Recommended guidelines for validation, quality control and reporting of TP53 variants in clinical practice

Bernard Leroy[1], Mandy L. Ballinger[2], Fanny Baran-Marszak[3], Gareth L. Bond[4], Antony Braithwaite[5],[6], Nicole Concin[7], Lawrence A. Donehower[8], Wafik S. El-Deiry[9], Pierre Fenaux[10], Gianluca Gaidano[11], Anita Langerød[12], Eva Hellstrom-Lindberg[13], Richard Iggo[14], Jacqueline Lehmann-Che[15], Phuong L. Mai[16], David Malkin[17], Ute M. Moll[18], Jeffrey N. Myers[19], Kim E. Nichols[20], Sarka Pospisilova[21], Patricia Ashton-Prolla[22], Davide Rossi[10] Sharon A. Savage[16], Louise C. Strong[23], Patricia N. Tonin[24], Robert Zeillinger[25], Thorsten Zenz[26],Joseph F. Fraumeni Jr.[27], Peter E.M. Taschner[28],Pierre Hainaut[29] and Thierry Soussi[1],[30],[31],[32]

[1] Sorbonne Université, UPMC Univ Paris 06, Paris F-75005, France
[2] Research Division, Peter MacCallum Cancer Centre, Melbourne, Australia
[3] Hôpital Avicenne, Assistance Publique—Hôpitaux De Paris, Bobigny, Service D'H ematologie Biologique, France
[4] Ludwig Institute for Cancer Research, University of Oxford, Nuffield Department of Clinical Medicine, Old Road Campus Research Building, Oxford OX3 7DQ, United Kingdom
[5] Dept of Pathology, School of Medicine, University of Otago, PO Box 913, Dunedin, New Zealand
[6] Children's Medical Research Institute, University of Sydney, Westmead NSW 2145, Australia
[7]Department of Gynecology and Obstetrics, Innsbruck Medical University, Anichstr. 35, 6020 Innsbruck, Austria
[8]Human Genome Sequencing Center, Baylor College of Medicine, USA
[9] Department of Hematology/Oncology and Molecular Therapeutics Program, Fox Chase Cancer Center, Philadelphia, PA, USA
[10] Service d'hématologie séniors, Hôpital St Louis / Université Paris 7, 1 avenue Claude Vellefaux, 75475 PARIS, France
[11] Division of Hematology, Department of Translational Medicine, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy
[12] Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway
[13] Karolinska Institute, Department of medicine, Center for Hematology and Regenerative medicine, Karolinska University Hospital, Stockholm, Sweden
[14] Bergonié Cancer Institute University of Bordeaux 229 cours de l'Argonne 33076 Bordeaux, France
[15] Molecular Oncology Unit, Department of Biochemistry, Hospital Saint Louis, AP-HP, Paris, France
[16] Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA
[17] Division of Hematology/Oncology, Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, ON. Canada
[18] Department of Pathology, Stony Brook University, Stony Brook NY 11794-8691, USA
[19] The University of Texas MD Anderson Cancer Center, USA
[20] Department of Oncology, Division of Cancer Predisposition, St. Jude Children's Research Hospital, Memphis TN 38120, USA
[21] Masaryk University, CEITEC – Molecular Medicine and University Hospital Brno, Department of Internal Medicine – Hematology and Oncology, Brno, Czech Republic
[22] Universidade Federal do Rio Grande do Sul (UFRGS) e Serviço deGenética Médica-HCPA, Rua Ramiro Barcelos 2350, Porto Alegre RS 90035-903 - Brasil
[23] University of Texas, M.D. Anderson Cancer Center, Houston TX 77030, USA
[24] Departments of Medicine & Human Genetics, McGill University and Cancer Research Program, The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada
[25] Molecular Oncology Group, Department of Obstetrics and Gynaecology, Medical University of Vienna, Vienna, Austria
[26] University of Heidelberg, Department of Medicine V, Heidelberg, Germany; Department of Translational Oncology, National Center for Tumor Diseases and German Cancer Research Center (dkfz), Heidelberg, Germany
[27] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892-9776, USA
[28] Generade Centre of Expertise Genomics and University of Applied Sciences Leiden, Zernikedreef 10, Room Z1.026, 2333 CL Leiden, The Netherlands
[29] Institut Albert Bonniot, Inserm 823, Université Grenoble Alpes, Rond Point de la Chantourne, 38700 La Tronche, France
[30] Department of Oncology-Pathology, Karolinska Institutet, Cancer Center Karolinska (CCK) R8:04, Stockholm SE-171 76, Sweden
[31] INSERM, U1138, Centre de Recherche des Cordeliers, Paris, France

[32] Corresponding author: thierry.soussi@ki.se

Abstract

Accurate assessment of *TP53* gene status in sporadic tumors and in the germline of individuals at high risk of cancer due to Li-Fraumeni Syndrome (LFS) has important clinical implications for diagnosis, surveillance and therapy.
Genomic data from more than 20,000 cancer genomes provide a wealth of unbiased information on cancer gene alterations and have confirmed TP53 as the most commonly mutated gene in human cancer. Analysis of a database of 70,000 *TP53* variants reveals that the two newly discovered exons of the gene, exons 9β and 9γ, generated by alternative splicing, are the targets of inactivating mutation events in breast, liver and head and neck tumors. Furthermore, germline rearrangements in intron 1 of *TP53* are associated with LFS and are frequently observed in sporadic osteosarcoma.
In this context of constantly growing genomic data, we discuss how screening strategies must be improved when assessing *TP53* status in clinical samples. Finally, we discuss how TP53 alterations should be described by using accurate nomenclature to avoid confusion in scientific and clinical reports.

Introduction

A major goal of cancer research is the identification of tumor-specific vulnerabilities that can be exploited to tailor treatment to the unique genetic and epigenetic tumor profile of individual patients.(1) This can be achieved as a result of the enormous progress in cancer genomics and the increasingly detailed knowledge of the genetic landscape of the most common tumor types. Single Nucleotide Variants (SNV)[33] as well as small insertions and deletions (indels) targeting cancer genes are among the most common deleterious genetic events that are scattered throughout the entire genome of the tumor.(2)

A unique three-phase pattern of variant description is observed following the discovery of a novel cancer gene: discovery, validation, and clinical practice. (3) (Figure 1) The duration of these phases, individually and globally, depends on the scientific "popularity" of the gene, the type of alteration and its clinical relevance (Figure 1). For several genes, such as *BRAF* for which the first variants were described in 2002, the three-phase workflow was rapidly completed due to the very limited diversity of the variants. The *BRAF* variant NM_004333.4:c.1799T>A (p.Val600Glu) is virtually the only deleterious variant reported in a wide variety of cancers including melanoma, papillary thyroid cancer, colorectal carcinoma, glioma and other cancers and successful targeted therapy has already been developed. (4) The three-phase workflow is also well illustrated by the analysis of the *TP53* suppressor gene. The discovery phase began in 1989 with the first description of *TP53* variants in lung and colorectal cancer. (5, 6) Over the following years, there was a steady increase in the number of publications describing novel *TP53* alterations in most cancer types, culminating in over 10,000 variants (encompassing about 2,500 distinct mutational events) reported in 300 publications by 2001.(7, 8) More than 85% of the different missense *TP53* variants reported in the various *TP53* databases were identified during the discovery phase. The decline in the number of published *TP53* variants began in 2002, corresponding to the beginning of the second, validation phase. The latest issue of the *TP53* variant database was released in 2015 and contains a total of

[33] In this manuscript, the term "variant will be used to describe genetic changes (see Box 1A for more information on terminology)

60,000 variants, encompassing 1,700 different missense and nonsense variants. (9) The number of novel single-base variants has not increased significantly for several years now, indicating that a saturation plateau has been reached with the discovery of all potential deleterious *TP53* variants.

TP53 mutation analysis has now reached the third phase with the development of clinical guidelines for *TP53* mutation testing in various settings. Germline *TP53* variants have emerged as a significant cause of genetic predisposition to cancer associated with LFS.(10) The most recent version of the National Comprehensive Cancer Network (NCCN) guidelines recommends *TP53* mutation testing in individuals with onset of breast cancer before 35 years of age, either concurrently with *BRCA1/2* testing or as a follow-up test after negative *BRCA1/2* testing (NCCN Guidelines Version 2.2015, http://www.nccn.org/professionals/physician_gls/pdf/genetics_screening.pdf). Somatic *TP53* mutation analysis is now widely used in clinical trials involving patient stratification based on *TP53* status and in trials of novel drugs targeting either wild-type or mutant TP53 in order to activate a TP53 antitumor response. *TP53* mutation screening is therefore rapidly becoming an integral part of many therapeutic or prevention strategies in clinical practice.

The TP53 network

The transcription factor, TP53 protein, is at the center of a network that integrates and transmits multiple signals, generated during various stress events to ensure cell and tissue homeostasis. (11–13). These pathways include two other members of the TP53 family, TP63 and TP73 (14, 15) as well as two negative regulators, MDM2 and MDM4. (16) TP53 also has transcription-independent functions via a direct interaction with pro- and anti-apoptotic factors in mitochondria, thereby regulating apoptosis. (17)

Under normal conditions, p53 protein is maintained at low levels as a result of rapid turnover mediated by MDM2, its main negative regulator. In response to various forms of stress, TP53 becomes activated and elicits a variety of activities including cell growth arrest, apoptosis or senescence to prevent the propagation of aberrant cells. Although these three cellular responses were originally associated with the tumor suppressor activity of TP53, their importance has recently been challenged in several mouse models. (18)

Recent evidence has also linked TP53 function to regulation of metabolism and the redox balance to maintain intracellular homeostasis. (19) Whether or not these functions are associated with the tumor suppressor effect of TP53 remains to be elucidated.

