NEOPLASIA
www.neoplasia.com

Volume 19 Number xx Month 2016 pp. 28–34 28

Unsupervised Analysis of Array Comparative Genomic Hybridization Data from Early-Onset Colorectal Cancer Reveals Equivalence with Molecular Classification and Phenotypes[1]

María Arriba[*,2], Juan L. García[†,‡,2], Daniel Rueda[§], Jessica Pérez[†,‡], Lorena Brandariz[¶], Oana A. Nutu[¶], Laura Alonso[¶], Yolanda Rodríguez[#], Miguel Urioste[,††], Rogelio González-Sarmiento[†,‡] and José Perea[*,¶]**

[*]Centre for Biomedical Research of the 12 de Octubre University Hospital, Avda. de Córdoba, S/N, 28041, Madrid, Spain; [†]Biomedical Research Institute of Salamanca (IBSAL). University Hospital of Salamanca-USAL-CSIC, P° de San Vicente, 58-182, 37007, Salamanca, Spain; [‡]Institute of Molecular and Cellular Biology of Cancer (IBMCC), University of Salamanca-CSIC, Campus Miguel de Unamuno, 37007, Salamanca, Spain; [§]Molecular Biology Laboratory, 12 de Octubre University Hospital, Avda. de Córdoba, S/N, 28041, Madrid, Spain; [¶]Department of Surgery, 12 de Octubre University Hospital. Avda. de Córdoba, S/N, 28041, Madrid, Spain; [#]Department of Pathology, 12 de Octubre University Hospital, Avda. de Córdoba, S/N, 28041, Madrid, Spain; [**]Familial Cancer Clinical Unit, Human Cancer Genetics Program. Spanish National Cancer Research Centre (CNIO), C/Melchor Fernández Almagro, 3, 28029, Madrid, Spain; [††]Center for Biomedical Network Research on Rare Diseases (CIBERER), Institute of Health Carlos III, C/Monforte de Lemos 3-5, Pabellón 11, 28029, Madrid, Spain

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/Rosas 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

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 depending on the CIMP/MSI status of the samples. *CONCLUSIONS*: Our findings may provide a basis for a new form of classifying EOCRC according to the genomic status of the tumors.

Neoplasia (2016) 19, 28–34

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 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 death 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 (χ^2) test. For comparisons between more than two groups, analysis of variance (ANOVA) (for normal distributions) or the Kruskal-Wallis test (for