

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Single Nucleotide Polymorphisms and Colorectal Cancer Risk: The First Replication Study in a South American Population

Israel Gomy, Ligia Petrolini de Oliveira and
Benedito Mauro Rossi

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.69994>

Abstract

Colorectal cancer (CRC) heritability is determined by the complex interaction between inherited variants and environmental factors. CRC incidence rates have been increasing specially in developing countries, such as Brazil, where CRC is the third most frequent cancer in both genders. Genome-wide association studies (GWAS), based on thousands of cases and controls typed at thousands of single nucleotide polymorphisms (SNPs), have identified several variants that associate with gastrointestinal cancer risk. Less of half of the familial risk has been elucidated through GWAS that identified common SNPs in almost exclusively European populations. Replication studies in admixed heterogeneous populations are scarce and most failed to replicate all the imputed SNPs. Population stratification by ethnic subgroups with different allele frequencies and so with different patterns of linkage disequilibrium may cause expurious associations. Here, we show the first replication study of CRC inherited susceptibility in South America and aimed to identify known SNPs, which are associated with CRC risk in European populations.

Keywords: cancer susceptibility, cancer risk, colorectal cancer, single nucleotide polymorphism, South America

1. Introduction

1.1. Epidemiology

Colorectal cancer (CRC) is one of the most prevalent cancers in both genders worldwide, responsible for about 10% of all neoplasms, mainly in developed industrialized countries, such as Australia and New Zeland, North America, and Europe [1].

The variation in the CRC global incidence rates is progressively decreasing, and the increase of new cases in historically low incidence populations, such as Asian, seems to come along with a greater economic development and, consequently, a greater exposure to risk factors so-called “Western lifestyle.” The highest incidence rates are in countries such as Australia, New Zealand, and in Western Europe, whereas the lowest incidence rates are in Africa and Southern Asia [2].

However, in developing countries such as Russia, Mexico, Chile, and Brazil, there is an increase of CRC incidence rates, possibly reflecting their population’s lifestyles and the increased life expectancy, as well as the improved data records, the insufficient public health policies for CRC screening, and the lack of population awareness.

Data from international registries indicate that the overall 5-year survival rate is on average 55% in developed countries and 40% in developing countries and is higher in women than in men [3].

According to estimates from the National Cancer Institute for the 2016/2017 biennium, in Brazil, 34,280 individuals will develop CCR (17,620 women and 16,660 men), corresponding to the second most common malignancy in women and third most common malignancy in men [4].

The detection and early removal of premalignant lesions reduces CRC mortality as confirmed by several studies that have screened populations at general risk using the fecal occult blood test. In addition, the use of flexible sigmoidoscopy for screening has shown promising results in randomized trials in the United Kingdom [5] and the United States [6], where significant reductions in both incidence and mortality were observed. Improved survival was also observed with this approach in genetically determined high-risk groups [7]. Therefore, for those determined high-risk individuals could be offered a more intensive surveillance, with colonoscopy or flexible sigmoidoscopy periodically at shorter intervals. Colonoscopy is already offered to individuals at high risk due to personal or familial history of CRC, as well as for families with Lynch syndrome and intestinal polyposis syndromes, for which more assiduous surveillance is recommended [8]. Therefore, stratifying the general population into risk categories would allow the individualization of screening and prevention strategies.

1.2. Molecular pathogenesis

The classic adenoma-carcinoma sequence has revealed an intricate molecular pathogenesis of CRC, where tumor suppressor genes are inactivated, and proto-oncogenes are activated through several signaling pathways, such as *APC-B-catenin*, *RAS-RAF*, *PIK3CA-PTEN*, and *TGF-B* [9].

Three main molecular mechanisms are involved in CRC pathogenesis: chromosomal instability, microsatellite instability, and serrated polyp pathway. The first one occurs in most sporadic cancers where the accumulation of mutations, rearrangements, and aneuploidy drives malignant transformation within decades [10]. The second one occurs in about 15% of sporadic CRC and in most hereditary CRC. In sporadic CRC, an epigenetic event—hypermethylation—occurs in CpG islands of MMR gene promoters, which silences them leading to a genetic instability in microsatellite regions of genome [11]. In Lynch syndrome, mutations in MMR genes lead to microsatellite instability and accelerate adenoma-carcinoma sequence

more rapidly, the reason why Lynch syndrome families develop cancer in their 40's or even earlier. The most recently serrated polyp pathway involves molecular mechanisms other than the classic adenoma-carcinoma sequence but has not yet been fully elucidated [12].

1.3. Risk factors

Colorectal carcinoma is a multifactorial disease, where complex interactions between genetic and environmental factors determine individual risk. Among the latter are diets rich in red meat and animal fat and lower in fiber, smoking, alcohol consumption, obesity, sedentary lifestyle, and chronic inflammatory bowel disease [13]. In addition to age, gender, and previous history of polyps, familial history is considered the main risk factor, being the relative risk between siblings two to three times higher than in the general population [14].

Traditionally, CRC has been classified into sporadic and hereditary. The concept of familial CRC reflects one end of a risk spectrum determined by the contribution of genetic variants of susceptibility. Most are sporadic with no family history and known genetic susceptibility. Most of the CRC susceptibility genes were identified in families affected by inherited syndromes, which are caused by mutations with high penetrance. These syndromes account for about 6% of CRC cases and can be classified as syndromes with or without gastrointestinal polyposis [15]. Among the main syndromes with polyposis are familial adenomatous polyposis (FAP) caused by mutations in the *APC* gene; Peutz-Jeghers syndrome, attributable to mutations in the *STK11* gene; Juvenile polyposis, associated with the *BMPR1A* gene, and Cowden's syndrome, related to the *PTEN* gene. Among non-polyposis syndromes, the most prevalent is Lynch's syndrome, accounting for about 3% of all the cases with CRC, caused by mutations in the mismatch repair genes during DNA replication (*MLH1*, *MSH2*, *MSH6*, *PMS2*, and *EPCAM*) [16].

Most of the mutations identified in familial CRC are highly penetrant, that is, with a high chance of manifesting cancer throughout the life. However, there are families with CRC clusters that do not have mutations in genes associated with hereditary syndromes. This raises the hypothesis that there are other variants or mutations with low penetrance that make certain individuals more susceptible to the CRC development. Studies with brothers with and without CRC, as well as several association studies, have identified regions in the human genome in which single nucleotide polymorphisms (SNPs) variants are associated with CRC susceptibility [17].

