Table 1. Associations between *TP53* mutation within exons 5 to 8 and clinicopathologic variables

n = 1,794	*n*	*TP53* wild-type*	*TP53* mutation†	Mutated (%)	χ^2 test (*P*)
Age group (quartiles)					
<49	464	373	91	19.6	0.037
49-59	458	368	90	19.7	
60-70	456	386	70	15.4	
>70	416	359	57	13.7	
Tumor subtype					
ductal	1,162	942	220	18.9	<0.001
lobular	175	165	10	5.7	
medullar	18	11	7	38.9	
tubular	34	34	0	0	
mucinous	27	27	0	0	
other	71	57	14	19.7	
missing	307				
Tumor grade					
1	118	117	1	0.8	<0.001
2	458	397	61	13.3	
3	345	267	78	22.6	
missing	873				
Tumor size					
T1	547	488	59	10.8	<0.001
T2	918	737	181	19.7	
T3	252	201	51	20.2	
T4	20	13	7	35.0	
missing	57				
Nodal status					
N0	955	811	144	15.1	0.016
N ≥ 1	690	504	136	19.7	
missing	149				
ER status					
ER+	1,152	1,022	130	11.3	<0.001
ER−	503	349	154	30.6	
missing	139				
PR status					
PR+	1,107	963	144	13.0	<0.001
PR−	547	408	139	25.4	
missing	140				

*No *TP53* mutation within exons 5 to 8.
†Any *TP53* mutation within exons 5 to 8.

receptor status were all associated with patient survival (Table 2). The effect of tumor histopathologic subtype was mainly due to tubular and medullar types that were associated with lower and higher rates of death, respectively. Large tumor size, high histopathologic grade, presence of nodes, and absence of hormone receptors were associated with high mortality rates. Patients with a *TP53* mutation had a relative risk of breast cancer–specific death of ~2 over a period of 10 years following surgery compared with patients with no mutation (Table 2).

In multivariate analysis including *TP53* mutation, tumor size, node status and ER/PR status, an interaction between *TP53* mutation and PR status was revealed (Table 3). Whereas the lack of PR expression increased the relative risk of breast cancer death by 2-fold in patients without *TP53* mutation, PR status did not affect the survival of patients with a *TP53* mutation in exons 5 to 8 (Table 3; Fig. 1A). These results were confirmed after multiple adjustments for tumor size and node status. By contrast, ER status, even when forced in the final multivariate model, was not significantly associated with outcome, either as the main effect or as a factor interacting with *TP53* mutation status.

A sensitivity analysis was carried out (Table 3, sensitivity analysis) on the whole data set by including patients with missing information on nodal status or tumor size (labeled as "missing" in Table 3). This analysis confirmed the result above, as did an analysis made using a more extended definition of the outcome (overall survival – data not shown), or using 5 years as censoring time (Table 3).

Because data on tumor grade was lacking for too many cases to be included in the multivariate analysis, a separate analysis

on a subset of patients with information on grading was carried out. Patients predicted to have a favorable outcome were selected (tumor grade <3, tumor size <5 cm, no node invasion, positive ER or PR receptor). Ten-year survival analysis of these patients stratified by *TP53* mutation status showed that patients with a *TP53* mutation had a large and significant reduction in survival compared with patients without mutation (close to 60% at 10 years; Fig. 1B).

Table 2. Univariate analysis of breast-specific mortality rates >10 years, according to clinical and molecular characteristics of patients with breast cancer and samples (n = 1,794)

n = 1,794	n (%)	Death (n)	Ten-year mortality rate (/1,000)	Relative risk	Log rank test (P)
Age group (quartiles)					
<49	464 (25.86)	121	40.15	1.00	0.1178
49-59	458 (25.53)	102	33.15	0.83	
60-70	456 (25.42)	98	31.99	0.80	
>70	416 (23.19)	96	42.75	1.06	
Tumor subtype					
ductal	1,162 (64.77)	271	37.26	1.00	0.0302
lobular	175 (9.75)	40	37.67	1.01	
medullar	18 (1.00)	6	49.11	1.32	
tubular	34 (1.90)	0	0.00	0.00	
mucinous	27 (1.51)	4	25.65	0.69	
other	71 (3.96)	14	27.89	0.75	
missing	307 (17.11)				
Tumor grade					
1	118 (6.58)	11	12.87	1.00	<0.0001
2	458 (25.53)	83	30.90	2.40	
3	345 (19.23)	98	52.52	4.08	
missing	873 (48.66)				
Tumor size					
T1	547 (30.49)	70	17.39	1.00	<0.0001
T2	918 (51.17)	211	37.63	2.16	
T3	252 (14.05)	115	85.85	4.94	
T4	20 (1.11)	14	191.78	11.03	
missing	57 (3.18)				
Nodal status					
N0	955 (53.23)	134	19.99	1.00	
N ≥ 1	690 (38.46)	243	62.12	3.11	<0.0001
missing	149 (8.31)				
ER status					
ER+	1,152 (64.21)	234	31.11	1.00	
ER−	503 (28.04)	148	49.35	1.59	<0.0001
missing	139 (7.75)				
PR status					
PR+	1,107 (61.71)	207	28.47	1.00	
PR−	547 (30.49)	174	53.95	1.90	<0.0001
missing	140 (7.80)				
TP53 mutation*					
none	1,460 (81.38)	292	30.53	1.00	<0.0001
outside exons 5-8	26 (1.45)	8	55.56	1.82	
exons 5-8	308 (17.17)	117	69.18	2.27	
TP53 mutation in exons 5-8 and PR status					
wild-type† PR+	963 (53.7)	155	23.87	1.00	<0.0001
wild-type† PR−	408 (22.7)	119	48.51	2.03	
mutant PR+	144 (8.0)	52	66.87	2.80	
mutant PR−	139 (7.7)	55	71.23	2.98	
missing	140 (7.8)				

NOTE: Relative risks and log rank test P values are shown for each clinical and molecular category.
*TP53 gene has been analyzed between exons 5 to 8 in 1143 tumors whereas the whole coding sequence has been analyzed in 651 cases.
†The 26 TP53 mutations found outside exons 5 to 8 are included in the wild-type group (see text).

Table 3. Multivariate Cox proportional hazards models and sensitivity analysis of 10-year mortality rates

	Complete case analysis	Sensitivity analysis	
	Cox model, 10 years (n = 1,470)	Cox model, 10 years including missing (n = 1,750)*	Cox model, 5 years (n = 1,470)
	HR† (95% CI)	HR† (95% CI)	HR† (95% CI)
Tumor size			
T1	1	1	1
T2	1.78 (1.32-2.42)	1.83 (1.38-2.43)	2.09 (1.40-3.11)
T3	3.24 (2.24-4.69)	3.37 (2.41-4.70)	3.48 (2.18-5.55)
T4	8.47 (4.15-17.29)	5.65 (3.02-10.54)	9.54 (4.37-20.84)
missing		0.68 (0.29-1.56)	
Nodal status			
N0	1	1	1
N ≥ 1	2.53 (2.00-3.21)	2.56 (2.04-3.20)	3.00 (2.34-4.02)
missing		2.52 (1.70-3.73)	
TP53 mutation in exons 5-8 and PR status			
wild-type‡ PR+	1	1	1
wild-type‡ PR−	1.83 (1.41-2.39)	1.80 (1.40-2.30)	2.06 (1.50-2.83)
mutant PR+	2.40 (1.70-3.38)	2.48 (1.79-3.44)	3.00 (2.03-4.44)
mutant PR−	2.63 (1.89-3.66)	2.58 (1.87-3.55)	2.67 (1.78-3.99)
missing		1.14 (0.77-1.67)	

*The total does not add to 1,794 as the stratified Cox model drops some observations from the analysis.
†Hazard rate ratio.
‡The 26 TP53 mutations found outside exons 5 to 8 are included in the wild-type group (see text).

Prognostic values of specific TP53 mutations. Mutations within exons 5 to 8 were classified in different groups according to the effect or position of the mutation in the primary or tertiary sequence of the protein (see Materials and Methods). Kaplan-Meier survival analysis of patients grouped according to the type of *TP53* mutation found in their tumor showed that non-missense mutations (any mutation other than missense) and missense mutations in the DNA-bindings motifs (DBM) were associated with a strong reduction in survival compared with patients without mutations, whereas missense mutations outside the DNA-bindings motifs (non-DBM) were associated with an intermediate reduction of survival (Fig. 2A). The 10-year mortality rates for non-DBM and DBM mutations were, respectively, 43.92 and 73.42 (per 1,000 persons, P = 0.0897; Table 4). If the non-DBL missense mutations were grouped with silent mutations (associated with similar mortality rate) and used as a reference group, the relative risk associated with DBM mutations was 1.8 (1.03-3.18) and the one of non-missense mutations was 2.9 (1.44-5.38). These results remained valid after adjustment for tumor size, node status, and hormone receptor status. Analysis of mortality rates for the most frequent (hotspots) missense mutations in this series identified mutations with higher or lower mortality rates compared with other missense mutations (Table 4). Figure 2B shows that missense mutations at codon 179 and the R248W mutation were associated with reduced survival, whereas the G245S and Y220C mutations were associated with better survival compared with any other missense mutation. Of note, other mutation hotspots, which are general hotspots for all breast cancers (R175H, R248Q, R273H/C, codons 163, 249, and 282), were associated with mortality rates similar to those of non-hotspot missense mutations (these mutations were included in the "other" category in Fig. 2).

Several groupings of missense mutations were done based on functional, structural, or conservation properties of mutant proteins (Table 4). First, mutations were classified as conserved or nonconserved if the affected residue is conserved or not in vertebrate species. Mutations were also classified according to their capacity to transactivate p53-responsive promoters in yeast-based assays (9). Groups were made according to mutant protein activity on a single promoter or to its global activity on seven different promoters. Finally, mutations were classified according to the predicted structural effects of amino acid substitution (20). Two categories were made, one for mutations predicted to affect protein folding or protein-DNA interaction, and another for all other mutations. None of these classifications detected significant differences in patient survival (Table 4).

Discussion

In this study, which is the largest to date, on the prognostic value of *TP53* gene mutation in breast cancer, we show that *TP53* gene mutation is an independent factor of prognosis for breast cancer survival after adjustment for tumor size, node status, and hormone receptor status (ER and PR). The relative risk of dying of breast cancer within 10 years following surgery for patients with a *TP53* mutation within exons 5 to 8 in their tumor was between 2 and 3 compared with patients without any mutation. In a subset of patients with available tumor grade information and favorable outcome (low grade, limited

size, no nodes, and the presence of hormone receptors), the presence of a *TP53* mutation was associated with a reduction in survival close to 60% at 10 years. This result shows the value of *TP53* mutation as an additional prognostic indicator in this group of patients. It is of note that our analysis was focused on mutations occurring within exons 5 to 8. These exons contain >90% of the mutations reported in breast cancer (19). Thus, most published studies have analyzed only these exons, as it is the case for 1,143 of the 1,794 tumors included in the present study. Analysis of the whole coding sequence (and splice junctions) in 651 cases revealed the presence of 26 mutations outside exons 5 to 8 (4%). These 26 mutations were associated with a relative risk of death close to 2 compared with wild-type cases. Thus, among the 835 mutation-negative breast cancers analyzed for exons 5 to 8 only, about 40 cases may contain a *TP53* mutation. Thus, the actual risk associated with *TP53* mutation may be greater than that estimated here. In clinical practice, it is therefore recommended to conduct mutation analysis on all coding exons and splice junctions.

There was a linear relationship between the size of the tumor and the frequency of *TP53* mutations, and a strong association

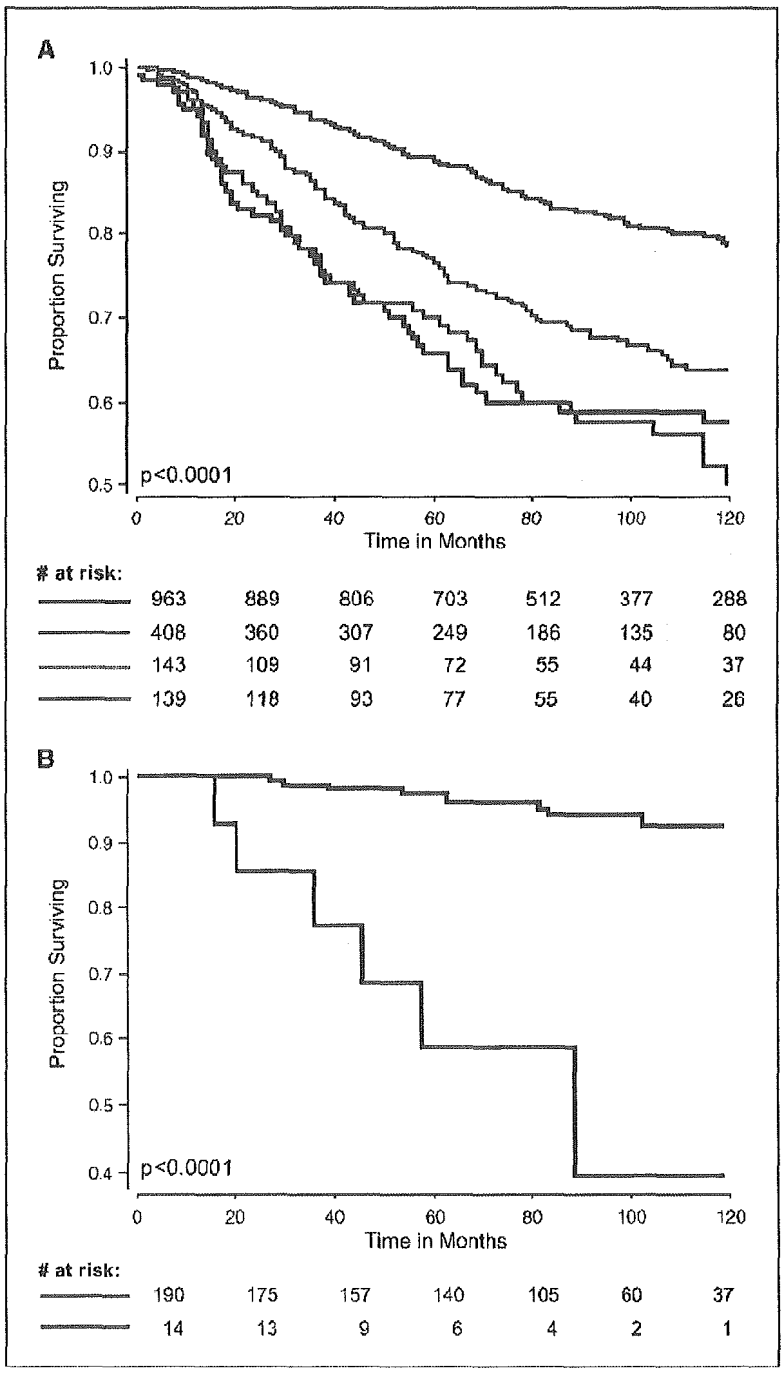


Fig. 1. Kaplan-Meier survival curves of patients with breast cancer stratified by *TP53* gene mutation status. *A*, entire cohort of 1,794 patients. Survival of patients without a mutation within exons 5 to 8 and with a positive PR status (*blue line*), without a mutation within exons 5 to 8 and with a negative PR status (*red line*), with a mutation within exons 5 to 8 and with a positive PR status (*green line*), with a mutation within exons 5 to 8 and with a negative PR status (*black line*). *B*, subset of 204 patients with favorable outcome (tumor grade <3, tumor size <5 cm, node negative and ER or PR positive cases). Survival of patients without a mutation within exons 5 to 8 (*blue line*), compared with patients with a mutation within exons 5 to 8 (*red line*).

