Recommendations for Analyzing and Reporting *TP53* Gene Variants in the High-Throughput Sequencing Era

Thierry Soussi,[1,2]* Bernard Leroy,[2] and Peter E.M. Taschner[3]

[1]*Department of Oncology-Pathology, Cancer Center Karolinska (CCK), Karolinska Institute, Stockholm, Sweden;* [2]*Université Pierre et Marie Curie-Paris 6, Paris 75005, France;* [3]*Department of Human Genetics S-4-P, Leiden University Medical Center, Leiden, The Netherlands*

ABSTRACT: The architecture of *TP53*, the most frequently mutated gene in human cancer, is more complex than previously thought. Using *TP53* variants as clinical biomarkers to predict response to treatment or patient outcome requires an unequivocal and standardized procedure toward a definitive strategy for the clinical evaluation of variants to provide maximum diagnostic sensitivity and specificity. An intronic promoter and two novel exons have been identified resulting in the expression of multiple transcripts and protein isoforms. These regions are additional targets for mutation events impairing the tumor suppressive activity of *TP53*. Reassessment of variants located in these regions is needed to refine their prognostic value in many malignancies. We recommend using the stable Locus Reference Genomic reference sequence for detailed and unequivocal reports and annotations of germ line and somatic alterations on all *TP53* transcripts and protein isoforms according to the recommendations of the Human Genome Variation Society. This novel and comprehensive description framework will generate standardized data that are easy to understand, analyze, and exchange across various cancer variant databases. Based on the statistical analysis of more than 45,000 variants in the latest version of the UMD *TP53* database, we also provide a classification of their functional effects ("pathogenicity").

Hum Mutat 35:766–778, 2014. © 2014 Wiley Periodicals, Inc.

KEY WORDS: *TP53*; p53: mutation nomenclature; cancer; recommendations; annotation

The *TP53* Gene 30 Years Later: From Simplicity to Complexity

Our increasing knowledge of the *TP53* gene (MIM #191170) can stand as a paradigm for our evolving perception of a transcriptional unit over the last decade. The roots of the idea that one gene encodes one protein are to be found, with few exceptions, in the analysis of prokaryotes and lower eukaryotes. However, large-scale analyses using high-throughput methodologies have turned this concept upside down. Indeed, the latest issue of the ENCODE project suggested that 10–12 protein isoforms could be expressed by each human gene [Gerstein et al., 2012]. The *TP53* gene is a model for this revolution in knowledge with the discovery of its complex architecture involving different mechanisms to transcribe at least eight mRNAs and translate up to 12 different protein isoforms (Table 1) [Bourdon et al., 2005]. The present review has three parts: in the first, we discuss how these novelties can affect the detection and analysis of *TP53* variant status in human tumors. In the second, we discuss the reporting and classification of *TP53* variants. In the third, we provide specific recommendations for the detection and reporting of *TP53* variants.

Although this review is focused on the *TP53* gene, the recommendations and many issues discussed here are equally valid for other cancer genes, which may also require renewed investigation and variant effect reassessment.

TP53 Role in Cancer

The *TP53* gene is not only the most frequently mutated gene in human cancer, it is also a gene that acts as through a signaling hub, integrating a plethora of upstream signals and orienting them to various effectors pathways (reviewed by [Vousden and Prives, 2009; Levine et al., 2011]). Furthermore, *TP53* activities are largely dependent on the cellular context and the tissue of origin. Homozygote *TP53* knockout mice do not die in utero, an attribute that is not found in other common tumor suppressor genes such as *APC*, *PTEN*, *RB1*, or *BRCA1/2* [Taneja et al., 2011]. Furthermore, the predisposition to cancer in *TP53* knockout mice starts largely after sexual maturity, suggesting that *TP53*, paradoxically, is not as essential as other tumor suppressors for individual survival [Kenzelmann Broz and Attardi, 2010; Jackson and Lozano, 2013].

Germ line *TP53* variants are associated with predisposition to various types of hereditary cancer, including familial breast cancer, Li–Fraumeni syndrome and pediatric adrenocortical carcinoma [Custodio et al., 2013; Kamihara et al., 2014]. The association of a specific *TP53* variant with pediatric adrenocortical carcinoma is currently unexplained and mouse models are urgently needed to assess its mechanisms in vivo. As reviewed by Donehower in this issue, knockin mice expressing different *TP53* hotspot mutants display heterogeneous tumor phenotypes [Donehower, 2014]

Table 1. TP53 Isoforms, Transcripts, and Proteins[a]

Common protein name[b]	LRG_t[c]	LRG_p[d]	NCBI transcript	NCBI_Protein	Residues (kDa)[e]	Promoter	Major splicing event[f]
Full-length p53, p53, p53α	t1	p1	NM_000546.5	NP_000537.3	393/43.6	P1/P1′	I
Full-length p53, p53, p53α	t2[g]	p1	NM_001126112.2	NP_001119584.1	393/43.6	P1/P1′	I
p53β, p53i9	t3	p3	NM_001126114.2	NP_001119586.1	341/37.9	P1/P1′	II
p53γ	t4	p4	NM_001126113.2	NP_001119585.1	346/38.5	P1/P1′	III
Δ40p53α, ΔNp53, p47	t1	p8	NM_001276760.1	NP_001263689.1	354/39.3	P1/P1′	I
Δ40p53α, ΔNp53, p47	t2	p8	NM_001276761.1	NP_001263690.1	354/39.3	P1/P1′	I
Δ40p53α, ΔNp53, p47	t8[h]	p8	NM_001126118.1	NP_001119590.1	354/39.3	P1/P1′	I
Δ40p53β	t3	p9	NM_001263625.1	NP_001263625.1	302/33.5	P1/P1′	II
Δ40p53γ	t4	p10	NM_001276695.1	NP_001263624.1	307/34.1	P1/P1′	III
Δ133p53α	t5	p5	NM_001126115.1	NP_001119587.1	261/59.6	P2	I
Δ133p53β	t6	p6	NM_001126116.1	NP_001119588.1	209/23.7	P2	II
Δ133p53γ	t7	p7	NM_001126117.1	NP_001119589.1	214/24.4	P2	III
Δ160p53α	t5	p11	NM_001276697.1	NP_001263626.1	234/26.6	P2	I
Δ160p53β	t6	p12	NM_001276698.1	NP_001263627.1	182/20.7	P2	II
Δ160p53γ	t7	p13	NM_001276699.1	NP_001263628.1	187/21.4	P2	III

[a]Equivalence between the various identifiers used to describe the various isoforms of the *TP53* gene.
[b]Other names found in the literature for the various TP53 isoforms.
[c]LRG identifiers for the various TP53 transcripts. A description of the LRG identifiers is described in the text and can be visualized at ftp://ftp.ebi.ac.uk/pub/databases/lrgex/LRG_321.xml.
[d]LRG identifiers for the various TP53 proteins.
[e]Theoretical molecular weight.
[f]Major splicing events: I: classical splicing with all exons including exon 9; II: splicing event including exon 9 β; III: splicing event including exon 9 γ.
[g]Transcript t2 has a deletion of 3 nucleotides at the beginning of exon 2.
[h]Transcript t8 retains intron 2.

The Structure of the *TP53* Gene

The human *TP53* gene organization as depicted over the past 20-plus years was simple: it produced a transcript containing 11 exons encoding a single protein of 393 amino acids [Soussi and May, 1996]. The structural organization of the gene is well conserved through evolution, although it comprises specific features that are still not fully understood. First, in humans and mice, the large intron (10 kb) between the noncoding exon 1 and exon 2 containing the first ATG codon encodes hp53int1, a small untranslated RNA of unknown function (Fig. 1) [Reisman et al., 1996]. A second particularity is the recently identified *WRAP53* gene (MIM #612661) that partially overlaps the 5′ region of the *TP53* gene in a head-to-head configuration (Fig. 1) [Mahmoudi et al., 2009]. Several *WRAP53* transcripts overlap *TP53* exon 1 as well as hp53int1, suggesting the possibility of regulation via the formation of double-stranded RNA. The consequences of this specific organization which is conserved in mammals are currently unknown. The fact that a similar gene is localized in the 5′ region of the *TP73* gene (but not in the *TP63* gene) suggests that it has some important regulatory function for *TP53* and *TP73* [Mahmoudi et al., 2009].

The full-length (393 aa) TP53 protein (TP53 also known as TP53 α or p1) translated from the major mRNA species initiated from promoter 1 (P1) upstream of exon 1 remains the most abundant isoform. The functional organization of the *TP53* gene is more complex than previously thought: the NCBI's RefSeq database now contains 15 different pairs of *TP53* transcript and protein records references due to the policy to associate only one RNA species to a single protein (Table 1 and Supp. Fig. S1). Thus, several mRNA species encoding more than one protein have been duplicated with different RefSeq NM-accession numbers and two protein isoforms are represented by multiple RefSeq NP-accession numbers. To solve this confusing situation, *TP53* specialists have joined forces with the Locus Reference Genomic (LRG) Consortium, which provides stable reference sequences and a coordinate system for permanent and unambiguous reporting of disease-causing variants in genes related to any pathology [Dalgleish et al., 2010; MacArthur et al., 2014]. LRGs already cover 715 genes associated with noncancerous or cancerous diseases.

Their records in Entrez Gene (http://www.ncbi.nlm.nih.gov/gene/) and Ensembl (http://www.ensembl.org/index.html) contain links to the corresponding LRGs. The UCSC Genome Browser Website (http://genome.ucsc.edu/) provides the LRG Regions track under the Mapping and Sequencing header and an LRG Transcripts track under the Genes and Gene Predictions header (January 2014). The joint effort resulted in a recently released stable *TP53* reference sequence, LRG_321 containing the genomic sequence from human genome build GRCh37.p13 (ftp://ftp.ebi.ac.uk/pub/databases/lrgex/LRG_321.xml). We believe that its annotation with precise labels and coordinates of eight different *TP53* transcripts (t1–t8) and 12 isoforms (p1 and p3 to p13) will be preferred to the RefSeq identifier pairs provided by the NCBI for genome build GRCh37.p13 (Table 1 and Supp. Fig. S1). Therefore, we will use LRG transcript and protein isoform numbers throughout this review. Without further specification, exon and intron numbering refers to transcript LRG_321t1; amino-acid positions to full-length LRG_321p1.

