53 residues that are essential for the tumor-suppressive effect of TP53.

## Curation of the TP53 Mutation Database

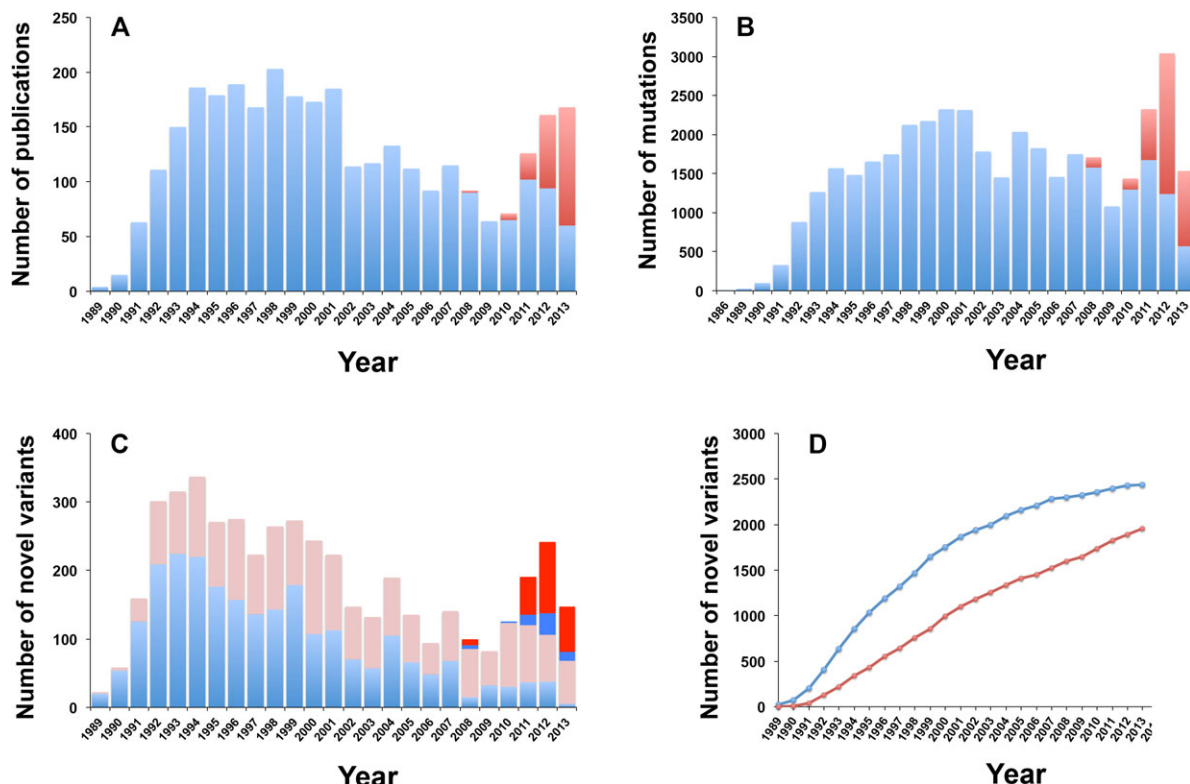
There is a marked difference in the frequency of the 45,000 somatic mutations included in the current updated issue of the database (<http://p53.fr>). Although the database is composed mainly of true driver mutations, it also contains a significant number of nondriver mutations; indeed, their identification is essential for providing an accurate database. Several algorithms have been developed to infer the pathogenicity of nsSNVs in human diseases. Some of them, such as Sift or Mut Assessor, can be used for any gene, whereas other are more specific to TP53 and rely on specific features of the protein [Ng and Henikoff, 2006; Reva et al., 2011]. It is beyond the scope of this article to review these procedures, but briefly their sensitivity does not exceed 75%–80% with a bias toward hot-spot mutations. Furthermore, they can have a high false-positive detection rate. For example, the passenger mutation, p.R175C, is identified as highly damaging by most algorithms because it is localized in a highly conserved residue.

For more than 10 years, we have been concerned by the quality of the data included in the TP53 database [Soussi and Beroud, 2001, 2003; Soussi et al., 2005, 2006a, 2006b; Edlund et al., 2012]. To circumvent obviously biased manual curation, we have developed and refined several statistical procedures to provide a highly curated set of TP53 mutations (Supp. Fig. S2).

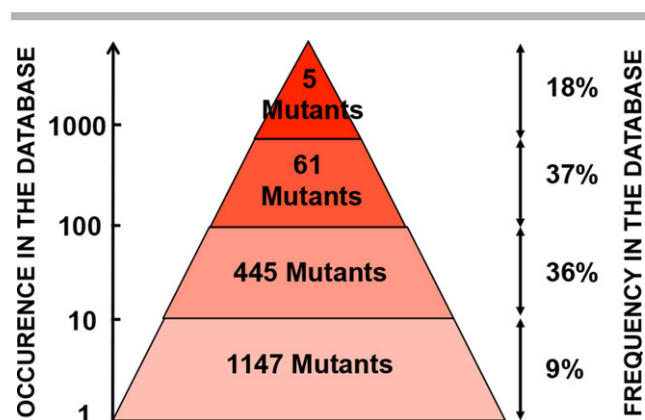
Mutant diversity in the database can be apprehended from two angles: TP53 variants can be considered irrespectively of the context of the mother publication, or they can be investigated in relation to the other variants described in the same publication.

## Heterogeneity of TP53 Variants

The most uncontested and solid information for TP53 variants is their frequency in the database. Five missense variants are found more than 1,000 times, 61 between 100 and 1,000 times, 445 between 10 and 100 times, and 1,147 less than 10 times (Fig. 5). Frequent mutants are undeniably driver mutations selected during the



**Figure 4.** Trends in publications on *TP53* mutation. Number of publications reporting *TP53* mutations (**A**) and number of mutations (**B**) published since 1989 with blue and red bars corresponding to conventional Sanger sequencing and NGS, respectively. **C**: Number of novel *TP53* variants published each year (mutant novelty) with blue and red corresponding to amino acid substitutions and indels, respectively, and light and dark colors corresponding to variants detected by conventional Sanger sequencing and NGS, respectively. **D**: Cumulated *TP53* novelty since 1989 with blue and red corresponding to amino acid substitutions and indels, respectively. Since 2008, the number of novel *TP53* variants detected in human tumors has plateaued, whereas the rate of detection of novel frameshift mutations has remained constant for more than 20 years. Data used for this analysis and those presented in the subsequent figures were issued from the 2014 release of the UMD *TP53* database (<http://p53.fr>).



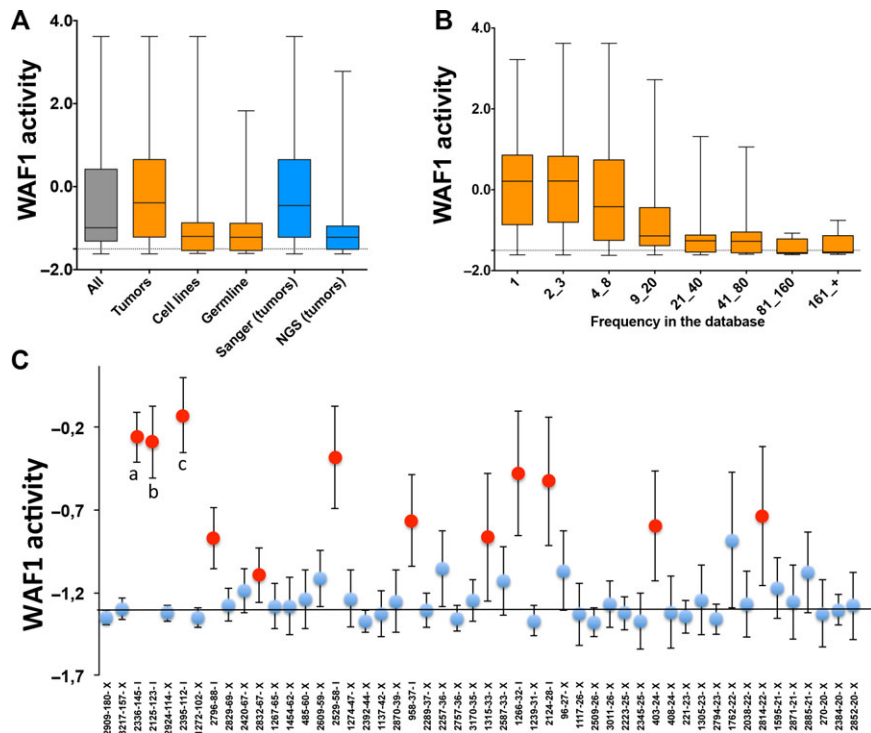
**Figure 5.** Frequency of *TP53* mutants in the UMD *TP53* database. Only missense and nonsense *TP53* variants are shown in this graph.

progression of neoplasia. Functional analyses that include both in vitro and in vivo analyses have shown that all *TP53* hot-spot mutations displayed a clear loss of activity (Fig. 6A–C) [Ory et al., 1994; Kato et al., 2003].

The situation is less clear for mutants appearing less frequently, and establishing the border between significant and nonsignificant mutants is not an easy task. Infrequent mutants, and more particularly those described only once or twice (630 variants, 2% of

the database), may be rare driver mutations with low penetrance, passenger mutations, or sequencing artifacts.

The transcriptional activity of more than 2,000 mutants was quantified in a yeast system using various response elements found in common *TP53* target genes such as *CDKN1A*, *BAX*, or *MDM2* [Kato et al., 2003]. A global analysis of all *TP53* mutants found in human cancer indicated that loss of activity is very heterogeneous, ranging from total inactivation to activity greater than that of the wild-type protein (Fig. 6A and B). *TP53* mutants can be categorized according to their origin, tumor, cell line, or germline. For the latter two, the detection of *TP53* mutations is less prone to sequencing artifacts, as the DNA is of better quality and the mutations are either heterozygote or homozygote, which eases analysis. Another advantage in cell lines is that *TP53* status benefits from accruing verification over multiple studies and this redundancy results in a good panel of certified *TP53* mutants. For germline mutations, sequencing is performed in clinical genetics laboratories with stringent quality controls that are not mandatory in research laboratories. The pattern of *TP53* germline mutations is similar to somatic mutations with a high frequency of transition at CpG dinucleotide (see the review of Kamihara et al. [2014] in this issue). For tumors, sample origin is highly variable, ranging from frozen cells or tissue to formalin-fixed paraffin-embedded (FFPE) specimens obtained after microdissection. These various features explain the differences in the distribution of *TP53* mutant activity: as shown in Figure 6A, the great majority of mutant *TP53* reported in cell lines or in constitutional DNA are inactive, whereas those observed in