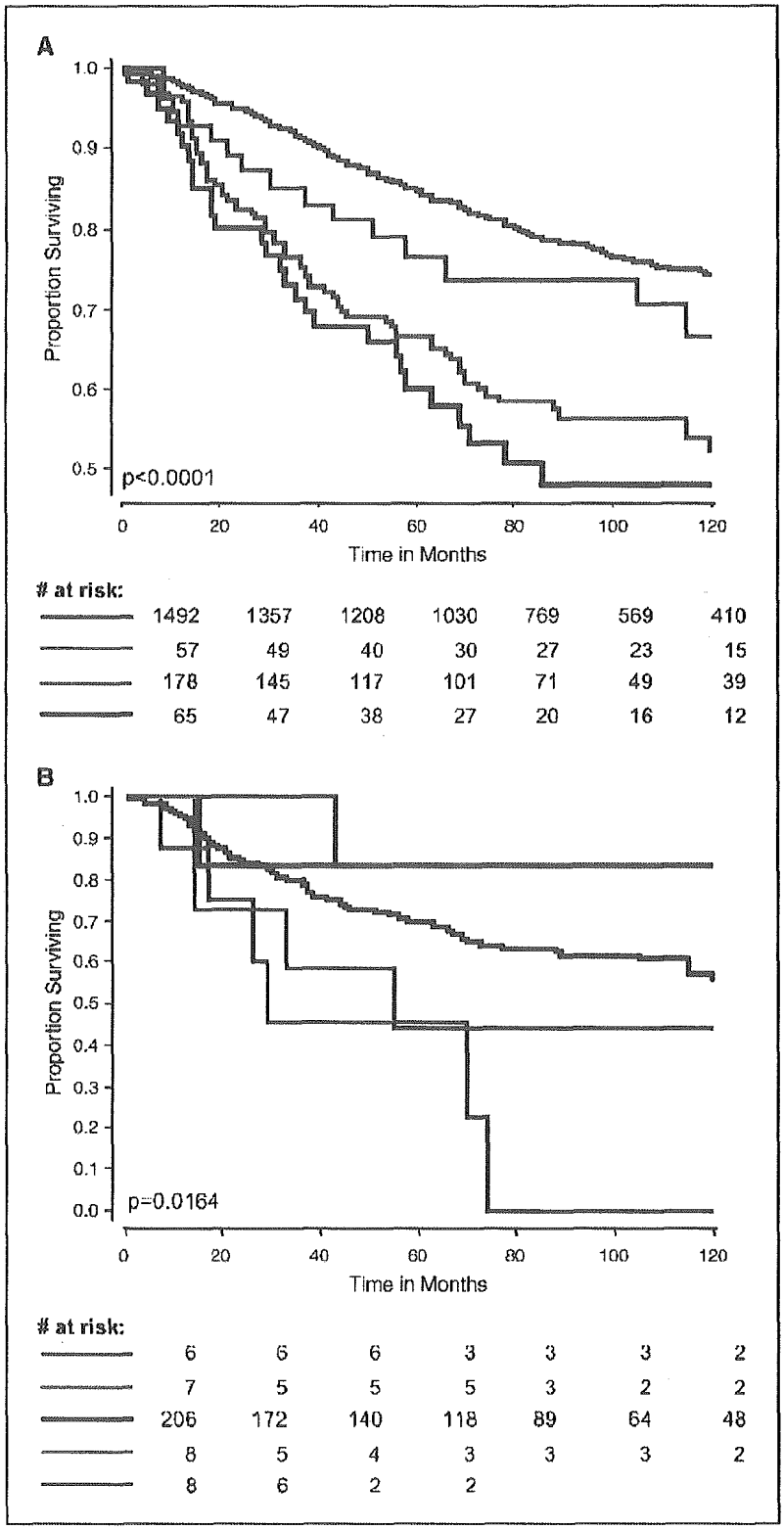


Fig. 2. Kaplan-Meier survival curves of patients with breast cancer stratified by the type of *TP53* gene mutation found in their tumor. *A*, survival of patients without mutation or with a silent mutation within exons 5 to 8 (*blue line*), with a missense mutation within exons 5 to 8 but outside the DBMs (*red line*), with a missense mutation in the DBMs (*green line*), with a mutation other than missense within exons 5 to 8 (*black line*). *B*, survival of patients with the Y220C mutation (*blue line*), the G245S mutation (*green line*), the R248W mutation (*red line*), with any missense mutation at codon 179 (*purple line*), or any other missense mutations (*black line*).

between the presence of a *TP53* mutation and high grade, positive node status, and loss of hormone receptors (ER and PR), which is in agreement with previous reports (http://www-p53.iarc.fr/Somatic.html). *TP53* mutations have also been found to be associated with increased global genomic instability (21–23) and with markers of increased cell proliferation such as high mitotic frequency, high expression of Ki-67, and high cyclin E expression (24, 25). These results show that *TP53* mutations are generally associated with an advanced and aggressive tumor phenotype.

TP53 mutations were significantly more frequent in young women and in medullar carcinoma as reported by others (25–27). Both early age at onset and medullar subtype may be indicative of inherited cancer due to BRCA1 germ line mutations (28, 29). BRCA1 breast cancers may represent up to 5% of the cases in a breast cancer series and these cancers usually present a high frequency of *TP53* somatic mutations (19). Because family history is not well documented in our series and <5% of the patients have been tested for BRCA1, we cannot exclude that some BRCA1 cases may contribute to the observed mutation frequency. To rule out any confounder effect of possible BRCA1 cases in subsequent analyses, results were systematically verified on subsets excluding patients <40 years and medullar cases.

The prognostic value of *TP53* mutation has been shown to be independent of tumor size, node status, or ER in a number of reports (13, 15, 25, 30, 31). The present study confirms these observations. In addition, we found an interaction between *TP53* mutation and PR content that has not been previously reported. *TP53* mutation combined with low PR had a very bad prognosis independently of tumor size, node status, and ER status. PR status, which reflects estrogen pathway integrity, has been shown to be more relevant than ER status for tumor response to tamoxifen and prediction of patient survival (32–34). In two independent retrospective series of patients (14, 35), one of them being included in our series (35), *TP53* mutation has been shown to affect tumor response to tamoxifen. These results suggest that *TP53* pathway may play a role in the response to antihormone therapy. However, the lack of information on treatment for a significant number of patients prevented us from exploring this hypothesis. Further studies are thus required to explore the possible interplay between *TP53* and ER pathways and the consequences on tumor development and behavior.

When comparing the prognostic value of different types of mutations, we found that the more severe mutations were non-missense mutations followed by missense mutations in the DBMs (L2/L3 and LSH), among which mutations at codon

Table 4. Univariate analysis of ten-years mortality rates and log rank test *P* values of breast cancer cases with specific *TP53* mutations within exons 5 to 8 ($n = 306$)

	n	Deaths (*n*)	Ten-year mortality rate (/1,000)	Log rank test (*P*)
Silent	6	2	43.17	0.3605
Non-missense	65	28	88.05	
Missense*	235	87	65.80	
Conserved	199	78	71.64	0.0944
Nonconserved	36	9	38.57	
Structurally explained†	151	57	70.44	0.3421
Not structurally explained	79	27	54.36	
Function 1	72	31	78.17	0.5165
Function 2	44	12	45.20	
Function 3	7	3	80.18	
Function 4	112	41	65.85	
DBM	178	72	73.42	0.0897
Non-DBM	57	15	43.92	
Zn	17	7	86.15	0.3325
DNA	61	27	83.38	
Non-Zn/DNA	157	53	57.79	
163	6	2	68.97	0.441
175	19	7	68.29	
179	8	6	254.42	
220	7	2	45.63	
245	9	2	37.68	
248	26	12	78.65	
249	7	3	71.29	
273	20	7	68.46	
282	7	3	71.01	
R175H	18	7	69.65	0.805
R248Q	18	8	69.06	
R248W	8	4	108.84	
R273H	9	3	67.29	
R273C	8	3	64.86	
G245S	7	1	24.49	

NOTE: Two patients were excluded because their *TP53* mutations were not precisely identified.
*Missense mutations were grouped as described in details in Materials and Methods. Survival analyses for the different subcategories are shown in the subsequent rows.
†Structural information was not available for five missense mutations.

179 and the R248W mutant were associated with the highest mortality rates. Grouping of missense mutations according to their loss of transcriptional activities measured by systematic yeast-based assays, or to their predicted effect on protein structure, did not correlate with patient outcome. Although these structural and functional analyses of p53 mutations are the most extensive that are currently available, they may not give an accurate assessment of the changes induced by mutation that really have an effect on clinical outcome. Indeed, other variables such as protein-protein interactions, transcriptional repression, and transactivation of other genes, not taken into account here, play an important role in the antiproliferative activity of p53 and in the activities of mutant proteins (36, 37). Mutations affecting the L2/L3 motif involved in specific DNA-binding and zinc coordination have been repeatedly described as "bad" mutations in breast cancer based on their association with poor tumor response to treatment and short patient survival (4, 26, 30, 38–41). Functional assessments of some of these mutations in human cells have shown not only loss of transcriptional activity and defects in the capacity to induce cell cycle arrest or apoptosis, but also gain of function properties and/or dominant-negative effects, resulting in growth-promoting activities and resistance to drug-induced apoptosis (http://www-p53.iarc.fr/p53MUTfunction.html). How these properties specifically affect tumor response to treatment and patient outcome is still under debate. Because deletions/insertions mutations are expected to result in a null phenotype (truncated and unfolded proteins), our results suggest that loss of transcriptional activity is the main determinant of the poor prognostic value of *TP53* mutations in breast cancer. Moreover, loss of transcriptional activity may be sufficient to promote breast tumor development, as suggested by studies on germ line mutations. First, non–missense mutations and missense mutations in the DBMs were associated with an earlier age at onset of breast tumors compared with missense mutations outside these motifs (32 versus 42 years; ref. 42). Second, *TP53* null mammary epithelium isolated from *TP53* null mice and transplanted into cleared mammary fat pads of *TP53* wild-type mice showed that the absence of *TP53* is sufficient to cause the development of primary tumors (43). If loss of transcriptional activity seems to be the main determinant of breast tumor development and behavior, it cannot be excluded that some specific mutants, such as codon 179 and R248W might exert dominant-negative effects and gain of function activities responsible for their very bad prognostic value.

Some limitations, mainly due to the multicenter structure of the study, have to be acknowledged. First, intercenter heterogeneity could be mostly controlled through stratification (44), as mentioned in Materials and Methods. In addition, differences of follow-up between hospitals were homogenized by censoring of follow-up. Second, *TP53* mutation detection was done in five different laboratories with four different prescreening methods that may differ in their sensitivities. Thus, hazard risk estimates for *TP53* mutation may have been underestimated. Third, the presence of missing values in known predictors of survival prevented the inclusion of some variables in multivariate analysis. It was particularly true for histopathologic grade and subtype with, respectively, 48% and 17% of missing values. However, histopathologic subtype and grade may reflect the presence of molecular alterations including *TP53* mutation and hormone receptor expression and thus may be collinear with *TP53* when entered in the multivariate model. The sensitivity analyses confirmed results when missing values for variables such as tumor size and nodal status were integrated in the analysis, and when censoring was made at the 5-year follow-up, or when using a more extended definition survival (overall mortality). These sensitivity analyses show the stability of the results. Nonetheless, the lack of information on family history, adjuvant treatments, and other markers such as *ERBB2* amplification, prevented us from investigating the influence of these variables on our results. In a recent study, Bull et al. have found that *TP53* gene mutations were more frequent than *ERBB2* amplification in women with node-negative breast cancer and that *TP53* mutation may be beneficial in identifying women at higher risk of disease recurrence and death when their tumor has *ERBB2* amplification (45). In two other studies, the use of expression microarrays has shown that *TP53* gene mutation is highly associated with groups of patients with similar gene expression profiles (46), and that tumor classification based on these profiles was a stronger predictor of outcome than any of the classical clinicopathologic markers, *TP53* mutation status being equally significant.[14] These results strongly support the fact that *TP53* mutations have a prognostic value in various groups of patients. However, further studies will be required to precisely identify which groups of patients would benefit or not from *TP53* mutation screening.

