A Novel Logistic Model Based on Clinicopathological Features Predicts Microsatellite Instability in Colorectal Carcinomas

Anna Colomer, PhD, Nadina Erill, PhD,* August Vidal, MD,*† Miquel Calvo, PhD,‡ Ruth Roman, PhD,* Montse Verdú, PhD,† Carlos Cordon-Cardo, MD, PhD,*†§ and Xavier Puig, MD, PhD*†*

Abstract: High-frequency microsatellite instability has been reported to be associated with good prognosis in colorectal adenocarcinoma. However, methods to assess microsatellite instability (MIN) are based on genetic assays and are not ideally suited to most histopathology laboratories. The aim of the present study was to develop a model for prediction of MIN status in colorectal cancer based on phenotypic characteristics. Clinicopathological features of a cohort of 204 patients with primary colon cancer were retrospectively reviewed following predetermined criteria. Genetic assessment of MIN status was performed on DNA extracted from sections of formalin-fixed, paraffin-embedded specimens by testing a panel of 11 microsatellite markers. Logistic regression analysis generated a mathematical tool capable of identifying colorectal tumors displaying MIN status with a sensitivity of 77.8% and a specificity of 96.8%. Features associated with instability included the proximal location of the lesions, occurrence of solid and/or mucinous differentiation, absence of cribriform structures, presence of peritumoral Crohn-like reaction, expansive growth pattern, high Ki67 proliferative index, and p53-negative phenotype. This approach predicts microsatellite instability in colorectal carcinoma with an overall assigned accuracy of 95.1% and a negative predictive value of 97.8%. Implementation of this tool to routine histopathological studies could improve the management of patients with colorectal cancer, especially those presenting with stage II and III of the disease. It will also assist in identifying a subset of patients likely to benefit from adjuvant chemotherapy.

Key Words: hereditary nonpolyposis colorectal cancer, sporadic colorectal cancer, microsatellite instability, DNA replication errors, logistic model

(*Diagn Mol Pathol* 2005;14:213–223)

From *BIOPAT, Grup Assistència, Barcelona, Spain; †HISTOPAT Laboratoris, Barcelona, Spain; ‡Statistics Department of the Universitat de Barcelona, Barcelona, Spain; and §Division of Molecular Pathology, Memorial Sloan-Kettering Cancer Center, New York, NY.

Reprints: Carlos Cordon-Cardo, Division of Molecular Pathology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021 (e-mail: cordon-c@mskcc.org).

Familiar and sporadic colorectal carcinomas arise through a multistep process from the transformation of normal colonic epithelial cells. Two main genetic pathways of tumorigenesis have been distinguished, known as chromosomal instability (CIN) and microsatellite instability (MIN).[1] CIN is the most common, affecting approximately 85% of sporadic colorectal cancers, and shares genetic alterations characteristic of the familial adenomatous polyposis (FAP).[2] It involves mutations of critical genes, such as *APC*, K-*RAS*, and *TP53*; DNA hypermethylation; and chromosomal aberrations, including loss of heterozygosity at chromosomes 17p and 18q, leading to aneuploidy.[2–4] MIN tumors share their molecular mechanisms with hereditary nonpolyposis colorectal cancer (HNPCC). They are characterized by alterations of the mismatch repair (*MMR*) genes[5,6] and secondary mutations affecting genes involved in growth signaling (eg, *TGFßRII*, *IGFRII*)[7,8] and apoptosis (eg, *BAX*).[9] Because cells with alterations in *MMR* genes do not properly repair spontaneous errors occurring during DNA replication, they are prone to accumulate frame-shift mutations and base-pair substitutions at microsatellite sequences. Therefore, colorectal cancers that belong to this mutator phenotype display a characteristic molecular pattern that has been termed "high-frequency microsatellite instability" or MSI-H.[5,6,10]

Clinical evidence let to the conclusion that patients carrying sporadic MIN colorectal cancers behave like those with HNPCC in terms of improved survival[10–12] and display a stage-by-stage better prognosis than those with sporadic CIN tumors.[13–16] Additional studies confirmed the good prognosis of stage II or III colorectal carcinoma patients with MSI-H.[13,14,16–18] Furthermore, it was reported that MSI-H tumors possessed greater chemosensitivity.[17–19] However, it was also found that 5-fluorouracil–based chemotherapy (5-FU) did not significantly increase overall and disease-free survival in patients with stage II or III MSI-H tumors.[20–22] Due to this different response to adjuvant chemotherapy, defining MIN status could assist in discriminating a subgroup of patients who may benefit from 5-FU regimens.

A distinct histologic appearance has been attributed to colorectal cancers arising from the MIN pathway. These lesions are typically right sided, more likely to be poorly differentiated, and often display mucinous-type or signet-ring cells.[23,24] Additionally, these lesions frequently exhibit pronounced intraepithelial lymphocytic infiltration[25,26] and

extensive peritumoral accumulation of lymph cells, similar to that observed in Crohn's disease.[27,28] Several studies were conducted in an attempt to predict the MIN status of colorectal tumors by their sole clinical and pathologic features. Jass et al[23] reported that tumors exhibiting less than 30% unstable microsatellite markers are phenotypically indistinguishable from stable lesions. They concluded that pathologic examination allows a sensitive and specific identification of tumors exhibiting MSI-H by means of a decision tree-based model.

The objective of this study was to provide pathologists with an inexpensive and versatile device to categorize which colorectal tumors should undergo MIN analysis on the basis of available clinical and pathologic data. A model based on stepwise logistic regression is presented by which non–MSI-H tumors could be identified with great accuracy.

MATERIALS AND METHODS

Patient Characteristics and Tissues

Between July 1996 and June 2003, 237 consecutive tumor samples were collected prospectively from an equal number of patients surgically treated for primary colorectal cancer in the area of Barcelona, Spain. Specimens were collected following specific institutional requirements, and the study was approved by the institutional review board. A final cohort of 204 patients was used for this study because the remaining 33 cases (14% of total specimens) were excluded due to poor DNA quality or deficient amplification from either normal or tumor samples. Table 1 summarizes demographic data and clinicopathological features on this cohort. Histologically, lesions were classified as adenocarcinomas or mucinous adenocarcinomas (ICD-O codes 8140/3 and 8480/3, respectively).[29] The extent of tumor invasion and regional lymph node involvement were assigned according to the TNM (UICC) system.[30] Clinical stage was based on the Astler-Coller's modification.[31] Finally, with regard to histologic grading,[29] tumors were classified as low-grade lesions (grade 1: glandular structures in >95% of the tumor), intermediate-grade (grade 2: glandular structures in 50%–95%), and high-grade tumors (grade 3; glandular structures in <50%), being mucinous adenocarcinomas considered high-grade lesions.

The same criteria were used to prospectively collect tumor samples from 68 additional patients surgically treated between July 2003 and February 2005. Table 1 summarizes, along with the original set, data on this independent cohort of patients, which we included to evaluate how the proposed novel tool of prediction works in the context of clinical practice.

Tissues included in the study were routinely fixed in 10% buffered formalin and embedded in paraffin. Representative hematoxylin and eosin (H&E) stained 5-μm thick sections of tumor, normal mucosa, associated lesions, margins (proximal, distal, and radial), and lymph nodes were used in each case to provide final diagnosis. The mean number of tumor slides per specimen was 4.0 ± 2.1 (range 1–18). Viable representative areas of paired normal and tumor tissue were selected for DNA extraction (see below). For the retrospective study following updated criteria, all the available H&E slides were reviewed by a pathologist (either AV or XP) without knowledge of MIN status.