Up to 25% of cases are familial CRC aggregations whose heritability has been partially uncovered by GWAS SNPs [18]. However, the large proportion of familial risk remains unexplained—so-called missing heritability.

1.4. Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) are variations of the human genome, where two or occasionally three alternative nucleotides are common in the population. In most cases, an SNP has two alternative forms, termed alleles, for example, A or G at a certain position in the genome.

There are 10 million SNPs estimated in the human genome, representing, along with other types of polymorphisms (such as copy number variations), about 90% of human genetic variation, including susceptibility to disease. Two individuals are 99.5% identical in their DNA sequences, and, every 1000 base pairs, there is one SNP [19].

Variants that have been deleterious during evolution are particularly rare due to natural selection. In turn, pathogenic variants that are deleterious in homozygosis may have become neutral or undergo a selection balance by conferring an advantage on asymptomatic heterozygotes. Therefore, alleles of frequent SNPs are not expected to have any significant phenotypic effect, either because natural selection would be in charge of eliminating it, - if it were detrimental by negative selection -, or fixing it, if beneficial by positive selection. Moreover, most SNPs are not located in either coding or regulatory sequences but in intergenic sequences [20].

1.5. Genome-wide association studies

Searching for population associations is an attractive option to identify disease susceptibility genes. Association studies are easier to conduct than linkage studies because they do not require multiple family cases segregating the phenotype. However, they depend on the linkage disequilibrium (LD)—the nonrandom association between alleles at different loci—with a susceptibility factor, which can only be identified by markers located in the same haplotype block (the set of alleles at a linked locus in a single chromosome) close to the factor [21].

SNPs are the markers of choice for studying LD for three reasons: (1) they are sufficiently abundant that they allow verifying very short chromosome segments; (2) compared to microsatellites, they have a lower mutation rate; (3) SNPs are easily large-scale genotyped on genome [20].

The structure of the human LD was investigated by the HapMap project, and the first result was a list of more than 3 million SNPs that captured most of the common genomic variation in some populations [22].

Genome-wide association studies (GWAS) are based on the LD principle at the population level, which is usually the result of a particular ancestral haplotype common in a population. Usually, loci that are physically close exhibit a stronger LD than those that are distant in a chromosome. The genomic distance at which LD decays determines how many genetic markers are required to “tag” a haplotype block, being the number of such markers much smaller than the total of segregating variants in the population. For example, the selection of about 500,000 common SNPs in the human genome is sufficient to “tag” the common variants in non-African population, even though the total SNPs are greater than 10 million [22]. These SNPs are called tagSNPs.

Although GWAS are not influenced by prior biological knowledge or genomic location of SNPs, they are influenced by LD between genotyped SNPs and non-genotyped causative variants. The strength of the statistical association between the alleles at the two loci in the genome depends, mainly, on their allelic frequencies. Thus, a rare variant—minor allele frequency (MAF) less than 0.01—will be low LD with a neighboring common variant, even

though they are in the same recombination range. However, most of the SNPs selected in the SNP arrays are common (MAF higher than 0.05), and therefore, GWASs have the power to detect association of variants that are relatively common in the population [21]. On the other hand, it is suggested that the observed association between a common SNP and a complex trait may result from LD of the SNP with rare variants at the same locus. Since common alleles and causal rare variants are correlated in a low LD, the hypothesis of a “synthetic association” implies that the magnitude of the effect of the causative variants is much greater than that of the common genotyped SNPs by the GWAS. For example, if an SNP explains 0.1% of the phenotypic variance in the population, the causal variant would account for 5–10% [23].

1.6. GWAS and CRC

Most of the studies to identify low penetrance alleles for CRC susceptibility were based on a candidate gene approach, whose role in CRC pathogenesis was supposedly known. However, without the real understanding of the biology of predisposition, the choice of genes was problematic. Thus, until the advent of GWAS, few or no association studies based on this approach were able to identify alleles of susceptibility unequivocally associated with the CRC risk [17].

To date, seven GWAS and three meta-analyses have identified about 20 independent loci associated with CRC susceptibility [24–34].

The number of common variants contributing with more than 1% of the inherited risk is very low, and it is very unlikely that there will be other SNPs with similar effects (greater than 1.2) for alleles with frequencies greater than 20% in European populations. In fact, the GWAS identified on average 80% of the common SNPs in this population but only 12% of SNPs with a minor allele frequency (MAF) between 5 and 10% [17].

However, variants with this profile, if taken collectively, can confer substantial risks due to their multiplicity, and in the case of CRC, to date, explain about 10% of heritability [33]. In a model built on data from the Scottish GWAS, about 170 common independent variants would explain all the genetic variance of the CRC [35]. Therefore, most of the genetic susceptibility to CRC still needs to be defined the so-called “missing heritability”. There are other possible causes of this unidentified portion: (1) the effect of rare variants; (2) failure to identify causal variants; and (3) allelic heterogeneity [36].

GWAS strategies to identify modest common risk alleles are not ideal for identifying rare variants (MAF below 1%) with potentially greater effects, as well as for capturing copy number variants and other structural variants, such as insertions, complex rearrangements, or expansions of microsatellite repeats, which may alter the risk of CRC. As efforts are made to scale up the GWAS meta-analyses, both in terms of sample size and coverage of SNPs, as well as to increase the number of SNPs considered for large-scale replication, it will be feasible to discover new variants. It is possible that a multiple loci approach based on haplotype markers identifies rare alleles. In addition, the use of exome sequencing may provide a more effective strategy for finding such variants [37].

2. Objectives

The overall objective of the present study was to replicate in individuals of the Brazilian population the 10 SNPs associated with CRC risk that are previously described in European populations. The specific objectives were to (1) calculate the allelic and genotype frequencies of the 10 SNPs in cases and controls; (2) analyze the association between the genotypes and alleles of the 10 SNPs and CRC risk; (3) calculate the magnitude of the effect on CRC risk; and (4) correlate the genotypes of the 10 SNPs with clinical-pathological characteristics and with familial history.

3. Research methods

3.1. Patient selection criteria

This is a retrospective study of case-control genetic association, whose sample comprised 727 cases and 740 controls, recruited from the Departments of Pelvic Surgery, Clinical Oncology, and Community Medicine at AC Camargo Cancer Center, in São Paulo, Brazil. All patients and controls authorized the present study by signing the informed consent form previously approved by the Research Ethics Committee of the institution under number 1231/09.