In the early 1990s, the rapid accumulation of data on *TP53* mutations in human cancer raised high expectations for clinical exploitation. However, most studies relied on immunohistochemistry to assess *TP53* status, a method prone to misclassification, as many *TP53* mutations do not correlate with protein accumulation. In the present study, the strongest association with poor survival was found for non–missense mutations, predicted to generate a negative immunostaining. It is thus highly recommended to perform gene sequencing to precisely identify the mutation. Several common polymorphisms in *TP53* (in exons 3, 4, and 6) may also deserve further investigation. Although a recent study argues against a role for these polymorphisms in breast cancer susceptibility (47), there is experimental evidence that codon 72 polymorphism (arginine to proline) may influence wild-type p53 activity in response to cytotoxic drugs (48). Mutation detection in our series was done by PCR-based prescreening methods followed by DNA sequencing, which are the gold standards for gene mutation identification, but they are labor-intensive, and thus, are not suitable for clinical practice. New techniques have been developed recently that may be more easily implemented for routine use, such as microarray-based methods (Affymetrix, arrayed primer extension assay; ref. 49). As these techniques will soon be available at an effective cost, *TP53* gene mutation may become an important marker for patient management.

Acknowledgments

We thank G. Martel-Planche for technical assistance.

[14] A. Langerød and A-L. Børresen-Dale, unpublished data.

References

1. Hainaut P, Hollstein M. p53 and human cancer: the first ten thousand mutations. Adv Cancer Res 2000; 77:81–137.
2. O'Connor PM, Jackman J, Bae I, et al. Characterization of the p53 tumor suppressor pathway in cell lines of the National Cancer Institute anticancer drug screen and correlations with the growth-inhibitory potency of 123 anticancer agents. Cancer Res 1997;57:4285–300.
3. Lowe SW, Bodis S, McClatchey A, et al. p53 status and the efficacy of cancer therapy *in vivo*. Science 1994;266:807–10.
4. Borresen AL, Andersen TI, Eyfjord JE, et al. TP53 mutations and breast cancer prognosis: particularly poor survival rates for cases with mutations in the zinc-binding domains. Genes Chromosomes Cancer 1995;14:71–5.
5. Kihara C, Seki T, Furukawa Y, et al. Mutations in zinc-binding domains of p53 as a prognostic marker of esophageal-cancer patients. Jpn J Cancer Res 2000; 91:190–8.
6. Temam S, Flahault A, Perie S, et al. p53 gene status as a predictor of tumor response to induction chemotherapy of patients with locoregionally advanced squamous cell carcinomas of the head and neck. J Clin Oncol 2000;18:385–94.
7. Skaug V, Ryberg D, Kure EH, et al. p53 mutations in defined structural and functional domains are related to poor clinical outcome in non-small cell lung cancer patients. Clin Cancer Res 2000;6:1031–7.
8. Resnick MA, Inga A. Functional mutants of the sequence-specific transcription factor p53 and implications for master genes of diversity. Proc Natl Acad Sci U S A 2003;100:9934–9.
9. Kato S, Han SY, Liu W, et al. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. Proc Natl Acad Sci U S A 2003;100:8424–9.
10. Pharoah PD, Day NE, Caldas C. Somatic mutations in the p53 gene and prognosis in breast cancer: a meta-analysis. Br J Cancer 1999;80:1968–73.
11. Andersen TI, Holm R, Nesland JM, Heimdal KR, Ottestad L, Borresen AL. Prognostic significance of TP53 alterations in breast carcinoma. Br J Cancer 1993;68:540–8.
12. Bukholm IK, Nesland JM, Karesen R, Jacobsen U, Borresen AL. Relationship between abnormal p53 protein and failure to express p21 protein in human breast carcinomas. J Pathol 1997;181:140–5.
13. Cuny M, Kramar A, Courjal F, et al. Relating genotype and phenotype in breast cancer: an analysis of the prognostic significance of amplification at eight different genes or loci and of p53 mutations. Cancer Res 2000;60:1077–83.
14. Bergh J, Norberg T, Sjogren S, Lindgren A, Holmberg L. Complete sequencing of the p53 gene provides prognostic information in breast cancer patients, particularly in relation to adjuvant systemic therapy and radiotherapy. Nat Med 1995;1:1029–34.
15. Thorlacius S, Borresen AL, Eyfjord JE. Somatic p53 mutations in human breast carcinomas in an Icelandic population: a prognostic factor. Cancer Res 1993;53: 1637–41.
16. Gretarsdottir S, Tryggvadottir L, Jonasson JG, et al. TP53 mutation analyses on breast carcinomas: a study of paraffin-embedded archival material. Br J Cancer 1996;74:555–61.
17. Sorlie T, Johnsen H, Vu P, Lind GE, Lothe R, Borresen-Dale AL. Mutation Screening of the TP53 gene by temporal temperature gradient gel electrophoresis. Methods Mol Biol 2004;291:207–16.
18. Cho Y, Gorina S, Jeffrey PD, Pavletich NP. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science 1994; 265:346–55.
19. Olivier M, Hainaut P. TP53 mutation patterns in breast cancers: searching for clues of environmental carcinogenesis. Semin Cancer Biol 2001;11:353–60.
20. Martin AC, Facchiano AM, Cuff AL, et al. Integrating mutation data and structural analysis of the TP53 tumor-suppressor protein. Hum Mutat 2002;19:149–64.
21. Jong YJ, Li LH, Tsou MH, et al. Chromosomal comparative genomic hybridization abnormalities in early- and late-onset human breast cancers: correlation with disease progression and TP53 mutations. Cancer Genet Cytogenet 2004;148:55–65.
22. Sigurdsson S, Bodvarsdottir SK, Anamthawat-Jonsson K, et al. p53 abnormality and chromosomal instability in the same breast tumor cells. Cancer Genet Cytogenet 2000;121:150–5.
23. Eyfjord JE, Thorlacius S, Steinarsdottir M, Valgardsdottir R, Ogmundsdottir HM, Anamthawat-Jonsson K. p53 abnormalities and genomic instability in primary human breast carcinomas. Cancer Res 1995;55:646–51.
24. Lindahl T, Landberg G, Ahlgren J, et al. Overexpression of cyclin E protein is associated with specific mutation types in the p53 gene and poor survival in human breast cancer. Carcinogenesis 2004;25:375–80.
25. Chappuis PO, Estreicher A, Dieterich B, et al. Prognostic significance of p53 mutation in breast cancer: frequent detection of non-missense mutations by yeast functional assay. Int J Cancer 1999; 84:587–93.
26. Kucera E, Speiser P, Gnant M, et al. Prognostic significance of mutations in the p53 gene, particularly in the zinc-binding domains, in lymph node- and steroid receptor positive breast cancer patients. Austrian Breast Cancer Study Group. Eur J Cancer 1999;35: 398–405.
27. De Cremoux P, Salomon AV, Liva S, et al. p53 mutation as a genetic trait of typical medullary breast carcinoma. J Natl Cancer Inst 1999;91:641–3.
28. Eisinger F, Jacquemier J, Charpin C, et al. Mutations at BRCA1: the medullary breast carcinoma revisited. Cancer Res 1998;58:1588–92.
29. Rebbeck TR. Inherited genetic predisposition in breast cancer. A population-based perspective. Cancer 1999;86:2493–501.
30. Powell B, Soong R, Iacopetta B, Seshadri R, Smith DR. Prognostic significance of mutations to different structural and functional regions of the p53 gene in breast cancer. Clin Cancer Res 2000;6:443–51.
31. Overgaard J, Yilmaz M, Guldberg P, Hansen LL, Alsner J. TP53 mutation is an independent prognostic marker for poor outcome in both node-negative and node-positive breast cancer. Acta Oncol 2000;39: 327–33.
32. Pichon MF, Pallud C, Brunet M, Milgrom E. Relationship of presence of progesterone receptors to prognosis in early breast cancer. Cancer Res 1980;40:3357–60.
33. Bernoux A, De Cremoux P, Laine-Bidron C, Martin EC, Asselain B, Magdelenat H. Estrogen receptor negative and progesterone receptor positive primary breast cancer: pathological characteristics and clinical outcome. Institut Curie Breast Cancer Study Group. Breast Cancer Res Treat 1998;49:219–25.
34. Bardou VJ, Arpino G, Elledge RM, Osborne CK, Clark GM. Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. J Clin Oncol 2003;21: 1973–9.
35. Berns EM, Foekens JA, Vossen R, et al. Complete sequencing of TP53 predicts poor response to systemic therapy of advanced breast cancer. Cancer Res 2000;60:2155–62.
36. Erster S, Moll UM. Stress-induced p53 runs a transcription-independent death program. Biochem Biophys Res Commun 2005;331:843–50.
37. Soussi T, Lozano G. p53 mutation heterogeneity in cancer. Biochem Biophys Res Commun 2005;331: 834–42.
38. Aas T, Borresen AL, Geisler S, et al. Specific P53 mutations are associated with *de novo* resistance to doxorubicin in breast cancer patients. Nat Med 1996; 2:811–4.
39. Geisler S, Lonning PE, Aas T, et al. Influence of TP53 gene alterations and c-erbB-2 expression on the response to treatment with doxorubicin in locally advanced breast cancer. Cancer Res 2001;61: 2505–12.
40. Geisler S, Borresen-Dale AL, Johnsen H, et al. TP53 gene mutations predict the response to neoadjuvant treatment with 5-fluorouracil and mitomycin in locally advanced breast cancer. Clin Cancer Res 2003;9: 5582–8.
41. Takahashi M, Tonoki H, Tada M, et al. Distinct prognostic values of p53 mutations and loss of estrogen receptor and their cumulative effect in primary breast cancers. Int J Cancer 2000;89:92–9.
42. Olivier M, Goldgar DE, Sodha N, et al. Li-Fraumeni and related syndromes: correlation between tumor type, family structure, and TP53 genotype. Cancer Res 2003;63:6643–50.
43. Jerry DJ, Kittrell FS, Kuperwasser C, et al. A mammary-specific model demonstrates the role of the p53 tumor suppressor gene in tumor development. Oncogene 2000;19:1052–8.
44. Glidden DV, Vittinghoff E. Modelling clustered survival data from multicentre clinical trials. Stat Med 2004;23:369–88.
45. Bull SB, Ozcelik H, Pinnaduwage D, et al. The combination of p53 mutation and neu/erbB-2 amplification is associated with poor survival in node-negative breast cancer. J Clin Oncol 2004;22:86–96.
46. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74.
47. Suspitsin EN, Buslov KG, Grigoriev MY, et al. Evidence against involvement of p53 polymorphism in breast cancer predisposition. Int J Cancer 2003;103: 431–3.
48. Sullivan A, Syed N, Gasco M, et al. Polymorphism in wild-type p53 modulates response to chemotherapy *in vitro* and *in vivo*. Oncogene 2004;23:3328–37.
49. Tonisson N, Zernant J, Kurg A, et al. Evaluating the arrayed primer extension resequencing assay of TP53 tumor suppressor gene. Proc Natl Acad Sci U S A 2002;99:5503–8.

106. Thun, M. J., Henley, S. & Patrono, C. Nonsteroidal anti-inflammatory drugs as anticancer agents: mechanistic, pharmacologic, and clincal issues. *J. Natl Cancer Inst.* **94**, 252–266 (2002).
107. Bernstein, L. *et al.* Tamoxifen therapy for breast cancer and endometrial cancer risk. *J. Natl Cancer Inst.* **91**, 1654–1662 (1999).
108. Bergman-Jungestrom, M., Gentile, M., Lundin, A. C. & Wingren, S. Association between *CYP17* gene polymorphism and risk of breast cancer in young women. *Int. J. Cancer* **84**, 350–353 (1999).
109. United States Department of Health and Human Services. in *Reducing Tobacco Use: A Report of the Surgeon General — Executive Summary*. 7 (United States Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Atlanta, Georgia, 2000).

Acknowledgements
We are most thankful for the critical contribution of peer reviewers.

Competing interests statement
The authors declare no competing financial interests.

DATABASES
The following terms in this article are linked online to:
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
ATM | *BRCA1* | *BRCA2* | PSA | *TP53*
National Cancer Institute: http://www.cancer.gov
breast cancer | colon cancer | endometrial cancer | lung cancer | non-Hodgkin lymphoma | prostate cancer

FURTHER INFORMATION
American Cancer Society Cancer Prevention Study II: http://www.cancer.org/docroot/RES/content/RES_6_1_History_and_Accomplishments.asp
IARC Handbooks on Cancer Prevention: http://www.iarc.fr/IARCPress/general/prev.pdf
IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: http://monographs.iarc.fr/
IARC Press: http://www.iarc.fr/IARCPress/index.php
National Center for Health Statistics, National Health and Nutrition Survey: http://www.cdc.gov/nchs/nhanes.htm
NCI Early-Detection Research Network: http://www3.cancer.gov/prevention/cbrg/edrn/
US Surgeon General Reports on Smoking and Health: http://profiles.nlm.nih.gov/NN/ListByDate.html
Access to this interactive links box is free online.

SCIENCE AND SOCIETY

Locus-specific mutation databases: pitfalls and good practice based on the p53 experience

Thierry Soussi, Chikashi Ishioka, Mireille Claustres and Christophe Béroud

Abstract | Between 50,000 and 60,000 mutations have been described in various genes that are associated with a wide variety of diseases. Reporting, storing and analysing these data is an important challenge as such data provide invaluable information for both clinical medicine and basic science. Locus-specific databases have been developed to exploit this huge volume of data. The p53 mutation database is a paradigm, as it constitutes the largest collection of somatic mutations (22,000). However, there are several biases in this database that can lead to serious erroneous interpretations. We describe several rules for mutation database management that could benefit the entire scientific community.