A discussion of all aspects of the various signaling pathways regulated by TP53 is beyond the scope of this article and recent reviews on this subject are available.(12, 20, 21)

Heterogeneity of TP53 variants

Among the 14 million new cases of cancer diagnosed in 2012, 7 to 8 million (50 to 60%) tumors harbored a somatic *TP53* variant (http://globocan.iarc.fr). With a few exceptions, such as testicular cancer, neuroblastoma or mesothelioma, *TP53* variants can be detected in all types of cancer with a high degree of heterogeneity (ranging from 10 to 90%) making *TP53* the most frequently mutated gene in human cancer (2, 22). Apart from variants, TP53 function can also be inactivated via other mechanisms such as

amplification of its negative regulators *MDM2* and *MDM4* (previously called *MDMX)* or by binding to viral oncoproteins such as E6, expressed by Human Papilloma Virus. (23, 24) In acute myeloid leukemia, hyperactivity of histone deacetylase HDAC8 prevents post-translational acetylation-mediated activation of the TP53 protein, which is essential for its tumor suppressor function.(25)

Among the 60,000 tumors that harbor TP53 modifications described to date, missense alterations in the coding region of the full-length protein are the most common alterations. Approximately 1,500 different missense TP53 variants have been identified, ranging from several hot spots at positions 175, 248 or 273, reported several thousand times in many different tumors, to infrequent mutants detected at very low frequencies. (9) On the other hand, more than 4,000 TP53 variants are frameshift events leading to incorrect protein synthesis. This observation raises two important issues that have not been fully resolved. The first issue concerns the pathogenicity of all of these variants. Although there is no longer any doubt about the loss of function of the various hot spot variants, the loss of function of less frequent variants, particularly those that have been described at very low frequencies, remains unclear. (26) This is a key issue for genetic counseling, as the use of NGS has led to the discovery of very rare novel TP53 variants of unknown significance (VUS) in the normal population. (27) Multiple methodologies have been developed to assess the pathogenicity of TP53 variants, but their specificities and sensitivities remain low for uncommon variants. (28–30)

The second issue concerns the heterogeneity of TP53 variants. Missense mutant proteins exhibit severely impaired transcriptional activity as well as a gain of oncogenic activities that promote tumorigenesis, leading to the notion that tumors are addicted to mutant TP53. (22, 31, 32) Furthermore, a wealth of *in vitro* data as well as data from animal models indicate that the oncogenic activities of TP53 variants are heterogeneous and can vary according to the tissue type and the genetic background of the cells. (33) (34) (35, 36) Classifying TP53 status as either “wild-type” or “mutant” is therefore an oversimplification, as TP53-null tumors due to loss of TP53 expression have a different phenotype compared to tumors overexpressing an oncogenic TP53 variant.

Clinical relevance of *TP53* mutation

Somatic TP53 mutation in human tumors

The predictive and prognostic value of TP53 status in various types of cancer has been the subject of several thousand studies with controversial findings and limited clinical application and a review of this literature is beyond the scope of this paper. (37–39) These discordant results are due to multiple causes, such as the methodology and strategy used to assess TP53 status, the heterogeneity of tumor types, the genetic background of the tumor and the large number of different TP53 variants.

To circumvent some of these problems, TP53 variants have been tentatively divided into multiple categories according to their localization on the protein, the type of mutation (missense versus indel) or the evolutionary conservation of the mutated residue. Although some of these classifications have improved the clinical value of TP53 status for head and neck cancer (30), breast carcinoma (40) or diffuse large B-cell lymphoma (41), no clear rationale to definitively score TP53 variants has yet been defined.

One of the best examples of the clinical value of TP53 status is chronic lymphocytic leukemia. Although the frequency of TP53 variants is very low in asymptomatic patients, the presence of TP53 variants is usually associated with poor prognosis characterized by advanced clinical stage, rapid disease progression, chemoresistance, and shorter overall survival. (42) The recent CLL8 trial identified *TP53* variants as one of the strongest prognostic markers in patients receiving standard-of-care first-line therapy.(43) A European consortium (European Research Initiative on CLL, ERIC) has been created to develop and standardize *TP53* mutation analysis in CLL to allow better patient stratification.(44)

The development of liquid biopsies and analysis of circulating cell-free tumor DNA (ctDNA) as a surrogate for tumor genotyping has raised renewed interest in *TP53* variants, as the high gene mutation rate makes TP53 an attractive biomarker.(45) ctDNA analysis during therapy can provide early information about treatment resistance related to the emergence of *TP53* variants in response to the selective pressure of therapies. In many tumor types, such as lung, gastric, high-grade serous ovarian, or breast carcinoma, *TP53* variants are an early event that can be detected in ctDNA from patients with early-stage disease.(46, 47) The possible role of detection of *TP53* variants in ctDNA from individuals at high risk of cancer, allowing early clinical diagnosis and resulting in a higher cure rate, constitutes an exciting challenge for the future.

Because of the extremely high frequency of missense variants, the oncogenic gain of function of many variants and the fact that cancer cells overexpress the mutant protein, TP53 is a promising target for the development of therapies designed to induce inhibition or restoration of TP53 function by small molecules. (48) This prospect is supported by recent studies showing that reconstitution of the function of TP53 leads to the suppression of established tumors in mouse models. (49, 50) Molecules targeting mutant TP53 have been developed and are currently at the stage of clinical trials. (51) Components of the various pathways leading to accumulation of mutant TP53 such as hsp90 (52) or gain of function such as TP73, can also be targeted.(53) More than 150 clinical trials related to TP53 pathways are currently underway, including the use of novel molecules that specifically target mutant TP53.(54)

Germline TP53 variants in hereditary cancer predisposition syndromes

Germline *TP53* variants were first identified in individuals from families with LFS (55). LFS is a rare autosomal dominant syndrome, in which patients are predisposed to a wide variety of cancer types, with a young age at onset of malignancies, and the potential for multiple primary cancer sites during the affected individual's lifetime. (56) A Li-Fraumeni Like (LFL) syndrome with less stringent criteria than LFS was subsequently described. The frequency of *TP53* variants in LFS and LFL is 70% and 20-40%, respectively. LFS and LFL present a similar spectrum of germline and somatic TP53 variants with missense and indel variants scattered throughout the gene. The frequency of *de novo TP53* mutation (creating variants in the germ cells of one of the parents or in the fertilized egg) has been estimated to be as high as 30%, which is very high compared to the frequency of other tumor suppressor genes such as *BRCA1/2* (less than 5%). (57) Identification of *TP53* germline variants in LFS and LFL could potentially be beneficial for individual patients by allowing early cancer detection and prevention. (58)

A specific pathogenic germline variant arising from a founder event (c.1010G>A, p.(Arg337His)) has been identified in Brazilian children with adrenal cortical carcinomas

(ACC).(59) The prevalence of this variant is particularly high in Southern Brazil, where it can be as high as 0.3% in the general population (60) and is also common in patients with LFS and LFL from this geographical region. (61)

Recent studies have detected germline *TP53* variants in various cohorts of BRCA1/2-negative patients with early onset of breast cancer, indicating that the TP53 gene should be added to the cancer gene panel used for screening in these patients.(62–65)

Assessing TP53 status in human cancer

TP53 mutation analysis has now reached the clinical practice phase, as cancer patients are likely to benefit from this information. Somatic *TP53* variants were initially reported to cluster within DNA sequences encompassing exons 5 through 8, encoding the core DNA-binding domain of the protein. (6, 66) The majority of subsequent studies therefore exclusively focused on these regions, introducing a major bias with under-representation of variants that may occur in other regions of the gene. Over the last decade, most sequencing centers encompass the entire coding region of the gene, and this expanded coverage, together with the recent use of next generation sequencing (NGS) that covers all *TP53* exons, has revealed that up to 10% of *TP53* variants are localized in exons 2–4 and exons 9–11. (9) Of note, the spectrum of these variants differs from that of variants occurring in exons 5-8, as they mostly consist of indels that usually lead to a TP53-null phenotype. (9) The discovery and validation phases have clearly demonstrated the pathogenicity of these variants as well as their clinical utility; screening exons 2 to 11 is now highly recommended (Figure 2).

For a long time, the *TP53* gene was considered to be expressed as a single protein of uniform size (mRNA derived from exons 2-11, encoding 393 amino acids). However, the more complex architecture and expression pattern of the *TP53* gene has only been recognized in recent years. *TP53* mobilizes various mechanisms to transcribe at least eight different mRNA isoforms, which are generated by alternative splicing or alternative promoter usage. (67, 68) Collectively, these mRNAs have the potential to give rise to up to 12 different proteins, although the exact expression level, tissue distribution and biological function of each of these protein variants are poorly understood. This complex expression pattern implies that sequences located in TP53 introns and involved in the production of alternative forms of the protein may have a critical impact on overall biological functions of p53, and may therefore be important target regions for somatic or germline variants. Mouse models have shown that constitutive expression of a short TP53 isoform lacking the transactivation domain (Δ122p53) leads to chronic inflammation and a different and more aggressive tumor spectrum compared to *TP53*-null mice, suggesting that this isoform could act as a dominant oncogene. (69)

Intron 9 of *TP53* is a typical example of this type of situation, as it has now been clearly established that intron 9 contains two novel alternative exons, each one encoding a different carboxy-terminus for the p53 protein (Figure 2).(70) The biological function of these novel p53 protein isoforms, p53β and p53γ, has not yet been elucidated. Both proteins lack part of the oligomerization domain and have different transcriptional activities compared to full-length p53.(71)

Because of the bias towards screening for somatic variants exclusively in exons 5-8, these alternative exons have been excluded from most studies that used Sanger sequencing to assess variants. The increasing use of NGS strategies that address a broader range of sequences within the *TP53* locus demonstrates that significant variants may occur within these alternative exons. The latest version of the UMD *TP53* database containing 66,000 *TP53* variants derived from 3,900 curated and annotated publications including recent whole genome sequencing studies was released in November 2015. Analysis of this database identified several somatic non-synonymous variants in the coding region of exon 9β and five non-synonymous variants in the coding region of exon 9γ (Figure 3 and Table 1). Furthermore, two variants in the untranslated region of exon 9β and four variants in a splice signal common to both alternative mRNA isoforms were also identified (Figure 3 and Table 1). Analysis of the latest issue of dbSNP also showed that numerous synonymous and non-synonymous germline variants are localized in introns 9β and γ. The clinical significance of these variants is unknown at the present time, but their discovery warrants further analysis to validate whether or not screening of this region could be important to determine *TP53* status.

