The Signatures of Autozygosity among Patients with Colorectal Cancer

Manny D. Bacolod,[1] Gunter S. Schemmann,[7] Shuang Wang,[2] Richard Shattock,[1] Sarah F. Giardina,[1] Zhaoshi Zeng,[3] Jinru Shia,[4] Robert F. Stengel,[7] Norman Gerry,[10] Josephine Hoh,[11] Tomas Kirchhoff,[5] Bert Gold,[9] Michael F. Christman,[10] Kenneth Offit,[5] William L. Gerald,[4] Daniel A. Notterman,[8] Jurg Ott,[6] Philip B. Paty,[3] and Francis Barany[1]

[1]Department of Microbiology, Weill Medical College of Cornell University; [2]Department of Biostatistics, Mailman School of Public Health, Columbia University; Departments of [3]Surgery, [4]Pathology, and [5]Medicine, Memorial Sloan-Kettering Cancer Center; [6]Laboratory of Statistical Genetics, Rockefeller University, New York, New York; [7]School of Engineering and Applied Science and [8]Department of Molecular Biology, Princeton University, Princeton, New Jersey; [9]Laboratory of Genomic Diversity, National Cancer Institute at Frederick, Frederick, Maryland; [10]Department of Genetics and Genomics, Boston University, Boston, Massachusetts; and [11]School of Public Health, Yale University, New Haven, Connecticut

Abstract

Previous studies have shown that among populations with a high rate of consanguinity, there is a significant increase in the prevalence of cancer. Single nucleotide polymorphism (SNP) array data (Affymetrix, 50K *Xba*I) analysis revealed long regions of homozygosity in genomic DNAs taken from tumor and matched normal tissues of colorectal cancer (CRC) patients. The presence of these regions in the genome may indicate levels of consanguinity in the individual's family lineage. We refer to these autozygous regions as identity-by-descent (IBD) segments. In this study, we compared IBD segments in 74 mostly Caucasian CRC patients (mean age of 66 years) to two control data sets: (*a*) 146 Caucasian individuals (mean age of 80 years) who participated in an age-related macular degeneration (AMD) study and (*b*) 118 cancer-free Caucasian individuals from the Framingham Heart Study (mean age of 67 years). Our results show that the percentage of CRC patients with IBD segments (≥4 Mb length and 50 SNPs probed) in the genome is at least twice as high as the AMD or Framingham control groups. Also, the average length of these IBD regions in the CRC patients is more than twice the length of the two control data sets. Compared with control groups, IBD segments are found to be more common among individuals of Jewish background. We believe that these IBD segments within CRC patients are likely to harbor important CRC-related genes with low-penetrance SNPs and/or mutations, and, indeed, two recently identified CRC predisposition SNPs in the 8q24 region were confirmed to be homozygous in one particular patient carrying an IBD segment covering the region. [Cancer Res 2008;68(8):2610–21]

Introduction

Colorectal cancer (CRC) is one of the four most prevalent cancers in the United States. In 2007, there will be 153,760 new cases of CRC in the United States, resulting in 52,180 deaths (1). According to a recent worldwide statistical compilation, over a million people suffered from the disease in 2002, with the majority of cases in industrialized countries (2). Genetics aside, the incidence of CRC correlates with diets rich in fat and calories, and low in vegetables, fruits, and fibers as well as alcohol consumption and smoking (3). Traditionally, CRC cases are divided into two categories: sporadic and familial (or hereditary; ref. 4). Approximately 70% of the cases are classified as sporadic, afflicting people with apparently no family history of the disease. Of the familial cases, the two most commonly occurring are familial adenomatous polyposis (FAP) and hereditary nonpolyposis CRC (HNPCC). FAP, characterized by formation of polyps within the gastrointestinal tracts of affected individuals, is caused by highly penetrant, autosomal dominant germ line mutations in the *adenomatous polyposis coli* (*APC*) gene, and can account for ~1% of all CRCs (5). HNPCC (Lynch syndrome) cases, seen in as many as 2.5% of all CRCs, are caused by highly penetrant mutations in DNA mismatch repair genes (primarily *MLH1* and *MSH2*; ref. 6). Much less common genetically linked CRCs are those arising from hamartomatous polyp syndromes such as juvenile polyposis, Peutz-Jeghers, and Cowden's, which are caused by mutations in *SMAD4* (7), *STK11* (8), and *PTEN* (9), respectively. However, the exact genetic causes of a great percentage of familial CRCs remain undiscovered and likely due to low penetrating alleles. Moreover, the distinction between spontaneous and familial CRCs may be understated. Some cancers classified as sporadic cases may in fact have underlying genetic components (4, 10). Several statistical analyses of huge cancer databases have attempted to quantify the heritable components of cancers. The cohort studies from Sweden (11) and Utah (12) showed that the CRC family risk ratios, which is a direct measure of heritability (13), are 4.41 (considered high) and 2.54 (considered moderate), respectively. In addition, the Scandinavian twin study (involving a little less than 45,000 pairs of twins) showed that hereditary factors affect colon cancer 35% of the time (14). Therefore, a significant proportion of heritable CRC remains unaccounted for.

Our research group aims to characterize CRCs using a variety of molecular techniques, including expression profiling (15, 16), methylation profiling (17), mutational scanning (18, 19), and single nucleotide polymorphism (SNP) array-based chromosomal analysis (15). The latter technique (Affymetrix Human Mapping array) readily reveals cancer tissue chromosomal aberrations such as amplifications and loss of heterozygosity (LOH). Indeed, we initially set out to identify regions of varying amplification and to determine if any correlation existed between these chromosomal

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Francis Barany, Department of Microbiology and Immunology, Cornell University Weill Medical College, New York, NY 10021. Phone: 212-746-6509; Fax: 212-746-7983; E-mail: barany@med.cornell.edu.

doi:10.1158/0008-5472.CAN-07-5250

aberrations and expression data generated from the same samples. When we began to examine the results of individual patients comparing their copy number and LOH between the tumor and the matched normal, we often saw chromosomal gains and losses in the tumor but not in the normal tissue (Fig. 1*A*). Surprisingly, we often identified samples where the regions of high homozygosity [identity-by-descent (IBD) segments] were found in both the tumor and the corresponding normal mucosa (Fig. 1*B*). These homozygous segments are most probably indicators of an individual's autozygosity—an indication that parents share a common ancestor