Progress has been made over recent years in the cloning of the genes involved in both monogenic and polygenic disorders, including complex diseases such as cancer[1]. For each of these genes, numerous alterations of various types have also been described, ranging from point mutations to large deletions. The future development of new high throughput methods for the detection of mutations will lead to an enormous increase in the detection of new mutations[2]. It is difficult to evaluate the number of mutations reported in the literature to date (more than 50,000 have been collected in various databases), but a similar number could remain unpublished. It is also impossible to predict how many new mutations will be detected over the next 10 years, and the reporting and analysis of these mutations will therefore constitute a major challenge for the future[3,4]. Nevertheless, a number of points can be predicted. First, knowledge of these mutations will be important for treatment decisions as well as for basic science. And second, changes in our environment will lead to variations in the mutational events that modify our genome. Such changes will alter the distribution and/or pattern of mutations leading to the discovery of new and specific hot-spot mutations, so databases will need to be constantly updated. A good example of this is the specific mutation of *TP53* at codon 249 that is only found in hepatocellular carcinoma (HCC) that occurs in countries with a high exposure to the food contaminant aflatoxin B1. In these regions, the high specificity of the 249 mutation has enabled the development of a very sensitive diagnostic procedure that should not be applied when exposure to aflatoxin B1 is not (or no longer) evident[5].

The practical value of mutation analysis

All studies performed to date show that mutations are, in general, not randomly distributed. Hot-spot regions have been demonstrated, corresponding to a region of DNA that is susceptible to mutations (such as CpG dinucleotides), a codon encoding a key residue in the biological function of the protein, or both (BOX 1). Identification of these hot-spot regions and natural mutants is essential to define crucial regions in an unknown protein. In large genes such as neurofibromin 1 (*NF1*; 59 exons, 2,818 amino acids), retinoblastoma 1 (*RB1*; 27 exons, 928 amino acids), adenomatosis polyposis coli (*APC*; 15 exons, 2,843 amino acids), breast cancer 1 (*BRCA1*; 24 exons, 1,863 amino acids) and the titin gene (*TTN*; 363 exons, approximately 25,000 amino acids), detection of point mutations by direct sequencing analysis is difficult because of the size of the target gene. Identification of a hot-spot region allows analysis to be focused on this region, keeping in mind that a negative result should be viewed with caution.

It has also been clearly demonstrated that alterations in a single gene can cause various types of disorders. For example, mutations in *RET* are associated with multiple endocrine neoplasia types IIA[6] and IIB[7], familial medullary thyroid carcinoma[8] and a non-cancerous disorder known as Hirschsprung disease[9,10]. Mutations seem to be localized in specific domains of the protein for each of these disorders. The site of specific alterations at various positions in a given gene is also associated with different clinical features, as in the case of colon cancer and mutations in *APC*. A mutation in the C-terminus of the protein is specifically associated with a secondary abnormality, congenital hypertrophy of the retinal pigment epithelium[11], whereas mutations in the N-terminus are associated with an attenuated phenotype[12]. Analysis of mutations can also lead to the definition of risk factors. For instance, von Hippel–Lindau (VHL) families with mutations in *VHL* that result in truncated proteins have an increased frequency of renal-cell carcinoma (83%) compared with families with *VHL* missense mutations (54%)[13]. In diseases that are characterized

by considerable variations in the clinical phenotype between families and also within the same family, such as Marfan syndrome (a connective tissue disorder), it is very important to confirm or formally exclude the diagnosis in high-risk family members as early as possible because of the potentially fatal cardiovascular complications of the disease.

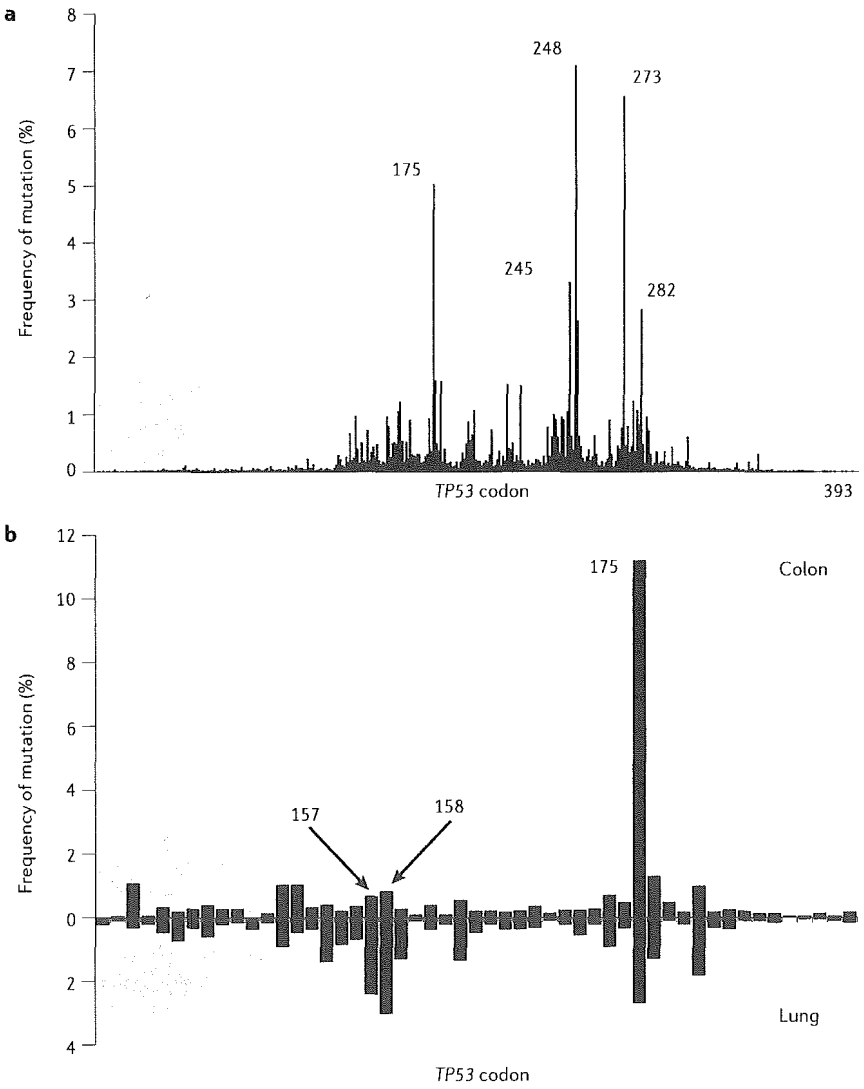
Therefore, it is important to consider mutation databases not only as tools that can provide essential information on protein structure and function, but also as an important framework for the development of new molecular-based diagnostic strategies and patient management.

Collecting mutations and LSDBs

Historically, collections of mutations and variations in human genes have been reported in the published literature. In the mid-1980s, several of these variations were available in the form of various databases, such as the Genome DataBase (GDB)[14], GenBank[15], the European Molecular Biology Laboratory (EMBL)[16] and Swiss-Prot[17]. However, because of the structure of these databases, the extraction of relevant information concerning mutations was almost impossible, so specific software had to be developed to facilitate this. Several teams started to develop specific databases to collect and document mutations in human genes. Today, several hundred locus-specific databases (LSDBs) are available through the Internet and have been recently reviewed[4]. Many of these databases are just a simple list of mutations that cannot be searched. They are also highly heterogeneous in terms of quality and content[4]. One of the main problems of these LSDBs concerns their follow-up. A recent survey of 138 known LSDBs of human gene mutations with available follow-up information found that 40 databases have not been updated since the year 2000 and that another 44 had only been updated between 2001 and 2003. Although the creation of a mutation database can be exciting and gratifying (in terms of publication), follow-up is time-consuming and less stimulating. As recently highlighted in a special report in *Nature*, financial issues are also involved and several databases, including the Asthma and Allergy gene database, were closed due to lack of funding[18]. These LSDBs are sponsored by only a few grants and they are usually developed 'on the side' (the Universal Mutation Database (UMD) for p53, created in 1991, only received one grant from a charity organization in 1995)[19]. Projects to generate central databases that regroup information from multiple LSDBs have been developed, but except for the Human Gene Mutation Database (HGMD), most of them have been abandoned for technical reasons such as incompatibility of the various LSDBs or lack of funding. The HGMD reports more than 40,000 mutations in 1,500 nuclear genes, but only germline

Box 1 | **Origin of p53 mutations**

Distribution of mutations in the p53 protein showing the various mutation hot spots (21,717 mutations in the entire database). Hot spots for somatic mutations can be explained both at the DNA level (the codon is highly susceptible to modification) and the protein level (the residue is essential for the function of the protein). In the case of p53, both explanations prevail. Codon 175 (similar to the other hot-spot codons 245, 248, 273 and 282, shown in part **a**) contains a CpG dinucleotide. In mammalian cells, the cytosine in this dinucleotide is often methylated and it has been shown that the 42 CpG sites of the *TP53* gene are methylated in normal tissue[50]. The higher deamination rate of 5-methylcytosine leading to a T–G mismatch that is not efficiently repaired leads to this high rate of transition in the *TP53* gene. Deamination of cytosine leads to a U–G mismatch that can be removed more efficiently. Various studies have also demonstrated that exogenous carcinogens have a higher affinity for methylated CpG dinucleotides than for their unmethylated counterparts[51,52]. Mutations at every CpG dinucleotide of the *TP53* gene have been reported in the p53 database, albeit at different frequencies[43]. Exogenous carcinogens can also target some specific residues of the *TP53* gene, such as benzo(a)pyrene [B(a)P] diolepoxide (BPDE), one of the carcinogens of tobacco smoke, which binds specifically to codons 157 and 158 *in vitro* (shown in part **b**). This observation explains the predominance of mutations found at these two codons in lung cancers from smokers compared with non-smokers or other cancers.



mutations are included (which excludes most cancers) and each mutation is reported only once[20].

In the early 1990s we decided to develop not just a simple repository of locus-specific mutations but a dynamic database that would include various software tools for the analysis of locus-specific mutations. This project ultimately led to the development of the UMD software[21], which is now recognized by the Human Genome Organization (HUGO) and the Human Genome Variation Society (HGVS) as a reference tool with which to build an LSDB. It was first used to create the UMD p53 database in 1991 and subsequently to develop databases for various genes involved in cancers, such as *APC* in colon cancer[22], *BRCA1* and *BRCA2* in breast cancer, *MEN1* in multiple endocrine neoplasia type 1 (REF. 23), the sulfonylurea receptor, *SUR1*, in hyperinsulinism[21], *RB1* in retinoblastoma, *VHL* in von Hippel–Lindau syndrome[24] and *WT1* in Wilms tumour[25]. UMD software is also used in other databases of genes that are involved in genetic disease, such as *FBN1* in Marfan syndrome[26], *LDLR* in hypercholesterolaemia[27], *VLCAD* in very-long-chain acyl-CoA dehydrogenase deficiency and *DMD* in Duchenne muscular dystrophy. The useful nature of these databases is illustrated by the findings that have been prompted by the information contained within them. The UMD APC LSDB is the only *APC* gene mutation database available to the scientific community. It was the basis for the discovery of the dependence of the second hit (somatic mutations) on the site of the first hit (germline mutations)[28]. It also led to the demonstration that in families with inherited colon cancer that showed high levels of somatic G:C–T:A transversions in *APC*, the disease was caused by Mut-Y homologue (*MYH*) germline mutations[29]. Moreover, analysis of the pattern of *TP53* mutations was the basis for a large number of molecular epidemiology studies demonstrating the relationship between exposure to exogenous carcinogens and p53 mutations in different cancer types[30,31] (BOX 1).

Germline versus somatic mutations

Broadly, DNA alterations have two origins: germline or somatic. Germline mutations (also called constitutional mutations) are found in all cells of the organism, including germline cells, and can be transmitted to the offspring. They are the cause of most hereditary genetic diseases, such as cystic fibrosis, and most myopathies or heamoglobinopathies. They are also involved in inherited cancer syndromes such as familial adenomatous polyposis (caused by mutations in *APC*), familial breast cancer (caused by mutations in *BRCA1* or *BRCA2*) or Li–Fraumeni syndrome (LFS, caused by mutations in *TP53*). Somatic mutations are acquired in somatic tissue during the subject's lifetime and predominantly give rise to neoplastic disease with mutations restricted to the tumour.

Germline mutations are easy to detect when using specific methodology and an adequate screening strategy, but this process is time-consuming and costly for large genes. DNA or RNA is usually extracted from blood cells, leading to large amounts of good quality genetic material. Depending on the type of disease — autosomal-dominant or recessive — the cells will carry either one or two mutations. The biological significance of deletions, insertions or nonsense mutations is usually obvious, but the main problem concerns germline missense mutations. It is difficult to determine whether the detected sequence variant is a causal mutation or a neutral (polymorphic) variation without any effect on phenotype[32]. Only extensive studies that include segregation analysis of the mutation inside the family, phylogenetic conservation of the targeted codon, or statistical analysis can unravel the pathogenicity of the mutation, but often the question remains unresolved. It is usually assumed that variants that are found in more than 1% in the normal population could correspond to polymorphisms, but this feature might be heterogeneous between various regions of the genome.