The inclusion criteria for cases were CRC diagnosis before age 75 years or with advanced colorectal adenoma (villous histology and/or greater than 1 cm and/or severe dysplasia) diagnosis before age 60 years, and controls were individuals without CRC who did not have first- or second-degree relatives with CRC. Controls were not matched with the cases in relation to the socioeconomic condition, ancestry, or self-referred ethnicity. The exclusion criteria were the presence of hereditary syndromes of predisposition to CRC, immunohistochemistry tests showing absence of proteins from DNA mismatch repair genes, the presence of high-penetrance germline mutations in susceptible genes to CRC, and appendix tumors and/or previous chronic inflammatory bowel disease.

3.2. SNP genotyping

After consent form was signed, 8 ml of peripheral blood was drawn to extract, purify, quantify, and amplify DNA. Quantitative PCRs for 10 tagSNPs (rs6983267, rs4939827, rs4779584, rs16892766, rs10795668, rs4444235, rs9929218, rs10411210, rs961253, and rs3802842) were performed using TaqMan® SNP Genotyping Assays and 7900HT Fast Real-Time PCR System (Thermo Scientific Fisher, CA, USA). Allelic discrimination analyses were made using the Sequence Detection Systems Automation Controller Software v2.3 and the auto-calling technique (Thermo Scientific Fisher, CA, USA).

As genotyping quality control, genotypes with the following criteria were excluded: (1) overall call <95%; (2) discordant duplicates; (3) Hardy-Weinberg imbalance ($p < 10^{-6}$ in cases and $p < 10^{-4}$ in controls); and (4) clusters of unsatisfactory genotypes in the X:Y axis inspection.

3.3. Statistics analysis

All tests were corrected for multiple analyses to avoid type I error. The allelic and genotypic frequencies of each SNP were calculated using the DeFinetti program [38], and the deviations of the genotype frequencies in cases and controls predicted by the Hardy-Weinberg equilibrium were calculated by the chi-square test with one degree of freedom or by Fisher's exact test, if the expected cell count was less than five.

Association analyses between the genotypes found in cases and controls for each SNP were performed with several types of genetic models, using the SNP and Variation Suite Version 7.6.10 program [39]. Multiple analyses were corrected by false discovery rate and Bonferroni.