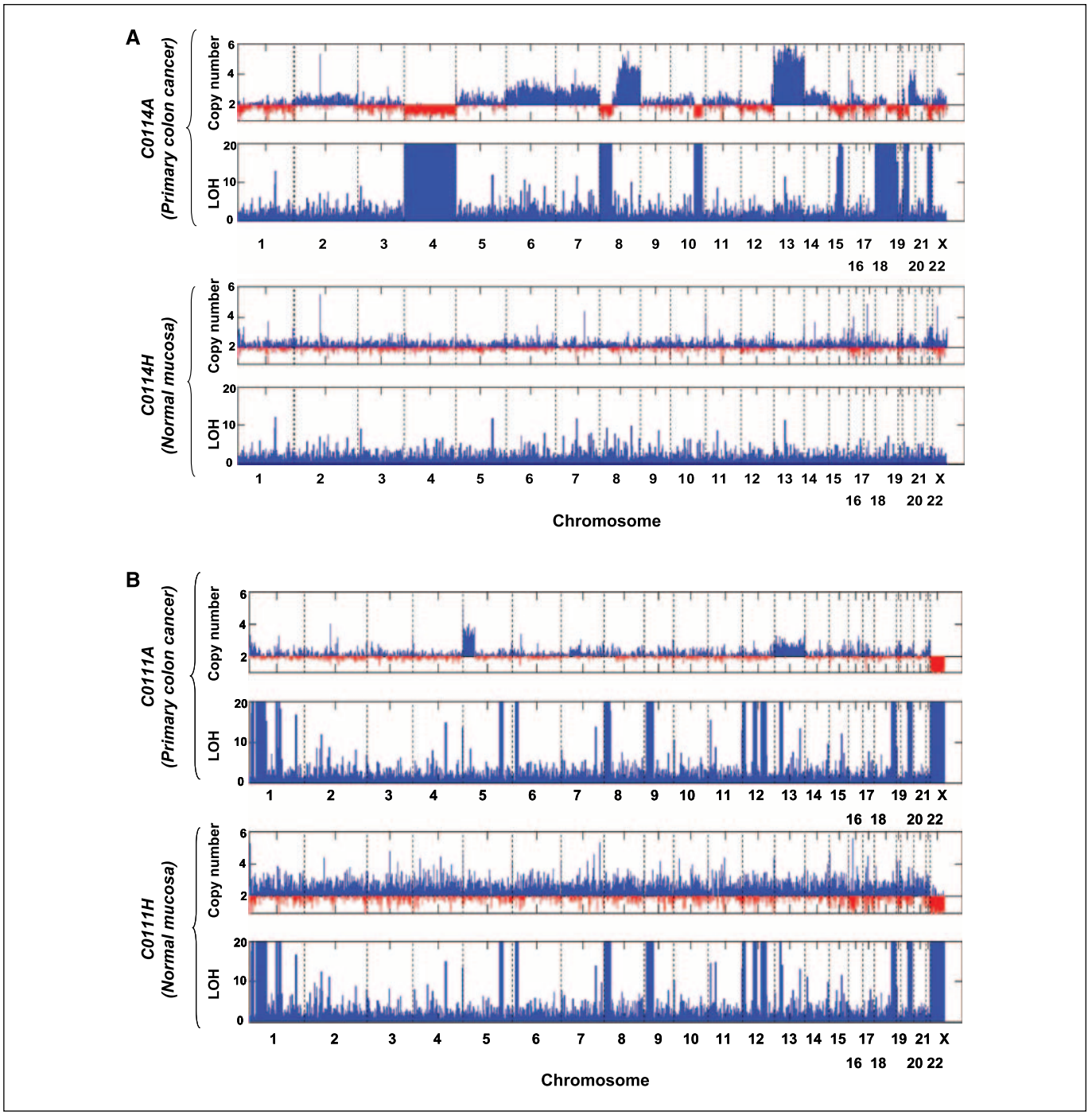


Figure 1. *A,* SNP array (Affymetrix Xba 240 50K) whole-genome analysis of the colon cancer tissue C0114A and its matching normal mucosa C0114H. The charts (copy number and LOH) were generated by Affymetrix CNAT version 3 (26). The aberrations in C0114A include losses in chromosomes 4, 22, 8p, and parts of 10 and 20p, as well as gains in chromosomes 13 and 20q. The copy number chart indicates deviations from the normal copy number of 2 (baseline of the chart). High LOH values (for the charts, the LOH value is capped at 20), indicated by tall blue bars represent segments in the chromosome of contiguous homozygous SNPs. In the CRC sample, regions of copy loss usually correspond to regions of high LOH. The matching normal (C0114H) indicates neutral copy number (equal populations of red and blue bars only represent noise) throughout the genome. *B,* SNP array (Affymetrix Xba 240 50K) whole-genome analysis of the colon cancer tissue C0111A and its matching normal mucosa C0111H. Unlike in C0114A (*A*), the regions of high homozygosity in C0111A can also be found in its corresponding normal mucosa (C0111H). These homozygous segments may in fact be indicator of genomic autozygosity in the patient C0111.

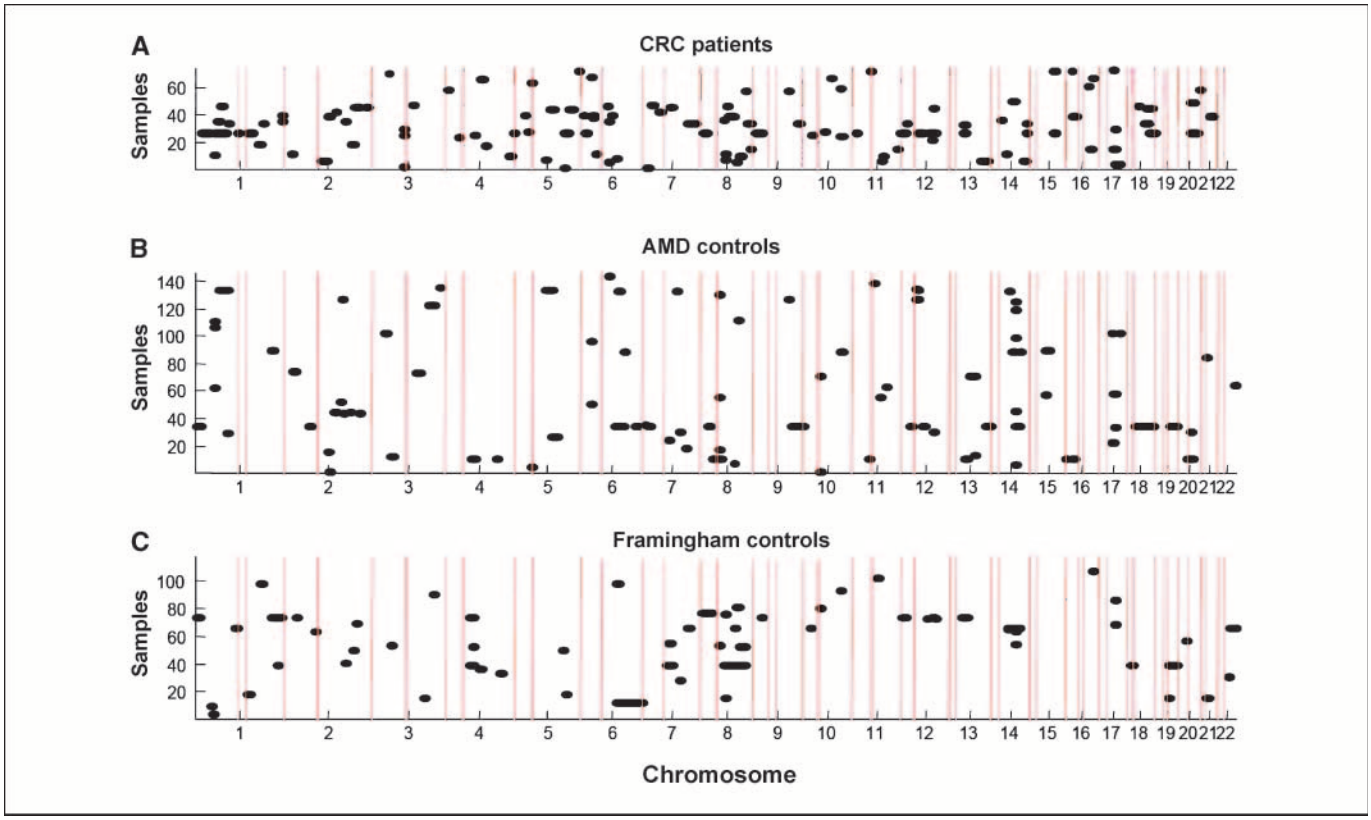


Figure 2. The locations of the identified IBD segments (*black horizontal bars*) among the genomes of the 74 colon cancer patients (*A*), and the AMD (*B*) and Framingham (*C*) control data sets. The threshold limit was set to a minimum of 4-Mb length encompassing at least 50 consecutive homozygous SNPs (but allowing at most 2% errors).

(20). The question was then raised whether these signatures of autozygosity occur more frequently among CRC patients. Several studies have shown that cancer occurs more frequently among groups with higher degrees of consanguinity, that is, groups that share a common ancestor. Among these studies is the comparison of incidence of cancer and other late-onset complex diseases between individuals from genetically isolated islands in middle Dalmatia, Croatia, and a control population (21). The investigators concluded that inbreeding can be a positive predictor for a number of late-onset diseases such as heart disease, stroke, and cancer. The same observations were noted in a Pakistani study where cancer patients, on average, have higher coefficient of inbreeding compared with the general population (22). In another study involving descendants of an Italian immigrant group in Wisconsin, 94% of the subjects with reported adenocarcinomas (mostly colorectal) were products of consanguineous parentage (23). The detrimental effects of inbreeding have been known throughout mankind's history, but most studies have focused on how inbreeding causes rare Mendelian diseases. The effect of inbreeding on cancer is likely more complex than a simple Mendelian genetics, with many more genetic components involved. Nonetheless, studying these genetically isolated populations may eventually lead to discovery of other genes that contribute to cancer predisposition. It is the same argument backed by a growing number of researchers who believe that studying the genetics of purebred dogs known to have high incidence of cancer may eventually help in the discovery of cancer-related genes in humans (24).

There are a number of ways to measure an individual's degree of consanguinity (25). Rudan and coworkers (21) used Wright's path method in measuring average inbreeding coefficients for both the case and control populations. Using short tandem repeat polymorphisms (STRP) as markers, Broman and Weber described the presence of homozygous segments in some individuals from reference families genotyped by Centre d'Etude du Polymorphisme Humain (CEPH). With the advent of high-density human SNP arrays (also known as genotyping arrays), the process of identifying the homozygous segments in the genome has become easier (26). In this study,[12] we show that signatures of autozygosity correlate to CRC incidence and that these IBD regions may be the locations of genes that contribute to CRC heritability.