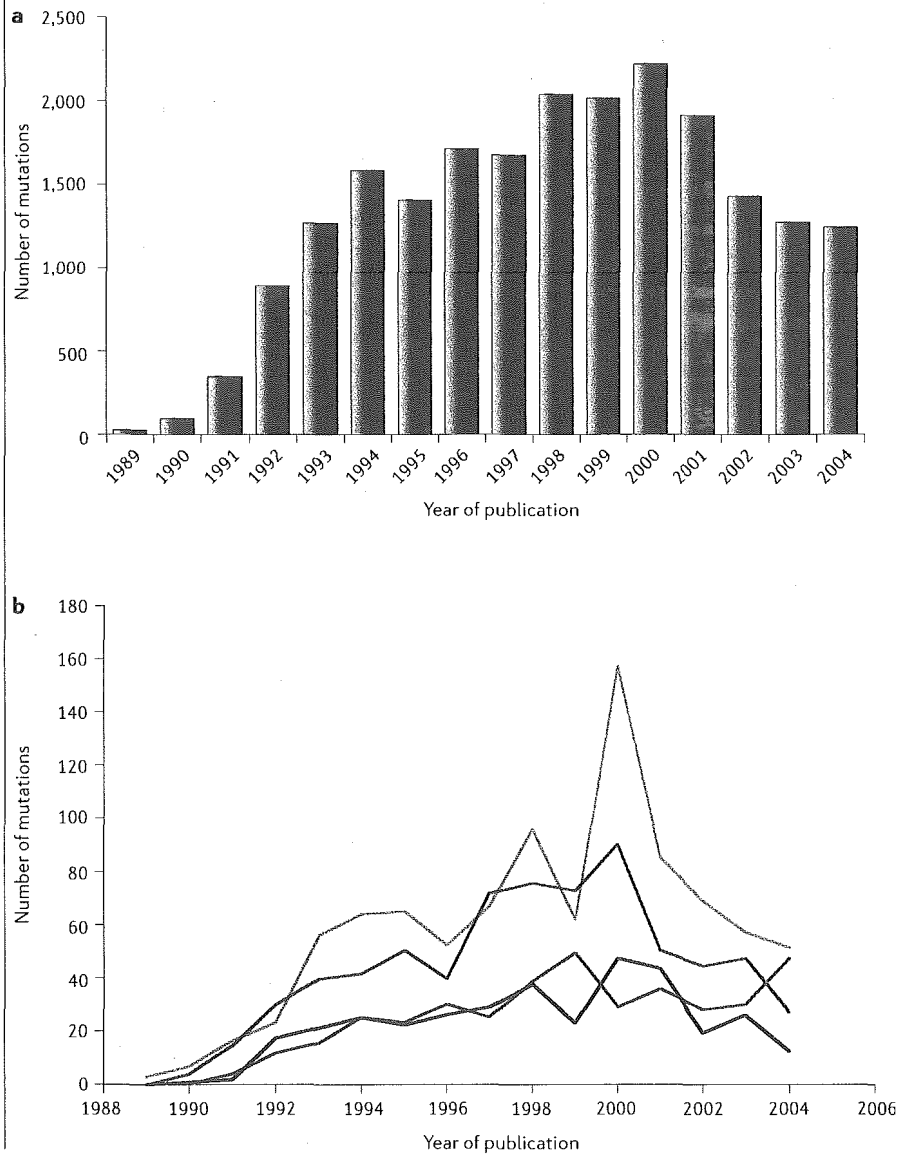
For somatic mutations, this problem usually does not arise, as comparison of normal and neoplastic tissue from the same patient will unambiguously identify the mutation. Although most primary tumours have a few somatic mutations in different genes — in other words, 'driver mutations' selected for a growth advantage — some tumours with a mutator phenotype can have hundreds of mutations. These mutations, termed either 'hitch-hiking mutations' or 'passenger mutations', do not confer any selective advantages and are co-selected with a driver mutation[33,34]. The frequency of these passenger mutations is unknown at the present time and could vary according to the type of cancer. As there is no easy way to distinguish between these two types of mutations, numerous passenger mutations will probably contaminate all mutation databases. Detection of these passenger mutations constitutes an important challenge, as these mutations do not have any clinical relevance and could interfere with database analysis.

An important problem for somatic mutations concerns the difficulty of their detection. Tumour samples can be highly heterogeneous in terms of content and origin, both of which have a profound impact on the diagnosis. The main problem is contamination of the tumour by normal cells such as stromal cells or infiltrating lymphocytes. This contamination can be low for surgical specimens (less than 20%), but can be more than 90% for biopsies. For biological specimens such as serum, sputum, faeces or urine, tumour cells can account for less than 0.01% of all cells. Pre-screening methodologies such as single-strand conformation polymorphism (SSCP), denaturing high-performance liquid chromatography (DHPLC) or denaturing gradient gel electrophoresis (DGGE) can increase the sensitivity of detection, but not when the target constitutes less than 5% of the original sample material. The second problem concerns the protocol used to conserve the sample. Although material frozen directly after collection generates good quality DNA, DNA from fixed and paraffin-embedded tissue is often degraded and analysis can lead to the artefactual discovery of mutations. Furthermore, the low yield of DNA obtained from these samples can necessitate the use of two rounds of PCR amplification (nested PCR), a procedure known to be error-prone if not carefully controlled.

Most tumour-suppressor genes can be mutated constitutionally and somatically in family syndromes and sporadic cancers. Nevertheless, for each gene, databases for germline and somatic mutations must be clearly dissociated, as they have different interpretations and applications. For somatic alterations, hot-spot regions have been demonstrated, corresponding to a DNA region that is susceptible to mutations, or a codon that encodes a key residue in the biological function of the protein, or both. Whatever the explanation, it is synonymous with a high rate of *de novo* mutations at this position. This information can be useful for pinpointing important regions of the protein. For germline mutations, mutation hot-spots can also be associated with a founder effect in which most carriers of the mutation descend from a single ancestor. Selection of some founder mutations over large periods of time can be explained by their association with certain specific selective advantages. The best example is the ΔF508 mutation found in the cystic fibrosis transmembrane conductance regulator gene (*CFTR*) in 70% of Caucasian patients with cystic fibrosis. Genetic analysis has shown that all of these patients originate from a single ancestor and that this mutation is a

Box 2 | **Growth of p53 mutation publications**

Since the first publication in 1989, there was a constant increase in the number of publications describing *TP53* mutations, culminating in 2000 (see part **a**). The decrease first observed in 2001 is continuing. This is not because of a lack of interest in p53, but because of the difficulty in publishing *TP53* mutations in peer-reviewed journals, owing to a lack of novelty. Furthermore, in recent publications, *TP53* mutations are not fully described owing to journal space considerations. Many laboratories have unpublished *TP53* mutations that are not included in the database. This problem is not specific to *TP53* — it also concerns other genes and raises an important issue related to the publication of mutations. It is estimated that 50% of all mutations in various databases are unpublished and have been collected by the curators. This unpublished information has not been peer-reviewed, which raises the problem of its accuracy and how it should be curated before being entered in a database. The trend for publishing p53 mutation data has also decreased in most instances. The publication trend for two hot-spot mutations (R175H and R248Q) and one moderate mutation (G245S) follows the publication trend observed above with a decrease since 2000 (see part **b**). The publication rate for unique missense mutations has not decreased and has even showed a slight increase in 2004. Several non-exclusive explanations can account for this progression: more thorough analysis of the whole *TP53* gene instead of exons 5 to 8 (50% of unique mutations lie outside exons 5 to 8), analysis of early neoplastic lesions that do not have full loss of p53 activity, and less stringent experimental procedures. The lines shown in part **b** represent the publication trends for R175H (lilac), R248Q (green) 245S (blue) and unique missense mutations (red).



unique molecular event[35]. A founder effect is therefore associated with social or geographical isolation, and the number of patients associated with the same mutation does not correspond to the mutation rate. The increasing rate of discovery of founder mutations in human populations and their heterogeneity in disease penetrance will lead to the development of mutation databases that will require extensive social and clinical information that might preclude their usefulness.

Structure and accessibility

LSDBs often originate as loosely organized compilations of data. Curators choose from the available database management systems or create their own system, depending on their abilities. As no standard has yet been adopted, the way data is presented in LSDBs varies enormously. Most curators use flat file, plain text databases or spreadsheet programs (such as Microsoft Excel) as a simple means to collate and store data on mutations, but neither the search nor the retrieval of specific data are possible. More sophisticated databases use MySQL, an open source database management software that runs on most platforms. A minority of curators use specialized or generic software such as the UMD[36], the Mutation Storage and Retrieval Program (MuStaR)[37], or the Leiden Open Source Variation Database (LOVD)[38]. The use of these complex relational databases allows specific analysis of either the entire database or any customized subset.

One of the greatest needs for the future is the capacity to link, merge and interrogate multiple gene mutation databases. It is now well-accepted that neoplastic transformation of normal cells requires mutations in multiple genes to achieve inactivation of the various pathways that control cell growth, apoptosis or genome integrity. Extensive analysis of well-characterized tumours indicates that there is a non-random pattern in gene inactivation. Alteration of *BRAF* and *KRAS* are mutually exclusive in various types of cancers. A similar situation can be observed in colon cancer with mutation occurring in either *CTNNB1* (the gene that encodes β-catenin) or *APC*. This situation is easily explained by the fact that such gene pairs belong to the same biological pathways and there is usually no need for redundant inactivation.

Although some observations are obvious and can be easily detected, more subtle correlations can be revealed by more extensive database searches. It would therefore be highly desirable to link LSDBs for various genes to evaluate the 'mutation profile' of tumours. Theoretically, this should be easy

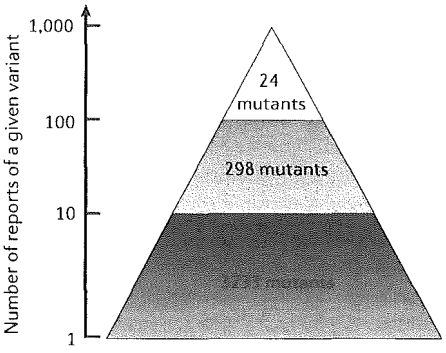


Figure 1 | **Frequency of p53 mutants in the Universal Mutation Database.** There are 3,555 different p53 mutants in the database, but their frequency is heterogeneous. 3,233 mutants are found at frequencies ranging from 1 to 10 times, with 1,874 mutants described only once. 298 mutants have been described at intermediate frequencies (between 11 and 99 times). Only 24 mutants are found more than 100 times, with the highest frequency of 979 times for the R175H mutant. Overall, the latest version of the p53 database includes 21,717 mutations.

because most clinical laboratories work on limited subsets of informative tumours that are used for multiple genetic analyses. Unfortunately, all of this information cannot be linked at the present time. Mutation analyses for various genes are published over different periods in various journals using different sample names for similar tumours, preventing cross-referencing. Only a few papers have performed multigene analysis, but most of the papers on mutations are restricted to single genes. It would be highly advisable to improve this situation to allow meta-analysis with linked LSDBs. A first step would be to ensure that authors always use the same nomenclature for sample labelling. Another step would be to define an international nomenclature for samples that could include various types of information such as country, clinical centre and unique ID. As this information is already indicated in the material section of publications, this type of nomenclature would not raise any ethical issues.

The other advantage of using a homogeneous label is that it would facilitate merging of information from databases developed on different platforms or with different generic software. Most of the fields of the various mutation databases are similar (codon position, exon, wild-type and mutant codon, mutation ID, and so on) and the nomenclature for each mutation is already well-established and many journals require the use of this name in their publications[39]. The feasiblilty of merging databases, once homogeneous standardization has been achieved, was illustrated by the work on the p53 databases performed at the EMBL European Bioinformatics Institute (EBI). Although six p53 databases are available, only two have been reproducibly updated. Despite being developed on different platforms, these two p53 databases can be easily merged by using certain fields such as tumour ID or references to identify common entries.

Access to most mutation databases has not been a crucial issue as they were developed in academic laboratories and most of them are freely available, whereas only a few require registration. However, the issue of intellectual property rights concerning biological databases has been a major concern and these issues are far from being resolved. A legal framework has been developed for the protection of databases both in Europe and in the USA, but some remaining differences and gaps have prevented perfect legislation[40,41].

A p53 mutation database

The first *TP53* mutations were published in 1989. At present, more than 2,000 publications have reported the description of p53 alterations in various neoplasms and also in other diseases, such as rheumatoid arthritis. The latest version of the UMD p53 database contains 21,717 mutations, which is approximately 30% of all mutations found in human diseases reported to date (April 2005 release). The decreasing number of reports describing *TP53* mutations since 2001 is mainly due to the difficulty of publication (BOX 2). Several thousand mutations in the *TP53* gene, and other genes, are currently unpublished and unavailable in mutation databases. To address this issue, the HGVS has initiated a project to collect all this information by means of newly developed web software called the Waystation Project. However, this software is still at the stage of testing and more volunteers are required for validation of this important project.

Bias in the p53 mutation database

In 2001 and then in 2003, several reservations were expressed concerning the biological significance of some *TP53* mutations[42,43]. First, there is a marked difference in the frequency between the various mutations, with occurrences ranging from once (401 mutants) to 979 times (mutant R175H) (FIG. 1). Structural and biological studies have

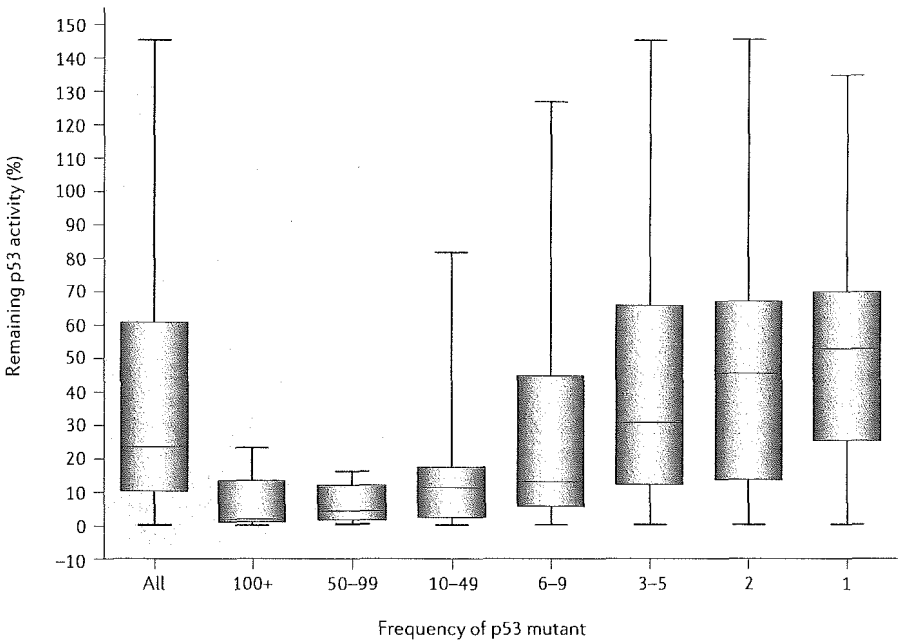


Figure 2 | **Activity of p53 mutants according to their frequency in the Universal Mutation Database.** The 3,555 p53 mutants have been classified into 7 categories according to their frequencies in the database, ranging from more than 100 times (hot-spot mutants) to rare mutants (frequency shown on the x-axis). The y-axis corresponds to the transcriptional activity of p53 mutants, in which 100% corresponds to the activity of the wild-type protein. Box and whisker plots show the upper and lower quartiles and range (box), median value (horizontal line inside the box), and full range distribution (whisker line). Loss of p53 activity is clearly observed for hot-spot mutants, whereas rare mutants behave more heterogeneously, with a high proportion of mutants that do not show any loss of activity. This figure is an updated version of the analysis described in REF. 46.

