BIOMARKERS, VOL. 8, NOS. 3–4 (MAY–AUGUST 2003), 311–332

Taylor & Francis
healthsciences

In situ quantification of aberrant p53 in colorectal neoplasia

STEVEN JAY SMITH[1]*, ALFRED NEUGUT[2], DANIEL HEITJAN[3], KENNETH FORDE[4], PETER HOLT[5], REGINA M. SANTELLA[6], LUO JIIN-CHYUAN[7], WALTER CARNEY[8], LLEWELYN WARD[9] and PAUL W. BRANDT-RAUF[6]

[1] VA-New York Harbor Health Care System, Brooklyn, New York, USA
[2] Departments of Medicine and Epidemiology, New York Presbyterian Medical Center, New York, USA
[3] Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, USA
[4] Department of Surgery, New York Presbyterian Medical Center, New York, USA
[5] Departments of Surgical Pathology and Gastroenterology, St Luke's-Roosevelt Hospital, New York, USA
[6] Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, USA
[7] Department of Medicine, Taiwan University, Taiwan, Republic of China
[8] Oncogene Science Diagnostics, Bayer Corporation, Tarrytown, New York, USA
[9] Department of Renal Pathology, New York Presbyterian Medical Center, New York, USA

Received 10 January 2003, revised form accepted 24 April 2003

Aberrant p53 protein accumulation was measured immunohistologically in 342 colorectal paraffin-embedded tissue sections from 115 patients (24 with adenocarcinoma, 59 with adenoma and 32 'hospital controls'). Subjective scoring was compared with quantitative cell imaging, including dichotomous ($p53^+/p53^-$) status, ng $p53^{mut}$ mg^{-1} enterocyte protein, and tumour burden and patient body 'burden' of aberrant p53. A total of 62.5% cancer patients, 23.7% adenoma patients and 3.1% hospital controls were accorded $p53^+$ status on the basis of p53 quantification. Quantitative $p53^+/p53^-$ assignment had a stronger inverse association with survival ($\chi^2 = 6.17$, $p = 0.013$, Kaplan–Meier test) than subjective 'visual estimation' ($\chi^2 = 0.57$, $p = 0.449$). There was a strong inverse relationship between the p53 'body burden' and the months of post-diagnosis survival (hazard ratio = 1.42, $p = 0.0004$, Cox proportional hazards). Absolute quantification for inactivated p53 permits objective and reproducible scoring, adjusts for intra-laboratory immunostaining 'batch effects', corrects for fixation artefacts, and standardizes for inter-laboratory differences in fixation, antibody selection and staining method. Clinically, *in situ* quantification of p53 will permit more accurate survival prognoses and will inform therapy selection and dose. Ultimately, accurate quantitative tissue/blood p53 correlations may provide a minimally invasive and systemic surrogate measure for these same clinical purposes.

Keywords: immunohistochemistry, p53, cell imaging densitometry, colorectal, protein quantitation, tumour/tumour tissue.

* Corresponding author: Steven J. Smith, VA-New York Harbor Health Care System, Room 3-210, 800 Poly Place, Brooklyn, NY 11209, USA. Tel: (+1) 718 439 4162; e-mail: smithstevenjay@netscape.net

Biomarkers ISSN 1354-750X print/ISSN 1366-5804 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/1354750031000138676

Introduction

Aberrant p53 in colorectal tumours

Sporadic adenoma and adenocarcinomas represent over 90% of all dysplastic colorectal polyps and cancers. Disabling the normally evanescent p53 wild-type protein through mutation and/or deletion in colorectal patients is associated with proliferation of dysplastic cells, gene heteroploidy and chromosomal instability, preferential survival of hypoxic-resistant diseased enterocytes, tumour angiogenesis, metastasis, multi-drug resistance and unresponsiveness to alkylation chemotherapy (Chin *et al*. 1992, Symonds *et al*. 1994, Harris *et al*. 1996, Vogelstein *et al*. 2000). Strong, multiple correlations between immunostained aberrant p53 protein on the one hand and molecular/clinical associations on the other suggests p53 immunohistochemistry (IHC) is a versatile and biologically valid biomarker for a wide range of inactivating p53 mutations. The average p53 mutation rate among colorectal adenocarcinomas, as estimated from DNA sequencing studies, is approximately 70%; loss of the remaining p53 wild-type allele – loss of homozygosity (LOH) – occurs in 75–83% of patients with a p53 mutation. Furthermore, the percentage LOH is proportional to the percentage of cancer cells that immunostain positive for aberrant p53 protein.

Adenomatous colorectal polyps are, typically, the precursor for adenocarcinoma. Adenomas lacking mutant p53 cells manifest 6–9% LOH; those with cells expressing mutant p53 ($p53^{mut}$) display up to 67% LOH. The pattern of correlation between tumour progression and prevalence rates of p53 inactivation/mutation in human studies is corroborated by animal research, and suggests the role of aberrant p53 role is to promote the progression of late stage dysplasia to cancer – as do tumour size, non-tubular histology and sessile configuration (Schottenfeld and Winawer 1982), risk factors all associated with loss of p53 function. Regardless of the attendant risks, the percentage of p53-positive ($p53^{+}$) immunostained cells is typically $\leq 10\%$, arrayed as discrete, contiguous foci. In colorectal adenocarcinomas the probability of metastasis, recurrence and survival are functions of the stage, grade, age of the patient and p53 status – not of tumour size *per se* (Fidler 1997). An aberrant p53 IHC focus in the interior of an adenoma identifies the subset of 'aggressive' dysplastic enterocytes that are in transition to cancer. Among adenocarcinoma patients the presence of a $p53^{+}$ cell population within the primary cancer is predictive of liver or intestinal lymph node metastasis; a high proportion of the metastatic cells are also $p53^{+}$ (Kimura *et al.* 1996, Sory *et al*. 1997). Aberrant p53 asserts its effects independently of the tumour stage and grade (Campo *et al.* 1991). Although the site of the primary tumour can influence the metastatic pathway, this does not confound the influence of p53 on survival, since there is no locational preference for p53 inactivation.

If the relative affinities and quantities of interacting proteins affect the divergent signal transduction kinetics and pathways of homoeostasis or disease, the measurement of disease-related proteins with accurate (IHC) quantification should aid in making diagnoses and assessing prognoses. This appears to be the case for mutated p53 and colorectal dysplasia:

i. Although it is true that the identity of the mutated p53 codon determines whether *in vitro* gene inactivation will be recessive or 'dominant negative', all p53 point mutations result in a single, cell-transforming phenotype (Ozbun and Butel 1997).
ii. *In vitro* and *in vivo* data suggest that it is elevated translation of mutant p53 protein, rather than DNA amplification or increased mRNA transcription, that leads to its intracellular accumulation.
iii. Inactivation of p53 is clinically detectable. Colonoscopy is sufficiently sensitive and accurate to be able to detect and remove intestinal tumours as small as 1–2 mm in diameter, and the aberrant p53 cell population is likely be demonstrated using IHC, since the percentage of $p53^+$ cancerous cells often exceeds 50% and nearly every *p53* mutation increases the stability of the protein, prolonging its half-life 21- to 22-fold (Kraiss *et al.* 1991) by interfering with its normal regulatory, degradational binding by mouse double minute-2 (MDM-2) protein (Lane and Hall 1997).
iv. Of the wide variety of reported p53 missense point mutations, 87% occur within its DNA-binding domain (exons 5–8), making the observed p53 accumulation a marker of lost tumour-suppression function.
v. Aberrant p53 protein accumulation is proportional to the loss of function (Levine and Momand 1990, Tominaga *et al.* 1992, Sharma *et al*. 1993, Lowe *et al.* 1994, Soussi *et al*. 1994).
vi. Despite the varying affinities and avidities of particular anti-p53 antibodies, IHC is highly specific. There is $>80\%$ concordance within a given tissue sample between aberrant p53 IHC and $p53^{mut}$ DNA status for breast (Van der Kooy *et al*. 1996), lung and colon/rectum (Soussi *et al*. 1994) cancers.
vii. Since p53 inactivation appears to play specific roles in cancer progression (Thomlinson and Gray 1955, Vogelstein *et al.* 1988, Kemp *et al.* 1993, Vaupel and Hockel 1995, Graeber *et al*. 1996, El-Diery 1998, Hendrix 2000), IHC should display histological specificity.

IHC of p53

Given the above experimental and clinical evidence, we were surprised to find highly diverse results regarding colorectal p53 IHC in the literature. There was a wide divergence in the reported percentage of patients diagnosed as $p53^+$, being 47–80% (mean 61%) among cancer studies and 3.3–25% among studies of adenoma. Furthermore, detection of p53 using IHC in adenocarcinoma has not been consistently prognostic. A review by Dowell and Hall (1995) of colorectal cancer patients concluded that only research projects with a large study population had any likelihood of finding an association between p53 loss and survival, yet our own review of 29 large, multivariate IHC studies, with lengthy follow-up, revealed that only 18 had found p53 immunostaining to be independently prognostic.

Due to such inconsistency, p53 IHC has failed to gain widespread acceptance for the routine evaluation of patients with colorectal or other cancers (American Society of Clinical Oncology 1996). Varying IHC conditions lead to non-standardized scoring of the immunostaining within and between clinical labora-

tories, making longitudinal and inter-laboratory studies problematic (Adams *et al*. 1999). Selection of p53 antibodies with different affinities and specificities (Gannon *et al.* 1990, Baas *et al.* 1994, Soussi *et al*. 1994, Vojtešek *et al.* 1995, Bonsing *et al.* 1997) requires internal staining standards. Fixative choice and use will differentially affect antigen 'masking', the intracellular location of the staining (Fisher *et al.* 1994), tissue morphology preservation and the tissue penetration rate. Antigen masking (loss) is roughly proportional to fixation duration. Simultaneous fixation of tissue of different types or densities, combining resected and biopsy tissue or even tissue and cultured cells, causes inconsistent and incommensurate preservation of antigenicity and morphology. Reliance on arbitrary 'positivity' thresholds, lack of antigen reference standards, subjective and irreproducible scoring of p53 (McShane *et al*. 2000) and other cancer antigens (O'Leary and Calvin 1998) present other obstacles. Overall, current immunostaining scoring methods result in incommensurable measurements, whether comparisons are made over time, between laboratories, between tumour proteins, or between a given patient's tumour burden and the corresponding blood concentration. This study investigates the use of an accurate, standardized, quantitative IHC method to measure aberrant p53 in colorectal neoplasia.

