Unsupervised Analysis of Array Comparative Genomic Hybridization Data from Early-Onset Colorectal Cancer Reveals Equivalence with Molecular Classification and Phenotypes

Abstract

**AIM:** To investigate whether chromosomal instability (CIN) is associated with tumor phenotypes and/or with global genomic status based on MSI (microsatellite instability) and CIMP (CpG island methylator phenotype) in early-onset colorectal cancer (EOCRC). **METHODS:** Taking as a starting point our previous work in which tumors from 60 EOCRC cases (≤45 years at the time of diagnosis) were analyzed by array comparative genomic hybridization (aCGH), in the present study we performed an unsupervised hierarchical clustering analysis of those aCGH data in order to unveil possible associations between the CIN profile and the clinical features of the tumors. In addition, we evaluated the MSI and the CIMP statuses of the samples with the aim of investigating a possible relationship between copy number alterations (CNAs) and the MSI/CIMP condition in EOCRC. **RESULTS:** Based on the similarity of the CNAs detected, the unsupervised analysis stratified samples into two main clusters (A, B) and four secondary clusters (A1, A2, B3, B4). The different subgroups showed a certain correspondence with the molecular classification of colorectal cancer (CRC), which enabled us to outline an algorithm to categorize tumors according to their CIMP status. Interestingly, each subcluster showed some distinctive clinicopathological features. But

Abbreviations: aCGH, array comparative genomic hybridization; CIMP, CpG island methylator phenotype; CIN, chromosomal instability; CNA, copy number alteration; CRC, colorectal cancer; DFS, disease-free survival; EOCRC, early-onset colorectal cancer; GII, genomic Instability Index; LS, Lynch syndrome; MACS, microsatellite and chromosome stable tumors; MMR, mismatch repair; MSI, microsatellite instability; MSS, microsatellite stability; OS, overall survival

Address all correspondence to: José Perea García, Department of Surgery, Hospital Universitario 12 de Octubre, C/Roxas de Aravaca, 82A, 1°Dcha, E-28023, Madrid, Spain.
E-mail: josepereag@hotmail.com

1 Financial support: This work was supported by Projects PI10/0683, PI13/01741, PI13/0127 from the Spanish Ministry of Health and Consumer Affairs; FEDER; and Mari Paz Jiménez Casado Foundation.
2 These two authors contributed equally to this work.
Received 20 June 2016; Revised 2 November 2016; Accepted 7 November 2016 © 2016 The Authors. Published by Elsevier Inc. on behalf of Neoplasia Press, Inc. This is an open access article under theCCBY-NC-NDlicense (http://creativecommons.org/licenses/by-nc-nd/4.0/). 1476-5586
http://dx.doi.org/10.1016/j.neo.2016.11.006
more interestingly, the CIN of each subcluster mainly affected particular chromosomes, allowing us to define chromosomal regions more specifically affected.

CONCLUSIONS: Our findings may provide a basis for a new form of classifying EOCRC according to the genomic status of the tumors.

Introduction

Colorectal cancer (CRC) has a great impact on the world population, since it represents the third most common malignancy and the second leading cause of death in developed countries [1,2]. Its pathogenesis is a multistep process in which the accumulation of different genetic and epigenetic alterations leads to the transformation of healthy colonic epithelial cells into malignant cells [3]. The loss of genomic stability is a key molecular pathogenic step that occurs early in tumorigenesis, and it can be caused by at least three major molecular pathways: chromosomal instability (CIN), microsatellite instability (MSI) and CpG island methylator phenotype (CIMP).

Early-onset CRC (EOCRC) represents a relatively unusual entity commonly related with hereditary forms of the disease. Thus, it is estimated that about 11% of colon cancers and 18% of rectal cancers arise in individuals younger than 50 years [4–7]. In comparison with late-onset CRC, EOCRC is more frequently associated with poor clinical features and it is considered as a high-risk group within CRC [8,9].