Another example of the importance of including *TP53* intronic sequences in mutation screening strategies is the identification of a hotspot region for rearrangements occurring in intron 1. More than 20 years ago, recurrent rearrangements in *TP53* intron 1 (~10 kb) were identified by Southern blot, but at the time this information was not included in guidelines for mutation screening.(72, 73) A recent study of intron 1 rearrangements found co-segregation with cancer risk in four generations of a family with LFS features, suggesting this genetic alteration may predispose to a wide range of cancers. (74) However, intron 1 rearrangements have been observed in only one type of sporadic cancer, osteosarcoma, where they occur in about 50% of cases. (74, 75) Of note, osteosarcoma has long been considered to be a type of cancer in which missense *TP53* variants were relatively rare, and it has been proposed that amplification of *MDM2*, rather than *TP53* variants, was a key mechanism for inactivation of the p53 protein in these cancers. (76) The detection of intronic rearrangements in a large subset of human osteosarcoma suggests that this cancer should also be considered to have a high rate of somatic *TP53* aberrations. In most cases analyzed to date, rearrangements in intron 1 led to balanced translocations involving different chromosomes, apparently without preference for a specific translocation product. The sites of breakpoints for rearrangements in intron 1 currently remain unclear. Ribi et al have documented seven rearrangements that all occurred within a defined region of 1.7 kb.(74) In contrast, in another study, Chen et al.(75) identified breakpoints occurring across the entire sequence of intron 1.

Several N-terminally truncated *TP53* isoforms are encoded by transcripts generated by a promoter localized in intron 4 of the *TP53* gene (Figure 2). It is conceivable that variants localized in this region would impair the synthesis of several TP53 isoforms. This question has yet to be resolved and a discovery phase will be necessary to investigate further.

<u>The importance of rigorous description of genetic variants and their effects</u>

Most researchers and clinicians like to describe genetic variants in a tangible way in terms of the protein. Amino acid names are more distinct and the numbers of amino acids in reference sequences is less than the number of nucleotides in the corresponding reference sequences. Due to the complexity of the human genome and the existence of

genes with multiple transcripts and protein isoforms, description of the numerous variants associated with genetic diseases has become complicated and can lead to erroneous descriptions and growing confusion in the genetics community. For more than 15 years, the Human Genome Variation Society (HGVS) has provided guidelines for variant terminology and nomenclature (den Dunnen JT and Antonarakis SE (2000). Hum. Mutat. 15:7-12)(Table 2A to D). The consistent use of a uniform nomenclature in the management of DNA sequence variations is critical for concise communication of diagnostic testing and genetic risk assessment. The importance of nomenclature has been recognized in the standards and guidelines for the interpretation of sequence variants recently published by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) (77). These guidelines are partly based on the HGVS guidelines for variant terminology (Table 2A) and take into account the fact that the meaning of certain terms may change as a result of changes in the contents of databases. The short genetic variation database, dbSNP, originally contained only high-frequency variants, but started to accept disease-causing variants and somatic variants with build 134 in 2011. Many researchers still regard dbSNP entries as variants with no functional effects, which has become a major source of confusion.

It is important to realize that genetic variants are mainly detected at the nucleotide level by DNA and RNA sequencing and should therefore be reported in terms of DNA and RNA sequences (Table 2B) to avoid the assumption that sequence variants do not alter gene expression or splicing. Reporting variants at a level other than where they were detected should be regarded as a form of interpretation. Variant descriptions in terms of protein and RNA (when RNA was not sequenced) should therefore reflect this fact by using parentheses flanking the description. Correct interpretation of variants and their effects as determined in functional assays is important for optimal patient care. It is currently unclear which transcripts and protein isoforms have been assessed by the various assays. The ability of assays to detect the different transcripts and protein isoforms must therefore be validated. After validation of the assay, the results of clinical and scientific reports should be accompanied by specification of the transcripts and protein isoforms potentially detected by the assay Table 2C).

Conclusions

Accurate assessment of *TP53* status is essential for optimal patient care, but several major questions remain unresolved (Box 1D). The recent discovery of *TP53* variants within regions outside the sequences encoding the canonical form of the p53 protein call for reconsideration of the guidelines for TP53 mutation screening in cancer patients. The use of NGS readily allows increased coverage of *TP53* sequences with no significant increase in cost or analysis time. However, implementing *TP53* intronic and alternative exonic sequences in NGS depends on the selection of appropriate regions by probes deduced from databases derived from the Consensus Coding Sequence Project (CCDS) or other similar databases. Only the recent versions of these databases include full information on alternative *TP53* exons, but it remains unclear whether this information is taken into account by the manufacturers of the various commercial products used for exome sequencing. Moreover, many standard bioinformatic pipelines used for the identification of somatic variants are tailored to exclude intronic *TP53* variants because they were not thought to have any functional significance. Therefore, large-scale studies on the precise clinical significance of *TP53* variants in introns and alternative exons are now required to

improve our understanding of the significance of these regions (Table 2D). In the meantime, a pragmatic recommendation would be to consider the entire sequence of the *TP53* gene for mutation screening strategies using NGS in sporadic cancers as well as in the germline of subjects who meet the criteria for *TP53* mutation testing. (78–80)

References

Nomenclature for the description of sequence variants: Human Genome Variation Society. https://www.hgvs.org/mutnomen. Accessed 5 November 2015

1. McDermott U, Downing JR, Stratton MR. Genomics and the continuum of cancer care. N Engl J Med. 2011;364:340-50.
2. Kandoth C, McLellan MD, Vandin F et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502:333-9.
3. Soussi T. Locus-specific databases in cancer: what future in a post-genomic era? The TP53 LSDB paradigm. Hum Mutat. 2014;35:643-53.
4. Holderfield M, Deuker MM, McCormick F, McMahon M. Targeting RAF kinases for cancer therapy: BRAF-mutated melanoma and beyond. Nat Rev Cancer. 2014;14:455-67.
5. Baker SJ, Fearon ER, Nigro JM et al. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. Science. 1989;244:217-21.
6. Takahashi T, Nau MM, Chiba I et al. p53: a frequent target for genetic abnormalities in lung cancer. Science. 1989;246:491-4.
7. Soussi T, Ishioka C, Claustres M, Beroud C. Locus-specific mutation databases: pitfalls and good practice based on the p53 experience. Nat Rev Cancer. 2006;6:83-90.
8. Olivier M, Hollstein M, Hainaut P. TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. Cold Spring Harb Perspect Biol. 2010;2:a001008.
9. Leroy B, Anderson M, Soussi T. TP53 mutations in human cancer: database reassessment and prospects for the next decade. Hum Mutat. 2014;35:672-88.
10. Nichols KE, Malkin D. Genotype Versus Phenotype: The Yin and Yang of Germline TP53 Mutations in Li-Fraumeni Syndrome. J Clin Oncol. 2015;33:2331-3.
11. Harris SL, Levine AJ. The p53 pathway: positive and negative feedback loops. Oncogene. 2005;24:2899-908.
12. Freed-Pastor WA, Prives C. Mutant p53: one name, many proteins. Genes Dev. 2012;26:1268-86.
13. Wang SJ, Gu W. To be, or not to be: functional dilemma of p53 metabolic regulation. Curr Opin Oncol. 2014;26:78-85.
14. Candi E, Agostini M, Melino G, Bernassola F. How the TP53 family proteins TP63 and TP73 contribute to tumorigenesis: regulators and effectors. Hum Mutat. 2014;35:702-14.
15. Stindt MH, Muller PA, Ludwig RL, Kehrloesser S, Dötsch V, Vousden KH. Functional interplay between MDM2, p63/p73 and mutant p53. Oncogene. 2014
16. Marine JC. MDM2 and MDMX in cancer and development. Curr Top Dev Biol. 2011;94:45-75.
17. Green DR, Kroemer G. Cytoplasmic functions of the tumour suppressor p53. Nature. 2009;458:1127-30.
18. Hock AK, Vousden KH. Tumor Suppression by p53: Fall of the Triumvirate? Cell. 2012;149:1183-5.
19. Lee P, Vousden KH, Cheung EC. TIGAR, TIGAR, burning bright. Cancer Metab. 2014;2:1.
20. Kruiswijk F, Labuschagne CF, Vousden KH. p53 in survival, death and metabolic health: a lifeguard with a licence to kill. Nat Rev Mol Cell Biol. 2015;16:393-405.
21. Bieging KT, Attardi LD. Cancer: A piece of the p53 puzzle. Nature. 2015;520:37-8.
22. Soussi T, Wiman KG. TP53: an oncogene in disguise. Cell Death Differ. 2015;22:1239-49.
23. Gembarska A, Luciani F, Fedele C et al. MDM4 is a key therapeutic target in cutaneous melanoma. Nat Med. 2012;18:1239-47.
24. Eischen CM, Lozano G. The Mdm network and its regulation of p53 activities: a rheostat of cancer risk. Hum Mutat. 2014;35:728-37.

25. Qi J, Singh S, Hua WK et al. HDAC8 Inhibition Specifically Targets Inv(16) Acute Myeloid Leukemic Stem Cells by Restoring p53 Acetylation. Cell Stem Cell. 2015;17:597-610.
26. Cooper GM. Parlez-vous VUS. Genome Res. 2015;25:1423-6.
27. Amendola LM, Dorschner MO, Robertson PD et al. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. Genome Res. 2015;25:305-15.
28. Petitjean A, Mathe E, Kato S et al. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. Hum Mutat. 2007;28:622-9.
29. Carlsson J, Soussi T, Persson B. Investigation and prediction of the severity of p53 mutants using parameters from structural calculations. FEBS J. 2009;276:4142-55.
30. Osman AA, Neskey DM, Katsonis P et al. Evolutionary Action Score of TP53 Coding Variants Is Predictive of Platinum Response in Head and Neck Cancer Patients. Cancer Res. 2015;75:1205-15.
31. Bieging KT, Mello SS, Attardi LD. Unravelling mechanisms of p53-mediated tumour suppression. Nat Rev Cancer. 2014;14:359-70.
32. Oren M, Rotter V. Mutant p53 gain-of-function in cancer. Cold Spring Harb Perspect Biol. 2010;2:a001107.
33. Jackson JG, Lozano G. The mutant p53 mouse as a pre-clinical model. Oncogene. 2013;32:4325-30.
34. Olive KP, Tuveson DA, Ruhe ZC et al. Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. Cell. 2004;119:847-60.
35. Lang GA, Iwakuma T, Suh YA et al. Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. Cell. 2004;119:861-72.
36. Strano S, Dell'Orso S, Di Agostino S, Fontemaggi G, Sacchi A, Blandino G. Mutant p53: an oncogenic transcription factor. Oncogene. 2007;26:2212-9.
37. Robles AI, Harris CC. Clinical outcomes and correlates of TP53 mutations and cancer. Cold Spring Harb Perspect Biol. 2010;2:a001016.
38. Petitjean A, Achatz MI, Borresen-Dale AL, Hainaut P, Olivier M. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. Oncogene. 2007;26:2157-65.
39. Bertheau P, Lehmann-Che J, Varna M et al. p53 in breast cancer subtypes and new insights into response to chemotherapy. Breast. 2013;22 Suppl 2:S27-9.
40. Olivier M, Langerod A, Carrieri P et al. The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer. Clin Cancer Res. 2006;12:1157-67.
41. Young KH, Leroy K, Moller MB et al. Structural profiles of TP53 gene mutations predict clinical outcome in diffuse large B-cell lymphoma: an international collaborative study. Blood. 2008;112:3088-98.
42. Gruber M, Wu CJ. Evolving understanding of the CLL genome. Semin Hematol. 2014;51:177-87.
43. Stilgenbauer S, Schnaiter A, Paschka P et al. Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. Blood. 2014;123:3247-54.
44. Pospisilova S, Gonzalez D, Malcikova J et al. ERIC recommendations on TP53 mutation analysis in chronic lymphocytic leukemia. Leukemia. 2012;26:1458-61.
45. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. Nat Rev Clin Oncol. 2013;10:472-84.
46. Izumchenko E, Chang X, Brait M et al. Targeted sequencing reveals clonal genetic changes in the progression of early lung neoplasms and paired circulating DNA. Nat Commun. 2015;6:8258.
47. Bettegowda C, Sausen M, Leary RJ et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014;6:224ra24.
48. Muller PA, Vousden KH. Mutant p53 in cancer: new functions and therapeutic opportunities. Cancer Cell. 2014;25:304-17.