The *TP53* gene produces all TP53 isoforms in a combinatorial manner using an alternative promoter, alternative splicing, and alternative translation start sites (Table 1 and Figs. 1 and 2) [Bourdon et al., 2005]. Their amino-termini are determined by the use of two different promoters each producing transcripts with two translation start sites, except for t8. Three different carboxy-termini can be generated by the exclusion or inclusion of two newly discovered exons, 9 β and 9 γ, localized in intron 9. Each of the four different amino-termini can be combined with each of the three carboxy-termini leading to 12 isoforms (Table 1 and Figs. 2 and 3). We will discuss this in more detail, starting with the amino-termini produced by P1 and P2 transcripts. The translation of the four P1-initiated *TP53* mRNAs t1–t4 can begin at codon 1 or codon 40 leading to the expression of either full-length TP53 protein (TP53) or TP53 protein truncated of the first transactivation domain (Delta40p53). Furthermore, Delta40p53α proteins can also be encoded by alternatively spliced *TP53* mRNA t8 still containing intron 2. It is important to note that intron 2 included in transcript t8 contains a stop codon that prevents expression of full-length p53. This alternative splicing of TP53 intron 2 has been described in several cell lines and normal
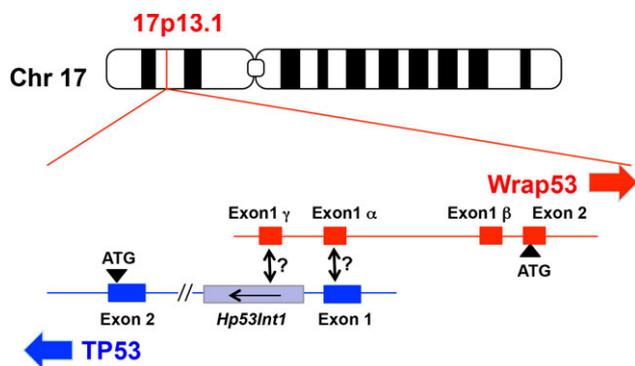
Figure 1. Organization of the 5′ ends of the *TP53* and *WRAP53* genes. The two genes have a head-to-head configuration (Mahmoudi et al., 2009). The overlap between the transcripts (depicted by a double arrow) would suggest that both genes could regulate the other via an antisense mechanism. Intron 1 of the *TP53* gene is 10 kb long.
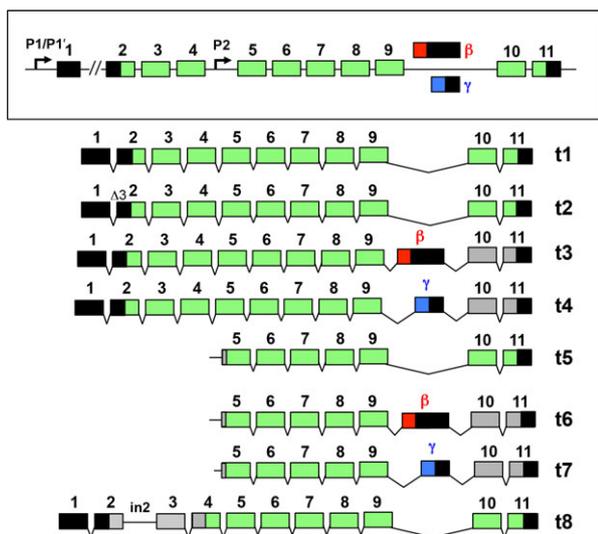

Figure 2. Transcriptional organization of the *TP53* gene. The *TP53* gene (upper part of the figure) is transcribed into eight different mRNAs. Transcripts t1 to t4 originate from promoter P1 and P1′ localized upstream of the gene. Transcripts t5 to t8 originate from promoter P2 localized in intron 4. Translated exons are shown in green. The two novel exons β and γ are shown in red and blue, respectively. Nontranslated region are shown in black. For transcripts t3, t4, t6, and t7, which include exon β or γ, exons 10 and 11 are noncoding (gray boxes). Transcript t8 encodes only p8 (DeltaTP53α) and exons 1 to 3 are noncoding (gray boxes). Proteins translated from the various transcripts are described in Table 1 and Figure 3.

human lymphocytes [Matlashewski et al., 1987; Ghosh et al., 2004]. The second group with different amino-termini contains *TP53* isoforms encoded by three novel mRNAs: t5, t6, and t7. These transcripts are produced by the P2 promoter localized in intron 4 and probably extending to exon 3 and start from a novel transcription initiation site at the 3′ end of intron 4 (Figs. 2 and 3). They generate *TP53* proteins with different amino-termini starting either at amino acid 133 or 160 (Delta133p53 and Delta160p53) Aoubala et al. (2011). Finally, each of the different amino-termini can be combined with a different carboxy-terminus. The alternatively spliced transcripts containing exon 9 β, t3 and t6, use two translation start sites each to encode 4 TP53 isoforms combining different amino-termini
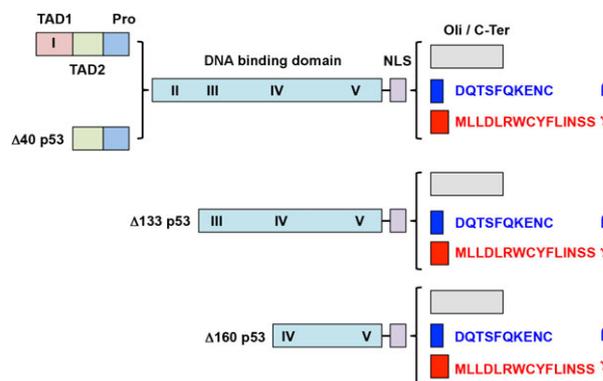

Figure 3. Comparison of domains in various *TP53* isoforms. TAD1: transactivation domain 1; TAD2: transactivation domain 2; Pro: proline-rich domain; NLS: nuclear localization signal; Oli: oligomerization domain; C-ter: carboxy-terminus domain of TP53 containing multiple post-translational modification sites. Only one of the three NLS is inside the core region common to all TP53 isoforms. The two other NLS are localized in the C-ter region.

with the β carboxy-terminus (TP53β, Delta40p53β, Delta133p53β, and Delta160p53β). Similarly, t4 and t7 containing exon 9 γ encode four TP53 isoforms combining different amino-termini with the γ carboxy-terminus (TP53γ, Delta40p53γ, Delta133p53γ, and Delta160p53γ). Although expression at the physiological level of intron 2 containing t8 equivalents with exons 9 β and 9 γ remains to be demonstrated experimentally, these would probably encode the existing Delta40p53β and Delta40p53γ isoforms.

TP53 Variant Screening in Human Cancer: a Necessity

In human cancer, *TP53* is the gene most frequently hit by somatic mutation events [Leroy et al., 2014a; Leroy et al., 2014b]. Questions concerning particular *TP53* variant spectra can be raised in several malignancies, such as ovarian cancer or basal subtypes of breast cancer, as they display more frameshift and nonsense variants than other cancer types [Curtis et al., 2012]. We currently do not know if these variants are the consequence of a specific defect in the DNA repair system of these tumors or more so the result of counter selection for the expression of mutant TP53. Osteosarcoma is one of the few cancers displaying a high frequency of *TP53* gene deletion, an observation already made in 1987 and confirmed more recently by whole genome analysis [Masuda et al., 1987; Barretina et al., 2010]. However, the basis of this gene deletion is currently unknown.

It is beyond the scope of this review to summarize the more than 20,000 publications describing correlations or lack of correlations between *TP53* variants and various clinical parameters such as response to treatment, overall or relapse-free survival and many others. The fact that tumors with *TP53* variants are associated with poorer prognosis has been repeatedly demonstrated for various types of cancer and recently confirmed in the pan-cancer study [Olivier and Taniere, 2011; Kandoth et al., 2013]. *TP53* variants have been shown to have value for prognosis and treatment orientation in chronic lymphocytic leukemia (CLL) and thus *TP53* screening has been recommended [Pospisilova et al., 2012; Malcikova et al., 2014]. The analysis of germline mutations in various cancer-prone families is also an essential clinical aspect of TP53 screening that goes beyond Li–Fraumeni syndrome [Kamihara et al., 2014]. TP53 germline mutations have also been observed in families at high risk

of breast cancer and TP53 screening has been recommended for individuals diagnosed with early onset breast cancer at age 35 years or younger, with a negative BRCA1/BRCA2 test.

Analyzing *TP53* variants in various types and subtypes of tumors will provide clues and permit the generation of working hypotheses on the pleiotropic functions of this gene and its products. The identification of variant hot and cold spots in specific cancer types is essential for a better understanding of both the structure/function relationship of the TP53 protein and tumor etiology. Combined with information about treatment outcome, this should improve the prognostic and predictive value of TP53 variant biomarkers.

TP53 Variant Spectrum and Isoforms

The first *TP53* variants discovered affect the highly conserved domain of the protein, encoded by exons 5 to 8 of transcript t1 [Nigro et al., 1989; Takahashi et al., 1989]. These early observations caused a strong bias as the majority of the subsequent studies were performed by screening only the central region of the gene. This vicious circle led to the general belief that only a few variants were localized outside of the exons 5–8. Most sequencing studies have focused their analyses on exons encoding the central regions of the proteins (exon 5–8). More recent studies have shown that at least 10%–15% of *TP53* variants are localized in exons 2–4 and exons 9–11 (Fig. 4) [Leroy et al., 2013; Leroy et al., 2014b]. Furthermore, the spectrum of these variants is different, as they consist mostly of small indels that usually lead to a *TP53* null phenotype.

The majority of *TP53* variants are localized in the central region of the full-length protein p1. Those localized in the amino- or carboxy-terminus will only affect the full-length TP53, sparing some isoforms (Fig. 4). Less than 1% of *TP53* variants will miss the three Delta40 isoforms (p8, p9, and p10), but 7.4% and 18.4% of the variants will lead to the synthesis of intact Delta133 (p5, p6, and p7) and Delta160 (p11, p12, and p13) isoforms, respectively (Fig. 4). Variants in exons 10 and 11 are less frequent (2.4%) but they do not target β (p3, p6, p9, and p12) and γ (p4, p7, p10, and p13) isoforms (Fig. 4). Somatic variants in exon 11 containing the shared 3′UTR might deregulate the *TP53* network [Li et al., 2013]. This could affect the expression of all transcripts. A few missense variants have been identified in exon 9 β and γ in several tumors.