Materials and Methods

Tissue Acquisition, Sample Selection, and DNA Extraction

Tissue acquisitions followed the protocols of the institutional review boards of Memorial Sloan Kettering Cancer Center (MSKCC) and Cornell University Weill Medical College (institutional review board nos. 0201005297 and 9807003424). Our initial objective was to study the chromosomal aberrations (copy number changes, LOH) in CRCs using high-density SNP mapping arrays (Affymetrix). Based on the pathologist reports (MSKCC), 74 CRC samples showing ≥70% pure tumor cells were chosen for SNP array analysis. Most of these samples were from Caucasian patients (average age

[12] We previously reported these observations in the following scientific meetings: (*a*) AACR special conference: Advances in Colon Cancer Research (poster presentation, Cambridge, MA, November 14–17, 2007); (*b*) AACR Colorectal Cancer: Molecular Pathways and Therapies (poster presentation, Dana Point, CA, October 19–23, 2005); and (*c*) Chips-to-Hits IBC Meeting (F. Barany as invited speaker; Boston, MA; September 27, 2006).

of 66 ± 12 years), described in detail in Supplements S-A1 and S-A2. We initially examined the chromosomal aberration profiles (copy number, LOH) of these 74 tumor samples using the Affymetrix copy number analysis tool (CNAT). The presence of long stretches of homozygous segments (with CNAT LOH values of ~20, examples of which are shown in Fig. 1*B*) in copy neutral regions in the genomes of these samples prompted us to examine the chromosomal profiles of the normal tissues matching these tumors as well. Evidently, 22 of the 74 CRC samples did not exhibit these long stretches of homozygosity. For cost-reducing purposes, it was decided not to run the matching normal tissues for these 22 samples. Therefore, the IBD segment analysis of the 74 patients would come from Affymetrix SNP array data from (*a*) the 22 low LOH CRC tissues and (*b*) normal tissues (normal mucosa, normal lung, or normal liver) of the remaining 52 patients. The inclusion of data from those 22 CRC samples is explained in Supplements S-B2 and S-C2. All genomic DNAs were extracted from snap-frozen tissues that had been prepared and stored at MSKCC as described in previous studies (17, 19).

SNP Array Procedure

The procedure for the Affymetrix GeneChip Human Mapping 50K SNP array was carried out according to the manufacturer's guidelines. Briefly, 0.25 μg of genomic DNA was digested with *Xba*I. The digests were then ligated to oligonucleotide adapters, PCR-amplified (such that the amplicons were in the range of 250–2,000 bp), fragmented, biotin-labeled, and hybridized to the array for 16 h. Following hybridization, the array chips were washed and then stained with streptavidin-phycoerythrin and a biotinylated anti-streptavidin antibody in the Affymetrix Fluidics Station 450. The arrays were scanned in GeneChip Scanner 3000 to generate the image (DAT) and cell intensity (CEL) files. The CEL files were imported to GeneChip Genotyping Analysis Software 4.0 (GTYPE 4.0, Affymetrix) to generate the SNP calls using the dynamic model mapping algorithm (27). It should be noted that the analyses of the current study were undertaken before the release of GTYPE 4.1 and its new Bayesian robust linear model with Mahalanobis distance classifier algorithm.

Use of Other Control Data Sets

To determine the frequency of IBD on the general population, we used the following controls: (*a*) The age-related macular degeneration (AMD) data set representing the 146 non-Hispanic Caucasian individuals who participated in AMD study (28). These include the 96 cases (mean age 79 ± 5.2 years old) and the 50 controls (mean age 82 ± 2.2 years old). (*b*) The Framingham data set: 118 Caucasian individuals who are a subset of the National Heart, Lung, and Blood Institute Framingham Heart Study (ages 61–81 years; ref. 29). Clinical data indicate that the individuals in the Framingham data set had no known cancer at the time they participated in the study. Other control data sets used were SNP array data from (*a*) 30 Ashkenazi Jewish group afflicted with breast cancer (AJBC), (*b*) 133 Ashkenazi Jewish group with no incidence of cancer (AJNC), and (*c*) the subgroup of 48 Caucasian individuals in the Affymetrix reference data set. The last control group did not have any available clinical information. Detailed analyses of the Ashkenazi Jewish data sets are to be described elsewhere.[13] We also examined the possibility of population stratification between our CRC cohort and either the AMD and Framingham data set using the EIGENSTRAT method (30).

Identifying the IBD Segments

Method 1: homozygosity detection. In the Affymetrix CNAT program, the fraction of homozygous SNPs (AA or BB) among all the Affymetrix reference samples at a given base position is the same as the probability the same SNP will be homozygous in the sample in question (26). With this premise, the measure of LOH in the CNAT program is specifically defined as "−log of the probability that contiguous SNPs from *m* to *n* are all homozygous." One of the shortcomings of this algorithm is its high sensitivity to erroneous calls. We therefore used an alternative way to measure homozygosity. Our algorithm looks for regions of autozygosity by searching for consecutive homozygous SNPs in the region, taking into account a 2% error (where at most 2% of SNP calls within the region are heterozygous). We also set the minimum length of the autozygous regions to be 4 Mb in length, with at least 50 probed SNPs (see Supplement S-D1 for further explanation). These values were chosen to provide adequate coverage of the genome (~75%) while also allowing for a low false discovery rate. When shorter regions are considered, they cannot be uniformly detected across the genome due to the SNP density of the chip used. This method is applied to both the CRC and control data sets. A filter is then applied to eliminate regions in the CRC patients' genomes, which are completely covered in the controls (i.e., the start of the control region is at or before the start of the patient's region and the end of the control region is at or after the end of the control region). This filter allows the isolation of regions that are unique or more frequent in the CRC patients compared with the controls. A second filter is applied to look for a given degree of overlap among the CRC regions (e.g., 2, 3, 4, etc., samples). This whole procedure will be discussed in detail in a companion article.[14]

Method 2: logarithm of the odds calculation. Another statistical method we used in identifying the IBD segments is an extension of the Broman and Weber approach (20), in which the autozygosity logarithm of the odds (LOD, base 10) score for a 5-Mb segment (ranging from SNP position *j* to SNP position *k*) in the genome was calculated. As defined in that article,

$$\mathrm{LOD}(j,k) = \sum_{i=j}^{k} \log R;$$

$$R = [P(g_i|\text{autozygous at } i)/P(g_i|\text{not autozygous at } i)] \quad (A)$$

P (g_i | autozygous at i) refers to the probability of the observed genotype g at the ith position in the genome, given that the ith position is autozygous, whereas P (g_i | not autozygous at i) refers to the probability of the observed genotype g at the ith position in the genome, given that the ith position is not autozygous. If the SNP call (genotype) at position i is AA or BB, then $R = (1 - \varepsilon) / P_A + \varepsilon$, or $(1 - \varepsilon) / P_B + \varepsilon$, respectively (20). On the other hand, if the genotype is AB, then $R = \varepsilon$, where ε denotes the combined rate of genotyping error and mutations (maximum of 2%). P_A and P_B are the frequencies of alleles A and B, respectively, in the study group (i.e., CRC or control population separately). Described in detail in a separate manuscript,[15] the algorithm used a sliding window method (5 Mb from one end of a chromosome to the other with 0.5 Mb step size) to form segments along the genome.

Verification of SNPs by Direct DNA Sequencing

Tumors containing IBD segments covering SNPs that are recently associated with colon cancer and Crohn's disease were subjected to dideoxy-sequencing to genotype the associated SNP and to verify homozygosity at that region. DNA sequencing was performed using the Applied Biosystems Automated 3730 DNA Analyzer, along with Big Dye Terminator chemistry and AmpliTaq-FS DNA Polymerase (Applied Biosystems). Universal primers (forward: 5′-CGTCACGACACGAAAAC-3′ and reverse: 5′-CGTCACGACACGAAACA-3′) were used for the sequencing and the following DNA-specific primers were used to amplify the DNA segment covering the SNP in question: rs9469220 (forward: 5′-CAGAGTCACTTGTCTCTGGCAGTCCAAGCTACTA-3′, reverse: 5′-AATAAGTCAGCCACTGCACCTGGA-3′), rs17234657 (forward: 5′-AGTGCTGAAGCGGAATTGAGCTCC-3′, reverse: 5′-AGGGACACAAGGGATTTGACTGTG-3′), rs11805303 (forward: 5′-AGTAGTGCCTTTCACCACCCATCA-3′, reverse: 5′-ACGTTGTTCCCAGGTGCTGTTATC-3′), rs10883365

[13] A. Olshen, et al. Analysis of genetic variation in Ashkenazi Jews by high density SNP genotyping, submitted for publication.

[14] G. Schemmann, et al., in preparation.

[15] S. Wang, et al. Genome-wide autozygosity mapping in human populations, submitted for publication.