TABLE 1. Characteristics of the Training Set and Validation Set

Clinicopathological Features	Training Set (n = 204)	Validation Set (n = 68)
Age*	68.7 ± 12.1	71.3 ± 10.4
Gender		
Male	119 (58.3)	45 (66.2)
Female	85 (41.7)	23 (33.8)
Histology		
Adenocarcinoma	174 (85.3)	65 (95.6)
Mucinous adenocarcinoma	30 (14.7)	3 (4.4)
Extent of invasion (TNM)		
pT1	13 (6.4)	6 (8.8)
pT2	26 (12.7)	19 (14.7)
pT3	102 (50.0)	26 (38.2)
pT4	63 (30.9)	26 (38.2)
Regional lymph node involvement (TNM)		
pN0	99 (48.5)	40 (58.8)
pN1	58 (28.4)	12 (17.6)
pN2	47 (23.0)	16 (23.5)
Stage (Astler-Coller's)		
A	10 (4.9)	14 (20.6)
B	86 (42.2)	25 (36.8)
C	90 (44.1)	21 (30.9)
D	18 (8.8)	8 (11.8)
Grade (WHO)		
1	76 (37.3)	52 (76.5)
2	82 (40.2)	11 (16.2)
3	46 (22.5)	5 (7.4)

*Arithmetic mean ± standard deviation.

Report and Definition of Histopathological Features

Colorectal cancers were considered proximal or distal when located, respectively, proximal or distal to the splenic flexure (including rectum). On the basis of configuration, tumors were classified as exophytic (or polypoid), ulcerated, and stenosing. A percentage estimation of different tumor patterns including solid, mucinous, and signet-ring cell differentiation were reported in each case. The percentage of cribriform structures, consisting of a growing pattern of glands exhibiting small-sized secondary lumina and rounded in shape, was scored as an independent feature. The proportion of micropapillary pattern was also assessed in every specimen, defined as infiltrating micropapillae without fibrovascular core admixed with tubuloalveolar clusters within clear spaces separated by a fine reticular to fibrocollagenous stroma lacking desmoplasia. The presence of small-sized infiltrating tubular structures of irregular shape, often angulated, and embedded in a sparse desmoplastic stroma—referred to here as infiltrating microglandular pattern—was also determined and expressed as a percent value. With regard to configuration of the advancing margin or growth pattern, tumors were split into

infiltrative when an irregular, infiltrating pattern of growth was demonstrated, as opposed to expansive when the tumor border was a smooth-pushing front.[32] The presence of tumor-infiltrating lymphocytes (TIL) was characterized by the finding of 4 or more unequivocal intraepithelial lymphocytes in a single high-power field (40× objective) of an H&E-stained section, identified and counted in areas displaying the highest content of TILs.[26] Peritumoral Crohn-like lymphoid response was considered positive when at least 3 nodular aggregates of lymphocytes deep to the advancing margin of the tumor were found within a single low-power field (4× objective).[28] Intramural and extramural thin-walled vessel invasion (TWVI), venous vessel invasion (VVI), and perineural invasion (PNI) were also evaluated and recorded. Occurrence of adenomas, nonadenomatous polyps, and presence of residual (contiguous) adenoma were features additionally reported in all cases.

Immunohistochemical Analysis

Immunohistochemistry was carried out using primary mouse monoclonal antibodies DO-7 and MIB-1 (both from DAKO, Glostrup, DK) detecting p53 and Ki67 proteins, respectively. The avidin-biotin immuno-peroxidase method was performed on 4-μm thick paraffin sections. Briefly, sections were treated with 0.1 mol/L citrate buffer (pH 6.0) as the antigen retrieval method and independently incubated overnight at 4°C with mAb DO-7 (1:2000 dilution, 200 ng/mL) or mAb MIB-1 (1:600 dilution, 60 ng/mL). Secondary reagents were biotinylated horse antimouse antibodies (Vector Laboratories Inc., Burlingame, CA) used at 1:100 final working dilution. Avidin-biotin complexes (Vectastain ABC PK 4000 ST, Vector Laboratories) were then applied at 1:100 dilution. 3,3′-diaminobenzidine was used as the final chromogen and Harris-modified hematoxylin as the nuclear counterstain.

Positive controls for DO-7 staining were colorectal tumors previously known to harbor *TP53* mutations displaying positive nuclear immunoreactivity. Positive controls for MIB-1 staining were tumors known to have a high proliferative index (over 20%) as measured by Ki67 nuclear antigen expression. The corresponding negative controls, omitting primary antibody, were also included. Immunohistochemical evaluation was conducted double-blind by scoring the estimated percentage of tumor cells showing nuclear staining. Consistent with previous reports,[33–35] certain intratumoral heterogeneity was observed in the distribution of both p53- and Ki67-stained neoplastic cells. In these cases, the score was produced by estimating the percentage of positive cells in microscopic fields displaying the most intense immunoreactive tumor cells. We avoided microscopic fields displaying weak immunoreactivities, which could represent artefactual loss of antigenicity.

DNA Extraction and Control Amplification

Eleven serial 5-μm thick sections were used for each case. One of these sections was used for tissue characterization after H&E staining. Unstained slides were then aligned by morphology to the stained slide and corresponding areas macrodissected. Tumor samples were microscopically examined to confirm the specificity of dissection.

Genomic DNA was obtained from the macrodissected sections using a proteinase-K extraction method, followed by purification with phenol/chloroform and ethanol precipitation, as described previously.[36] To assess the quality of the genomic DNA extracted, 1 μL of a 200 ng/μL dilution was used as template for the amplification of a 268-bp fragment of the human β-globin gene,[37] using a GeneAmp PCR System 2400 thermal cycler (PE Biosystems, Foster City, CA). The amplification profile is summarized in Table 2. Efficiency of PCR reactions was assessed in 2% agarose gels.

TABLE 2. Size of PCR Products, Primer Labeling, and Profiles of Amplification

Locus	Product Size (pb)	Labeling	Amplification Profile
β-globin gene			
PC04/GH20	268	None	94°C 5 min; hold/94°C 30 sec; 57°C 30 sec; 72°C 30 sec; 35 cycles/72°C 7 min; hold
NCI panel			
BAT25	120	6-FAM	94°C 2 min; hold/94°C 10 sec; 55°C 30 sec; 72°C 30 sec; 30 cycles/72°C 7 min; hold
BAT26	116	TET	
D5S346 (APC)	96–122	HEX	
D2S123	197–227	6-FAM	
D17S250 (Mfd 15CA)	151–169	TET	
18q panel			
D18S55	134–152	6-FAM	94°C 5 min; hold/94°C 30 sec; 55°C 30 sec; 72°C 30 sec; 30 cycles/72°C 30 min; hold
D18S58	144–160	TET	
D18S61	157–183	HEX	
D18S64	188–208	6-FAM	
D18S69	194–210	TET	
TP53 locus			
P53 CA	103–135	6-FAM	94°C 5 min; hold/94°C 30 sec; 60°C 30 sec; 72°C 30 sec; 27 cycles/72°C 30 min; hold

Analysis of Microsatellite Instability

All selected tumors were studied to assess their MIN status using 11 microsatellite markers; 5 of them were chosen following the guides recommended by the American Join Commission on Cancer, the International Collaborative Group on HNPCC, and the HNPCC Cancer Study Group in Germany (BAT25, BAT26, D5S346, D2S123 and D17S250).[38] The 6 additional microsatellites represent a consensus panel originally aimed at elucidating the LOH status of chromosome 18q (D18S55, D18S58, D18S61, D18S64, and D18S69)[39] and LOH at the TP53 locus on 17p (P53CA).[40]

Microsatellite markers from the consensus panel (NCI panel) were amplified in multiplex reactions using the HNPCC Microsatellite Instability Test (Roche Molecular Biochemicals, Mannheim, Germany) according to the manufacturer. Briefly, 200 ng of genomic DNA from both paired normal and tumor samples were used as template for amplification, mixed along with 5 μL of 5× multiprimer mix and 5 μL of 5× enzyme master mix up to 25 μL. PCR conditions and fluorescence dyes for primer labeling are summarized in Table 2.