**Figure 6.** TP53 mutation heterogeneity. Each TP53 mutant is associated with a quantitative value assessed in a yeast assay [Kato et al., 2003; Soussi et al., 2006a]. The activity without p53 or with wild-type p53 was  $-1.5$  and  $2.5$ , respectively, and the activity of the majority of TP53 mutants was situated between these two values. **A:** Analysis of mutant TP53 activity according to the origin of the sample. The y-axis corresponds to the remaining transcriptional activity of TP53 mutants [Kato et al., 2003]. Box-and-whisker plots show the interquartile range (boxes), median values (horizontal lines inside the boxes), and full-range distributions (whisker lines) for TP53 activity. All: entire database; tumors: tumors only; cell lines: cell lines only; germline: germline only; Sanger (tumor): tumors analyzed by conventional DNA sequencing; NGS (tumors): tumors sequenced by NGS (mostly from frozen samples). For germline mutations, the R337H mutation, very frequently found in patients with adrenocortical carcinoma in Brazil, was only added once to the database because it has been shown to be a founder mutation. **B:** Distribution of the remaining activity of all mutant TP53. Mutant TP53 are classified into eight categories according to their frequencies in the database. The y-axis corresponds to the remaining transcriptional activity of TP53 mutants. **C:** Meta-analysis of publications reporting TP53 mutations in breast carcinoma. Dots: for each publication, the quantitative value of all TP53 mutants was averaged. Bars: 95% C.I. The mean and 95% CI of p53 activity for all studies and all breast carcinomas is shown on the far left of the graph. Horizontal line, mean of the combined studies. The publication code is indicated on the x-axis: the first number is an anonymous ID for the publication and the second is the number of p53 mutants included in that study. Studies are presented from left to right in decreasing order of the number of TP53 variants they described. The y-axis corresponds to TP53 transactivation activity. Only studies reporting 20 or more TP53 mutations are shown on this graph. **a:** Publication associated with a high frequency of tumors with more than one mutation (30%), mutants with significant activities (60%), and an unusual hotspot of mutations. **b:** Publication associated with a high frequency of sSNVs (36%), tumors with more than one mutation (45%), and mutants with significant activities (50%). **c:** Publication associated with a high frequency of sSNVs (22%), tumors with more than one mutation (70%), and mutants with significant activities (50%).

tumors display a larger heterogeneity. The updated TP53 database now includes 6,500 mutations (250 variants) obtained via NGS of DNA extracted from frozen tissue. Loss-of-function in this specific set of mutants displays a distribution similar to that observed for cell lines or germ cells, a good indicator of the high quality of these NGS studies (Fig. 6A).

Combining a functional analysis with a determination of the frequency of each mutant in the entire database leads to the striking observation of a clear inverse correlation between the frequency of TP53 mutants and their activity: frequent TP53 mutants are always inactive, whereas approximately half of the mutants reported only once have activity greater than 50% compared with wild-type TP53, further underlining the very limited importance of these mutations during transformation (Fig. 6B).

### Heterogeneity of TP53 Publications

In 2006, we developed a novel method to rank publications reporting TP53 mutations in various cancers [Soussi et al., 2006a]. For

each article, the mean and 99% confidence interval of the residual activity of the reported TP53 mutants were calculated and displayed graphically among those of other publications for similar types of cancer. For this method, mutations were analyzed in a context of study homogeneity, that is, studies sharing a similar origin for the DNA sample and employing similar methodologies to detect mutations. In an ideal context, mutant distribution should be similar among the various publications with rare mutations being evenly dispersed. These analyses led to the identification of multiple outlier studies that reported a high frequency of TP53 mutants that did not sustain loss-of-activity (Fig. 6C).

In 2012, we developed a novel multiparametric analysis of all mutation reports using several independent quality criteria and showed that outlier studies cumulated a higher number of recurring independent singularities such as multiple mutations in a given tumor, rare mutants, sSNVs, or identical mutations, all of which are known to be associated with sequencing artifacts [Edlund et al., 2012]. Some criteria, such as multiple TP53 mutations in a given tumor or a high frequency of sSNVs, may be observed in a few patients

having a specific genetic background or a particular exposure to a mutagen, but they should concern only a few tumors in large cohorts. No studies using NGS and frozen tissue have shown any of these singularities, which furthermore have never been observed in cell lines. The majority of the outlier studies used DNA extracted from FFPE specimens, which are known to be highly heterogeneous in quality and vulnerable to DNA damage caused by the various fixation procedures [Soussi et al., 2014].

This problem is present throughout the database, but stands out particularly for breast cancer. In that setting, we noticed that eight studies, corresponding to 20% of the published *TP53* breast cancer mutations, were classified as outliers in the statistical analysis (Fig. 6C).

As is the case for the missclassification or the cross-contamination of cell lines, the publication of erroneous somatic mutation sequences is a widespread problem extending beyond *TP53* mutation reports [Berglind et al., 2008; Leroy et al., 2014]. We hope that researchers and journal editors will stop “burying their heads in the sand” and start taking full responsibility for their actions. Keeping these false sequences out of the literature and removing them when they get there requires fastidious attention. Indeed, once in the literature this pollution can do great harm by skewing the results of meta-analyses or participating in the dissemination of distorted information. For example, one of the most significant outlying *TP53* studies in breast cancer has been cited more than 150 times despite the publication of several letters of concern about its analyses.

The recent increase in the number of retracted articles and the discovery of cases of massive fraud in science will be an acute concern in the coming years [Gupta, 2013]. It is relatively easy to construct fraudulent mutation reports and relatively difficult to identify them; how many there may be, if any, is currently unknown.

To conclude on a note of optimism, we can hope that the bulk of data being produced by large-scale sequencing projects will not only improve the quality of the various cancer mutation databases but also dilute artifactual studies, reducing their contributions to insignificance.

## The Signature of the Mutational Process in Human Cancer: The *TP53* Contribution

For many years, the molecular epidemiology of human cancer was restricted to a few genes. Early work on the *ha-ras* gene showed that mutations had tumor-specific spectrums [Rodenhuis et al., 1988]. Unfortunately, missense mutations that activate the oncogenic activity of *ha-ras* are restricted to just a few codons (12, 13, and, to a lesser extent, 61 and 146), consequently limiting the impact of studies. In contrast, analyses of *TP53* mutations have been successfully used to establish links between exposure to carcinogens and various types of cancer (Supp. Fig. S3). Many reviews detailing these various findings are available [Brash, 1997; Giglia-Mari and Sarasin, 2003; Staib et al., 2003; Toyooka et al., 2003; Slade et al., 2009; Boffetta, 2010; Soussi, 2011b; Schetter and Harris, 2012].

In vitro and animal studies performed during the twentieth century identified specific mutational signatures associated with various mutagens [Soussi, 2011b]. Numerous sequencing studies performed on the *TP53* gene made it possible to identify these signatures in human cancers associated with specific exposures. The most remarkable example is that of tandem mutations, which are only observed in skin cancers and specifically induced by ultraviolet radiation (Supp. Fig. S3). The relationship between G→T transversion and lung cancer in smokers is also very noteworthy as is the mutation of codon 249 observed in aflatoxin B1-induced liver can-

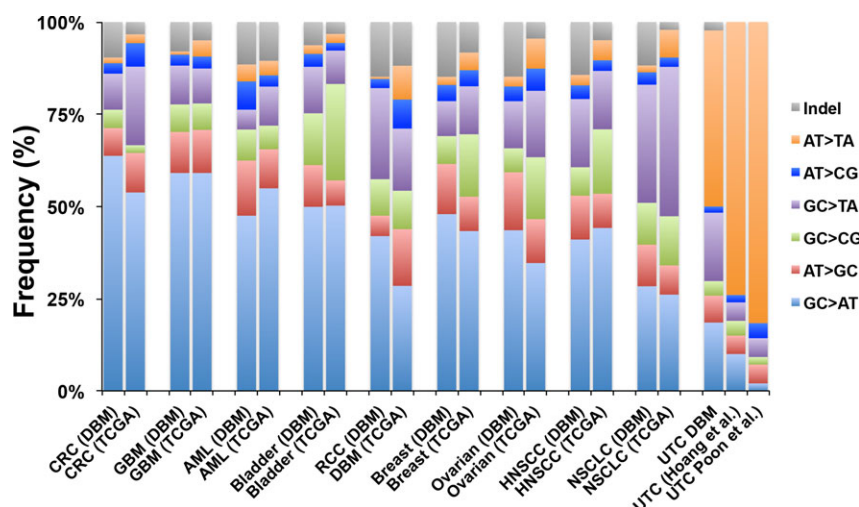
cers (Supp. Fig. S3). The studies leading to these discoveries were possible because *TP53* was the only gene that united several specific features used to study the origin of carcinogenesis in humans: (1) mutation in many types of cancer; (2) high mutation frequency; (3) predominantly modified by point mutations; (4) small enough to be relatively easy to analyze; (5) large number of codons that can be targeted by mutations (more than 200); and (6) multiple types of substitution can be identified in a single codon.

With the recent progress in sequencing methodologies, it is now possible to extend this type of analysis to the entire genome, a process that not only confirms the information derived from *TP53* studies, but also demonstrates that all types of alterations, passenger and driver, can be used to track mutagen fingerprints (Fig. 7) [Stratton et al., 2009]. Therefore, various biases that could interfere with *TP53* studies, such as the low frequency of *TP53* mutations in some cancers or the small size of the gene, can now be overcome. The use of passenger mutations for spectrum analysis offers the following advantages: (1) they can be found in large numbers in a single tumor (current analyses indicate that several tens of thousands of passenger mutations can be found per tumor genome), thus enabling single tumor profiling; (2) the large size of the human genome makes the analysis independent of the target size or sequence, allowing a detailed analysis of sequence context that was impossible in small genes; (3) due to their origin, they are true neutral mutations preventing any mutational profile bias that could occur during selection; (4) the lack of selective pressure allows an analysis of each type of cancer, avoiding the problem of specific gene inactivation in particular cancers [Alexandrov et al., 2013a, 2013b; Kandoth et al., 2013; Lawrence et al., 2013].

In lung cancer, the patterns of mutational events are very similar and show a striking similarity to the mutational events observed in the *TP53* gene (Fig. 7) [Pleasant et al., 2010]. The predominance of GC→TA transversions is a typical signature of carcinogen exposure associated with smoking. These observations are similar regardless of the strategy used to sequence the cancer genome (exome or whole-genome sequencing). For other types of cancer, such as colorectal cancer or brain tumors, the pattern of mutations found in the whole genome is also similar to that described for *TP53* (Fig. 7). They are predominantly GC→AT transitions, thus implicating an endogenous methylation-driven process (most probably deamination of 5-methylcytosine [5-mC]) as a major causative factor. Some upper urinary tract urothelial cell carcinomas (UTC) have been associated with exposure to aristolochic acid (AA), a natural compound found in traditional herbal preparations used as health supplements and remedies. The analysis of *TP53* mutations in UTC associated with AA exposure shows a high frequency of AT→TA transversion, a signature specific to the mutagenic effect of AA [Grollman et al., 2007]. Further sequencing of the entire genome of similar tumors confirms this particular landscape (Fig. 7) [Hoang et al., 2013; Poon et al., 2013]. Considering this association between cancer and exposition to carcinogens, the sequencing of genomes from liver cancer associated with aflatoxin B1 exposure will be very interesting, as the signature on the *TP53* gene is specific to codon 249. The identification—or lack of identification—of other mutation hotspots will provide more insight on the specificity of DNA adduct formation.

Recent analyses have uncovered other patterns of mutations associated with apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) cytidine deaminases, with a high frequency of C→T or C→G substitutions in the nucleotide context sequence TCW (W: A or T) [Burns et al., 2013; Roberts et al., 2013]. Using the *TP53* gene for specific patterns of mutations in various trinucleotide contexts is not really feasible since exonic mutations





**Figure 7.** Spectrum of mutations in human cancer deduced from sequencing studies of the TP53 gene (left columns) or full genome sequencing (right columns). The high frequency of GC>TA transversions in lung cancer is associated with tobacco smoking. The three spectrums on urothelial cancer (UTC) are developed from analyses of patients exposed to AAs, a setting that presents a very particular landscape of mutations dominated by A to T transversion. Full genome sequencing data were calculated from the pan-cancer studies [Kandoth et al., 2013] and urothelial data from the work of Poon et al. (2013) and Hoang et al. (2013).