The clinicopathological features of tumors can differ significantly depending on the type of genomic alterations, which makes CRC a heterogeneous disease in which it is difficult to determine the clinical consequences of individual alterations. Although some studies have attempted to correlate the clinicopathological features and the molecular profile in late-onset tumors [10–12], this relationship has not been fully investigated in EOCRC, possibly because of the low frequency of CRC in young people [4,5]. In our previous work, we performed a comprehensive analysis of the DNA copy number alterations (CNAs) that occur in two heterogeneous disease in which it is difficult to determine the clinical consequences of individual alterations. Although some studies have attempted to correlate the clinicopathological features and the molecular profile in late-onset tumors [10–12], this relationship has not been fully investigated in EOCRC, possibly because of the low frequency of CRC in young people [4,5]. In our previous work, we performed a comprehensive analysis of the DNA copy number alterations (CNAs) that occur in two groups of patients differing in age at onset, and observed substantial dissimilarities regarding the CIN pattern as well as the most frequent CNAs arising in each group [13]. Taking this as a starting point, the purpose of the present study was to investigate whether the CIN profile is also associated with the biological characteristics and/or with the global genomic status (based on MSI and CIMP) in EOCRC, when an unsupervised hierarchical clustering analysis is performed according to the similarity of the CNAs detected.

Materials and Methods

Patients, Samples and Data Collection

A total of 88 individuals diagnosed with CRC at an age of 45 years or younger (range: 16–45 years) were collected at the 12 de Octubre University Hospital in Madrid. Family history of cancer (including at least three generations) and clinicopathological information was obtained for each patient, with a follow-up of at least 60 months from surgery. All patients (or a first degree relative in case of the index case) provided written consent, and the study was approved by the Ethics Committee of our Institution.

Six patients were excluded because familial adenomatous polyposis was diagnosed. Material for array comparative genomic hybridization (aCGH) analysis could be obtained from 60 of the remaining 82 patients. In our series, the left location was considerably more common than the right one (53.3% vs. 20%) (Supplementary Table S1). Moreover, and as expected given the early-onset of the disease, the percentage of sporadic cases was lower than the percentage of patients who had a familial component or fulfilled the clinical criteria for Lynch syndrome (LS). Additional clinical, pathological and familial features are shown in Supplementary Table S1.

Assessment of Genomic Instability: Molecular Classification

A tissue specimen was obtained from each index case. Microscopic inspection of paraffin-embedded samples was performed by a pathologist, and samples with more than 70% of tumor cells in the neoplastic material were considered adequate for further analysis. The protocol for DNA isolation was as previously reported [13].

We used the Bethesda panel to assess the MSI status, and considered a result positive when two or more markers were altered. Blood samples were taken from the MSI index cases to assess germline mutations in MLH1, MSH2 and MSH6. Moreover, MSI tumors were analyzed for the BRAF V600E mutation in order to identify possible sporadic cases. For the assessment of CIMP, we investigated the methylation status of the promoter regions of CACNA1G, CDKN2A, CRABP1, IGF2, MLH1, NEUROG1, RUNX3 and SOCS1. CIMP-High was defined as the presence of ≥6/8 methylated promoters, CIMP-Low as 1/8 to 5/8 methylated promoters and CIMP-0 as the absence of methylated promoters [14]. We classified tumors into four categories according to the MSI and CIMP status as described by Ogino and Goel: (1) MSI/CIMP-High; (2) MSI/CIMP-Low/0; (3) MSS/CIMP-High; (4) MSS/CIMP-Low/0 [15]. Finally, the degree of CIN was evaluated by aCGH, considering tumors with more than 3 whole chromosomes affected as CIN+. Tumors with 1–3 whole chromosomes affected as MACS (microsatellite and chromosome stable), and tumors with no whole chromosome affected as CIN-.

The procedures for the evaluation of CIN, MSI, and CIMP were as previously reported [9,13].

Unsupervised Analysis of aCGH Data

Tumors were clustered based on the copy number states of their windowed probes [13]. Unsupervised analysis was performed using hierarchical clustering algorithms (squared Euclidean distances) implemented in Multi Experiment Viewer 4.8.1 (www.tm4.org/mev.html).