The new transcripts and their isoforms raise important questions: Are any novel variant effects specific for the tumor they were observed in? Are these effects extendable and of importance to other types of cancer? What are the functional effects of a particular variant on each of the isoforms? Are all isoforms equally important for clinical decision making? Do variants targeting only a limited number of p53 isoforms lead to heterogeneous complexes with different consequences for the *TP53* network?

Variants appearing before codon 133 or after codon 331 of full-length *TP53* would lead to combined expression of variant and wild type (WT) TP53 protein isoforms from the same allele. Although loss of WT TP53 activity has been described, the biological consequences of the combined expression of both WT and variant isoforms in tumors are totally unknown. Because these isoforms can regulate gene expression, some may retain tumor suppressor activities in a mutant *TP53* gene context, whereas others may confer gain of function. TP53 isoforms may thus explain inconsistencies in the different biological activities described for mutant p53. It is essential to further investigate this interplay with the goal of improving the prognostic and predictive value of p53 variant biomarkers.

TP53 Gene Analysis

Use of Prescreening Strategies

Prescreening strategies such as SSCP, DGGE, or dHPLC have been of tremendous value for the detection of variants in clinical samples (Supp. Table S1). Although their specificity is variable, several offer better sensitivity than direct sequencing to detect variants in tumor tissue heavily contaminated with normal cells. These strategies can be easily extended to incorporate the new target regions. However, the spectacular decrease in the cost of conventional and next-generation sequencing (NGS), as well as automation and use of high-throughput capillary electrophoresis have alleviated the need of prescreening. It may even be advantageous to avoid the latter in order not only to get the best specificity but also to reduce costs and delays, which are important considerations in routine clinical practice.

Immunohistochemistry

It is commonly accepted that somatic *TP53* missense variants lead to the accumulation of the TP53 protein in the tumor cell nucleus. Thus, tumor staining using various TP53 monoclonal antibodies has developed as a surrogate for *TP53* variant analysis [Bartek et al., 1991]. Several thousand articles describing TP53 staining in a wide variety of tumors have been published and several reviews have addressed the various problems associated with this indirect screening technique [Hall and Lane, 1994; Hall and McCluggage, 2006]. Reviewing these studies is beyond the scope of this article, but it has been largely demonstrated that the relation between TP53 accumulation and *TP53* variants is not straightforward. Although truncated TP53 proteins resulting from nonsense and frameshift variants cannot be assessed via immunohistochemistry, it has been repeatedly reported that some tumors with missense variants could also be negative for TP53 accumulation [Casey et al., 1996; Alsner et al., 2008; Gluck et al., 2012]. In contrast, tumors without *TP53* variants have been reported with TP53 protein accumulation. We strongly advocate against TP53 staining as a screening methodology for *TP53* variants. Any possible clinical value for the analysis of TP53 accumulation per se, without inferring the origin of such accumulation, remains to and should be determined, but this issue should not be conflated with *TP53* variant screening. It is also important to consider the monoclonal antibody used for the detection of TP53 accumulation. Most commercialized antibodies recognize an epitope localized in the amino-termini of the proteins (1–40) and will thus miss most isoforms [Legros et al., 1994; Tenaud et al., 1994]. Because of the presence of immunodominant epitopes in the TP53 proteins, polyclonal sera raised against human TP53 are highly biased toward the amino-terminus. Because several *TP53* isoforms have different subcellular localizations, it would be interesting to investigate, via IHC using a panel of antibodies, any possible associations between their subcellular localization (nuclear, cytoplasmic, or speckled staining) and disease outcome, and any possible role for subcellular localization in the stratification of cancers into subtypes [Ghosh et al., 2004].

Variant Arrays and Sensitivity of TP53 Analysis

Variant arrays mainly detect known sequence alterations. The *TP53* variant spectrum is quantitatively and qualitatively very heterogeneous across the various types of cancer. In colorectal carcinoma, the high frequency of transitions localized at methylated CpG dinucleotides leads to a high clustering of variants at codons 175, 248, 273, and 285. This clustering is also observed
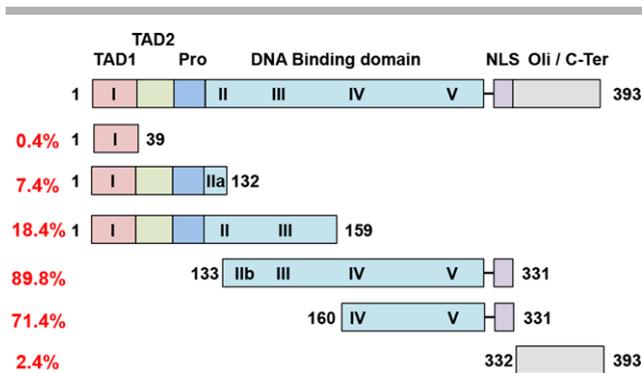
Figure 4. Frequency of *TP53* variants in the various domains that characterize the different TP53 isoforms.
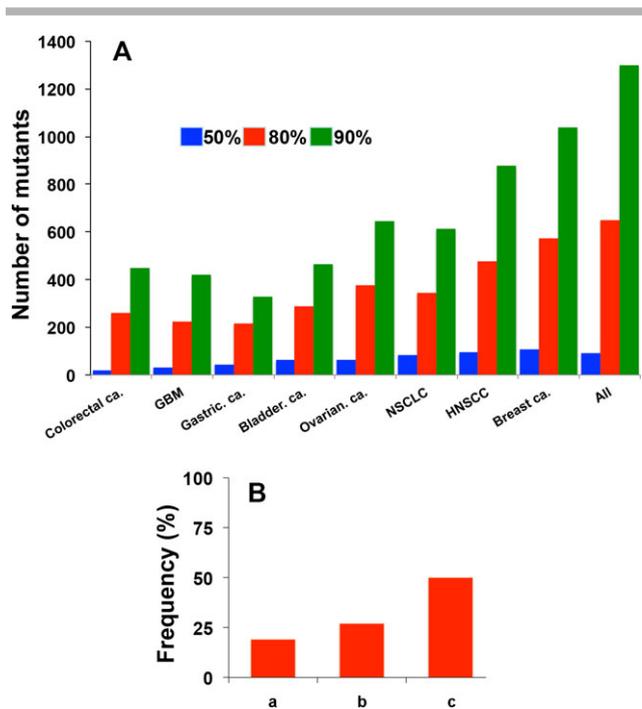

Figure 5. Sensitivity index for the detection of *TP53* variants. **A**: Number of variants to be assessed for each type of cancer (or for the entire database), to reach sensitivities of 50%, 80%, and 90% (GBM: glioblastoma; NSCLC: nonsmall cell lung cancer; HNSCC: head and neck squamous cell carcinoma; All: entire TP53 database). **B**: Sensitivity of three different commercially available arrays (**a**–**c**, December 2013).

in astrocytomas and glioblastomas, but not for other types of cancer. Using the 45,000 *TP53* variants included in the database, it is possible to calculate a sensitivity index to establish the number of variants that need to be screened for in order to attain a given sensitivity (Fig. 5). Looking at all cancers in the *TP53* database shows that 1,400 variations must be assessed to reach a sensitivity of 90% (Fig. 5). However, when looking at any one type of cancer, the variant screening number can vary widely, from 450 in colorectal carcinoma to 1,040 in breast carcinoma, depending on variant spread and the heterogeneity of the various mutation events. Several commercial arrays are currently available for the detection of *TP53* variants but none attain a sensitivity higher than 50%. Arrays will have to be redesigned to cover new target regions. Whether an array interrogating every position to cope with the highly scattered

nature of *TP53* variants would be more effective than sequencing in terms of specificity, sensitivity, or cost is still an open question.

Sample Origin

Formalin fixation plus paraffin embedding (FFPE) is one of the most widely practiced methods of clinical sample preservation and archiving. It has been estimated that there are over 400 million FFPE tissue samples archived in tissue banks worldwide, a gold mine for genomic analyses [Sah et al., 2013]. Although changing therapeutic protocols have reduced the clinical value of some collections, their intrinsic value is inestimable for the definition of variant profiles. These samples will be essential for novel studies such as the 10K project, which is focused on sequencing more than 100,000 tumor specimens (http://news.sciencemag.org/biology/2013/03/ready-more-10000-cancer-genomes-projects). However, DNA extraction from FFPE samples is challenging. Common problems such as formaldehyde cross-linking, degradation, and mixing of single-stranded and double-stranded DNA result in fragmented DNA of variable quality. The age of the sample and the quality of storage have also been shown to be important parameters for DNA quality. Compared with the frozen tissue, sequencing DNA extracted from FFPE samples requires more accurate quality controls; indeed several studies have reported artifact variants associated with this type of material [Williams et al., 1999; Marchetti et al., 2006; Soussi et al., 2006].

Although novel methodologies are available to improve the quality and the yield of DNA extracted from FFPE tissue, the risk of sequencing artifacts will remain greater than when using DNA from frozen tissue. For NGS, it will be essential to develop accurate algorithms to distinguish method- or sample quality-related sequencing background errors before assessing variant calling procedures. Another key point with FFPE tissue is the size of the specimen. Large tissue samples, harvested for example during surgery, permit the detection of most tumor variants, including those in the various subclones that sustain tumor heterogeneity. In contrast, PFFE samples, often obtained from local biopsies, can be very small and thus may provide only a limited view of the various genetic alterations. This aspect is not genuinely problematic for somatic *TP53* variants that occur early during transformation because they spread throughout the tumor. However, it is problematic for somatic variants arising by mutation events at later stages and thus only present in subclones that may not be represented in a small specimen of the tumor.

The Genetic Material

Both genomic DNA and cDNA derived from mRNA have been used to infer *TP53* mutational status in human tumor or cell lines. Sequencing genomic DNA allows unequivocal identification of any primary change, also in promoter and intronic regions. Sequencing cDNA limits variant detection to transcribed regions, but will elucidate the effects of primary changes at the RNA level supporting better predictions of effects at the isoform level. As discussed by Leroy et al. (2014a), cDNA sequencing is associated with potential artifactual results as point variants that affect splicing lead to abnormally spliced RNA species usually quoted as deletions.