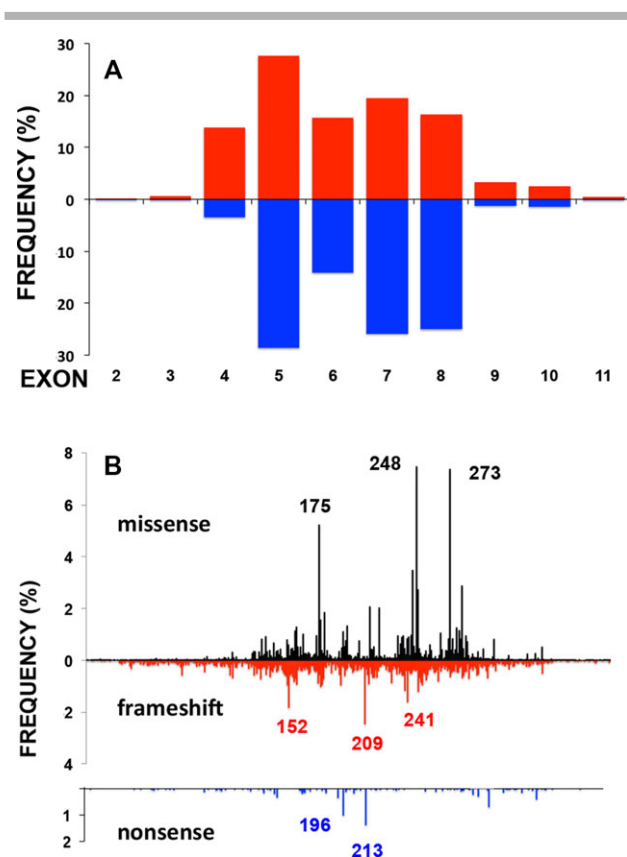
are shaped by both the nucleotide sequences and the resulting amino acids that must be selected by the mutation. Other specific patterns have also been uncovered but it is still unclear as to whether they have endogenous or exogenous origins; in vitro and in vivo mutagenesis assays will be required to uncover their etiology [Alexandrov et al., 2013a; Lawrence et al., 2013]. Further large-scale analyses of tumor genomes, such as the cancer 10K project focused on sequencing the genomes of 10,000 tumors of each of the most common cancers, may be of great help in increasing specific cancer etiology knowledge highly valuable for the development of preventive medicine.

## Distribution of TP53 Mutations

The high concentration of mutations in exons 5–8 encoding the DNA-binding domain of the TP53 protein was already obvious 25 years ago [Hollstein et al., 1991; Caron de Fromental and Soussi, 1992]. Although bias could occur due to the systematic analysis of these exons, sequencing the full-length gene has just softened the distribution (Fig. 8A and B).

The distribution of mutations in the TP53 protein is unique among all cancer genes, both oncogenes and tumor suppressors. All but five residues of the 393 amino acid long TP53 protein have been the target of at least one mutation in human cancer and each residue in the core domain has been found to be mutated a minimum of five times. This vast scattering of TP53 mutations is due to the high fragility of the core region; indeed, large-scale mutational analysis of each of its residues confirms its extreme sensitivity to substitutions.

It is important to keep in mind that the present analysis focuses on the major isoforms of the TP53 protein. As extensively discussed in an accompanying paper, a substantial number of mutations do not target all TP53 isoforms [Soussi et al., 2014]. Furthermore, beta and gamma exons localized in intron 9 have never been formally analyzed for mutations and only a handful of mutations have been described. A more intensive analysis of this region in human tumors must be undertaken. Any potential discoveries of mutation targeting will shed light on a putative tumor-suppressive activity for its isoforms.



**Figure 8.** TP53 mutations distribution. **A:** Occurrence of SNVs (missense and nonsense, blue bars) and frameshift mutations (red bars) among the various exons of the TP53 gene. A lack of screening for exon 4 in multiple studies created a partial analytical bias that could not be circumvented by immunohistochemistry since frameshift mutations do not lead to TP53 accumulation. **B:** Distribution of TP53 mutations in the TP53 protein. For nonsense mutations, hotspot codons at positions 196 and 213 are easily explained as the codons, both CGA, contain a CpG dinucleotide that will sustain a high frequency of C>T events.



The TP53 protein contains 64 amino acid residues that are conserved in all vertebrates. Two of them, resulting from codons 19 and 23, have never been the target of a single substitution in human cancer. This is not surprising since Phe19 and Trp23 are essential for the binding of MDM2 to TP53, and any substitutions at these positions would be lethal for the cell [Kussie et al., 1996]. Residue Leu26, also important for MDM2 binding, is not fully conserved in every TP53 but similarly to Phe19 and Trp23, it has never been found mutated. Among the other conserved residues, all but three, Lys121, Val97, and Pro98, are frequently substituted in human cancer (between 25 and 2,000 times).

Lysine 121 is localized in the L1 loop of the protein (residues 113–123), a cold spot region for mutation. Neighboring residue 120 is essential for DNA binding as it makes direct contact with a guanine base in the major groove of the DNA [Cho et al., 1994]. Several authors have shown that artificial mutations in the L1 domain changed the specificity of DNA binding and mutants at positions 120, 121, or 123 displayed an increased affinity for some p53RE and increased apoptotic activity [Saller et al., 1999; Zupnick and Prives, 2006]. Although these observations explain why such mutants are not selected during tumorigenesis, the modular role of this L1 loop in distinguishing various TP53 target genes has not yet been identified. The function of the two other cold spot codons at positions 97 and 98 is currently unclear.

## Frameshift Mutations

Frameshift mutations account for 11% of all mutations found in the TP53 database but their distribution differs from that of missense mutations (Fig. 8B). Small insertions, deletions, or indels are found more frequently than SNVs in exons 4, 9, and 10, which do not encode the core regions of the protein (Fig. 8B). Immunohistochemical analyses performed either on tumors or in cell lines clearly show that these mutations lead to a total absence of TP53 protein expression and could therefore be considered as TP53 null. Three positions could be characterized as hotspot mutations since they correspond to more than 1.5% of the total indel events in the database (Fig. 8B). Codons 151 and 152 contain a monotonic run of five cytosines that is known to induce strand slippage during DNA replication. The major event detected at this position is a single-nucleotide insertion or deletion (c.454\_455ins1 or c.454del1C) (mutation description in this manuscript uses the full-length TP53 transcript [LRG\_321t1 or NM\_000546.5] and the full-length protein [LRG\_321p1 or NP\_000537.3] as a reference). See Soussi et al., 2014 in the present issue and “<http://www.lrg-sequence.org>” for more information on TP53 mutation nomenclature. Codons 240 and 241, associated with a c.721delT or a c.723delC, contain a TTCC sequence but whether it induces replication errors remains to be determined. Codon 209 is associated with the c.625\_626delAG mutational event, although juxtaposing an AAA triplet sequence, it does not contain an obvious sequence to explain its high frequency of mutation.

## Silent Mutations

The UMD TP53 database contains 1560 sSNVs that target 265 different residues in the TP53 protein. This number does not include sSNVs identified in tandem mutations across two adjacent codons as they are the result of a single mutational event that generates two TP53 variants, and thus cannot be considered as independent events. The frequency of sSNVs is largely underestimated because this type of modification is often discarded before publication as an “irrelevant change that does not affect the protein.” This is highly

unfortunate as it is now well established that sSNVs can have critical consequences on RNA splicing or stability as well as on translation efficiency [Sauna and Kimchi-Sarfaty, 2011].

Nineteen sSNVs in the TP53 gene are found more than 10 times in the most recent issue of the database (Table 1). They can be divided in three categories: those that are truly pathogenic (P), those that are truly sequencing artifacts (A), and those that have an unknown status (U). The sSNVs at codon 125 that encode a threonine can be issued from two different mutation events (c.375G>A or c.375C>T) and is found 54 times in the database, both as somatic and germline mutations. This nucleotide is localized at the end of exon 4, just before the donor site in intron 4, and impairs TP53 splicing (see the section *Splicing Mutations*) [Varley et al., 1998]. This mutation is the most frequent sSNV in the TP53 database. Similarly, sSNV c.672G>A (codon 224 encoding a glutamic acid), found 17 times in the database and localized at the end of exon 5, is likely to be detrimental for TP53 splicing.

Several other sSNVs are obviously sequencing artifacts. c.477C>T (codon 159 encoding an alanine), is reported 19 times in the database and has been described in 12 different publications, six of which were found to be outlier studies as defined in the previous section. One of these articles reported the mutation seven times. The 19 cases of mutation were identified in FFPE tumor tissue. Similarly, mutation c.477C>T (codon 177 encoding a proline), reported 17 times in the database, is associated with 13 publications, 11 of which are tagged as outliers.

Having categorized the sSNVs that are deleterious for TP53 function and those that are obviously sequencing artifacts, there remains seven sSNVs that have an unknown status but are reported more than 10 times in the database.

The remaining uncommon SNVs are more difficult to assess but it is likely that they comprise a mix of passenger mutations and sequencing artifacts (Supp. Table S1); the probability of a rare driver mutation is unlikely.

## Splice Mutations

Eukaryotic gene splicing is highly complex, involving signal sequences localized not only in introns but also in exon [Wang and Cooper, 2007]. Although mutations at donor and acceptor splice sites or branch point sequences localized in introns are well known to impair splicing, it is more difficult to associate splicing defects with exonic mutations. Exonic splicing enhancer (ESE) and exonic splicing silencer (ESS) are DNA sequence motifs (six bases) localized within an exon [Crotti and Horowitz, 2009]. They modulate the splicing of pre-mRNA either positively (ESE) or negatively (ESS). These regulatory sequences were first described more than 15 years ago, but their identification remains difficult, although there are several computer programs that predict ESE and ESS in the TP53 gene. We currently do not know whether mutations at these positions impair splicing in addition to other defects such as DNA binding.

Splice mutations in the TP53 gene are underestimated but they remain nonetheless uncommon. In the current version of the TP53 database, they represent 2% of all mutations but this number rises to 4% if we select only NGS studies that target the entire TP53 gene. These numbers take into account only those mutations that target the two conserved intronic nucleotides GT and AG localized at the exon/intron boundaries (Figs. 9 and 10). We currently do not know whether mutations can be localized at other positions in the splice sites or in the branching point signals, but if it is possible, the event should remain relatively infrequent as there is a strong selection for the tumor to express a mutant TP53.