Methods

Study population

The study population was drawn from patients who underwent colorectal colonoscopy at the Columbia-Presbyterian Medical Center (CPMC) and the St Luke's-Roosevelt Medical Center in New York City between October 1990 and March 1993. The CPMC study subjects had newly diagnosed colorectal adenocarcinomas or adenomas, no prior colorectal cancer, and no prior or current ulcerative colitis. The signed patient consent form requested blood samples, questionnaire information, access to their medical records, and permission to analyse their blood and any surgically removed colorectal tissue. The St Luke's-Roosevelt Hospital provided non-dysplastic archived tissue biopsies from gastroenterology patients with typical non-dysplastic conditions such as diverticulosis. These 'hospital controls' had no prior colorectal dysplasia, and no current or past ulcerative colitis. In addition to gastroenterological and oncological investigations, blood chemistry and radiology reports from the two main hospitals, follow-up data sources included various state and hospital tumour registries, death certificates, and phone interviews. Information was available on age, gender, race, medical history, size of lesion and, for the cancers, degree of cellular differentiation (grade) and tumour penetration (stage); both hospitals employed a modified Duke's scoring system.

The study population consisted of 24 colorectal adenocarcinoma patients, 59 colorectal adenoma patients and 32 colonoscopy controls for whom there was analysable tissue. The average ages for these groups were 65.7, 62.5 and 59.8 years, respectively, which reflects the characteristic age-specific risks for these conditions. The percentages of female patients were also typical, being 59%, 49% and 57%, respectively. The percentages of Caucasians were 92% and 89% in the adenocarcinoma and adenoma patients, respectively, but this data was not available for the controls. The 115 patients had a total of 161 tumours, providing 182 fixed, paraffin-embedded tissue blocks comprising 342 histological samples of adenocarcinomatous, adenomatous, hyperplastic and normal tissue. After using several thousand sectioned and immunostained cultured cells and colonoscopy control cells to define a $p53^+$-staining threshold, approximately 16 000 $p53^+$ cells were quantified for aberrant p53.

The distribution of the tumours in this study according to adenoma dysplastic grade, histological type (tubular, villous, tubulovillous) and size was typical of colorectal polyp patients. The distribution of stage and grade of the cancers was also typical, although stage D cancers were slightly under-represented. Cancer-containing adenomas were much larger and much less likely to have tubular histology than purely adenomatous lesions, whether or not the dysplastic patient had a synchronous additional cancer.

Clinical follow-up information until March 1997 was available for 23 of the 24 (96%) cancer cases, 45 of the 59 (76%) adenoma cases, and 10 of the 32 (31%) colonoscopy controls. In order to avoid possible selection bias, follow-up analysis was limited to the cancer subjects. The median follow-up period for the cancer cases was 42 months (range 6.5–76 months). The criteria for deducing underlying cause of death were serum concentrations of liver enzymes and carcinoembryonic antigen (CEA),

computed tomography (CT), ultrasound and/or X-ray imaging. Attempted patient follow-up began with patient's surgery (earliest date November 1990) until March 1997. Follow-up endpoints were recurrence of colorectal adenoma or adenocarcinoma, progression from adenoma to cancer, and death attributable to colorectal cancer. Months of survival duration were tallied until the patient was lost to follow-up or died.

Scoring of tissue

Computerized cell image densitometry (CAS 200; Cell Analysis Systems, Elmhurst, Illinois, USA) was used to analyse immunostained tissue sections and immunostained cultured calibration cells, both stained with brown antigen stain and cyan nuclear counterstain. Paired templates from haematoxylin (blue) and eosin (pink) (H&E) histology tissue sections were used to segregate the constituent regions (or 'tissue classes') in the sample according to their respective histopathology, cell morphology and tissue architecture. Picograms of aberrant p53 per patient (the 'body burden') were compared with patient-matched presurgical plasma concentrations of aberrant p53 in a subset of mutant p53 cancer and adenoma cases, measured using a mutant-specific p53 enzyme-linked immunosorbent assay (ELISA) (QAIA03; Oncogene Research Products, Cambridge, Massachusetts, USA) that has been shown to give reliable results on plasma/sera concentrations (Greco *et al*. 1994, Luo *et al.* 1995) and to reveal declining p53 blood levels among colon (Shim *et al*. 1998) and breast (Rosanelli *et al*. 1993) cancer patients soon after surgical removal of their tumour(s).

Both the tissue samples and the cultured 'calibration' cells were formalin-fixed and paraffin-embedded. Loss of immunoreactivity to the anti-p53 pantropic IHC antibody (DO-1), sustained by both the calibration cells and the tissue during fixation/embedding, was largely recovered in their respective 5 μm tissue sections by 'antigen retrieval' in heated citrate buffer. Inclusion of similarly treated calibration cells allows the IHC tissue scoring to be standardized within or between laboratories for variations in factors such as antibodies, reagents, incubation times and antigen retrieval methods. Since a preferable, previously established clinical standard (Galen and Gambino 1975) to define $p53^+$ IHC is lacking, a statistical criterion of the upper 2.5% of the log-normalized frequency distribution of the average optical density (OD)/cell was used. After segregating a tissue section according to its constituent H&E-defined histological classes, the $p53^+$ area was measured in any tissues containing cell foci exceeding this cut-off. Previously determined, and differing, concentrations of aberrant p53 in multiple calibration tumour cell lines were used to translate summed OD pixels of immunoreactive p53 in the tissue sections into absolute quantities of aberrant p53 protein.

The calibration cells were used to create batch-specific standard curves, which allowed us to correct for batch effect differences in tissue staining intensity. Furthermore, aggregated frequency distributions of cell staining intensity in both histologically normal tissue (hospital controls) and cultured cells containing homozygous wild-type normal p53 made it possible to provide a common statistical criterion for the definition of $p53^+$. In order to be certain the cultivated control cells were valid tissue staining standards, we subjected them to comparable, but not identical, tissue processing and immunostaining conditions. Optimized IHC parameters for the cells and the tissue were experimentally (and separately) determined according to five criteria:

i. equal ΣOD/cell signal to noise ratios for $p53^+$ and $p53^-$ nuclei
ii. an identical $p53^+$ cut-off value, using the same criterion (mean+2SD)
iii. similar (dynamic) ranges of ΣOD/cell expression among $p53^+$ nuclei
iv. equal cell sizes and nuclear/cytoplasmic ratios
v. normal (Gaussian) ΣOD/cell frequency distributions for each calibration cell line and tissue class.

Tissue and calibration cells were tested for staining reproducibility. Multiple tissue sections were 'bread-loafed' from tissue blocks to verify similar levels of p53 inactivation throughout. Based on these results, we concluded that the fixed, paraffin-embedded calibration cells were valid and effective quantitative surrogates for the fixed, paraffin-embedded tissue blocks.

The following quantities of p53 were measured: ng p53 mg^{-1} total enterocyte protein, number of p53 molecules per cell, pg p53 per tumour (p53 tumour burden) and pg p53 per patient (p53 body burden). The intensity of aberrant p53 expression per cell was measured using cell imaging densitometry (using the cell measurement program) and the calibration curve was used to derive ng p53 mg^{-1} total enterocyte protein. The extent of aberrant p53 expression was measured directly (using the micrometer program), and the $p53^+$ nuclear area per lesion was converted to the nuclear volume (mm^3) per lesion. The two measures were multiplied to derive pg p53 per lesion. This is a biologically specific measure: it is limited to the affected region of diseased cells, for example the nuclei of the enterocyte cell population within the intestinal crypts and segregated by histological tissue class. It is also a highly stable metric, being unaffected by unpredictable and varying dilution effects from the inclusion of irrelevant cellular

Table 1. Subjective scoring of the extent and intensity of aberrant p53 immunostaining.

Scoring	Criteria
+++	>50% of cells intensely stained
++	10–50% of cells intensely stained
+	1–10% of cells intensely stained, or 10–50% of cells moderately stained
+/−	1–10% of cells moderately stained, or 10–50% of cells lightly stained
−	1–10% of cells lightly stained, or absence of any staining

and stromal material. One or more histological lesions comprised a tumour, and one or more tumour burdens equalled the patient's $p53^+$ body burden.

Prior to quantifying p53 tissue immunostaining, all the samples were scored subjectively (visual estimation). The tissue sections were assigned ordinal ranks according to explicit levels and percentages of immunostaining intensity/extent (Table 1). This made it possible to validate p53 quantification by testing the overall concordance between it and the subjective rankings using Spearman's rank-order correlation (ρ), but also highlighted the deficiency of subjective scoring with respect to its wide and overlapping scoring ranges (data not shown). For the purposes of comparing the subjective versus the quantitative tissue scoring with respect to patient-matched p53 plasma concentrations, the continuous quantitative tissue and plasma scores were reduced to comparable binary categories. *Cohen's kappa* (κ) coefficient statistic was used to test for correlations between tissue and plasma scores [−0.0–1.0 range, correlation] $\kappa > 0.50$ considered a strong true (observed-expected) (Cohen J.A., 1960). The p53 visual estimation and quantitative evaluations were compared with respect to tissue histology, patient diagnosis and follow-up. A third, objective, purely densitometric measure termed p53 positivity ($p53^+$ nuclear area × average OD/cell) was created in order to evaluate the subjective ordinal ranks, independently of our absolute quantification method.

