Colorectal cancer risk is not associated with increased levels of homozygosity in Saudi Arabia

Abdul K. Siraj, PhD[1], Hanif G. Khalak, PhD[2], Mehar Sultana, MSc[1], Maha Al-Rasheed, BSc[1], Prashant Bavi, MD[1], Nasser Al-Sanea, MD[3], Fouad Al-Dayel, MD[4], Shahab Uddin, PhD[1], Fowzan S. Alkuraya, MD, FACMG[5], Khawla S. Al-Kuraya, MD, FCAP[1]

Purpose: Runs of homozygosity (ROHs) represent a measure of the extent of autozygosity and are correlated with the extent of inbreeding. Recently, it has been suggested that ROHs may contribute to the risk of colorectal cancer (CRC). The high rate of consanguinity and CRC in the Saudi population prompted us to test the role of autozygosity in the CRC risk.

Methods: We compared 48 Saudi CRC patients to 100 ethnically matched controls, processed on the Affymetrix 250K *Sty*I SNP GeneChip platform and analyzed using the plink package.

Results: We could find no evidence of a significant relationship between autozygosity and CRC risk.

Conclusion: The negative results in our study add additional significance to what has been previously reported in literature, as this is the first study to address these questions in an inbred population. Our subgroup analysis of patients with microsatellite unstable–positive tumors as compared with other groups did not significantly change our results. Although these results do not rule out the presence of recessively acting CRC-predisposing genes in a small percentage of patients, which our relatively small sample size could not capture, they suggest that such genes are unlikely to account for the disturbingly high incidence of CRC in our consanguineous population.

Genet Med 2012:14(8):720–728

Key words: colorectal cancer; homozygosity; risk

INTRODUCTION

Colorectal cancer (CRC) is a major cause of mortality and morbidity worldwide. In Saudi Arabia, the incidence of CRC is increasing. According to the latest statistics, CRC is the second most common cancer among Saudi males and the third most common among Saudi females.[1,2]

Genetic aspects of CRC range from Mendelian forms as in familial CRC syndromes (hereditary nonpolyposis colorectal cancer and familial adenomatous polyposis) at one end of the spectrum to sporadic occurrence, which is believed to be the result of interaction of genetic and environmental factors. Although the genetic factors underlying familial CRC syndromes have been delineated, little is known about the genetic risk determinants of CRC in the general population. Current clues that suggest the involvement of recessively acting genes are based on the data associated with consanguinity and from populations that are characterized by a high degree of inbreeding[3–8] as well as from studies in animals.[9] The Saudi population is known for having relative genetic homogeneity due to particular demographic, historic, and tribal characteristics and is known for high consanguinity.[10] The effect of inbreeding on cancer is likely more complex than simple Mendelian genetics, with many genetic components involved. Nonetheless, studying these genetically isolated populations may eventually lead to discovery of other genes that contribute to cancer predisposition.

Autozygosity is a term that is used to denote the presence of two identical haplotypes that are derived from an ancestor shared by both parents. It essentially represents a special type of homozygosity. Runs of homozygosity (ROHs) in the genome can be used to detect autozygosity and are directly correlated with the extent of inbreeding. Although the role of ROHs in unmasking recessively acting mutations is well established in Mendelian genetics, much less is known about their contribution to more complex disorders such as cancer. Assessment of ROHs on a genome-wide basis, therefore, provides a measure of extent of autozygosity and ultimately exposing recessively acting disease genes.[11] Previously, a significant increase in the frequency of homozygosity in combined cases as compared with controls was reported in patients with breast, prostate, or head and neck cancer of Northern/Western European ancestry by whole-genome loss of heterozygosity analysis using microsatellite markers.[12]

In addition, Bacolod et al. demonstrated, using Affymetrix SNP arrays (Santa Clara, CA), that cases with CRC harbored significantly more homozygous regions than did healthy individuals.[13] However, this observation could not be replicated.[14–16] Findings from these studies support the hypothesis that there exist multiple, recessive, cancer-predisposing loci that are not readily detected using a conventional genome-wide association approach based on analysis of individual single-nucleotide polymorphisms (SNPs). Although the efforts of genome-wide association studies can help identify common variants, there are likely to be rare variants that may be uncovered through whole-genome homozygosity analysis.

To examine whether homozygosity is associated with an increased risk of developing CRC and to search for novel recessively acting disease loci, we conducted a whole-genome homozygosity analysis of 48 cases with CRC and 100 controls using the Affymetrix GeneChip Human Mapping 250K Sty Array SNP genotyping platform, asking four specific questions. First, do patients with CRC have enrichment for ROHs in particular chromosomal regions as compared with controls? Second, do patients with CRC have longer ROHs as compared with controls? Third, is there a particular SNP that is more likely to be homozygous in patients with CRC as compared with controls? And fourth, are patients with CRC more inbred than controls? To each of these questions in our study, the answer was negative, and thus we found no evidence to support the existence of an association between CRC and increased levels of homozygosity in our study population.

MATERIALS AND METHODS
Patient selection and DNA extraction
Blood samples from 48 cases with CRC were provided by the Colorectal Unit, Department of Surgery, King Faisal Specialist Hospital and Research Centre with long-term follow-up data. A total of 100 matched controls were available from the Blood Bank at King Faisal Specialist Hospital and Research Centre. An on-staff pathologist (P.B.) reviewed all tumors for grade and histological subtype. The institutional review board of the King Faisal Specialist Hospital and Research Centre approved the study.