Screening and Analyzing the *TP53* Gene in the Postgenomic Era

The discovery of the various *TP53* isoforms, the two novel exons localized in intron 9 and the promoter in intron 4 suggests that the

conventional screening strategy is inadequate and must be extended with these regions [Bourdon et al., 2005] (Fig. 6). The 3′UTR of the *TP53* gene forms another target for potential alterations. In patients with B-cell lymphoma, somatic variants in the 3′UTR disrupt the interaction between *TP53* mRNA and *miR-125b*, possibly resulting in deregulation of the *TP53* network [Li et al., 2013]. These new functional elements raise many questions, some of which can be answered using new technology: What is the spectrum of variants hitting them? Are all transcripts and isoforms expressed in all tissues and relevant for all types of cancer? What are the qualitative and quantitative effects of variants on each of the transcripts and isoforms?

A more complete picture might emerge using RNAseq with NGS and accurate bioinformatics analysis. Although historically, splicing *TP53* variants have been thought to occur infrequently, recent large-scale analyses have suggested that they may be underestimated and may represent 2%–4% of total variants. Variants localized at the border of exons and predicted to be synonymous have been shown to affect normal splicing [Leroy et al., 2014a]. The combination of DNA and RNA analysis should permit the detection of variants and identification of normally spliced and all aberrant mRNA species resulting from them. The expression pattern of the various *TP53* transcripts is likely tissue-specific, necessitating clear definition for each tissue sample and annotation of all relevant experimental conditions. Whether or not variants can qualitatively or quantitatively modify this complete transcript and isoform profile is currently unknown. Identification of transcript variants caused by splice defects may also be possible depending on how efficiently the new splice site and alternative sites are used, and the sequencing coverage. Such studies will be very informative as they will establish if any TP53 proteins, abnormal or normal, can be potentially translated or if these tumors should be considered as *TP53* null not expressing TP53 gain of function variants. RNAseq will also be very useful for the identification of exonic splicing enhancer or exonic splicing silencer regions, which modulate splicing [Wang and Cooper, 2007; Sauna and Kimchi-Sarfaty, 2011]. Their positions could be inferred from exonic variants hitting them and their consequences on *TP53* splicing. This would help to establish possible associations between their alteration and gene misregulation in human cancer.

Toward an Adequate TP53 Sequence Analysis Pipeline

Although several thousand tumor genomes from various types of cancer have been sequenced, data for intronic sequences, including the two novel exons of *TP53*, have not been analyzed and are not easily accessible. This is due to the filtering pipeline used by most studies, which causes variants in newly discovered exons awaiting annotation in the various databases to be missed (Fig. 7). Current strategies are largely biased toward coding regions, with the separation of the various variants in four tiers as follows: tier 1: variants altering coding sequences (nonsynonymous or synonymous) splice site, or noncoding RNA; tier 2: variants targeting conserved or regulatory sequences; tier 3: variants occurring in nonrepetitive regions of the human genome, including introns; and tier 4: variants occurring in repetitive noncoding regions [Ding et al., 2010]. In most studies, only tiers 1 and 2 are used to profile the mutational landscape of tumors. Mining various databases, which are not always freely available, for variants in specific regions to analyze raw sequence data will require considerable time and expertise. Exomic analysis via NGS is also currently biased since (commercial) exome capture kits must be upgraded to include β and γ exons.

The classical NGS sequence analysis pipeline for clinical samples has three phases (Fig. 7). After the base calling and variant calling phases, the variant filtering phase includes numerous steps requiring access to external references. Several tools can be used to annotate variants on all different transcripts. The last phase could use the LRG for unambiguous descriptions of variants at the different levels contributing to harmonization of *TP53* variant reporting as shown below.

Reporting *TP53* Variants

A Note on Terminology

For more than 15 years, the Human Genome Variation Society (HGVS) has provided guidelines for variant terminology and nomenclature. The HGVS recommends the use of the term "variant" instead of "mutation," "SNP" or "polymorphism" for sequence variants in general, regardless of their functional consequence or tissue of origin (see http://www.hgvs.org/mutnomen/). We would like to suggest the *TP53* community to embrace this recommendation for the following reasons.

Originally, SNP described a germ line variation that exists at a frequency of at least 1% in the general population (http://www.ncbi.nlm.nih.gov/books/NBK21088/). Such variations are the roots of diversity in species and tremendously useful as markers for genetic studies. Created in 1998, the dbSNP database maintained by the NIH keeps track of SNPs (http://www.ncbi.nlm.nih.gov/SNP/). In the literature, the term SNP has obtained an additional meaning associated with low or very limited risk on disease or tumor formation. Since 2011 (build 134), dbSNP started accepting submissions of germ line and somatic variations associated with various types of diseases and changed its name to "database of Short Genetic Variation" keeping the dbSNP acronym. Several frequent *TP53* variants (e.g., rs28934578) are included in dbSNP, but other hot spot variants are missing, whereas rare somatic variants can be found. This heterogeneity caused by biased dbSNP submissions is misleading, as it does not reflect the true occurrence and frequencies of *TP53* variants. Therefore, without further distinction, we can no longer assume that variants in dbSNP are associated with the lack of effect on disease and tumor characteristics. The mix of neutral and disease-causing variants in dbSNP has led to confusion and ambiguities in the *TP53* field and many others. The use of "SNP" and "polymorphism", which is normally associated with low or very limited risk on disease or tumor formation, for all variants in dbSNP could be detrimental for various types of analysis, potentially leading to the wrong clinical diagnosis. It is also one source of discrepancies between *TP53* variant databases, fueling discussions about variants being "true SNPs," "natural SNPs" or disease-causing "mutations." As all of them can no longer be regarded to meet the original definition, it would be better to refer to them as "dbSNP entries".

Classification of TP53 Variants

Although all *TP53* germ line variants have been detected in tumors, there is a clear need to distinguish them from somatic variants, as *TP53* variants may have different effects in normal tissue and tumors because of the various complex roles of TP53 isoforms. We would like to propose simplifying variant descriptions by indicating all variants observed in the germ line as a "germ line variants" and variants observed in tumors, but not present in normal tissue, as a "somatic variants." If normal tissue has not been examined, variants observed in tumor tissue could be labeled as "variant detected in
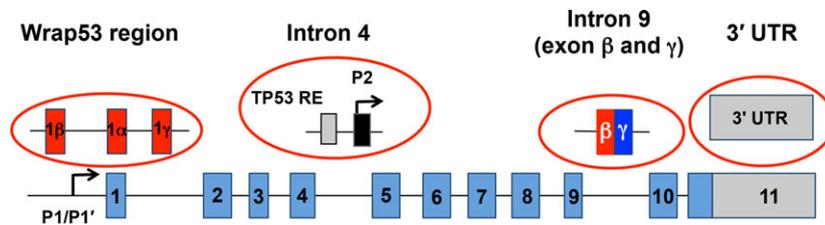
Figure 6. *TP53* gene screening in the postgenomic era. Four novel regions of the *TP53* gene must be included in screens for alterations to define their importance in tumorigenesis; i: the *WRAP53* region that includes overlapping exons with the *TP53* gene; ii: intron 4 with a *TP53* response element (*TP53* RE) and the P2 promoter expressing transcripts t5, t6, and t7; iii: intron 9 with the two novel exons β and γ; iv: the 3′UTR region containing sequences targeted by microRNA mir-125b.
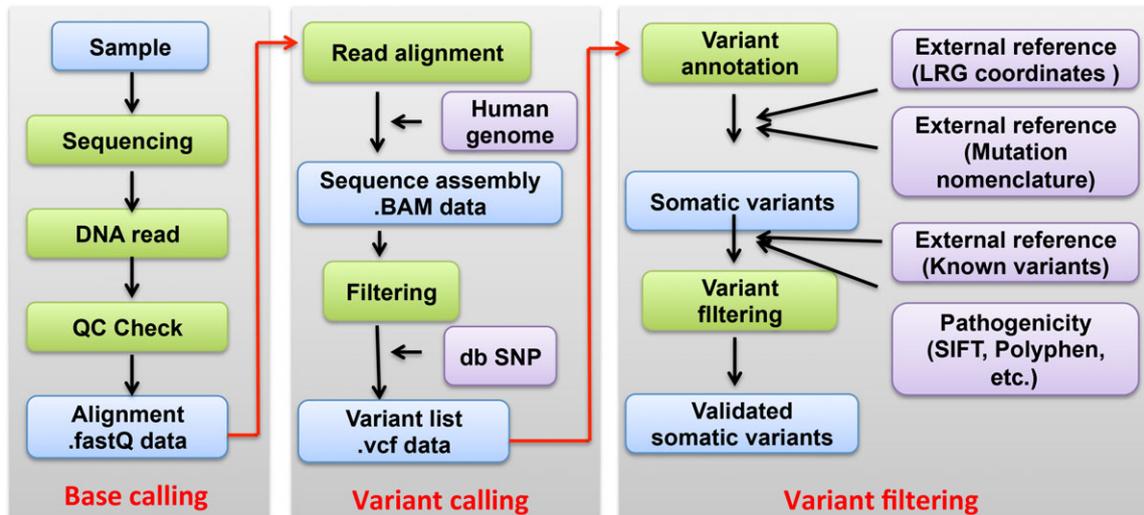
Figure 7. DNA sequencing pipeline used to identify somatic variants. The first phase (Base calling) is highly specific and linked to the sequencing procedure used to generate sequence reads. The second phase (Variant calling) aligns the sequence reads to the human reference genome (or to the sequence derived from normal tissues from the same patient) to construct a fully assembled sequence and infer all new variations by filtering against dbSNP. Most of these steps are very similar across various sequencing platforms. The final step annotates these variations using multiple external sources of references.

tumor." Thus, germ line variants detected in one individual can be described in others as "detected in tumor" or as somatic variants. Many clinical geneticists are using the five-class system for germ line variants [Plon et al., 2008]. We suggest using the same classification for somatic variants ranging from Class 5 (has functional consequences, "pathogenic") via Class 3 (unknown, Variant of Uncertain Significance [VUS]) to Class 1 (benign). Variants previously described as "passenger" or "hitch-hiking mutation" could be assigned to Classes 1–3, depending on supporting evidence. Clearly, this classification describes the functional consequences of the variant in isolation. Future refinement may be required to describe modifier effects when different variants occur in combination on the same or different alleles. Although the predicted functional consequences for germ line and somatic variants at the RNA and protein level will be the same, their clinical effects may differ. Additional classification may be necessary to distinguish between increased predisposition to tumor formation for the first and potential effects on tumor progression, prognosis, and treatment outcome for the latter.

