Hierarchical clustering analysis of pathologic and molecular data identifies prognostically and biologically distinct groups of colorectal carcinomas

Daniela Furlan[1,*], Ileana W Carnevali[1,*], Barbara Bernasconi[1], Nora Sahnane[1], Katia Milani[1], Roberta Cerutti[1], Valentina Bertolini[2], Anna Maria Chiaravalli[1], Francesco Bertoni[3], Ivo Kwee[3], Roberta Pastorino[4] and Capella Carlo[1]

[1]Department of Human Morphology, Section of Anatomic Pathology, Centro Insubre di Biotecnologie per la Salute Umana, University of Insubria and Ospedale di Circolo, Varese, Italy; [2]IRCCS-MultiMedica, Section of Anatomic Pathology, Milan, Italy; [3]Laboratory of Experimental Oncology, Oncology Institute of Southern Switzerland, Bellinzona, Switzerland and [4]Department of Applied Health Sciences, University of Pavia, Pavia, Italy

This work has evaluated the potential superiority of a morphomolecular classification based on the combination of clinicopathologic and molecular features of colorectal cancers. A cohort of 126 colorectal carcinomas was investigated by unsupervised hierarchical clustering analysis to combine 13 routinely assessed clinicopathologic features and all five molecular markers recently suggested by Jass' classification to distinguish four molecular subtypes of sporadic colorectal carcinomas. Survival analysis was assessed by a Cox proportional hazards model. A clear separation into three prognostically significant groups was identified: cluster A and cluster C were associated with good prognosis and cluster B with poor prognosis ($P = 0.006$). Clinicopathologic and molecular features of cluster A and cluster B tumors were strongly concordant with colorectal cancer profiles characterized by microsatellite instability or by chromosomal instability, respectively. The clinicopathologic features of cluster C tumors were suggestive of a less aggressive disease than cluster B tumors. Genetically, they appeared intermediate between cluster A and cluster B tumors, as they were mainly microsatellite stable tumors showing high levels of both MGMT methylation and loss of heterozygosity. Chromosomal instability was significantly lower in cluster C than in cluster B tumors. A more accurate tumor classification should combine the prognostic power of clinicopathologic parameters with molecular biomarkers that provide information regarding the natural history of the cancer. Hierarchical clustering seems to be a useful, promising and powerful tool for further translational studies and should lead us to define a diagnostic and prognostic signature for different carcinomas.

The classification of tumors from all sites is continuously evolving and improvements are central to advances in cancer prognosis as well as in

Correspondence: Dr D Furlan, BSc, Department of Human Morphology, Anatomic Pathology Unit, University of Insubria, Via O. Rossi, 9, 21100 Varese, Italy.
E-mail: daniela.furlan@uninsubria.it
*These authors contributed equally to this work.

clinical decisions about the best-treatment plan for patients. Currently, tumor-node-metastasis (TNM) staging remains the gold standard for prognostic classification of colorectal cancer even though this system offers little information about response to treatment in individual patients or reasons for treatment failure.

During the last 20 years, there have been significant advances in our understanding of colorectal cancer pathogenesis, and it is now clear that these tumors encompass a heterogeneous complex of

disease that reflects different underlying mechanisms of carcinogenesis.[1] Recently, some authors have used this new knowledge about colorectal carcinoma pathogenesis to construct a molecular classification of colorectal carcinomas by integrating genetic and epigenetic data.[2,3] However, very few data about the clinicopathologic profiles of the colorectal carcinomas examined were reported in these studies, so it is not clear whether a molecular classification would yield more accurate prognosis analyses than the traditional TNM staging system. Many molecular markers are actively being investigated for their potential prognostic value but none of them has consistently reached the prognostic significance of TNM staging in multivariate analysis and there is no consensus about their clinical use.[4] In addition, no study has so far evaluated the potential superiority of a classification based on the combination of clinicopathologic and genetic/epigenetic parameters.

Jass[5] proposed a new classification of colorectal carcinomas based on the correlation of morphological and molecular features. The authors proposed four molecular subtypes of sporadic colorectal carcinomas with specific morphological features. These groups are defined according to (1) the underlying types of genetic instability (microsatellite instability or chromosomal instability), (2) *MGMT* and *hMLH1* methylation status, and (3) *BRAF* and *KRAS* mutations. Although this system is the first example of a colorectal carcinoma classification including both morphological and molecular aspects, no information about the potential prognostic significance of the classification is yet available.

In this work, we propose a translational study using unsupervised hierarchical clustering analysis to identify an informative colorectal carcinoma subclassification from a combination of routinely assessed clinicopathologic features and a limited number of specific molecular markers, which have been shown to provide information about the underlying types of genetic instability. The aim of this analysis was to determine whether combined pathologic and molecular factors should identify prognostically and biologically distinct groups of colorectal carcinoma patients, and provide useful criteria for a diagnostic setting of colorectal carcinomas.

Materials and methods

Patient Selection

The series comprised 126 patients with primary colorectal carcinoma who underwent curative surgical resection between 2000 and 2004 at the Ospedale di Circolo, Varese, Italy. Patients with stages III and IV received an adjuvant chemotherapy based on 5-fluorouracil according to a protocol in use at the Ospedale di Circolo, Varese. All seven patients with cancer of the infraperitoneal rectum

underwent postoperative radiation therapy. One hundred fifteen colorectal carcinomas were consecutive cases, and the only criterion for inclusion in the study was the availability of sufficient tumor tissue for histopathological and molecular analysis. The other 11 tumors were sporadic colorectal carcinomas exhibiting microsatellite instability at high frequency that had previously been collected in our laboratory.[6] These samples were included in order to increase the number of tumors characterized by this type of genetic instability.

The patients ranged in age from 40 to 91 years, with a mean age of 67 years. Outcome data were collected by consulting clinical records and/or the Tumor Registry of the Lombardy region (Italy) and were available for all but six patients, with median follow-up of 60 months (range 0–110 months). Ethical approval was obtained according to the rules of the Ethics Committee of the Ospedale di Circolo, Varese, Italy.

All cases were histologically reviewed by two independent pathologists (A.M.C., C.C.) according to the WHO classification of tumors of the digestive system[7] and the TNM staging system.[8] The clinicopathologic features of the tumor series are summarized in Table 1.