SNP array procedure
The procedure for the Affymetrix GeneChip Human Mapping 250K Sty SNP array was carried out according to the manufacturer's guidelines. Briefly, 0.25 μg of genomic DNA was digested with *Sty*I. The digests were then ligated to oligonucleotide adapters, polymerase chain reaction (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 on the Affymetrix Fluidics Station 450. The arrays were scanned in the GeneChip Scanner 3000 to generate image (DAT) and cell intensity (CEL) files. All CEL files were imported into the Affymetrix Genotyping Console analysis software 3.0 (Affymetrix) for quality control and to generate SNP calls using the Bayesian robust linear model with Mahalanobis distance classifier algorithm (http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf).

Sequencing analysis
PCR amplification of coded regions with intron–exon boundaries for selected genes (see **Supplementary Table S1** online) and direct sequencing of both strands was performed. The efficiency and quality of the amplification PCR were confirmed by running PCR products on a 2% agarose gel. The PCR products were subsequently subjected to direct sequencing PCR with BigDye terminator V 3.0 cycle sequencing reagents (Applied Biosystems, Foster City, CA). The samples were finally analyzed on an ABI PRISM 3100xl Genetic Analyzer (Applied Biosystems).

Statistical methods
SNP genotyping quality control and allele-calling of the 48 CRC and 100 normal control samples were performed using the Affymetrix Genotyping Console 3.0 analysis software (http://www.affymetrix.com/products_services/software/specific/genotyping_console_software.affx). Only samples for which at least 95% of the full SNP panel had genotype calls were included in further analysis—this yielded 44 CRC and 95 normal samples. Power analysis performed for the 44 CRC and 95 control samples demonstrated that all percentages of homozygous frequencies in **Table 1** for both cases (CRC) and controls (normal) are at least 30%, with odds ratio ≥3.5, suggesting that the power of the comparison in our study is close to 70% (25% when corrected for multiple testing, see **Supplemental Table S2** online). For further downstream analyses of individual SNP homozygosity, inbreeding, and individual ROH, using both full and low-linkage disequilibrium (LD) SNP panels, the plink package (http://pngu.mgh.harvard.edu/purcell/plink/) was used, with default parameters except for minimum length of ROH, minimum number of SNPs per ROH, and maximum number of heterozygous SNPs per ROH. In addition, the frequency of ROHs in multiple samples was verified, along with any co-occurring copy-number variation (CNV) regions, using Genotyping Console 3.0. All statistical tests performed in this study were two-tailed. Comparison and association test statistics and plots were generated using the R Project for Statistical Computing Software (http://www.r-project.org/).

RESULTS
To address the four questions that pertain to the relationship between autozygosity and CRC risk, we performed several tests to assess the homozygosity for SNPs and ROHs between cases and controls. We further analyzed the associations between homozygosity and CRC in subgroups of CRC. All analyses were performed using both the 250K and the low-LD SNP panels.

Genome-wide assessment of associations between homozygosity at individual SNPs and CRC risk
We initially tested the association between homozygosity (for either major or minor allele) and CRC risk for individual SNPs

Table 1 Genome-wide assessment of associations between homozygosity at individual SNPs and CRC risk

Probe set	dbSNP rs ID	Gene	Region	Chromosome	Physical position (bp)	Strand	deCODE genetic map position	Average het rate	Hom CRC	Het CRC	% Hom CRC	Hom normal	Het normal	% Hom normal	Hom OR	Hom Fisher P value	Hom FDR
SNP_A-2193544	rs17581972		Intergenic	18	8,536,673	+	27.95	0.49	30	14	68	34	60	36	3.74	5.16×10^{-4}	4.69×10^{-2}
SNP_A-2260604	rs8090286	*EPB41L3*	Intron	18	5,552,155	−	16.98	0.46	32	12	73	37	57	39	4.06	4.48×10^{-4}	4.12×10^{-2}
SNP_A-4275449	rs7936589		Intergenic	11	23,012,445	+	38.56	0.49	39	5	89	42	52	45	9.51	7.28×10^{-7}	5.79×10^{-4}
SNP_A-1885017	rs4748453	*CACNB2*	Intron	10	18,686,992	+	43.31	0.34	39	2	95	63	30	68	9.17	1.90×10^{-4}	2.19×10^{-2}
SNP_A-1826563	rs2623822		Intergenic	8	68,850,389	+	78.14	0.33	39	5	89	54	40	57	5.71	1.96×10^{-4}	2.19×10^{-2}
SNP_A-4299782	rs10283397	*GFRA2*	Intron	8	21,632,983	+	37.66	0.38	32	12	73	34	60	36	4.65	9.88×10^{-5}	1.18×10^{-2}
SNP_A-4290984	rs7810009		Intergenic	7	56,312,199	+	77.68	0.33	33	11	75	39	55	41	4.19	2.54×10^{-4}	2.35×10^{-2}

CRC, colorectal cancer; FDR, false-discovery rate ; Het, heterozygote; Hom, homozygote; OR, odds ratio; SNP, single-nucleotide polymorphism.