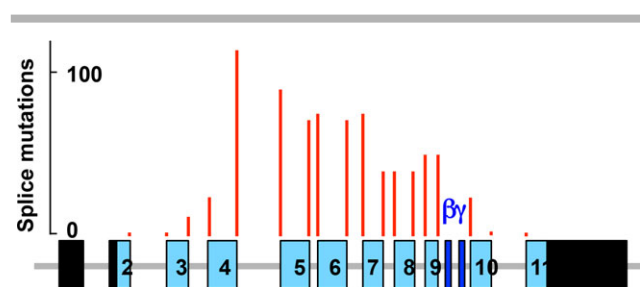
**Table 1. Synonymous Variants in the TP53 Database**

cDNA variant	Codon	Reports <sup>a</sup>	Frequency <sup>b</sup>	CpG <sup>c</sup>	Comment	Significance
c.375G>A	125 (Thr)	24/1	29/1	Yes	Splice mutation	Deleterious
c.375G>T	125 (Thr)	14/0	24/0	No	Splice mutation	Deleterious
c.477C>T	159 (Ala)	7/6	7/12	No	DNA used in the 13 publications was extracted from FFPE tissue Six out of 13 publications were classified as outliers Seven mutants were reported in a single publication	Sequencing artifact
c.531C>T	177 (Pro)	6/7	6/11	No	High frequency in outlier studies	Sequencing artifact
c.672G>A	224 (Glu)	12/2	14/3	No	Splice mutation	Deleterious
c.483C>T	161 (Ala)	9/6	9/7	No	High frequency in outlier studies	Sequencing artifact
c.741C>T	247 (Asn)	10/6	10/6	No	High frequency in outlier studies	Sequencing artifact
c.465C>A	155 (Thr)	7/4	10/6	No		Unknown; passenger mutation?
c.750C>T	250 (Pro)	9/4	10/5	No		Unknown; passenger mutation?
c.459C>T	153 (Pro)	10/4	10/5	Yes		Unknown; passenger mutation?
c.678C>T	226 (Gly)	2/5	2/12	No	High frequency in outlier studies	Sequencing artifact
c.462C>T	154 (Gly)	10/4	10/4	No		Unknown; passenger mutation?
c.456G>A	152 (Pro)	6/5	7/6	Yes	High frequency in outlier studies	Sequencing artifact
c.732C>T	244 (Gly)	6/5	7/6	Yes	High frequency in outlier studies	Sequencing artifact
c.519G>A	173 (Val)	8/4	9/4	No	High frequency in outlier studies	Sequencing artifact
c.744G>A	248 (Arg)	4/4	4/8	Yes	High frequency in outlier studies	Sequencing artifact
c.423C>T	141 (Cys)	7/3	8/4	No		Unknown; passenger mutation?
c.417G>A	139 (Lys)	8/3	8/3	No		Unknown; passenger mutation?
c.861G>A	287 (Glu)	8/2	8/3	No		Unknown; passenger mutation?

<sup>a</sup>Number of publications that report each TP53 variant. The two numbers correspond to validated and outliers studies, respectively [Edlund et al., 2012].

<sup>b</sup>Frequency of each mutant in the 2014 issue of the UMD TP53 database. The two numbers correspond to variants from validated and outliers studies, respectively.

<sup>c</sup>Indicates whether the mutation is localized at a CpG dinucleotide.



**Figure 9.** Distribution of splice mutations in the *TP53* gene. Substitutions targeting the two conserved intronic nucleotides at the donor site (GT) or the acceptor site (AG) have been summed for each site. For intron 4 and 5, exonic sSNVs that impair donor sites have also been included, which explains the high frequency of mutations at these positions. For introns 2 and 3 as well as for the two novel exons  $\beta$  and  $\gamma$  in intron 9, the absence of mutation reflects predominantly the lack of studies having analyzed these regions.

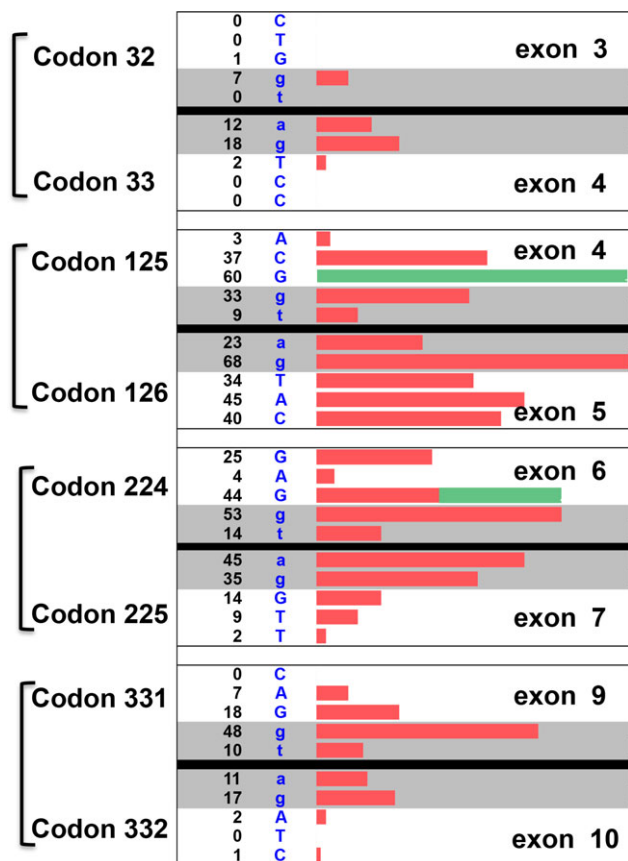
As discussed in the previous section, exonic mutations localized close to an intron/exon boundary can be detrimental to correct splicing (Fig. 10). These events are easy to spot when the nucleotide is at the third position of a codon as most mutations will lead to an sSNV (Fig. 10). Analysis of the *TP53* gene revealed that four introns (3, 4, 6, and 9) cause no codon interruption. This includes codons 125 and 224 described in the previous section as well as codons 261 and 331 that display sSNVs at low frequencies in the database (three and six times, respectively) (Table 1 and Supp. Table S2).

The distribution of splice mutations is similar to that of missense mutations (Fig. 9). However, more research is needed to determine whether this is due to a bias associated with the high number of studies that have focused on these regions or representative of true selection. Although NGS studies encompassing the whole *TP53* gene share similar distributions, more studies are needed to ascertain this observation. This landscape of mutations does not include potential alteration of splicing effects in exons  $\beta$  and  $\gamma$  localized inside intron 9 as they were blind from analysis.

## Transition at CpG Dinucleotides

Across the entire *TP53* gene, 425 CpG dinucleotides have been observed, 42 of which are localized in coding exons. The cytosines in these 42 CpG sites are methylated in normal tissue but the exact purpose of this modification for *TP53* regulation is currently unknown [Tornaletti and Pfeifer, 1995]. Cytosine methylation also causes genome instability and probably a third of all transition mutations responsible for genetic diseases and cancers in humans [Cooper and Yousoufian, 1988]. These effects are due to the much higher spontaneous deamination of 5-mC as compared with cytosine. Deamination of 5-mC leads to T:G mismatches, whereas deamination of cytosine generates U:G mismatches. Both are recognized by mismatch-specific DNA glycosylases within the base excision repair system. Several lines of evidence suggest that other mechanisms may also be associated with the high frequency of substitution at CpG dinucleotides. Exogenous carcinogens, such as Benzo(a)pyrene or UV sunlight, have greater affinity for methylated CpG dinucleotides than for their unmethylated counterparts [Denissenko et al., 1997; You et al., 2000]. It is also possible that endogenous mutagens, derived from an altered cell metabolism, target methylated CpG dinucleotides resulting in a high rate of transition.

CpG in coding regions can take three forms depending on their location inside a codon: CGN, NCG, or NNC-GNN, referred here as type I, II, and III, respectively (Fig. 11). The symmetrical nature of a CpG dinucleotide induces methylation of the involved cytosines on the two DNA strands. Transition in type I CpG always results in amino acid substitution or the synthesis of a nonsense mutant regardless of whether it is the first or the second base that is changed (Fig. 11A). Specific targeting of the C on the coding strand will lead to a C to T transition, whereas targeting of the C on the noncoding strand will lead to a similar event that will be translated as a G to A transition on the coding strand. Transition in type II CpG will result in mutation only when the C on the coding stand is modified; a mutation of the opposite stand will alter the third base of the codon, which generally does not change the amino acid residue due to degeneration of the genetic code (Fig. 11B). Inversely, transition

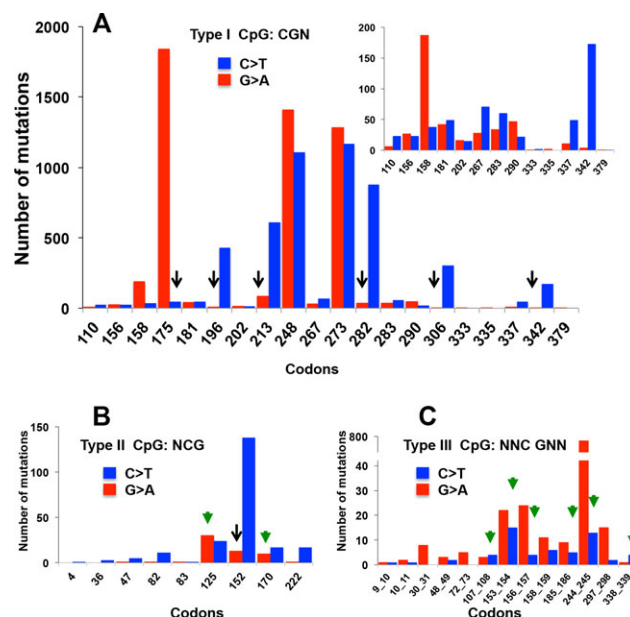


**Figure 10.** Distribution of *TP53* mutation in and around introns 3, 4, 6, and 9. For each intron, the preceding and following codons (blue upper-case letters in white boxes) are shown, as are splice junctions (blue lower-case letters in gray boxes). Codon numbers are shown on the left margin. The number of substitutions found at each position is indicated on the left and illustrated graphically by red or green bars, representing nsSNVs and sSNVs, respectively.

in type III CpG will lead to an amino acid change only when the C on the noncoding strand is modified (Fig. 11C).

All but one of the 42 “coding” CpG dinucleotides in the *TP53* gene have been found to be mutated at least once in human cancer, corresponding to 25% of all the mutations in the database, but there is a high degree of heterogeneity in their frequency (Fig. 11 and Supp. Fig. S4). The inverse relationship between this frequency and the loss-of-activity of the resulting variant is striking and demonstrates that many mutants resulting from these substitutions are not truly selected during neoplasia because they are still active. In contrast, the most frequent mutants issued from transition at CpG dinucleotides correspond to the *TP53* hotspot of mutations and display a clear loss-of-function (Supp. Fig. S4).

Examination of the frequency of transition at the two hotspot codons, 248 and 273 (type I CpG, more than 1,000 hits), shows a roughly equal frequency of C>T and G>A transitions (Fig. 11A), which is not surprising since the deamination of C residues in both strands of a symmetrical CpG dinucleotide will occur and be repaired at a similar rate. Therefore, the two *TP53* variants resulting from these substitutions should occur at the same frequency if they both act as driver mutations with similar properties. The four *TP53* variants issued from transitions at codon 248 or 273 (p.R248Q, p.R248W, p.R273H, and p.R273C) have been extensively analyzed



**Figure 11.** Distribution of G>A (red) and C>T (blue) transitions at each CpG dinucleotide of the *TP53* gene. Black or green arrows indicate codons that show disequilibrium in the frequency of their transitions or that lead to an sSNV. For type I CpG, the insert displays CpG mutated less than 200 times for a better view of infrequent mutations.

and their loss-of-function has been confirmed both in vitro and in vivo.

In contrast, codon 175, also including a type I CpG, shows a marked disequilibrium in the distribution of mutations, with 1889 G>A (p.R175H) and only 45 C>T (p.R175C) (Fig. 11A).