Correct Sequence Variant Nomenclature

To keep up with our increasing knowledge of gene architecture, the HGVS has regularly published recommendations and guide-lines for variant nomenclature [Cotton and Malcolm, 1991; Claustres et al., 2002; Cotton et al., 2008; Auerbach et al., 2011]. The latest update of these recommendations, compiled by den Dunnen and Antonarakis (2000), is available on the HGVS Website (http://www.hgvs.org/mutnomen/). Nonetheless, the literature is plagued with fancy—but meaningless—or incomplete variant descriptions. During *TP53* variant database curation, we noticed a large degree of heterogeneity in published *TP53* variant descriptions with less than 20% following the official nomenclature despite numerous contacts with editors and publishers. Many "exotic" nomenclatures often hampered the accurate identification of variants. Furthermore, numerous studies contained typographical errors or incorrect reference sequences due to manual manipulation of the data. Several years ago, we contacted more than 20 journal editors to discuss the use of correct variant descriptions with them. Although they acknowledged the problem, currently merely four journals have initiated efforts to solve this for new manuscripts by making the use of HGVS variant nomenclature mandatory. Thus, we fear that incorrect variant descriptions, mostly related to manual sequencing, will remain permanently in existing literature. Ultimately these will be diluted by correct variant descriptions from computerized NGS data analyses including tools describing variant according to the official nomenclature. Although much less, mistakes can still occur because of the heterogeneous gene nomenclature

and numbering systems, combined with the very high number of automatically processed genes. We have noticed a number of "scrambled" *TP53* variant descriptions, mixing both the coordinates of the full-length protein and those of a particular isoform in a single list. Thus, harmonization of *TP53* variant reporting is urgently needed to accurately perform comparative cross-study analyses, and fully appreciate their pleiotropic effects and establish their relevance in clinical practice.

Numerous recent publications describe *TP53* variants only at the protein level, such as p.R175H or p.R248W. This trend is usually associated with the use of commercial or custom-made arrays specific to cancer gene variants. Using protein variant descriptions is highly confusing because the true genetic event cannot be correctly inferred. Because of the codon degeneracy, several mutation events can lead to the same amino-acid substitution. HGVS variant nomenclature in combination with the stable LRG_321 reference sequence can generate unambiguous *TP53* variant descriptions at the DNA level. In case no other variants are described, one-time specification of the LRG_321 reference sequence suffices to use transcript and protein isoform numbers. Thus, the TP53 hotspot variant LRG_321p1:p.R249S (in short p1:p.R249S), which can result from two different transversion events should have been described as either t1:c.747G>T or t1:c.747G>C. Unambiguous variant descriptions at the DNA level ensures that information can easily be transferred between various sources, for example, from publications to databases, or from one database to another. The LRG annotation supports automatic conversion of these descriptions to define the consequences of any variant for the eight transcripts and 12 protein isoforms. They would also support reconstruction of the sequence observed as input for bioinformatics tools predicting effects on splicing and other downstream processing at the RNA and protein level. For instance, the t1:c.314G>T variant previously predicted to result in amino-acid substitution p.(G105D) creates a new donor splice site (CAGGGCAGC to CAGg*tc*agc). This splices out the end of exon 4 and intron 4, leading to an in-frame deletion of the end of exon 4. The new transcript would be translated into a p53 protein with its second conserved domain deleted as observed in breast tumors (J. C. Bourdon, personal communication). Describing and reporting these changes at the RNA and protein level as t1:r.313_375del and p1:p.G105_T125del would help to train bioinformatics tools and get better predictions.

Variants in dbSNP

In the human population, hundreds of dbSNP entries describe variants in the *TP53* gene or in its vicinity and several haplotype blocks have also been identified [Mechanic et al., 2007; Phang et al., 2011; Ortiz-Cuaran et al., 2013]. A large number of studies have focused on the association between common *TP53* germ line variants and cancer risk (reviewed in [Whibley et al., 2009]). Several dbSNP entries such as rs78378222 (t1:c.*1175A>C, localized in the 3′UTR of the gene and responsible for changing the AATAAA polyadenylation signal to AATACA, RNA change: r.*1175a>c) or rs17878362 (t1:c.96+41_97−54del, 16 bp deletion in intron 3) have been associated with an increased risk of cancer but more studies are needed to establish any possible clinical value [Stacey et al., 2011].

Classification of dbSNP Entries Associated with TP53 Amino-Acid Changes

Combing the literature, we have identified 14 dbSNP entries (previously called "natural SNPs") associated with amino-acid changes in the *TP53* protein (Table 2 and Supp. Table S2). Unfortunately, the literature search did not reveal the princeps publications defining several of these variants. We have classified the variants described in these dbSNP entries as Class 1 (benign, previously "certified": "C"), Class 3 ("uncertain": "U"), or Class 5 variant ("mutation": "M") (Table 2 and Supp. Table S2).

rs1042522 (t1:c.215C>G, p1:p.P72R, previously described as p.R72P due to a reference sequence based on the other allele) and rs1800371 (t1:c.139C>T, previously: p.P47S) are the two most frequently detected exonic *TP53* variants. They have been extensively analyzed and their status as a Class 1 germ line variant is clear (Table 2 and Supp. Table S2); they will thus not be further discussed here.

Four variants in dbSNP, rs11540654 (t1:c.329G>C, previously: p.R110P and t1:c.329G>T, previously: p.R110L), rs55832599 (t1:c.799C>T, previously: p.R267W), and rs17849781 (t1:c.832C>G, previously: p.P278A) have been detected in the germ line of cancer families (Table 2 and Supp. Table S2). Their dbSNP records show that two of them are very rare (rs11540654 [p.R110L and p.R110P]) and no population frequency data are available for the two others. These four variants have no transcriptional activity and lack proapoptotic function [Kakudo et al., 2005; Soussi et al., 2005; Wang et al., 2013]. It is therefore likely that these dbSNP entries represent rare deleterious germ line variants (Class 5). Their detection in various types of cancer described in multiple entries in the UMD *TP53* database suggests they are also Class 5 somatic variants (Table 2 and Supp. Table S2).

rs72661117 (t1:c.550G>A, previously: p.D184N) represents a dbSNP entry of a somatic variant, which has not been detected in the population. Found in 36 tumors in the UMD *TP53* database, its frequency is higher than the passenger mutation background, but it is not significantly associated with dubious studies. Variant p.D184N does not display any obvious deficiencies in its transcriptional activity. It is located close to Ser183, a residue phosphorylated by Aurora B, but not within its consensus sequence [Gully et al., 2012]. More information about potential effects on TP53 posttranslational modification is needed to justify another classification than somatic Class 3.

dbSNP entry rs35163653 (t1:c.649G>A, previously: p.V217M) is a very rare germ line variant described in two independent populations. The UMD *TP53* database contains 14 descriptions of this variant being somatic, but mostly from articles containing dubious data [Edlund et al., 2012]. Although this variant can be considered a Class 1 benign germ line variant, further research is needed to assess its functional consequences in tumors.

The status of rs55819519 (t1:c.869G>A, previously: p.R290H) is more ambiguous. This variant changed the G of a CpG dinucleotide compatible with the deamination of the methylated cytosine on the other DNA strand. No loss of activity has been associated with this variant but it is localized close to posttranslationally modified residues K291 (ubiquitination site) and K292 (ubiquitination and acetylation site). It has been described 47 times in the UMD *TP53* database, including germ line variants in seven independent cancer families. No population data are available, suggesting this is an extremely rare germ line variant (Class 3). The impact on *TP53* function remains to be investigated, so the somatic variant has to be classified as Class 3 for the moment. The reciprocal transition caused by deamination of methylated cytosine t1:c.868C>T appears 21 times in the *TP53* variant database, does not lead to *TP53* loss of function and can be assigned to Class 1.

The three dbSNP entries rs17882252 (t1:c.1015G>A, previously: p.E339K), rs35993958 (t1:c.1079G>C, previously: p.G360A), and rs17881470 (t1:c.1096T>G, previously: p.S366A) share similar

Table 2. Nonsynonymous dbSNP Entries in the *TP53* Gene[a]

SNP[b]	cDNA variant[c]	Protein variant[c]	Common name	Database frequency[d]	Activity[e]	Population analysis[f]	Class[g]
rs1800371	t1:c.139C>T	p1:p.P47S	p.P47S	_[h]	Wt	5	1
rs1042522	t1:c.2151>G	p1:p.P72R	p.P72R	_[h]	Wt	>10	1
rs11540654	t1:c.329G>T	p1:p.R110L	p.R110L	58/3	Null	1	5
rs11540654	t1:c.329G>C	p1:p.R110P	p.R110P	25/1	Null	0	5
rs72661117	t1:c.550G>A	p1:p.D184N	p.D184N	36/0	Wt	0	3
rs35163653	t1:c.649G>A	p1:p.V217M	p.V217M	14/0	Wt	2	1
rs72661119	t1:c.787A>G	p1:p.N263D	p.N263D	8/0	Wt	0	3
rs55832599	t1:c.799C>T	p1:p.R267W	p.R267W	65/2	Null	0	5
rs17849781	t1:c.832C>G	p1:p.P278A	p.P278A	48/1	Null	0	5
rs55819519	t1:c.869G>A	p1:p.R290H	p.R290H	47/7	Wt	0	3
rs56184981	t1:c.932A>G	p1:p.N311S	p.N311S	2/0	Wt	0	3
rs17882252	t1:c.1015G>A	p1:p.E339K	p.E339K	1/0	Wt	4	1
rs35993958	t1:c.1079G>C	p1:p.G360A	p.G360A	3/0	Wt	3	1
rs17881470	t1:c.1096G>T	p1:p.S366A	p.S366A	4/2	Wt	1	1

[a]An extended version of this table is available (Supp. Table S2).
[b]Only dbSNP entries describing exonic changes resulting in amino-acid substitutions. The database document contains a full list of all *TP53* dbSNP entries (Supp. Table S3).
[c]Description of variants using the LRG_321 reference sequence.
[d]Frequency of each variant in the 2014 issue of the UMD *TP53* database. The two numbers correspond to somatic and germline variants respectively.
[e]Functional activity of each variant defined by Kato et al. (2003) and from the UMD *TP53* database (Hamroun et al., 2006).
[f]Number of large-scale sequencing projects that have described this SNP (data from http://www.ncbi.nlm.nih.gov/SNP/, build 139).
[g]Classification of each dbSNP entry as benign, Class 1 (1), VUS, Class 3 (3) and deleterious, Class 5 (5). Difference between 1 and 3 is based on population analysis. See text for more details.
[h]These dbSNP entries have never been included in the *TP53* database.

observations. They are present at very low frequency in the UMD *TP53* database, do not affect *TP53* activity and have been validated in multiple population analyses. Therefore, they can be reasonably considered as Class 1 germ line and somatic variants.

