Genomics xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Identification of chromosomal copy number variations and novel candidate loci in hereditary nonpolyposis colorectal cancer with mismatch repair proficiency

Weixiang Chen [a,b,c,1], Lin Yuan [a,b], Ying Cai [a,b], Xiaocheng Chen [a,b], Yayun Chi [a,b], Ping Wei [a,b], Xiaoyan Zhou [a,b,*], Daren Shi [a,b]

[a] Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai, People's Republic of China
[b] Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, People's Republic of China
[c] Department of Pathology, Shanghai Gongli Hospital, Shanghai, People's Republic of China

ARTICLE INFO

Article history:
Received 29 October 2012
Accepted 7 February 2013
Available online xxxx

Keywords:
Colorectal neoplasms, hereditary nonpolyposis
Copy number variation
CytoScan HD Array
Genome-wide analysis

ABSTRACT

The pathogenesis of microsatellite stable hereditary non-polyposis colorectal cancers (MSS HNPCC) is unclear. To identify genomic regions that might be involved in MSS HNPCC pathogenesis, we selected 20 pairs of MSS HNPCC for a genome-wide study using copy number variation targeted (CNV-targeted) CytoScan HD Array. A remarkably increased frequency of 20q gain (70%) and high levels of copy-neutral loss of heterozygosity (40%) were observed. The most frequent tumor-specific CNVs included amplifications (7p21.3-15.1, 8q13.3-24.3, 13q14.1-33.3 and 20q12-13.33) and deletions (8p11.23-23.1, 15q11.2-26.1, 17p13.1-13.3 and 18q11.2-21.33). In addition, 10 novel CNVs were discovered and led to identification of WDR16 and RAPGEF5 as candidate genes involved in tumorigenesis, displaying a robust correlation between expression and genomic alterations. Moreover, WDR16 and RAPGEF5 exhibited altered protein expression levels as assessed by immunohistochemistry (IHC) in 41 other independent samples. Finally, high consistency (68–84%) was observed between CNVs by Array and quantitative PCR. These findings are important for further elucidating MSS HNPCC pathogenesis.

1. Introduction

Hereditary non-polyposis colorectal cancer (HNPCC) syndrome was initially defined according to data based on family history. Over the last several years, germline mutations in mismatch repair (MMR) genes (MSH2, MLH1, MSH6, PMS1, and PMS2) have been identified as the causative factors for this syndrome. Families with cancers that segregate with germline MMR gene mutations are defined as having Lynch syndrome. However, approximately half of the families that fulfill the more stringent revised Amsterdam criteria for HNPCC do not have evidence of MMR deficiency, and therefore, their tumors are considered microsatellite stable (MSS) [1]. Clinically defined HNPCC patients with the MSS phenotype (MSS HNPCC) have clinicopathological characteristics and molecular features that are distinct from those of HNPCC patients with the microsatellite instability (MSI) phenotype (MSI HNPCC) and from those of patients with sporadic colorectal cancers; these factors suggest that other unknown genes could be involved [2,3]. Because of the gap between the clinical classification and current molecular diagnostic strategies for familial colorectal cancers and the limited treatment options available to these patients, there is an urgent need to further investigate the molecular events involved in MSS HNPCC.

Comprehensive knowledge of the genomic alteration events responsible for cancer is a critical foundation for its diagnosis and prognosis and for developing targeted therapeutics. Recently, copy number variation (CNV) has been recognized as one of the most important genomic alterations that plays a role in cancer pathogenesis [4]. In addition, somatic CNVs can be used to identify regions of the genome that are involved in disease phenotypes [5,6]. Array comparative genome hybridization (aCGH) is widely used to identify copy number variations in genomes [7,8]. However, this technique is limited to identifying homozygous deletions and cannot distinguish between paternal and maternal recombination events [9,10]. SNP genotyping array technology allows for the combined analysis of both copy number and loss of heterozygosity (LOH) (including classical LOH and copy-neutral loss of heterozygosity (cnLOH)) throughout the genome, overcoming the limitations of aCGH [11,12]. However, the commercial SNP genotyping arrays focus on variants that are present in 5% or more of the population and feature a limited number of CNV probes. Therefore, submicroscopic structural variants are poorly captured by

Abbreviations: HNPCC, hereditary non-polyposis colorectal cancers; MMR, mismatch repair; MSS, microsatellite stable; MSI, microsatellite instability; CNV, copy number variation; LOH, loss of heterozygosity; cnLOH, copy-neutral loss of heterozygosity; qPCR, quantitative PCR.