We reported the disequilibrium for codon 175 as early as 1994 as it was already obvious in a mutation database composed of only a few hundred entries [Ory et al., 1994]. To analyze the role of this codon in *TP53* function and/or loss-of-function, we constructed a library of 15 different p53 mutations at position 175 [Ory et al., 1994]. We found that the G>A transition led to variant p.R175H, the most frequent mutant in the database. The biochemical and biological function of p.R175H mutant (c.524G>A) are impaired entirely, and it is furthermore associated with various gain-of-function activities [Blandino et al., 1999]. On the other hand, the p.R175C mutant (c.523C>T) is not impaired for any *TP53* function [Ory et al., 1994]. Considering this body of evidence, we hypothesize that p.R175C may be a passenger mutation coselected during neoplastic transformation. This hypothesis is supported by the following observations: (1) 30% of tumors with the p.R175C variant also have other *TP53* mutations; (2) the variant is highly uncommon in cell lines; and (3) it has never been described as a germline mutation.

Disequilibrium has also been observed at codons 196 (432 p.R196\* vs. 11 p.R196Q), 213 (611 p.R213\* vs. 84 p.R213Q), 267 (71 p.R1267W vs. 28 p.R267Q), 282 (881 p.R282W vs. 35 p.R282Q), 306 (302 p.R306\* vs. 3 p.R306Q), and 342 (173 p.R342\* vs. 4 p.R342Q) with ratios ranging from 2.5 (for codon 213) to 100 (for codon 306) (Fig. 11A). It is worth noting that an arginine to glutamine substitution is observed among all the variants found at low frequencies. This substitution manifests as the tolerated transformation of the positively charged polar amino acid arginine to another polar residue, glutamine, suggesting that these mutations have only a mild effect on *TP53* activity.

Three of the above-mentioned variants are found at very low frequencies (p.R196Q, p.R306Q, and p.R342Q) and are obviously passenger mutations, whereas two others, p.R213Q and p.R267Q, are found with significant frequency, suggesting that they are low-penetrant mutants. It is also noteworthy that four of the five frequent variants are nonsense mutations but whether this is due to coincidence or a greater aptitude for selection during transformation for these variants is an open question.

For type II and III CpG, only two mutants, p.P152L (type II, found 138 times) and p.G245S (type III, found 689 times), are obviously driver mutations associated with impaired transcriptional activity (Fig. 11B and C). sSNVs localized at codon 125 are discussed in previous sections.

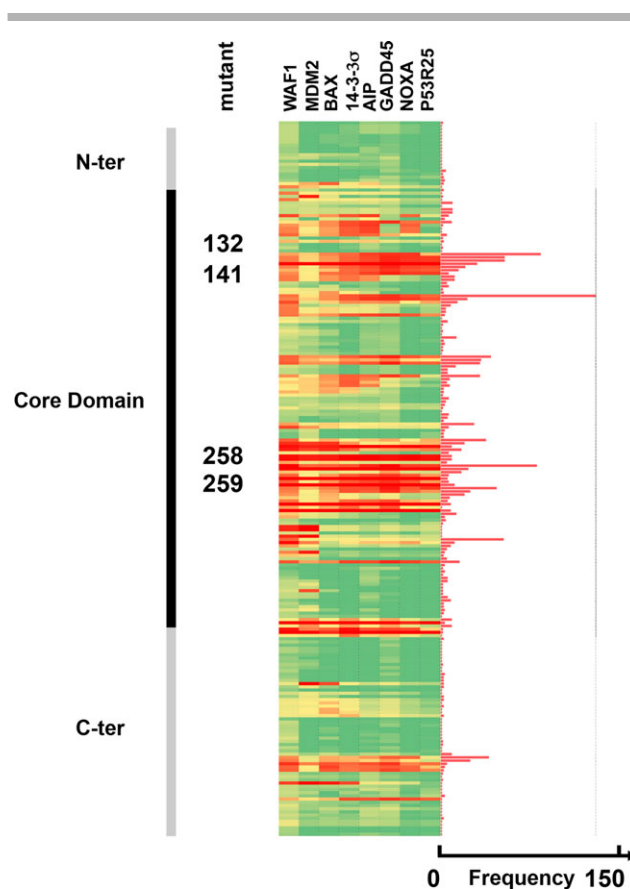
Other transitions, some resulting in sSNVs others in nsSNVs, occur from one to 24 times in the database and their status as true mutations is questionable (Supp. Fig. S4). Most of the nsSNVs retain a significant residual activity and several of them are found in tumors with multiple TP53 mutations.

## Posttranslational Modification

Like many eukaryotic transcription factors, the TP53 protein is the target of multiple posttranslational modifications (PTMs) that modulate its activity [Meek and Anderson, 2009], (see also the review of Nguyen et al. [2014] in the present issue). Several proteins coded from driver genes such as *CCND1* or *CTNNB1* are impaired by mutations that target residues modified by phosphorylation. Analyzing the impact of mutations on PTM is rather tricky as these protein sites include several continuous residues around the modified amino acid, and mutations changing the net charge or the hydrophobicity of these neighboring residues can also be detrimental for modifications.

Twelve different types of modification targeting 62 out of the 393 residues of the TP53 protein have been uncovered, the most frequent being phosphorylation and ubiquitination [Meek and Anderson, 2009] and Nguyen et al. (2014) in the present issue. All but six of the modification-susceptible residues have been found to be mutated in various cancers, corresponding to 5% of the mutants in the database (Fig. 12). The frequency of these mutations is highly heterogeneous but as discussed later in this review, assessing the impact on PTM for residues in the core region of TP53 is difficult to conceptualize, as these mutations will also impair DNA binding and transcriptional activity (Fig. 12). In contrast, most PTM sites localized outside the core domain are not mutated at a high frequency and those mutations that do occur do not impair TP53 transcriptional activity significantly, suggesting that they are either passenger mutations or sequencing errors (Fig. 12). The only exception is residue 337 localized in the oligomerization domain and discussed in the next section. It should be noted that PTMs on a specific residue are modulated by the neighboring residues, which are important for recognition by modifying enzymes. Therefore, mutations close to PTM sites could be also deleterious. Analysis of the distribution of TP53 mutations close to the various PTM sites does not reveal any specific clustering.

This lack of selection for PTM sites as driver mutations is in line with observations made in murine models [Olsson et al., 2007]. For example, mice expressing single TP53 mutants at various phosphorylation sites displayed only very mild—if any—specific phenotypes and were not highly prone to cancer ([Donehower, 2014; Nguyen et al., 2014] in this special issue). Similarly, mice with triple mutations at the three acetylation sites localized in the DNA binding domain did not display early onset of tumors, suggesting that acety-



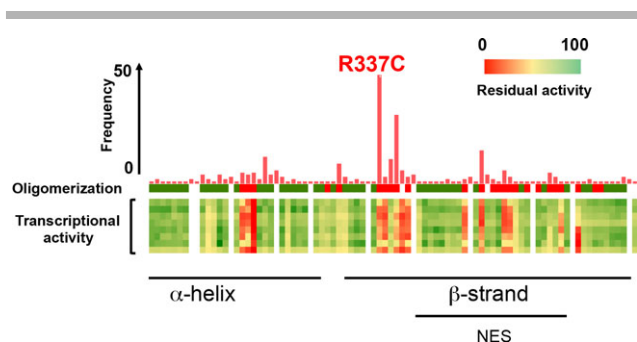
**Figure 12.** Impact of mutations in the various PTM sites of the TP53 protein. The heat map corresponds to the residual transcriptional activity of TP53 variants ranging from 0% to 100% compared with wild-type TP53. Each column represents a transcription promoter and each row represents a mutant TP53 at a PTM site. Activities are displayed from red (lowest) to red (highest) [Kato et al., 2003]. The frequency of each variant in the database is shown on the right side of the heat map as a red bar ranging from one to 160. Position for frequent mutants at codon 132, 141, 258, and 259 is shown on the left (see Supp. Fig. S5 for a more detailed figure) The most frequent mutants are also impaired for transactivation, making it difficult to evaluate the contribution of a deficient PTM.

lation at these sites is not essential for the tumor-suppression activity of TP53 [Li et al., 2012].

## Oligomerization Domain

Tetramerization is an important feature of TP53 and required for its DNA-binding activity. An unstressed cell contains predominantly dimers of TP53 that oligomerize quickly after DNA damage [Rajagopalan et al., 2011; Gaglia et al., 2013]. The oligomerization domain (residues 326–355) consists of a beta-sheet (residues 326–333), followed by an alpha-helix (residues 335–356) [Clare et al., 1994]. This domain is encoded by exon 9 and 10 and the splitting occurs between codons 331 and 332. As TP53 isoforms beta and gamma do not include the coding region of exon 10, mutations in this region will target predominantly TP53 proteins that express full-length carboxy terminus. The formation of TP53 tetramers follows a particular kinetic with the quick formation of dimeric molecules followed by a slow association of these dimers into tetramers. In tumor cells expressing both wild-type and mutant TP53, poised dimers containing a wild and a mutant subunit have not been





**Figure 13.** Mutation in the oligomerization domain of TP53. The lower heat map corresponds to the residual transcriptional activity of TP53 variants ranging from 0% to 100% compared with wild-type TP53 on eight different transcription promoters. Each column represents a TP53 mutant and each row represents a transcription promoter. Activities are displayed from red (lowest) to red (highest) [Kato et al., 2003]. The upper heat map corresponds to oligomerization status of the various mutant measured by cross-linking experiments indicating whether the protein is still a tetramer (green colors) or formed only monomers (red colors) [Kato et al., 2003]. The frequency of each variant is shown above the upper heat map as red bars ranging from one to 50 (see Supp. Fig. S6 for a more detailed figure). The beta strand and the alpha-helix as well as the nuclear export signal are shown below the heat map.

observed and consequently it appears as if only wt2/Mut2 tetramers are formed [Natan et al., 2009]. The oligomerization domain also includes a nuclear export signal (residue 340–351) exposed only when the protein is dimeric, suggesting a link between TP53 transport in the nucleus and DNA damage [Stommel et al., 1999].

At first glance, the importance of this domain for TP53 activity should lead to the observation of a high frequency of TP53 mutation. In fact, the opposite trend is observed: only a few mutations have been observed with the exception of codon 337 (Fig. 13). The oligomerization domain has been extensively analyzed and multiple functional data, such as transactivation or oligomeric state defined via multiple criteria or by biological activity, are available for many substitutions whether or not they were found in human cancer (Fig. 13) [Kawaguchi et al., 2005; Kamada et al., 2011]. The heat map summarizing these data shows that the most frequent mutants display a significant loss of transcriptional activities, a dysfunction in tetramerization, and impairment for apoptosis (Fig. 13). The only exceptions are mutants at codon 331 (significantly mutated in human cancer) that do not display any particular loss-of-function. An explanation for this may be that the codon is at the extremity of exon 9 and any mutation could impair splicing (see *Pleiotropic Effects of TP53 SNVs in Human Cancer*).

Arginine 337 forms an intermolecular salt bridge with aspartic acid at position 352 in TP53 dimers and is involved in the stabilization of the dimer [Clare et al., 1995]. Several variants at this position (p.R337C, p.R337L, or p.R337G) display impaired oligomerization and biological activities, and are frequently mutated in human cancer (Fig. 13). This position also includes p.R337H, found as a germline variant associated with pediatric adrenocortical carcinomas in Brazil, where this low-penetrant mutation is thought to be carried by thousands of people [Ribeiro et al., 2001; Custodio et al., 2013]. Haplotype analysis indicates a founder effect for p.R337H [Letouze et al., 2012]. As a somatic event, it is quite infrequent in the database (10 occurrences). Although the transcriptional activity of this mutant is close to that of the wild-type protein, it is very sensitive to pH change due to the protonation of the histidine residue [DiGiammarino et al., 2002].