Molecular Study

Normal and tumor DNA of each patient was obtained from formalin fixed and paraffin-embedded tissues using representative 8-μm sections of tumor samples and of matched normal lymph nodes. Three sections of every specimen were treated twice with xylene, and then washed twice with ethanol. DNA was extracted using a QIAamp® DNA FFPE Tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Neoplastic areas were manually microdissected for DNA extraction and contained at least 70% of tumor cells, to minimize contamination by normal cells.[9]

Microsatellite instability status was assessed by PCR amplification of the five markers recommended by the National Cancer Institute, following the Bethesda guidelines.[10] The PCR fragments were separated using an ABI Prism 310 Genetic Analyser and further analyzed using GeneMapper 4.0 (Applied Biosystems, Foster City, CA, USA). Microsatellite instability at high frequency was scored when at least 40% of the microsatellites analyzed were unstable.

Chromosomal instability status was assessed by loss of heterozygosity analysis of four microsatellite markers (*APC*, *TP53*, D18S363, and D18S474) and by interphasic fluorescent *in situ* hybridization (FISH) analysis to evaluate aneusomies of chromosomes 5, 7, 13, 17, and 18. All microsatellite stable colorectal carcinomas were investigated by loss of heterozygosity analysis using published sequences of primers for each marker (http://www.ncbi.nlm.nih.gov). PCR and post-PCR analyses were carried

Table 1 Clinicopathologic data of the 126 colorectal cancer patients

	No.	*%*
Number of cases	126	
Gender		
Male	69	55
Female	57	45
Age (years)		
≤65	43	34
66–75	51	40
>75	32	26
Site		
Right	48	38
Left[a]	78	62
Histological type		
MUC, MED, PAP, CRIB	45	36
ADC, SRC; UND, SQUAM	81	64
Grade of differentiation		
G1–G2	96	76
G3	30	24
pT stage		
1–2–3	109	86
4	17	14
Lymph node metastasis		
pN0	57	45
pN+	66	55
TNM stage		
Stage I	6	5
Stage II	44	39
Stage III	45	41
Stage IV	17	15
Macroscopic growth		
Polypoid/vegetant	46	36
Ulcerative/infiltrative	80	64
Microscopic growth		
Infiltrative	101	80
Expansive	25	20
Neuroinvasion		
Present	44	35
Absent	82	65
Angio/lymphoinvasion		
Present	47	37
Absent	79	63
Fibrosis		
Present	83	66
Absent	43	34
Tumor infiltrating lymphocytes		
Present	80	63
Absent	46	37
Overall survival		
Alive with no evidence of disease	77	64
Died of disease	38	32
Postoperative death	5	4

Abbreviations: ADC, adenocarcinoma not otherwise specified; CRIB, cribriform; MED, medullary; MUC, mucinous; PAP, papillary; SQUAM, squamous; SRC, signet ring cell; UND, undifferentiated.
[a]27 Rectal carcinomas and 51 cancers from left colon.

out under previously published conditions.[9] An allelic imbalance, suggestive of either an allelic loss or gain, was scored by the ratio of relative allelic peak height in the tumor DNA to relative allelic peak height in the corresponding normal DNA.

Interphasic FISH was performed on extracted nuclei of 29 colorectal carcinoma samples, selected after unsupervised hierarchical clustering analysis. Nuclei were obtained from $15\,\mu$m-thick sections from paraffin blocks using a previously described method.[11] The probe panel included EGR1(5q31)/D5S23, D5S721(5p15.2), *TP53*(17p13)/alpha17, D13S319(13q14.3)/13q34, alpha18 (Abbott Molecular Europe, Wiesbaden, Delkenheim, Germany) and *EGFR*(7p11.2)/alpha7 (Dakocitomation, Copenhagen, Denmark A/S). FISH analyses were carried out using the method previously described by Tibiletti *et al.*[12] Each FISH experiment was evaluated blindly by two independent operators and only experiments with 90% hybridization efficiency were considered. In each case and for each probe at least 200 nuclei were examined. FISH signals were counted with a Leica DMRA microscope equipped with single- and triple-band pass filters. The FISH evaluation was performed using the recommendations of the Italian Society of Human Genetics (http://www.sigu.net). Nuclei extracted from 10 normal colic mucosa were used as a control. The cutoff values were calculated as the mean value plus 3 s. d. assuming a binomial distribution of the spots. Initially, the cutoff value was evaluated for each probe; however, because no differences were observed, a common cutoff value was used, namely 3% for trisomy, tetrasomy, and polysomies.

DNA sequencing was used to evaluate mutations in the *KRAS* and *BRAF* genes. Genomic DNA from tumor samples was amplified by PCR using primers of the exon 15 of *BRAF* gene (5′-TGTTTTCCTTTAC TTACTACACCTCA and 5′-GGCCAAAAATTTAATC AGTGGA) and of the exon 2 of *KRAS* gene (5-TCA TTATTTTTATTATAAGGCCTGCT and 5′-GGTCCT GCACCAGTAATATGC). The PCR products were purified by using Microcon YM-50 Centrifugal Filter Units (Millipore, Billerica, MA, USA) according to the manufacturer's instructions. The purified PCR products were sequenced on an ABI Prism 310 Genetic Analyser (Applied Biosystems) with the ABI Prism Big Dye Terminator Cycle Sequencing kit version 1.1 (Applied Biosystems). The primers used for amplification were also used for sequencing. The sequencing results were analyzed using DNA Sequencing Analysis Software version 5.3 (Applied Biosystems). All mutations were confirmed by analysis in both the forward and reverse directions.

The methylation status of the CpG-rich 5′-region of the *MGMT* gene was assessed using a semiquantitative fluorescence-based real-time detection method.[13] Briefly, 300–600 ng of genomic DNA of tumor samples were subjected to sodium bisulphite treatment using the Epitect Bisulfite Kit (Qiagen) according to the manufacturer's protocol. Real-time

PCR analysis was carried out under previously published conditions.[14]

Statistical Analysis

Colorectal cancer-specific survival was computed from the date of cancer diagnosis up to the date of death or the end of follow-up. Patients who died of causes unrelated to colorectal cancer were censored at the time of death, whereas patients who died within 1 month of surgery were excluded from the analyses.