in the 250K panel. Details of the results for the seven SNPs with false-discovery rate <0.05 are shown in **Table 1** (**Figure 1a**). The most strongly associated SNP was rs7936589 (chr11:23,012,445 base; *P* value = 7.28×10^{-7}; odds ratio 9.51). No single SNP reached globally significant association (considered to be *P* value $<10^{-7}$, using the widely accepted threshold for genome-wide association studies); **Figure 1a**). To assess the correlation of genome-wide homozygosity with CRC, we aggregated homozygous SNP counts in cases versus controls, without considering the minor allele frequency. For the SNPs from the entire 250K panel, the median number of homozygous SNPs in cases was 130,613 (interquartile range (IQR) = 2,677), as compared with 130,498 (IQR = 2,930) in controls (Wilcoxon *P* value = 0.987). This lack of association was confirmed when we repeated the analysis using the low-LD panel, with medians of 32,782 (IQR = 695) and 32,804 (IQR = 769), respectively (*P* = 0.957, Wilcoxon test; **Table 2**). In addition, we calculated the inbreeding coefficient (*F*) across all samples (see **Supplementary Tables S3** and **S4** online). The median (and IQR) for *F* in cases and controls using all SNPs were 0.029 (0.042) and 0.023 (0.063), respectively, not significantly different from each other (Wilcoxon *P* value = 0.151), with similar results for the low-LD SNP panel (**Table 2**). Therefore, we could find no evidence to suggest that cases were, in general, more inbred than controls.

Analysis of ROHs in cases and controls

The existence of LD blocks means that relatively short ROHs, from tens to hundreds of kilobases, are common across the genome.[17] Evidently, most of these regions probably do not result from true autozygosity. We therefore set thresholds to define an ROH based on genomic regions where either a minimum number of (50) consecutive, nonmissing SNPs were homozygous (allowing for miscalls) or in which homozygosity extended for a minimum length (250 kb) along the chromosome. We also calculated total ROH length per individual (i.e., the sum of the lengths of the ROHs in their genome) as a more robust measure of autozygosity than counting the number of ROHs per genome. For example, with a threshold ROH size of 4 Mb, the latter method would twice score two ROHs of 4 Mb but would only score a single region of 8 Mb once; in terms of indicating autozygosity, however, a single 8-Mb region would be at least as important as two 4-Mb regions.

To provide a comparison with the work of Bacolod and colleagues,[13] we initially analyzed ROHs that were ≥50 SNPs in size using the 250K SNP panel. Every individual had at least 22 ROHs and the median number of these regions per individual was approximately 148. There was no evidence of an association between the total ROH size in each individual and CRC (*P* = 0.201, Wilcoxon test; **Table 3**). To determine whether this result was robust, we repeated the analysis using a number of different criteria to define a ROH (≥30 SNPs, ≥40 SNPs, ≥60 SNPs, ≥1 Mb, ≥2 Mb, ≥4 Mb, and ≥10 Mb; **Figure 1b–d**). The only evidence found for an association between total size of the ROHs in each individual and CRC was for ROHs containing ≥30 SNPs (**Table 3**). However, this result was not confirmed on repeating
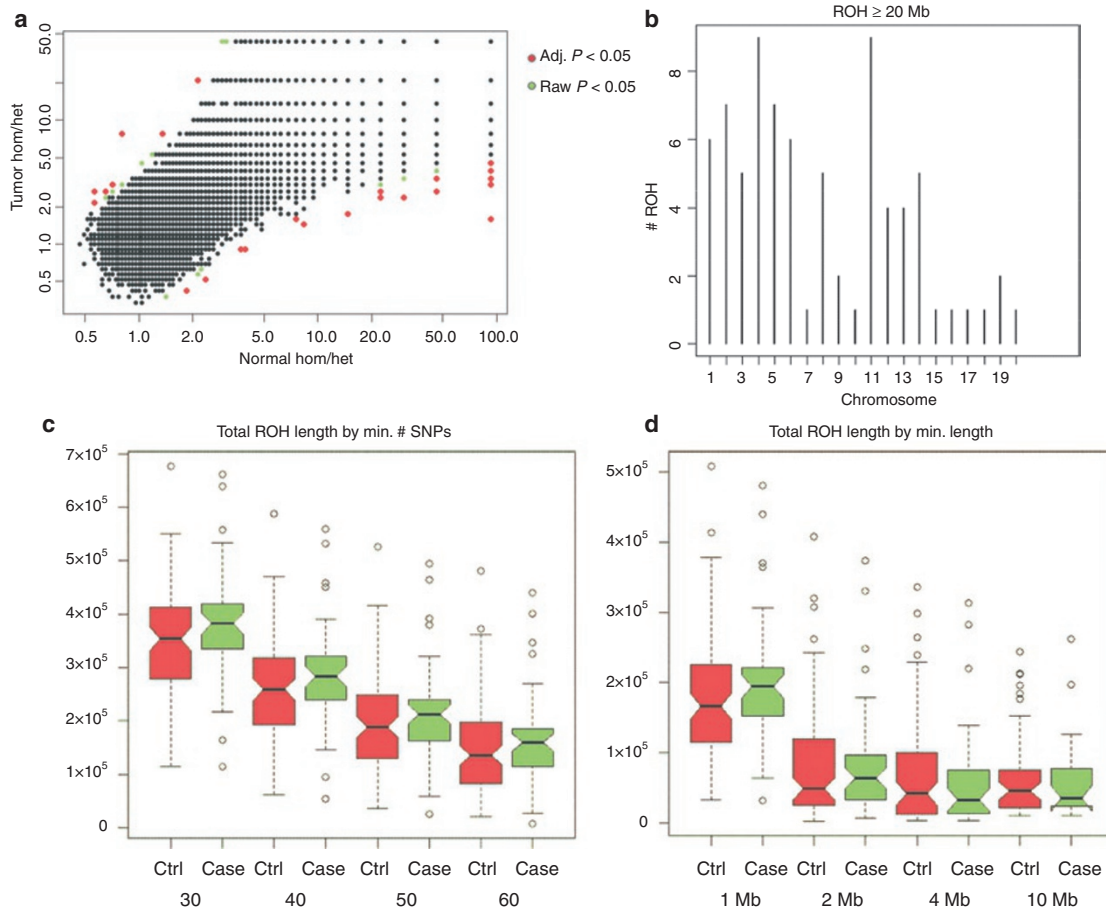
Figure 1 SNP- and ROH-based association of CRC cases and controls. (**a**) SNP-based homozygosity association. (**b**) ROH counts by chromosomes (≥20 Mb). (**c**) Boxplot showing total ROH length by minimum number of SNPs between cases and controls. (**d**) Boxplot showing total ROH length by minimum length between cases and controls. Ctrl, control; het, heterozygote; hom, homozygote; ROH, run of homozygosity; SNP, single-nucleotide polymorphism.