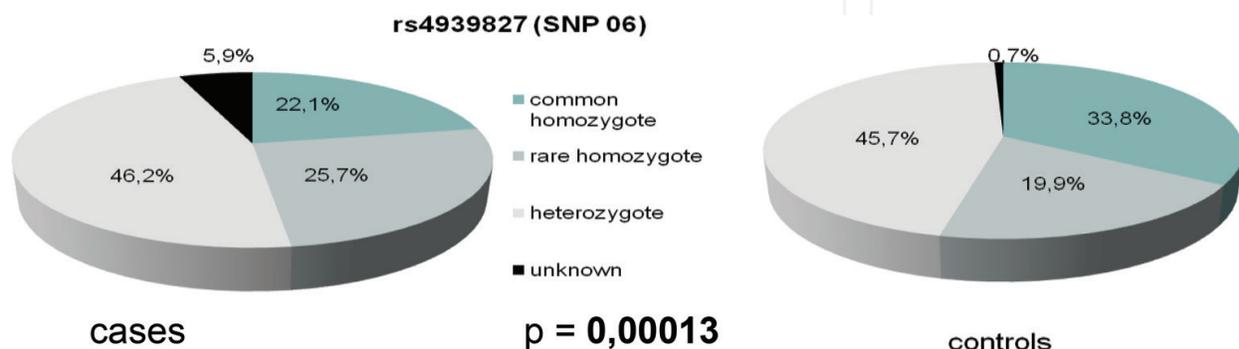
4. Results

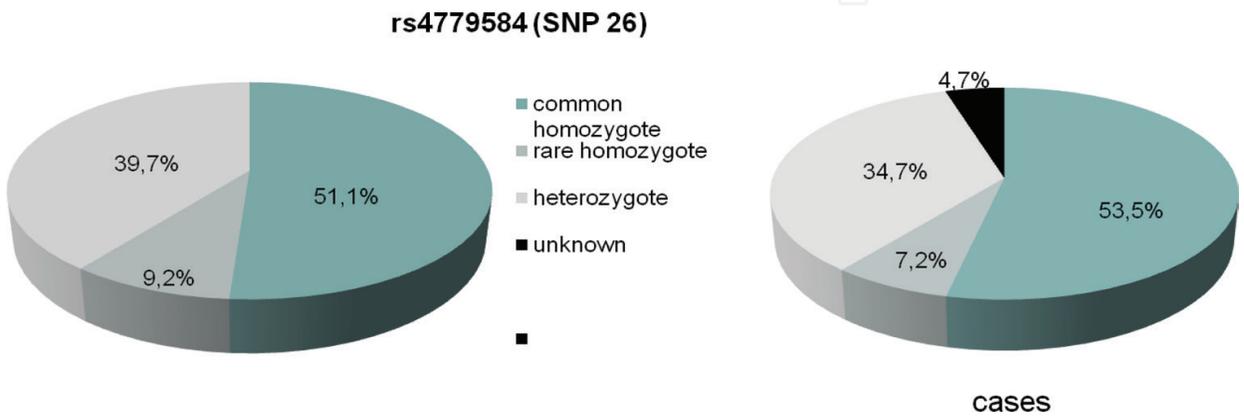
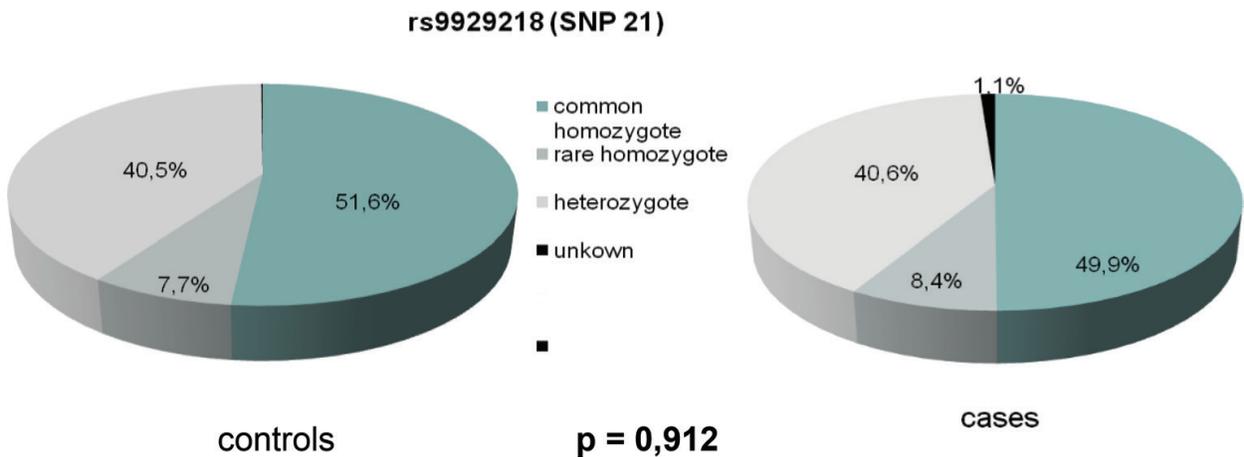
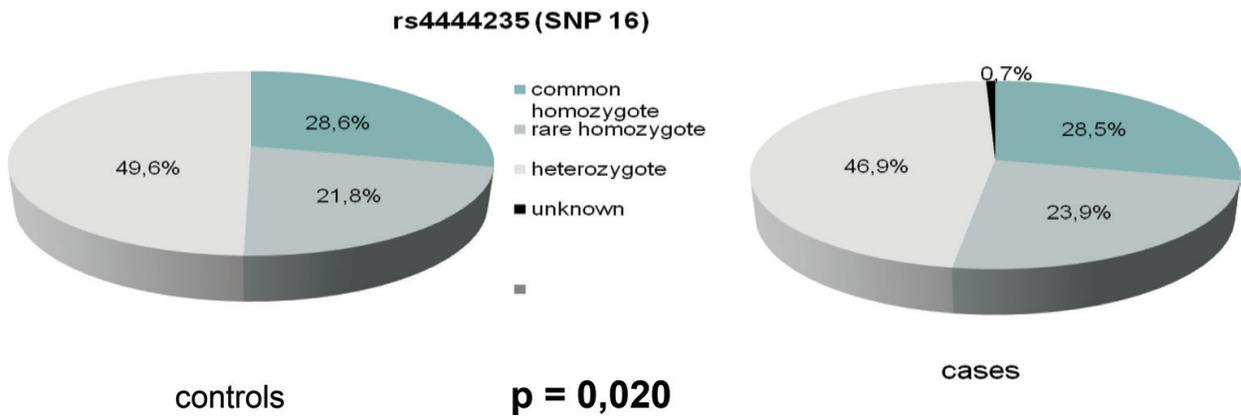
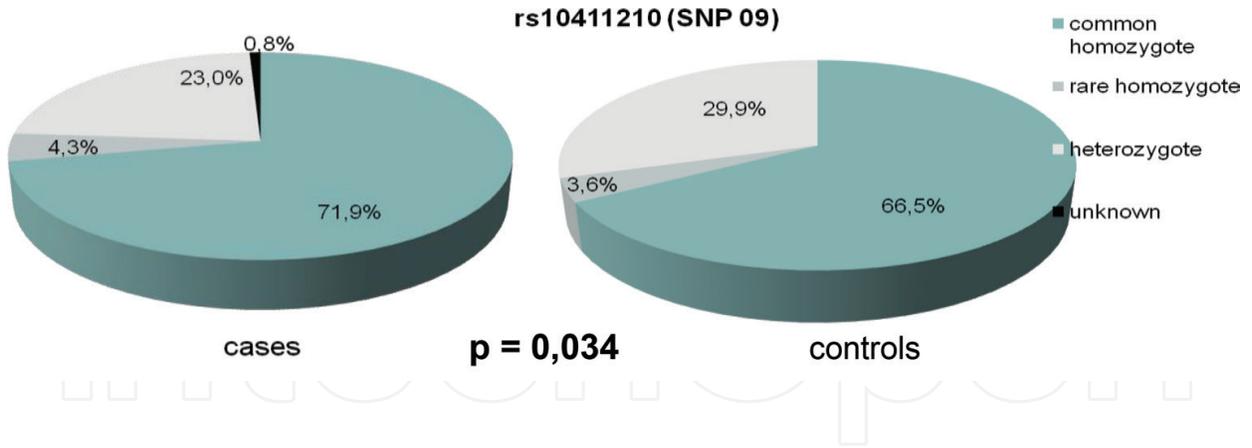
4.1. Clinical characteristics of cases and controls