A univariable analysis was performed using a Cox proportional hazards model by including one variable at a time. Multivariable analysis was carried out with a Cox model that includes all the variables. The final set of variables were selected using a forward–backward stepwise method. The survival analysis was performed using STATA 10 Software (http://www.stata.com/).

Unsupervised hierarchical cluster analysis were done using a Euclidean distance measure and Ward linkage with the R statistical package (http://www.r-project.org/).[15] This analysis was done to identify prognostically relevant groups of colorectal carcinoma patients based on combined molecular and clinicopathologic variables. The statistical analysis for association and correlation was performed using Fisher's exact test, analysis of variance, and the independent sample t test.

Results

Survival Analysis According to Clinicopathologic and Molecular Features in Tumors

Table 2 illustrates the frequency of each molecular marker analyzed as well as specific associations between clinicopathologic features and molecular markers. Briefly, colorectal carcinomas with microsatellite instability were positively associated with location in the right colon ($P < 0.001$), specific histological types (mucinous, medullary, papillary, or cribriform types) ($P = 0.001$), a poor grade of

Table 2 Correlation between clinicopathologic and molecular markers

CLINICO-PATHOLOGIC AND MOLECULAR FEATURES	MICROSATELLITE INSTABILITY		BRAF MUTATION		MGMT METHYLATION		KRAS MUTATION		LOSS OF HETEROZYGOSITY (5q, 17p, 18q)	
	n= 27		n= 15		n= 37		n= 26		n= 82	
GENDER (female)		59%		47%		43%		50%		43%
SITE (right colon)	p=5.3*10⁻⁷	81%	p=8.1*10⁻⁶	87%		49%		31%	p=0.05	26%
MACROSCOPIC GROWTH (V/P)		30%		40%		38%		38%		33%
HISTOLOGICAL TYPE (MUC,MED, PAP,CRIB)	p=0.001	63%		47%		35%		39%		32%
HISTOLOGICAL GRADE (3)	p=3*10⁻⁴	52%	p=0.001	60%		22%		23%	p=0.006	16%
TNM STAGE (T1-T2)	p=0.03	68%		45%	p=0.03	63%		40%	p=0.02	36%
LYMPH NODE METASTASES (pN+)		38%		57%	p=0.009	35%		58%	p=0.03	62%
NEUROINVASION (present)		33%		33%		38%		50%		34%
ANGIGO-LYMPHOINVASION (present)		33%		33%		27%		46%		43%
MICROSCOPIC GROWTH (infiltrative)		70%		80%		84%		85%		82%
STROMAL FIBROSIS (present)		41%		60%		65%		77%	p=4.4*10⁻⁴	76%
LYMPHOID INFILTRATE (present)		74%		60%		73%		46%		59%
MICROSATELLITE INSTABILITY				87%		35%		8%		0%
BRAF MUTATION	p=2.2*10⁻⁸	48%				11%	p=3.9*10⁻²	0%		2%
MGMT METHYLATION	p=0.03	48%		27%				20%		19%
KRAS MUTATION		7%	p=3.9*10⁻²	0%		12%			p=0.03	30%
LOSS OF HETEROZYGOSITY (5q, 17p, 18q)		0%		13%		45%		81%		

CRIB, cribriform; MED, medullary; MUC, mucinous; PAP, papillary.
Loss of heterozigosity in at least one of the three chromosome analyzed (5q, 17p, 18q).

Table 3 Results of the univariate Cox regression analysis

Clinicopathologic variable	Hazard ratio (95% CI)	Significance (P-value)
Left site (*vs* right site)[a]	1.64 (0.79–3.37)	0.181
Macroscopic growth[b]	0.33 (0.69–2.57)	0.4
Histological type[c]	1.34 (0.67–2.71)	0.408
Histological grade (G3 *vs* G1–G2)	**2.16 (1.09–4.30)**	**0.027**
pT stage (pT4 *vs* pT1–pT3)	**3.42 (1.69–6.93)**	**0.001**
Lymph node metastasis (pN+ *vs* pN0)	**1.93 (0.99–3.78)**	**0.054**
TNM stage (IV *vs* III *vs* II *vs* I)	**3.13 (1.96–4.99)**	**10^{-6}**
Microscopic growth (infiltrative *vs* expansive)	2.19 (0.78–6.18)	0.138
Neuroinvasion (present *vs* absent)	**2.25 (1.19–4.26)**	**0.013**
Angio/lymphoinvasion (present *vs* absent)	**2.10 (1.11–3.96)**	**0.023**
Stromal fibrosis (present *vs* absent)	1.13 (0.57–2.24)	0.728
Tumor infiltrating lymphocytes (present *vs* absent)	0.55 (0.29–1.04)	0.066
Molecular variable		
Microsatellite instability status (absent *vs* present)[d]	2.08 (0.74–5.88)	0.165
Loss of heterozygosity status (present *vs* absent)[e]	1.69 (0.74–3.87)	0.214
MGMT methylation	0.9 (0.44–1.86)	0.781
BRAF mutation	0.45 (0.11–1.86)	0.269
KRAS mutation	1.34 (0.64–2.81)	0.434

[a]27 Rectal carcinomas and 51 cancers from distal colon.

[b]Polypoid/exophytic *vs* ulcerative/infiltrative.

[c]Common-type adenocarcinoma *vs* mucinous, medullary, papillary, cribriform histology.

[d]Microsatellite instability at high frequency.

[e]Loss of heterozygosity in at least one of the three chromosome region analyzed (5q, 17p, 18q).

Prognostically significant values are indicated in bold.

Table 4 Results of the multivariable analysis with a Cox model

Variable	HR	95% CI	P-value
Histological grade	1.97	0.95–4.06	0.068
TNM stage	3.37	2.01–5.65	10^{-6}
Angio/lymphoinvasion	1.63	0.84–3.16	0.15

CI confidence interval; HR, hazard ratio.

clinicopathologic parameters were correlated with a poor patient outcome: high (G3) tumor grade ($P = 0.027$), deep tumor invasion (pT4, $P = 0.001$), presence of lymph node metastases ($P = 0.054$), advanced TNM stage ($P = 10^{-6}$), presence of neuroinvasion ($P = 0.013$), and of angio/lymphoinvasion ($P = 0.023$).