the analysis in the low-LD SNP panel using the same ROH defining criteria as above. Every individual carried at least one homozygous region (median 4.0) detected by the low-LD SNP panel and virtually all of these regions were large (1–42 Mb in size). Analysis of the total size of ROHs in each individual determined again that there was no significant difference between cases and controls (**Table 3**). The total length of ROHs detected in each person using the 250K and low-LD SNP panel are summarized in (**Figure 1c,d**) and detailed in **Supplementary Tables S3 and S4** online, respectively. **Figure 1b** illustrates the overall distribution of large ROHs (≥20MB) across all chromosomes.

To provide a further comparison between our results and those of Bacolod and colleagues,[13] we calculated the frequencies of cases and controls in which we detected one or more ROHs of ≥4 Mb in length. Using the 250K SNP panel, 39 of 44 (88.6%) cases and 72 of 95 (75.8%) controls had these ROHs ($P = 0.11$, Fisher's exact test). For the low-LD panel, 29 of 44 (65.9%) cases and 47 of 95 (49.5%) controls had ROHs of ≥4 Mb ($P = 0.10$, Fisher's exact test). We thus failed to detect the marked significant difference between cases (62.2%) and two sets of controls (35.6% and 28.8%) that was seen in the study by Bacolod and colleagues. The longest homozygous regions were derived from individuals

in our study who were found to have higher levels of inbreeding ($F > 0.06$) than the median ($F = 0.026$ for all samples) from a population with strong evidence of consanguineous marriage.

Recurrent ROHs

Although most of the ROHs examined thus far were individually uncommon, some occurred in >10% of cases or controls when assessed using the 250K panel. We therefore addressed whether any of these specific, relatively common homozygous regions were associated with CRC risk. Using ROHs with ≥1 Mb of consecutive homozygous SNPs detected in the 250K panel, we searched for overlapping ROHs that were found in more than five individuals (cases and/or controls). This resulted in a total of 4,169 ROHs that met the inclusion criteria. After taking multiple testing into account, 12 ROHs reached global significance for an association between homozygosity and CRC risk (**Table 4**), all of which were more common in cases than controls. One ROH was found in five different samples at chromosome 11q near the centromere.

Confirmation of homozygous regions

Homozygous regions might result from chance, autozygosity, uniparental isodisomy, or hemizygosity. To determine whether

Table 2 Genome-wide assessment of associations between aggregate SNP homozygosity for either allele, and calculation of inbreeding coefficient (*F*) for all samples using the 250K and the low-LD SNP set

Group	Median no. hom SNPs	IQR for no. hom SNPs	Median *F*	IQR *F*	LD-pruned median no. hom SNPs	LD-pruned IQR for no. hom SNPs	LD-pruned median *F*	LD-pruned IQR *F*
Case	130,613	2,677	0.029	0.042	32,782	695	0.032	0.045
Control	130,498	2,930	0.023	0.063	32,804	769	0.021	0.066
Wilcoxon *P* value	0.987	—	0.151	—	0.957	—	0.177	—

hom, homozygous; IQR, interquartile range; LD, linkage disequilibrium; SNP, single-nucleotide polymorphism.

Table 3 Comparison of the total ROH size (in Mb) in cases versus controls

Total ROH length	Min. ROH, length (Mb)	Case Median	Case IQR	Control Median	Control IQR	Wilcoxon *P* value	Total ROH length	Min. ROH, length (Mb)	Case Median	Case IQR	Control Median	Control IQR	Wilcoxon *P* value
All SNPs	≥1	194.7	66.2	166.3	110.7	1.07×10^{-1}	LD SNPs	≥1	23.8	31.4	18.3	31.4	1.78×10^{-1}
	≥2	64.3	61.0	48.9	94.0	3.01×10^{-1}		≥2	17.0	24.2	17.2	29.3	9.92×10^{-1}
	≥4	32.9	62.4	41.9	85.3	6.10×10^{-1}		≥4	12.9	18.8	16.5	31.8	2.77×10^{-1}
	≥10	35.1	52.7	45.9	53.9	8.48×10^{-1}		≥10	16.8	19.0	20.0	19.1	9.33×10^{-1}