The dichotomous inactivation status ($p53^+$ or $p53^-$) as determined by subjective or quantitative p53 scoring of the cancer patients was compared with patient survival (alive or dead) over the follow-up period using the Fisher-Irwin exact test. When dichotomizing the visual estimation scores, the ambiguous '$p53^{+/-}$' staining rank was grouped with the '$p53^+$' rank. The more sensitive Kaplan–Meier test was used to compare the relative ability of the two methods to contrast cumulative survival duration for $p53^+$ versus $p53^-$ cancer patients. Absolute quantification of p53 permitted testing for a dose–response effect between the degree of p53 inactivation and the survival duration using Cox's *Proportionate Hazards* method.

Results

Aberrant p53 diagnosis

$p53^+$ immunostaining was intensely brown, entirely nuclear, strictly limited to the enterocytes of the intestinal crypts, and focal (rather than 'mosaic') in its staining pattern. On average, <9% of the adenoma nuclear area was $p53^+$, and often less than 1%. The left-hand panel of Figure 1 (inset) shows the immunohistological staining of a typical 1 cm tubular adenoma tissue block: the lower half of the inset shows staining with H&E and the upper half is the next 5 μm tissue section taken from the same block, but immunostained for p53. Aberrant p53, typically, first develops in the base of the crypt and then spreads upward within the crypt toward the intestinal lumen. Combining H&E with IHC defines p53 inactivation in terms of the stage of disease progression and illustrates the clonal nature of that inactivation, that is, one or more small contiguous 'islands' are seen within the 'sea' of the overall dysplasia. The right-hand panel of Figure 1 shows the multifocal expression of aberrant p53 within adenomas and the change that occurs as they first become transitional adenoma/adenocarcinoma hybrids and then progressively invasive cancers: due to either the dominance of one $p53^{mut}$ clone or to the coalescence of multiple clones, the number of $p53^+$ foci decrease.

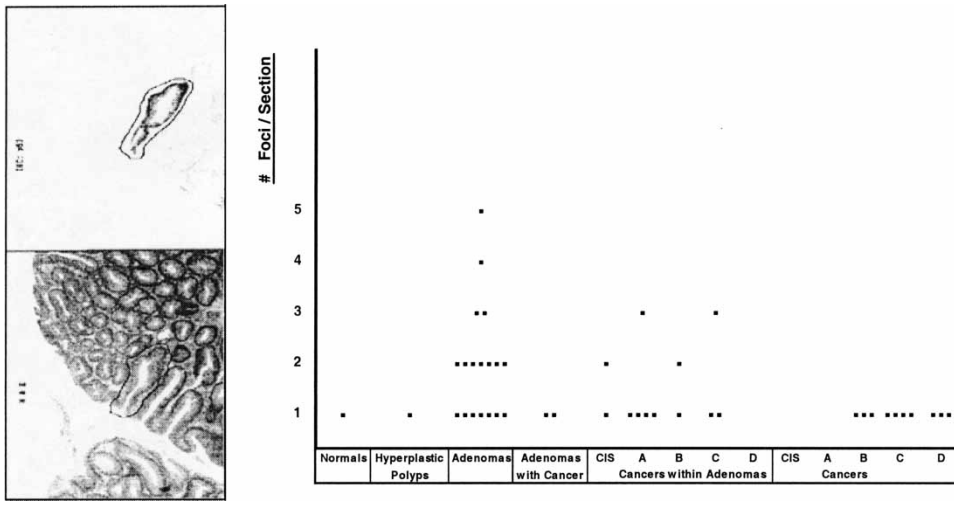

Figure 1. Left-hand panel: Typical 1 cm tubular adenoma tissue section stained with H&E (lower half) and the next 5 μm section from the same block immunostained for p53 (upper half). Right-hand panel: Number of $p53^{+}$ foci per tissue block by histological class. CIS, carcinoma in-situ.

In either case, the normal, wild-type p53 neighbouring cells are progressively displaced. There was no association within the study population between the location of the cancer within the large intestine/rectum (proximal versus distal) and the $p53^{+}$ status ($p = 0.417$, Fisher-Irwin exact test).

The mean aberrant $p53^{mut}$ protein concentration among 13,357 $p53^{+}$ cancerous enterocytes from the cancer patient cohort was 5.5 ng mg^{-1}. The mean concentration of aberrant p53 among the 7854 $p53^{+}$ adenomatous cells scored in all those with one or more adenomas was slightly lower at 3.91 ng mg^{-1} adenomatous enterocyte protein. Among the 1563 $p53^{+}$ adenomatous cells scored within transitional adenomas containing an interior cancer focus, the mean p53 concentration was the greatest of the three tissue classes at 6.81 ng mg^{-1} total adenomatous crypt protein.

The average amounts of aberrant p53 in the specific histological classes within the tumours are shown in Figure 2. The molar measure of the aberrant protein, that is, p53 molecules per cell, shows the same peaking in intensity of expression in the adenomatous cells progressing to cancer (adenoma containing/with cancer) as did the ng mg^{-1} measure of p53 concentration. Note, however, the tremendous increase in accumulated total aberrant p53 in the cancerous lesions from the proliferation of the 'aggressive fraction' of neoplastic cells lacking any tumour suppression: the immunostained volume for the typical cancer was 44 times greater than that for the average adenoma. For cancer tissue the number of $p53^{+}$ cells is far more significant than the degree of p53 inactivation per cell.

If the loss of p53 tumour suppressor function is a marker of the 'aggressive fraction' within the dysplastic polyp, there may be a threshold level of p53 inactivation that facilitates the progression to malignancy. Even though larger adenomas are already known to correlate with synchronous cancers, this study revealed that $p53^{+}$ adenomas having a tumour burden of less than 1 ng of $p53^{mut}$

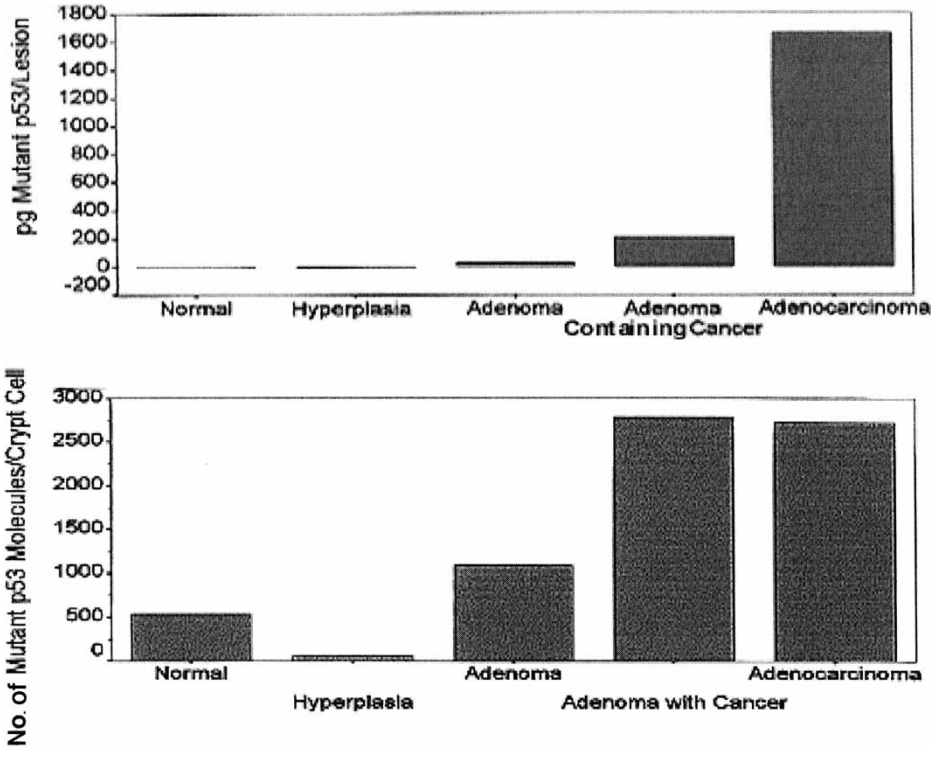

Figure 2. Total inactivated p53 per lesion. $n = 161$ for tumours; $n = 35$ for normal tissue.

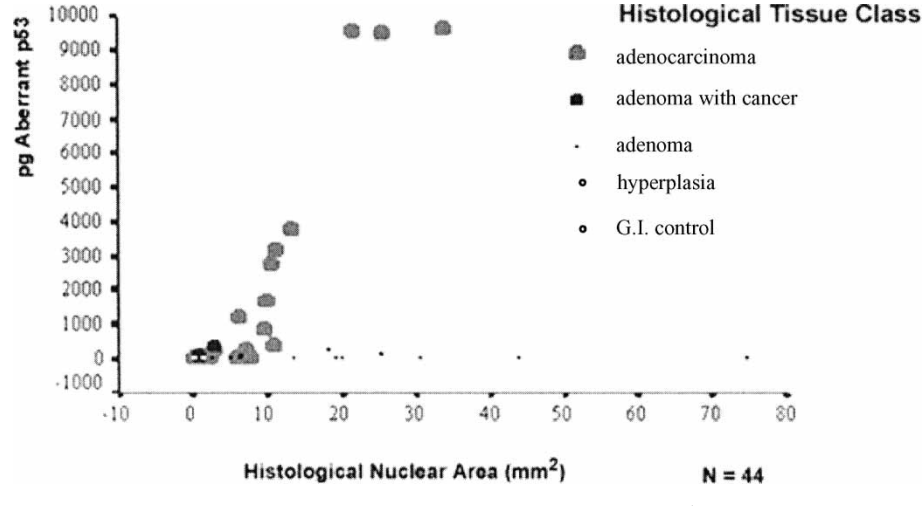

Figure 3. Quantity of aberrant p53 versus histological nuclear area in $p53^+$ colorectal tissue samples, showing that the amount of aberrant p53, rather than adenoma size, determines progression to adenocarcinoma.

had dysplastic cell populations with nuclear areas ≤ 80 mm^2, yet no cancer histology. On the other hand, even among the adenocarcinomas that were far smaller than the adenomas, nearly every such lesion contained > 1 ng of aberrant p53 (Figure 3).