A multivariable analysis with a Cox model has been used for exploring the relationship between the survival of a patient and several explanatory variables. This analysis was carried out using a forward–backward stepwise strategy. The best-selected model includes the following variables: TNM stage, angio/lymphoinvasion, and grade of differentiation. TNM stage has a deep prognostic value, instead grade of differentiation shows a borderline significance (Table 4).

Unsupervised Hierarchical Clustering Analysis of Clinicopathologic and Molecular Data

Unsupervised hierarchical clustering allowed the classification of 126 colorectal carcinomas on the basis of their similarities in clinicopathologic and molecular features. A clear separation into three prognostically significant groups was found. Cluster A and cluster C (25 and 37 patients, respectively) were associated with good prognosis and cluster B (64 patients) was correlated with poor prognosis ($P = 0.006$; Figure 1).

In cluster A, 77% of patients were alive at the census point, with a median survival of 60 months (mean, 57 months; 95% CI, 46–67). All colorectal carcinomas in this group exhibited microsatellite instability and high frequency of *MGMT* methylation (48% of cases). *BRAF* mutations observed in our series were almost exclusively in this cluster and were observed in 48% of cases (Figure 2a). The clinicopathologic profile of these tumors was described above for colorectal carcinomas with microsatellite instability.

In cluster C, 79% of patients were alive at the census point, with a median survival of 80 months (mean 76 months; 95% CI, 69–83). There was no significant survival difference between cluster A and cluster C patients. Colorectal carcinomas in cluster C were characterized by a very low frequency of microsatellite instability (5% of cases) and no *BRAF* mutations. By contrast, loss of heterozygosity of at least one of the three regions investigated and
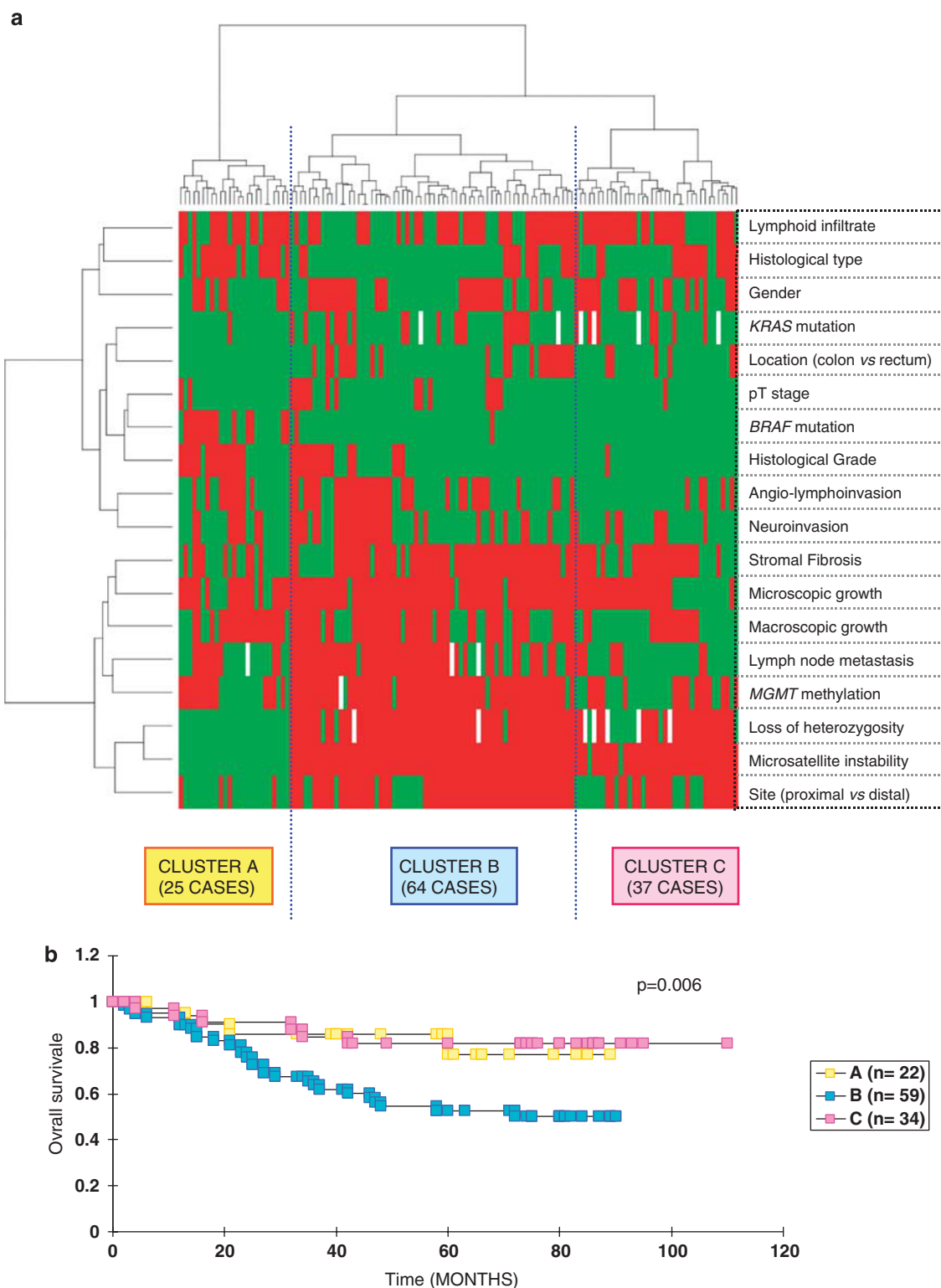
tumor differentiation ($P < 0.001$), stage I and stage II disease ($P = 0.03$), *BRAF* mutation ($P < 0.001$), and *MGMT* methylation ($P = 0.03$). By contrast, colorectal carcinomas with loss of heterozigosity were positively associated with location in the left colon or rectum ($P = 0.05$), with a low (G1) or intermediate (G2) tumor grade ($P = 0.006$), with the presence of lymph node metastases ($P = 0.03$), with stage III and stage IV disease ($P = 0.02$) and with the presence of stromal fibrosis ($P < 0.001$). *KRAS* mutation was significantly associated with loss of heterozygosity ($P = 0.03$) and was found to be mutually exclusive with *BRAF* mutation ($P = 0.04$). Finally, *MGMT* methylation was positively associated with the absence of lymph node metastases ($P = 0.009$) and with stages I and II ($P = 0.03$).