Total ROH length	Min. ROH, SNP count	Case Median	Case IQR	Control Median	Control IQR	Wilcoxon *P* value	Total ROH length	Min. ROH, SNP count	Case Median	Case IQR	Control Median	Control IQR	Wilcoxon *P* value
All SNPs	≥30	382.3	79.6	353.8	133.5	3.75×10^{-2}	LD SNPs	≥30	27.3	32.5	20.2	30.9	2.08×10^{-1}
	≥40	283.6	78.7	259.1	125.2	9.98×10^{-2}		≥40	19.6	29.8	18.2	28.4	6.86×10^{-1}
	≥50	212.6	72.5	188.3	119.4	2.01×10^{-1}		≥50	18.0	22.0	15.4	30.5	6.65×10^{-1}
	≥60	159.9	70.1	135.4	115.7	1.56×10^{-1}		≥60	17.3	21.3	16.9	30.9	9.66×10^{-1}

Size data shown are for the 250K SNP panel for ≥ 60 SNPs, ≥ 50 SNPs, ≥ 40 SNPs, and ≥ 30 SNPs within an ROH, and for ≥ 10 Mb,≥ 4 Mb, ≥ 2 Mb, and ≥ 1 Mb minimum ROH size.

IQR, interquartile range; LD, linkage disequilibrium; Min., minimum; ROH, runs of homozygosity; SNP, single-nucleotide polymorphism.

the common homozygous regions were actually hemizygous CNVs, the positions of the common ROHs were compared with those of CNVs identified using Genotyping Console 3.0. We searched for CNVs that covered at least 90% of the detected ROH. However, none of the common ROHs we found could be explained by a CNV (data not shown).

Additionally, a subgroup of patients with microsatellite unstable (microsatellite unstable–positive) tumors were compared with the negative group. Molecular diagnosis showed that 7 of 48 cases were microsatellite unstable–positive. After correcting for multiple testing, eight ROHs reached global significance for an association between homozygosity and CRC risk in microsatellite unstable–positive cases, which were shared by five of the six samples. For selected genes from these regions, the encompassing coding region with intron–exon boundaries were sequenced; however, no mutations were detected (see **Supplementary Table S5** online). Analysis with subgroups did not significantly change our results with respect to our four hypothesized questions (data not shown).

DISCUSSION

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. That these regions resulted from uniparental disomy (an instance when an offspring inherits both copies or segments of chromosomes from a single parent), although possible, is highly unlikely. Recent studies[12,13] have reported an increased frequency of homozygous microsatellites or ROHs in cancer cases as compared with controls and that these regions showing identity by descent may be the locations of genes contributing to tumor heritability.[12,13] Moreover, these data have been interpreted as providing an explanation for the increased cancer rates in populations with higher degrees of consanguinity. There are studies that have compared the incidence of cancer and other late-onset complex diseases between individuals from genetically isolated islands in Middle Dalmatia, Croatia, and a control population.[18] These studies suggest that inbreeding can be a positive predictor for a number of late-onset diseases such as heart disease, stroke, and cancer. Similar observations were published in a Pakistani study where cancer patients, on average, have a higher coefficient of inbreeding as compared with the general population.[7] Another study demonstrated that 94% of the subjects with reported adenocarcinomas (mostly colorectal) originated from a consanguineous population of descendants of an Italian immigrant group in Wisconsin,[6] suggesting an explanation for increased cancer incidence with higher degree of inbreeding. Thus, an explanation for increased

Table 4 The recurrent ROHs that were identified in cases and controls using the 250K SNP panel