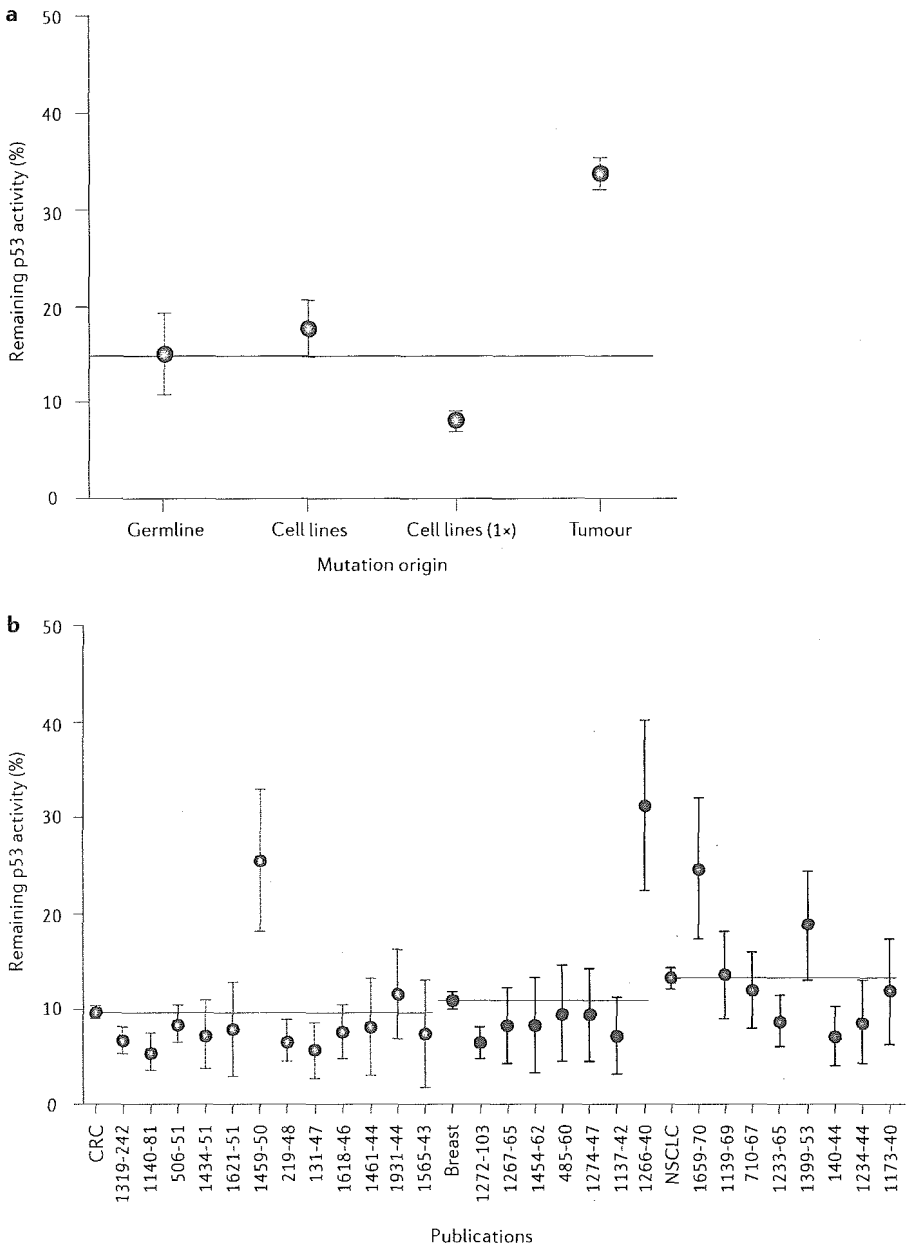


Figure 3 | **p53 loss of function. a** | Distribution of p53 loss of function. The dot and bars show the mean and 95% confidence interval (CI) of p53 activity as measured by transactivation of the *CDKN1A* (which encodes p21) promoter. The y-axis corresponds to p53 transactivation activity as described in FIG. 2, and the x-axis lists the origin of the p53 mutant. The horizontal line represents the mean of the loss of p53 function in germline mutations. Many cell lines have more than one mutation that can be either the same allele or on two different alleles. It has been previously shown that only one of the mutants shows loss of activity, whereas the second mutant is either more leaky or shows no loss of activity, indicating that only one of the mutants is important for the development of the tumour. Cell lines (1x), cell lines with only one mutation; tumour, all tumours in the database. **b** | Meta-analysis of p53 loss of function. For each cancer, the remaining activity of the p53 mutant in each publication is graphically displayed. The dot and bars show the mean and 95% CI of p53 activity, as measured by transactivation of the *CDKN1A* promoter. The horizontal line shows the mean of the combined studies. The publication code is indicated on the x-axis: the first number is an anonymous ID for the publication, and the second number indicates the number of p53 mutants included in this study. The reference value corresponding to the mean and 95% CI of all studies for the specific cancer is shown on the far left of for each cancer type. Studies are presented from left to right in decreasing order of the number of p53 mutants. The y-axis corresponds to p53 transactivation activity as described in FIG. 2. Only studies with 40 or more p53 mutations are shown on the graph. Red, colorectal cancer (CRC); blue, breast cancer; green, non-small-cell lung cancer (NSCLC). Adapted from REF. 53.

shown that mutations at hot-spot positions inactivate p53 growth-suppressive and apoptotic activities, which explains why they are found at a high frequency in human cancers. On the other hand, the significance of the rare mutants remained elusive as they were not the subject of intense scrutiny. The trend for publication of rare mutants has not decreased compared with that of other mutants (BOX 2). Second, the inclusion of reports with unusual and non-reproducible patterns of p53 mutations can alter the quality of the database. Although an unbiased database should contain all publications in the literature, some reports are notoriously dubious. Inclusion of artefactual results has a number of harmful effects, both intellectually when they are quoted indiscriminately by non-specialists, but also for the integrity of the databases. This is clearly illustrated by the debate on the origin of p53 mutations in lung cancer, challenged by the tobacco industry, as recently discussed by Bitton and colleagues[44].

The recent construction and biological analysis of 2,300 p53 mutants shed some new light on p53 mutant activity and allowed some unique analyses[45,46]. The remaining activity of mutants that are often found in the *TP53* database is usually low, ranging from 0% to 20% compared with the normal protein (FIG. 2). For rare mutants, the scatter is very heterogeneous, ranging from 0% to 160%. There is a clear inverse correlation between the frequency of p53 mutants and their activity. Of the mutants that have been found only once, approximately half have an activity greater than 50% when compared with wild-type p53, indicating that the importance, if any, of these mutations is very low. Analysis of the origin of p53 mutations indicates that several methodological biases are responsible for this observation.

TP53 germline mutations that are found in patients with LFS sustain the most drastic loss of activity (FIG. 3a). LFS is a rare autosomal-dominant syndrome in which patients are predisposed to a wide variety of cancers. Diagnosis of *TP53* germline mutations is performed by trained staff in specific laboratories using robust protocols that require analysis both at the DNA and RNA level to identify anomalies that are important for diagnosis and follow-up of these families. The situation is similar for cell lines because good quality genetic material can be analysed and mutations are usually present in 100% of cells. For tumours, the mean loss of activity in p53 mutants is lower than that observed in samples from patients with LFS or from cell lines, and there is a wide range of distribution that reflects a considerable

heterogeneity in p53 mutant activity. This is because of the large number of rare mutations that show no loss of activity.

Meta-analysis of the 2,000 publications describing *TP53* mutations led to the discovery of reports that are characterized by having a high frequency of rare variants with no loss of activity[53] (FIG. 3b). Close examination of these publications showed several anomalies that have cast serious doubts on these results, such as multiple mutations in each tumour, a high frequency of 'neutral' mutations or an unusual pattern of mutations in other genes. A methodological bias has also been demonstrated, as 55% of these studies use nested PCR versus 8% of all other studies that have identified a common p53 mutant with loss of activity[53]. Although this methodology is powerful in amplifying minute amounts of DNA, it is error-prone if not carefully controlled. It is usually used for DNA extracted from paraffin-embedded tissue, which is often fragmented. Furthermore, fixation procedures can lead to chemical modification of the DNA that can generate misincorporation during PCR amplification. The use of archival tissue enables the rapid and convenient assay of a large number of tissues from specific diseases that would take years to accumulate prospectively, but it has some drawbacks. In the p53 database, data from these ambiguous reports account for 4% to 8% of all reported mutations depending on the type of cancer. p53 alterations in lung cancers from one publication (1659-70 in FIG. 3) has been a constant problem in the heated debate on the origin of mutations in smokers. Its inclusion in statistical analyses lead to serious bias and a change in the outcome of any study[47].

All these points raise the important question of how to include this information in the various databases. Two different approaches have been adopted. In the p53 database of the International Agency for Research on Cancer (IARC), all p53 mutations are included irrespective of the quality of the study. In the UMD p53 database, a manual curation eliminates ambiguous reports. In both databases, however, careful curation ensures removal of duplicate data, the other major plague of p53 mutation databases. Both approaches have advantages and disadvantages. Combining all publications ensures an unbiased database, but includes all dubious data; whereas curation, although it removes dubious data, could be dangerous because a meaningful report of an unusual p53 mutation could be discarded. The knowledge of p53 mutant activity and the meta-analysis of the p53 mutant database now provide a more objective approach to distinguish dubious studies. In the latest version of the UMD p53 database, all publications have been included, but those that differ statistically from the range of other studies have been tagged with a warning. This procedure will allow accurate analysis to be performed with greater confidence.

Guidelines for LSDB curation

We think that the problems described above are not limited to the p53 databases and extend to other databases. LSDBs are tools developed by the scientific community for the scientific community. As already indicated above, the value of these databases has now been clearly established, but it is essential to guarantee their quality. The pollution by artefactual results has a number of harmful effects, both intellectually when they are quoted indiscriminately by non-specialists, but also for analysis of the databases. Their inclusion in LSDBs can also mask other original studies describing real differences in mutation profiles. Quality control must therefore be applied at all levels. This problem concerns not only laboratories but also journals and anonymous reviewers involved in the publication of this information.

We suggest the following guidelines for database curation. All mutations should be reported, including 'neutral' mutations that do not change the amino acid as they can affect splicing or RNA stability. Polymorphisms, such as those at codons 72 or 213 for the *TP53* gene, should not be included as mutations. When mutations are unusual in terms of frequency, multiplicity or profile, confirmation by independent analyses should be performed. This concept of independent repetition has a very broad definition from one laboratory to another. The most rigorous approach consists of repeating the experiment by starting as early as possible in the process — for example, with a second DNA extraction from the sample when available. When using archival samples (paraffin-embedded tissue), a negative control should be performed with DNA extracted from the same sample using the same extraction procedure as for tumour DNA. The properties of most p53 mutants are now available on the UMD website and should be checked for validation. An excessive number of mutations with wild-type activity or rarely reported in the literature should be considered carefully. Typographical errors in mutation tables (approximately 10% to 20% of all reports of *TP53* mutations) should be avoided by carefully checking the data. New tools to generate accurate *TP53* mutation tables are now available on the p53 and UMD website. Using the official nomenclature for description of mutations is a good way to prevent ambiguities and errors[39]. Reviewers and editors must carefully check that all of these procedures have been followed. To facilitate this process, journal editors who are actively involved in the publication of *TP53* mutations should refer authors to the UMD p53 website to check their data. Although these guidelines specifically concern p53, they can obviously be applied to other genes involved in cancer.

Recently, several studies have addressed the biological significance of *BRCA1* germline mutations to identify variants that are causally linked to breast and ovarian cancer[48,49]. Inclusion of this information in the BRCA1 mutation database will be essential to improve the value of genetic counselling.

Application of these simple rules will be beneficial for the entire scientific community. Apart from ensuring the author's compliance with a rigorous scientific and technological approach, reviewers and editors must also act as gatekeepers to ensure that the quality of the information published is maintained at a level of excellence.

Thiery Soussi is at the Université P.M. Curie, 4 place Jussieu, 75005 Paris, France, and the Karolinska Institute, Department of Oncology-Pathology, Cancer Center Karolinska (CCK), SE-171 76 Stockholm, Sweden.

Chikashi Ishioka is at the Department of Clinical Oncology, Institute of Development, Aging, and Cancer, Tohoku University, Sendai 980-8575, Japan.

Mireille Claustres and Christophe Béroud are at the Laboratoire de Génétique Moléculaire et Chromosomique, Institut Universitaire de Recherche Clinique et CHU, CNRS UPR 1142, 641, avenue du Doyen Gaston Giraud, 34093 Montpellier Cedex 5, France.

Correspondence to T.S.
e-mail: thierry.soussi@free.fr
doi:10.1038/nrc1783

1. Collins, F. S. Positional cloning moves from perditional to traditional. *Nature Genet.* **9**, 347–350 (1995).
2. Cotton, R. G. Mutation detection 2001: novel technologies, developments and applications for analysis of the human genome. *Hum. Mutat.* **19**, 313–314 (2002).
3. Paalman, M. H., Cotton, R. G. & Kazazian, H. H. Jr. Variation, databases, and disease: new directions for human mutation. *Hum. Mutat.* **16**, 97–98 (2000).
4. Claustres, M., Horaitis, O., Vanevski, M. & Cotton, R. G. Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. *Genome Res.* **12**, 680–688 (2002).
5. Montesano, R., Hainaut, P. & Wild, C. P. Hepatocellular carcinoma: from gene to public health. *J. Natl Cancer Inst.* **89**, 1844–1851 (1997).
6. Mulligan, L. M. *et al.* Germ-line mutations of the *RET* proto-oncogene in multiple endocrine neoplasia type 2A. *Nature* **363**, 458–460 (1993).
7. Hofstra, R. M. *et al.* A mutation in the *RET* proto-oncogene associated with multiple endocrine neoplasia type 2B and sporadic medullary thyroid carcinoma. *Nature* **367**, 375–376 (1994).
8. Xue, F. *et al.* Germline *RET* mutations in MEN 2A and FMTC and their detection by simple DNA diagnostic tests. *Hum. Mol. Genet.* **3**, 635–638 (1994).
9. Edery, P. *et al.* Mutations of the *RET* proto-oncogene in Hirschsprung's disease. *Nature* **367**, 378–380 (1994).
10. Romeo, G. *et al.* Point mutations affecting the tyrosine kinase domain of the *RET* proto-oncogene in Hirschsprung's disease. *Nature* **367**, 377–378 (1994).