Statistical Analysis

Comparison of continuous variables was done using Student’s two-tailed t test (for normal distributions) or the Mann–Whitney U test (for nonparametric distributions), whereas comparison of categorical variables was done using Pearson’s chi square ($\chi^2$) test. For comparisons between more than two groups, analysis of variance (ANOVA) (for normal distributions) or the Kruskal-Wallis test (for
nonparametric distributions) were used. Statistical analysis was performed using SPSS 17.0 (SPSS Inc., Chicago, IL, USA), and differences were considered statistically significant when the $P < .05$.

The study of association and correlation was performed employing R Statistical Software [16] using Fisher’s exact test, analysis of variance and the independent sample $t$ test. $P < .05$ was considered statistically significant.

Results

Unsupervised Analysis According to the CIN Profile

The unsupervised hierarchical clustering stratified samples into two main clusters. The first of them (cluster A) was composed of 38 cases while the second one (cluster B) was composed of 22 cases (Fig. 1; Supplementary Fig. 1). Both clusters, in turn, were divided into two secondary clusters: subcluster A1 ($n = 25$) and A2 ($n = 13$) derived from A, and subcluster B3 ($n = 10$) and B4 ($n = 12$) derived from B (Fig. 1; Supplementary Fig. 1).

Clinicopathological Features in the Different Clusters

Cluster A included a larger percentage of very young patients (<36 years at diagnosis) and all tumor recurrences ($P = .035$), and was associated with a worse prognosis (Fig. 2; Table 1). In the comparison of subgroups A1, A2, B3 and B4, statistical significance was only reached when tumor histology was considered (Table 1). However, it should be emphasized that there were some distinctive

Fig. 1. Unsupervised hierarchical clustering based on the genomic instability profiles of the studied tumors. The studied samples were stratified into two main clusters (A, B) and four secondary clusters (A1, A2, B3, B4) according to the similarity of the CNAs detected.

Fig. 2. Kaplan–Meier survival plots: (A) Comparison of the DFS between the two main clusters; (B) comparison of the DFS between the four subclusters; (C) comparison of the OS between the two main clusters; (D) comparison of the OS between the four subclusters. Note the significantly higher proportion of tumor recurrence within the cluster A.
clinopathological features between the different subgroups. On the one hand, subgroup A1 was associated with the highest percentage of tumor recurrences. On the other hand, the subgroup including the highest proportion of stage IV tumors (B4) showed the highest death rates as well as the highest prevalence of rectal location (Table 1). As expected, the right-sided location (typically related with LS) was predominant within the subcluster with the highest percentage of individuals with familial antecedents of the disease (A2) (Table 1).

**Distribution of MSI and CIMP**

Thirteen patients (16%) of a total of 81 showed MSI and the concomitant loss of expression of one or more of the DNA mismatch repair (MMR) proteins. Of the 13 patients with MSI, 7 had a pathogenic germline mutation in one of the MMR genes, 4 had variants of unknown significance in at least one of these genes, and 1 showed lack of expression of MLH1 and hypermethylation of its gene promoter (data not shown).

Of the 60 tumors that could be evaluated by aCGH, MSI was present in 10 cases, 7 of which were confirmed as LS patients (1 had a mutation in MLH1, 4 in MSH2 and 2 in MSH6) (Table 2). There was only 1 patient with hypermethylation of the MLH1 promoter; however, the BRAF V600E mutation was absent in this patient. MSI was homogeneously distributed among clusters A and B, with a complete correlation with the immunohistochemical results (Table 2). Regarding the methylator phenotype, we observed a clear prevalence of CIMP-High tumors within cluster A, with all but one of the tumors contained therein showing this genotype (Table 2). When we compared the subclusters derived from the main clusters, we observed that only B4 showed absence of MSI tumors (being also the cluster with the lowest proportion of right-sided tumors), and that B3 was the only subgroup entirely composed of CIMP-Low/0 tumors (Table 2).