Taken together, these data suggest that the oligomerization domain is not a primary target of the selection forces that drive neoplasia and inactivate TP53. Further structural and in vivo analyses of residue 337, which is found as a fascinating founder germline mutation and targeted by several somatic events, will provide valuable information on the relation between oligomerization and TP53 function in normal and tumor cells.

## Pleiotropic Effects of TP53 SNVs in Human Cancer

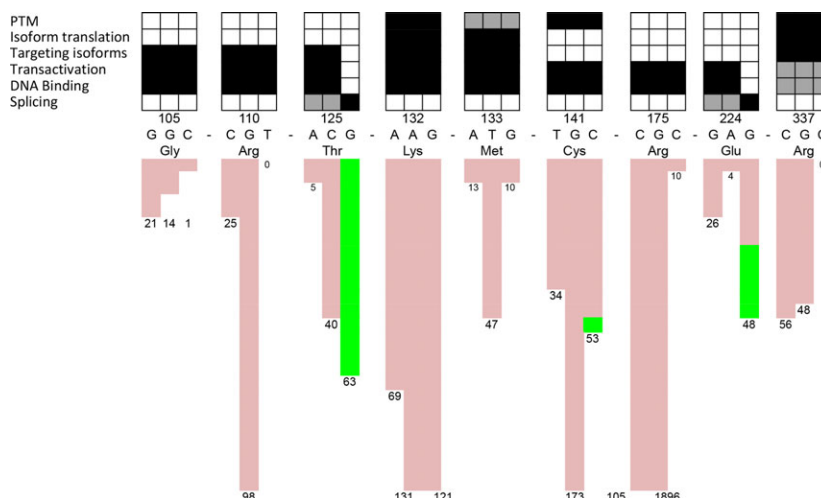
As most TP53 mutations are localized in the core region of the protein that contains the DNA-binding domain (residues 100–300), a binary relationship between TP53 mutations and inactivation of the transcriptional activity of the protein has been established. This relation is supported by multiple in vitro and in vivo studies and it is highly likely that the tumor-suppression function of TP53 is linked to a complex transcriptional program. Recent studies in mouse models have suggested that growth arrest, apoptosis, and senescence were not essential for tumor-suppression function [Brady et al., 2011; Li et al., 2012; Valente et al., 2013]. Consequently, novel target genes and novel pathways must be identified to obtain a more accurate understanding of the heterogeneous behavior of the diverse TP53 variants.

Sequencing of the coding regions of cancer genes has led to a protein-centric view of the consequences of the various SNVs included in the TP53 database. The concept of a eukaryotic gene and its various products must evolve to take into account and fully understand the multiple consequences of SNVs. Although the protein is the final product after translation, a single SNV can have multiple effects on posttranscriptional modifications, translation efficiency, or PTMs, leading to the synthesis of RNA and/or protein variants that carry a combination of multiple defects.

We identified at least five different effects that can be caused by SNVs on the TP53 gene or protein, such as splicing, DNA binding/transactivation, differential targeting of the various isoforms, isoform translation, and PTMs (Fig. 14). Most substitutions can impair DNA binding and transcriptional activity and additional effects can pile up. The TP53 gene is transcribed into eight different mRNAs and translated into 12 protein isoforms [Khoury and Bourdon, 2010]. Several isoforms lack the amino terminus and translation starts at codons 40, 133, or 160. Others do not contain the carboxy terminus due to alternative splicing with two newly discovered exons localized in intron 9. Depending on the position of the mutation, not all isoforms will be impaired. Indeed, mutants at codons 110, 127, or 342 are found more than 100 times in the database despite the fact that they do not target all TP53 isoforms (Fig. 14). To what extent these mutations contribute to the inactivation of the TP53 pathway is currently unknown, as is the significance of the expression of the remaining wild-type isoforms.

The ATG codons 133 and 160 assign amino acids to the DNA-binding site of the TP53 protein but they also act as start codons for isoforms Delta133 and Delta160. Mutations at these positions not only impair the transcriptional activity of the full-length protein, they also prevent the translation of the isoforms (Fig. 14). Translation efficiency is modulated by an overall sequence context, the Kozack sequence, that includes nucleotides adjacent to the start codon and therefore mutations in contiguous codons can increase or decrease the level of the deltaN isoforms.

As discussed in the previous section, several PTMs such as ubiquitination or neddylation contribute to the stability of the protein.



**Figure 14.** Potential cumulative effects of TP53 mutations. Nine representative residues of the TP53 protein are shown in this figure. The multiple possible consequences of a mutation on various TP53 characteristics are listed on the left. Black boxes: nucleotides leading to a confirmed modification of the characteristics; gray boxes: nucleotide substitutions leading to a potential modification of the characteristics. The number of substitutions at each position is shown as a bar below the sequence. Red: substitutions that change the residue; green: substitutions that do not change the residue. For PTM, substitutions will change the residue and therefore abolish TP53 modification, but whether this change will impair the tumor-suppressive effect of TP53 is unknown. Codon 132 cumulates multiple potential defects such as defects in DNA binding and transactivation and impaired ubiquitination, and will furthermore not target all isoforms. Base substitutions could affect the efficiency of translation starting at codon 133. Moreover, codon 133 mutations will impair the synthesis of delta 133 isoforms, DNA binding, and transactivation of full-length proteins as well as delta40 isoforms. It may also impair PTM at residue 132 by changing the recognition site of the ubiquitin enzyme.

Frequent mutations localized in the core region of the protein (residues Lys132 or Cys141) are modified by PTM and it is difficult to assess whether loss of DNA binding or PTM or both impair TP53 function (Fig. 14). For codon 337, the situation is even more complex as the residue, localized in the oligomerization domain and translated via exon 10, is not present in eight out of 12 isoforms and has furthermore been shown to be methylated. Therefore, the net consequence of any change at this residue will be complex with multiple potential consequences targeting only a subset of TP53 proteins. This is not a trivial point; as discussed previously, it is estimated that thousands of individuals carry the low-penetrant p.R337H TP53 mutation in Brazil. Structural and functional analyses have suggested that DNA binding and transactivation are moderately impaired, but most of the experiments were performed in vitro using the full-length isoform. More in vivo analyses are needed to fully understand how this mutation impairs the TP53 pathway.

Most of the information needed for correct splicing is not at the splice site itself and exonic sequences can have a profound impact on splicing efficiency. For a gene such as TP53 that expresses eight differentially spliced mRNA translated into 12 isoforms with sometimes opposing effects, a delicate equilibrium of the ratio of each protein is required. We can expect that several exonic mutations will be deleterious for both RNA splicing and protein function (Fig. 14). RNAseq technology will be of great aid in addressing this question as it will enable the quantitative and qualitative assessment of the status of all TP53 mRNAs in a tumor.

It is possible that these sidecar effects carry little weight compared with mutations affecting the main activity of TP53 directly, or that the loss of TP53 function is indeed the cumulative result of multiple defects. Exploring the contributions of the diverse variants will be very laborious but also surely mandatory if we hope to understand how each mutant ultimately affects TP53 function.

## Conclusion

The analysis of TP53 mutations in human cancer has been at the forefront of molecular epidemiology and led to the identification of specific patterns of mutation associated with carcinogen exposure [Schetter and Harris, 2012]. Whole-genome sequencing of tumors has confirmed that the global mutation landscape is similar to that previously described in the TP53 gene [Alexandrov et al., 2013a; Kandath et al., 2013]. Thanks to the large size of the human genome, novel original mutation signatures have been uncovered in specific tumor types and will contribute greatly to our understanding of cancer etiology. Novel patterns of mutation specific to TP53 are also emerging and in turn raising interesting questions. For example, the observation that basal-like breast cancers and ovarian cancer share a high frequency of TP53 mutations (more than 80%) with a high number of frameshift mutations needs to be confirmed and investigated. Analyzing whether this is a consequence of deficient DNA repair or specific selection for null TP53 status will be of interest.

The end of the twentieth century was dominated by the concept that single gene alterations could be used as efficient clinical biomarkers. As a result, much time was invested looking for the "Holy Grail" gene. TP53 was one of the genes at the forefront of these studies, resulting in thousands of publications, and yet only a handful of biomarkers are currently in clinical testing [Hutchinson and DeVita, 2009]. Great expectations were put on the use of TP53 mutations as clinical biomarkers for predicting prognosis or response to treatments, hopes supported by the central function of TP53 in the cellular response to DNA damage induced by chemotherapeutic drugs. After 25 years spent scrutinizing TP53 status in all types of cancer and after thousands of analyses and meta-analyses, we now have a much clearer vision of TP53 and can thus nuance our expectations and focus on specific niches where TP53 mutations do indeed have clinical value. Monitoring TP53 mutations in CML or in MDS before relapse or transformation is an example of a direct and clinically pertinent application.

The analysis of germline mutations in various cancer-prone families is also an essential clinical aspect of *TP53* and novel observations suggest that its value goes beyond Li–Fraumeni syndrome [Kamihara et al., 2014]. Indeed, *TP53* germline mutations have also been observed in families at high risk of breast cancer, albeit at very low frequency, suggesting that the spectrum of tumors associated with these mutations may be shaped by some currently unknown genetic factor.

For many other cancer types, the value of *TP53* mutations per se has not been powerful enough to warrant its use in the clinic. Before all else, we need to improve the diagnosis of *TP53* alterations qualitatively, taking into account the entire sequence of the gene, including the newly discovered exons. Furthermore, we can acquire a more limpid picture of the landscape of *TP53* inactivation in various types of cancer by looking at the whole *TP53* pathway, including the miRNA network that regulates *TP53* function. One of the predominant issues that has emerged from the cancer genome projects is the identification of specific landscapes of genetic and epigenetic alterations targeting specific pathways, beyond the question of tissue specificity. In this information-rich context, it should be possible to define specific niches where *TP53* alterations can act as strong biomarkers. Finally, if we wish to define the true clinical value of the molecule and its mutations, we must identify the exact nature of the tumor-suppressive function(s) of *TP53* that are impaired in human cancer.

New methodologies offering increased sensitivity for the detection of mutations in an excess of normal DNA have made feasible the use of circulating tumor DNA as a source of material for genetic analyses [Diehl et al., 2005; Dawson et al., 2013]. Thanks to their high frequency, *TP53* mutations will be among the most useful targets in this novel setting.

*TP53*-based therapy has been in the pipeline of numerous academic and pharmaceutical laboratories [Lane et al., 2010; Selivanova, 2010; Hong et al., 2014]. In the wake of the disappointment surrounding gene and viral therapies, a new generation of chemical drugs targeting mutant *TP53* is under development and several candidates are currently in clinical trials [Lehmann et al., 2012]. Drugs to inhibit the interaction between wild-type *TP53* and MDM2, thus increasing *TP53* activity, are also in development. In particular, these therapies hold the promise of offering efficacy in all types of tumors, whether they express wild-type or mutant *TP53* [Vassilev, 2004; Zhao et al., 2010].

For all of the various settings and perspectives described above, the *TP53* databases have been invaluable resources. They have played—and continue to play—key roles in linking clinical and basic research and developing numerous working hypotheses. Disabling the *TP53* pathway is ubiquitous in human cancer and perhaps even mandatory. Understanding how the various mutations shape tumors and deciphering the heterogeneity of the various targets in different tissues will be precious not only for treating cancer but also for understanding the roles of *TP53* in normal cells.

## Acknowledgments

Work in our laboratory was supported by Cancerföreningen iStockholm and Cancerfonden. While this paper was in press, Supsek et al. established experimentally that the synonymous mutations adjacent to splice identified in the presnet manuscript were indeed detrimental for *TP53* splicing.