* Corresponding author at: Department of Pathology, Fudan University Shanghai Cancer Center, 270 Dong'an Road, Shanghai 200032, People's Republic of China.
E-mail addresses: cwx7511@sohu.com (W. Chen), xyzhou100@yahoo.com (X. Zhou).
[1] Department of Pathology, Fudan University Shanghai Cancer Center, 270 Dong'an Road, Shanghai 200032, People's Republic of China.

0888-7543/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2013.02.003

available SNP genotyping arrays that were designed to evaluate SNPs [13]. The recent introduction of the Affymetrix® CytoScan™ HD Array (CNV-targeted array), which is based on the validated Genome-Wide Human SNP Array 6.0 and contains more than 2.6 million markers for copy number variants and approximately 750,000 SNPs, has enabled the detection of copy number aberrations with high resolution across the genome. In addition, the CytoScan™ HD Array provides allelic imbalance information from SNPs. This array has great power to detect known and novel chromosome aberrations across the entire human genome and features unbiased whole-genome coverage, with excellent performance across the entire genome.

In the current study, we performed Genome-Wide CNV analysis on MSS HNPCC and paired normal colorectal samples with the Affymetrix® CytoScan™ Human CytoScan HD Array to identify abnormal regions of the genome that might be involved in MSS HNPCC pathogenesis. The ultimate goal of this study was to improve the understanding of the molecular mechanisms of MSS HNPCC and to develop more effective molecular markers for the diagnosis, treatment and prevention of this type of cancer. To our knowledge, this is the first comprehensive study of the genetic alterations found in MSS HNPCC to be performed using a powerful CNV-targeted array platform.

2. Materials and methods

2.1. Patient selection and DNA extraction

A total of 126 colorectal cancer patients with a family history suggestive of HNPCC syndrome and who underwent surgery between 2008 and 2011 at the Shanghai Cancer Center (Shanghai, China) were enrolled in this study. All of these patients were from unrelated families and fulfilled stringent clinical criteria: Amsterdam I/II or Amsterdam borderline (families fulfilling all Amsterdam I/II criteria with the exceptions that the maximum allowed age of diagnosis was extended to 60 years old and that gastric cancer could be present). Fresh frozen colorectal carcinomas and corresponding normal tissues (12 blood lymphocyte samples and 8 normal mucosa samples from more than 5 cm away from primary tumor of the same individual) were collected from the tissue bank of the Shanghai Cancer Center (Shanghai, China). The samples used contained more than 90% cancerous or mucosal cells. Genomic DNA was extracted using a QIAGEN DNA purification kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. This study was approved by the institutional review board of the Fudan University Shanghai Cancer Center, and patient consent was obtained for the release of all medical records and tissues.

2.2. Microsatellite analysis and immunohistochemical analysis of MMR proteins

According to the international guidelines, a panel of microsatellite markers, including the mononucleotide repeat markers BAT25 and BAT26 and the dinucleotide repeat markers D5S346, D2S123, and D17S250, were used to classify MSI. Tumors were defined as MSI-H if two or more of the markers exhibited instability and as MSI-L if fewer than two markers were positive for instability. Tumors with no instability in any of the markers were defined as MSS [14].