Chr	Map loc	Left (Mb)	Right (Mb)	Length (Kb)	Case	Control	CI-left	OR	CI-right	Fisher P value	Bonf. adj. P value	n Genes	Genes
1	q21.3	149.63	149.67	40	19	5	1.8	5.8	22.1	8.54×10^{-4}	4.27×10^{-2}	2	PSMB4, POGZ
1	q21.3	149.76	149.92	160	17	1	3.7	26.1	1,138.5	1.15×10^{-5}	5.77×10^{-4}	3	CGN, TUFT1, SNX27
2	p13.11	74.39	74.88	490	24	2	5.1	23.2	218.4	1.24×10^{-7}	6.19×10^{-6}	18	DCTN1, WDR54, **RTKN**, ZNHIT4, WBP1, GCS1, MRPL53, CCDC142, TTC31, LBX2, PCGF1, **TLX2**, DQX1, AUP1, **HTRA2**, LOXL3, **DOK1**, C2orf65
3	p24.3	17.21	17.70	490	24	0	11.0	100.0	Inf	7.02×10^{-10}	3.51×10^{-8}	1	**TBC1D5**
3	p21.32	44.65	44.74	90	18	3	2.4	9.2	52.9	1.53×10^{-4}	7.67×10^{-3}	3	ZNF197, ZNF35, ZNF502
5	q31.3	139.68	140.22	540	11	0	3.1	100.0	Inf	2.02×10^{-4}	1.01×10^{-2}	18	HBEGF, SLC4A9, ANKHD1, ANKHD1-EIF4EBP3, EIF4EBP3, SRA1, APBB3, SLC35A4, CD14, TMCO6, NDUFA2, IK, WDR55, DND1, HARS, HARS2, ZMAT2, PCDHA1
5	q33.1	150.67	150.92	250	17	1	3.7	26.1	1,138.5	1.15×10^{-5}	5.77×10^{-4}	2	SLC36A2, SLC36A1
7	q33.1	134.25	134.37	120	19	4	2.2	7.4	32.8	2.66×10^{-4}	1.33×10^{-2}	1	CALD1
10	p15.2	3.03	3.07	37	31	5	5.0	15.8	61.1	1.64×10^{-8}	8.21×10^{-7}	0	Upstream of PFKP
10	p11.21	35.26	35.60	340	14	1	2.8	19.7	866.3	1.50×10^{-4}	7.51×10^{-3}	2	CUL2, CREM
21	q22.13	38.28	38.43	150	10	0	2.7	100.0	Inf	4.67×10^{-4}	2.34×10^{-2}	1	DSCR4
22	q13.1	37.23	37.52	290	15	2	2.3	10.7	102.4	3.72×10^{-4}	1.86×10^{-2}	7	DMC1, **CBY1**, TOMM22, JOSD1, **GTPBP1**, UNC84B, DNAL4
11	q22.1	100.83	100.92	90	17	4	1.8	6.2	27.7	1.17×10^{-3}	5.84×10^{-2}	1	TRPC6
7	q21.2	91.20	91.61	410	24	9	1.7	4.5	12.9	1.21×10^{-3}	6.03×10^{-2}	3	MTERF, AKAP9, CYP51A1
22	q13.1	39.38	39.69	310	13	2	1.8	8.7	84.8	1.69×10^{-3}	8.45×10^{-2}	6	MCHR1, SLC25A17, **ST13**, XPNPEP3, DNAJB7, RBX1
12	q24.31	119.83	120.13	300	13	2	1.8	8.7	84.8	1.69×10^{-3}	8.45×10^{-2}	6	HNF1A, C12orf43, OASL, OASL, P2RX7, P2RX4
16	q21	63.50	63.75	250	13	2	1.8	8.7	84.8	1.69×10^{-3}	8.45×10^{-2}	1	CDH11
2	q23.3	152.29	152.50	210	13	2	1.8	8.7	84.8	1.69×10^{-3}	8.45×10^{-2}	3	NEB, ARL5A, CACNB4
19	p13.11	19.36	19.52	160	13	2	1.8	8.7	84.8	1.69×10^{-3}	8.45×10^{-2}	5	GATAD2A, TSSK6, NDUFA13, FLJ44968, CILP2
6	p21.1	41.09	41.23	140	13	2	1.8	8.7	84.8	1.69×10^{-3}	8.45×10^{-2}	8	UNC5CL, BZRPL1, APOBEC2, C6orf130, NFYA, NFYA, TREML1, TREM2
19	p13.2	9.22	9.81	590	17	5	1.5	4.9	18.7	3.38×10^{-3}	1.69×10^{-1}	12	ZNF699, ZNF559, ZNF177, ZNF266, ZNF560, ZNF426, ZNF121, ZNF561, ZNF562, ZNF846, FBXL12, UBL5
10	q23.31	98.84	99.10	260	12	2	1.6	7.8	76.4	3.48×10^{-3}	1.74×10^{-1}	4	SLIT1, **ARHGAP19**, FRAT1, FRAT2
20	q13.13	46.86	47.28	420	21	8	1.5	4.0	12.1	3.75×10^{-3}	1.88×10^{-1}	5	PREX1, ARFGEF2, CSE1L, STAU1, DDX27
1	p31.1	75.72	76.05	330	18	6	1.4	4.3	15.0	4.52×10^{-3}	2.26×10^{-1}	4	SLC44A5, ACADM, RABGGTB, MSH4

All regions had a minimum ROH size of 1 Mb and overlapped with ROHs in more than five individuals. *P* values for significant difference in ROH frequency between cases and controls were calculated using Fisher's exact test.
Bonf., Bonferroni; Chr, chromosome; CI, confidence interval; Inf, infinity; loc, loci; OR, odds ratio; ROH, run of homozygosity; SNP, single-nucleotide polymorphism.
Boldface values represent the genes selected for direct sequencing.

Table 4 Continued on next page.

Table 4 Continued.