11. Olschwang, S. *et al.* Restriction of ocular fundus lesions to a specific subgroup of APC mutations in adenomatous polyposis coli patients. *Cell* **75**, 959–968 (1993).
12. Spirio, L. *et al.* Alleles of the *APC* gene: an attenuated form of familial polyposis. *Cell* **75**, 951–957 (1993).
13. Gallou, C. *et al.* Mutations of the *VHL* gene in sporadic renal cell carcinoma: definition of a risk factor for VHL patients to develop an RCC. *Hum. Mutat.* **13**, 464–475 (1999).
14. Pearson, P. L. The genome data base (GDB) — human gene mapping repository. *Nucleic Acids Res.* **19** (Suppl.), 2237–2239 (1991).
15. Bilofsky, H. S. *et al.* The GenBank genetic sequence databank. *Nucleic Acids Res.* **14**, 1–4 (1986).
16. Hamm, G. H. & Cameron, G. N. The EMBL data library. *Nucleic Acids Res.* **14**, 5–9 (1986).
17. Bairoch, A. & Boeckmann, B. The SWISS-PROT protein sequence data bank. *Nucleic Acids Res.* **19** (Suppl.), 2247–2249 (1991).
18. Merali, Z. & Giles, J. Databases in peril. *Nature* **435**, 1010–1011 (2005).
19. Horaitis, O. & Cotton, R. G. The challenge of documenting mutation across the genome: the human genome variation society approach. *Hum. Mutat.* **23**, 447–452 (2004).
20. Stenson, P. D. *et al.* Human Gene Mutation Database (HGMD): 2003 update. *Hum. Mutat.* **21**, 577–581 (2003).
21. Béroud, C., Collod-Béroud, G., Boileau, C., Soussi, T. & Junien, C. UMD (Universal Mutation Database): a generic software to build and analyze locus-specific databases. *Hum. Mutat.* **15**, 86–94 (2000).
22. Béroud, C. & Soussi, T. *p53* and *APC* gene mutations: software and databases. *Nucleic Acids Res.* **25**, 138–138 (1997).
23. Wautot, V. *et al.* Germline mutation profile of *MEN1* in multiple endocrine neoplasia type 1: search for correlation between phenotype and the functional domains of the MEN1 protein. *Hum. Mutat.* **20**, 35–47 (2002).
24. Béroud, C. *et al.* Software and database for the analysis of mutations in the *VHL* gene. *Nucleic Acids Res.* **26**, 256–258 (1998).
25. Jeanpierre, C., Béroud, C., Niaudet, P. & Junien, C. Software and database for the analysis of mutations in the human *WT1* gene. *Nucleic Acids Res.* **26**, 271–274 (1998).
26. Collod, G., Béroud, C., Soussi, T., Junien, C. & Boileau, C. Software and database for the analysis of mutations in the human *FBN1* gene. *Nucleic Acids Res.* **24**, 137–140 (1996).
27. Varret, M. *et al.* Software and database for the analysis of mutations in the human LDL receptor gene. *Nucleic Acids Res.* **25**, 172–180 (1997).
28. Lamlum, H. *et al.* The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis. *Nature Med.* **5**, 1071–1075 (1999).
29. Al-Tassan, N. *et al.* Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nature Genet.* **30**, 227–232 (2002).
30. Soussi, T. The *p53* tumour suppressor gene: a model for molecular epidemiology of human cancer. *Mol. Med. Today* **2**, 32–37 (1996).
31. Harris, C. C. *p53* tumor suppressor gene: at the crossroads of molecular carcinogenesis, molecular epidemiology, and cancer risk assessment. *Environ. Health Perspect.* **104**, 435–439 (1996).
32. Nelson, D. R. 'A variant of uncertain significance' and the proliferation of human disease gene databases. *Hum. Genomics* **2**, 70–74 (2005).
33. Rodin, S. N., Holmquist, G. P. & Rodin, A. S. CPG transition strand asymmetry and hitch-hiking mutations as measures of tumorigenic selection in shaping the p53 mutation spectrum. *Int. J. Mol. Med.* **1**, 191–199 (1998).
34. Futreal, P. A. *et al.* A census of human cancer genes. *Nature Rev. Cancer* **4**, 177–183 (2004).
35. Morral, N. *et al.* The origin of the major cystic fibrosis mutation (ΔF508) in European populations. *Nature Genet.* **7**, 169–175 (1994).
36. Beroud, C. *et al.* UMD (Universal Mutation Database): 2005 update. *Hum. Mutat.* **26**, 184–191 (2005).
37. Brown, A. F. & McKie, M. A. MuStaR and other software for locus-specific mutation databases. *Hum. Mutat.* **15**, 76–85 (2000).
38. Fokkema, I. F., den Dunnen, J. T. & Taschner, P. E. LOVD: easy creation of a locus-specific sequence variation database using an 'LSDB-in-a-box' approach. *Hum. Mutat.* **26**, 63–68 (2005).
39. den Dunnen, J. T. & Antonarakis, S. E. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum. Mutat.* **15**, 7–12 (2000).
40. Maurer, S. M., Hugenholtz, P. B. & Onsrud, H. J. Intellectual property. Europe's database experiment. *Science* **294**, 789–790 (2001).
41. Greenbaum, D. & Gerstein, M. A universal legal framework as a prerequisite for database interoperability. *Nature Biotechnol.* **21**, 979–982 (2003).
42. Soussi, T. & Béroud, C. Assessing TP53 status in human tumours to evaluate clinical outcome. *Nature Rev. Cancer* **1**, 233–240 (2001).
43. Soussi, T. & Béroud, C. Significance of TP53 mutations in human cancer: a critical analysis of mutations at CpG dinucleotides. *Hum. Mutat.* **21**, 192–200 (2003).
44. Bitton, A., Neuman, M. D. & Glantz, S. A. The *p53* tumour suppressor gene and the tobacco industry: research, debate and conflict of interest. *Lancet* **365**, 1–10 (2005).
45. Kato, S. *et al.* 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 (2003).
46. Soussi, T., Kato, S., Levy, P. P. & Ishioka, C. Reassessment of the *TP53* mutation database in human disease by data mining with a library of *TP53* missense mutations. *Hum. Mutat.* **25**, 6–17 (2005).
47. Krawczak, M. & Cooper, D. N. p53 mutations, benzo[a]pyrene and lung cancer. *Mutagenesis* **13**, 319–320 (1998).
48. Phelan, C. M. *et al.* Classification of BRCA1 missense variants of unknown clinical significance. *J. Med. Genet.* **42**, 138–146 (2005).
49. Fleming, M. A., Potter, J. D., Ramirez, C. J., Ostrander, G. K. & Ostrander, E. A. Understanding missense mutations in the *BRCA1* gene: an evolutionary approach. *Proc. Natl Acad. Sci. USA* **100**, 1151–1156 (2003).
50. Tornaletti, S. & Pfeifer, G. P. Complete and tissue-independent methylation of CpG sites in the *p53* gene: implications for mutations in human cancers. *Oncogene* **10**, 1493–1499 (1995).
51. You, Y. H., Li, C. & Pfeifer, G. P. Involvement of 5-methylcytosine in sunlight-induced mutagenesis. *J. Mol. Biol.* **293**, 493–503 (1999).
52. Denissenko, M. F., Chen, J. X., Tang, M. S. & Pfeifer, G. P. Cytosine methylation determines hot spots of DNA damage in the human *p53* gene. *Proc. Natl Acad. Sci. USA* **94**, 3893–3898 (1997).
53. Soussi, T. *et al.* Meta-analysis of the p53 mutation database for mutant p53 biological activity reveals a methodological bias in mutation detection. *Clin. Cancer Res.* (in the press).

Acknowledgements
Development of the UMD software is supported by grants from AFM to C.B. and M.C. Work of T.S. is supported by the ARC and ligue contre le cancer (comité de Paris). We are grateful to Richard Cotton for critical reading of this manuscript.

Competing interests statement
The authors declare no competing financial interests.

DATABASES
The following terms in this article are linked online to:
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
APC | *BRCA1* | *BRCA2* | *FBN1* | *NF1* | *RB1* | *RET* | *TP53* | *TTN* | *VHL*
National Cancer Institute: http://www.cancer.gov
breast cancer | colon cancer | hepatocellular carcinoma | lung cancer
OMIM: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
Marfan syndrome | von Hippel–Lindau syndrome

FURTHER INFORMATION
EMBL: http://www.ebi.ac.uk/embl/
GDB Human Genome Database: http://gdbwww.gdb.org/
GenBank: http://www.psc.edu/general/software/packages/genbank/genbank.html
Human Gene Mutation Database: http://www.hgmd.cf.ac.uk/hgmd0.html
Mutation Storage and Retrieval Program: http://www.hgu.mrc.ac.uk/Softdata/Mustar/
Leiden Open Source Variation Database: http://www.dmd.nl/LOVD/1.1.0/
MySQL: http://www.mysql.com/
p53 database of the International Agency for Research on Cancer: http://www.iarc.fr/p53/
Swiss-Prot: http://www.expasy.org/sprot/
The p53 Database: http://www.p53.free.fr
Universal Mutation Database for p53: http://www.umd.be:2072/index.shtml
Waystation Project: http://www.centralmutations.org
Access to this interactive links box is free online.

ONLINE CORRESPONDENCE ✉
Nature Reviews Cancer publishes items of correspondence online. Such contributions are published at the discretion of the Editors and can be subject to peer review. Correspondence should be no longer than 500 words with up to 15 references and should represent a scholarly attempt to comment on a specific Review or Perspective article that has been published in the journal. To view correspondence, please go to our homepage and select the link to New Correspondence, or use the URL indicated below.

The following correspondence has recently been published:

Time to harness the pro-apoptotic property of NFκB?

Radhakrishnan, S. K. and Kamalakaran, S.

http://www.nature.com/nrc/archive/correspondence.html

This correspondence relates to the article:

Nuclear factor-κB inhibitors as sensitizers to anticancer drugs

Nakanishi, C. and Toi, M.

Nature Rev. Cancer **5**, 297–309 (2005)

原　著

大腸癌に対する *l*-LV＋5-FU療法の効果および有害反応の検討

千葉奈津子　吉岡　孝志　加藤　俊介　高橋　信　柴田　浩行
加藤　誠之　酒寄　真人　大塚　和令　角道　祐一　大堀　久詔
山浦　玄悟　高橋　雅信　安田　勝洋　石岡千加史*

〔*Jpn J Cancer Chemother* **32**(7): 991-995, *July,* 2005〕

Retrospective Analysis on Efficacy and Toxicity of 5-Fluorouracil (5-FU) and *l*-leucovorin (*l*-LV) in Advanced or Recurrent Colorectal Cancer: Natsuko Chiba, Takashi Yoshioka, Shunsuke Kato, Shin Takahashi, Hiroyuki Shibata, Satoshi Kato, Masato Sakayori, Kazunori Otsuka, Yuichi Kakudo, Hisatsugu Ohori, Gengo Yamaura, Masanobu Takahashi, Katsuhiro Yasuda and Chikashi Ishioka (*Dept. of Clinical Oncology, Institute of Development, Aging and Cancer (IDAC) and Tohoku University Hospital, Tohoku University*)

Summary

The aim of this study was to evaluate the efficacy and toxicity of 5-fluorouracil (5-FU) and *l*-leucovorin (*l*-LV) in 50 patients with advanced or recurrent colorectal cancer in our institute. The dose of 5-FU was 600 mg/m² and the dose of *l*-LV was 250 mg/m². Objective response were 36.8% of patients who had administration of full-dose and 14.8% of patients who had the administration of reduced dose or prolonged interval. No significant difference was observed in clinical benefit rates between patients administrated in full-dose and patients in reduced dose or prolonged interval. Median survival time (MST) of patients in reduced dose or prolonged interval is longer than patients in full-dose. These data suggest that 5-FU/*l*-LV can be given in the outpatient and yields improved prognosis and minimal adverse reactions even in patients in reduced dose or prolonged interval. **Key words**: Colorectal cancer, 5-Fluorouracil, *l*-Leucovorin (*Received Nov. 18, 2004/Accepted Dec. 24, 2004*)

要旨　*l*-LV/5-FU療法が施行された切除不能，転移・再発大腸癌50症例において，前治療の有無，投与間隔の延長，投与量の減量の効果，安全性に与える影響をretrospectiveに検討した。前治療の有無では，前治療のない群が前治療のある群より高い奏効率を示したが，SD症例も含めた病勢コントロール可能例を含めると前治療の有無による大きな差は認められなかった。しかし，MSTでは差を認め前治療のない群のほうが長期であった。また，減量投与した群では標準的投与法を行った群の4割程度の奏効率となったが，SD症例まで含めた病勢コントロールとMSTはむしろ標準投与を行った群より高く，減量投与した群でも高い治療効果が得られた。以上より，全身状態・副作用などにより標準治療が行えない場合でも，減量によって治療を継続することで病気の進行を制御でき，予後を改善する可能性が示唆された。

はじめに

5-fluorouracil (5-FU) は大腸癌の化学療法において，中心的な役割を果たしている。その作用機序は主に，5-FUの代謝産物であるFdUMPと還元型葉酸がthymidylate synthase (TS) に強力に結合して複合体を形成し，TS活性が阻害されることによると考えられている[1,2]。