The subjective ordinal ranking of positivity showed a rough overall correlation with the independent cell imaging densitometry measure of positivity. However, these ranks had certain deficiencies: (i) the arbitrary and ambiguous +/− rank for borderline staining; (ii) the substantial overlap in the amount of positivity between '+', '++' and '+++'; and (iii) the '+++' lesions encompassed a nearly eight-fold range of 'positivity'.

p53 absolute quantification was used as a basis to dichotomize the p53 status of each patient. Patients with multiple histological lesions were classified according to their most advanced tumour. Aberrant p53 was detected in 15 out of 24 (62.5%) of the adenocarcinoma patients, 14 out of 59 (23.7%) of the adenoma patients, and one out of 32 (3.1%) of the non-dysplastic control population. If aberrant p53 is treated as a risk factor or exposure and adenoma and adenocarcinoma as disease outcomes in a 2×2 table, the exposure odds ratio (OR) is 5.36 ($p < 0.005$, Yates's corrected χ^2), that is colorectal cancer patients were more than five times more likely to have lost some degree of p53 tumour suppression function than were purely adenomatous patients. For adenoma patients versus the non-dysplastic controls, the aberrant p53 exposure OR was 9.6 ($p = 0.025$), and for adenocarcinoma patients versus the controls the OR was 51.6 ($p < 0.001$).

Besides patient categorization, aberrant p53 status was also scored according to tissue class ($n = 157$ tumours). Overall, 66.7% of the cancers were $p53^+$. No p53 staining was detected in the histologically normal tissue adjacent to the adenocarcinomatous or adenomatous lesions. When these neighbouring normal tissue samples were combined with the paraffin sections from the non-dysplastic control patients, the rate of aberrant p53 in non-dysplastic tissue samples was 0.6% (one out of 168). The $p53^+$ rate among all the hyperplastic 'lesions' contained in tissue sections from all patients was 4.5% (one out of 22), highlighting the important distinction between hyperplasia and dysplasia.

Among the adenocarcinoma tumours, $p53^+$ rates varied somewhat by tumour grade, being 57.1% in well-differentiated tumours, 68.8% in moderately differentiated tumours and 75% in poorly differentiated tumours, but the trend was not statistically significant ($\chi^2 = 0.41$, $p = 0.52$). Similarly, $p53^+$ rates varied slightly by tumour stage, being 66.7% in Duke's A, 57.1% in Duke's B, 62.5% in Duke's C and 100% in Duke's D, but again the differences were not statistically significant ($\chi^2 = 0.07$, $p = 0.78$). Among the adenoma tumours, $p53^+$ was twice as prevalent among those lesions with a villous component (villous or tubulovillous) (33.3%) than among those that were purely tubular (16.7%), although this difference was not statistically significant (OR = 2.5, $p = 0.16$). $p53^+$ incidence increased with increasing adenoma size, being 9% in lesions ≤ 5 mm in diameter, 21% in those 6–10 mm in diameter and 44.4% in those > 10 mm in diameter, a trend that was highly significant ($\chi^2 = 9.7$, $p = 0.002$).

The body burden among the $p53^+$ adenocarcinoma patients ranged from 0 to 9910 pg (mean $\pm$ SD 2515 $\pm$ 3423 pg) and among the $p53^+$ adenoma patients ranged from 0 to 229 pg (mean $\pm$ SD 23 $\pm$ 65 pg). The only $p53^+$ hospital control patient had a very small p53 body burden of 0.2 pg. With respect to tumour grade, the mean p53 body burden was 1140 pg in the well-differentiated cases, 2661 pg in the moderately differentiated cases, and 3957 pg in the poorly differentiated cases.

With respect to tumour stage, the mean p53 body burden was 2583 pg in the Duke's A cases, 1578 pg in the Duke's B cases, 2444 pg in the Duke's C cases, and 3479 pg in the Duke's D cases. In $p53^+$ adenomas the mean p53 body burden was 50 pg in those with any villous component and 4 pg in those that were purely tubular. The mean p53 body burden varied dramatically with increasing adenoma size, being 0.13 pg in lesions ≤ 5 mm in diameter, 0.41 pg in those 6–10 mm in diameter and 34.77 pg in those > 10 mm in diameter.

There was good agreement in the assignment of $p53^+/p53^-$ status to tissue between p53 quantification and visual estimation. When sections scored as +/− on visual estimation were included in with the $p53^+$ sections, all the adenocarcinoma tissue sections accorded $p53^+$ status using the p53 quantification algorithm were also $p53^+$ on visual estimation. However, three adenocarcinomatous sections assigned $p53^-$ status using quantification were designated $p53^+$ on visual estimation (overall: $\kappa = 0.81$, $p < 0.001$). Similarly, for the adenomatous sections, 17 $p53^+$ sections were scored as $p53^+$ according to visual estimation, but 14 sections assigned $p53^-$ status on quantification were designated $p53^+$ by visual estimation ($\kappa = 0.62$, $p < 0.001$). When visual estimation +/− sections were included with the $p53^-$ sections, the concordance between the subjective and quantitative scoring methods among the cancerous tumours did not change, but improved for the adenomatous lesions ($\kappa = 0.90$, $p < 0.001$). There was also good agreement between the ordinal ranking of lesions by visual estimation and p53 quantity, using Spearman's rank-order correlation ($\rho = 0.78$, $p = 0.01$). Once again, this correlation improved if the +/− ordinal rank was grouped together with the $p53^-$ rank ($\rho = 0.90$, $p = 0.01$). Overall, there were no lesions scored as $p53^-$ on visual estimation that were $p53^+$ according to quantification.

When the 16 matched adenoma and adenocarcinoma patient tissue and blood samples were compared according to the two different IHC scoring systems, visual estimation provided a weak but significant correlation with plasma levels ($\kappa = 0.36$, $p = 0.025$). In contrast, the p53 tissue quantification method assigned 12 of the 16 patients concordance in p53 status between their tissue and plasma samples: eight patients ranked as $p53^-$, three cases ranked as + and one case ranked as ++ for both tissue and blood. This represented fairly good agreement ($\kappa = 0.54$) between the p53 body burden and the $p53^{mut}$ plasma concentration levels, with a strong statistical significance ($p = 0.002$).

Patient survival

Increasing patient age is often associated with inferior overall health and/or immune resistance to intestinal dysplasia. If age is also correlated with the p53 status of dysplastic cases and non-dysplastic controls, it could cause a spurious association between p53 inactivation and shortened patient survival. This was not the case, however, since there were no statistically significant differences (Student's t-test) in the average age among the three patient groups. Among the 23 of the 24 colorectal cancer patients who were followed up, the average ages of the $p53^-$ and $p53^+$ subjects were nearly identical, and patient age did not correlate with survival duration ($R^2 = 0.03$, $p = 0.426$).

The adenocarcinoma mortality rate was 48%, which is typical for US hospitals. In order to accommodate the varying duration of follow-up, the Kaplan–Meier cumulative survival and Cox proportionate hazards statistics were used to examine the importance of the known major prognostic risk factors of stage and grade. Univariate Kaplan–Meier analysis demonstrated the expected prognostic relationship for both tumour grade (log rank = 11.54, $p = 0.003$) and tumour stage (log rank = 9.34, $p = 0.025$), according to the expected rank hierarchy. Similarly strong associations in the expected hierarchy were found using univariate Cox's proportionate hazards statistics. The incremental risk of death over time was significant for tumour grade ($p = 0.003$) and stage ($p = 0.015$).

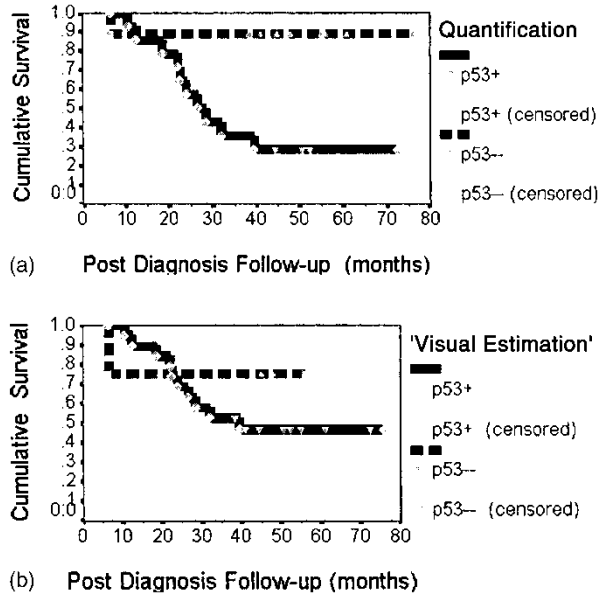

Figure 4. Adenocarcinoma survival curves by aberrant p53 status. (a) p53 quantification: mean survival was 68.3 months for $p53^-$ patients and 37.6 months for $p53^+$ patients (log rank = 6.17, $p = 0.013$). (b) Visual estimation of p53: mean survival was 43.6 months for $p53^-$ patients and 48.2 months for $p53^+$ patients (log rank = 0.57, $p = 0.4493$). Survival analysis, such as Kaplan–Meier and Cox Proportionate Hazards measures 'time to terminal event'. Once the event, in this article, death, has occurred the subject is no longer reflected in the follow-up statistics beyond the point in time that the event occurred. The same is true of 'censored' cases. These are people for whom death has not occurred over the duration that their post diagnostic survival was followed. The points on the graphs that mark a point in time at which they were censored does not necessarily mean the point at which they were lost to follow-up; it marks the follow-up duration at which they were excluded for further calculations of relative cumulative survival duration, which includes those alive and still followed at the end of the study's observation period. Up to the end of their follow-up they are in the overall treatment cohort denominator. Beyond their follow-up duration they are removed from the denominator because one can say nothing further about the timing of the onset of the event for them because it is unknown: they have not died and one cannot say when or if they would have if they had been followed longer. Although the graphs have a common origin and a single time scale, their follow-up began and ended at different times. The marks on the survival curves indicate the point at which they were removed from calculations of the cumulative survival time. I chose Cox and Kaplan–Meier in part because they incorporate censored cases properly and are more robust to unequal durations of follow-up than alternative survival statistics such as Logistic Regression.