Table 3 provides a summary of the prognostic significance of all markers investigated by univariable analysis. Some clinicopathologic parameters, but no molecular markers, were prognostically significant ($P \leq 0.05$). Specifically, the following

a

Figure 1 Unsupervised two-dimensional hierarchical clustering based on clinicopathologic and molecular markers identifies three distinct clusters of colorectal carcinomas. (**a**) Full-length view of the cluster diagram with cases orientated along the horizontal axis and clinicopathologic/molecular markers orientated along the vertical axis. (**b**) Kaplan–Meier curves show significant differences in overall survival of patients in clusters A, B, and C.

MGMT methylation were found in 62% and 57% of these tumors, respectively. *KRAS* mutations were identified in 19% of these cases. Compared with cluster A tumors, colorectal carcinomas in cluster C

were more likely to originate in the left colon, and they mostly were common-type adenocarcinomas of low (G1) or intermediate (G2) grade. Analogously to cluster A colorectal carcinomas, they were
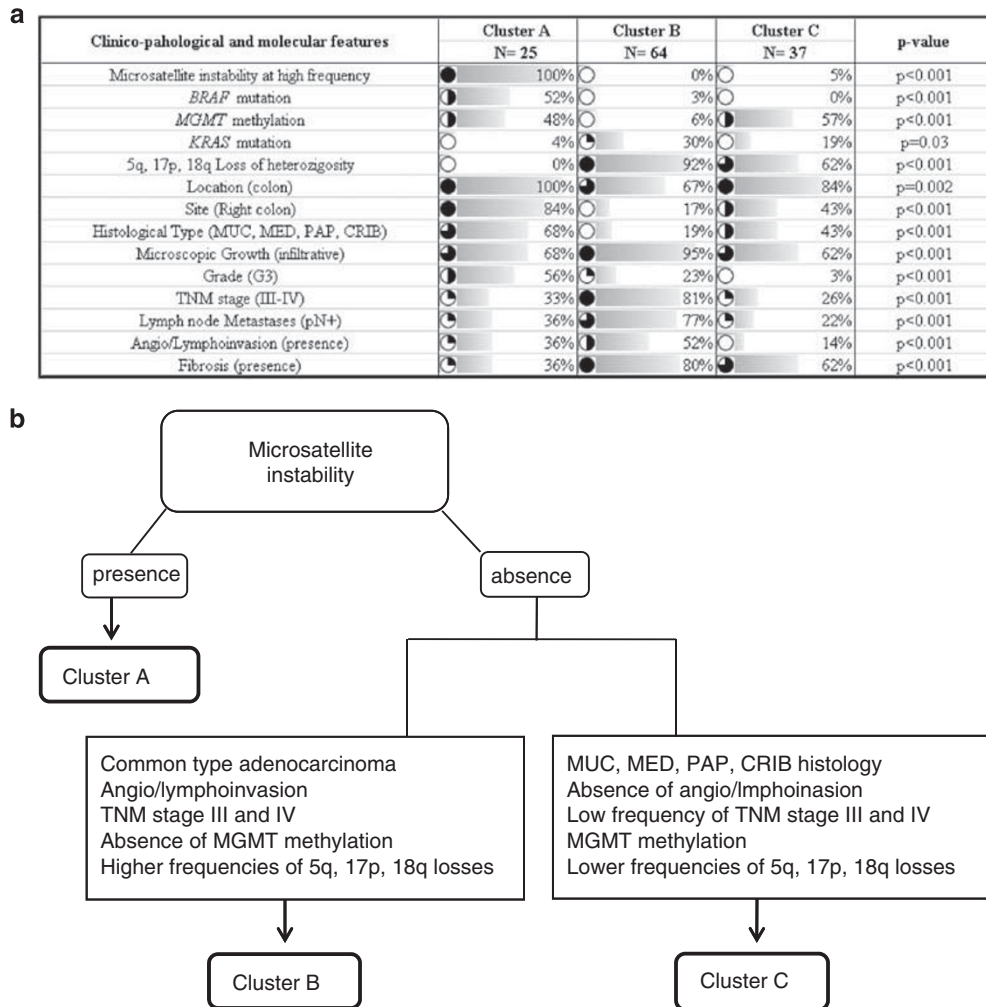
Figure 2 (**a**) Frequencies of molecular and clinicopathologic features in clusters A, B, and C tumors. CRIB, cribriform; MED, medullary; MUC, mucinous; PAP, papillary. (**b**) Derivation of clusters A, B, and C based on the significantly different clinicopathologic and molecular features observed in each group.

mainly stage I or stage II tumors, and they were characterized by very low frequencies of angio/lymphoinvasion, neuroinvasion and lymph node metastases (Figure 2a).

In cluster B, 52% of patients were alive at the census point, with a median survival of 75 months (mean, 67 months; 95% CI, 61–73). Cluster B comprised the most aggressive tumors in our series. Compared with colorectal carcinomas in clusters A and C, these tumors were significantly associated with stage III or stage IV ($P<0.001$), presence of lymph node metastases ($P<0.001$), infiltrative growth ($P<0.001$), angio/lymphoinvasion ($P<0.001$), and presence of stromal fibrosis ($P<0.001$). Genetically, they were characterized by the highest frequencies of *KRAS* mutation and loss of heterozygosity observed in our series (30% and 92% of cases, respectively). Particularly, loss of heterozygosity frequency was significantly higher in cluster B than in cluster C tumors ($P<0.01$), with the most remarkable differences for 17p loss (75% of cluster B tumors *vs* 34% of cluster C tumors showed 17p loss, $P=0.004$). Finally,

cluster B showed rare *MGMT* methylation, no *BRAF* mutations, and absence of microsatellite instability (Figure 2a).

Figure 2b summaries a derivation of the three different clusters based on the significantly different clinicopathologic and molecular features observed in each group.