**Distribution of CIN**

Clusters A and B did not differ in terms of CNAs nor with respect to the total genomic instability index (GII), even though cluster B

**Table 1.** Clinicopathological correlation of the different subgroups obtained by unsupervised hierarchical clustering analysis

| Clusters | Cluster A1 | Cluster A2 | Cluster B3 | Cluster B4 | P-value† |
|----------|------------|------------|------------|------------|----------|
| No. of tumors | 25 | 13 | 10 | 12 | - |
| 36–45 years old | 19 | 10 | 9 | 11 | NS |
| <36 years old | 6 (54.5%) | 3 (27.3%) | 1 (9.1%) | 1 (9.1%) | NS |
| Location | | | | | |
| Right colon | 5 (20%) | 5 (38.4%) | 1 (10%) | 1 (8.3%) | NS |
| Left colon | 15 (60%) | 4 (30.8%) | 7 (70%) | 6 (50%) | NS |
| Rectum | 5 (20%) | 4 (30.8%) | 2 (20%) | 5 (41.7%) | NS |
| Histology | | | | | |
| Adenocarcinoma | 22 (88%) | 7 (53.8%) | 9 (90%) | 11 (91.7%) | 0.034 |
| Malignant adenoma | 3 (12%) | 6 (46.2%) | 1 (10%) | 1 (8.3%) | NS |
| T. differentiation: | | | | | |
| High | 10 (45.5%) | 3 (42.9%) | 4 (44.4%) | 2 (18.2%) | NS |
| Moderate | 8 (36.4%) | 4 (57.1%) | 5 (55.6%) | 9 (81.8%) | NS |
| Low | 4 (18.1%) | - | - | - | NS |
| Mucin production | 7 (31.8%) | 1 (14.3%) | 3 (33.3%) | 2 (18.2%) | NS |
| "Signet ring" cells | 1 (4.5%) | - | 3 (33.3%) | 2 (18.2%) | NS |
| TNM stage | | | | | |
| I | 6 (24%) | 6 (46.2%) | 2 (20%) | 3 (25%) | NS |
| II | 9 (36%) | 4 (30.8%) | 7 (70%) | 1 (8.3%) | NS |
| III | 5 (20%) | 1 (7.6%) | 1 (10%) | - | NS |
| IV | 5 (20%) | 2 (15.4%) | - | 5 (41.7%) | NS |
| Average No. of polyps | 3.76 [1.00] | 6.00 [4.00] | 2.40 [2.00] | 1.33 [0.00] | NS§ |
| Synchronous T. | 2 (8%) | 1 (7.7%) | - | - | NS |
| Metachronous T. | - | 2 (15.4%) | - | - | NS |
| OS ± SD | 87.12 ± 50.85 | 99.68 ± 63.18 | 86.10 ± 10.03 | 59.50 ± 36.07 | NS§ |
| DFS ± SD | 72.60 ± 53.90 | 78.38 ± 65.68 | 84.50 ± 9.48 | 51.42 ± 44.26 | NS§ |
| T. recurrence | 5 (20%) | 2 (15.4%) | - | - | NS |
| Mortality | 7 (28%) | 3 (23.1%) | 5 (50%) | 4 (33.3%) | NS |
| Sporadic | 13 (52%) | 3 (23.1%) | 5 (50%) | 6 (50%) | NS |
| Familial aggregation | 9 (36%) | 6 (46.2%) | 1 (10%) | 4 (33.3%) | NS |
| Amsterdam II positive | 3 (12%) | 4 (30.7%) | 4 (40%) | 2 (16.7%) | NS |
| Lynch syndrome | 2/7 (28.6%) | 3/7 (42.8%) | 2/7 (28.6%) | - | NS |

Data shown in brackets represent median values. † Statistical comparison was performed using Pearson's Chi Square (χ²) test. ‡ Statistical comparison was performed using the Kruskal-Wallis test. § Statistical comparison was performed using analysis of variance (ANOVA). DFS: Disease-free survival. No.: Number. NS: Not significant. OS: Overall survival. SD: Standard deviation. T.: Tumor.