The two dbSNP entries rs56184981 (t1:c.932A>G, previously: p.N311S) and rs72661119 (t1:c.787A>G, previously: p.N263D) are very infrequent in the UMD *TP53* database, do not affect *TP53* activity and lack population information. Therefore, they have been assigned to Class 3.

Considered in its totality, this analysis indicates that several *TP53* dbSNP entries are indeed somatic Class 5 variants, but others need more verification. This "pollution" of dbSNP with variants not belonging to the low-risk Classes 1 and 2 clearly could result in removal of deleterious variants from NGS data when crudely filtering against the whole database. The problem becomes more consequential when the database is used by private companies to infer disease risk. For example, rs55819519, discussed above, has been used to infer potential risk and labeled as "The change R<>H is uncommon and the homozygous form could be significant." Such annotation in dbSNP is extremely alarming and only curation of dbSNP and careful use of its data can prevent problems. An annotated list of all *TP53* dbSNP entries is available in the Supp. Table S3.

Annotating Variants in the *TP53* Gene

TP53 variants can be annotated independently with variant descriptions, as well as their consequences at the RNA and protein level using information from the *TP53* database (Fig. 8). Specific information, such as the protein domain location, posttranslational modifications, phylogenetic conservation, or properties associated with the WT residue provides insights on the importance of the residue. Further information includes the frequency of the variant in the database, a functional analysis and predictions of deleterious amino-acid substitutions using popular algorithms. Any analytical pipeline for calling *TP53* variants can easily incorporate these data (Supp. Table S4). Four examples below demonstrate how LRG coordinates can help to illustrate effects on different transcripts and isoforms.

chr17:g.7578406C>T (t1:c.524G>A) is a hotspot variant that leads to the expression of the p.R175H TP53 protein. This variant, localized in exon 5, hits all eight TP53 mRNAs and 12 TP53 isoforms. The changes of a specific transcript and isoform can be described using their t and p numbers (e.g., t5:c.128G>A resulting in p5:p.R43H and p11:p.R16H). For t1:c.524G>A, the analysis is straightforward and indicates that this variant inactivates TP53 tumor suppressive functions.

chr17:g.7579358C>A (t1:c.329G>T; p1:p.R110L) is a good example of a frequent somatic Class 5 variant that targets only a subset of TP53 isoforms. Indeed, as it is localized in exon 4, it does not impair all delta133 and delta160 TP53 isoforms. Although this variant impairs the activity of the full-length protein, its biological consequences for the various isoforms are unknown.

chr17:g.7579312C>T (t1:c.375G>A) predicted to be a synonymous variant has been often described as p.T125T. The nucleotide substitution is localized at the end of exon 4 and has been shown to lead to aberrant splicing, affecting 5 transcripts produced by the P1 promoter [Leroy et al., 2014a]. If a normally spliced transcript with the synonymous codon is expressed, its protein should now be described as p1:p. = . Transcripts t5, t6, and t7, transcribed from the internal P2 promoter in intron 4, should be normal.

chr17:g75794222C>T (t1:c.265C>T; p1:p.P89S) is a nonsynonymous variant reported 28 times in the 2014 version of the UMD *TP53* database. This variant does not display a significant loss of activity. Several prediction programs do not identify it as deleterious (Fig. 8). Twenty-two of these 28 variants were described in a single publication that has been shown to be artifactual [Patocs et al., 2007; Edlund et al., 2012]. The remaining six variants were found in tumors that contained multiple *TP53* variants. This variant is therefore a typical example of an artifactual result, which has been tagged in the database.

Assessing the Functional Effects of These Variants

An in-depth discussion of the assessment of the *TP53* variant effects is beyond the scope of this review, but we would like to emphasize that high quality data are essential if variant
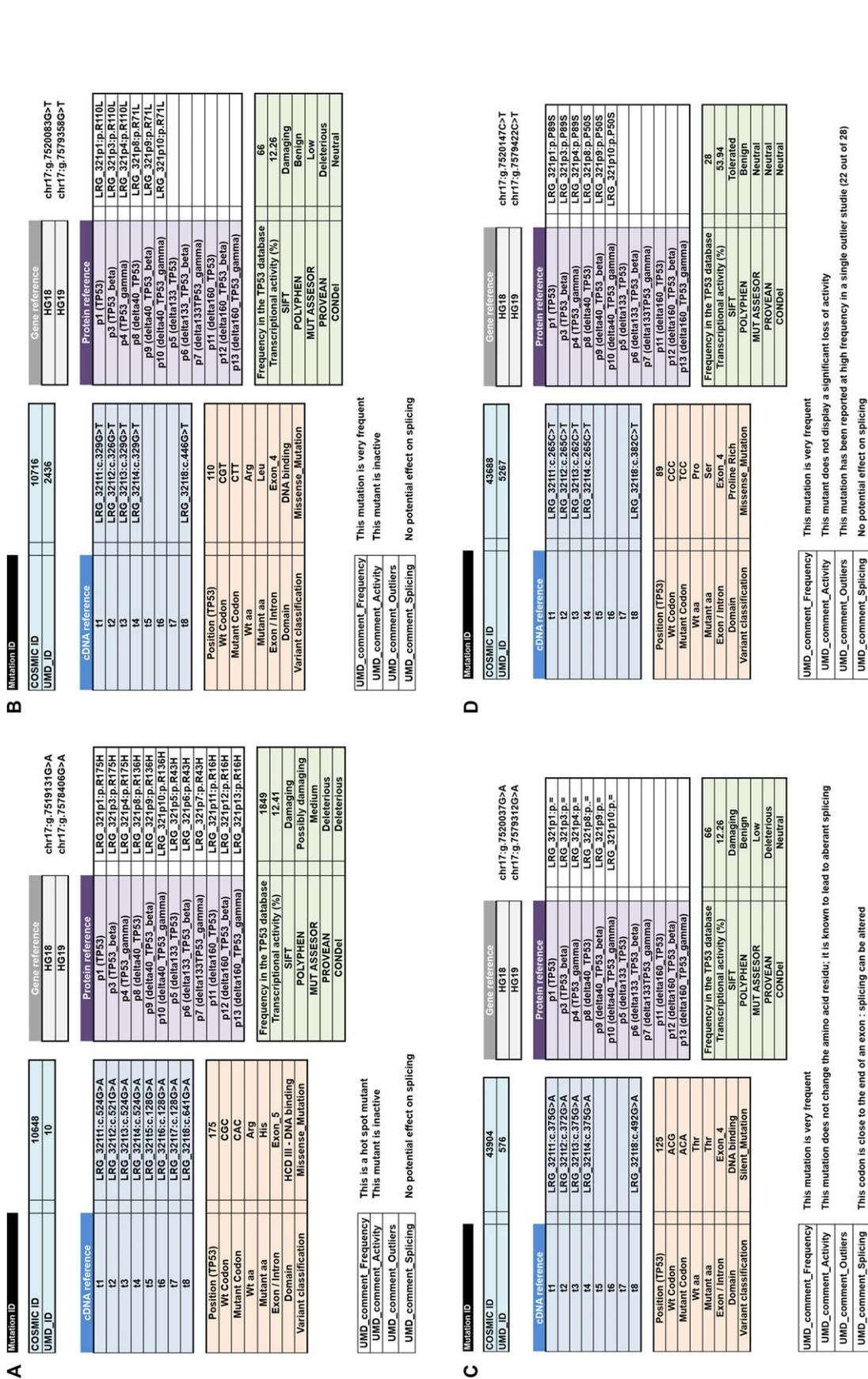
Figure 8. Extensive description of *TP53* variants. Four different variants present in the COSMIC and UMD database are described relative to two human genome builds. For each variant, annotations on all affected *TP53* transcripts and protein isoforms are shown (**A–D**). In addition, the tables provide UMD comments, codon information and protein effect predictions for full-length TP53 (p1). The information shown here is derived from the *TP53* data table (See Supp. Table S4 for regular updates). **A**: chr17:g.7578406C>T (T1:c.524G>A; LRG_321p1:p.R175H) is a hotspot variant that targets all TP53 isoforms as indicated by the different descriptions at the protein level. The deleterious effect of this variant on TP53 activity predicted by several tools has been confirmed by numerous experimental analyses. **B**: chr17:g.7579358C>A (T1:c.329G>T; LRG_321p1:p.R110L). This variant does not impair six out of the 12 putative TP53 isoforms. Several tools predict a deleterious effect on TP53 activity, which has been confirmed by numerous experimental analyses. **C**: chr17:g.7579312C>T (T1:c.375G>A; LRG_321p1:p. =). This variant, previously reported as p.T125T suggesting a synonymous change, affects the splicing of five out of the eight *TP53*transcripts. This splice defect was not predicted by specific splice site prediction algorithms. The unknown effect at the protein level can be indicated as p.? relative to the affected transcripts. **D**: chr17:g7579422C>T (T1:c.265C>T; LRG_321p1:p.P89S) is considered to be purely artifactual (See text for details).
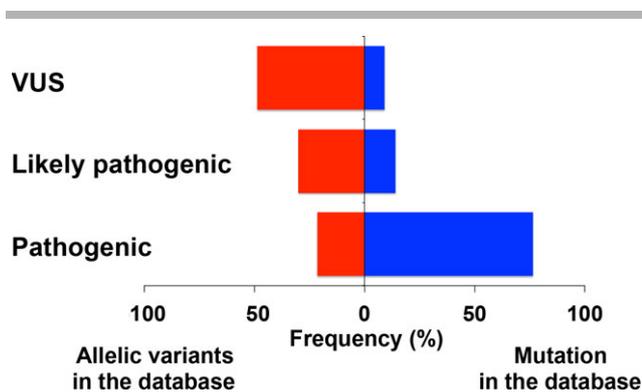
Figure 9. Classification of *TP53* missense variants. The 1,603 missense variants included in the *TP53* variant database (2014 release) were categorized in three classes according to their functional consequences ("pathogenicity") (Soussi et al., unpublished results). Class 5 ("Pathogenic") variants: Three-hundred forty four variants (21% of the variants, red bar) corresponding to 76% of all variants included in the database (blue bar). Class 4 ("Likely pathogenic") variants: Four-hundred eighty variants (30% of the variants, red bar) corresponding to 14% of all variants included in the database (blue bar). Class 3 ("VUS"): Seven-hundred seventy nine variants (49% of the variants, blue bar) corresponds to 9% of all variants included in the database (blue bar).