49. Ventura A, Kirsch DG, McLaughlin ME et al. Restoration of p53 function leads to tumour regression in vivo. Nature. 2007;445:661-5.
50. Martins CP, Brown-Swigart L, Evan GI. Modeling the therapeutic efficacy of p53 restoration in tumors. Cell. 2006;127:1323-34.
51. Lehmann S, Bykov VJ, Ali D et al. 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. 2012;30:3633-9.
52. Alexandrova EM, Yallowitz AR, Li D et al. Improving survival by exploiting tumour dependence on stabilized mutant p53 for treatment. Nature. 2015;523:352-6.
53. Zhang S, Zhou L, Hong B et al. Small-Molecule NSC59984 Restores p53 Pathway Signaling and Antitumor Effects against Colorectal Cancer via p73 Activation and Degradation of Mutant p53. Cancer Res. 2015;75:3842-52.
54. Cheok CF, Verma CS, Baselga J, Lane DP. Translating p53 into the clinic. Nat Rev Clin Oncol. 2011;8:25-37.
55. Malkin D, Li FP, Strong LC et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science. 1990;250:1233-8.
56. McBride KA, Ballinger ML, Killick E et al. Li-Fraumeni syndrome: cancer risk assessment and clinical management. Nat Rev Clin Oncol. 2014;11:260-71.
57. Gonzalez KD, Buzin CH, Noltner KA et al. High frequency of de novo mutations in Li-Fraumeni syndrome. J Med Genet. 2009;46:689-93.
58. Villani A, Tabori U, Schiffman J et al. Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: a prospective observational study. Lancet Oncol. 2011;12:559-67.
59. Ribeiro RC, Sandrini F, Figueiredo B et al. An inherited p53 mutation that contributes in a tissue-specific manner to pediatric adrenal cortical carcinoma. Proc Natl Acad Sci U S A. 2001;98:9330-5.
60. Legal EF, Ascurra M, Custódio G et al. Prevalence of an inherited cancer predisposition syndrome associated with the germ line TP53 R337H mutation in Paraguay. Cancer Epidemiol. 2015;39:166-9.
61. Giacomazzi J, Selistre SG, Rossi C et al. Li-Fraumeni and Li-Fraumeni-like syndrome among children diagnosed with pediatric cancer in Southern Brazil. Cancer. 2013;119:4341-9.
62. McCuaig JM, Armel SR, Novokmet A et al. Routine TP53 testing for breast cancer under age 30: ready for prime time? Fam Cancer. 2012;11:607-13.
63. Maxwell KN, Wubbenhorst B, D'Andrea K et al. Prevalence of mutations in a panel of breast cancer susceptibility genes in BRCA1/2-negative patients with early-onset breast cancer. Genet Med. 2015;17:630-8.
64. Li J, Meeks H, Feng BJ et al. Targeted massively parallel sequencing of a panel of putative breast cancer susceptibility genes in a large cohort of multiple-case breast and ovarian cancer families. J Med Genet. 2016;53:34-42.
65. Tung N, Battelli C, Allen B et al. Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. Cancer. 2015;121:25-33.
66. Nigro JM, Baker SJ, Preisinger AC et al. Mutations in the p53 gene occur in diverse human tumour types. Nature. 1989;342:705-8.
67. Bourdon JC, Fernandes K, Murray-Zmijewski F et al. p53 isoforms can regulate p53 transcriptional activity. Genes Dev. 2005;19:2122-37.
68. Marcel V, Dichtel-Danjoy ML, Sagne C et al. Biological functions of p53 isoforms through evolution: lessons from animal and cellular models. Cell Death Differ. 2011;18:1815-24.
69. Slatter TL, Hung N, Campbell H et al. Hyperproliferation, cancer, and inflammation in mice expressing a Δ133p53-like isoform. Blood. 2011;117:5166-77.
70. Flaman JM, Waridel F, Estreicher A et al. The human tumour suppressor gene p53 is alternatively spliced in normal cells. Oncogene. 1996;12:813-8.

71. Surget S, Khoury MP, Bourdon JC. Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. Onco Targets Ther. 2013;7:57-68.
72. Masuda H, Miller C, Koeffler HP, Battifora H, Cline MJ. Rearrangement of the p53 gene in human osteogenic sarcomas. Proc Natl Acad Sci U S A. 1987;84:7716-9.
73. Miller CW, Aslo A, Tsay C et al. Frequency and structure of p53 rearrangements in human osteosarcoma. Cancer Res. 1990;50:7950-4.
74. Ribi S, Baumhoer D, Lee K et al. TP53 intron 1 hotspot rearrangements are specific to sporadic osteosarcoma and can cause Li-Fraumeni syndrome. Oncotarget. 2015;6:7727-40.
75. Chen X, Bahrami A, Pappo A et al. Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. Cell Rep. 2014;7:104-12.
76. Oliner JD, Kinzler KW, Meltzer PS, George DL, Vogelstein B. Amplification of a gene encoding a p53-associated protein in human sarcomas. Nature. 1992;358:80-3.
77. Richards S, Aziz N, Bale S et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17:405-24.
78. Tinat J, Bougeard G, Baert-Desurmont S et al. 2009 version of the Chompret criteria for Li Fraumeni syndrome. J Clin Oncol. 2009;27:e108-9; author reply e110.
79. Hettmer S, Archer NM, Somers GR et al. Anaplastic rhabdomyosarcoma in TP53 germline mutation carriers. Cancer. 2014;120:1068-75.
80. Bougeard G, Renaux-Petel M, Flaman JM et al. Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers. J Clin Oncol. 2015;33:2345-52.
81. Soussi T, Leroy B, Taschner PE. Recommendations for analyzing and reporting TP53 gene variants in the high-throughput sequencing era. Hum Mutat. 2014;35:766-78.
82. Kern SE, Winter JM. Elegance, silence and nonsense in the mutations literature for solid tumors. Cancer Biol Ther. 2006;5:349-59.
83. Stacey SN, Sulem P, Jonasdottir A et al. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. Nat Genet. 2011;43:1098-103.
84. Iggo R, Rudewicz J, Monceau E et al. Validation of a yeast functional assay for p53 mutations using clonal sequencing. J Pathol. 2013;231:441-8.

Table 2 / Box 2A How to avoid confusion - definitions of genetic variation terminology according to the Human Genome Variation Society

Recommended standard terms:
1 *Variant*: every permanent *genetic change*.
Connotation: Neutral. Recommended by ACMG and AMP.
Avoid using *Mutation* to indicate the *variant* itself, because of its negative connotation due to frequent use as *disease-causing variant* instead of the broader concept *variant*. *Mutation* can be used to describe the **process** or **event** generating genetic variation.
Avoid using the term *Polymorphism*. In its original meaning: *a variant with a frequency of 1% or higher in the population.* Due to its frequency, considered to be *non-disease-causing.*
Avoid using the term *Single Nucleotide Polymorphism (SNP)*: *variant present in dbSNP*. dbSNP now contains other types of short sequence variants. In addition, rare variants causing hereditary disease and somatic variants are accepted. In conclusion: no longer synonymous with *non-disease-causing.*
2 Affects function: HGVS recommended modifier alternative for the term *Pathogenic* used to indicate a *disease-causing* effect. The term *pathogenic* is inappropriate for use with traits and creates confusion when used without mentioning *specific context (in combination with a similar variant on the same allele)* or *conditions (when inherited from the father, imprinted)* necessary to observe the functional effect causing disease. Germline variants in tumor suppressor genes can only be considered as having functional effects when somatic second hits inactivate the second allele in tumors.
HGVS recommended five-tier variant classification system: *affects function, probably affects function, unknown, probably does not affect function (or probably no functional effect), or does not affect function (no functional effect).*
The ACMG and AMP guidelines still include the five-tier variant classification system relevant to Mendelian disorders *pathogenic, likely pathogenic, uncertain significance, likely benign, or benign*, but recommend providing the condition and inheritance pattern to clarify the context in reports.