| Table 2. **Global distribution of MSI and CIMP among the different subgroups obtained by unsupervised hierarchical clustering analysis**

| Clusters | Cluster A1 | Cluster A2 | Cluster B3 | Cluster B4 | P-value† |
|----------|------------|------------|------------|------------|----------|
| No. of tumors | 25 | 13 | 10 | 12 | - |
| MSI | 2 (8%) | 3 (23.1%) | 5 (50%) | - | 0.007 |
| Expression of MMR: | | | | | |
| Normal | 23 | 10 | 5 | - | 0.007 |
| Absence | 2 | 3 | 5 | - | NS |
| Lynch Syndrome | 2 | 3 | 2 | - | NS |
| MMR genes affected: | | | | | |
| MLH1 | 1 | - | - | - | - |
| MSH2 | 1 | - | - | - | - |
| MSH6 | 1 | - | - | - | - |
| BRAF mutation | - | - | 1 (100%) | - | NS |
| CIMP-High | 7 (28%) | 4 (30.8%) | - | 1 (8.3%) | NS |
| CIMP-Low/0 | 18 (72%) | 9 (69.2%) | 10 (100%) | 11 (91.7%) | NS |

† Statistical comparison was performed using Pearson's Chi Square (χ²) test.

CIMP: CpG island methylator phenotype. -MMR: Mismatch repair. (Mol.): Molecular. (MSI): Microsatellite instability. (MSS): Microsatellite instability. No.: Number. NS: Not significant.
showed a slightly higher value of GII total and a lower number of CNAs per sample. However, the average aneuploidy frequency was significantly higher in cluster B \( (P = .026; \) Supplementary Table S2). Interestingly, subgroup A1 showed significantly higher values for GII total \( (P = .001) \), GII gains \( (P = .0001) \), GII losses \( (P = .030) \) and average number of CNAs per sample \( (P = .002) \), and contained almost the entire instability associated with cluster A (Supplementary Table S3). By contrast, we did not find significant differences between B3 and B4, despite the higher degree of instability in the first subgroup (Supplementary Table S4). In the comparison of the four secondary subgroups, A2 demonstrated significantly lower values for almost all parameters suggestive of CIN. As a result, it was the subgroup with the highest percentage of CIN- tumors \( (P = .003; \) Table 3).

Concerning aneuploidy, we observed that the vast majority of MACS (also MSS) were included in A1, whereas the vast majority of CIN- tumors were included in A2 (Table 3). As expected, the most unstable subgroup (A1) included the highest proportion of MSS tumors and the most stable subgroup (A2) included the highest proportion of MSI tumors (Tables 2 and 3); this is not surprising, given that CIN and MSI are carcinogenesis pathways traditionally considered as mutually exclusive. Moreover, we observed some aneuploid events that seemed to be more frequent in some subgroups than in others. Thus, alterations in chromosomes 19 and 22 were predominant in subcluster A1 (60%), whereas alterations in chromosomes 4, 11 and 15 were predominant in B3 (40%) and alterations in chromosome 13 were predominant in B4 (67%).

### Recurrent CNAs

Among the most frequent alterations in cluster A were gains at 19p13 (72%) and 9q33-q34 (62%), and losses at 5q13 (67%) and 1q12-q21 (62%), while the most frequent alterations in cluster B were gains at 8q22-q24 and 13q13-q14 (41%), and losses at 12q24 (50%). In addition, there were 63 recurrent alterations (frequencies greater than or equal to 30%) occurring with statistically significantly different frequencies in clusters A and B (Supplementary Table S5). On the other hand, gains at 19p13-p12 and 19q13 (96%) were common in A1, but only losses at 2q12-q13 (64%) were recurrent in A2. Finally, gains at 4q31-q32, 5q14, 12p13-q13 and 13q12 (100%), and losses at 17q11 (80%) were frequent in B3, whereas gains at 1p36-p34, 2p25-p24, 2q11-q14, 2q37, 3p25, 3q21-q23, 7p14, 7q32 and 16p13-p11 (92%) were frequent in B4.

### Molecular Classification

Based on MSI and CIMP status, we defined four molecular groups as described in Materials and Methods [15]. MSS/CIMP-Low/0 tumors were the most frequent whereas MSI/CIMP-High tumors were uncommon (Table 2). Moreover, MSI/CIMP-High tumors were related with LS patients and/or Amsterdam II-positive families.