In order to evaluate the presence of abnormal chromosomal content and/or the presence of chromosome instability, interphasic FISH on extracted nuclei was performed on a selected subset of 29 cases (8, 10, and 11 cases from cluster A, cluster B, and cluster C, respectively). FISH analysis was performed using centromeric and locus-specific probes selected from chromosomal regions frequently involved in colorectal carcinogenesis. Detailed results of FISH analysis are shown in Table 5. FISH analysis revealed the presence of different numerical chromosome alterations in all the tumors analyzed and also demonstrated the presence of chromosomal instability in a large subset of these cases. Representative results of FISH analysis in the

Table 5 Frequencies of chromosome gains and losses observed in clusters A, B, and C tumors by interphasic FISH analysis on extracted nuclei

GAIN	CLUSTER A N = 8	CLUSTER B N = 10	CLUSTER C N = 11
5p15.2	9.9%	39.5%	19.1%
EGR1 (5q31)	6.0%	38.5%	19.8%
CEN 7	28.3%	52.7%	42.1%
EGFR (7p21)	31.7%	55.2%	40.9%
CEN 17	2.2%	42.8%	20.6%
TP53	7.6%	28.8%	23.8%
13q14	7.6%	57.9%	34.3%
13q34	10.2%	42.8%	41.4%
CEN 18	4.2%	16.3%	10.9%
LOSS	CLUSTER A N = 8	CLUSTER B N = 10	CLUSTER C N = 11
TP53	0.0%	27.8%	11.4%
CEN 18	22.3%	26.4%	35.8%

three clusters are shown in Figure 3. Comparing the different clusters, colorectal carcinomas in cluster A showed a significantly lower percentage of aneuploid cells than cluster B and cluster C tumors (mean percentage of aneuploid cells was 16.6% in cluster A tumors *vs* 41% and 31.9% in cluster B and cluster C tumors, respectively; $P < 0.001$). Both cluster B and cluster C tumors showed different cell populations with trisomies, tetrasomies and polisomies suggestive of a high degree of genetic heterogeneity. Nevertheless, colorectal carcinomas in cluster B showed a significantly higher percentage of aneuploid cells than cluster C tumors ($P = 0.0015$). Gain of chromosomes 7 and 13 were the most frequent aneusomies observed in clusters C and B, whereas loss of chromosome 18 as well as gain of chromosome 7 and/or of *EGFR* (7p11.2) region were the most relevant chromosome alterations detected in cluster A tumors. In addition, scoring FISH signals for chromosome 17 and for the *TP53* gene, a loss of at least one copy of p53 was observed only in colorectal carcinomas in clusters B and C regardless of the ploidy and the genetic heterogeneity of chromosome 17 aneusomies. The percentage of cells with p53 loss is higher in tumors from cluster B (mean of cells with p53 loss for cluster B: 27.8%) than in tumors from cluster C (mean of cells with p53 loss for cluster C: 11.4%).

Discussion

This work has evaluated the potential superiority of a morphomolecular classification based on the combination of clinicopathologic and molecular features of colorectal cancers using an unsupervised hierarchical clustering analysis and a multivariable analysis with a Cox model.

Unsupervised hierarchical clustering analysis is a common method to profile gene expression data.[16–19] More recently, there have been attempts to classify cancers based on hierarchical clustering analysis of immunohistochemical markers[20–22] as well as of genetic and epigenetic features.[2]

For the first time, this study used unsupervised hierarchical clustering analysis to combine 13 routinely assessed clinicopathologic features and all five molecular markers recently suggested by Jass' classification to distinguish four molecular subtypes of sporadic colorectal carcinomas.[5] Clustering analysis resulted in a clear separation into three prognostically and biologically significant groups. Features of cluster A and cluster B were strongly concordant with the molecular subtypes indicative of the two major types of genomic instability recognized as alternative mechanisms of colorectal carcinogenesis,[1] and they overlap with groups 1 and 4 from Jass' classification, respectively.

CHROMOSOMAL REGION	CLUSTER A	CLUSTER B	CLUSTER C

Rows (left to right columns A, B, C):
- EGR1(red)/5p15.2 (green)
- EGFR (red)/Cen7 (green)
- 13q14 (red)/13q34 (green)
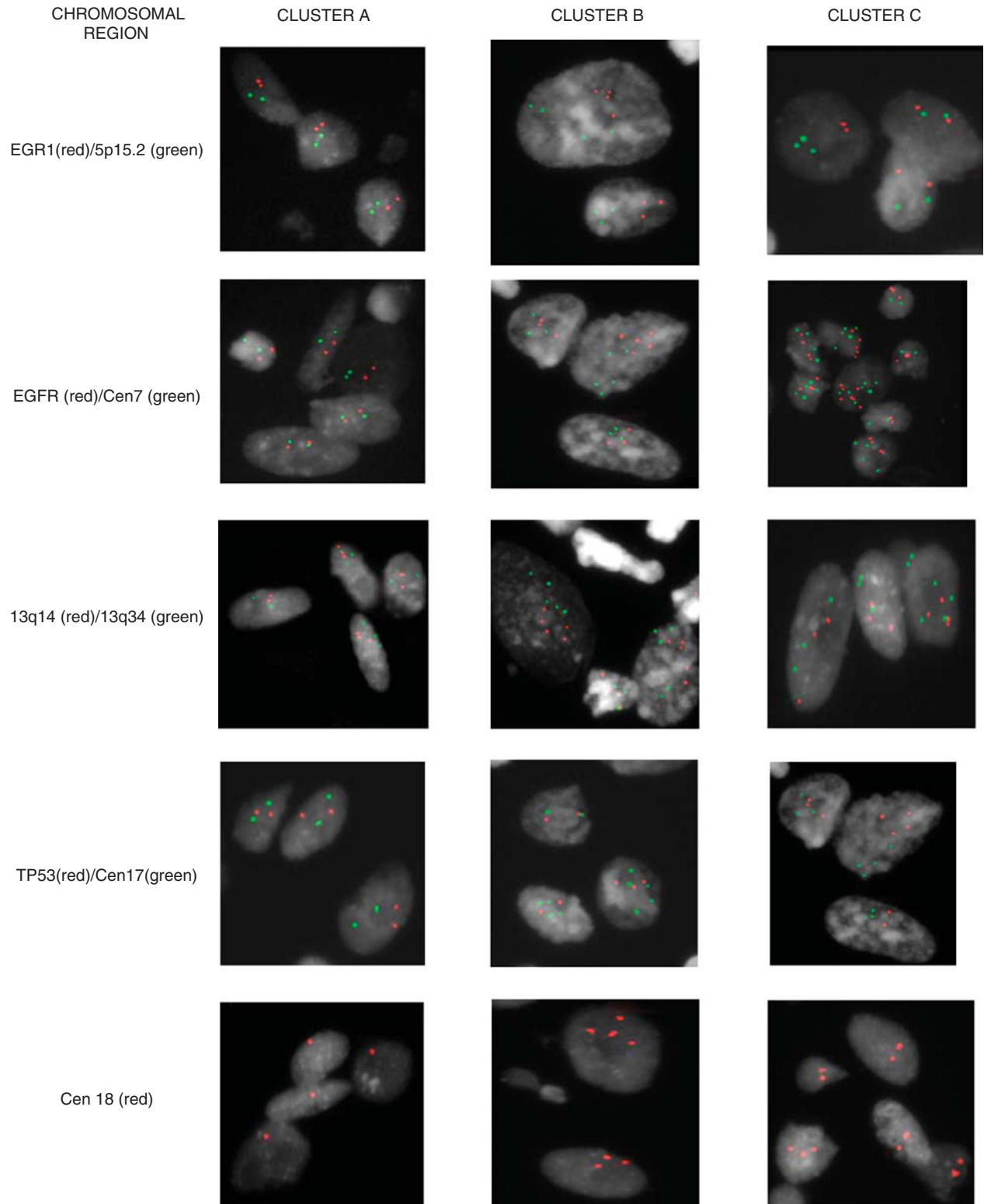- TP53(red)/Cen17(green)
- Cen 18 (red)