Table 1. Partial list of the IBD segments in the CRC patients

Chromosome	Patient ID	IBD segments (all)		
		Homozygosity analysis (method 1; Mb-Mb)	Same region as identified by autozygosity analysis (method 2; Mb-Mb)	Length* (Mb)
1	C0111	14.837–41.8932	15.4744–41.9744	27.0562
1	C0111	51.3405–84.5851	51.9744–84.9744	33.2446
1	C0253	68.1948–76.2018	66.9744–76.4744	8.0069
1	C0111	142.406–157.5006	144.9744–159.9744	15.0945
2	02308	97.1053–118.0407	98.6008–119.1008	20.9355
2	C0221	191.3785–205.8568	191.1008–207.1008	14.4783
2	C0221	218.8516–228.8536	218.1008–229.6008	10.0021
5	C0181	94.2878–108.8576	93.2605–110.2605	14.5697
5	C0111	131.4386–148.2447	131.7605–149.2605	16.8061
5	C0181	148.1866–158.304	148.2605–158.7605	10.1174
6	C0111	9.6167–20.0816	8.6506–21.6506	10.4649
7	00485	5.1196–14.9583	3.6512–16.1512	9.8387
7	C0265	20.0919–29.2056	19.6512–30.1512	9.1137
7	C0221	70.9568–80.6667	73.1512–81.6512	9.7099
7	C0153	120.9134–136.5389	120.6512–138.1512	15.6255
8	C0111	2.7994–17.945	0.6806–19.1806	15.1456
8	C0161	74.6258–91.0269	73.6806–90.6806	16.4012
8	C0153	126.8113–136.872	125.1806–137.6806	10.0607
9	C0111	2.9818–23.1483	3.2394–23.2394	20.1666
9	C0153	121.2258–130.8439	121.2394–130.7394	9.6181
10	A7223	97.3554–106.2828	97.2672–106.2672	8.9275
12	C0111	0.0939–12.4867	0.0937–13.5937	12.3928
12	C0111	42.3234–59.9483	42.5937–60.0937	17.625
12	C0111	71.049–92.2312	71.5937–92.5937	21.1821
13	C0111	33.6001–42.5475	33.8211–43.3211	8.9475
13	02308	81.8705–105.3628	81.3211–105.3211	23.4923
16	C0161	12.2388–26.0158	12.7052–25.7052	13.777
17	02050	40.2645–57.9614	42.4512–57.9512	17.6968
18	C0153	45.2834–55.8778	44.1499–56.1499	10.5944
18	C0192	50.5348–68.1576	50.6499–68.6499	17.6228
18	C0111	58.4454–73.4044	58.1499–74.1499	14.959
20	C0111	29.31–54.4835	27.5957–53.5957	25.1735
20	C0329	32.0727–43.4094	32.0957–43.0957	11.3367
21	C0161	20.7118–31.7663	20.0748–32.0748	11.0544

NOTE: Shown are IBD segments of at least 8 Mb in length. The complete list is found on Supplement S-B7. The segments are identified through both homozygosity (method 1) and LOD calculation (method 2) analyses (described in Materials and Methods).

*The length of the IBD segment identified through homozygosity analysis.

†The average LOD of the segments determined by autozygosity analysis.

‡The number of SNPs covered by the IBD region identified through homozygosity analysis.

§Portion of IBD segment (homozygosity analysis) of at least 4 Mb in length and not overlapping any IBD segment found in the two control data sets (AMD and Framingham).

‖Portion of IBD segment (method 1) of at least 4 Mb in length and not overlapping any IBD segment found in only the AMD control data set.

¶The IBD segments found to be specific to CRC patients when compared with the AMD controls using LOD calculations (method 2).

(forward: 5′-TGCTGTTCCCTGGCTGATTCTGA-3′, reverse: 5′-ACGTTGTTCCCAGGTGCTGTTATC-3′), rs10505477 (forward: 5′-GTGGTGAACTTTGCAGTGGTCCAA-3′, reverse: 5′-GACTCCTTGTTCCTCCACTTCTGCCAAA-3′).

The PCR reaction (25 μL) contained 20 mmol/L Tricine (pH 8.7), 16 mmol/L $(NH_4)_2SO_4$, 2.5 mmol/L $MgCl_2$, 0.2 mmol/L deoxynucleotide triphosphate, 0.2 μmol/L of each gene-specific primer, 2.5 units of AmpliTaq Gold DNA polymerase, and 100 ng of genomic DNA. Thermocycling conditions were as follows: 95°C for 10 min to activate AmpliTaq Gold polymerase; followed by 25 cycles of 94°C for 30 s, 60°C for 1 min, 72°C for 1 min; followed by a final extension step at 72°C for 30 min.

Table 1. Partial list of the IBD segments in the CRC patients (Cont'd)

IBD segments (all)		IBD segments (CRC patient specific)		
Average [†] LOD	No. SNPs covered [‡]	No overlap with AMD and Framingham data sets[§] (Mb-Mb)	No overlap with AMD data set[‖] (Mb-Mb)	Cancer patient-specific IBD region (Mb-Mb)[¶]
8.9985	289	14.837–41.89323	14.837–41.89323	20.4744–25.4744
23.5478	957	51.34045–63.01534	51.34045–63.01534	57.9744–62.9744
13.9603	233			
17.2482	236	148.9734–157.5006	142.406–157.5006	
13.7592	337	97.10526–118.0407	97.10526–118.0407	100.1008–106.6008
18.5575	274	196.1016–205.8568	191.3785–205.8568	
20.8216	198	218.8516–228.8536	218.8516–228.8536	222.6008–229.6008
20.29	372	94.28781–97.90896	94.28781–97.90896	93.2605–98.2605; 104.2605–109.2605
14.2176	350	131.4386–148.2447	131.4386–148.2447	131.7605–137.2605
20.9462	288	148.1866–158.304	148.1866–158.304	150.7605–156.2605
20.7318	266	9.616674–20.08158	9.616674–20.08158	13.6506–18.6506
30.1294	396	6.951365–11.58131	6.951365–11.58131	10.1512–15.6512
20.3819	247	20.09189–29.20558	20.09189–29.20558	20.6512–30.1512
15.0587	167		72.97821–80.66672	
20.4	400	128.6356–136.5389	120.9134–136.5389	
25.9458	563	2.799396–17.945	2.799396–17.945	0.6806–19.1806
16.3239	410		74.62577–88.92999	74.1806–79.1806
22.6335	240	126.8113–136.872	126.8113–136.872	
30.2339	641	2.981779–18.52022	2.981779–23.14833	4.7394–12.7394
6.1826	74			
11.2501	121			
16.5157	231		0.093917–12.48667	0.0937–13.0937
10.3238	268	45.00302–55.27138	45.00302–55.27138	
17.2552	472	72.89781–80.71947	71.04904–92.23119	83.0937–92.0937
15.2037	271		33.60007–38.7675	33.8211–38.8211
24.4355	686	81.8705–93.5177	81.8705–93.5177	81.3211–99.3211
6.6682	158	19.1283–26.01577	19.1283–26.01577	
17.7118	286	46.1909–52.27807	46.1909–52.27807	48.9512–56.9512
22.7059	274			47.1499–56.1499
24.3056	511	55.57051–59.7204	55.57051–59.7204	50.6499–67.1499
21.3315	453			59.1499–67.1499
11.2064	359	39.59348–54.48347	39.59348–54.48347	
13.4314	167	39.59348–43.4094	39.59348–43.4094	
25.2767	398	20.71183–31.76627	20.71183–31.76627	20.0748–32.0748

Results

IBD segments as extended runs of homozygous SNPs. The first approach to score regions of autozygosity (method 1) was to isolate the long stretches of homozygosity in the genomic DNAs taken from mostly noncancerous tissues (explained in Materials and Methods). We set the threshold limit to a minimum of 4-Mb length encompassing at least 50 consecutive homozygous SNPs, but allowing at most 2% heterozygous SNPs. These identified IBD regions (on chromosomes 1 to 22) are indicated as bars in Fig. 2 (*A*, CRC patients; *B*, AMD controls; *C*, Framingham controls). A partial list (at least 8 Mb in length) of these IBD segments are shown in Table 1. The longest IBD segment is the chromosome 1 region from 51.3405 to 84.5851 Mb found in patient C0111, who has a total of 271.6 Mb of homozygous segments distributed in 19 IBD segments (see Supplements S-B3, S-B4, and S-B5) for the total IBD segment lengths of every CRC patient, as well as all the AMD, and Framingham control individuals. As shown in Table 1, the same IBD regions in chromosome 1 of patient C0111 is also identified by the autozygosity (method 2) analysis, having an average LOD score of 23.5. Within this segment is a region (51.34045–63.01534 Mb) not overlapping with any of the IBD regions in both control data sets. In all, the homozygosity analysis identified a total of 117 IBD segments of at least 4 Mb in length (Supplement S-B7). In another approach (Method 2), a LOD score was calculated to compare the strength of autozygosity versus nonautozygosity for a defined genomic region. This calculation was able to identify the 5 Mb regions (many regions were overlapping, and further inspection identified these autozygous regions as contiguous) in the CRC patients' genome with LOD values of at least 5 (see Supplement S-B8). Of the 34 IBD segments identified by homozygosity analysis

Table 2. Summary of the results of IBD segment analysis using the homozygosity mapping and autozygosity (LOD) approach

Data set	Homozygosity analysis (method 1)				LOD approach (Method 2)	
	Average IBD ≥ threshold (Mb)	Average IBD without max IBD sample	% Samples with IBD ≥4 Mb (threshold)	% Samples with IBD ≥8 Mb	Average LOD*	% Samples with LOD ≥5 †
All CRC patients	12.7	9.2	62.2	36.4	2.76	87.8
AMD control	5.3	3.7	35.6	12.3	0.67	16.4
Framingham control	5.5	4.5	28.8	13.6	1.64	56.8

*Average of positive LOD.
†Samples that have at least one segment with LOD ≥ 5 are eligible.