Univariate p53 analysis based on quantitative p53 assignment of binary $p53^+$/$p53^-$ displayed an equally strong and significant prognostic ability. In a simple 2 × 2 cross tabulation, the relationship between inactivated p53 and survival was statistically significant, with five of the 15 (33%) $p53^+$ cases remaining alive at follow-up compared with eight of the nine (89%) $p53^-$ cases (Fisher's exact test, $p = 0.01$). Kaplan–Meier analysis demonstrated a significant and strongly inverse correlation between $p53^+$ status and cumulative survival (log rank = 6.17, $p = 0.013$) (Figure 4a), with a mean post-diagnostic survival of 68.3 months for $p53^-$ patients but only 37.6 months for $p53^+$ patients. However, Kaplan–Meier analysis failed to detect such an association when $p53^+$ status was determined by visual estimation (log rank = 0.57, $p = 0.45$) (Figure 4b). Such IHC scoring indicated a slightly protective effect for the loss of the p53 tumour suppressor: on average the $p53^+$ patients slightly outlived the $p53^-$ patients (48.2 months versus 43.6 months). The superior discriminatory ability of quantification to assign binary p53 status was due to its ability to eliminate subjective false positives among those patients alive at the conclusion of follow-up.

Using quantification to dichotomize p53 status for univariate Cox's proportionate hazards analysis revealed that, although quite not as prognostically powerful as the extremes of grade or stage ranks, the effect of loss of p53 tumour suppression was statistically significant ($p = 0.039$) and conferred a nearly nine-fold increased relative risk of death due to the cancer over the follow-up period for any given time (t) since diagnosis:

$$\log_e h(t) = h_0(t) + 2.177x$$

where, $h(t) = [h_0(t)][e^{\beta}]$, $h_0(t)$ is the baseline hazard rate for $p53^-$ patients, and $e^{\beta} = e^{2.177} = 8.82$.

The number of cancer patients was too small to achieve statistically significant independent multivariate risk for p53 inactivation when controlling for stage or grade. However, within every level of both stage and grade, the Cox's proportional hazards curves revealed accelerated rates of death for $p53^+$ patients compared with

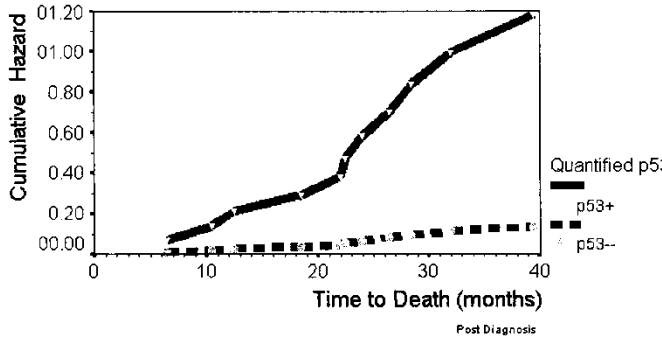

Figure 5. Cumulative proportional risk of $p53^+$ status in adenocarcinoma patients as assessed by Cox's proportional hazards, showing a constant, added risk of death. Cumulative proportional hazard: $h(t) = [h_0(t)]\ [e^{(\beta_1 x_1 + \beta_2 x_2 \ldots \beta_n x_n)}]$. Proportional hazard: $\log_e$ [death rate] $= y = a(t) = \beta_1 x_1 + \beta_2 x_2 \ldots \beta_n x_n$, where t is the time since surgery, $a(t)$ is the baseline ($p53^-$) incidence rate of death, $[h_0(t)]$ is the baseline ($p53^-$) cumulative hazard rate, x is the only risk variable specified (i.e. p53 status; $p53^- = 0$, $p53^+ = 1$), and β is the proportional hazard increase per increase in x.

$p53^-$ patients. When holding grade constant, the 8.82-fold elevated relative risk associated with $p53^+$ status decreased only slightly to 8.61, and the $p53^+$ regression coefficient remained significant ($p = 0.024$). Controlling for tumour stage, the $p53^+$ relative risk dropped slightly to 7.18; again, the $p53^+$ R coefficient remained significant ($p = 0.027$).

The added risk of death due to $p53^+$ status was almost perfectly constant (linear) relative to the baseline risk for $p53^-$ patients over the entire follow-up period (Figure 5). In this univariate comparison the $p53^+$ and the $p53^-$ cancer patients shared similar distributions of stage and grade, indicating that the influence of the other risk factors were constant over the follow-up period.

Besides its application to binary p53 status, the Cox's proportionate hazards statistic (Kelsey *et al*. 1986) was also used to model the cumulative cancer patient survival according to the absolute quantity of aberrant p53 per patient. Analysis of the aberrant p53 body burden – a measure of immunostaining intensity and extent and number of mutant p53 loci – revealed a highly significant ($p < 0.0004$) inverse regression. The slope of this shortened survival became steeper with increasing dose of aberrant p53. The survival pattern of $p53^-$ patients' (0 ng body burden) provided the baseline hazard function. Where the p53 risk factor is a continuous independent variable, e^{β} is not a relative risk index; instead it measures the percentage increase in the hazard rate for a given amount of aberrant p53 compared with that of the $p53^-$ patients (Table 2). For example, a 1 ng p53 body burden had a corresponding hazard rate (e^{β}) of 1.42 ($p = 0.0004$) and a downward survival curve. The projected survival patterns in Table 2 for body burdens of 0 ng (baseline) and 5 ng p53 show the dramatic shortening of survival duration as the amount of p53 inactivation progresses: there was a 42% decrement in survival rate

Table 2. Proportional hazards regression equation: adenocarcinoma patient survival duration is inversely proportional to p53 body burden.

		Survival		
Time (months)	Baseline cumulative hazard	Baseline	1 ng body burden	5 ng body burden
6.5	0.0083	0.9917	0.9882	0.9531
10.5	0.0179	0.9821	0.9747	0.9014
12.5	0.0316	0.9684	0.9555	0.8315
18.5	0.0544	0.9456	0.9237	0.7251
22.0	0.0875	0.9125	0.8728	0.5909
22.5	0.1240	0.8760	0.8288	0.4673
24.0	0.1652	0.8348	0.7740	0.3543
26.5	0.2089	0.7911	0.7172	0.2602
28.5	0.2546	0.7454	0.6591	0.1848
32.0	0.3136	0.6864	0.5864	0.1151
39.5	0.3990	0.6010	0.4856	0.0536

Overall score: $\chi^2 = 19.860$, d.f. $= 1$, $p < 0.0001$ (β_0 and $\beta_1 = 0$, χ^2 approximation to likelihood-ratio test).

Change from baseline hazard: $\chi^2 = 12.391$, d.f. $= 1$, $p = 0.0004$ ($\beta_0 = \beta_1$, likelihood-ratio test).

Proportionate hazards regression equation: variable = ng p53 per patient, $\beta_1 = 0.3497$ (SD 0.0993), d.f. $= 1$, $p = 0.0004$, $R = 0.4062$, $e^{\beta_1} = 1.42$ (compared to $p53^-$ patients, a 1 ng aberrant p53 body burden implies a 42% decrease and a 5 ng aberrant p53 body burden implies a 476% decrease in survival rate per added month follow-up).

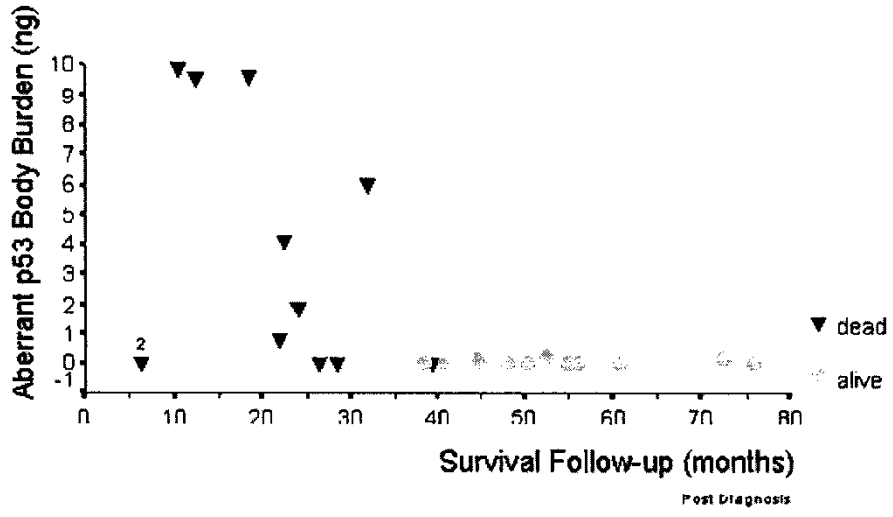

Figure 6. Aberrant p53 body burden versus survival in adenocarcinoma patients. [2] Sole patient for whom an initial biopsy (100% $p53^+$) and eventual diagnosis (stage C) were available but not resected tissue; this is probably an 'outlier', understating the true p53 body burden.

per month of follow-up compared with $p53^-$ patients for a body burden of 1 ng, but a 476% decrement for those with a body burden of 5 ng. The square of the correlation coefficient R is often interpreted to be the percentage of variation of the dependent variable (survival) accounted for by variation in the independent variable. Here, that would imply that $\sim 17\%$ of the variability in patient survival duration was 'explained' by the patient's quantity of aberrant p53. The mean body burden for the 15 cancer patients who were $p53^+$ patients was 2.51 ng (1.6 ng per tumour), although the variability (SD 3.42 ng) and range (10–9.91 ng) were very large.