## References

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Boyault S, Burkhardt B, et al. 2013a. Signatures of mutational processes in human cancer. *Nature* 500:415–421.

Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. 2013b. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* 3:246–259.

Aylon Y, Oren M. 2011. New plays in the p53 theater. *Curr Opin Genet Dev* 21:86–92.

Baker SJ, Fearon ER, Nigro JM, Hamilton SR, Preisinger AC, Jessup JM, vanTuinen P, Ledbetter DH, Barker DF, Nakamura Y, White R, Vogelstein B. 1989. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* 244:217–221.

Barretina J, Taylor BS, Banerji S, Ramos AH, Lagos-Quintana M, Decarolis PL, Shah K, Socci ND, Weir BA, Ho A, Chiang DY, Reva B, et al. 2010. Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. *Nat Genet* 42:715–721.

Bell D, Berchuck A, Birrer M, Chien J, Cramer DW, Dao F, Dhir R, Disaia P, Gabra H, Glenn P, Godwin AK, Gross J, et al. 2011. Integrated genomic analyses of ovarian carcinoma. *Nature* 474:609–615.

Berglind H, Pawitan Y, Kato S, Ishioka C, Soussi T. 2008. Analysis of p53 mutation status in human cancer cell lines: a paradigm for cell line cross-contamination. *Cancer Biol Ther* 7:699–708.

Berkers CR, Maddocks OD, Cheung EC, Mor I, Vousden KH. 2013. Metabolic regulation by p53 family members. *Cell Metab* 18:617–633.

Blandino G, Levine AJ, Oren M. 1999. Mutant p53 gain of function: differential effects of different p53 mutants on resistance of cultured cells to chemotherapy. *Oncogene* 18:477–485.

Boffetta P. 2010. Biomarkers in cancer epidemiology: an integrative approach. *Carcinogenesis* 31:121–126.

Brady CA, Jiang D, Mello SS, Johnson TM, Jarvis LA, Kozak MM, Kenzelmann Broz D, Basak S, Park EJ, McLaughlin ME, Karnezis AN, Attardi LD. 2011. Distinct p53 transcriptional programs dictate acute DNA-damage responses and tumor suppression. *Cell* 145:571–583.

Brash DE. 1997. Sunlight and the onset of skin cancer. *Trends Genet* 13:410–414.

Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, Refsland EW, Kotandeniya D, Tretyakova N, Nikas JB, Yee D, Temiz NA, et al. 2013. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 494:366–370.

Caron de Fromental C, Soussi T. 1992. TP53 tumor suppressor gene: a model for investigating human mutagenesis. *Genes Chromosomes Cancer* 4:1–15.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, et al. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2:401–404.

Chin L, Hahn WC, Getz G, Meyerson M. 2011. Making sense of cancer genomic data. *Genes Dev* 25:534–555.

Cho Y, Gorina S, Jeffrey PD, Pavletich NP. 1994. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science* 265:346–355.

Ciriello G, Cerami E, Sander C, Schultz N. 2012. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res* 22:398–406.

CLCG Project. 2013. A genomics-based classification of human lung tumors. *Sci Transl Med* 5:209ra153.

Clore GM, Ernst J, Clubb R, Omichinski JG, Kennedy WM, Sakaguchi K, Appella E, Gronenborn AM. 1995. Refined solution structure of the oligomerization domain of the tumour suppressor p53. *Nat Struct Biol* 2:321–333.

Clore GM, Omichinski JG, Sakaguchi K, Zambrano N, Sakamoto H, Appella E, Gronenborn AM. 1994. High-resolution structure of the oligomerization domain of p53 by multidimensional NMR. *Science* 265:386–391.

Cooper DN, Youssoufian H. 1988. The CpG dinucleotide and human genetic disease. *Hum Genet* 78:151–155.

Cooper WA, Lam DC, O'Toole SA, Minna JD. 2013. Molecular biology of lung cancer. *J Thorac Dis* 5(Suppl 5):S479–S490.

Crotti LB, Horowitz DS. 2009. Exon sequences at the splice junctions affect splicing fidelity and alternative splicing. *Proc Natl Acad Sci USA* 106:18954–18959.

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, et al. 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486:346–352.

Custodio G, Parise GA, Kiesel Filho N, Komechen H, Sabbaga CC, Rosati R, Grisa L, Parise IZ, Pianovski MA, Fiori CM, Ledesma JA, Barbosa JR, et al. 2013. Impact of neonatal screening and surveillance for the TP53 R337H mutation on early detection of childhood adrenocortical tumors. *J Clin Oncol* 31:2619–2626.

Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, Dunning MJ, Gale D, Forshew T, Mahler-Araujo B, Rajan S, Humphray S, et al. 2013. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* 368:1199–1209.

Denissenko MF, Chen JX, Tang MS, Pfeifer GP. 1997. Cytosine methylation determines hot spots of DNA damage in the human P53 gene. *Proc Natl Acad Sci USA* 94:3893–3898.

Diehl F, Li M, Dressman D, He Y, Shen D, Szabo S, Diaz LAJ, Goodman SN, David KA, Juhl H, Kinzler KW, Vogelstein B. 2005. Detection and quantification of

- mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci USA* 102:16368–16373.
- 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.
- Donehower LA. 2014. Insights into wild-type and mutant p53 functions provided by genetically engineered mice. *Hum Mutat* 35:715–727.
- Dumay A, Feugeas JP, Wittmer E, Lehmann-Che J, Bertheau P, Espie M, Plassa LF, Cottu P, Marty M, Andre F, Sotiriou C, Pusztai L, et al. 2013. Distinct tumor protein p53 mutants in breast cancer subgroups. *Int J Cancer* 132:1227–1231.
- Edlund K, Larsson O, Ameur A, Bunikis I, Gyllenstein U, Leroy B, Sundstrom M, Micke P, Botling J, Soussi T. 2012. Data-driven unbiased curation of the TP53 tumor suppressor gene mutation database and validation by ultra-deep sequencing of human tumors. *Proc Natl Acad Sci USA* 109:9551–9556.
- Gaglia G, Guan Y, Shah JV, Lahav G. 2013. Activation and control of p53 tetramerization in individual living cells. *Proc Natl Acad Sci USA* 110:15497–15501.
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, et al. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6:p11.
- Giglia-Mari G, Sarasin A. 2003. TP53 mutations in human skin cancers. *Hum Mutat* 21:217–228.
- Grollman AP, Shibutani S, Moriya M, Miller F, Wu L, Moll U, Suzuki N, Fernandes A, Rosenquist T, Medverec Z, Jakovina K, Brdar B, et al. 2007. Aristolochic acid and the etiology of endemic (Balkan) nephropathy. *Proc Natl Acad Sci USA* 104:12129–12134.
- Gupta A. 2013. Fraud and misconduct in clinical research: a concern. *Perspect Clin Res* 4:144–147.
- Hartmann A, Blaszyk H, Kovach JS, Sommer SS. 1997. The molecular epidemiology of p53 gene mutations in human breast cancer. *Trends Genet* 13:27–33.
- He L, He X, Lowe SW, Hannon GJ. 2007. microRNAs join the p53 network—another piece in the tumour-suppression puzzle. *Nat Rev Cancer* 7:819–822.
- Hermeking H. 2012. MicroRNAs in the p53 network: micromanagement of tumour suppression. *Nat Rev Cancer* 12:613–626.
- Hoang ML, Chen CH, Sidorenko VS, He J, Dickman KG, Yun BH, Moriya M, Niknafs N, Douville C, Karchin R, Turesky RJ, Pu YS, et al. 2013. Mutational signature of aristolochic acid exposure as revealed by whole-exome sequencing. *Sci Transl Med* 5:197ra102.
- Hollstein M, Sidransky D, Vogelstein B, Harris CC. 1991. p53 mutations in human cancers. *Science* 253:49–53.
- Hong B, van den Heuvel AP, Prabhu VV, Zhang S, El-Deiry WS. 2014. Targeting tumor suppressor p53 for cancer therapy: strategies, challenges and opportunities. *Curr Drug Targets* 15:80–89.
- Hussain SP, Schwank J, Staib F, Wang XW, Harris CC. 2007. TP53 mutations and hepatocellular carcinoma: insights into the etiology and pathogenesis of liver cancer. *Oncogene* 26:2166–2176.
- Hutchinson L, DeVita VTJ. 2009. The Holy Grail of biomarkers. *Nat Rev Clin Oncol* 6:553.
- Kamada R, Nomura T, Anderson CW, Sakaguchi K. 2011. Cancer-associated p53 tetramerization domain mutants: quantitative analysis reveals a low threshold for tumor suppressor inactivation. *J Biol Chem* 286:252–258.
- Kamihara J, Rana HQ, Garber JE. 2014. Germline TP53 mutations and the changing landscape of Li-Fraumeni syndrome. *Hum Mutat* 35:654–662.
- Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MD, Miller CA, et al. 2013. Mutational landscape and significance across 12 major cancer types. *Nature* 502:333–339.
- 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 USA* 100:8424–8429.
- Kawaguchi T, Kato S, Otsuka K, Watanabe G, Kumabe T, Tominaga T, Yoshimoto T, Ishioka C. 2005. The relationship among p53 oligomer formation, structure and transcriptional activity using a comprehensive missense mutation library. *Oncogene* 24:6976–6981.
- Khoury MP, Bourdon JC. 2010. The isoforms of the p53 protein. *Cold Spring Harb Perspect Biol* 2:a000927.
- Kress M, May E, Cassingena R, May P. 1979. Simian virus 40-transformed cells express new species of proteins precipitable by anti-simian virus 40 tumor serum. *J Virol* 31:472–483.
- Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, Pavletich NP. 1996. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 274:948–953.
- Lane DP, Cheok CF, Lain S. 2010. p53-based cancer therapy. *Cold Spring Harb Perspect Biol* 2:a001222.
- Lane DP, Crawford LV. 1979. T antigen is bound to a host protein in SV40-transformed cells. *Nature* 278:261–263.
- Laurie NA, Donovan SL, Shih CS, Zhang J, Mills N, Fuller C, Teunisse A, Lam S, Ramos Y, Mohan A, Johnson D, Wilson M, et al. 2006. Inactivation of the p53 pathway in retinoblastoma. *Nature* 444:61–66.
- Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, et al. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499:214–218.
- Lehmann S, Bykov VJ, Ali D, Andren O, Cherif H, Tidefelt U, Uggla B, Yachnin J, Juliusson G, Moshfegh A, Paul C, Wiman KG, et al. 2012. Targeting p53 in vivo: a first-in-human study with p53-targeting compound APR-246 in refractory hematologic malignancies and prostate cancer. *J Clin Oncol* 30:3633–3639.
- Leroy B, Girard L, Minna JD, Gazdar AF, Soussi T. 2014. Analysis of p53 mutation status in human cancer cell lines: a reassessment. *Hum Mutat* 35:756–765.
- Letouze E, Rosati R, Komechen H, Doghman M, Marisa L, Fluck C, de Krijger RR, van Noesel MM, Mas JC, Pianovski MA, Zambetti GP, Figueiredo BC, et al. 2012. SNP array profiling of childhood adrenocortical tumors reveals distinct pathways of tumorigenesis and highlights candidate driver genes. *J Clin Endocrinol Metab* 97:E1284–E1293.
- Levine AJ. 2009. The common mechanisms of transformation by the small DNA tumor viruses: the inactivation of tumor suppressor gene products: p53. *Virology* 384:285–293.
- Levine AJ, Tomasini R, McKeon FD, Mak TW, Melino G. 2011. The p53 family: guardians of maternal reproduction. *Nat Rev Mol Cell Biol* 12:259–265.
- Li J, Poi MJ, Tsai MD. 2011. Regulatory mechanisms of tumor suppressor P16(INK4A) and their relevance to cancer. *Biochemistry* 50:5566–5582.
- Li T, Kon N, Jiang L, Tan M, Ludwig T, Zhao Y, Baer R, Gu W. 2012. Tumor suppression in the absence of p53-mediated cell-cycle arrest, apoptosis, and senescence. *Cell* 149:1269–1283.
- Linzer DI, Levine AJ. 1979. Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell* 17:43–52.
- Malcikova J, Pavlova S, Kozubik KS, Pospisilova S. 2014. TP53 mutation analysis in clinical practice: lessons from chronic lymphocytic leukemia. *Hum Mutat* 35:663–671.
- Masuda H, Miller C, Koeffler HP, Battifora H, Cline MJ. 1987. Rearrangement of the p53 gene in human osteogenic sarcomas. *Proc Natl Acad Sci USA* 84:7716–7719.
- Meek DW, Anderson CW. 2009. Posttranslational modification of p53: cooperative integrators of function. *Cold Spring Harb Perspect Biol* 1:a000950.
- Moody CA, Laimins LA. 2010. Human papillomavirus oncoproteins: pathways to transformation. *Nat Rev Cancer* 10:550–560.
- Natan E, Hirschberg D, Morgner N, Robinson CV, Fersht AR. 2009. Ultraslow oligomerization equilibria of p53 and its implications. *Proc Natl Acad Sci USA* 106:14327–14332.
- Ng PC, Henikoff S. 2006. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* 7:61–80.
- Nguyen TA, Menendez D, Resnick MA, Anderson CW. 2014. Mutant TP53 posttranslational modifications: challenges and opportunities. *Hum Mutat* 35:738–755.
- Ogino S, Fuchs CS, Giovannucci E. 2012. How many molecular subtypes? Implications of the unique tumor principle in personalized medicine. *Expert Rev Mol Diagn* 12:621–628.
- Oliner JD, Pietenpol JA, Thiagalingam S, Gyuris J, Kinzler KW, Vogelstein B. 1993. Oncoprotein MDM2 conceals the activation domain of tumour suppressor p53. *Nature* 362:857–860.
- Olsson A, Manzl C, Strasser A, Villunger A. 2007. How important are post-translational modifications in p53 for selectivity in target-gene transcription and tumour suppression? *Cell Death Differ* 14:1561–1575.
- Ory K, Legros Y, Auguin C, Soussi T. 1994. Analysis of the most representative tumour-derived p53 mutants reveals that changes in protein conformation are not correlated with loss of transactivation or inhibition of cell proliferation. *EMBO J* 13:3496–3504.
- Ozturk M. 1991. p53 mutation in hepatocellular carcinoma after aflatoxin exposure. *Lancet* 338:1356–1359.
- Pleasant ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin ML, Beare D, Lau KW, Greenman C, Varela I, Nik-Zainal S, et al. 2010. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 463:184–190.
- Poon SL, Pang ST, McPherson JR, Yu W, Huang KK, Guan P, Weng WH, Siew EY, Liu Y, Heng HL, Chong SC, Gan A, et al. 2013. Genome-wide mutational signatures of aristolochic acid and its application as a screening tool. *Sci Transl Med* 5:197ra101.
- Pugh TJ, Weeraratne SD, Archer TC, Pomeranz Krummel DA, Auclair D, Bochicchio J, Carneiro MO, Carter SL, Cibulskis K, Erlich RL, Greulich H, Lawrence MS, et al. 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487:330–337.
- Rajagopalan S, Huang F, Fersht AR. 2011. Single-molecule characterization of oligomerization kinetics and equilibria of the tumor suppressor p53. *Nucleic Acids Res* 39:2294–2303.