(method 1) to be at least 8 Mb in length, all 34 (100%) segments were also identified by autozygosity analysis (method 2; Table 1). For the 56 segments of at least 6 Mb in length identified by method 1, 53 (95%) were also identified by method 2. The concordance was down to 80% (94 of 117) if all the method 1–identified IBD regions are considered. Copy number/LOH analysis (Supplements S-B1 and S-B11) showed that IBD segments can easily be distinguished from actual LOH and uniparental disomy (UPD) regions, with the latter two occurring frequently in tumor samples but not in the matching normals. Only 17 of the 117 identified IBD segments were from the 22 CRC samples and all of these were shorter than 6 Mb. In Supplements S-B2 and S-C2, we present a clear explanation to justify the use of the data from these 22 CRC samples.

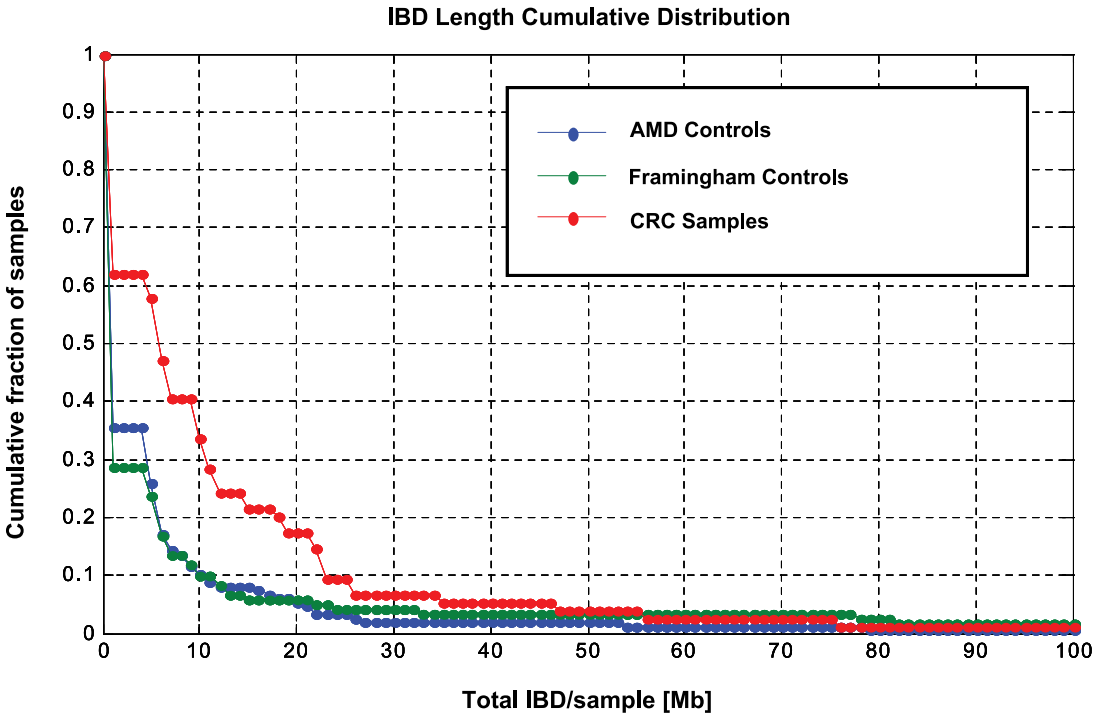
Higher percentage of IBD regions in CRC patients compared with the control data sets. Using the homozygosity approach, we identified 46 of 74 CRC patients (62%) to have at least one IBD segment satisfying the set threshold (4 Mb). In contrast, 34 of 118 (29%) and 52 of 146 (36%) of the Framingham and AMD control individuals, respectively, have detectable IBD segments (Table 2). When the analysis was performed using the threshold limit for IBD segments, the CRC patients showed average IBD lengths of 12.7 Mb, whereas AMD and Framingham data sets showed average IBD segment lengths of 5.3 and 5.5 Mb, respectively. When we removed the patient with the longest IBD segment from each data set, the average total IBD segment length was reduced to 9.2, 4.5, and 3.7 Mb for CRC patients, for AMD, and for Framingham data sets, respectively. This finding is also shown in Fig. 3*A* which shows the cumulative distribution of the total IBD segment lengths for the CRC patients and the two control data sets.

The graph is presented in such a way that each data point represents the cumulative fraction (y axis) of the samples with the corresponding minimum cumulative IBD segment length (x axis). In other words, $Y = f(X \geq x)$. For example, the graph tells us that ~35% of the CRC patients have total IBD length of at least 10 Mb, whereas it is only 10% for both controls. The clear difference between the CRC patients and the control data sets can be seen even up to a cumulative frequency of 20 Mb IBD segment/sample. The Kolmogorov-Smirnov test (31) showed significant difference between the CRC and AMD ($P = 1.28 \times 10^{-5}$) and CRC and Framingham ($P = 1.13 \times 10^{-5}$) distributions. On the other hand, there was no significant difference between the distributions of AMD and Framingham data sets ($P = 0.91$). The use of LOD calculations (method 2) also identified most of the IBD segments detected by the homozygosity (method 1) analysis. On average, CRC patients have LOD of 2.76, which is significantly higher than either the AMD (0.67) or Framingham (1.64) controls (Table 2). Eighty-eight percent of the CRC patients had LOD score of at least 5, whereas it is 16% and 57% for the AMD and Framingham controls, respectively.

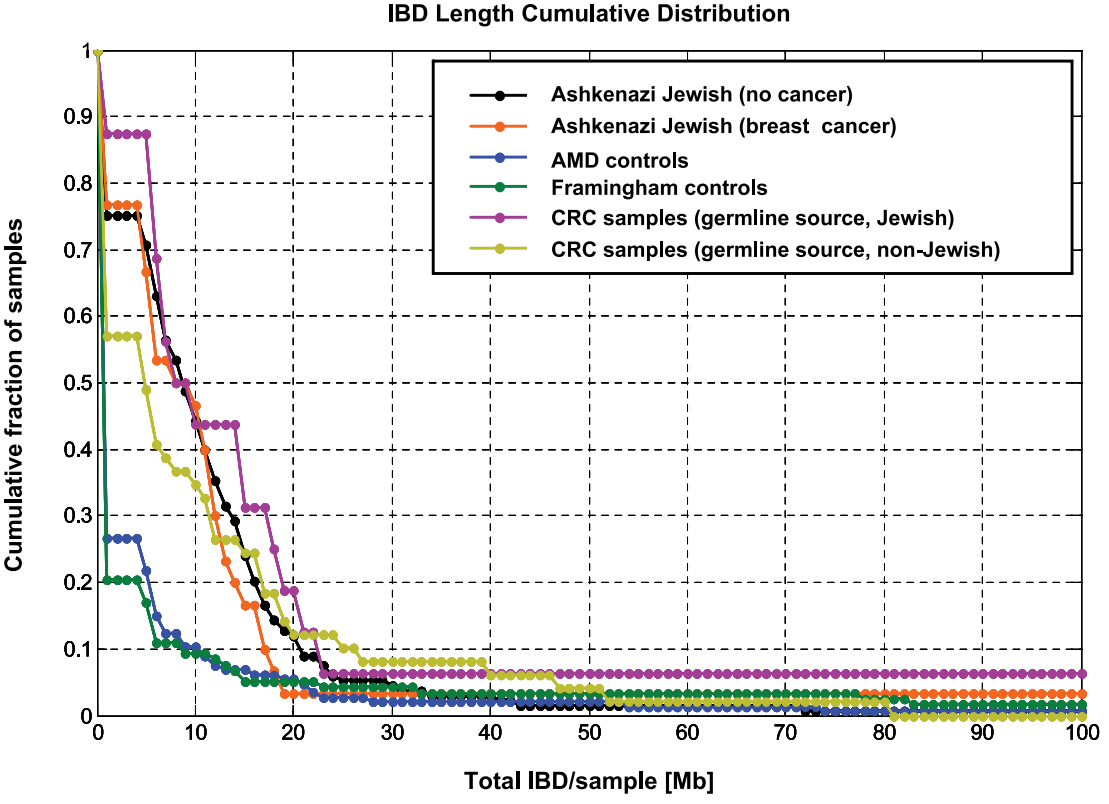
CRC patients of Jewish ancestry have higher percentage of IBD regions compared with the rest of the cohort, and the control groups. If the CRC patients are divided into Jewish and non-Jewish groups, 94% of the former and 35% of latter have IBD regions. There is also a disparity in IBD segment size—8.3 Mb for the Jewish and 5.1 Mb among non-Jewish patients (calculated based on information and data listed in Supplements S-A2 and S-B3). Statistical comparison (Kolmogorov-Smirnov analysis) also showed a clear difference between the CRC Jewish and