information is to be used for clinical decisions. The identification of the founder variant p.R337H in Brazil is a perfect example of this reciprocal complementarity between clinical and basic research [Ribeiro et al., 2001]. LRG_321t1:c.1010G>A (p.R337H) is a germ line variant associated with a high predisposition to pediatric adrenocortical carcinoma, with prevalence reaching 5 per 1,000 in certain districts of the state of Paraná in Brazil [Custodio et al., 2013]. This variant, localized in the oligomerization domain of exon 10, does not target all TP53 isoforms. Functional analyses have shown that full-length TP53 protein variant is transcriptionally active but somewhat sensitive to changes in pH [DiGiammarino et al., 2002].

TP53 variants have been classified using multiple criteria [T. Soussi et al., unpublished results] (Fig. 9). Variants included in the TP53 database were categorized in three classes: (1) Class 5—variants having functional effects ("pathogenic"); (2) Class 4—variants likely having functional effects ("likely pathogenic"); and (3) Class 3—VUS. Although the number of VUS is high, most are infrequent and correspond to 9% of the total number of variants in the database. On the other hand, Classes 5 and 4 variants correspond to 76% and 14% of the total number of variants, respectively.

Multiple generic bioinformatics tools developed to assess variant effects have often used *TP53* variants as a paradigm to check their specificity and sensitivity, both of which rarely exceed 80%. Most of these tools are only efficient for variants whose functional effects were otherwise obvious with simple criteria such as frequency or association with a loss of activity. We believe that it will be impossible to improve their rate of detection substantially in the future. The next generation of tools should be tailored for each gene by including information about the disease mechanism, specific data related to transcript and protein structure and function as well as information related to variant frequency and interactions between variants. Collecting this information in gene variant databases (locus-specific databases) will be invaluable for *TP53*, but also for other genes.

Recommendations

It is important to emphasize that most recommendations are not specific to the detection and description of variants in the *TP53* gene. They are meant to harmonize approaches and variant reporting and applicable to all other genes.

(1) The complete gene (promoter, exons, and introns), including the region that overlaps with the *WRAP53* gene should be screened at the DNA level in different tissues and tumors to identify variants of importance. This unbiased analytical approach is vital and irreplaceable for defining unambiguously the regions of *TP53* that are of importance in human cancer and generating important working hypotheses to understand the tumor suppressive function of *TP53* and it various isoforms, including those encoded by the new β and γ exons. Although conventional approaches can be adapted to cover the new regions, new strategies using NGS platforms might be more effective, at least for research purposes.

(2) The terms "mutation" and "SNP" are ambiguous and should not be used. The HGVS recommends using the term "variant," regardless of its origin or frequency. Variants can be distinguished according to their origin and their functional consequences. Germ line variants are those inherited or arising de novo before fertilization. In the context of cancer, somatic variants are observed in tumor cells, but not in normal cells. Variants detected in tumor material should not be labeled as somatic, unless their absence in normal cells of the same individual has been confirmed. Without this confirmation, variants can be annotated as "detected in tumor"

(3) We would like to propose describing functional consequences at the molecular level using the same classification system for germ line and somatic variants. Both germ line and somatic variants can have functional consequences ("pathogenic variants") or not ("benign variants" including so-called "somatic passenger mutations"). In case insufficient evidence exists, variants are classified as VUS. The five-class system including the intermediate terms "Likely having functional effects" (Class 4) and "Likely benign" (Class 2) is already applied for calculations of cancer susceptibility risks of inherited variants and takes information about recurrent somatic variants into account [Plon, 2008]. We recommend specification of variant origin to assess potential differences between the functional consequences of germ line and somatic variants. This may help to reconcile and refine different classifications and help translating this information into a clinical outcome probability score.

(4) All researchers and clinicians working on *TP53* are recommended to:

    (a) Specify the reference sequences used to describe data at the different levels (genome build, transcript and protein reference sequence accession numbers and version numbers). For data standardization, the stable Locus Reference Genomic sequence LRG_321 is preferred. In case, RefSeq gene or transcript records are used, their version numbers should be included.

    (b) Describe germ line and somatic variants according to the HGVS sequence nomenclature guidelines. LRG_321t1 should be used at the coding DNA and RNA levels; the full-length TP53 protein LRG_321p1 at the protein level. In case variants in multiple genes are described, the full format (e.g., LRG_321t1:c.215C>G, LRG_321t1:r.215c>g, LRG_321p1:p.P72R) is recommended. In publications, after specification of the reference sequence record, the

transcript or protein number is sufficient (e.g., t1:c.215C>G, t1:r.215c>g; p1:p.P72R). To report changes in the P1 and P2 promoter sequences, either chromosomal positions or positions relative to transcripts should be used. In general, promoter studies indicate the position of the first nucleotide upstream of the transcription start site as –1, counting backward to the position of transcription factor binding sites. The HGVS nomenclature would indicate this as t1:c.-203. A more informative description format similar to the one used for intron variants still awaits HGVS approval. In this format, t1:c.-202-u1 contains the start of the first exon and the prefix u indicating an upstream location which upon removal results in the traditional position –1.

(c) Support further evaluation of the classification of variants and their clinical consequences by submitting variant data to publicly available gene variant databases.

(5) In discussions about functional effects of variants on other *TP53* transcripts and protein isoforms, changes at the coding DNA, RNA, and protein levels can be indicated using the corresponding LRG_321 transcript and protein numbers. Observed effects on splicing should be described at the RNA level using t1:r. Predicted effects on RNA and protein should be between parentheses (e.g., t1:r.(215>g); p1:p.(P72R)).

Conclusions

Novel large-scale analyses have confirmed and extended the unchallenged leadership of *TP53* as the most frequently mutated gene in human cancer [Kandoth et al., 2013; Lawrence et al., 2014]. The value of *TP53* variants as predictive or prognostic markers in various types of cancer has been extensively analyzed. The repeatedly shown association of Classes 4 and 5 *TP53* variants with a poor prognosis in CLL resulted in published recommendations for *TP53* screening [Malcikova et al., 2014]. In other tumor types, novel studies using molecular classification of cancer subtypes have also associated Classes 4 and 5 *TP53* variant with poor prognosis. All these research efforts point toward a need to standardize the analysis of *TP53* alterations so that their clinical significance in all subtypes of cancer can be clearly established. The high frequency of *TP53* alterations across all types of cancer indicates that they can be useful as molecular biomarkers for monitoring tumor progression. Indeed, several recent studies have shown that *TP53* variants can be detected in tumor DNA circulating in the plasma of patients with breast or ovarian cancer [Murtaza et al., 2013; Bettegowda et al., 2014]. Methods that detect *TP53* alterations with high sensitivity even in the presence of large amounts of normal DNA have been developed. Their specificity may be disappointing as they target only a few hotspot regions. Advances in genomic methodologies will alleviate the various sensitivity and specificity issues but only intricate knowledge of the significance of the various alterations of the target gene will lead to efficient diagnostics.

Acknowledgment

References

Aoubala M, Murray-Zmijewski F, Khoury MP, Fernandes K, Perrier S, Bernard H, Prats AC, Lane DP, Bourdon JC. 2011. p53 directly transactivates Delta 133p53alpha, regulating cell fate outcome in response to DNA damage. Cell Death Differ 18:248–258.

Alsner J, Jensen V, Kyndi M, Offersen BV, Vu P, Borresen-Dale AL, Overgaard J. 2008. A comparison between p53 accumulation determined by immunohistochemistry and TP53 mutations as prognostic variables in tumours from breast cancer patients. Acta Oncol 47:600–607.

Auerbach AD, Burn J, Cassiman JJ, Claustres M, Cotton RG, Cutting G, den Dunnen JT, El-Ruby M, Vargas AF, Greenblatt MS, Macrae F, Matsubara Y, et al. 2011. Mutation (variation) databases and registries: a rationale for coordination of efforts. Nat Rev Genet 12:881; discussion 881.

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.

Bartek J, Bartkova J, Vojtesek B, Staskova Z, Lukas J, Rejthar A, Kovarik J, Midgley CA, Gannon JV, Lane DP. 1991. Aberrant expression of the p53 oncoprotein is a common feature of a wide spectrum of human malignancies. Oncogene 6:1699–1703.

Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, Bartlett BR, Wang H, Luber B, Alani RM, Antonarakis ES, Azad NS, et al. 2014. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med 6:224ra24.

Bourdon JC, Fernandes K, Murray-Zmijewski F, Liu G, Diot A, Xirodimas DP, Saville MK, Lane DP. 2005. p53 isoforms can regulate p53 transcriptional activity. Genes Dev 19:2122–2137.

Casey G, Lopez ME, Ramos JC, Plummer SJ, Arboleda MJ, Shaughnessy M, Karlan B, Slamon DJ. 1996. DNA sequence analysis of exons 2 through 11 and immunohistochemical staining are required to detect all known p53 alterations in human malignancies. Oncogene 13:1971–1981.

Claustres M, Horaitis O, Vanevski M, Cotton RG. 2002. Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. Genome Res 12:680–688.

Cotton RG, Auerbach AD, Axton M, Barash CI, Berkovic SF, Brookes AJ, Burn J, Cutting G, den Dunnen JT, Flicek P, Freimer N, Greenblatt MS, et al. 2008. GENETICS. The Human Variome Project. Science 322:861–862.