5-FUの増強効果をめざして，5-FUをeffectorとしたbiochemical modulationとしてLeucovorin (LV) の併用療法が多く報告されており[3-6]，5-FUの静注，点滴静注，*l*-LV低用量，高用量と様々な方法が行われている。一般的に奏効率は20～30%，全生存期間の中央値(以下MST) は10～13か月とされている。

本邦では，1999年*l*-LVを用いた*l*-LV/5-FU療法の毎週投与法が承認を得ており，5-FU 600 mg/m²静注，*l*-LV 250 mg/m²の2時間点滴静注が保険適応となっている。本療法は週1回の外来通院での化学療法が可能で，

*　東北大学病院・腫瘍内科
東北大学加齢医学研究所・癌化学療法研究分野

連絡先: 〒980-8575 仙台市青葉区星陵町4-1　東北大学加齢医学研究所・癌化学療法研究分野
石岡千加史

0385-0684/05/¥500/論文/JCLS

表1 患者背景

背景因子		例数	年齢
性	男性	26	30～79(中央値59)
	女性	24	30～74(中央値63)
原発巣の有無	あり	12	
	なし	38	
原発部位	回盲部	5	
	上行結腸	10	
	横行結腸	4	
	下行結腸	2	
	S状結腸	5	
	直腸	22	
	FAP	2	
転移巣	肝	26	
	肺	26	
	リンパ節	11	
	骨	8	
	腹膜播種	8	
	局所	2	
	その他	5	
先行化学療法	あり(5-FU系)	18	
	なし	32	

患者のQOLの向上をもたらすことが期待されている。

今回われわれは，*l*-LV/5-FU療法を行った切除不能，転移・再発大腸癌症例50例について，先行化学療法の有無の治療効果への影響，標準治療を行った群と，患者の全身状態，副作用などの理由により投与量を減量した群の治療効果の比較，および同治療の安全性・外来化学療法の認容性についてretrospectiveに検討を行った。

I．対象および方法

1．対象（表1）

対象症例は，1999年12月から2003年12月までに東北大学加齢医学研究所付属病院および東北大学病院腫瘍内科において，*l*-LV/5-FU療法が施行された切除不能もしくは転移・再発大腸癌症例50例である。内訳は，年齢は30～79歳（中央値59歳），男性26例・女性24例，原発巣の残存例が12例・切除例が38例で，いずれもPS 0～2の化学療法可能と判断された症例であった。原発巣別，転移巣別症例数は表1のとおりで，主な転移巣は肺26例，肝26例，リンパ節11例，腹膜播種8例，骨8例であった。また，先行化学療法を受けている症例は18例で，いずれも5-FU系の薬剤が使用されており，うち3例はアジュバント療法として*l*-LV/5-FU療法が行われていた。

2．方法（表2）

投与法は，*l*-LV 250 mg/m²を生理食塩水500 m*l*に溶解し2時間で点滴静注し，開始1時間後に5-FU 600 mg/m²の静注を6週投与2週休薬する保険適応投与法を標準とした。

表2 投与方法

投与方法	例数
5-FU 600 mg/m² *l*-LV 250 mg/m²週1回投与	21
投与量・投与法の変更により減量投与 (例: 隔週投与 3W3休/m²→/bodyなど)	29

平均投与回数	12.9回
投与回数	例数
1～5	9
6～10	16
11～20	15
21～	10
入院での平均投与回数	2.6回

標準的投与法を行った症例は21例，3週投与3週休薬などの投与間隔の延長や5-FUを600 mg/bodyとするなど1回投与量の減量など，何らかの減量投与した例は29例であった。平均投与回数は12.9回で，投与回数の内訳は表3のとおりである。

これらのうち入院にて行った回数の平均は2.6回で，ほとんどの投与が外来で行われた。

腫瘍縮小効果は，固形がんの治療効果判定のための新ガイドライン（RECISTガイドライン）[7]に従い判定した。

II．結　果

1．治療効果（表3，4）

評価可能症例数は47例で，CRは認められず，PR 11例，SD 28例，PD 8例であった。全評価症例の奏効率は23.4%，SD症例まで含めた病勢コントロールの可能な例は82.9%であった。前治療の有無で分けると，奏効率では，前治療のない群では27.6%で，前治療のある群は16.7%と，前治療のない群でより高い奏効率を示したが，SD症例も含めた病勢コントロール可能例は，前治療のない群で82.8%，前治療のある群でも83.3%で，両群間に差を認めなかった。MSTでは，前治療のある群では10か月，前治療のない群では12か月で，前治療のない群のほうが長期であった。

また，標準的な治療を行った群と，全身状態，副作用などにより投与間隔の延長や1回投与量の減量など，何らかの減量投与した群での治療効果の比較は表4に示したとおりである。減量投与を行った群の奏効率は14.8%で，標準的投与法を行った群の36.8%の4割程度の奏効率となったが，SD症例まで含めた病勢コントロール可能例は，減量投与した群でも90%以上となり高い治療効果を認めた。

表 3　臨床効果・奏効率

臨床効果	例数	%
CR	0	0.0
PR	11	23.4
SD	28	59.5
PD	8	17.0
計（評価可能症例）	47	100.0

	CR＋PR		CR＋PR＋SD		MST
	例数	%	例数	%	
全評価症例	11/47	23.4	39/47	80.9	11.5か月
前治療あり（全身化学療法）	3/18	16.7	15/18	83.3	10か月
前治療なし	8/29	27.6	24/29	82.8	12か月

表 4　標準投与群と減量投与群の比較

投与法別		標準投与群	減量投与群
年齢中央値		59	59
効果	CR＋PR	7/19（36.8%）	4/27（14.8%）
	CR＋PR＋SD	14/19（73.8%）	25/27（92.6%）
MST		8か月	12か月
平均投与回数		12.5	13.6
先行化学療法		10/21（47.6%）	8/27（29.6%）
原発残存		6/21（28.6%）	5/27（18.5%）
転移巣	肝	12	12
	肺	10	15
	リンパ節	3	7
	骨	3	3
	腹膜播種	2	5

表 5　有害反応

副作用	発現例数	grade			
		1	2	3	4
白血球減少	10	1	8	1	0
好中球減少	4	1	0	2	1
血小板減少	1	0	1	0	0
嘔気・嘔吐	5	1	2	2	0
食欲不振	1	1	0	0	0
下痢	10	5	2	3	0
手足皮膚反応	5	0	5	0	0
腎不全	1	0	0	1	0
倦怠感	5	4	1	0	0
色素沈着	1	0	1	0	0

また，標準的投与法を行った群のMSTは8か月で，減量投与した群でのMSTは12か月で，減量投与した群のMSTのほうが長期であった。追跡可能な全症例のMSTは11.5か月であった。両群の累積生存率はKaplan-Meier法にて図1に示したとおりで，減量投与群が標準投与群に比べて高い効果を示した。

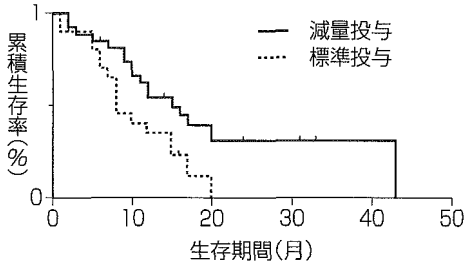


図 1　生 存 率

2．有害事象（表5）

grade 3以上の有害反応として，白血球減少が1例（2.2%），好中球減少が3例(6.5%)，嘔気・嘔吐が2例（4.3%），下痢が3例（6.5%），腎不全が1例（2.2%）認められた。その他，手足皮膚反応，倦怠感，食欲不振，血小板減少，色素沈着も認められたが，いずれもgrade 1～2と軽微であった。全症例を通じ副作用に対してはG-CSFや制吐剤，止痢剤の投与など，および5-FU投与量の減量，投与間隔の延長などの投与法の変更により対処可能であり，治療継続は可能であった。本治療での治療関連死はなかった。

III．考　察

近年，患者のQOLの向上のため外来化学療法が推奨されており，臨床効果を維持しつつ毒性の軽減を図り，外来においても安全に施行できる投与法を検討することが重要になっている。

本邦において*l*-LV/5-FU療法は切除不能・再発大腸癌の標準的化学療法とされてきた。しかし，実際の化学療法の現場では，副作用により休薬や治療の中断，入院を余儀なくされるケースも多い。よって，治療効果を維持しつつ副作用を軽減し，投与の継続性を高めることは今後の重要な目標である。

今回の検討において*l*-LV/5-FU療法は，副作用の発

現した場合も投与間隔の延長・投与量の減量により外来治療の継続が可能であった。

長谷川らの3週投与1週休薬と6週投与2週休薬を比較した検討では，3週投与1週休薬では奏効率が0%で，6週投与2週休薬の23.5%に比べて治療効果が劣るとされた[8]。

本研究においても，減量投与を行った群では奏効率は低下していた。しかしながら，SD症例まで含めた腫瘍制御率，またMSTについては，むしろ減量投与を行った群のほうが良好な結果を示した。これは，表4に示したように標準投与群にPD症例が多かったこと，先行化学療法を行っている例が多かったことが影響している可能性がある。しかしながら，本研究における減量投与群の良好なMSTは，本治療が副作用，全身状態などから投与間隔の延長や1回投与量の減量などを行い，減量投与となったとしても標準投与と遜色のない治療効果があることを示しており，治療を継続することで病勢の進行を制御でき，予後を改善すると考えられる。

LV/5-FU療法では5-FU投与法に急速静注（bolus）と点滴静注（infusional）があり，LV投与法には高用量と低用量がある[9]。本邦ではLV高用量に5-FU急速静注のみが保険適応とされてきた。しかし，LV高用量では5-FU点滴静注がやや良好という成績が示されており[10]，毒性に関しても5-FU急速静注では好中球減少の発症頻度が高いのに対し点滴静注では低く，手足皮膚反応が多く認められるもののその他の非血液毒性に差はなく，安全性においても点滴静注のほうが優れていることが示唆され[11]，そのため本年より5-FUの点滴静注も認められた。

近年，大腸癌の化学療法は大きく進歩している。まず，irinotecan hydrochloride（CPT-11）の導入により転移性大腸癌の治療成績は改善した。*l*-LV/5-FU+CPT-11併用療法は，5-FUの急速静注でも[12]点滴静注でも[13]*l*-LV/5-FU療法よりも延命効果を認め，現在わが国において行える大腸癌に対して最も有効な化学療法と考えられる。

また，欧米では第三世代の新規プラチナ誘導体であるoxaliplatin（Oxa）の臨床試験が進み，*l*-LV/5-FUとの併用療法が標準療法と考えられるようになってきた[14]。OxaはCDDPと同様DNAに対してアルキル化剤として作用し，少なくとも*in vitro*ではCDDP，carboplatin（CBDCA）とは交差耐性を示さず[15-18]，腎毒性がCDDPより低く血液毒性はCBDCAより低い。*l*-LV/5-FUとOxaの併用療法FOLFOXは奏効率40～50%，MST 17～20か月という優れた効果を示しており[19-21]，わが国でも2005年4月に臨床導入された。

分子標的薬の導入もひかえ，さらなる転移性大腸癌の治療の進歩が期待されているところである。

なお，本研究結果は第37回制癌剤適応研究会にて発表した。

文　献

1) 相羽惠介: 化学療法．癌と化学療法 **31**(5): 706-711, 2004.
2) 相羽惠介: フッ化ピリミジン系薬剤およびプリン代謝拮抗剤．V．各薬剤に特異的な耐性機構とその克服．癌の薬剤耐性とその克服―基礎と臨床―(大沼尚夫，竹村譲・編)，宇宙堂八木書店，東京，2001，pp 211-221.
3) 植竹宏之，樋口哲郎，榎本雅之・他: 大腸癌に対する化学療法．癌と化学療法 **30**(12): 1889-1894, 2003.
4) O'Connell MJ: A phase III trial of 5-fluorouracil and leucovorin in the treatment of advanced colorectal cancer. A Mayo Clinic/North Central Cancer Treatment Group study. *Cancer* **63**(6 *suppl*): 1026-1030, 1989.
5) Petrelli N, Herrera L, Rustum Y, *et al*: A prospective randomized trial of 5-fluorouracil versus 5-fluorouracil and high-dose leucovorin versus 5-fluorouracil and methotrexate in previously untreated patients with advanced colorectal carcinoma. *J Clin Oncol* **5**(10): 1559-1565, 1987.
6) De Gramont A, Krulik M, Cady J, *et al*: High-dose folinic acid and 5-fluorouracil bolus and continuous infusion in advanced colorectal cancer. *Eur J Cancer Clin Oncol* **24**(9): 1499-1503, 1988.
7) Therasse P, Arbuck SG, Eisenhauer EA, *et al*: New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* **92**(3): 205-216, 2000.
8) 長谷川伸一，赤池　信，山本裕司・他: *l*-leucovorin, 5-FU併用療法の進行結腸・直腸癌に対する投与法別の有用性に関する検討．癌と化学療法 **31**(5): 729-733, 2004.
9) Jager E, Heike M, Bernhard H, *et al*: Weekly high-dose leucovorin versus low-dose leucovorin combined with fluorouracil in advanced colorectal cancer: results of a randomized multicenter trial. Study Group for Palliative Treatment of Metastatic Colorectal Cancer Study Protocol 1. *J Clin Oncol* **14**(8): 2274-2279, 1996.
10) Efficacy of intravenous continuous infusion of fluorouracil compared with bolus administration in advanced colorectal cancer. Meta-analysis Group In Cancer. *J Clin Oncol* **16**(1): 301-308, 1998.
11) Toxicity of fluorouracil in patients with advanced colorectal cancer: effect of administration schedule and prognostic factors. Meta-Analysis Group In Cancer. *J Clin Oncol* **16**(11): 3537-3541, 1998.
12) Saltz LB, Cox JV, Blanke C, *et al*: Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. Irinotecan Study Group. *N Engl J Med* **28**: 343(13): 905-914, 2000.
13) Douillard JY, Cunningham D, Roth AD, *et al*: Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. *Lancet* **25**: 355(9209): 1041-1047, 2000.
14) de Gramont A, Bosset JF, Milan C, *et al*: Randomized trial comparing monthly low-dose leucovorin and