Figure 3. *A,* the cumulative distributions of the lengths of IBD segments for the CRC patients, as well as AMD and Framingham control individuals. The graph is presented in such a way that each data point represents the cumulative fraction (y axis) of the samples with the corresponding minimum cumulative IBD segment length (x axis). In other words, $Y = f(X \geq x)$. The clear difference between the CRC patients and the control data sets can be seen even up to a cumulative frequency of 20 Mb IBD segment/sample. The Kolmogorov-Smirnov test showed significant differences between the CRC and AMD ($P = 1.28 \times 10^{-5}$), as well as between CRC and Framingham ($P = 1.13 \times 10^{-5}$) distributions. On the other hand, there was no significant difference between the distributions of AMD and Framingham data sets ($P = 0.91$). *B,* the cumulative distributions of the lengths of IBD segments for Jewish and non-Jewish subgroups of the CRC patients, the AMD and Framingham controls, along with AJBC and AJNC patients. Statistical comparison (Kolmogorov-Smirnov test) also showed a clear difference between the CRC Jewish and non-Jewish distributions ($P = 0.0170$). Nonetheless, both the percentages of samples with IBD segments and the average IBD segment size are significantly higher for non-Jewish patients compared with either the AMD ($P = 4.30 \times 10^{-4}$) or Framingham controls ($P = 1.08 \times 10^{-4}$; *B*). We then compared the IBD segment distributions in the Ashkenazi Jewish (AJBC and AJNC) data sets with those of our CRC and control (AMD and Framingham) data sets. The IBD segment distributions of AJBC and AJNC are indistinguishable from each other ($P = 0.922$). However, it is very clear that the fraction of samples with at least 5 Mb total IBD length is higher in both Ashkenazi Jewish data sets than in the CRC non-Jewish, as well as AMD and Framingham data sets. Statistical comparisons show that AJBC versus AMD, AJNC versus AMD, AJBC versus Framingham, and AJNC versus Framingham have P values of 1.31×10^{-6}, 9.48×10^{-17}, 2.09×10^{-7}, and 2.54×10^{-17}, respectively. The data from AJBC and AJNC groups were generated using the more dense Affymetrix 500K SNP array. Before the comparing the IBD segments identified from the 500K and the 50K Xba array data, we identified the SNPs whose genomic positions are closely matched in the two sets (maximum separation of 10,000 bp, although 9,360 SNPs are identical, in the two array sets; see Supplement S-D2). Thus, the IBD regions identified and plotted for *B* were from the analyses of 39,097 SNPs.

A


IBD Length Cumulative Distribution
Cumulative fraction of samples
Total IBD/sample [Mb]
AMD Controls
Framingham Controls
CRC Samples
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80
90
100

B


IBD Length Cumulative Distribution
Cumulative fraction of samples
Total IBD/sample [Mb]
Ashkenazi Jewish (no cancer)
Ashkenazi Jewish (breast cancer)
AMD controls
Framingham controls
CRC samples (germline source, Jewish)
CRC samples (germline source, non-Jewish)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80
90
100

non-Jewish distributions (P = 0.0170). Nonetheless, both the percentages of samples with IBD segments and the average IBD segment size are significantly higher for non-Jewish patients compared with either the AMD ($P = 4.30 \times 10^{-4}$) or Framingham controls ($P = 1.08 \times 10^{-4}$; Fig. 3*B*). This observation also led us to examine additional data sets generated specifically for a genome-wide association study at MSKCC: 30 AJBC, along with 133 AJNC. We then compared the IBD segment distributions in the Ashkenazi Jewish (AJBC and AJNC) data sets with those of our CRC and control (AMD and Framingham) data sets. The IBD segment distributions of AJBC and AJNC are virtually indistinguishable from each other (P = 0.922). It is very clear that the fraction of samples with at least 5 Mb total IBD length is higher in both Ashkenazi Jewish data sets than in the CRC non-Jewish, as well as AMD and Framingham data sets. Statistical comparisons show that AJBC versus AMD, AJNC versus AMD, AJBC versus Framingham, AJNC versus Framingham have P values of 1.31×10^{-6}, 9.48×10^{-17}, 2.09×10^{-7}, and 2.54×10^{-17}, respectively.

Autozygosity increases CRC risk. From the data plotted in Fig. 3*A*, it is possible to calculate the extent to which autozygosity adds to CRC risk by using Bayes' rule, a formula of conditional probabilities: $P(B|A) = P(A|B) \times P(B)/P(A)$. If we assume that A refers to IBD $\geq x$, where x is the IBD length, and B refers to CRC incidence, then:

$$P(\text{CRC}|\text{IBD} \geq x) = P(\text{IBD} \geq x|\text{CRC}) \times P(\text{CRC})/P(\text{IBD} \geq x) \qquad (B)$$

From Fig. 3*A*, we can see that P(IBD $\geq$ 10 Mb|CRC) = 0.3 and P(IBD $\geq$ 20 Mb|CRC) = 0.18. Furthermore, the data from control data sets (which represents 95% of the population) suggest that P(IBD $\geq$ 10 Mb) = 0.1; P(IBD $\geq$ 20 Mb) = 0.05. Therefore

$$P(\text{CRC}|\text{IBD} \geq 10\text{Mb}) = 3 \times P(\text{CRC}); P(\text{CRC}|\text{IBD} \geq 20\ \text{Mb}) = 3.6 \times P(\text{CRC}) \qquad (C)$$

Equation B shows that having total IBD of at least 10 Mb increases CRC risk 3-fold, whereas having a total IBD of at least 20 Mb increases the risk almost 4-fold.

Discussion and Conclusion

The most plausible explanation for the presence of long stretches of homozygous regions in an individual's genome is that his or her parents can trace their lineage to a common ancestor. UPD (an instance when an offspring inherits both copies or segments of chromosomes from a single parent), although possible, is highly unlikely. In cancer tissues, the appearance of a UPD can be manifested in events of gene conversions when a copy or segments of a chromosome are lost and the remaining copy gets duplicated (32, 33). In their analysis of STRPs in the genomes of individuals from CEPH reference families, Broman and Weber (20) discovered that long homozygous segments are quite common and that these may be attributed to autozygosity. In one particular family, all the progeny showed 4 to 12 autozygous segments with an average length of 19 cM per segment. The fact that both parents did not show any significant homozygosity suggests that the parents can trace their ancestry to a common individual. Using the publicly available SNP genotype data for 209 individuals from the International Hapmap Project (34), Gibson and coworkers identified 1,393 homozygous segments (with at least 1-Mb length and minimum SNP density of 1 SNP per 5 kb; ref. 35). The longest identified homozygous segment (17.91 Mb) is that of a Japanese individual whom the authors consider to be a progeny of related parents. Yorubas from Ibadan, Nigeria, have the fewest long tracts of homozygosity when compared with Han Chinese from Beijing, Japanese from Tokyo, and CEPH Utah individuals of Northern and Western Europe ancestry. This observation is consistent with the belief that the African race has been established earlier (thus higher incidence of recombination subdividing the haplotype regions) than the Asiatic and Caucasian races. Another important conclusion from their study is that these homozygous segments are more prevalent in regions of high linkage disequilibrium (and thus, of low recombination). Based on the analysis of Li and coworkers, the genomes of 34 of 515 (6.6%) unrelated Han Chinese individuals also contained these homozygous segments (which they referred to as long contiguous stretches of homozygosity). The segment size ranged from 2.94 to 26.27 Mbp (36). Using the publicly available Affymetrix data sets, they also found out that 26.2% of Caucasians and 4.76% of African Americans also have these IBD segments in their genomes. When they analyzed the genomes of siblings of a consanguineous marriage, they found out that the genomes of all the siblings exhibited multiple long contiguous stretches of homozygosity ranging from 3.06 to 52.17 Mb. This served as clear proof that genomic IBD regions result from inbreeding. Most recently, the International Hapmap Project (phase 2; ref. 37) was able to identify these extended runs of homozygosity among 51 of 270 individuals (19%). Although they used more dense SNP arrays, and set different specifications (minimum of 3 Mb), the percentages of individuals with long homozygous segments were comparable with what we found in the AMD and Framingham controls. The authors also contended that these were most probably due to recent co-ancestry in the individuals' parents. We then examined the possibility that the IBD segments among our CRC subjects may actually be haploblocks or groups of alleles (or SNPs) that are usually in linkage disequilibrium. However, of the 117 IBD segments identified by the homozygosity analysis, only 11 (9%) have at least 30% overlap (see Supplement S-B9) with the long haplotype regions identified by the International Hapmap project (phase I; ref. 34).