Table 2 / Box 2B Describing *TP53* variants unambiguously

1. Use the official HGNC gene symbol: *TP53*
2. Specify the **genomic** reference sequence. For next generation sequencing, use the chromosomal accession and version number NC_000017.10 for genome build GRCh37.p13 or NC_000017.11 for genome build GRCh38.p2. Do not replace by chr 17!
For diagnostic purposes, preferably use the stable Locus Reference Genomic sequence LRG_321. See Dalgleish et al. (http://www.genomemedicine.com/content/pdf/gm145.pdf)
3. Use HGVS nomenclature (http://www.hgvs.org/mutnomen) to describe genetic variants at all different levels
4. All variants **must** be reported at the genomic DNA (g.) and coding DNA level (c.). The genomic reference sequence **must** cover the entire gene including the promoter and the 5′ and 3′ untranslated regions.
Example: genomic description LRG_321t1:g.18749G>A, coding DNA: LRG_321t1:c.818G>A
Alternative: the **accession and version number** of the corresponding RefSeq Gene NG_017013.2. Note: the annotation of this reference sequence may change without version update.
5. All variants **should** be reported at the RNA level (r.).
Example: LRG_321t1:r.818g>a (cDNA sequenced) or LRG_321t1:r.(818g>a) (cDNA not sequenced)
6. All variants **should** be reported using HGVS nomenclature at the protein level (p.)
Example: LRG_321p1:p.Arg273His (cDNA sequenced) or LRG_321p1:p.(Arg273His) (cDNA not sequenced)
7. **Predicted** effects at the RNA and protein level should be indicated **in parentheses**
8. A *dbSNP entry* (rs number) is insufficient to unambiguously describe the genetic variant found in an individual, since the alleles are not specified.
9. *Somatic variant*: Variant generated by a somatic *mutation event*. Variants should **only** be labeled as *somatic* when normal tissue from the same individual tested negative. When normal tissue from the same individual tested positive, the test has revealed a *germline variant*. When normal tissue from the same individual was unavailable and the variant has not been transmitted by one of the parents, the variant should be labeled as *detected in tumor (tissue)*.

Table 2/ Box 2C Eliminating sources of confusion when reporting assay results in the literature

Specification of transcripts and protein isoforms.
TP53 transcripts should be specified using LRG_321 followed by the t1 to t8 suffixes (http://ftp.ebi.ac.uk/pub/databases/lrgex/LRG_321.xml). TP53 protein isoforms should be specified using LRG_321 followed by the p1 to p13 suffixes. See Soussi et al., 2014 for examples. (81)
Variants in the *TP53* gene may affect its eight transcripts and twelve protein isoforms in different ways. Researchers should be aware of this variation when analyzing *TP53* functional effects using different assays at the RNA and protein level. If possible, the transcripts and protein isoforms analyzed by the assays should be specified when describing their results to avoid confusion. validation of the ability of common TP53 assays to detect various transcripts and protein isoforms could help to resolve existing discrepancies and seemingly contradictory data in the literature and databases.

Table 2/ Box 2D TP53 variants in human cancer: unresolved questions

Which cancer types and/or subtypes will benefit the most from determination of TP53 status?

To what degree does TP53 loss of function induced by targeting certain upstream or downstream components, such as MDM2 or MDM4 amplification or microRNA dysregulation, resemble that induced by TP53 variants?

What is the contribution of TP53 variant heterogeneity to the phenotype of the tumor? (missense versus indel variants, hot spot versus non-hot spot variants)

What is the contribution of TP53 isoforms to tumor phenotype?

What is the contribution of germline TP53 mutations in familial cancer unrelated to LFS and LFL?

Is there any tumor or cell type specificity for loss and/or gain of function of TP53 variants?

Which drugs would be the most effective on tumors with functional and non-functional TP53 pathways?

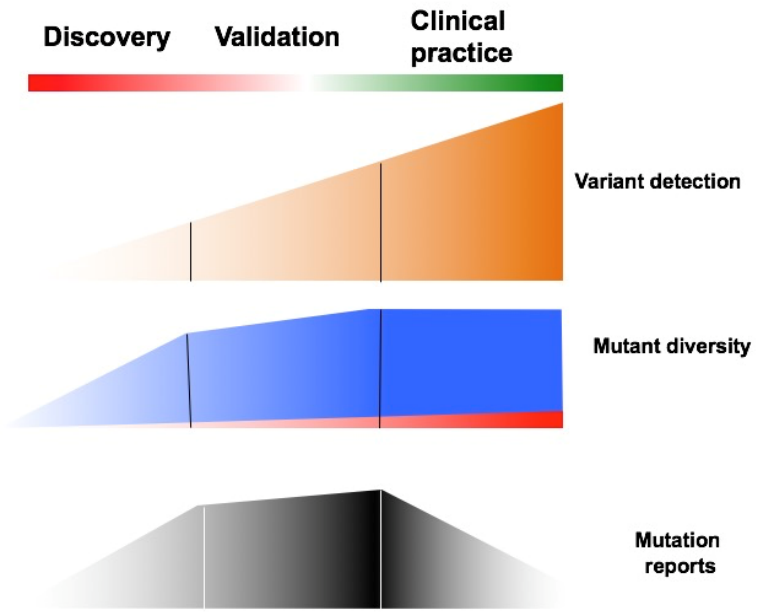


Figure 1. Cancer gene variant discovery and validation: pathway to clinical practice:

The discovery and validation of cancer gene variants follow several phases that resemble those used for other cancer biomarkers. During the discovery phase, publications precisely describe novel variants and discuss their potential pathogenicity in relation to the disease. A burst of studies then leads to the identification of novel and generally diverse mutants. This phase is commonly associated with parallel reports on the mutation rate and/or clinical novelties, often published in journals with a high impact factor.

Transition to the validation phase occurs rapidly when genetic and clinical data start to become redundant. During this phase, sequencing of multiple new clinical specimens mostly reveals variants that have already been described and mutant diversity will begin to reach a plateau. The length of this phase is highly dependent on the number of genetic events needed to modify the targeted gene. For oncogenes that require specific events to change their function, this number tends to be limited, since most of them will be missense variants targeting a critical functional region. In contrast, tumor suppressor genes may harbor a large number of genetic events, including nonsense variants, splice variants, as well as indels of varying size scattered throughout the gene.

This validation phase is vital, as it adds nuance and validates data from the discovery phase in a wide variety of clinical and/or geographical settings. Consequently, variants are either described in supplementary materials or quoted as unpublished data, leading to a decrease in reported variants. Except for a few very specific cases, the validation phase is accompanied by a decrease in the impact factor of the publishing journals. This decrease in descriptions of variants does not reflect their frequency in the disease or the incidence of their analysis, but rather a lack of interest and lack of novelty, introducing a bias against their publication. If the variant has no clinical significance, the number of studies will decrease rapidly, and then stop. It has also been observed that the validation phase is associated with an increase in inconsistent studies. An extensive analysis of the various flaws associated with the publication of variants was provided by Kern and Winter in their 2006 review. (82)

Finally, transition to the clinical practice phase then begins for variants of clinical interest. However, publications fall off, as service laboratories do not consider reporting them in the literature to be an essential part of their work. Descriptions of novel variants then become scarce.

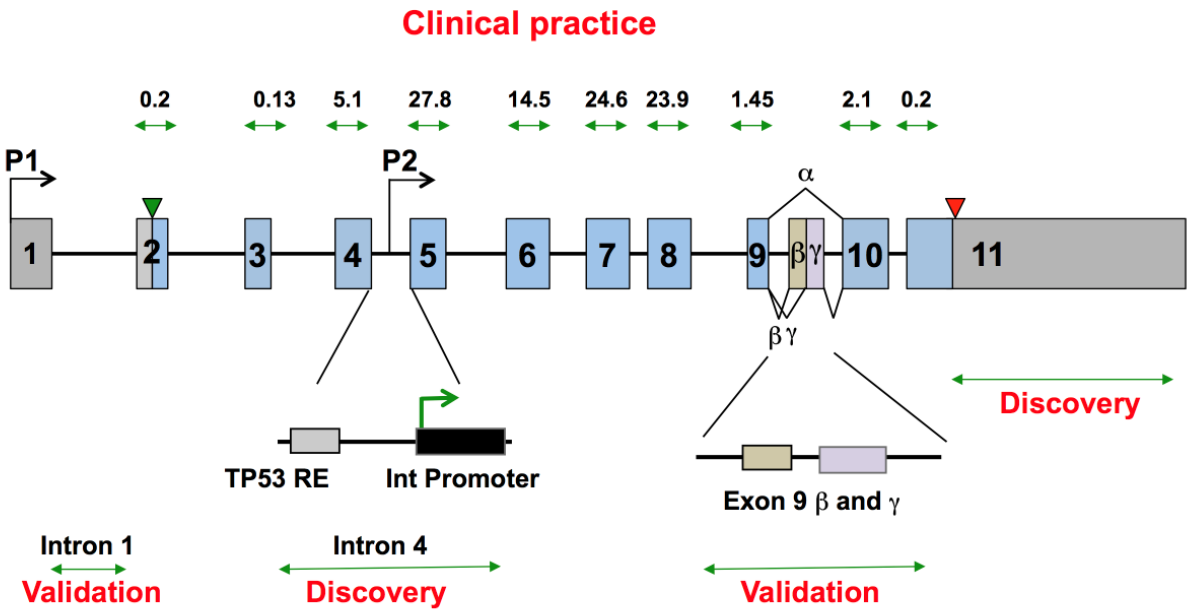


Figure 2. Assessment of *TP53* status.

Sequencing of the classical exons 2 to 11 (including splice junctions) is now mandatory, as discovery and validation phases have both demonstrated that variants are scattered throughout these exons. A few variants have been discovered in exons 9β and 9γ, their pathogenicity is currently unknown and more data are needed to assess their recurrence (Figure 1C and Table 1). Additional studies will need to be performed to ensure that these variants are true driver variants and not simply rare or passenger variants. The region of intron 4 contains the second internal promoter P2 that leads to the expression of 9 additional protein isoforms as well a TP53 response element (TP53RE). Whether or not variants in this region can alter *TP53* status and impair its tumor suppressor function remains unknown. Recent studies have shown that a rare variant, rs78378222 SNP, localized in the Poly A signal of the TP53 genes, leads to impaired 3'-end processing of mRNA and confers susceptibility to various types of cancer (red triangle) (83).

The validation phase for these various events could be performed *in silico* using data obtained from whole-genome sequencing of tumors, as these regions were most probably sequenced, but discarded by the various filtering processes used in the analytical pipelines.
The frequency of *TP53* variants in each coding exon is shown above each exon.