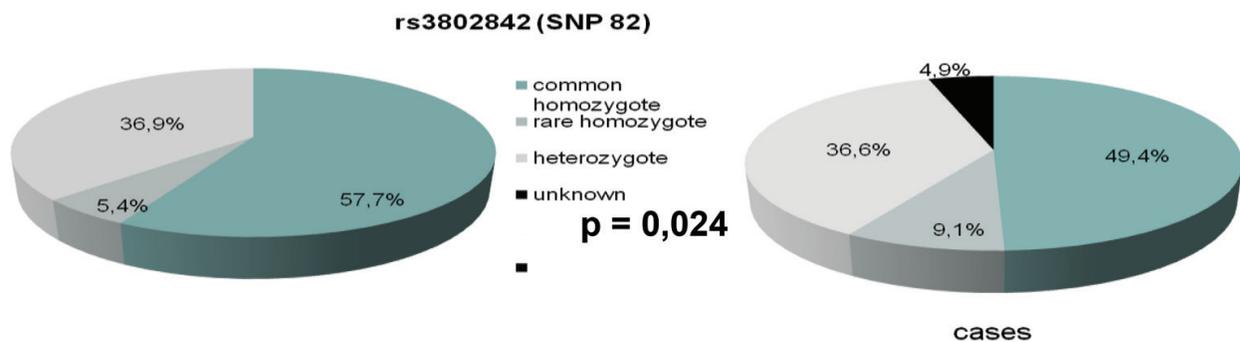
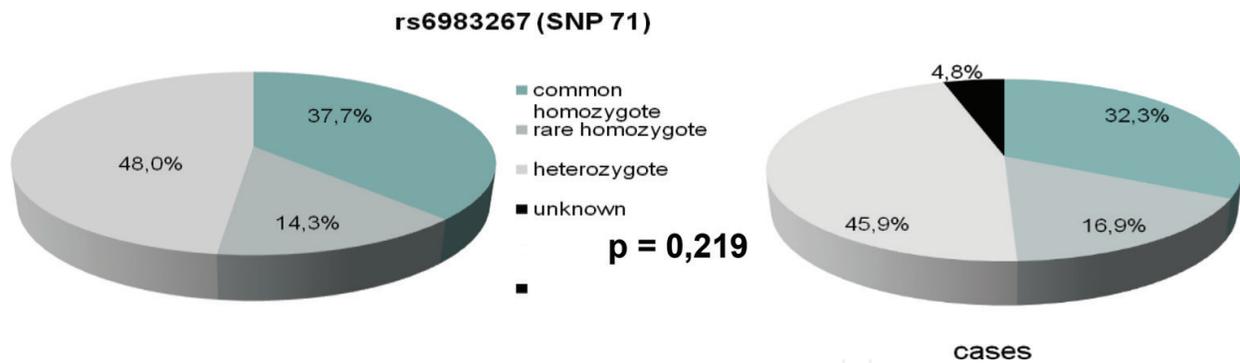
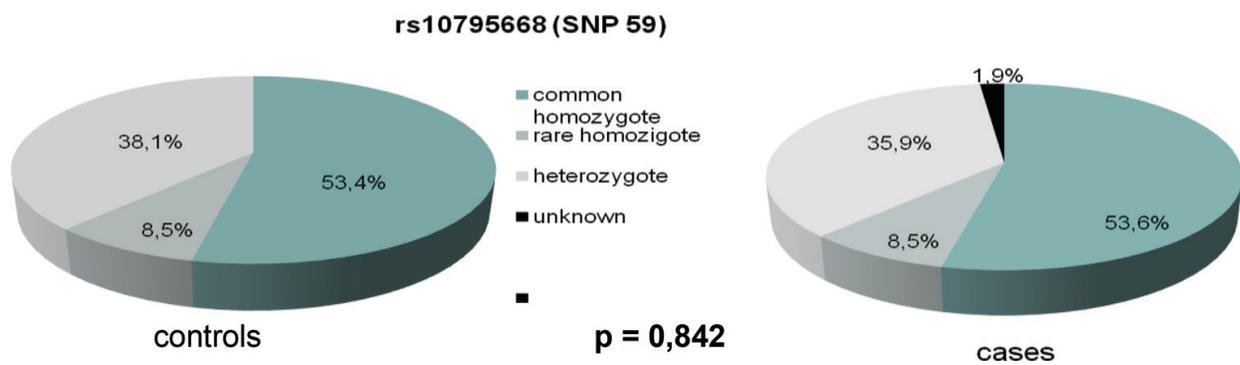
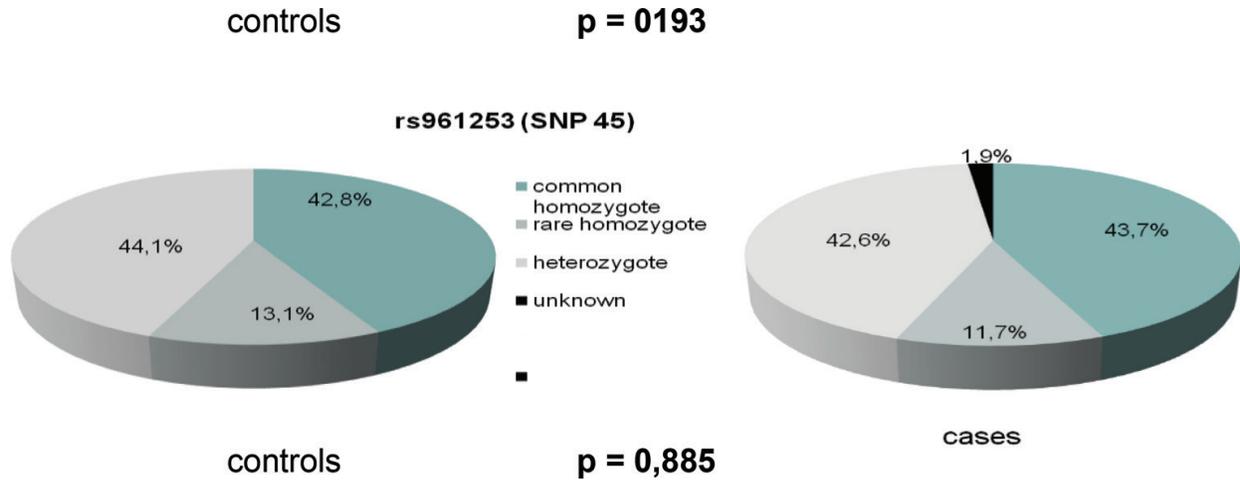
Of the 727 cases included in this study, 51% were male, with a median age of diagnosis of 56.9 ± 10.1 SD years old; 30% fulfilled Bethesda criteria; 3% of tumors were high-risk adenomas; the most common site of CRC was the rectum and in about 10%, there was an extra-colonic second primary tumor; tubular adenocarcinomas moderately differentiated at clinical stage III was mostly diagnosed. The majority of patients was alive disease-free until data collection and about 30% of cases had no familial history of CRC, despite almost 20% did not know about affected relatives. Of the 740 controls included in this study, 52% were female, with a median age of 51.9 ± 12.3 SD years old. Cases and controls were age and sex matched ($p = 0.126$ and 0.193 , respectively).

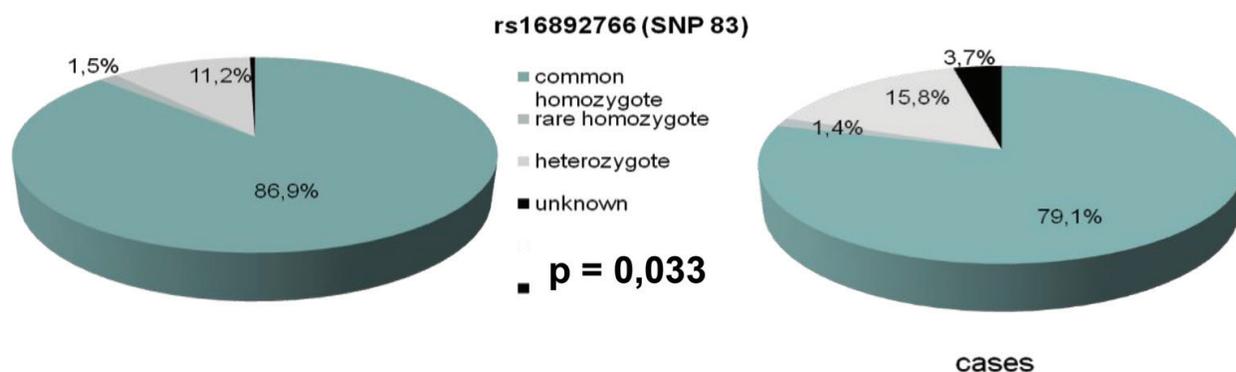
4.2. SNP genotyping and association tests

The genotypic frequencies of each SNP in cases and controls and their p-values are depicted in the following graphics:









The allelic frequencies, the number of alleles, the genotyping rate, and the Hardy-Weinberg equilibrium (EHW) test are represented in **Table 1**.