Amplification of the five 18q microsatellite markers was also performed with 200 ng of genomic DNA in 25-μL reaction mixtures containing 200 μmol of each dNTP, 1.5 μmol/L Mg^{2+}, and 1.05 U of the Expand High Fidelity PCR System (Boehringer Mannheim Corp, Indianapolis, IN). Two to 7 pmols of the 5 pairs of primers reported elsewhere[39] were used per multiplex reaction. Product sizes and fluorescence dyes used to modify the primers are also summarized in Table 2, along with PCR conditions. To avoid polymerase shuttering, a 5′ tail was added at the nonlabeled counterpart of each pair of primers. A final 30-minute extension was performed to avoid incomplete 3′ A nucleotide addition.

Amplification of the P53CA dinucleotide repeat[40] was set up as a single reaction using the same amounts of PCR mix compounds than those for the 18q markers but the polymerase which was diminished to 0.7 U. PCR conditions and primer labeling are also summarized in Table 2.

One microliter of each fluorescent product previously diluted 1:10 in distilled water (except those from the consensus panel of markers) was separately denatured for 2 minutes at 90°C in 12.5 μL of deionized formamide to which 0.5 μL of GeneScan-350 (TAMRA) size marker (PE Applied Biosystems, Foster City, CA) had been previously added. Once denatured, products were run by capillary electrophoresis on an ABI PRISM 310 Genetic Analyzer (PE Applied Biosystems). Analysis was performed using the GeneScan software (PE Applied Biosystems). Instability was assigned to a microsatellite marker if its fragment pattern displayed either additional peaks or the appearance of separated novel fragments when the fluorescence profiles of the normal and tumor tissue were compared with each other. In accordance with consensus definitions of the US National Cancer Institute, tumors were classified as being microsatellite stable (MSS) when none of the tested markers exhibited instability, microsatellite instability-low (MSI-L) when less than 30% of the loci screened displayed instability (up to 3 out of 11), and microsatellite instability-high (MSI-H) when 30% or more of the tested loci resulted unstable.[38]

Statistical Analysis

Twenty-nine independent variables were subdivided into either categorical (gender, tumor location, tumor configuration, histologic grade, extent of invasion, clinical stage, intramural and extramural TWVI, intramural and extramural VVI, intramural and extramural PNI, growth pattern, Crohn-like lymphoid response, TIL, occurrence of adenomas and nonadenomatous polyps, and presence of residual adenoma; n = 18) or numerical (age, tumor size, percentage of solid, mucinous and signet-ring cell differentiations, cribriform growth, micropapillar and infiltrating microglandular patterns, nodal involvement, Ki67 proliferative index, and p53 phenotype; n = 11). Categorical features were investigated for their possible association with MIN status by χ^2 test or Fisher exact test of contingency tables as appropriate. The relationship between MIN status and the numerical variables was approached with a logistic regression-based model that included each of the variables one at a time. For all 3 tests, probability values (*P* values) were considered statistically significant when less than .05. Data were analyzed using the SPSS software package v.11.01 (Statistical Package for the Social Sciences Inc, Chicago, IL).

Two alternative multivariate statistical modeling approaches were considered in a primary stage, both aimed to reach the better predictor set for MSI-H. The tree program of the R package v.1.7.1 (R Development Core Team)[41,42] was used to build decision trees. This hierarchical structure of classification consists of a series of tests carried out in specific order, each one splitting data samples into subsets (nodes) that depend on the possible outcomes of the tests (branches). The tree sequentially flows from the root node toward the leaf nodes, which represent a class of event.

The logistic regression approach was applied to the whole set of 29 variables, with independence of the results obtained from the univariate analysis. Variables were entered into a stepwise procedure to assign a regression coefficient for each factor and to develop a mathematical formula which serves to estimate the probability of a tumor exhibiting MSI-H according to the following equation:

$$P = 1 - \left(\frac{1}{1 + e^{X}}\right)$$

where *P* is the expected probability and

$$X = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

is the development of the linear component of the formula.[27] β_0 is the independent term of the regression equation, β_i is the regression coefficient for the i-th explanatory variable, and x_i is the value of the i-th variable for any individual tumor. Notice that for dichotomous variables, x_i assumes value 1 or 0.

The robustness of both models derived from our cohort of patients was evaluated using a 10-fold cross-validation technique,[43] which is a halfway compromise between the standard procedure that uses a training/validation pair of sets and the one leave-out (jackknife) method. The 10-fold cross-validation technique is a resampling procedure based on a randomly partition of the full dataset into 10 disjoint subsets

of approximately equal size. The prediction model is trained 10 times, each one discarding a different subset, on the remaining 9. Actually, 10 estimations were computed with this approach on reasonably different subsets of data, instead of the single estimation in the training/validation set, or the numerous, but most probably overfitted, estimations in the jackknife. The 10-fold cross-validation technique also allowed comparison of node's position within the trees, something not feasible with the training/validation set approach, that was a key point to choose a best-fitting model in our study (see Results section).

In a subsequent stage, once the modeling approach had been decided, a second independent cohort of patients was used to test the performance of the novel model to validate it as a clinical tool for MSI-H prediction.

RESULTS

Frequency of Microsatellite Instability

Of the 237 tumors initially included in the series, 204 were suitable for microsatellite instability analysis. In 9 cases (3.8%), the study was not performed due to lack of amplification of the β-globine control in either normal or tumor samples. Another 24 cases (10.1%) fail to demonstrate amplification in more than 4 loci, so they were also excluded from the series. Out of 204 evaluable colorectal tumors, 18 (8.8%) exhibited MSI-H, 33 (16.2%) were MSI-L, and 153 (75.0%) failed to demonstrate instability (MSS). Overall, 186 tumors (91.2%) were classified as non–MSI-H (MSI-L plus MSS).

Tumors with MSI-H included cases with instability in all loci tested (n = 2), in 10 loci (n = 3), in 9 loci (n = 5), in 8 loci (n = 5), and in 7 or less loci (n = 3). Sixteen out of 18 tumors were unstable at both BAT25 and BAT26 (Fig. 1), whereas an additional case showed no signs of instability in those markers. The most frequently unstable dinucleotide marker of the NCI panel was D2S123 (n = 17), followed by D17S250 (n = 16) and D5S346 (n = 13) (Fig. 1). Among the 6 spare loci screened, D18S58 was the most commonly affected by instability (n = 16), followed by D18S55 (n = 14), D18S64 (n = 13), D18S61 (n = 11), and PCA53 (n = 9).

Among the 33 MSI-L tumors, 26 were found to be unstable at just 1 locus, 6 at 2 loci, and a single one exhibited instability at 3 loci. Instability involved BAT25 (n = 2), BAT26 (n = 2), D5S346 (n = 3), D2S123 (n = 7), and D17S250 (n = 8), and 12 tumors exhibited unstable markers other than those of the NCI panel.