The *Micrometer* densitometry computer program was also used to measure the size of the enterocytes in the tissue sections. When combined with the absolute p53 quantification, this permitted estimation of the number of aberrant p53 molecules per cell. Since this measure of the cellular intensity of disease-related protein expression is unaffected by the protein's molecular weight, it represents a 'molar' measure of tissue damage. When the Cox's proportionate hazards regression was applied to this approach to p53 quantification, aberrant p53 was again statistically significant ($p = 0.014$) in predicting cancer patient survival duration.

Figure 6 shows the actual survival pattern in relation to the p53 body burden, divided according to whether the patient was alive or dead at the end of the follow-up period. There was a clear demarcation between those whose tumours had lost p53 tumour suppressor function and those whose tumours had not: the former died much sooner and with a survival duration inversely proportional to their 'dose' of aberrant p53.

The accumulation of aberrant p53 could have been a spurious cause of death, being simply a consequence or intervening variable between the presumed 'true' cause of death, that is, the growth of the cancer. Among the cancer patients who died there was indeed a strong association between the accumulation of p53 per tumour (the tumour burden) and size of their tumours, according to a different cell

imaging measure (the total nuclear area) distinct from that used to calculate tumour burden ($R^2 = 0.81$, $p < 0.001$). However, the aggregate tumour nuclear area for the $p53^-$ patients was actually larger than for the $p53^+$ patients, being 41.9 mm^2 and 12.45 mm^2, respectively. Furthermore, while there was an inverse relationship between survival duration and total cancer nuclear area for the $p53^+$ cancer patients ($R^2 = 0.39$, $p = 0.018$), there was no such association among the $p53^-$ cohort ($R^2 = 0.001$, $p = 0.93$).

Discussion

Study design

The rates of p53 inactivation in the adenocarcinoma patients/tumours (63%/67%) were consistent with previous colorectal IHC studies (47–80%) that had used sensitive detection techniques, that is, DNA sequencing or IHC antigen retrieval. The rates of $p53^+$ among adenoma patients/tumours (24%/20%) were towards the high end of most previous reports (3.3–25%). Our use of an adenocarcinomatous calibration cell line containing no mutant p53 cells and our identification of contiguous 'islands' of aberrant p53-expressing cells in tissue samples helped us to avoid false-positive, so-called 'mosaic' staining patterns. Isolated, diffuse immunostained nuclei were sometimes present in the calibration cells lacking inactivated p53. These are cells transiently expressing wild-type p53 during endogenous tumour suppression of the cancer cells, captured 'in the act' by the pantropic anti-p53 monoclonal antibody. IHC sensitivity and specificity were good: the staining distributions for the $p53^-$ cancerous colorectal cell line and the hospital controls were nearly identical; no normal tissue adjacent to tumours was $p53^+$; the possibly false-positive hyperplastic samples and the single $p53^+$ normal tissue were very low incidence, very lightly stained and had very small $p53^+$ volumes, and were probably due to excess tissue exposure to the diaminobenzidine chromagen. Agreement with other assays applied to the tissue and/or calibration cells, with the matched blood samples, and with the paired histological templates suggests that the IHC method and its scoring were valid. The results of various quality controls showed quantitative IHC scoring to be fairly homogeneous and reproducible, and the sections to be representative of the entire tumour.

The risk profile of this study population was similar to that of the US average for both races and genders for the period 1992–1997 (35% local, 39% regional and 21% distant) (Ries *et al*. 1999), although this study had a smaller share of patients at both extremes. The 48% death rate over the 6.5–72 month follow-up period (median 42 months), along with the high rate of $p53^+$, permitted sufficient statistical power to analyse p53-related survival. This rate was somewhat below the 39% 60 month death rate for US colorectal cancer patients reported for the period 1992–1997 (Ries *et al*. 1999), but the average ages of the cancer, adenoma and hospital control patients were quite representative. Overall, the study had good 'external' validity, that is it can be generalized to US colorectal dysplasia patients.

We believe the internal epidemiological validity was adequate, since the targeted risk factors were those actually measured. Prior colorectal cancer studies have revealed that tumour stage, grade, size, site and patient age and gender do not

confound p53 status. The size and direction of the relative risk of death from these cancers and association of that risk with tumour stage and grade suggests that the diagnoses in these patients were correct. The original H&E diagnostic slides underwent repeat evaluation, and additional H&E tissue was sampled and scored at the time of the study to be sure changes had not occurred in the archival tissue block since surgery. The hospital controls were similar to the cases in potential risk factors, and the source and sampling of the tissue was identical. The quality of molecular diagnosis can be no better than the quality of the clinical sampling and diagnosis. The level of surgical skill and accuracy/sensivity of the colonoscopy procedure assured that tissue sample collection would not circumscribe the potential validity/sensitivity/specificity of the p53 quantification.

Some aspects of the design compensated for the modest number of cancer patients: all the tissue blocks for each patient were scored; hundreds of cells were scored per section; there were high rates of $p53^+$; the 'misclassification error' rates were low; and the small percentage (4%) of cancer patients lost to follow-up assured insignificant potential sampling bias. There was sufficient univariate statistical power to detect several significant correlations with quantified aberrant p53. Nonetheless, a larger, multivariate study could better control for and weigh any potentially confounding influences on p53 inactivation.

The findings

This study sheds some light on the natural history of colorectal neoplasia. Often multiple adenomas occur within polyps, some of which lose p53 tumour suppression function to varying degrees. Those reaching a threshold in aberrant p53 accumulation (~ 1 ng) are very likely to become cancers, particularly if they are large and of non-tubular histology. While there is a very large difference between adenomas and cancers with respect to the extent or volume of p53 inactivation, the intensity of $p53^{mut}$ expression (concentration/expression per cell) varies much less, and is greatest within those adenomas in transition to adenocarcinoma, that is, adenomatous regions of serrated adenoma/adenocarcinoma polyps. Whether grouped by patients or tumours, the incidence of $p53^+$ for combined adenomas and adenocarcinomas is consistent with the prospective tumour progression animal models. In comparison with those patients whose tumours are purely adenomatous, cancer patients with synchronous adenomas actually had a lower percentage of $p53^+$ adenomatous tissue on cell imaging scoring (15% versus 23.7%). It is possible that some of the $p53^+$ pre-cancers had already progressed to cancer, leaving the $p53^-$ adenomatous foci behind. The percentage of $p53^+$ adenomatous lesions (19.8%) was slightly lower than the percentage of $p53^+$ patients (23.7%). We commonly saw multiple adenomas per patient, thereby increasing the risk that at least one adenoma would undergo a $p53^+$-induced progression to cancer. By calculating simple univariate ORs and attributable risk in the (cancer) population, we estimate that 50.8% of the risk of progression from adenoma to cancer was due to loss of the p53 tumour suppressor gene. There was no demonstrated elevated risk of recurrence among our overall $p53^+$ adenoma cohort, although the limited follow-up duration reduced the probability of observing all such recurrences (Lev 1990).

Once the progression to cancer had occurred, tumour size affected patient survival, but only if the cancer was $p53^+$ (also see Cohen *et al*. 1993). p53 loss greatly worsened survival outcome and sped the onset of death in a dose-dependent manner, probably due to the spread of the 'aggressive' $p53^+$ fraction of the adenocarcinomatous cells. *In vitro* and animal studies have shown that cells expressing aberrant p53 lack the ability to repair DNA damage, perform apoptosis/inhibit cell division, enjoy a selective survival advantage in the primary tumour, have the greatest propensity to metastasize, and resist immune response or drug treatment.

The significant dichotomous and rank-order statistical associations between the two IHC scoring methods show that the two methods were not wildly disparate. Visual estimation scoring was inferior, done as well as possible; its subjective ordinal ranking hierarchy was considerably more rigorous, graduated and explicit in its categories than is typically the case. The results show why quantification-based scoring of p53 IHC is superior. It offers better quality control (standardized scoring, greater accuracy and precision, reproducibility and consistency) and greater statistical power to detect differences. It provides a rational statistical or clinical basis to define positivity, eliminating the ambiguity of borderline $p53^+$/$p53^-$ immunostaining, prevents overlapping ordinal ranks, and is sensitive to extremes of p53 inactivation. Our incorporation of tissue class scoring templates in quantitative immunostaining scoring ('immunohistology') could also be advantageously applied to subjective scoring.

The ultimate test of comparative IHC scoring approaches is their relative clinical effectiveness in the diagnostic and prognostic evaluation of patients and their tissue sections. The Kaplan–Meier analysis of cancer patient survival showed that quantification has a greater ability to discriminate between the survival of $p53^+$ and $p53^-$ patients. Applying visual estimation, we misclassified several of the surviving patients as being $p53^+$. Subjective false positives from amongst the ambiguous +/− sections created a large misclassification error; several of the cancer patients with truly functional p53, that is $p53^-$, who were still alive at the end of follow-up were incorrectly scored as $p53^+$. Inappropriately combining them with the true $p53^+$ patients, most of whom who had died, effectively diluted the true risk attributable to the loss of p53 tumour suppression. Using p53 visual estimation, the average survival duration was only 43.6 months for $p53^-$ patients, but was 68.3 months using p53 quantification. Absolute p53 quantification in tissue may facilitate the establishment of minimally invasive surrogate measures. Among a subset of 16 adenoma and adenocarcinoma patients for whom we had matched tissue and blood samples, p53 tissue quantification correlated more strongly and significantly with mutant p53 concentrations in plasma as tested by Cohen's kappa coefficient than did IHC visual estimation ($\kappa = 0.54$, $p = 0.002$ and $\kappa = 0.36$, $p = 0.025$, respectively). The ability of quantitative p53 to correlate with histological tissue classes, quantitative survival data and patient-matched blood samples, and its suggested pattern of independence from stage and grade, demonstrates the superiority of aberrant p53 quantification.