The genetic association tests and their genetic models are shown in **Table 2**.

Of the 10 SNPs, 5 (06, 09, 16, 82, and 83) were statistically significant ($p \leq 0.05$) associated with the risk of CRC and 2 (26 and 71) showed a trend to association ($p < 0.1$). SNP 06 showed the more significant association among all SNPs in all genetic models. SNP 09 was the only predictor of low risk, mainly in the dominant model (25% lower), whereas SNPs 16 and 82 were associated with high risk in the recessive model (45 and 85% higher, respectively). SNP 83 had higher risk principally in the dominant model (almost 50%). SNPs 26 and 71, on the other hand, obtained a marginally significant association only in the allelic and additive models, with a trend toward a higher risk with SNP 71 and lower risk with SNP 26. To sum up, of five SNPs associated with CRC risk, two (SNP 16 and 82) conferred higher risk among rare homozygotes than among heterozygotes and common homozygotes together through the recessive model and three (06, 09, and 83) showed higher risk among heterozygotes and rare homozygotes together than among common homozygotes through the dominant model.

Table 3 shows the five SNPs associated with CRC risk with their respective wild and variant alleles, their risk allele frequencies in comparison with the European population, the effect size of the variant allele and their populational attributable risks, which is the incidence decrease of disease if the population was not exposed to the risk allele.

5. Discussion

In common diseases, such as CRC, it is estimated that the most part of its genetic risk is due to the inherited multiple *loci* following polygenic model, each one with a common allelic frequency (MAF greater than 5%), whose effects show small sizes, between odds ratios 1.0 and 1.5. [17] Thus, to detect those small effects, it is necessary a big sample size. This strategy was validated by meta-analyses of GWAS data from European populations with tens of thousands genotyped individuals through high throughput platforms, followed by validation by multiple phases with independent series of cases and controls. Even though only about 20 common SNPs with modest effects were identified so far, each one with a p value corrected by multiple tests ($< 5.0 \times 10^{-8}$). In **Table 4**, GWAS data from European populations are compared to this study.

	EHW test		Allelic frequencies				Number of alleles				Frequencies		Genotyping rate		
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Global		
SNP	p	p	Major allele	Minor allele	Major allele	Minor allele	Major allele	Minor allele	Major allele	Minor allele	Major allele	Minor allele	%	%	%
rs4939827 (SNP06)	0.7018	0.0978	0.48	0.52	0.57	0.43	658	710	838	632	0.53	0.47	94.1	99.3	96.7
rs10411210 (SNP09)	0.0007	0.7158	0.84	0.16	0.81	0.19	1213	229	1205	275	0.83	0.17	99.2	100.0	99.6
rs4444235 (SNP16)	0.1565	0.9411	0.48	0.52	0.53	0.47	689	755	791	689	0.51	0.49	99.3	100.0	99.7
rs9929218 (SNP21)	0.9279	0.9273	0.71	0.29	0.72	0.28	1021	417	1064	414	0.72	0.28	98.9	99.9	99.4
rs4779584 (SNP26)	0.2319	0.3270	0.74	0.26	0.71	0.29	1030	356	1050	430	0.73	0.27	95.3	100.0	97.7
rs961253 (SNP45)	0.5022	0.3750	0.66	0.34	0.65	0.35	946	480	960	520	0.66	0.34	98.1	100.0	99.0
rs10795668 (SNP59)	0.0576	0.2311	0.73	0.27	0.72	0.28	1041	385	1072	408	0.73	0.27	98.1	100.0	99.0
rs6983267 (SNP71)	0.8151	0.7557	0.58	0.42	0.62	0.38	804	580	913	567	0.60	0.40	95.2	100.0	97.6
rs3802842 (SNP82)	0.1146	0.7613	0.71	0.29	0.76	0.24	984	398	1127	353	0.74	0.26	95.0	100.0	97.5
rs16892766 (SNP83)	0.1287	0.0005	0.9	0.10	0.93	0.07	1265	135	1369	105	0.92	0.08	96.3	99.6	98.0

Table 1. Allelic frequencies, number of alleles, genotyping rate and Hardy-Weinberg equilibrium (EHW) test.

Models SNP	Risk allele (minor)	RAF	Allelic OR [95% CI]	p	Dominant OR p [95% CI]	Recessive OR p [95% CI]	Additive OR p [95% CI]			
rs4939827 (SNP 06)	C	0.47	1.43 [1.23–1.66]	2.12×10^{-5}	1.67 [1.33–2.11]	0.0001	1.51 [1.18–1.93]	0.011	1.54 [1.20–1.98]	3.49×10^{-5}
rs10411210 (SNP 09)	T	0.17	0.83 [0.68–1.00]	0.080	0.75 [0.60–0.94]	0.040	1.19 [0.70–2.01]	0.749	0.71 [0.56–0.90]	0.083
rs4444235 (SNP16)	C	0.49	1.26 [1.09–1.45]	0.011	1.26 [1.00–1.60]	0.097	1.45 [1.14–1.83]	0.012	1.38 [1.07–1.78]	0.012
rs3802842 (SNP 21)	A	0.26	1.29 [1.09–1.53]	0.010	1.26 [1.02–1.56]	0.072	1.85 [1.23–2.78]	0.009	1.69 [1.1–2.60]	0.011
rs16892766 (SNP 26)	C	0.08	1.39 [1.07–1.82]	0.038	1.49 [1.11–1.99]	0.035	0.96 [0.40–2.27]	0.920	1.55 [1.14–2.10]	0.049
rs6983267 (SNP 45)	T	0.40	1.16 [1.00–1.35]	0.086	1.18 [0.95–1.46]	0.2	1.29 [0.97–1.72]	0.188	1.23 [0.91–1.66]	0.082
rs4779584 (SNP 59)	T	0.27	0.84 [0.72–1.00]	0.089	0.82 [0.66–1.00]	0.092	0.80 [0.55–1.17]	0.498	0.83 [0.67–1.04]	0.097
rs9929218 (SNP 71)	A	0.28	1.05 [0.89–1.23]	0.629	1.05 [0.85–1.29]	0.645	1.11 [0.76–1.62]	0.737	1.03 [0.83–1.28]	0.628
rs961253 (SNP 82)	A	0.34	0.94 [0.80–1.09]	0.515	0.93 [0.76–1.15]	0.623	0.90 [0.66–1.22]	0.824	0.95 [0.76–1.18]	0.524
rs10795668 (SNP 83)	A	0.27	0.97 [0.83–1.14]	0.739	0.95 [0.77–1.17]	0.682	1.02 [0.71–1.48]	1	0.94 [0.75–1.17]	0.739

RAF, risk allele frequency.