Of the 68 tumors suitable for microsatellite instability analysis that were used to evaluate the performance of the eligible model into clinical practice, 64 tumors (94.1%) were classified as non–MSI-H.

Correlation of Clinicopathological Features With MIN Status

Table 3 summarizes clinical data and histopathologic variables of the patients analyzed. Univariate analysis revealed that 7 out of the 18 categorical variables studied were significantly associated with MIN status (Table 3). The factors more tightly linked to MSI-H colorectal cancers were proximal tumor location, poor differentiation (grade 3), presence of TIL, and peritumoral Crohn-like reaction ($P < 0.001$). Other traits also related to unstable tumors were expansive tumor growth pattern ($P = .003$), Astler-Coller's tumor stage B ($P = 0.043$), and absence of intramural TWVI ($P = 0.045$). Regarding numerical features, 5 out of 11 were found to be related to MIN status (Table 3). Microsatellite unstable lesions were significantly larger than stable ones ($P = 0.022$), more solid ($P = 0.045$), and more mucinous ($P < 0.001$). MSI-H tumors also exhibited a higher Ki67 proliferative index ($P = 0.002$) and a lower p53 positive phenotype ($P = 0.016$) when compared with non–MSI-H colorectal carcinomas.

Decision Tree–Based Analyses

An initial training approach to reproduce Jass' model[23] with our data provided the following results: for tumors exhibiting less than 50% of mucinous differentiation (adenocarcinomas), the algorithm yielded an overall accuracy of 86.8%, with a sensitivity of 77.8%, a specificity of 87.3%, and a negative predictive value of 98.3%. For those tumors with 50% or more mucinous differentiation (mucinous adenocarcinomas), accuracy was estimated to be 56.7%, with a sensitivity 88.9%, a specificity 42.9%, and a negative predictive value of 90.0%. Finally, the global accuracy for all the tumors was 82.3%.

Table 4 summarizes the results obtained from the 10-fold cross-validation analyses of the decision tree generated using our series of tumors, after pruning the redundant nodes. Five categorical plus 6 numerical variables were found to be associated with MIN status at least once after analyzing the 10 different partitions. The node's positions of these variables in each of the trees are also summarized in Table 4. This preliminary approach pointed at Crohn-like reaction and mucinous differentiation as the main factors implicated in the

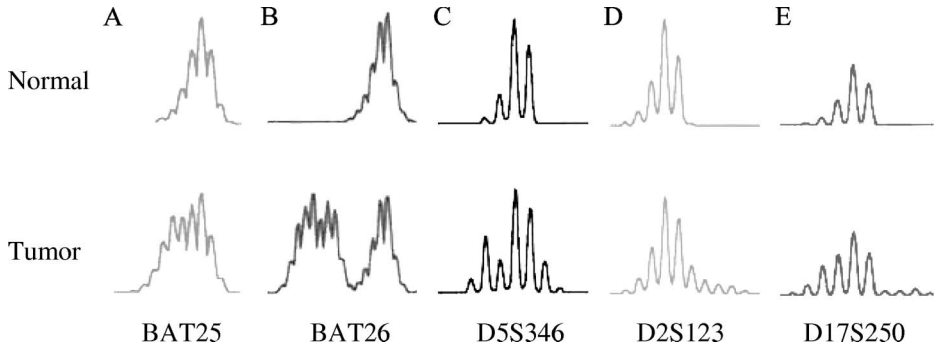

FIGURE 1. Electropherograms obtained from microsatellite instability testing using the five microsatellite markers of the NCI panel. Representative appearance of a tumor exhibiting MSI-H. For each marker, the top graph represents normal DNA and the bottom graph represents tumor DNA. (A) BAT25, (B) BAT26, (C) D5S346, (D) D2S123, and (E) D17S250.

definition of MIN status in the colorectal tumors studied. Nevertheless, the decision tree-based approach was soon discarded due to its lack of robustness. In Figure 2, trees from sets 3 and 8 have been plotted to illustrate the inconsistency of the variables involved in the algorithms and the swinging of node's position within the trees.

Logistic Regression-Based Model

After disposing of the histologic grade (see Discussion), 8 out of the 28 variables tested were found to be related to MIN status. Representative examples of some of these morphologic and immunohistochemical related features are illustrated in Figures 3 and 4, respectively. Among categorical features, proximal tumor location, expansive growth pattern (Fig. 3A), and peritumoral Crohn-like reaction (Fig. 3B) resulted in association with MSI-H tumors. Regarding numerical features, solid differentiation, mucinous differentiation (Fig. 3C), and Ki67 proliferative index (Fig. 4A) were found to be directly proportional to instability. On the contrary, occurrence of cribriform structures (Fig. 3D) and p53-positive phenotype (Fig. 4B) were both traits inversely proportional to MSI-H cancers. Regression coefficients (β_i) assigned to the 8 variables that exhibited relationship to MIN are summarized in Table 5. The best-fitting model was obtained by substitution of the β_i coefficients for their estimates into formula,[2] as follows:

$$X = -2.648 - 2.401x_1 + 0.021x_2 + 0.045x_3 - 1.884x_4 + 0.050x_5 - 0.029x_6 - 0.118x_7 - 1.600x_8$$

An estimated probability of a tumor being MSI-H higher than 29%, that is, a value of $P > 0.29$ in formula, was empirically found to be the best cut point to discriminate between MSI-H and non–MSI-H cancers. Table 5 shows the probability calculation of a tumor being MSI-H as an example of implementation of the logistic regression model into clinical practice.

This regression equation was capable of correctly classify 194 out of 204 tumors, which represents an accuracy of 95.1%. Six of the 10 misclassified tumors resulted false positives and the other 4 corresponded to false negatives. Overall, the model had a sensitivity of 77.8% and a specificity of 96.8%, being the positive and negative predictive values of 70.0% and 97.8%, respectively. After the 10-fold cross-validation analysis, the equation was still able to correctly classify all but 19 tumors (90.7%), including 8 false-negative (negative predictive value of 95.6%) and 11 false-positive (positive predictive value of 47.6%) lesions.

Finally, when the model was prospectively tested using a validation cohort of 68 naive tumors (4 MSI-H and 64 non–MSI-H), a sensitivity of 75.0%, a specificity of 93.8%, a positive predictive value of 42.9%, and a negative predictive value of 98.4% were obtained, with an overall accuracy of 92.7%.