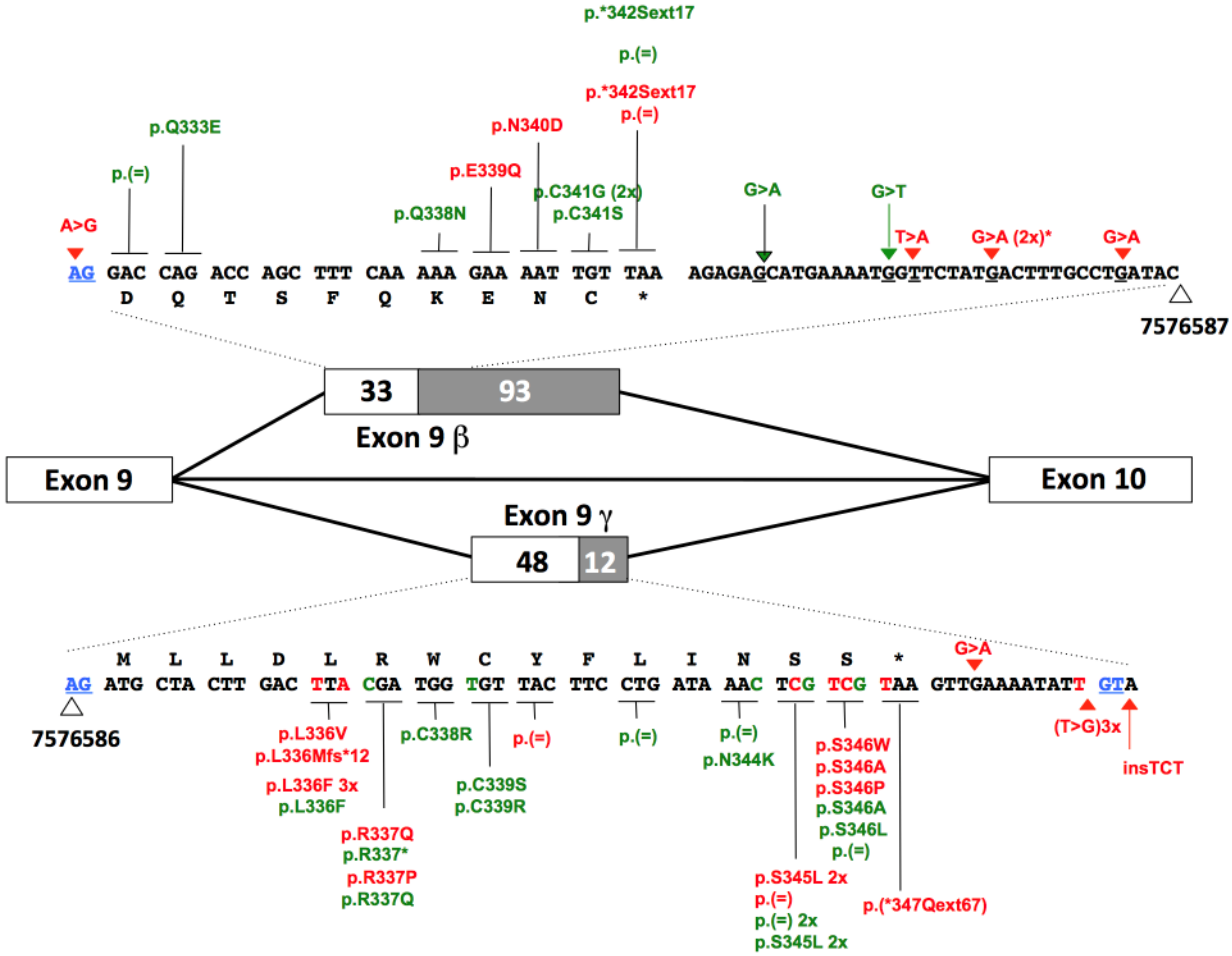


Figure 3. *TP53* variants detected in exon 9β and exon 9γ

Alternative splicing events in *TP53* intron 9 leading to the expression of β and γ isoforms. Exons 9β and 9γ have different splice acceptor sites, but share the same donor site (blue AG and GT sequence, respectively). Due to the marked overlap between the two exons, it is possible that substitutions localized in the translated region of exon 9γ impair exon 9β splicing. The main splicing event occurs between exon 9 and 10. Somatic variants detected in human cancer are shown in red, whereas germline SNPs from dbSNP are shown in green. See Table 1 for more information. The intron 9 splice donor site contains three single nucleotide substitutions and one insertion in three different tumors. The NC_000017.10:g.7576525A>C substitutions modify *TP53* splicing by leading to an unbalanced ratio of the various *TP53* mRNAs and a greater abundance of β isoforms. (84) Similarly, the three-nucleotide insertion detected in a lymphoma (NC_000017.10:g.7576522_7576523insCTT) probably has a deleterious effect on splicing. The substitution at the stop codon (NM_000546.5:c.993+314T>C) is predicted to add 17 novel amino acid residues to the 10 residues encoded by exon 9β. All these variations are concentrated within a narrow 150 bp region of intron 9, which is 2,800 nucleotides long. A total of 22 germline variants, shown in green, have been identified by mining the most recent versions of the dbSNP and the 1000-Genome databases. Several of these variants are identical to those found as somatic events in human tumors. It is not known whether these variants are neutral or associated with an increased risk of cancer. Overall, these data indicate that *TP53* exon 9β and exon 9γ are targeted by substitutions in human cancer and contain numerous germline variants. Due to the importance of *TP53* status in the evaluation of patients with multiple primary cancers or a strong family history of cancer, analysis of exons 9β and 9γ is now warranted to more clearly determine the clinical significance of germline variants in this region.

Table 1: *TP53* variants targeting specifically exon 9β and exon 9γ sequences and their splice sites.

Sample ID [a]/ **SNP_ID** [b]	Disease Name	Chromosomal reference NC_000017.10 [c]	Main TP53 transcript NM_000546.5 [d]	Exon 9β **transcript NM_001126114.2** [e]	Exon 9γ **transcript NM_001126113.2**[f]	Exon 9β **protein NP_001119586.1** [g]	Exon 9γ **protein NP_001119585.1** [h]
V6	Skin SCC	g.7576659A>G	c.993+194A>G	c.994-2A>G		p.(?)	
12a	Head and Neck Squamous Cell Carcinoma	g.7576636G>C	c.993+217G>C	c.1015G>C		p.(Glu339Gln)	
SA500637	Hepatocellular carcinoma	g.7576633A>G	c.993+220A>G	c.1018A>G		p.(Asn340Asp)	
SLN2522	Burkitt lymphoma	g.7576626A>C	c.993+227A>C	c.1025A>C		p.(*342Sext*17)	
RPMI	T-ALL	g.7576626A>G	c.993+227A>G	c.1025A>G		p.(=)	
XHDG17	Gallbladder carcinoma	g.7576601G>A	c.993+252G>A	c.*24G>A		p.(=)	
BRCA_UK_1	Breast carcinoma	g.7576591G>A	c.993+262G>A	c.*34G>A		p.(=)	
CLL117	B-Chronic Lymphocytic Leukemia	g.7576572T>G	c.993+281T>G		c.1006T>G		p.(Leu336Val)
3129	Esophageal Adenocarcinoma	g.7576570 A>C	c.993+283A>C		c.1008A>C		p.(Leu336Phe)
LINC-JP-DO23214	Hepatocellular carcinoma	g.7576570A>C	c.993+283A>C		c.1008A>C		p.(Leu336Phe)
LINC-JP-DO23506	Hepatocellular carcinoma	g.7576570A>C	c.993+283A>C		c.1008A>C		p.(Leu336Phe)
YUMUL	Melanoma	g.7576544C>T	c.993+309C>T		c.1034C>T		p.(Ser345Leu)
SJNBL197[i]	Neurobastoma	g.7576544C>T	c.993+309C>T		c.1034C>T		p.(Ser345Leu)
5-VS065-T1	Skin BCC	g.7576544C>T	c.993+309C>T		c.1034C>T		p.(Ser345Leu)
WD_02	Skin SCC	g.7576543G>A	c.993+310G>A		c.1035G>A		p.(Ser345Ser)
12-RS	Richter syndrome	g.7576542T>G	c.993+311T>G		c.1036T>G		p.(Ser346Ala)
TCGA-HT-A616	Glioma	g.7576541C>G	c.993+312C>G		c.1037C>G		p.(Ser346Trp)
SC-9007	Prostate carcinoma	g.7576539T>C	c.993+314T>C		c.1039T>C		p.(*347Glnext*67)
MEL-Ma-Mel-94	Melanoma	g.7576525T>G	c.993+328T>G		c.*12T>G		p.(=)
83[j]	Breast carcinoma	g.7576525T>G	c.993+328T>G		c.*12T>G		p.(=)
OV_AU_1	Ovarian carcinoma	g.7576525T>G	c.993+328T>G		c.*12T>G		p.(=)
ATH-2	Splenic marginal zone lymphoma	g.7576522_7576523insAGA	c.993+330_993+331insAGA		c.*12+2_*12+3insAGA		p.(=)
rs3021068 [k]		g.7576630T>G	c.993+223T>G	c.1021T>G		p.(Cys341Gly)	
rs17883348 [l]		g.7576619G>A	c.993+234G>A	c.*6G>A		p.(=)	
rs576532147 [k]		g.7576609G>T	c.993+244G>T	c.*16G>T		p.(=)	
rs554738122 [k]		g.7576569C>T	c.993+284C>T		c.1009C>T		p.(Arg337Ter)
rs1642789 [l]		g.7576563T>A	c.993+290T>A		c.1015T>A		p.(Cys339Ser)
rs200274944 [k]		g.7576546C>T	c.993+307C>T		c.1032C>T		p.(=)
rs201293647 [k]		g.7576543G>A	c.993+310G>A		c.1035G>A		p.(=)
rs372821099 [k]		g.7576540G>A	c.993+313G>A		c.1038G>A		p.(=)