Chr	Map loc	Left (Mb)	Right (Mb)	Length (Kb)	Case	Control	CI-left	OR	CI-right	Fisher P value	Bonf. adj. P value	n Genes	Genes
3	q23	140.53	140.73	200	18	6	1.4	4.3	15.0	4.52×10^{-3}	2.26×10^{-1}	4	MRPS22, COPB2, RBP2, RBP1
8	q24.23	138.59	139.74	1,150	15	4	1.5	5.1	23.2	4.60×10^{-3}	2.30×10^{-1}	2	COL22A1, FAM135B
12	q21.31	83.95	84.25	300	13	3	1.4	5.7	33.6	5.91×10^{-3}	2.95×10^{-1}	3	TSPAN19, LRRIQ1, ALX1,
19	p13.11	39.33	39.97	640	9	1	1.4	11.1	504.2	7.19×10^{-3}	3.59×10^{-1}	9	LSM14A, KIAA0355, GPI, PDCD2L, UBA2, WTIP, LOC284402, ZNF302, ZNF181
12	q13.11	46.70	47.05	350	14	4	1.3	4.7	21.2	8.77×10^{-3}	4.39×10^{-1}	8	TMEM106C, COL2A1, SENP1, PFKM, ASB8, C12orf68, OR10AD1, H1FNT
13	q31.1	78.77	79.03	260	12	3	1.3	5.1	30.5	1.14×10^{-2}	5.68×10^{-1}	2	RBM26, NDFIP2
5	q14.1	77.90	78.00	100	12	3	1.3	5.1	30.5	1.14×10^{-2}	5.68×10^{-1}	1	LHFPL2
1	p36.21	15.56	15.65	90	12	3	1.3	5.1	30.5	1.14×10^{-2}	5.68×10^{-1}	3	FHAD1, EFHD2, CTRC
1	p36.13	19.43	19.53	100	15	5	1.2	4.0	15.6	1.19×10^{-2}	5.96×10^{-1}	6	KIAA0090, MRTO4, AFAR3, AKR7A3, AKR7A2, PQLC2
13	q31.1	85.18	85.48	300	13	4	1.2	4.2	19.3	1.63×10^{-2}	8.13×10^{-1}	1	SLITRK6
3	p14.1	68.79	68.98	190	13	4	1.2	4.2	19.3	1.63×10^{-2}	8.13×10^{-1}	1	FAM19A4
18	q21.2	48.85	49.20	350	11	3	1.1	4.6	27.5	2.12×10^{-2}	1.00×10	1	**DCC**
22	q12.3	33.20	33.50	300	15	6	1.1	3.3	11.5	2.67×10^{-2}	1.00×10	0	
15	q21.1	39.30	39.64	340	10	3	1.0	4.1	24.6	3.87×10^{-2}	1.00×10	11	INOC1, EXDL1, CHP, OIP5, NUSAP1, NDUFAF1, RTF1, ITPKA, LTK, RPAP1, TYRO3
15	q25.1	79.26	79.54	280	10	3	1.0	4.1	24.6	3.87×10^{-2}	1.00×10	3	IL16, STARD5, TMC3
11	p11.2	46.30	47.30	1,000	18	9	1.0	2.7	7.8	4.18×10^{-2}	1.00×10	20	DGKZ, **MDK**, CHRM4, AMBRA1, HARBI1, KIAA0652, ARHGAP1, ZNF408, F2, CKAP5, LRP4, C11orf49, C11orf49, ARFGAP2, PACSIN3, PACSIN3, DDB2, ACP2, NR1H3, MADD
17	q24.1	61.22	61.65	430	11	4	0.9	3.4	15.8	5.13×10^{-2}	1.00×10	2	CCDC46, APOH
6	q24.3	145.80	146.20	400	16	8	0.9	2.6	7.9	6.07×10^{-2}	1.00×10	2	EPM2A, FBXO30
12	q24.31	121.80	122.58	780	14	7	0.8	2.5	8.2	8.64×10^{-2}	1.00×10	17	DENR, CCDC62, HIP1R, VPS37B, ABCB9, OGFOD2, ARL6IP4, PITPNM2, MPHOSPH9, C12orf65, **CDK2AP1**, SBNO1, SETD8, RILPL2, U1SNRNPBP, U1SNRNPBP, RILPL1
5	q22.2	112.05	112.34	290	14	7	0.8	2.5	8.2	8.64×10^{-2}	1.00×10	3	**APC**, SRP19, **REEP5**
22	q12.2	29.86	30.17	310	7	2	0.7	4.0	42.0	8.83×10^{-2}	1.00×10	7	PIB5PA, PLA2G3, RNF185, LIMK2, PIK3IP1, **PATZ1**, **DRG1**
20	q13.31	55.24	55.54	300	7	2	0.7	4.0	42.0	8.83×10^{-2}	1.00×10	6	BMP7, SPO11, RAE1, RBM38, HMG1L1, CTCFL
5	q23.1	115.45	115.70	250	17	10	0.8	2.2	6.1	1.14×10^{-1}	1.00×10	1	COMMD10

All regions had a minimum ROH size of 1 Mb and overlapped with ROHs in more than five individuals. *P* values for significant difference in ROH frequency between cases and controls were calculated using Fisher's exact test.

Bonf., Bonferroni; Chr, chromosome; CI, confidence interval; Inf, infinity; loc, loci; OR, odds ratio; ROH, run of homozygosity; SNP, single-nucleotide polymorphism.

Boldface values represent the genes selected for direct sequencing.

ORIGINAL RESEARCH ARTICLE

cancer risk based on the frequency of homozygous regions and consanguinity has formed the basis of a potentially new model of cancer progression.[19]

In this study, we have used Affymetrix GeneChip 250K SNP arrays to compare the structure of genetic variation in patients with CRC to that of healthy controls. Our analyses analyzed SNPs from both the full panel as well as those in low pairwise LD. Overall, evidence for an association between homozygosity and CRC risk was limited. Our results demonstrated a total of 4,169 ROHs that were found in more than five individuals (either cases or controls) and met the inclusion criteria. After taking multiple testing into account, only 12 ROHs reached significance for an association between homozygosity and CRC risk (Table 4), all of which were more common in cases than controls. Sequencing of selected genes with regions, including coding and intron/exon boundaries yielded no mutations (see **Supplementary Table S1** online). We did not find cases to be significantly more or less inbred than controls. Furthermore, our ROH analysis provided no evidence for an association between total ROH size per individual and increased risk of CRC, under any of several size criteria, using either the 250K or the low-LD SNP panel.