The unsupervised analysis clustered the samples into four groups according to the similarity of their CNAs. These groups showed a certain degree of correspondence with the molecular classification proposed by Ogino and Goel [15], in such a way that we were able to outline an algorithm by which tumors might be categorized according to their genomic status based on CIMP. Thus, the CIMP-High tumors were mostly contained within clusters A1 or A2 depending on the CIN degree, whereas the CIMP-Low/0 tumors were mostly contained within clusters B3 or B4 depending on the presence/absence of MSI (Fig. 3). Interestingly, the genomic instability of each subcluster mostly affected particular chromosomes allowing us to define chromosomal regions more specifically affected depending on the CIMP/MSI status of the samples. The most frequent events of CIN in A1 involved chromosomes 19 and 22, whereas in B3 they involved chromosomes 4, 11 and 15, and in B4 they involved chromosome 13 (Fig. 3).

### Discussion

In a previous study, we confirmed the existence of substantial differences between the CIN pattern of early-onset and late-onset colorectal tumors, suggesting that genomic profiles may be associated with age of onset in CRC [13]. In the present study, our aim was to investigate whether the CIN profile is also associated with the clinicopathological implications and/or with the global genomic status based on MSI and CIMP in EOCRC. For this purpose, we analyzed a cohort of 60 EOCRCs and performed an unsupervised hierarchical clustering analysis of the aCGH data.

The unsupervised analysis organized samples into two main clusters (A and B) and four secondary clusters (A1, A2, B3, B4) (Fig. 1). No
...the entire instability linked to cluster A. Therefore, the lack of statistical differences in the comparison of the GII \( \text{total} \) between A and B may be understood as a consequence of the masking of the high degree of instability associated to A1, because in a normal distribution the mean of the sampling distribution is equal to the mean of the population. Despite the fact that high levels of CIN have been related to a worse prognosis in CRC\[35\]. In addition, chromosome 13 was the autosome most frequently affected by aneuploidy within subcluster B4 (67%), which may be important given the large clinical and molecular heterogeneity of this subgroup.

Our unsupervised analysis revealed a certain parallelism with the molecular classification proposed by Ogino and Goel\[15\], which on the one hand allowed us to outline an algorithm by which tumors may be categorized according to their genomic status based on CIN, MSI and CIMP. The CIMP-High tumors were mostly contained within clusters A1 or A2 depending on the CIN degree, whereas the CIMP-Low/0 tumors were mostly contained within clusters B3 or B4 depending on the MSI status. Significant differences were observed regarding the clinicopathological features of the tumors included in clusters A and B, and they were overall those expected considering the familial component of each one of them (Table 1). Both MSI and CIMP showed a heterogeneous distribution across the clusters, with CIMP-High tumors mostly stratified into cluster A and CIMP-Low/0 tumors mostly stratified into cluster B (Table 2).

There were no differences either between the subgroups derived from the main clusters, and it was not surprising that the cluster with the highest percentage of metastasis at diagnosis (B4) exhibited the lowest values of OS (overall survival) and DFS (disease-free survival), that is, they had the worst prognosis (Table 1).

It is well known that sporadic tumors commonly develop via the CIN pathway, which is characterized by chromosomal imbalances (aneuploidy) and a high frequency of loss of heterozygosity \[17\]. On this basis, it would be expected that tumors contained in cluster B showed a higher degree of CIN, as has in fact been observed (Table 3). Contrary to expectations, the average number of CNAs per sample was higher in cluster A.

The average of whole chromosomes affected was significantly higher within the cluster with the lowest hereditary component (cluster B; \( P = .026 \)) even though we did not observe statistical differences regarding the GII \( \text{total} \) (Table 3). Interestingly, we observed that A1 contained almost the entire instability linked to cluster A. Therefore, the lack of statistical differences in the comparison of the GII \( \text{total} \) between A and B may be understood as a consequence of the masking of the high degree of instability associated to A1, because in a normal distribution the mean of the sampling distribution is equal to the mean of the population. Despite the fact that high levels of CIN have been related to a worse prognosis in CRC\[18\], in our series the worst clinical outcomes were linked to the cluster with a lower average of GII \( \text{total} \) (cluster A) which showed a significantly higher proportion of tumor recurrence (\( P = .035 \); Fig. 2). Probably, this contradiction is a consequence of the high discrepancy between A1 and A2 in terms of CIN being neutralized when cluster A was evaluated as a whole entity.