TABLE 3. Univariate Analysis to Correlate Clinicopathological Features With MIN Status of Colorectal Adenocarcinomas: Categorical Variables and Numerical Variables

Categorical Variable (n = 204)	MSI-H (n = 18)	Non-MSI-H (n = 186)	*P* Value*
Gender			
Male	8 (44.4)	111 (59.7)	0.222
Female	10 (55.6)	75 (40.3)	
Location			
Proximal	14 (77.8)	58 (31.2)	<0.001
Distal	4 (22.2)	128 (68.8)	
Configuration			
Exophytic	9 (50.0)	74 (39.8)	
Ulcerated	7 (38.9)	73 (39.2)	0.545
Stenosing	2 (11.1)	39 (21.0)	
Grade (WHO)			
1	3 (16.7)	73 (39.2)	
2	2 (11.1)	80 (43.0)	<0.001
3	13 (72.2)	33 (17.7)	
Extent of invasion (TNM)			
pT1	0	13 (7.0)	
pT2	3 (16.7)	23 (12.4)	0.482
pT3	11 (61.1)	91 (48.9)	
pT4	4 (22.2)	59 (31.7)	
Stage (Astler-Coller's)			
A	0	10 (5.4)	
B	13 (72.2)	73 (39.2)	0.043
C	5 (27.8)	85 (45.7)	
D	0	18 (9.7)	
Intramural TWVI			
Present	3 (16.7)	77 (41.4)	0.045
Absent	15 (83.3)	109 (58.6)	
Extramural TWVI			
Present	3 (16.7)	66 (35.5)	0.124
Absent	15 (83.3)	120 (64.5)	
Intramural VVI			
Present	2 (11.1)	11 (5.9)	0.321
Absent	16 (88.9)	175 (94.1)	
Extramural VVI			
Present	3 (16.7)	38 (20.4)	1.000
Absent	15 (83.3)	148 (79.6)	
Intramural PNI			
Present	0	7 (3.8)	1.000
Absent	18 (100.0)	179 (96.2)	
Extramural PNI			
Present	0	22 (11.8)	0.228
Absent	18 (100.0)	164 (88.2)	
Growth pattern			
Expansive	12 (66.7)	56 (30.1)	0.003
Infiltrative	6 (33.3)	130 (69.9)	
Crohn-like lymphoid reactivity			
Present	13 (72.2)	37 (19.9)	<0.001
Absent	5 (27.8)	149 (80.1)	
TIL			
Present	9 (50.0)	21 (11.3)	<0.001
Absent	9 (50.0)	165 (88.7)	

DISCUSSION

The assessment of MIN status in colorectal cancers has been based on genetic assays, mainly PCR tests. However, DNA extraction, amplification, and subsequent analysis of PCR products from primary tumors are time-consuming steps,

TABLE 3. (*continued*) Univariate Analysis to Correlate Clinicopathological Features With MIN Status of Colorectal Adenocarcinomas: Categorical Variables and Numerical Variables

Categorical Variable (n = 204)	MSI-H (n = 18)	Non-MSI-H (n = 186)	*P* Value*
Adenomas			
Present	5 (27.8)	49 (26.3)	1.000
Absent	13 (72.2)	137 (73.7)	
Non adenomatous polyps			
Present	3 (16.7)	21 (11.3)	0.451
Absent	15 (83.3)	165 (88.7)	
Residual adenoma			
Present	6 (33.3)	53 (28.5)	0.786
Absent	12 (66.7)	133 (71.5)	
Numerical variable (n = 204)	**MSI-H ($\bar{x}$ ± SD)†**	**Non-MSI-H ($\bar{x}$ ± SD)†**	***P* Value***
Age (years)	67.6 ± 14.1	67.8 ± 11.9	0.940
Tumor size (mm ø maximum)	51.3 ± 10.5	30.9 ± 20.6	0.022
Solid carcinoma (%)	19.4 ± 28.3	10.5 ± 16.7	0.045
Mucinous carcinoma (%)	40.8 ± 31.4	10.9 ± 24.1	<0.001
Signet-ring cell carcinoma (%)	0.8 ± 2.6	0.5 ± 3.7	0.672
Cribriform structures (%)	2.2 ± 5.2	6.9 ± 15.1	0.193
Micropapillary pattern (%)	1.1 ± 3.2	3.6 ± 10.7	0.320
Infiltrating microglandular pattern (%)	0.8 ± 2.6	4.4 ±11.8	0.204
Nodal involvement (n)	3.6 ± 8.8	2.2 ± 3.3	0.170
Ki67 proliferative index (%)	70.8 ± 13.9	53.9 ± 22.1	0.002
p53 overexpression (%)	15.3 ± 22.6	35.5 ± 34.5	0.016

**P* values of χ^2 test, Fisher test, and logistic regression were considered statistically significant when less than 0.05.

†$\bar{x}$ is the arithmetic mean, while SD relates to standard deviation.

involving complex processes that require specialized equipment and trained personnel. Because such molecular techniques are not implemented in most histopathology laboratories, MIN testing in most colorectal cancers is not performed. Nevertheless, MSI phenotype remains a predictive marker of good prognosis[1,2,13,14,19,28,29,44,45] and differential response to adjuvant chemotherapy.[20–22] The present study was conducted to generate a model for prediction of MIN status in colorectal cancers that could minimize routinely molecular MIN testing.

Jass et al[23] had reported that histopathological examination of colorectal cancer could define MSI-H tumors. These investigators proposed a decision tree–based model that performed with reasonable specificity and sensitivity. Nevertheless, they also underlined that the relative small sample size (n = 120) analyzed and the subjectivity of some of the variables used could limit the clinical applicability of the algorithm. Yet, when we used such approach in our cohort, we obtained similar results to those previously published. In our series, colorectal adenocarcinomas other than mucinous lesions could be correctly classified as non–MSI-H (negative predictive value of 98.1%), whereas mucinous adenocarcinomas failed to exhibit a reliable negative predictability (90.0%). It was based on these findings that we resolved to generate a novel model addressing previous limitations. However, despite an improvement in sensitivity, specificity, and accuracy, the initial device did not display enough robustness, as deduced from the 10-fold cross-validation analysis performed (Table 4, Fig. 2). As an alternative, logistic regression was used as an approach to obtain a best-fitting model. The rationale to apply mathematical formulation was to routinely estimate the probability of a tumor being stable by the time the histopathology report was being assembled.

In the present logistic approach, tumor histology grade was not included as a variable in the explanatory set. This was due, in part, to the functional relationship of this variable with solid and mucinous patterns because these 2 features are actually part of the WHO grading system of colorectal tumors.[29] As it is well known in regression analysis, a high dependence between multiple variables, termed collinearity, may produce undesired effects in the stepwise procedure, leading to bias of coefficient estimates.

Results obtained through the stepwise analysis did not substantially differ from those reported by Jass et al[23] using a decision tree–based model. Four out of the 6 factors involved

TABLE 4. Ten-Fold Cross-Validation Results of Decision Tree Trained on Our Series

Variable	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Set 10
Tumor location		3								
Solid differentiation	3		3					3	5	
Mucinous differentiation	4	2	4	1	1	1	1	2	1 (4)	1
Cribriform pattern		4		4	5					
Nodal involvement				3	3	3	3	4		
Crohn-like lymphoid reactivity	1	1	1							
Ki67 proliferative index		3 (4) (5)		2	4	2	2	3	2	2
Residual adenoma	5		5							
Tumor size	2	2	2		2	4		1 (5)	3	4 (5)
Adenomas										6
Infiltrating microglandular pattern										3

Note. Each of the columns shows the results of a disjoint subset of randomly selected tumors. The numbers found in each cell correspond to the node's position (eg, number 1 is the root node, numbers 2 are the nodes immediately below the root, etc), being the numbers in brackets lower node's positions of a determinate variable within the same tree.