Figure 3 Representative results of interphasic FISH analysis in clusters A, B, and C tumors. Colorectal carcinomas in cluster A showed a lower level of aneuploid cells than cluster B and cluster C cases. Both cluster B and cluster C tumors showed different cell populations, suggesting genetic heterogeneity. Loss of *TP53* is evident in nuclei from cluster B and cluster C tumors regardless of the ploidy level of chromosome 17.

Cluster A is completely composed of colorectal carcinomas with microsatellite instability at high frequency. As expected, these tumors had distinct molecular and clinicopathologic features, including low percentages of chromosome aneusomies, high frequencies of both *BRAF* mutation and *MGMT*

methylation. Clinically, they were associated with a better prognosis and showed a clinicopathologic profile consistent with the reported phenotype for tumors with defective DNA mismatch repair.[5,23]

Moreover, preliminary data in our laboratory examining a panel of tumors from clusters A, B, and C for the methylation status of 35 tumor suppressor genes, demonstrated that cluster A tumors showed the highest frequencies of *de novo* methylation (data not shown). This finding is consistent with previous studies that found a strong positive association between microsatellite instability and the presence of a CpG island methylator phenotype and outlines more precisely the molecular profile of these tumors.[3]

By contrast, cluster B consisted of colorectal carcinomas with chromosomal instability assessed by loss of heterozigosity analysis and by interphasic FISH. These tumors showed the highest frequencies of *TP53* LOH and of *KRAS* mutations, but rare *MGMT* methylation and no *BRAF* mutations. Clinically, they had the worst outcome and comprised the most aggressive tumors showing all the well-established clinicopathologic features associated with a poor prognosis in colorectal cancer. Cluster B tumors were associated with a poorer prognosis than cluster A tumors, in agreement with large meta-analyses,[24,25] which unequivocally established that chromosomal instability and microsatellite instability are useful prognostic variables for colorectal carcinomas. However, despite their prognostic value is no longer in question, chromosomal instability and microsatellite instability are not reliable markers to predict prognosis in colorectal carcinomas, especially when rather small cohorts of patients are examined.[26–31] Our data are consistent with these studies, showing that both markers were not significant prognostic variables. Univariable and multivariable analyses in our work demonstrated that the aggressive behavior of cluster B tumors cannot be explained by a relative contribution of the molecular markers included in the study (Tables 3 and 4). By contrast, TNM stage alone remains the most significant independent marker in predicting outcome.

An important finding of our analysis was the identification of a third subgroup of colorectal carcinomas, namely cluster C, that was associated with a good prognosis, without a significant survival difference to cluster A patients (Figure 1). The clinicopathologic profile of cluster C tumors was suggestive of a less aggressive disease than cluster B tumors considering TNM stage and lower frequencies of angio/lymphoinvasion and neuroinvasion.

Genetically, they appeared intermediate between cluster A and cluster B, as they lacked microsatellite instability (only two cases showed microsatellite instability) as well as BRAF mutation and they showed both *MGMT* methylation and frequent allelic losses. Interphasic FISH with centromeric probes added important information about the abnormal chromosomal content in cluster B *vs* cluster C tumors. This analysis has proven to be extremely useful in distinguishing a state of aneuploidy (an abnormal chromosomal content reflected in allelic imbalance at the molecular level) from a condition of chromosomal instability (an accelerated rate of gains or losses of whole or large portions of chromosomes).[32] Although the FISH approach does not directly measure an increased rate of aberrations, the chromosomal heterogeneity detectable by this analysis may serve as a good indicator of chromosomal instability.[33] Our analysis demonstrated a high grade of chromosomal heterogeneity suggestive of a phenotype with chromosomal instability in both clusters but significantly lower percentages of aneusomies were detected in cluster C than in cluster B tumors (Figure 3). This finding, together with the observation of significantly lower frequencies of 17p losses in cluster C than in cluster B, suggests a less advanced genetic progression in cluster C tumors.[34–36] This hypothesis is consistent with the clinicopathologic profile of these tumors significantly associated with a good prognosis.

The identification of cluster C tumors is the main discrepancy with Jass' classification. In our analysis, these tumors share molecular features observed in both the remaining two groups of sporadic colorectal cancers described by the authors, namely group 2 and group 3. In order to validate Jass' classification, we also considered the possibility that different levels of gene hypermethylation in cluster C may be useful to better stratify these tumors in a high-CpG island methylator phenotype and low-CpG island methylator phenotype (as observed in group 2 and in group 3, respectively). However, cluster C tumors were characterized by low levels of hypermethylation and no significant differences between cluster B and cluster C tumors were observed. Considering recent data, we inferred that cluster C tumors may overlap with the subset of colorectal carcinomas termed 'CIMP-low'[3] or 'CIMP 2'[2] showing low levels of methylation as well as chromosomal instability and high frequencies of *MGMT* methylation and of *KRAS* mutation.

In conclusion, our study demonstrates that a more accurate tumor classification should combine the prognostic power of clinicopathologic parameters with molecular biomarkers that provide information regarding the natural history of the cancer.