To assess the expression status of the MMR proteins (MLH1, MSH2, MSH6 and PMS2), immunohistochemical staining was performed on sections of formalin-fixed and paraffin-embedded tumor tissues. The whole tissue slides were stained with antibodies against MLH1 (clone G168-15, dilution 1:30; BD Pharmingen, San Diego, CA), MSH2 (clone GB-12, dilution 1:100; Oncogene Research Products, Boston, MA), MSH6 (clone 44, dilution 1:400; BD Biosciences, San Jose, CA), and PMS2 (clone A16-4, 1:50; BD Biosciences, San Jose, CA) according to standard protocols and procedures as indicated by the manufacturer. Although few tumor cells exhibited nuclear staining equivalent to that of normal tissue, mismatch repair proteins were considered to be intact and normally expressed.

2.3. CytoScan HD Array and copy number analysis

The Genome-Wide Human CytoScan HD Array (Affymetrix, CA, USA) was used to analyze genomic alterations according to the manufacturer's protocol. Briefly, 250 ng of genomic DNA from tumor and matched normal samples was digested with the restriction enzyme *Nsp*I and then ligated to an adapter, followed by PCR amplification using a single pair of primers that recognized the adapter sequence. The PCR products were run on a 2% TBE gel to confirm that the majority of products were between 150 and 2000 bp in length. To obtain a sufficient quantity of PCR product for further analysis, all products from each sample were combined and purified using magnetic beads (Agencourt AMPure, Beckman Coulter, Beverly, MA). The purified PCR products were fragmented using DNase I and visualized on a 4% TBE agarose gel to confirm that the fragment sizes ranged from 25 to 125 bp. The fragmented PCR products were subsequently end-labeled with biotin and hybridized to the array. Arrays were then washed and stained using a GeneChip® Fluidics Station 450 and scanned using an Affymetrix GeneChip® Scanner 3000 7G. Scanned data files were generated using Affymetrix GeneChip Command Console Software, version 1.2, and analyzed with Affymetrix® Chromosome Analysis Suite v1.2 (ChAS) (Affymetrix Inc., USA).

To calculate copy numbers, the data were normalized to baseline reference intensities using 270 HapMap samples and another 90 healthy normal individuals. The Hidden Markov Model (HMM) available within the software package was used to determine the copy number states and their breakpoints. Thresholds of $\log_2$ ratio ≥ 0.58 and ≤ -1 were used to categorize altered regions as CNV gains (amplification) and copy number losses (deletions), respectively.

To prevent the detection of false positive CNVs arising due to inherent microarray "noise", only alterations that involved at least 25 consecutive probes and that were more than 50 kbp in length were considered in the analysis of gains or losses in our study. Amplifications and deletions were analyzed separately. To exclude aberrations representing common normal CNVs, all the identified CNVs were compared with those reported in the Database of Genomic Variants (DGV, http://projects.tcag.ca/variation/).

To identify the genes involved in the CNVs further, we queried the UCSC database (http://genome.ucsc.edu), Ensemble (http://www.ensembl.org), and BioGPS (http://biogps.gnf.org). Gene annotation and gene overlap were determined using the human genome build 19 (hg19) and several widely used online databases (Ensembl: http://www.ensembl.org; UCSC: http://genome.ucsc.edu; and NetAffx: http://www.affymetrix.com).

2.4. Candidate CNV analysis

We compared overlapping regions in tumor and in normal samples to identify tumor-specific CNVs. Tumor-specific CNVs were considered possible candidate variants for MSS HNPCC if they fulfilled all of the following criteria: (a) there was overlap between the CNV and any RefSeq gene or Affymetrix-predicted promoter, in particular those genes/promoters associated with carcinogenesis or involved in pathways suspected to be affected in colorectal cancer; and (b) there was a statistically significant difference between the frequency of the CNV in the tumor and normal samples ($p<0.05$). If the candidate variants did not appear in the DGV database, they were considered rare novel candidate CNVs.

2.5. LOH analysis

For LOH analysis, we utilized the algorithm incorporated in the ChAS software. Because matched normal DNA was available for all

Please cite this article as: W. Chen, et al., Identification of chromosomal copy number variations and novel candidate loci in hereditary nonpolyposis colorectal cancer with mismatch repair proficiencyGenomics (2013), http://dx.doi.org/10.1016/j.ygeno.2013.02.003