There are clear correlations between the incidence of cancer and degrees of inbreedings on a number of population-based studies (21, 22). The results of our own study clearly show the difference in degrees (both the percentage and lengths) of autozygous segments between the MSKCC CRC patients and the control data sets. However, it is important to note that of all the 74 CRC patients in our study, 16 (22%) indicated Judaism as their religious affiliation. This is greatly due to the location of MSKCC (New York City). According to a 2002 survey, there are 1.4 million individuals of Jewish ancestry (constituting 15% of all the households) living in the five New York City boroughs plus three surrounding counties (38). Unfortunately, we do not have any information on the religous affiliations of the subjects making up the AMD and Framingham controls data sets. It is very likely that the incidence of autozygosity among people of Jewish ancestry are more prevalent compared with the average Caucasian population. Historically, Jewish communities have maintained high degree of endogamy (marrying within its own group) for cultural and religious reasons, thus

Table 3. The recently identified predisposition SNPs for colon cancer and Crohn's disease whose locations are covered by the IBD regions in some of the CRC patients in the study

Identified IBD region			SNP	Associated disease	Reference	Actual SNP genotype	Allele associated to added risk (Y/N)
CRC patient	Chromosomal region	Position (Mb)					
C0153	8q24	126.8113–136.872	rs10505477	Colon cancer	(46)	TT	Y
C0153	8q24	126.8113–136.872	rs6983267	Colon cancer (and prostate cancer)	(47, 48)	GG	Y
C0111	9p24	2.9818–23.1483	rs719725	Colon cancer	(46)	CC	N
C0253	1p31	68.1948–76.2018	rs11805303	Crohn's disease (strong association)	(49)	CC	N
10216	5p13	41.3047–45.7769	rs17234657	Crohn's disease (strong association)	(49)	TT	N
C0111	5q23	131.4386–148.2447	rs6596075	Crohn's disease (moderate association)	(49)	CC	Y
00485	5q23	132.3024–136.3673	rs6596075	Crohn's disease (moderate association)	(49)	CC	Y
C0181	5q33	148.1866–158.304	rs1000113	Crohn's disease (strong association)	(49)	CC	N
C0170	6p21	30.0114–34.5504	rs9469220	Crohn's disease (moderate association)	(49)	GG	N
C0159	6p21	31.6969–37.0289	rs9469220	Crohn's disease (moderate association)	(49)	GG	N
C0111	6p22	9.6167–20.0816	rs6908425	Crohn's disease (moderate association)	(49)	CC	Y
A7223	10q24	97.3554–106.2828	rs10883365	Crohn's disease (strong association)	(49)	AA	N
07061	10q24	100.2231–104.8909	rs10883365	Crohn's disease (strong association)	(49)	GG	Y

NOTE: The actual genotypes of the SNPs were verified by direct DNA sequencing.

increasing the chances of autozygotic signatures in their genomes. The patient C0111 who has the most IBD segments of all the MSKCC patients is of Jewish descent. We can only speculate whether the incidence of autozygosity is a contributing factor to the fact that Ashkenazi Jews have the highest incidence of colon cancer of any ethnic group in the world (39). Aside from dietary factors, genetics can also play a major role. The APC variant I1307K, almost unique to Ashkenazi Jews, has been identified as a CRC susceptibility factor among this group (40). The results of our principal components analysis (EIGENSTRAT method) did not identify population stratification between those CRC patients of Jewish ancestry and the rest of the CRC cohort (Supplement S-C4). Whereas there is no clear genetic variation between the CRC patients and Framingham control group, the opposite is true when comparing the CRC patients and the AMD control group. Although all of the individuals in the AMD data set are Caucasians, as is the majority of our CRC patients, the results of principal components analysis suggest that the Framingham data set is the more appropriate control group. The observed difference in IBD incidence between the CRC patients and the AMD control group may then be partly due to population stratification. It also appears that there is practically no difference between the AJBC and AJNC group in terms of the incidence of autozygosity (both of which have more IBD segments compared with either AMD or Framingham data sets). It should be noted that the AJBC individuals were chosen to be part of a genome-wide association study because of their family history of breast cancer. The increased predisposition to the disease for this group may have been brought by a dominant genetic factor and that longer IBD segments may not have played a major role in breast cancer predisposition. On the other hand, when comparing CRC Jewish and non-Jewish patients, we were essentially comparing two groups in which family history of CRC was much less common (Supplement S-A2). Among non-Jewish CRC patients, 7 of 49 (14.3%) had at least one first-degree relative, and 4 of 49 (8.2%) had at least one second-degree relative who also suffered from CRC. For the Jewish CRC patients, these numbers are 2 of 16 (12.6%) and 0 of 16 (0%), respectively. Nonetheless, these observations are not necessarily contradictory to our hypothesis that the cumulative effects of autozygosity may contribute to the incidence of spontaneous CRCs.

Is there a simple model to explain how autozygosity increases CRC risk at the molecular level? One approach requires us to distinguish between the high- and low-penetrance classes of cancer-predisposing genes. The former includes the dominantly inherited mutations in *APC, MLH1*, and *MSH2*. Such mutations only need to be heterozygous to contribute to cancer predisposition (reviewed in ref. 6) and have been identified with much help from classic genetic analyses. On the other hand, finding low-penetrance cancer-predisposing mutations often requires genetic association studies (4). A short list of genes identified to have variants belonging to the latter category includes *APC* (I1307K variant), *TGFBR1* (6 Ala variant), *HRAS1* (variable number of tandem repeats variant), and *MTHFR* (677V variant). *TGFBR1* (6 Ala variant), which is classified as a tumor suppressor, is found to be dose dependent, meaning the allele is more effective in predisposing cancer in homozygous than in heterozygous state (41). The base excision repair gene *MYH*, which has been linked to an FAP-like syndrome (42), can also have variants that can be transmitted in a dose-dependent manner, albeit differently. In the case of this gene, two mutations (Y165C and G382D) have been identified to be highly penetrant when in biallelic state (either homozygous or compound heterozygous; reviewed in ref. 43). However, it has also been shown that monoallelic mutations of *MYH* can also predispose for CRC at lower penetrance (44). Likewise, if dose-dependent, low-penetrance genes are located in IBD regions, the influence on cancer initiation or progression would be doubled. Longer IBD segments would have a higher probability of containing such alleles in homozygous

state. Moreover, longer IBD segments may also cover multiple low-penetrance, dose-dependent genes that have additive effects, which is now believed to occur in both sporadic and familial types of CRC (4). We can only presume that functionally, such low-penetrance, dose-dependent alleles do not necessarily have to be associated with tumor suppressors. For instance, it is possible that a mutation in the regulatory region of a proto-oncogene may result in protein overexpression or the dysregulation during stress.

This is an exploratory study on the possibility that autozygosity increases the risk of cancer, and there are limitations in our study. First, there is the lack of information on the cancer status of the AMD control subjects. The average age of AMD subjects is 80 years. According to the statistics provided by National Cancer Institute Surveillance Epidemiology and End Results, which is accessible through the Web site,[16] the incidence of CRC in the United States (for all races between 1975 and 2003) is 0.322%, 0.377%, and 0.416% for age groups 75 to 79, 80 to 84, and 85+, respectively (45). Thus, there is only a small chance that an AMD study participant has also been afflicted with CRC. All of the subjects in the other control data set (Framingham) did not have any cancer at the time of their participation according to clinical records. Second, our cohort was enriched for patients of Jewish ancestry, and the work would have benefited from availability of another sizable set of control individuals of Jewish ancestry who had not been diagnosed of CRC at a late age (75 years or older). Third, we chose to evaluate copy number and alleles only among matched normal and tumor samples, where the tumor samples had <30% stromal infiltrates. Such samples may have had more homogenous tumors, which, in turn, may have had a higher incidence of underlying genetic factors.