- Reva B, Antipin Y, Sander C. 2011. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* 39:e118.
- Ribeiro RC, Sandrini F, Figueiredo B, Zambetti GP, Michalkiewicz E, Lafferty AR, DeLacerda L, Rabin M, Cadwell C, Sampaio G, Cat I, Stratakis CA, et al. 2001. An inherited p53 mutation that contributes in a tissue-specific manner to pediatric adrenal cortical carcinoma. *Proc Natl Acad Sci USA* 98:9330–9335.
- Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G, Harris S, Shah RR, et al. 2013. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* 45:970–976.
- Rodenhuis S, Slebos RJ, Boot AJ, Evers SG, Mooi WJ, Wagenaar SS, van Bodegom PC, Bos JL. 1988. Incidence and possible clinical significance of K-ras oncogene activation in adenocarcinoma of the human lung. *Cancer Res* 48:5738–5741.
- Saller E, Tom E, Brunori M, Otter M, Estreicher A, Mack DH, Iggo R. 1999. Increased apoptosis induction by 121F mutant p53. *EMBO J* 18:4424–4437.
- Sauna ZE, Kimchi-Sarfaty C. 2011. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* 12:683–691.
- Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM. 1990. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* 63:1129–1136.
- Schetter AJ, Harris CC. 2012. Tumor suppressor p53 (TP53) at the crossroads of the exposome and the cancer genome. *Proc Natl Acad Sci USA* 109:7955–7956.
- Schlomm T, Iwers L, Kirstein P, Jessen B, Kollermann J, Minner S, Passow-Drolet A, Mirlacher M, Milde-Langosch K, Graefen M, Haese A, Steuber T, et al. 2008. Clinical significance of p53 alterations in surgically treated prostate cancers. *Mod Pathol* 21:1371–1378.
- Selivanova G. 2010. Therapeutic targeting of p53 by small molecules. *Semin Cancer Biol* 20:46–56.
- Slade N, Moll UM, Brdar B, Zoric A, Jelakovic B. 2009. p53 mutations as fingerprints for aristolochic acid: an environmental carcinogen in endemic (Balkan) nephropathy. *Mutat Res* 663:1–6.
- Soussi T. 2011a. TP53 mutations in human cancer: database reassessment and prospects for the next decade. *Adv Cancer Res* 110:107–139.
- Soussi T. 2011b. Advances in carcinogenesis: a historical perspective from observational studies to tumor genome sequencing and TP53 mutation spectrum analysis. *Biochim Biophys Acta* 1816:199–208.
- Soussi T. 2014. Locus-specific databases in cancer: what future in a post-genomic era? The TP53 LSDB paradigm. *Hum Mutat* 35:643–653.
- Soussi T, Asselain B, Hamroun D, Kato S, Ishioka C, Claustres M, Beroud C. 2006a. 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.
- Soussi T, Beroud C. 2001. Assessing TP53 status in human tumours to evaluate clinical outcome. *Nat Rev Cancer* 1:233–240.
- Soussi T, Beroud C. 2003. Significance of TP53 mutations in human cancer: a critical analysis of mutations at CpG dinucleotides. *Hum Mutat* 21:192–200.
- Soussi T, Ishioka C, Claustres M, Beroud C. 2006b. Locus-specific mutation databases: pitfalls and good practice based on the p53 experience. *Nat Rev Cancer* 6:83–90.
- 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, Leroy B, Taschner PEM. 2014. Recommendations for analyzing and reporting TP53 gene alterations in the high throughput sequencing era. *Hum Mutat* 35:766–778.
- Staib F, Hussain SP, Hofseth LJ, Wang XW, Harris CC. 2003. TP53 and liver carcinogenesis. *Hum Mutat* 21:201–216.
- Stommel JM, Marchenko ND, Jimenez GS, Moll UM, Hope TJ, Wahl GM. 1999. A leucine-rich nuclear export signal in the p53 tetramerization domain: regulation of subcellular localization and p53 activity by NES masking. *EMBO J* 18:1660–1672.
- Stratton MR. 2011. Exploring the genomes of cancer cells: progress and promise. *Science* 331:1553–1558.
- Stratton MR, Campbell PJ, Futreal PA. 2009. The cancer genome. *Nature* 458:719–724.
- Supek F, Minana B, Valcarcel J, Gabaldon T, Lehner B. 2014. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* 156:1324–1335.
- Takahashi T, Nau MM, Chiba I, Birrer MJ, Rosenberg RK, Vinocour M, Levitt M, Pass H, Gazdar AF, Minna JD. 1989. p53: a frequent target for genetic abnormalities in lung cancer. *Science* 246:491–494.
- Terrier O, Bourdon JC, Rosa-Calatrava M. 2013. p53 protein isoforms: key regulators in the front line of pathogen infections? *PLoS Pathog* 9:e1003246.
- Tornaletti S, Pfeifer GP. 1995. Complete and tissue-independent methylation of CpG sites in the p53 gene: implications for mutations in human cancers. *Oncogene* 10:1493–1499.
- Toyooka S, Tsuda T, Gazdar AF. 2003. The TP53 gene, tobacco exposure, and lung cancer. *Hum Mutat* 21:229–239.
- Valente LJ, Gray DH, Michalak EM, Pinon-Hofbauer J, Egle A, Scott CL, Janic A, Strasser A. 2013. p53 efficiently suppresses tumor development in the complete absence of its cell-cycle inhibitory and proapoptotic effectors p21, puma, and noxa. *Cell Rep* 3:1339–1345.
- Varley JM, Chapman P, McGown G, Thorncroft M, White GR, Greaves MJ, Scott D, Spreadborough A, Tricker KJ, Birch JM, Evans DG, Reddel R, et al. 1998. Genetic and functional studies of a germline TP53 splicing mutation in a Li-Fraumeni-like family. *Oncogene* 16:3291–3298.
- Vassilev LT. 2004. Small-molecule antagonists of p53-MDM2 binding: research tools and potential therapeutics. *Cell Cycle* 3:419–421.
- Vousden KH, Lane DP. 2007. p53 in health and disease. *Nat Rev Mol Cell Biol* 8:275–283.
- Wade M, Li YC, Wahl GM. 2013. MDM2, MDMX and p53 in oncogenesis and cancer therapy. *Nat Rev Cancer* 13:83–96.
- Wang GS, Cooper TA. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 8:749–761.
- Weigelt B, Geyer FC, Reis-Filho JS. 2010. Histological types of breast cancer: how special are they? *Mol Oncol* 4:192–208.
- Wrighton KH. 2009. Small RNAs: p53 makes microRNAs mature. *Nat Rev Cancer* 9:612.
- Xu-Monette ZY, Medeiros LJ, Li Y, Orlowski RZ, Andreeff M, Bueso-Ramos CE, Greiner TC, McDonnell TJ, Young KH. 2012. Dysfunction of the TP53 tumor suppressor gene in lymphoid malignancies. *Blood* 119:3668–3683.
- You YH, Szabo PE, Pfeifer GP. 2000. Cyclobutane pyrimidine dimers form preferentially at the major p53 mutational hotspot in UVB-induced mouse skin tumors. *Carcinogenesis* 21:2113–2117.
- Zhao CY, Grinkevich VV, Nikulenkov F, Bao W, Selivanova G. 2010. Rescue of the apoptotic-inducing function of mutant p53 by small molecule RITA. *Cell Cycle* 9:1847–1855.
- Zheng R, Blobel GA. 2010. GATA transcription factors and cancer. *Genes Cancer* 1:1178–1188.
- Zupnick A, Prives C. 2006. Mutational analysis of the p53 core domain L1 loop. *J Biol Chem* 281:20464–20473.