Each subcluster manifested prevalence of certain events of CIN. The most frequent aneuploidies in A1 involved chromosomes 19 and 22 (60%) where some genes related to CRC are located (\( \text{MMP11, GSTT2, CHEK2 and EP300} \)). Moreover, about 96% of tumors included in A1 showed gains at 19p13-p12 and 19q13 while only losses at 2q12-q13 (64%) were recurrent in A2. This is an interesting finding because 19p13-p12 and 19q13 harbor genes previously related with the development of malignancy and growth progression of certain tumors such as \( \text{STK11} \) (a tumor suppressor related to Peutz-Jeghers syndrome) or \( \text{MADCAM1} \), whose relationship with inflammatory bowel disease has been reported\[19\]. More interesting, the region 2q12-q13 harbors \( \text{BUB1} \), a gene encoding a mitotic checkpoint serine/threonine-protein kinase whose alterations have been related to early-onset CRC\[20\]. The most frequent aneuploidies observed in B3 affected chromosomes 4, 11, 15 (40%). In this respect, it is important to mention that losses of large parts of chromosome 4 (where some genes important for carcinogenesis appear to be located) are repeatedly reported in several cancers\[21–28\]. Also supporting our results, there are many studies that identify different loci in chromosomes 11 and 15 as candidates for CRC susceptibility\[29–34\]. Finally, it is worth mentioning chromosome 13 since all tumors of B3 showed alterations in 13q12. This region contains \( \text{CDX2} \), a specific transcription factor expressed during intestinal development whose lack of expression appears to be associated with a poor prognosis in CRC\[35\]. In addition, chromosome 13 was the autosome most frequently affected by aneuploidy within subcluster B4 (67%), which may be important given the large clinical and molecular heterogeneity of this subgroup.

Acknowledgements
We thank the Tumor Registry of the Pathology Department of the 12 de Octubre University Hospital for providing the paraffin-embedded tissues, and Ron Hartong for his help with the English revision of this manuscript.

**Fig. 3.** Algorithm by which tumors may be categorized according to their genomic status based on CIN, MSI and CIMP. The CIMP-High tumors were mostly contained within clusters A1 or A2 depending on the CIN degree, whereas the CIMP-Low/0 tumors were mostly contained within clusters B3 or B4 depending on the MSI status.
Appendix A. Supplementary Data
Supplementary data to this article can be found online at http://dx.
doi.org/10.1016/j.neo.2016.11.006.