None of the widely recognized CRC predisposing genes (*APC, MLH1, MSH2*) fall within our identified IBD regions. Most recently, several laboratories have performed large-scale, genome-wide association studies and identified several loci associated with increased risk to colon cancer (46–48). All of the three newly identified colon cancer–associated SNPs of highest risk (46, 47) are within the IBD regions of two of our CRC patients: rs10505477 (8q24) in C0153; rs6983267 (8q24) in C0153; and rs719725 (9p24) in C0111 (Table 3). The SNP rs6983267 has also been identified to be a common risk factor for CRC and prostate cancer (48). Direct DNA sequencing of C0153 DNA revealed that the two 8q24 SNPs (rs10505477 and rs6983267) are indeed homozygous for the CRC-predisposing alleles. However, the 9p24 SNP was found to be homozygous for the non-CRC predisposition SNP. According to our clinical records, both patients C0111 and C0153 did not have any family history of CRC. We also examined the genotypes of IBD region SNPs that have been associated to Crohn's disease (49), a possible precursor of colon cancer (50). The Crohn's disease–associated SNPs were interrogated despite the fact that none of our CRC patients had any clinical documentation for the disease (Supplement S-A2). Of the seven Crohn's disease–associated SNPs located within the identified IBD regions among the CRC patients, three (among four patients) were found to be homozygous for the Crohn's disease–predisposing SNPs: rs6596075 (within IBD regions of C0111H and 00485K), rs6908425 (within IBD region of C0111H), and rs10883365 (within IBD region of 07061). However, SNPs rs11805303 (C0253K), rs17234657 (10216H), rs1000113 (C0181H), rs9469220 (C0170H and C0159H), and rs10883365 (A7223H) were all genotyped to be homozygous for the non–disease-associated alleles.

We have shown that there is a higher frequency and a longer length of IBD segments within our CRC patients compared with a number of control groups. Whether these IBD segments result in cancer or lead to the progression of cancer has yet to be determined. There is clearly a need to expand this study to include the sampling of a wider cohort and just as importantly to examine the identified IBD regions for potential cancer-causing genes.

[16] http://seer.cancer.gov

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Received 9/6/2007; revised 1/15/2008; accepted 1/30/2008.

Grant support: National Cancer Institute grant P01-CA65930, the Gilbert Family Foundation, and generous funding from the Ludwig Institute for Cancer Research/Conrad N. Hilton Foundation joint Hilton-Ludwig Cancer Metastasis Initiative.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Dr. Jenny Xiang and her team at the Microarray Core Facilities, Department of Microbiology, Cornell University Weill Medical College, New York, NY, for the help and services and Shoshana Rosenberg of MSKCC for all the valuable assistance.

References

1. Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. CA Cancer J Clin 2007;57:43–66.
2. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55:74–108.
3. Ahmed FE. Colon cancer: prevalence, screening, gene expression and mutation, and risk factors and assessment. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 2003;21:65–131.
4. de la Chapelle A. Genetic predisposition to colorectal cancer. Nat Rev Cancer 2004;4:769–80.
5. Segditsas S, Tomlinson I. Colorectal cancer and genetic alterations in the Wnt pathway. Oncogene 2006;25:7531–7.
6. Nagy R, Sweet K, Eng C. Highly penetrant hereditary cancer syndromes. Oncogene 2004;23:6445–70.
7. Howe JR, Roth S, Ringold JC, et al. Mutations in the SMAD4/DPC4 gene in juvenile polyposis. Science 1998;280:1086–8.
8. Hemminki A, Markie D, Tomlinson I, et al. A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. Nature 1998;391:184–7.
9. Liaw D, Marsh DJ, Li J, et al. Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. Nat Genet 1997;16:64–7.
10. Naccarati A, Pardini B, Hemminki K, Vodicka P. Sporadic colorectal cancer and individual susceptibility: a review of the association studies investigating the role of DNA repair genetic polymorphisms. Mutat Res 2007;635:118–45.
11. Dong C, Hemminki K. Modification of cancer risks in offspring by sibling and parental cancers from 2,112,616 nuclear families. Int J Cancer 2001;92:144–50.
12. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. J Natl Cancer Inst 1994;86:1600–8.
13. Risch N. The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. Cancer Epidemiol Biomarkers Prev 2001;10:733–41.
14. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78–85.
15. Tsafrir D, Bacolod M, Selvanayagam Z, et al. Relationship of gene expression and chromosomal abnormalities in colorectal cancer. Cancer Res 2006;66:2129–37.
16. Wen Y, Giardina SF, Hamming D, et al. GROα is highly expressed in adenocarcinoma of the colon and down-regulates fibulin-1. Clin Cancer Res 2006;12:5951–9.
17. Cheng YW, Shawber C, Notterman D, Paty P, Barany F. Multiplexed profiling of candidate genes for CpG island methylation status using a flexible PCR/LDR/Universal Array assay. Genome Res 2006;16:282–9.
18. Pincas H, Pingle MR, Huang J, et al. High sensitivity EndoV mutation scanning through real-time ligase proofreading. Nucleic Acids Res 2004;32:e148.
19. Favis R, Huang J, Gerry NP, et al. Harmonized microarray/mutation scanning analysis of TP53

mutations in undissected colorectal tumors. Hum Mutat 2004;24:63–75.
20. Broman KW, Weber JL. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. Am J Hum Genet 1999;65:1493–500.
21. Rudan I, Rudan D, Campbell H, et al. Inbreeding and risk of late onset complex disease. J Med Genet 2003;40:925–32.
22. Shami SA, Qaisar R, Bittles AH. Consanguinity and adult morbidity in Pakistan. Lancet 1991;338:954.
23. Lebel RR, Gallagher WB. Wisconsin consanguinity studies. II: Familial adenocarcinomatosis. Am J Med Genet 1989;33:1–6.
24. Sutter NB, Ostrander EA. Dog star rising: the canine genetic system. Nat Rev Genet 2004;5:900–10.
25. Rousset F. Inbreeding and relatedness coefficients: what do they measure? Heredity 2002;88:371–80.
26. Huang J, Wei W, Zhang J, et al. Whole genome DNA copy number changes identified by high density oligonucleotide arrays. Hum Genomics 2004;1:287–99.
27. Di X, Matsuzaki H, Webster TA, et al. Dynamic model based algorithms for screening and genotyping over 100 K SNPs on oligonucleotide microarrays. Bioinformatics 2005;21:1958–63.
28. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. Science 2005;308:385–9.
29. Kannel WB. The Framingham Study: ITS 50-year legacy and future promise. J Atheroscler Thromb 2000;6:60–6.
30. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–9.
31. Riffenburgh RH. Statistics in Medicine. San Diego (CA): Academic Press; 1999. p. 581.
32. Andersen CL, Wiuf C, Kruhoffer M, Korsgaard M, Laurberg S, Orntoft TF. Frequent occurrence of uniparental disomy in colorectal cancer. Carcinogenesis 2007;28:38–48.
33. Teh MT, Blaydon D, Chaplin T, et al. Genomewide single nucleotide polymorphism microarray mapping in basal cell carcinomas unveils uniparental disomy as a key somatic event. Cancer Res 2005;65:8597–603.
34. International_Hapmap_Consortium. A haplotype map of the human genome. Nature 2005;437:1299–320.
35. Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. Hum Mol Genet 2006;15:789–95.
36. Li LH, Ho SF, Chen CH, et al. Long contiguous stretches of homozygosity in the human genome. Hum Mutat 2006;27:1115–21.
37. International_Hapmap_Consortium. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449:851–61.
38. Ukeles J, Miller R. Jewish Community Study of New York: 2002; 2004.
39. Feldman GE. Do Ashkenazi Jews have a higher than expected cancer burden? Implications for cancer control prioritization efforts. Isr Med Assoc J 2001;3:341–6.
40. Laken SJ, Petersen GM, Gruber SB, et al. Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. Nat Genet 1997;17:79–83.
41. Kaklamani VG, Hou N, Bian Y, et al. TGFBR1*6A and cancer risk: a meta-analysis of seven case-control studies. J Clin Oncol 2003;21:3236–43.
42. Al-Tassan N, Chmiel NH, Maynard J, et al. Inherited variants of MYH associated with somatic G:C->T:A mutations in colorectal tumors. Nat Genet 2002;30:227–32.
43. Lipton L, Tomlinson I. The genetics of FAP and FAP-like syndromes. Fam Cancer 2006;5:221–6.
44. Peterlongo P, Mitra N, Sanchez de Abajo A, et al. Increased frequency of disease-causing MYH mutations in colon cancer families. Carcinogenesis 2006;27:2243–9.
45. Ries L, Harkins D, Krapcho M, et al. SEER cancer statistics review, 1975-2003; 2006.
46. Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet 2007;39:989–94.
47. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 2007;39:984–8.
48. Haiman CA, Le Marchand L, Yamamato J, et al. A common genetic risk factor for colorectal and prostate cancer. Nat Genet 2007;39:954–6.
49. Wellcome_Trust_Case_Control_Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–78.
50. Bernstein CN, Blanchard JF, Kliewer E, Wajda A. Cancer risk in patients with inflammatory bowel disease: a population-based study. Cancer 2001;91:854–62.