Quantitative p53 methods

Microscope-based cell imaging densitometry provided a foundation for p53 tissue quantification. It provided the needed objectivity, reproducibility (repeat scoring $R^2 = 0.87$, $p = 0.006$), consistency and speed, and the ability to easily acquire, aggregate and interpret IHC data. Its ability to combine tissue morphology and immunostaining has made *immunohistology* with tissue class 'templates' possible. Histology summarizes the accumulated effects of all the activated oncogenes and inactivated tumour suppressor genes and their signal transduction effectors (Hunter 2000). *In situ* scoring allowed the identification and segregation of those cell types affected by the disease. Scoring p53 expression solely in $p53^+$ enterocytes targeted by colorectal dysplasia avoided the 'ecological fallacy' at the cellular level: the 'exposure' and disease were observed among the relevant cell population. This stands in contrast to *ex situ* p53 tissue quantification performed on colorectal tumour lysate (Joypaul *et al*. 1993). In this method the physiologically relevant concentration of any $p53^+$ tissue focus is diluted in its surrounding environment of stromal material and non-enterocyte cells, with the dilution occurring to a variable, unpredictable degree.

In the present study, the tissue quantification was valid and accurate, and standardization was achieved. The five empirical criteria listed above for valid and effective quantitative tissue surrogates were all satisfied by the calibration cells. Furthermore, not only were the $p53^+/p53^-$ signal to noise immunostaining ratios for the calibration cells and the cases/controls very similar (~ 23), this common signal to noise ratio was nearly identical to the ratio of the half-lives of aberrant/wild-type p53 (~ 22). This suggests that the average immunostaining scores per cell were proportional to the cell's underlying loss of biological function, transforming IHC into an accurate dosimeter for lost p53 tumour suppressor protection. This may inform the derivation of the proper *in situ* therapeutic dose.

Cell imaging pixels were initially converted to mass amounts of inactivated p53 per mass of enterocyte protein. Such mass amounts can be easily and accurately calculated with existing techniques and freely translated to different levels of human measurement: tissue sections, tumours, tissue or blood. Use of approximated cellular volumes, the molecular weight of p53 and the average p53 expression in the calibration cells permitted IHC staining to be converted to the number of p53 molecules per cell. This 'molar equivalent' measure of aberrant p53 tissue expression also significantly correlated with patient survival. This alternative form of IHC quantification allows the concentrations of disease-related proteins to be compared, irrespective of their molecular bulk.

We did not attempt to identify particular p53 point mutations, such as dominant negatives, in the tissue or blood of the patients (Caron de Fromental and Soussi 1992, Levine *et al*. 1994), nor did we consider p53 LOH. Most sporadic colorectal cancers have two to four somatic gene defects, which are predictable in their combination and less so in their sequence (Fearon and Vogelstein 1990, Fearon and Jones 1992). We did not control for the status or timing of these oncogenes and tumour suppressors. For example, incorporation of oncogenic K-ras^{mut} (Morin *et al*. 1997) – a correlate but not an effect of p53 loss – would likely have improved our prediction of the adenoma to adenocarcinoma transition.

Likewise, analysis of the expression of the deleted in colon cancer (*DCC*) gene and/or of the various apoptosis-related genes would probably improve prognostic accuracy.

IHC tissue analysis with a high avidity/affinity pantropic antibody such as DO-1 will not capture all relevant diagnostic, prognostic and therapeutic information. However, performed quantitatively, it can become a very efficient, simple and powerful assay. The value of quantitative IHC will depend on the molecular pathology of the particular disease-related antigen. The reasons this approach was especially apt for inactivated p53 were outlined in the Introduction: there is a single phenotype with known functional effects for nearly all mutations, there is convergence of the multiple and severe consequences of p53 inactivation, and there is an associated survival advantage. The strong proportionality between the accumulation of inactive p53 protein, its antigenicity and the degree of its dysfunction may not be universal for all tumour antigens. Colorectal neoplasia was a favoured test system in view of the existence of a well-established and accurate tissue sampling procedure, that is, colonoscopy. The strength of the tissue/blood connection benefited only modestly from the degree of venous perfusion particular to the large intestine.

Implications

The increasing automation of fixation, immunostaining and cell imaging densitometry has improved the speed, reliability and cost of IHC, as well as the objectivity of its interpretation. The transition from research tool to clinical biomarker, however, requires intra- and inter-laboratory standardization. Incorporation of physiologically appropriate immunostaining calibration standards undergoing IHC processing equivalent to that of properly fixed tissue sections is the next step. Batch variations and protocol differences need no longer preclude data aggregation and unified statistical analysis. Standard curves, common criteria for positive immunostaining, and an algorithm for estimation of total amounts of tumour/patient antigen will improve prospective, multicentre clinical trials. Measurement of the absolute amount of disease-related proteins offers improvements in diagnosis, treatment and prognosis that are specific to the actual genetic lesion and the individual patient. Undeniably, the validity of IHC quantification must be demonstrated for each antibody/antigen/tissue combination by application of the type of empirical criteria described in this study. Replication of results for a particular combination will make it possible to establish clinically relevant ranges of antigen expression in place of statistical ones. The *in situ* nature of IHC measurement suggests the possible creation of 'molecular scalpels' that determine the necessary dosage of therapeutic molecules, tailoring the treatment to the actual amount of disease-related protein present. Measuring the degree of molecular injury may rationalize the assignment of patients to particular treatment options, which is currently a recognized dilemma (O'Leary and Calvin 1988). Establishment of a reliable and accurate correlation between levels measured in tissue and those in body fluids may allow the inverse association to be applied: levels can be measured in body fluids and assumptions drawn about levels in tissue. This could

be applied after surgery or adjuvant therapy in order to determine treatment effectiveness and to monitor recurrence or relapse.

Acknowledgements

The authors would like to thank Gail Garbowski, Administrative Assistant, for help with patient selection/data and sample collection, the Harlem Hospital Tumour Registry and Stuart Herna of the CPMC Tumour Registry for help with patient follow-up, and CPMC oncologists Martin Oster and Mark Stoopler.

References

Adams, E. J., Green, J. A., Clark, A. H. and Youngson, J. H. 1999, Comparison of different scoring systems for immunohistochemical staining. *Journal of Clinical Pathology*, **52** (1), 75–77.

American Society of Clinical Oncology 1996, Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. *Journal of Clinical Oncology*, **14** (10), 2843–2877.

Baas, I. O., Mulder, J.-W. R., Offerhaus, G. J. A., Vogelstein, B. and Hamilton, S. R. 1994, An evaluation of six antibodies for immunohistochemistry of mutant *p53* gene product in archival colorectal neoplasms. *Journal of Pathology*, **172**, 5–12.

Bonsing, B. A., Corver, W. E., Gorsira, M. C. B., van Vliet, M., Oud, P. S., Cornelisse, C. J. and Fleuren, G. 1997, Specificity of seven monoclonal antibodies against p53 evaluated with Western blotting, immunohistochemistry, confocal laser scanning microscopy, and flow cytometry. *Cytometry*, **28**, 11–24.

Campo, E., de la Calle-Martin, O., Miquel, R., Palacin, A., Romero, M., Fabregat, V., Vives, J., Cardesa, A. and Yague, J. 1991, Loss of heterozygosity of *p53* gene and p53 protein expression in human colorectal carcinomas. *Cancer Research*, **51**, 4436–4442.

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

Chin, Khew-Voon, Kazumitsu, Veda, Pdstan, Ira and Gottesman, Michael M. 1992, Modulation of activity of the promoter of the human MDR1 gene by Ras and p53. *Science*, **255**, 459–462.

Cohen, J. 1960, A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.

Cohen, A. M., Minsky, B. D. and Schilsky, R. 1993, Colon cancer. In *Cancer: Principles and Practice of Oncology*, edited by V. T. Devita, S. Hellman and S. Rosenberg, 4th edition (Philadelphia: J. B. Lippincott), pp. 929–977.

Cunningham, J., Lust, J. A., Schaid, D., Bren, G. D., Carpenter, H. A., Rissa, E., Kovach, J. S. and Thibodeau, S. A. 1992, Expression of p53 and 17p allelic loss in colorectal carcinoma. *Cancer Research*, **52**, 1974–1980.

Dowell, S. P. and Hall, P. A. 1995, The *p53* tumor suppressor gene and tumor prognosis: is there a relationship? *Journal of Pathology*, **177**, 221–224.

El-Diery, W. S. 1998, Regulation of p53 downstream genes. *Seminars in Cancer Biology*, **8**, 345–357.

Fearon, E. R. and Vogelstein, B. 1990, A genetic model for colorectal tumorigenesis. *Cell*, **61**, 759–767.

Fearon, E. R. and Jones, P. A. 1992, Progressing toward a molecular description of colorectal cancer development. *FASEB Journal*, **6**, 2783–2790.

Fidler, I. 1997, Molecular biology of cancer invasion and metastasis. In *Cancer: Principles and Practice of Oncology*, edited by V. T. Devita, S. Hellman and S. Rosenberg, 5th edition (Philadelphia: Lippincott-Raven), pp. 135–152.