In our study, TNM stage alone remains the most significant independent marker in predicting outcome. However, the future utility of the TNM staging depends on its ability to deal with the increase in population screening for colorectal cancer, the discovery of new therapies, and the use of new molecular biomarkers. Indeed, because the system depends on the temporal progression of tumors for its predictive accuracy, anything that reduces the temporal dimension reduces its accuracy.

Hierarchical clustering seems to be a useful, promising, and powerful tool for further translational

studies and should lead us to define a diagnostic and prognostic signature for different carcinomas.

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

1 Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology 2008;135:1079–1099.

2 Shen L, Toyota M, Kondo Y, *et al.* Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. Proc Natl Acad Sci USA 2007;104:18654–18659.

3 Ogino S, Goel A. Molecular classification and correlates in colorectal cancer. J Mol Diagn 2008;10: 13–27.

4 Walther A, Johnstone E, Swanton C, *et al.* Genetic prognostic and predictive markers in colorectal cancer. Nat Rev Cancer 2009;9:489–499.

5 Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology 2007;50:113–130.

6 Furlan D, Sahnane N, Carnevali I, *et al.* Up-regulation of the hypoxia-inducible factor-1 transcriptional pathway in colorectal carcinomas. Hum Pathol 2008;39: 1483–1494.

7 Hamilton SR, Vogelstein B, Kudo S, *et al.* Carcinoma of the colon and rectum. In: Hamilton SR, Aaltonen LA (eds). World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Digestive System. IARC Press: Lyon, 2000, pp 105–119.

8 Sobin L GM, Wittekind C. TNM Classification of Malignant Tumours. 6th edn. Wiley-Liss: New York, 2002.

9 Furlan D, Cerutti R, Genasetti A, *et al.* Microallelotyping defines the monoclonal or the polyclonal origin of mixed and collision endocrine-exocrine tumors of the gut. Lab Invest 2003;83:963–971.

10 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.

11 Paternoster SF, Brockman SR, McClure RF, *et al.* A new method to extract nuclei from paraffin-embedded tissue to study lymphomas using interphase fluorescence *in situ* hybridization. Am J Pathol 2002;160: 1967–1972.

12 Tibiletti MG, Martin V, Bernasconi B, *et al.* BCL2, BCL6, MYC, MALT 1, and BCL10 rearrangements in nodal diffuse large B-cell lymphomas: a multicenter evaluation of a new set of fluorescent *in situ* hybridization probes and correlation with clinical outcome. Hum Pathol 2009;40:645–652.

13 Eads CA, Danenberg KD, Kawakami K, *et al.* MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res 2000;28:E32.

14 Uccella S, Cerutti R, Placidi C, *et al.* MGMT methylation in diffuse large B-cell lymphoma: validation of quantitative methylation-specific PCR and comparison with MGMT protein expression. J Clin Pathol 2009;62: 715–723.

15 R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, 2009.

16 Sorlie T, Tibshirani R, Parker J, *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 2003;100:8418–8423.

17 Van't Veer LJ, Dai H, van de Vijver MJ, *et al.* Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–536.

18 Alizadeh AA, Eisen MB, Davis RE, *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503–511.

19 Garber ME, Troyanskaya OG, Schluens K, *et al.* Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci USA 2001;98:13784–13789.

20 Au NH, Cheang M, Huntsman DG, *et al.* Evaluation of immunohistochemical markers in non-small cell lung cancer by unsupervised hierarchical clustering analysis: a tissue microarray study of 284 cases and 18 markers. J Pathol 2004;204:101–109.

21 Makretsov NA, Huntsman DG, Nielsen TO, *et al.* Hierarchical clustering analysis of tissue microarray immunostaining data identifies prognostically significant groups of breast carcinoma. Clin Cancer Res 2004;10:6143–6151.

22 Alkushi A, Clarke BA, Akbari M, *et al.* Identification of prognostically relevant and reproducible subsets of endometrial adenocarcinoma based on clustering analysis of immunostaining data. Mod Pathol 2007; 20:1156–1165.

23 Benatti P, Gafa R, Barana D, *et al.* Microsatellite instability and colorectal cancer prognosis. Clin Cancer Res 2005;11:8332–8340.

24 Walther A, Houlston R, Tomlinson I. Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. Gut 2008;57:941–950.

25 Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. J Clin Oncol 2005;23:609–618.

26 Jernvall P, Makinen MJ, Karttunen TJ, *et al.* Microsatellite instability: impact on cancer progression in proximal and distal colorectal cancers. Eur J Cancer 1999;35:197–201.

27 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.

28 Feeley KM, Fullard JF, Heneghan MA, *et al.* Microsatellite instability in sporadic colorectal carcinoma is not an indicator of prognosis. J Pathol 1999;188:14–17.

29 Salahshor S, Kressner U, Fischer H, *et al.* Microsatellite instability in sporadic colorectal cancer is not an

independent prognostic factor. Br J Cancer 1999;81: 190–193.

30 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.

31 Guidoboni M, Gafa R, Viel A, *et al.* Microsatellite instability and high content of activated cytotoxic lymphocytes identify colon cancer patients with a favorable prognosis. Am J Pathol 2001;159:297–304.

32 Rajagopalan H, Nowak MA, Vogelstein B, *et al.* The significance of unstable chromosomes in colorectal cancer. Nat Rev Cancer 2003;3:695–701.

33 Wilkens L, Flemming P, Gebel M, *et al.* Induction of aneuploidy by increasing chromosomal instability during dedifferentiation of hepatocellular carcinoma. Proc Natl Acad Sci USA 2004;101:1309–1314.

34 Kern SE, Fearon ER, Tersmette KW, *et al.* Clinical and pathological associations with allelic loss in colorectal carcinoma [corrected]. JAMA 1989;261:3099–3103.

35 Graziano F, Cascinu S. Prognostic molecular markers for planning adjuvant chemotherapy trials in Dukes′ B colorectal cancer patients: how much evidence is enough? Ann Oncol 2003;14:1026–1038.

36 Russo A, Bazan V, Iacopetta B, *et al.* The TP53 Colorectal Cancer International Collaborative Study on the prognostic and predictive significance of *p53* mutation: influence of tumor site, type of mutation, and adjuvant treatment. J Clin Oncol 2005;23: 7518–7528.