The assertion that increased autozygosity correlates with cancer incidence provides an attractive explanation for reported increased cancer risk in inbred populations. However, several criticisms can be leveled at this assertion. The observation of an increased cancer risk associated with consanguinity has often been based on studies of a small number of individuals in an isolated community or a single large family with a high level of inbreeding.[20] Thus, the relevance of inbreeding to the population risk of cancer is unclear, as inbreeding and founder effects may be confounded. Lack of confirmation of these results by our study could partially be attributable to sample size, which was relatively small. Nonetheless, the negative results in this study of matched cases and controls from an inbred population carry more significance than what has been previously reported in the literature. Previous molecular studies have sought to establish a relationship between ROHs and cancer risk with case–control groups that have been ethnically heterogeneous or unmatched.[12,13] In addition these studies made use of relatively sparse microsatellite data, whereby, in our small cohort case–control study of CRC we have addressed these shortcomings by analyzing samples using a genome-wide 250K SNP platform and imposed a high level of quality control both in terms of genotyping and sample ancestry. Furthermore, our data from a subgroup of patients with microsatellite unstable–positive tumors did not significantly change our results. Of note, our sample set did not identify homozygosity across *MYH*, perhaps the only known recessively acting CRC gene. It is possible that our study sample size is not large enough to detect this or similarly recessive genes.

Our results are concordant with similar studies carried out in different cancers, such as breast, prostate, leukemia, and colon cancer,[14–16] in a predominantly consanguineous population.

Because our analysis suggests that whole-genome homozygosity analysis of inbred populations may not provide a robust methodology for identifying novel cancer susceptibility loci, it would be worthwhile to apply new strategies such as exome or whole-genome sequencing in the future to unveil these underlying, predisposing recessive alleles. It is unlikely that large numbers of recessive alleles exist that predispose to CRC and would be unmasked by autozygosity in inbred populations, such as that in Saudi Arabia.

In conclusion, our findings do not provide evidence that increased levels of homozygosity confer an increased risk of developing CRC. Although these results do not rule out the potential presence of recessively acting CRC-predisposing genes in a small percentage of patients that our relatively small sample size could not capture, they do suggest that such genes are unlikely to account for the disturbingly high incidence of CRC in our consanguineous population and that future research should consider other mechanisms.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/gim

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. Bavi PP, Abubaker JA, Jehan ZD, et al. Colorectal carcinomas from Middle East. Molecular and tissue microarray analysis of genomic instability pathways. *Saudi Med J* 2008;29:75–80.
2. Al-Eid HS. Cancer Incidence Report Saudi Arabia 1999–2000. King Faisal Specialist Hospital and Research Centre, 2004. http://www.scr.org.sa/reports/SCR2000.pdf.
3. Bener A, El Ayoubi HR, Chouchane L, et al. Impact of consanguinity on cancer in a highly endogamous population. *Asian Pac J Cancer Prev* 2009;10:35–40.
4. Feldman JG, Lee SL, Seligman B. Occurrence of acute leukemia in females in a genetically isolated population. *Cancer* 1976;38:2548–2550.
5. Abramson JH, Pridan H, Sacks MI, Avitzour M, Peritz E. A case-control study of Hodgkin's disease in Israel. *J Natl Cancer Inst* 1978;61:307–314.
6. Lebel RR, Gallagher WB. Wisconsin consanguinity studies. II: Familial adenocarcinomatosis. *Am J Med Genet* 1989;33:1–6.
7. Shami SA, Qaisar R, Bittles AH. Consanguinity and adult morbidity in Pakistan. *Lancet* 1991;338:954.
8. Simpson JL, Martin AO, Elias S, Sarto GE, Dunn JK. Cancers of the breast and female genital system: search for recessive genetic factors through analysis of human isolate. *Am J Obstet Gynecol* 1981;141:629–636.
9. Demant P. Cancer susceptibility in the mouse: genetics, biology and implications for human cancer. *Nat Rev Genet* 2003;4:721–734.
10. el-Hazmi MA, al-Swailem AR, Warsy AS, al-Swailem AM, Sulaimani R, al-Meshari AA. Consanguinity among the Saudi Arabian population. *J Med Genet* 1995;32:623–626.
11. Alkuraya FS. Autozygome decoded. *Genet Med* 2010;12:765–771.
12. Assié G, LaFramboise T, Platzer P, Eng C. Frequency of germline genomic homozygosity associated with cancer cases. *JAMA* 2008;299:1437–1445.

13. Bacolod MD, Schemmann GS, Wang S, et al. The signatures of autozygosity among patients with colorectal cancer. *Cancer Res* 2008;68:2610–2621.

14. Enciso-Mora V, Hosking FJ, Houlston RS. Risk of breast and prostate cancer is not associated with increased homozygosity in outbred populations. *Eur J Hum Genet* 2010;18:909–914.

15. Hosking FJ, Papaemmanuil E, Sheridan E, et al. Genome-wide homozygosity signatures and childhood acute lymphoblastic leukemia risk. *Blood* 2010;115:4472–4477.

16. Spain SL, Cazier JB, Houlston R, Carvajal-Carmona L, Tomlinson I; CORGI Consortium,. Colorectal cancer risk is not associated with increased levels of homozygosity in a population from the United Kingdom. *Cancer Res* 2009;69:7422–7429.

17. Abecasis GR, Ghosh D, Nichols TE. Linkage disequilibrium: ancient history drives the new genetics. *Hum Hered* 2005;59:118–124.

18. Rudan I, Rudan D, Campbell H, et al. Inbreeding and risk of late onset complex disease. *J Med Genet* 2003; 40:925–932.

19. Bacolod MD, Schemmann GS, Giardina SF, Paty P, Notterman DA, Barany F. Emerging paradigms in cancer genetics: some important findings from high-density single nucleotide polymorphism array studies. *Cancer Res* 2009;69:723–727.

20. Akasaka T, Balasas T, Russell LJ, et al. Five members of the CEBP transcription factor family are targeted by recurrent IGH translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL). *Blood* 2007;109:3451–3461.