Table 2. Genetic association tests and genetic models.

SNP	Wild allele*	Risk allele	Risk allele frequencies(RAF)	Minor allele frequency (Europeans)*	Effect size	Populational attributable risk** (%)
rs4939827 (SNP 06)	<i>T (major)</i>	<i>C (minor)</i>	0.47	0.47	43% higher	16.8
rs10411210 (SNP 09)	<i>C (major)</i>	<i>T (minor)</i>	0.17	0.10	25% lower	–
rs4444235 (SNP 16)	<i>T (major)</i>	<i>C (minor)</i>	0.49	0.48	26% higher	11.3
rs3802842 (SNP 82)	<i>C (minor)</i>	<i>C (minor)</i>	0.26	0.27	29% higher	7.0
rs16892766 (SNP 83)	<i>A (major)</i>	<i>C (minor)</i>	0.08	0.09	39% higher	3.0

* From 1000 genomes.
 ** (RAF(OR – 1))/(1 + RAF(OR – 1)).

Table 3. The five SNPs associated with CCR risk.

SNP	Risk allele	RAF	Effect size (odds PAR (%) ratio)		Sample size (total)	References
rs4939827 (SNP 06)	<i>C (major)</i>	0.52	1.18	8.6	15,362	Broderick et al. [27]
rs10411210 (SNP 09)	<i>T (minor)</i>	0.10	0.87	–	41,259	Houlston et al. [31]
	<i>C (major)</i>	0.90	1.15	11.9		
rs4444235 (SNP 16)	<i>C (minor)</i>	0.46	1.11	4.8	41,259	Houlston et al. [31]
rs9929218 (SNP 21)	<i>G (major)</i>	0.71	1.10	6.6	41,259	Houlston et al. [31]
rs4779584 (SNP 26)	<i>T (minor)</i>	0.18	1.26	4.5	14,663	Jaeger et al. [28]
rs961253 (SNP 45)	<i>A (minor)</i>	0.36	1.12	4.0	41,259	Houlston et al. [31]
rs10795668 (SNP 59)	<i>G (major)</i>	0.67	1.12	7.4	37,371	Tomlinson et al. [30]
rs6983267 (SNP 71)	<i>G (major)</i>	0.51	1.21	9.7	14,470	Tomlinson et al. [25]
rs3802842 (SNP 82)	<i>C (minor)</i>	0.29	1.12	3.4	27,794	Tenesa et al. [29]
rs16892766 (SNP 83)	<i>C (minor)</i>	0.07	1.25	1.7	37,371	Tomlinson et al. [30]

RAF, risk allele frequency; PAR, populational attributable risk.

Table 4. GWAS data from European populations compared to this data.

In this study, there was an association with CRC risk in half of SNPs (06, 09 16, 82 and 83), whose risk alleles revealed similar frequencies as to European GWAS, except SNP 06. Effect sizes were modest as well as European GWAS. SNP 06 was the variant that resulted the greatest effect with the most statistically significant association ($p_{\text{trend}} = 3.49 \times 10^{-5}$), which conferred the highest populational risk, whereas in European GWAS, the risk augmented up to 23%, representing 8.6% of the populational risk [35]. In the original study, the same SNP also showed the greatest association ($p_{\text{trend}} = 1.0 \times 10^{-12}$) [27]. SNP 09 (rs10411210) was associated with a low risk in a dose-dependent way [31]. In this study, however, this effect was detected only in the dominant model. It is noteworthy that in European studies the *major* allele (C) confers 15% higher risk, responsible for 12% of the populational risk [35]. In the present study, there was a trend to a higher risk but not statistically significant ($p = 0.08$). Moreover, SNP 16 was associated with a higher risk in this study than from European GWAS, as well as the populational attributable risk [35]. Likewise, SNPs 82 and 83 also increased more the CRC risk and populational risk in the present study than from European GWAS [35].

In the present study, the populational stratification by ancestry was not investigated. However, the Brazilian population, although greatly admixed, has a high prevalence of individuals from European ancestry, whose the great majority is located in the South (79.5%) and Southeast (74.2%) [40].

6. Conclusion

This study partially replicated European GWAS in Brazilian Southeastern population with a predominantly European genetic background. Small sample size and lack of stratification by ancestry are prone to type I and II errors, respectively. Further studies in admixed populations would certainly aid to uncover the missing heritability of CRC and help to build the genetic architecture of CRC susceptibility.

Author details

Israel Gomy^{1*}, Ligia Petrolini de Oliveira² and Benedito Mauro Rossi³

*Address all correspondence to: isgomy@gmail.com

1 Universidade Federal do Paraná, Curitiba, Brazil

2 Universidade Federal do ABC, Santo André, Brazil

3 Hospital Sírio Libanês, São Paulo, Brazil

References

- [1] Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer*. 2010;**127**: 2893-2917. DOI: 10.1002/ijc.25516