Sample ID / SNP_ID [a]	Origin [b]	Chromosomal reference NC_000017.10 [c]	Main TP53 transcript NM_000546.5 [d]	Exon 9β transcript NM_001126114.2 [e]	Exon 9γ transcript NM_001126113.2 [f]	Exon 9β protein NP_001119586.1 [g]	Exon 9g protein NP_001119585.1 [h]
V6	Skin SCC	g.7576659A>G	c.993+194A>G	c.994-2A>G		p.(?)	
rs750031971	dbSNP	g.7576655C>T	c.993+198C>T	c.996C>T		p.(=)	
rs764851816	dbSNP	g.7576654C>G	c.993+199C>G	c.997C>G		p.Q333E	
rs761303879	dbSNP	g.7576637A>C	c.993+216A>C	c.1014A>C		p.K338N	
12a	Head and Neck SCC	g.7576636G>C	c.993+217G>C	c.1015G>C		p.E339Q	
SA500637	HCC	g.7576633A>G	c.993+220A>G	c.1018A>G		p.N340D	
JEN9 (rs3021068)	dbSNP	g.7576630T>G	c.993+223T>G	c.1021T>G		p.C341G	
rs3021068	dbSNP	g.7576630T>G	c.993+223T>G	c.1021T>G		p.C341G	
rs3021068	dbSNP	g.7576630T>A	c.993+223T>A	c.1021T>A		p.C341S	
SLN2522	Burkitt lymphoma	g.7576626A>C	c.993+227A>C	c.1025A>C		p.*342Sext17	
Au3	Melanoma	g.7576626A>G	c.993+227A>G	c.1025A>G		p.(=)	
rs764562217	dbSNP	g.7576626A>C	c.993+227A>C	c.1025A>C		p.*342Sext17	
rs761121529	dbSNP	g.7576625A>G	c.993+228A>G	c.1026A>G		p.(=)	
rs17883348	dbSNP	g.7576619G>A	c.993+234G>A				
rs576532147	dbSNP	g.7576609G>T	c.993+244G>T				
HPB-ALL [i]	T-ALL	g.7576607T>A	c.993+246T>A				
RPMI [i]	T-ALL	g.7576607T>A	c.993+246T>A				
XHDG17	Gallbladder ca.	g.7576601G>A	c.993+252G>A				
SA6251-BRCA-UK	Breast carcinoma	g.7576591G>A	c.993+262G>A				
BRCA_UK_1	Breast carcinoma	g.7576591G>A	c.993+262G>A				
117	B-Chronic lymphocytic leukemia	g.7576572T>G	c.993+281T>G		c.1006T>G		p.L336V
19	Uterine carcinosarcoma	g.7576572_7576568del	c.993+281_993+285del5		c.1010el5		p.L336Mfs*12
3129	Esophageal Adc	g.7576570A>C	c.993+283A>C		c.1008A>C		p.L336F
SA529425-LINC-JP	Hepatocellular carcinoma	g.7576570A>C	c.993+283A>C		c.1008A>C		p.L336F
SA529505-LINC-JP	Hepatocellular carcinoma	g.7576570A>C	c.993+283A>C		c.1008A>C		p.L336F
COSM1610827	dbSNP	g.7576570A>C	c.993+283A>C		c.1008A>C		p.L336F
rs554738122	dbSNP	g.7576569C>T	c.993+284C>T		c.1009C>T		p.R337*
DS-54750	Colorectal carcinoma	g.7576568G>A	c.993+285G>A		c.1010G>A		p.R337Q
15	Bladder carcinoma	g.7576568G>C	c.993+285G>C		c.1010G>C		p.R337P
rs771319678	dbSNP	g.7576568G>A	c.993+285G>A		c.1010G>A		p.R337Q
rs749361930	dbSNP	g.7576566T>C	c.993+287T>C		c.1012T>C		p.W338R
rs1642789	dbSNP	g.7576563T>C	c.993+290T>C		c.1015T>C		p.C339R
rs1642789	dbSNP	g.7576563T>A	c.993+290T>A		c.1015T>A		p.C339S
DS-53453	Colorectal carcinoma	g.7576558C>T	c.993+295C>T		c.1020C>T		p.(=)
rs770028766	dbSNP	g.7576554C>T	c.993+299C>T		c.1024C>T		p.(=)
rs200274944	dbSNP	g.7576546C>T	c.993+307C>T		c.1032C>T		p.(=)
rs200274944	dbSNP	g.7576546C>G	c.993+307C>G		c.1032C>G		p.N344K
YUMUL	Melanoma	g.7576544C>T	c.993+309C>T		c.1034C>T		p.S345L
SJNBL197 [j]	Neuroblastoma	g.7576544C>T	c.993+309C>T		c.1034C>T		p.S345L
5-VS065-T1	Skin basal cell carcinoma	g.7576544C>T	c.993+309C>T		c.1034C>T		p.S345L

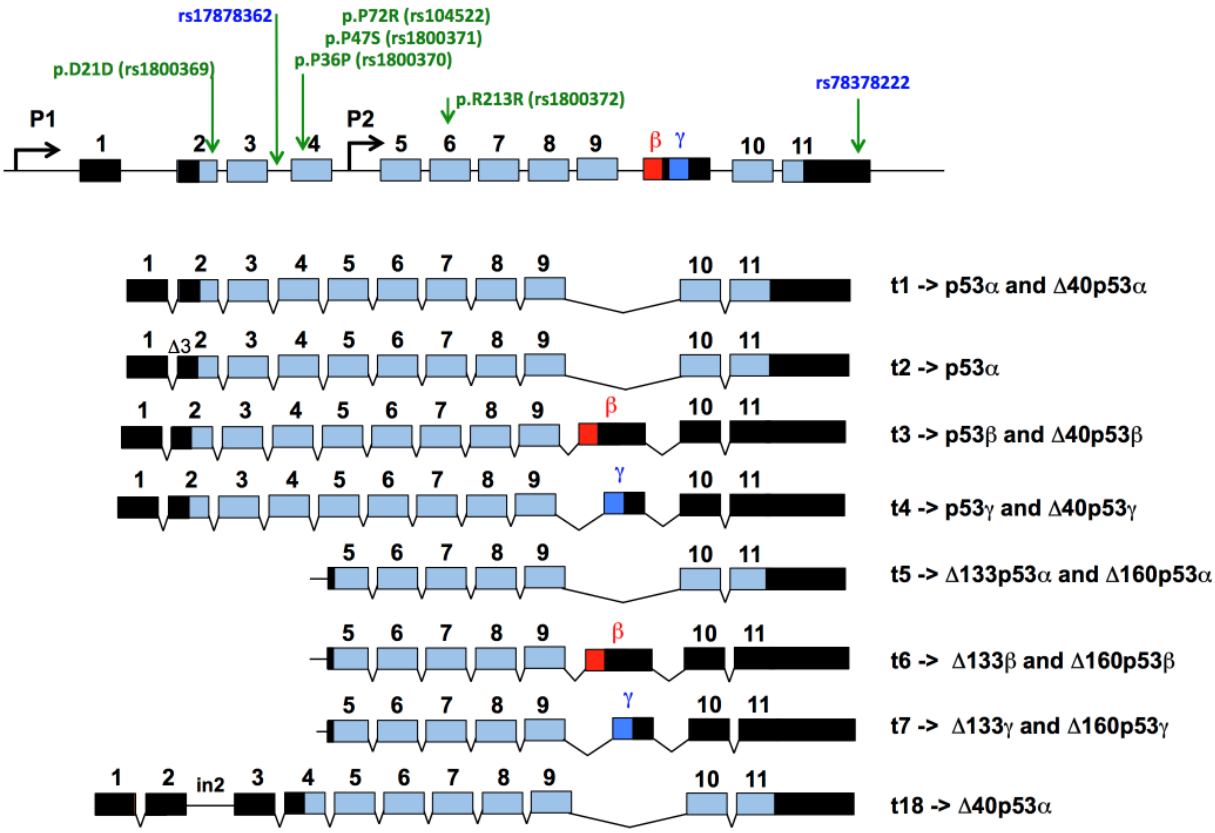
rs758194998	dbSNP	g.7576544C>T	c.993+309C>T	c.1034C>T	p.S345L
WD_06	Skin squamous cell carcinoma	g.7576543G>A	c.993+310G>A	c.1035G>A	p.(=)
rs201293647	dbSNP	g.7576543G>T	c.993+310G>T	c.1035G>T	p.(=)
rs201293647	dbSNP	g.7576543G>A	c.993+310G>A	c.1035G>A	p.(=)
12-RS	Richter syndrome	g.7576542T>G	c.993+311T>G	c.1036T>G	p.S346A
DS-53382	Colorectal carcinoma	g.7576542T>C	c.993+311T>C	c.1036T>C	p.S346P
COSM1731910	dbSNP	g.7576542T>G	c.993+311T>G	c.1036T>G	p.S346A
TCGA-HT-A616	Glioma (low grade)	g.7576541C>G	c.993+312C>G	c.1037C>G	p.S346W
rs756952434	dbSNP	g.7576541C>T	c.993+312C>T	c.1037C>T	p.S346L
rs372821099	dbSNP	g.7576540G>A	c.993+313G>A	c.1038G>A	p.(=)
SC_9007-Tumor	Prostate carcinoma	g.7576539T>C	c.993+314T>C	c.1039T>C	p.*347Qext67
CCRF-CEM	T-Acute lymphoblastic leukemia	g.7576533G>A	c.993+320G>A		
rs730882013	dbSNP	g.7576527_7576512del	c.993+326_993+341del16		
MEL-Ma-Mel-94	Melanoma	g.7576525T>G	c.993+328T>G		
83[k]	Breast carcinoma	g.7576525T>G	c.993+328T>G		
SA505836-OV-AU	Ovarian carcinoma	g.7576525T>G	c.993+328T>G		
ATH-2	Splenic marginal zone lymphoma	g.7576522TCT	c.993+331_993+332insTCT		