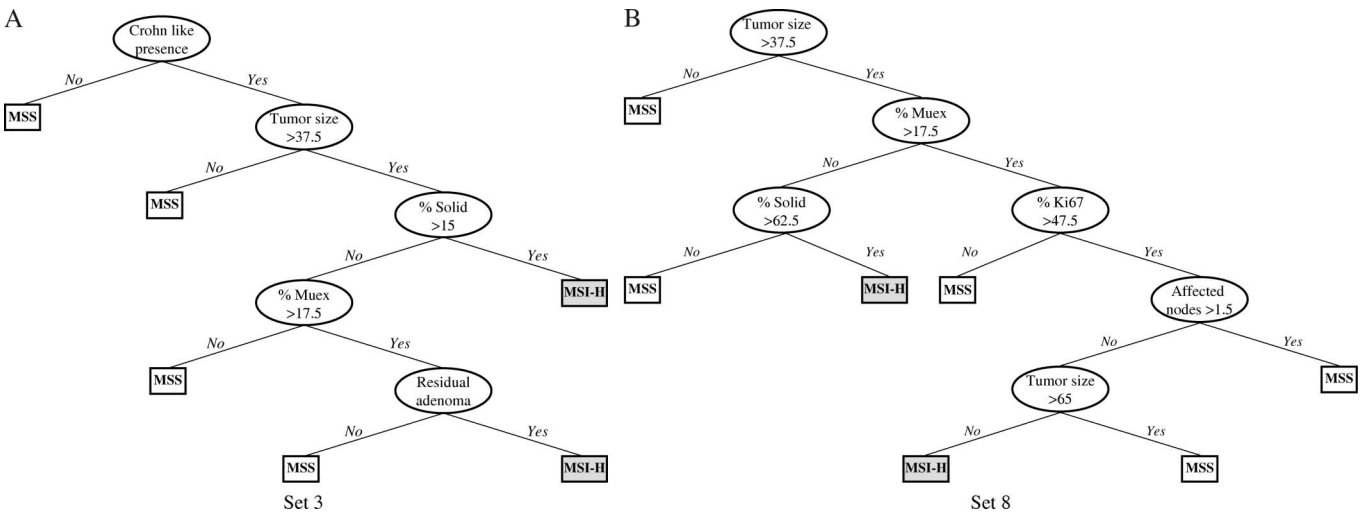

FIGURE 2. Two examples of the algorithms obtained from the 10-fold cross-validation analysis of the decision tree-based model. Data sets 3 and 8 share approximately 80% of cases. The inconsistency of the variables involved in each algorithm, as well as the swinging of node's position within the trees, evidences the lack of robustness of this approach. A, Tree from set 3. B, Tree from set 8.

in Jass' algorithm were found to be explanatory variables in our model of best fit, including tumor type (equivalent to our percentage of mucinous pattern), histologic grade (corresponding to our percentage of solid tumor), peritumoral Crohn-like reaction, and tumor location. Neither TIL nor clinical staging were found to be predictive factors in our model. Accumulating evidence indicates that MSI-H tumors are characterized by pronounced TILs.[25,26,46,47] Certainly, univariate analysis of our dataset revealed that TIL correlates extremely well with MIN status by itself but not when other variables are included in the model. This is consistent with the fact that TIL is related to other tested variables, mainly Crohn-like reactivity and tumor location, becoming insignificant when better predictors are placed onto the model. In any event, we estimate this finding particularly advantageous because TIL is a difficult feature to assess by simply H&E staining. As a matter of fact, TIL evaluation can be dramatically improved with the use of immunohistochemical assays using antibodies to T-lymphocyte surface markers (eg, $CD3^+$, $CD8^+$).[24,26]

Four explanatory variables arose from our multivariate logistic analysis that were not identified in Jass' model: growth pattern, cribriform pattern, Ki67 proliferative index, and identification of a p53 phenotype. Of these variables, the first 2 were included in Jass' study as potential predictive factors. The expanding growth at the advancing tumor margin had been found to be associated with MSI-H tumors.[44,45,48,49]

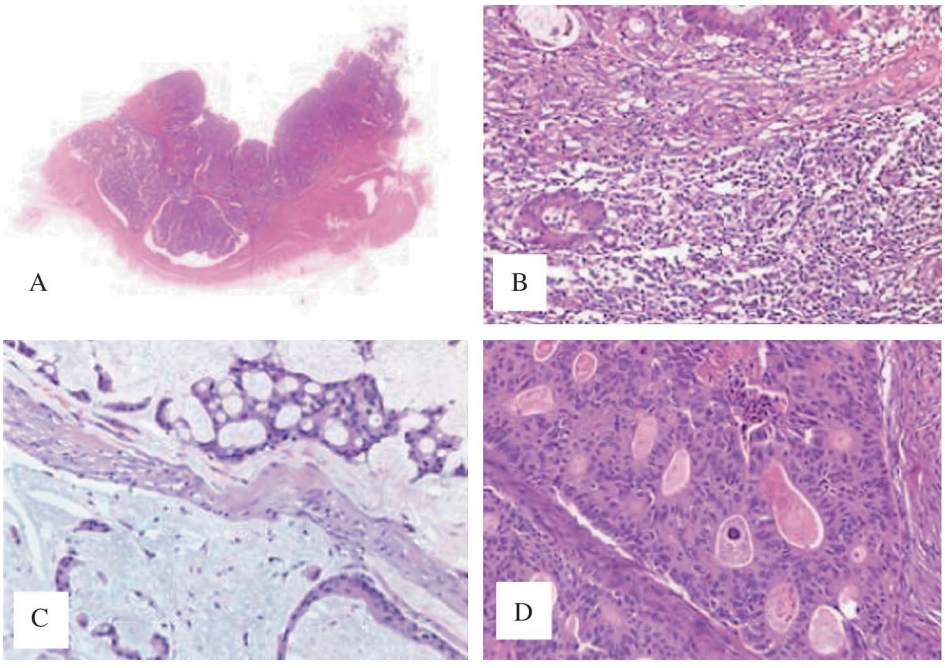

FIGURE 3. Representative examples of morphologic features found to be associated to MSI-H colorectal cancers. A, Expansive growth pattern: well-defined smooth margin (H&E, whole mounting). B, Dense diffuse peritumoral lymphoid infiltrate in a Crohn-like pattern (H&E, 200×). C, Mucinous differentiation: acinar structures and strips of neoplastic cells in a background of extracellular mucin (H&E, 200×). D, Glandular structures lacking intervening stroma growing in a cribriform pattern (H&E, 200×).

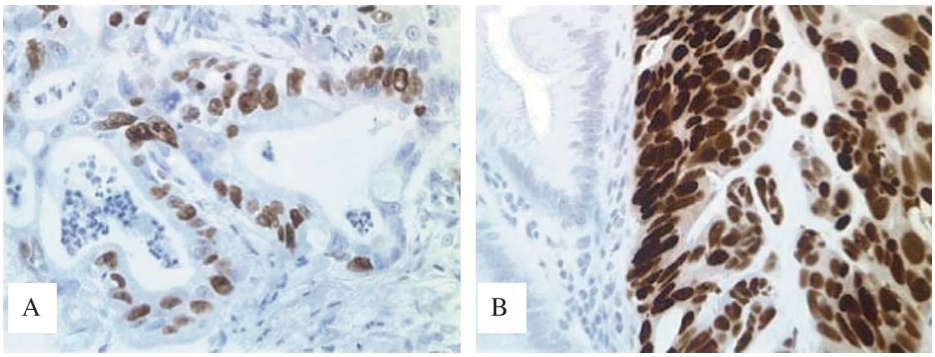

FIGURE 4. Representative examples of immunohistochemical features found to be associated to MSI-H colorectal cancers. A, High proliferative index in a MSI-H case (MIB-1 immunostaining, 200×). B, p53 overexpression in about 100% of neoplastic cells (right side) in contrast with negative nuclei in non tumoral crypts (DO-7 immunostaining, 200×).