- [2] Edwards BK, Ward E, Kohler BA, et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer*. 2010;**116**:544-573. DOI: 10.1002/cncr.24760
- [3] Center MM, Jemal A, Ward E. International trends in colorectal cancer incidence rates. *Cancer Epidemiology Biomarkers & Prevention*. 2009;**18**:1688-1694. DOI: 10.1158/1055-9965.EPI-09-0090
- [4] Ministério da Saúde. Estimativa 2016: Incidência de Câncer no Brasil. Instituto Nacional de Câncer José Alencar Gomes da Silva, Coordenação de Prevenção e Vigilância. Rio de Janeiro: INCA; 2015
- [5] Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: A multicentre randomised controlled trial. *Lancet*. 2010;**375**:1624-1633. DOI: 10.1016/S0140-6736(10)60551-X
- [6] Schoen RE, Pinsky PF, Weissfeld JL, et al. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. *The New England Journal of Medicine*. 2012;**366**:2345-2357. DOI: 10.1056/NEJMoa1114635
- [7] de Jong AE, Hendriks YM, Kleibeuker JH, et al. Decrease in mortality in Lynch syndrome families because of surveillance. *Gastroenterology*. 2006;**130**:665-671. DOI: 10.1053/j.gastro.2005.11.032
- [8] Levin B, Lieberman DA, McFarland B, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: A joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Gastroenterology*. 2008;**134**:1570-1595. DOI: 10.1053/j.gastro.2008.02.002
- [9] Vogelstein B, Fearon ER, Hamilton SR, et al. Genetic alterations during colorectal-tumor development. *The New England Journal of Medicine*. 1988;**319**:525-532. DOI: 10.1056/NEJM198809013190901
- [10] Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature*. 1998;**396**:643-649. DOI: 10.1038/25292
- [11] Cunningham JM, Christensen ER, Tester DJ, et al. Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. *Cancer Research*. 1998;**58**:3455-3460
- [12] Noffsinger AE. Serrated polyps and colorectal cancer: New pathway to malignancy. *Annual Review of Pathology*. 2009;**4**:343-364. DOI: 10.1146/annurev.pathol.4.110807.092317
- [13] Cunningham D, Atkin W, Lenz HJ, et al. Colorectal cancer. *Lancet*. 2010;**375**:1030-1047. DOI: 10.1016/S0140-6736(10)60353-4
- [14] Cannon-Albright LA, Skolnick MH, Bishop DT, Lee RG, Burt RW. Common inheritance of susceptibility to colonic adenomatous polyps and associated colorectal cancers. *The New England Journal of Medicine*. 1988;**319**:533-537. DOI: 10.1056/NEJM198809013190902

- [15] Aaltonen L, Johns L, Jarvinen H, Mecklin JP, Houlston R. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. *Clinical Cancer Research*. 2007;**13**:356-361. DOI: 10.1158/1078-0432.CCR-06-1256
- [16] Aaltonen LA, Salovaara R, Kristo P, et al. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *The New England Journal of Medicine*. 1998;**338**:1481-1487. DOI: 10.1056/NEJM199805213382101
- [17] Tomlinson IP, Dunlop M, Campbell H, et al. COGENT (COlorectal cancer GENEtics): An international consortium to study the role of polymorphic variation on the risk of colorectal cancer. *British Journal of Cancer*. 2010;**102**:447-454. DOI: 10.1038/sj.bjc.6605338
- [18] St John DJ, McDermott FT, Hopper JL, Debney EA, Johnson WR, Hughes ES. Cancer risk in relatives of patients with common colorectal cancer. *Annals of Internal Medicine*. 1993;**118**:785-790. DOI: 10.7326/0003-4819-118-10-199305150-00005
- [19] Sachidanandam R, Weissman D, Schmidt SC, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. 2001;**409**:928-933. DOI: 10.1038/35057149
- [20] Strachan T, Read A. Human genetic variability and its consequences. In *Human Molecular Genetics*. 4th ed. New York: Garland Science; 2011. pp. 405-440
- [21] Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *American Journal of Human Genetics*. 2012;**90**:7-24. DOI: 10.1016/j.ajhg.2011.11.029
- [22] International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005;**437**:1299-1320. DOI: 10.1038/nature04226
- [23] Dickson SP, Wang K, Krantz I, et al. Rare variants create synthetic genome-wide associations. *PLoS Biology*. 2010;**8**:e1000294. DOI: 10.1371/journal.pbio.1000294
- [24] Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nature Genetics*. 2007;**39**:989-994. DOI: 10.1038/ng2089
- [25] 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. *Nature Genetics*. 2007;**39**:984-988. DOI: 10.1038/ng2085
- [26] Haiman CA, Le Marchand L, Yamamoto J, et al. A common genetic risk factor for colorectal and prostate cancer. *Nature Genetics*. 2007;**39**:954-956. DOI: 10.1038/ng2098
- [27] Broderick P, Carvajal-Carmona L, Pittman AM, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nature Genetics*. 2007;**39**:1315-1317. DOI: 10.1038/ng.2007.18
- [28] Jaeger E, Webb E, Howarth K, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nature Genetics*. 2008;**40**:26-28. DOI: 10.1038/ng.2007.41

- [29] Tenesa A, Farrington SM, Prendergast JG, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nature Genetics*. 2008;**40**:631-637. DOI: 10.1038/ng.133
- [30] Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nature Genetics*. 2008;**40**:623-630. DOI: 10.1038/ng.111
- [31] Houlston RS, Webb E, Broderick P, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nature Genetics*. 2008;**40**:1426-1435. DOI: 10.1038/ng.262
- [32] Houlston RS, Cheadle J, Dobbins SE, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nature Genetics*. 2010;**42**:973-977. DOI: 10.1038/ng.670
- [33] Dunlop MG, Dobbins SE, Farrington SM, et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nature Genetics*. 2012;**44**:770-776. DOI: 10.1038/ng.2293
- [34] Peters U, Hutter CM, Hsu L, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Human Genetics*. 2012;**131**:217-234. DOI: 10.1007/s00439-011-1055-0
- [35] Tenesa A, Dunlop MG. New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nature Reviews Genetics*. 2009;**10**:353-358. DOI: 10.1038/nrg2574
- [36] Ehret GB, Lamparter D, Hoggart CJ, et al. A Multi-SNP locus-association method reveals a substantial fraction of the missing heritability. *American Journal of Human Genetics*. 2012;**91**:863-871. DOI: 10.1016/j.ajhg.2012.09.013
- [37] Fletcher O, Houlston RS. Architecture of inherited susceptibility to common cancer. *Nature Reviews Cancer*. 2010;**10**:353-361. DOI: 10.1038/nrc2840
- [38] DeFinetti program [Internet]. Available from: <http://ihg2.helmholtz-muenchen.de/cgi-bin/hw/hwa2.pl> [Accessed: 30 January 2013]
- [39] SNP & Variation Suite™ (Version 7.6.10) [Software]. Bozeman, MT: Golden Helix, Inc. Available from: <http://www.goldenhelix.com> [Accessed: 30 January 2013]
- [40] Pena SD, Di Pietro G, Fuchshuber-Moraes M, et al. The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One*. 2011;**6**:e17063. DOI: 10.1371/journal.pone.0017063