Table 1 legend

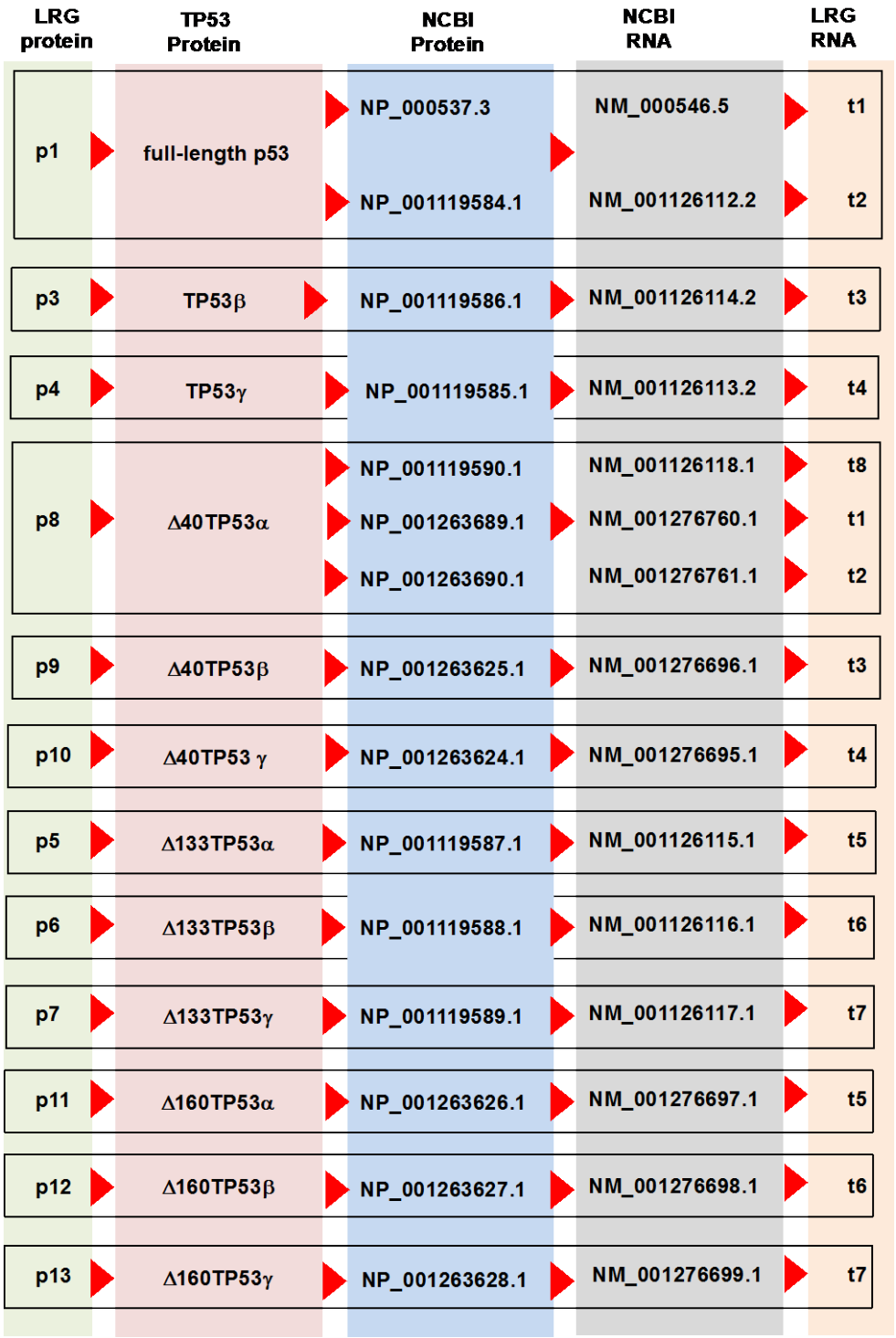
a Sample identifier used in the original publication or in dbSNP.
b Origin: somatic data were extracted from tumors included in the latest version of the TP53 mutation database (http://p53.fr) or from dbSNP (http://www.ncbi.nlm.nih.gov/snp). For most variants described in dbSNP, the minor allele frequency is less than 0.001.
c *TP53* genomic variant descriptions according to chromosomal reference sequence NC_000017.10 of the GRCh37 (hg19) genome assembly.
d RefSeq transcript NM_000546.5 encodes the TP53 full-length protein (NP_000537.3).
e *TP53* coding DNA variant descriptions according to RefSeq transcripts encoding TP53β (NM_001126114.2).
f *TP53* coding DNA variant descriptions according to RefSeq transcripts encoding TP53γ (NM_001126113.2).
g TP53 protein variant descriptions according to RefSeq proteins TP53β (NP_001119586.1).
h TP53 protein variant descriptions according to RefSeq proteins TP53γ (NP_001119585.1).
i Both T-ALL cell lines carry the same rare variant, suggesting cross-contamination.
j This variant was described as germline in a patient with a pediatric neuroblastoma
k The variant in this patient was shown to modify *TP53* splicing.

Supplementary material

Figure S1



TP53 gene and p53 mRNAs. The TP53 gene (upper part of the figure) is transcribed into eight different mRNAs. Transcripts t1 to t4 originate from promoter P1 localized upstream from the gene. Transcripts t5 to t8 originate from promoter P2 localized in intron 4. Translated exons are shown in blue. The two novel exons β and γ are shown in red and blue, respectively. Untranslated regions are shown in black. For transcripts t3, t4, t6, and t7, which include exons β or γ, exons 10 and 11 are noncoding. Transcript t8 encodes only p8 (DeltaTP53α) and exons 1 to 3 are noncoding (gray boxes). Proteins translated from the various transcripts are described on the right. Frequent TP53 germline SNP are shown on the TP53 gene with both coding (red) and non-coding (blue) snp. This figure describes the eight major TP53 transcripts as reported in LRG_321. Other transcripts have been identified, but it is currently unknown whether they are expressed in normal cells or in all cell types. (http://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000141510;r=17:7661779-7687550)


LRG protein
TP53 Protein
NCBI Protein
NCBI RNA
LRG RNA
p1
full-length p53
NP_000537.3
NM_000546.5
t1
NP_001119584.1
NM_001126112.2
t2
p3
TP53β
NP_001119586.1
NM_001126114.2
t3
p4
TP53γ
NP_001119585.1
NM_001126113.2
t4
p8
Δ40TP53α
NP_001119590.1
NM_001126118.1
t8
NP_001263689.1
NM_001276760.1
t1
NP_001263690.1
NM_001276761.1
t2
p9
Δ40TP53β
NP_001263625.1
NM_001276696.1
t3
p10
Δ40TP53 γ
NP_001263624.1
NM_001276695.1
t4
p5
Δ133TP53α
NP_001119587.1
NM_001126115.1
t5
p6
Δ133TP53β
NP_001119588.1
NM_001126116.1
t6
p7
Δ133TP53γ
NP_001119589.1
NM_001126117.1
t7
p11
Δ160TP53α
NP_001263626.1
NM_001276697.1
t5
p12
Δ160TP53β
NP_001263627.1
NM_001276698.1
t6
p13
Δ160TP53γ
NP_001263628.1
NM_001276699.1
t7

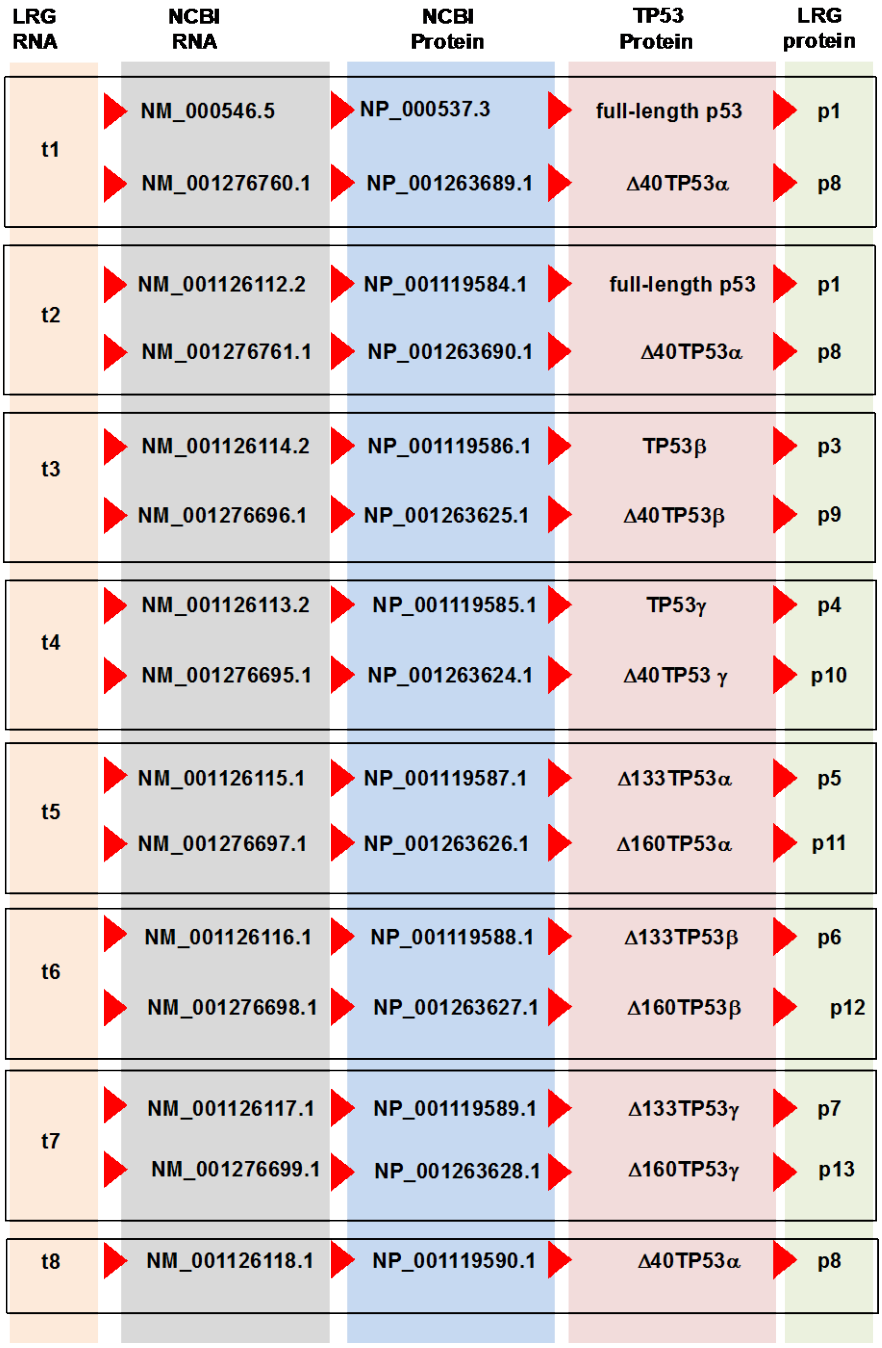


Figure S2A and B

Figure S2A. Correspondence between the LRG, the NCBI RefSeq TP53 protein and their corresponding transcripts. All TP53 transcripts (except one) have alternative translation start sites and can encode two TP53 isoforms. Each TP53 isoform has a single LRG_321 p (protein) number, but might be encoded by more than one transcript with its own t (transcript) number. In contrast, full-length TP53 protein (p1) and the Delta40p53α isoform (p8) have multiple RefSeq protein accession numbers. As RefSeq transcripts can be linked to only one RefSeq protein accession number in the NCBI's data model, a different RefSeq transcript accession number is therefore assigned to the same mRNA sequence for each TP53 isoform it encodes.
Figure S2B. Correspondence between the LRG, the NCBI RefSeq TP53 transcripts and their corresponding putative proteins.