Furthermore, this variable is accepted as an independent favorable prognosis factor.[32] Cribriform growth had not been considered in the context of MIN status. Results from our logistic approach revealed the significant association between the presence of cribriform structures and non–MSI-H tumors, as reported previously by Alexander et al.[27] This is not surprising if such cribriform pattern is defined as a glandular growth and, in consequence, a sign of differentiation. Conversely, other reports[23,48] suggested that the occurrence of cribriform structures existed within some poorly differentiated TIL-rich and well-circumscribed adenocarcinomas, thus suggesting an indirect linkage of this trait with MSI-H cancers.[44,49,50] We assign the disparity of reports to the misuse of the term "cribriform," mainly when dealing with undifferentiated solid tumors.

Identification of a p53-positive phenotype and Ki67 proliferative index, both biomarkers routinely assessed in most laboratories when studying colorectal cancers, were immunophenotypical features included in our logistic approach. These features had been previously reported to significantly correlate with MIN status. Results from the univariate analysis performed showed that MSI-H cancers had a lower percentage of p53-positive phenotype than non–MSI-H cancers, in accordance with previously reported studies.[24,44,49,51] Regarding Ki67, we found that MSI-H cancers exhibited a significantly higher proliferative index than non–MSI-H, consistent with other reports.[34,52] This higher proliferative index may be due to large proportion of poorly differentiated cancers within this group of unstable lesions. This finding matches with the good behavior ascertained to MSI-H tumors because the increased cell growth of such cancers appears counterbalanced by the high apoptotic index.[34]

In sum, the logistic model proposed posses a high overall accuracy in predicting microsatellite instability. The

TABLE 5. Explanatory Variables of the Logistic Regression Model

Explanatory Variable	i*	β_i	95% Confidence Limits	*P* Value†
Crohn-like lymphoid reactivity‡	1	−2.401	0.019–0.426	0.002
Mucinous differentiation	2	0.021	0.999–1.044	0.065
Solid differentiation	3	0.045	1.009–1.085	0.016
Growth pattern§	4	−1.884	0.032–0.727	0.018
Ki67 proliferative index	5	0.050	1.008–1.097	0.061
p53 overexpression	6	−0.029	0.943–1.001	0.021
Cribriform pattern	7	−0.118	0.786–1.005	0.061
Tumor location‖	8	−1.600	0.040–1.011	0.052
Constant	—	−2.648 (β_0)	—	—

*i relates to the step at which a particular variable is introduced into the model.

†*P* value at the last step of the backward stepwise analysis ($P_{in} = 0.05$ and $P_{out} = 0.10$). The forward approach had equivalent outcome ($P_{in} = 0.10$ and $P_{out} = 0.15$).

‡Crohn-like lymphoid reactivity: (0 = presence; 1 = absence).

§Growth pattern: (0 = expansive; 1 = infiltrative).

‖Tumor location: (0 = proximal; 1 = distal).

Example: A MSS proximal tumor ($x_8 = 0$) that exhibits Crohn like lymphoid reactivity ($x_1 = 0$) and expansive growth pattern ($x_4 = 0$), with 15% mucinous differentiation ($x_2 = 15$), no solid differentiation ($x_3 = 0$), 80% occurrence of cribriform structures ($x_7 = 80$), has 70% Ki67 immunostaining ($x_5 = 70$), and a negative p53 phenotype ($x_6 = 0$). The probability of this tumor being MSI-H can be estimate according to the formula:

$$P = 1 - \left(\frac{1}{1 + e^{X}}\right),$$

where the development of $X = -2.648 - (2.401 \times 0) + (0.021 \times 15) + (0.045 \times 0) - (1.884 \times 0) + (0.050 \times 70) - (0.029 \times 0) - (0.118 \times 80) - (1.600 \times 0)$

The use of the predictive tool gives a $P = 0.0003$, which is below the optimal cut point of 0.29, thus leading to a prediction of MSS tumor.

new equation was able to correctly classify all but 10 tumors (95.1%), including 6 false-positive (positive predictive value of 70.0%) and 4 false-negative (negative predictive value of 97.8%) lesions. In this context, we consider that clinicopathological features may not be used as a substitute of molecular methods to assess MIN status. Some authors have reported[27,46] that there is not only a considerable interlaboratory variability with regard to tissue processing and DNA testing but also a significant interobserver variability in the interpretation of certain morphology findings, which may account for the failure of predictions. Thus, our goal was to restrict MIN testing to lesions with suspected instability. The 6 false-positive cases of this series could skip misclassification. The committed error would be reduced to 4 out of 204 tumors, which represents less than 2% of misclassified cases, by just testing 9.8% of the series (20 cases). Moreover, this error would only rise to 8 out of 204 tumors in the worse case, as deduced from the 10-fold cross-validation analysis, which represents 3.9% of misclassified cases, by testing 10.3% of the series (21 cases).

Furthermore, the value of our logistic model is its capacity of well-predicting new cases. As a matter of fact, the model was prospectively tested in 68 cases, rendering a sensitivity of 75.0%, a specificity of 93.8%, a positive predictive value of 42.9%, and a negative predictive value of 98.4%. The approach combining the logistic model with instability testing of positively predicted cases resulted thus in less than 2% of misclassified new cases. The easy of handling the new equation (as shown in the example below Table 5), once implemented in a program, makes this approach attractive for routine practice. We believe that this logistic model can assist in discerning the profile of colorectal cancers, leading to the identification of patients that are more likely to benefit from adjuvant therapy, and thus tailoring individual treatments. This tool, in concert with Amsterdam criteria, may also contribute in a cost-benefit approach to the identification and clinical management of HNPCC patients.

ACKNOWLEDGMENT

The authors thank Eva Torija from BIOPAT for her secretarial assistance in data collection.