Fisher, C. J., Gillett, C. E., Vojtešek, B., Barnes, D. M. and Millis, R. R. 1994, Problems with p53 immunohistochemical staining: the effect of fixation and variation in the methods of evaluation. *British Journal of Cancer*, **69**, 26–31.

Galen, R. S. and Gambino, S. R. 1975, *Beyond Normality: The Predictive Value and Efficiency of Medical Diagnoses* (New York: Wiley & Sons).

Gannon, J. V., Iggo, R. and Lane, D. P. 1990, Activating mutations in p53 produce a common conformational effect. A monoclonal antibody specific for the mutant form. *EMBO Journal*, **9** (5), 1595–1602.

Graeber, T. G., Osmanian, C., Jacks, T., Housman, D. E., Koch, C. J., Lowe, S. W. and Giaccia, A. J. 1996, Hypoxia-mediated selection of cells with diminished apoptotic potential in solid tumours. *Nature*, **379**, 88–91.

Greco, C., Gandolfo, G. M., Mattei, F., Gradilone, A., Silvana, A., Pastore, L. I., Casale, V., Casole, P., Grassi, A., Cianciulli, A. M. and Agliano, A. M. 1994, Detection of *c-myb* genetic alterations and mutant p53 serum protein in patients with benign and malignant colon lesions. *Anticancer Research*, **14**, 1433–1440.

Harris, L. C., Remack, J. S., Houghton, P. J. and Brent, T. P. 1996, Wild-type p53 suppresses transcription of the human O^6-methylguanine-DNA methyltransferase gene. *Cancer Research*, **56**, 2029–2032.

Hendrix, M. J. 2000, De-mystifying the mechanism(s) of maspin. *Nature Medicine*, **6**, 374–376.

Hunter, T. 2000, Signaling – 2000 and beyond. *Cell*, **100**, 113–127.

Joypaul, B. V., Vðjtesek, B., Newman, L., Hopwood, D., Grant, A., Lane, D. P. and Cuschieri, A. 1993, Enzyme-linked immunosorbent assay for p53 in gastrointestinal malignancy: comparison with immunohistochemistry. *Histopathology*, **23**, 465–470.

Kelsey, J. L., Thompson, W. D. and Evans, A. S. (editors) 1986, *Methods in Observational Epidemiology* (New York: Oxford University Press).

Kemp, C. J., Donehower, L. A., Bradley, A. and Balmain, A. 1993, Reduction of *p53* gene dosage does not increase initiation of promotion but enhances malignant progression of chemically induced skin tumors. *Cell*, **74**, 813–822.

Kimura, O., Sugamura, T., Kijima, M., Makino, H., Shirai, S., Tatebe, H., Ito, H. and Kaibara, N. 1996, Flow cytometric examination of p53 protein in primary tumors and metastases to the liver and lymph nodes of colorectal cancer. *Diseases of the Colon and Rectum*, **39** (12), 1428–1433.

Kraiss, S., Spiess, S., Reihsaus, E. and Montenarh, M. 1991, Correlation of metabolic stability and altered quaternary structure of oncoprotein p53 with cell transformation. *Experimental Cell Research*, **192**, 157–164.

Lane, D. and Hall, P. 1997, MDM2 – arbiter of p53's destruction. *Trends in Biochemical Sciences*, **10**, 371–410.

Lev, R. (editor) 1990, *Adenomatous Polyps of the Colon* (New York: Springer-Verlag).

Levine, A. J. and Momand, J. 1990, Tumor suppressor genes: the *p53* and retinoblastoma sensitivity genes and gene products. *Biochimica et Biophysica Acta*, **1032**, 119–136.

Levine, A. J., Perry, M. E., Chang, A., Silver, A., Dittmer, D., Wu, M. and Welsh, D. 1994, The role of the *p53* tumor suppressor gene in tumorigenesis. *British Journal of Cancer*, **69**, 409–416.

Lowe, S. W., Jacks, T., Housman, D. E. and Ruley, H. E. 1994, Abrogation of oncogene-associated apoptosis allows transformation of p53-deficient cells. *Proceedings of the National Academy of Sciences USA*, **91**, 2026–2030.

Luo, J.-C., Neugut, A. I., Garbowski, G., Forde, K., Treat, M., Smith, S., Carney, W. and Brandt-Rauf, P. 1995, Levels of p53 antigen in the plasma of patients with adenomas and adenocarcinomas of the colon. *Cancer Letters*, **91** (2), 235–240.

McShane, L. M., Aamodt, R., Cordon-Cardo, C., Cote, R., Faraggi, D., Fradet, Y., Grossman, H. B., Peng, A., Taube, S. E. and Waldman, F. M. 2000, Reproducibility of p53 immunohistochemistry in bladder tumors. *Clinical Cancer Research*, **6** (5), 1854–1864.

Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. and Kinzler, K. W. 1997, Activation of β-catenin/Tcf signaling in colon cancer by mutations in β-catenin or APC. *Science*, **275**, 1787–1790.

O'Leary, T. and Calvin, V. 1998, U.S. Food and Drug Administration (FDA) approval of HERCEPTEST assay (HER-2). Joint Meeting of the 'Hematology and Pathology Devices' and 'Immunology Devices' Panels. http://www.fda.gov/, click on Transcript: Part 2 pages 1–113).

Ozbun, M. A. and Butel, J. S. 1997, *p53* tumor suppressor gene: structure and function. In *Encyclopedia of Cancer*, Vol. II, edited by J. R. Bertino (New York: Academic Press), pp. 1240–1257.

Ries, L. A. G., Kosary, C. L., Hankey, B. F., Miller, B. A., Clegg, L. and Edwards, B. K. (editors) 1999, *SEER Cancer Statistics Review, 1973–1998* (Bethesda: Cancer Statistics Branch, Cancer Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute).

Rosanelli, G. P., Wirnsberger, G. H., Purstner, P., et al. 1993, DNA flow cytometry and immunohistochemical demonstration of mutant p53 protein versus TPs and mutant p53 protein serum levels in human breast cancer. *Proceedings of the American Association for Cancer Research*, **34**, A1353.

Ruddon, R. W. (editor) 1987, *Cancer Biology* (New York: Oxford University Press).

Schottenfeld, D. and Winawer, S. J. 1982, Cancer epidemiology and prevention. In *Cancer Epidemiology*, edited by D. Schottenfeld and D. Fraumeni (Philadelphia: Saunders).

Sharma, S., Schwarte-Waldhoff, I., Oberhuber, H. and Schafer, R. 1993, Functional interaction of wild-type and mutant p53 transfected into human tumor cell lines carrying activated *RAS* genes. *Cell Growth and Differentiation*, **4**, 861–869.

SHIM, K.-S., KIM, K.-H., PARK, B.-W., LEE, S.-Y., CHOI, J.-H., AN, W.-D. and PARK, E.-B. 1998, Increased serum levels of mutant p53 proteins in patients with colorectal cancer. *Journal of the Korean Medical Society*, **13**, 44–48.

SHIN, D. M., KIM, J., RO, J. Y., HITTELMAN, J., ROTH, J. A., HONG, W. K. and HITTELMAN, W. N. 1994, Activation of p53 gene expression in premalignant lesions during head and neck tumorigenesis. *Cancer Research*, **54**, 321–326.

SORY, A., MINAMOTO, T., OHTA, K., YAMASHITA, K., SAWAGUCHI, M., MAI, M. and MISIUNA, P. 1997, Does p53 overexpression cause metastases in early invasive colorectal adenocarcinoma? *European Journal of Surgery*, **163** (9), 685–692.

SOUSSI, T., LEGROS, Y., LUBIN, R., ORY, K. and SCHLICHTHOLZ, B. 1994, Multifactorial analysis of *p53* alteration in human cancer: a review. *International Journal of Cancer*, **57**, 1–9.

SYMONDS, H., KRALL, L., REMINGTON, L., SAENZ, ROBLES M., JACKS, T. and VAN DYKET, T. 1994, p53-dependent aptosis in vivo: impact of p53 inactivation on tumorigenesis. *Cold Spring Harbor Symposium on Quantative Biology*, **59**, 247–257.

THOMLINSON, R. H. and GRAY, L. H. 1955, The histological structure of some human lung cancers and the possible implications for radiotherapy. *British Journal of Cancer*, **9**, 539–549.

TOMINAGA, O., HAMELIN, R., REMVIKOS, Y., SALMAON, R. and THOMAS, G. 1992, p53 from basic research to clinical applications. *Critical Reviews in Oncogenesis*, **3**, 257–282.

VAN DER KOOY, K., ROOKUS, M. A., PETERSE, H. L. and LEEUWEN, F. E. 1996, p53 overexpression in relation to risk factors for breast cancer. *American Journal of Epidemiology*, **144** (10), 924–933.

VAUPEL, P. W. and HOCKEL, M. 1995, Oxygenation status of human tumors; a reappraisal using computerised pO_2 histography. In *Tumor Oxygenation*, edited by P. W. Vaupel, D. K. Kelleher and M. Gunderoth (Stuttgart: Gustav Fischer), pp. 219–232.

VOGELSTEIN, B., FEARON, E. R., HAMILTON, S. R., KERN, S. E., PRESINGER, A. C., LEPPERT, M., NAKAMURA, Y., WHITE, R., SMIT, A. M. and BOS, J. L. 1988, Genetic alterations during colorectal tumor development. *New England Journal of Medicine*, **319** (9), 525–532.

VOGELSTEIN, B., LANE, D. and LEVINE, A. J. 2000, Surfing the p53 network. *Nature*, **408** (16), 307–310.

VOJTEŠEK, B., DOLEZALOVA, H., LAUEROVA, L., SVITAKOVA, M., HAVLIS, P., KOVARIK, J., MIDGLEY, C. A. and LANE, D. P. 1995, Conformational changes in p53 analyzed, using new antibodies to the core DNA binding domain of the protein. *Oncogene*, **10**, 389–393.