Cotton RG, Malcolm AD. 1991. Mutation detection. Nature 353:582–583.

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.

Dalgleish R, Flicek P, Cunningham F, Astashyn A, Tully RE, Proctor G, Chen Y, McLaren WM, Larsson P, Vaughan BW, Beroud C, Dobson G, et al. 2010. Locus Reference Genomic sequences: an improved basis for describing human DNA variants. Genome Med 2:24.

den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15:7–12.

DiGiammarino EL, Lee AS, Cadwell C, Zhang W, Bothner B, Ribeiro RC, Zambetti G, Kriwacki RW. 2002. A novel mechanism of tumorigenesis involving pH-dependent destabilization of a mutant p53 tetramer. Nat Struct Biol 9:12–16.

Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, Abbott RM, Hoog J, et al. 2010. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature 464:999–1005.

Donehower LA. 2014. Insights into wild-type and mutant p53 functions provided by genetically engineered mice. Hum Mutat 35:715–727.

Edlund K, Larsson O, Ameur A, Bunikis I, Gyllensten 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 ultradeep sequencing of human tumors. Proc Natl Acad Sci USA 109:9551–9556.

Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, et al. 2012. Architecture of the human regulatory network derived from ENCODE data. Nature 489:91–100.

Ghosh A, Stewart D, Matlashewski G. 2004. Regulation of human p53 activity and cell localization by alternative splicing. Mol Cell Biol 24:7987–7997.

Gluck S, Ross JS, Royce M, McKenna EFJ, Perou CM, Avisar E, Wu L. 2012. TP53 genomics predict higher clinical and pathologic tumor response in operable early-stage breast cancer treated with docetaxel-capecitabine +/– trastuzumab. Breast Cancer Res Treat 132:781–791.

Gully CP, Velazquez-Torres G, Shin JH, Fuentes-Mattei E, Wang E, Carlock C, Chen J, Rothenberg D, Adams HP, Choi HH, Guma S, Phan L, et al. 2012. Aurora B kinase phosphorylates and instigates degradation of p53. Proc Natl Acad Sci USA 109:E1513–E1522.

Hall PA, Lane DP. 1994. p53 in tumour pathology: can we trust immunohistochemistry?–Revisited! J Pathol 172:1–4.

Hall PA, McCluggage WG. 2006. Assessing p53 in clinical contexts: unlearned lessons and new perspectives. J Pathol 208:1–6.

Hamroun D, Kato S, Ishioka C, Claustres M, Beroud C, Soussi T. 2006. The UMD TP53 database and website: update and revisions. Hum Mutat 27:14–20.

Jackson JG, Lozano G. 2013. The mutant p53 mouse as a pre-clinical model. Oncogene 32:4325–4330.

Kakudo Y, Shibata H, Otsuka K, Kato S, Ishioka C. 2005. Lack of correlation between p53-dependent transcriptional activity and the ability to induce apoptosis among 179 mutant p53s. Cancer Res 65:2108–2114.

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.

Kenzelmann Broz D, Attardi LD. 2010. In vivo analysis of p53 tumor suppressor function using genetically engineered mouse models. Carcinogenesis 31:1311–1318.

Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G. 2014. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 505:495–501.

Legros Y, Lafon C, Soussi T. 1994. Linear antigenic sites defined by the B-cell response to human p53 are localized predominantly in the amino and carboxy-termini of the protein. Oncogene 9:2071–2076.

Leroy B, Anderson M, Soussi T. 2014a. TP53 mutations in human cancer: database reassessment and prospects for the next decade. Hum Mutat 35:672–688.

Leroy B, Fournier JL, Ishioka C, Monti P, Inga A, Fronza G, Soussi T. 2013. The TP53 website: an integrative resource centre for the TP53 mutation database and TP53 mutant analysis. Nucleic Acids Res 41:D962–D969.

Leroy B, Girard L, Minna JD, Gazdar AF, Soussi T. 2014b. Analysis of p53 mutation status in human cancer cell lines: a reassessment. Hum Mutat 35:756–765.

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 Y, Gordon MW, Xu-Monette ZY, Visco C, Tzankov A, Zou D, Qiu L, Montes-Moreno S, Dybkaer K, Orazi A, Zu Y, Bhagat G, et al. 2013. Single nucleotide variation in the TP53 3′ untranslated region in diffuse large B-cell lymphoma treated with rituximab-CHOP: a report from the International DLBCL Rituximab-CHOP Consortium Program. Blood 121:4529–4540.

MacArthur JA, Morales J, Tully RE, Astashyn A, Gil L, Bruford EA, Larsson P, Flicek P, Dalgleish R, Maglott DR, Cunningham F. 2014. Locus Reference Genomic: reference sequences for the reporting of clinically relevant sequence variants. Nucleic Acids Res 42:D873–D878.

Mahmoudi S, Henriksson S, Corcoran M, Mendez-Vidal C, Wiman KG, Farnebo M. 2009. Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. Mol Cell 33:462–471.

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.

Marchetti A, Felicioni L, Buttitta F. 2006. Assessing EGFR mutations. N Engl J Med 354:526–528; author reply 526.

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.

Matlashewski G, Pim D, Banks L, Crawford L. 1987. Alternative splicing of human p53 transcripts. Oncogene Res 1:77–85.

Mechanic LE, Bowman ED, Welsh JA, Khan MA, Hagiwara N, Enewold L, Shields PG, Burdette L, Chanock S, Harris CC. 2007. Common genetic variation in TP53 is associated with lung cancer risk and prognosis in African Americans and somatic mutations in lung tumors. Cancer Epidemiol Biomarkers Prev 16:214–222.

Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, Piskorz AM, Parkinson C, Chin SF, Kingsbury Z, Wong AS, Marass F, Humphray S, et al. 2013. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature 497:108–112.

Nigro JM, Baker SJ, Preisinger AC, Jessup JM, Hostetter R, Cleary K, Bigner SH, Davidson N, Baylin S, Devilee P, et al. 1989. Mutations in the p53 gene occur in diverse human tumour types. Nature 342:705–708.

Olivier M, Taniere P. 2011. Somatic mutations in cancer prognosis and prediction: lessons from TP53 and EGFR genes. Curr Opin Oncol 23:88–92.

Ortiz-Cuaran S, Cox D, Villar S, Friesen MD, Durand G, Chabrier A, Khuhaprema T, Sangrajrang S, Ognjanovic S, Groopman JD, Hainaut P, Le Calvez-Kelm F. 2013. Association between TP53 R249S mutation and polymorphisms in TP53 intron 1 in hepatocellular carcinoma. Genes Chromosomes Cancer 52:912–919.

Patocs A, Zhang L, Xu Y, Weber F, Caldes T, Mutter GL, Platzer P, Eng C. 2007. Breast-cancer stromal cells with TP53 mutations and nodal metastases. N Engl J Med 357:2543–2551.

Phang BH, Chua HW, Li H, Linn YC, Sabapathy K. 2011. Characterization of novel and uncharacterized p53 SNPs in the Chinese population–intron 2 SNP co-segregates with the common codon 72 polymorphism. PLoS One 6:e15320.

Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, Hogervorst FB, Hoogerbrugge N, Spurdle AB, Tavtigian SV. 2008. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat 29:1282–1291.

Pospisilova S, Gonzalez D, Malcikova J, Trbusek M, Rossi D, Kater AP, Cymbalista F, Eichhorst B, Hallek M, Dohner H, Hillmen P, van Oers M, et al. 2012. ERIC recommendations on TP53 mutation analysis in chronic lymphocytic leukemia. Leukemia 26:1458–1461.

Reisman D, Balint é, Loging WT, Rotter V, Almon E. 1996. A novel transcript encoded within the 10-kb first intron of the human p53 tumor suppressor gene (D17S2179E) is induced during differentiation of myeloid leukemia cells. Genomics 38:364–370.

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.

Sah S, Chen L, Houghton J, Kemppainen J, Marko AC, Zeigler R, Latham GJ. 2013. Functional DNA quantification guides accurate next-generation sequencing mutation detection in formalin-fixed, paraffin-embedded tumor biopsies. Genome Med 5:77.

Sauna ZE, Kimchi-Sarfaty C. 2011. Understanding the contribution of synonymous mutations to human disease. Nat Rev Genet 12:683–691.

Soussi T, Asselain B, Hamroun D, Kato S, Ishioka C, Claustres M, Beroud C. 2006. Meta-analysis of the p53 mutation database for mutant p53 biological activity reveals a methodologic bias in mutation detection. Clin Cancer Res 12:62–69.

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, May P. 1996. Structural aspects of the p53 protein in relation to gene evolution: a second look. J Mol Biol 260:623–637.

Stacey SN, Sulem P, Jonasdottir A, Masson G, Gudmundsson J, Gudbjartsson DF, Magnusson OT, Gudjonsson SA, Sigurgeirsson B, Thorisdottir K, Ragnarsson R, Benediktsdottir KR, et al. 2011. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. Nat Genet 43:1098–1103.

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.

Taneja P, Zhu S, Maglic D, Fry EA, Kendig RD, Inoue K. 2011. Transgenic and knockout mice models to reveal the functions of tumor suppressor genes. Clin Med Insights Oncol 5:235–257.

Tenaud C, Negoescu A, Labat-Moleur F, Legros Y, Soussi T, Brambilla E. 1994. p53 immunolabeling in archival paraffin-embedded tissues: optimal protocol based on microwave heating for eight antibodies on lung carcinomas. Mod Pathol 7:853–859.

Vousden KH, Prives C. 2009. Blinded by the light: the growing complexity of p53. Cell 137:413–431.

Wang B, Niu D, Lam TH, Xiao Z, Ren EC. 2013. Mapping the p53 transcriptome universe using p53 natural polymorphs. Cell Death Differ 21–521–532.

Wang GS, Cooper TA. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet 8:749–761.

Whibley C, Pharoah PD, Hollstein M. 2009. p53 polymorphisms: cancer implications. Nat Rev Cancer 9:95–107.

Williams C, Ponten F, Moberg C, Soderkvist P, Uhlen M, Ponten J, Sitbon G, Lundeberg J. 1999. A high frequency of sequence alterations is due to formalin fixation of archival specimens. Am J Pathol 155:1467–1471.