REFERENCES

1. Dunican DS, McWilliam P, Tighe O, et al. Expression differences between the microsatellite instability (MIN) and chromosomal instability (CIN) phenotypes in colorectal cancer revealed by high-density cDNA array hybridization. *Oncogene*. 2002;21:3253–3257.
2. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990;61:759–767.
3. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. *Cell*. 1996;87:159–170.
4. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature*. 1998;396:643–649.
5. Ionov Y, Peinado MA, Malkhosyan S, et al. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature*. 1993;363:558–561.
6. Shibata D, Peinado MA, Ionov Y, et al. Genomic instability in repeated sequences is an early somatic event in colorectal tumor genesis that persists after transformation. *Nat Genet*. 1994;6:273–281.
7. Markowitz S, Wang J, Myeroff L, et al. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science*. 1995;268:1336–1338.
8. Souza RF, Appel R, Yin J, et al. Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumours. *Nat Genet*. 1996;14:255–257.
9. Rampino N, Yamamoto H, Ionov Y, et al. Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science*. 1997;275:967–969.
10. Thibodeau SN, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon. *Science*. 1993;260:816–819.
11. Lothe RA, Peltomaki P, Meling GI, et al. Genomic instability in colorectal cancer: Relationship to clinicopathological variables and family history. *Cancer Res*. 1993;53:5849–5852.
12. Lukish JR, Muro K, DeNobile J, et al. Prognostic significance of DNA replication errors in young patients with colorectal cancer. *Ann Surg*. 1998;227:51–56.
13. Gryfe R, Kim H, Hsieh ET, et al. Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. *N Engl J Med*. 2000;342:69–77.
14. Halling KC, French AJ, McDonnell SK, et al. Microsatellite instability and 8p allelic imbalance in stage B2 and C colorectal cancers. *J Natl Cancer Inst*. 1999;91:1295–1303.
15. Samowitz WS, Curtin K, Ma KN, et al. Microsatellite instability in sporadic colon cancer is associated with an improved prognosis at the population level. *Cancer Epidemiol Biomarkers Prev*. 2001;10: 917–923.
16. Watanabe T, Wu TT, Catalano PJ, et al. Molecular predictors of survival after adjuvant chemotherapy for colon cancer. *N Engl J Med*. 2001;344: 1196–1206.
17. Hemminki A, Mecklin JP, Jarvinen H, et al. Microsatellite instability is a favorable prognostic indicator in patients with colorectal cancer receiving chemotherapy. *Gastroenterology*. 2000;119:921–928.
18. Wright CM, Dent OF, Barker M, et al. Prognostic significance of extensive microsatellite instability in sporadic clinicopathological stage C colorectal cancer. *Br J Surg*. 2000;87:1197–1202.
19. Elsaleh H, Joseph D, Grieu F, et al. Association of tumour site and sex with survival benefit from adjuvant chemotherapy in colorectal cancer. *Lancet*. 2000;355:1745–1750.
20. Carethers JM, Smith EJ, Behling CA, et al. Use of 5-fluorouracil and survival in patients with microsatellite-unstable colorectal cancer. *Gastroenterology*. 2004;126:394–401.
21. Parc Y, Gueroult S, Mourra N, et al. Prognostic significance of microsatellite instability determined by immunohistochemical staining of MSH2 and MLH1 in sporadic T3N0M0 colon cancer. *Gut*. 2004;53:371–375.
22. Ribic CM, Sargent DJ, Moore MJ, et al. Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. *N Engl J Med*. 2003;349:247–257.
23. Jass JR, Do KA, Simms LA, et al. Morphology of sporadic colorectal cancer with DNA replication errors. *Gut*. 1998;42:673–679.
24. Kim H, Jen J, Vogelstein B, et al. Clinical and pathological characteristics of sporadic colorectal carcinomas with DNA replication errors in microsatellite sequences. *Am J Pathol*. 1994;145:148–156.
25. Dolcetti R, Viel A, Doglioni C, et al. High prevalence of activated intraepithelial cytotoxic T lymphocytes and increased neoplastic cell apoptosis in colorectal carcinomas with microsatellite instability. *Am J Pathol*. 1999;154:1805–1813.
26. Michael-Robinson JM, Biemer-Huttmann A, Purdie DM, et al. Tumour infiltrating lymphocytes and apoptosis are independent features in colorectal cancer stratified according to microsatellite instability status. *Gut*. 2001;48:360–366.
27. Alexander J, Watanabe T, Wu TT, et al. Histopathological identification of colon cancer with microsatellite instability. *Am J Pathol*. 2001;158: 527–535.
28. Graham DM, Appelman HD. Crohn's-like lymphoid reaction and colorectal carcinoma: A potential histologic prognosticator. *Mod Pathol*. 1990;3:332–335.
29. Hamilton SR, Aaltonen LA, eds. *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Digestive System*. Lyon; IARC Press; 2000:103–143.
30. Sobin LH, Wittekind Ch, eds. *TNM Classification of Malignant Tumours*. 6th edition. New York: Wiley-Liss; 2002.

31. Astler VB, Coller FA. The prognostic significance of direct extension of carcinoma of the colon and rectum. *Ann Surg*. 1954;139:846–852.
32. Compton CC. Colorectal carcinoma: Diagnostic, prognostic, and molecular features. *Mod Pathol*. 2003;16:376–388.
33. Colomer A, Erill N, Verdu M, et al. Lack of p53 nuclear immunostaining is not indicative of absence of TP53 gene mutations in colorectal adenocarcinomas. *Appl Immunohistochem Mol Morphol*. 2003;11:130–137.
34. Michael-Robinson JM, Reid LE, Purdie DM, et al. Proliferation, apoptosis, and survival in high-level microsatellite instability sporadic colorectal cancer. *Clin Cancer Res*. 2001;7:2347–2356.
35. Purdie CA, O'Grady J, Piris J, et al. p53 expression in colorectal tumors. *Am J Pathol*. 1991;138:807–813.
36. Shibata DK, Arnheim N, Martin WJ. Detection of human papilloma virus in paraffin-embedded tissue using the polymerase chain reaction. *J Exp Med*. 1988;167:225–230.
37. Bauer HM, Ting Y, Greer CE, et al. Genital human papillomavirus infection in female university students as determined by a PCR-based method. *JAMA*. 1991;265:472–477.
38. Boland CR, Thibodeau SN, Hamilton SR, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: Development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res*. 1998;58:5248–5257.
39. Jen J, Kim H, Piantadosi S, et al. Allelic loss of chromosome 18q and prognosis in colorectal cancer. *N Engl J Med*. 1994;331:213–221.
40. Jones MH, Nakamura Y. Detection of loss of heterozygosity at the human TP53 locus using a dinucleotide repeat polymorphism. *Genes Chromosomes and Cancer*. 1992;5:89–90.
41. Breiman L, Friedman JH, Olshen RA, et al. In: Bickel P, Cleveland W, Dudley R, eds. *Classification and Regression Trees*. Belmont: Wadsworth International Group; 1984.
42. Ripley BD. *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press; 1996.
43. Weiss SM, Kulikowski CA. *Computer Systems That Learn, Classification and Prediction Methods From Statistics, Neural Nets, Machine Learning and Expert Systems*. San Mateo: Morgan Kaufmann Publishers; 1991.
44. Edmonston TB, Cuesta KH, Burkholder S, et al. Colorectal carcinomas with high microsatellite instability: Defining a distinct immunologic and molecular entity with respect to prognostic markers. *Hum Pathol*. 2000; 31:1506–1514.
45. Gafa R, Maestri I, Matteuzzi M, et al. Sporadic colorectal adenocarcinomas with high-frequency microsatellite instability. *Cancer*. 2000;89: 2025–2037.
46. Shia J, Ellis NA, Paty PB, et al. Value of histopathology in predicting microsatellite instability in hereditary nonpolyposis colorectal cancer and sporadic colorectal cancer. *Am J Surg Pathol*. 2003;27:1407–1417.
47. Smyrk TC, Watson P, Kaul K, et al. Tumor-infiltrating lymphocytes are a marker for microsatellite instability in colorectal carcinoma. *Cancer*. 2001;91:2417–2422.
48. Jass JR. Towards a molecular classification of colorectal cancer. *Int J Colorectal Dis*. 1999;14:194–200.
49. Suh JH, Lim SD, Kim JC, et al. Comparison of clinicopathologic characteristics and genetic alterations between microsatellite instability-positive and microsatellite instability-negative sporadic colorectal carcinomas in patients younger than 40 years old. *Dis Colon Rectum*. 2002;45:219–228.
50. Parc YR, Halling KC, Burgart LJ, et al. Microsatellite instability and hMLH1/hMSH2 expression in young endometrial carcinoma patients: Associations with family history and histopathology. *Int J Cancer*. 2000; 86:60–66.
51. Samowitz WS, Holden JA, Curtin K, et al. Inverse relationship between microsatellite instability and K-ras and p53 gene alterations in colon cancer. *Am J Pathol*. 2001;158:1517–1524.
52. Takagi S, Kumagai S, Kinouchi Y, et al. High Ki-67 labeling index in human colorectal cancer with microsatellite instability. *Anticancer Res*. 2002;22:3241–3244.