References

[1] Jemal A, Bray F, Center MM, Ferlay J, Ward E, and Forman D (2011). Global cancer statistics. CA Cancer J Clin 61(2), 69–90. http://dx.doi.org/10.3322/caac.201070.
[2] Ferlay J, Shin HR, Bray F, Forman D, Mathers C, and Parkin DM (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 127(12), 2893–2917. http://dx.doi.org/10.1002/ijc.25516.
[3] Leslie A, Carey FA, Peart NR, and Steele RJ (2002). The colorectal adenoma-carcinoma sequence. Br J Surg 89(7), 845–860. http://dx.doi.org/10.1053/bjs.2002.02102.
[4] Boyle P and Ferlay J (2005). Cancer incidence and mortality in Europe. 2004. Ann Oncol 16(3), 481–488. http://dx.doi.org/10.1093/annonc/mdi098.
[5] Siegel RL, Jemal A, and Ward EM (2009). Increase in incidence of colorectal cancer among young men and women in the United States. Cancer Epidemiol Biomarkers Prev 18(6), 1695–1698. http://dx.doi.org/10.1158/1055-9966.EPI-09-0186.
[6] Sia CS, Paul E, Wale RJ, Lynch AC, Heriot AG, and Warren SR (2014). No increase in colorectal cancer in patients under 50 years of age: a Victorian experience from the last decade. Colorectal Dis 16, 690–695. http://dx.doi.org/10.1111/cod.12146.
[7] Ahnen DJ, Wade SW, Jones WF, Sifri R, Mendoza Silveiras J, Greenamyer J, Guiuiffe S, Axilbund J, Spiegel A, and You YN (2014). The increasing incidence of young-onset colorectal cancer. A call to action. Mayo Clin Proc 89, 216–224. http://dx.doi.org/10.1016/j.mayocp.2013.09.006.
[8] Banerjee A, Hands RE, Powar MP, Bustin SA, and Dorudi S (2009). Microsatellite and chromosomal stable colorectal cancers demonstrate poor immunogenicity and early disease recurrence. Colorectal Dis 11(6), 601–608. http://dx.doi.org/10.1111/j.1463-1318.2008.01639.x.
[9] Pereja J, Rueda D, Canal A, Rodriguez Y, Alvaro E, Osorio I, Alegre C, Rivera B, Martinez J, and Benitez J et al (2014). Age at Onset Should Be a Major Criterion for Subclassification of Colorectal Cancer. J Mol Diagn 16(1), 116–126. http://dx.doi.org/10.1016/j.jmoldx.2013.07.010.
[10] Hawkins NJ, Tomlinson I, Meagher A, and Ward RL (2001). Microsatellite-stable diploid carcinoma: a biologically distinct and aggressive subset of sporadic colorectal cancer. Br J Cancer 84(2), 232–236. http://dx.doi.org/10.1054/bjoc.2000.1554.
[11] Oliveira C, Velho S, Mostinlho C, Ferreia A, Preto A, Domingo E, Capelinha AF, Duval A, Hamelin R, and Machado JC, et al (2007). KRAS and BRAF sequence. http://dx.doi.org/10.1054/bjoc.2000.1554.
[12] Hawkins NJ, Tomlinson I, Meagher A, and Ward RL (2001). Microsatellite-stable diploid carcinoma: a biologically distinct and aggressive subset of sporadic colorectal cancer. Br J Cancer 84(2), 232–236. http://dx.doi.org/10.1054/bjoc.2000.1554.
[13] Arribas M, García JL, Inglada-Pérez L, Rueda D, Osorio I, Rodríguez Y, Álvaro E, et al (2008). Deletion mapping of chromosome 4q22-35 and identification of four frequently deleted regions in head and neck cancers. Neoplasia 10(4), 299–304.
[14] Krona C, Carén H, SJöberg RM, Sandstedt B, Laureys G, Kogner P, and Osterlund P (2010). Frequent loss of heterozygosity in the 9p21.3 region. Cancer Res 70(1), 348–357. http://dx.doi.org/10.1158/0008-5472.CAN-09-3388.
[15] Benitez J, Garcia de la Concha JM, and Navarro C (2002). Cytogenetic analysis of colorectal cancer. Genes Chromosomes Cancer 33(1), 1–15. http://dx.doi.org/10.1002/gcc.10021.
[16] de Voer RM, Geurts van Kessel A, Weren RD, Ligtenberg MJ, Smeets D, Fu L, Vreede L, Kamminga J, Verwel IE, and Hahn MM, et al (2013). Germline mutations in the spindle assembly checkpoint genes BUB1 and BUB3 are risk factors for colorectal cancer. Gut 62(6), 898–906. http://dx.doi.org/10.1136/gutjnl-2012-302885.
[17] de Voer RM, Geurts van Kessel A, Weren RD, Ligtenberg MJ, Smeets D, Fu L, Vreede L, Kamminga J, Verwel IE, and Hahn MM, et al (2013). Germline mutations in the spindle assembly checkpoint genes BUB1 and BUB3 are risk factors for colorectal cancer. Gut 62(6), 898–906. http://dx.doi.org/10.1136/gutjnl-2012-302885.
[18] de Voer RM, Geurts van Kessel A, Weren RD, Ligtenberg MJ, Smeets D, Fu L, Vreede L, Kamminga J, Verwel IE, and Hahn MM, et al (2013). Germline mutations in the spindle assembly checkpoint genes BUB1 and BUB3 are risk factors for colorectal cancer. Gut 62(6), 898–906. http://dx.doi.org/10.1136/gutjnl-2012-302885.
[19] de Voer RM, Geurts van Kessel A, Weren RD, Ligtenberg MJ, Smeets D, Fu L, Vreede L, Kamminga J, Verwel IE, and Hahn MM, et al (2013). Germline mutations in the spindle assembly checkpoint genes BUB1 and BUB3 are risk factors for colorectal cancer. Gut 62(6), 898–906. http://dx.doi.org/10.1136/gutjnl-2012-302885.