Genetic ancestry is associated with colorectal adenomas and adenocarcinomas in Latino populations

Title: Genetic ancestry is associated with colorectal adenomas and adenocarcinomas in Latino populations

Running title: Genetic susceptibility of colorectal cancer in Latinos

Gustavo Hernandez-Suarez[1,2,*], MariaCarolina Sanabria[1,4], Marta Serrano[1,4], Oscar F. Herran[5], Jesus Perez[6], Jose L. Plata[7], Jovanny Zabaleta[3] Albert Tenesa*[2,8]

1. Grupo de Investigacion Epidemiologica, Instituto Nacional de Cancerologia de Colombia, Bogota, Colombia.

2. The Roslin Institute, The University of Edinburgh, Edinburgh, UK,

3. Department of Pediatrics and Stanley S. Scott Cancer Center, Louisiana State University Health Sciences Center, New Orleans, USA,

4. Universidad Nacional de Colombia,

5. Universidad Industrial de Santander, Bucaramanga, Colombia,

6. Universidad Libre, Barranquilla,Colombia,

7. Fundacion Oftalmologica de Santander, Bucaramanga, Colombia

8. MRC-Human Genetics Unit at the MRC IGMM, University of Edinburgh, Edinburgh UK

* To whom correspondence should be addressed.

Instituto Nacional de Cancerologia. Av 1 No.9-85, Bogota, Colombia Tel: +57 13341360; Fax: +57 13341360; Email: gustasuarez@gmail.com

The Roslin Institute and Royal (Dick) School of Veterinary Studies. University of Edinburgh. Easter Bush, Midlothian EH25 9RG. Tel: +44 (0)131 651 9226 Email: albert.tenesa@ed.ac.uk

Please cite this article as: Hernandez-Suarez, G, Sanabria, MC, Serrano, M, Herran, OF, Perez, J, Plata, JL, Zabaleta, J & Tenesa, A 2014, 'Genetic ancestry is associated with colorectal adenomas and adenocarcinomas in Latino populations' European Journal of Human Genetics, vol E-pub 12 February., 10.1038/ejhg.2013.310

Abstract

Colorectal cancer rates in Latin American countries are less than half of those observed in the United States. Latin-Americans are the resultant of generations of admixture of Native American, European, and African individuals. The potential role of genetic admixture in colorectal carcinogenesis has not been examined. We evaluate the association of genetic ancestry with colorectal neoplasms in 190 adenocarcinomas, 113 sporadic adenomas and 243 age-sex matched controls enrolled in a multicentric case-control study in Colombia. Individual ancestral genetic fractions were estimated using the STRUCTURE software, based on allele frequencies and assuming three distinct population origins. We used the Illumina Cancer Panel to genotype 1,421 sparse SNPs, and CEU, LWJ and CHB populations from the HapMap project as references. A total of 678 autosomal SNPs overlapped with the HapMap datasets SNPs and were used for ancestry estimations. African mean ancestry fraction was higher in adenomas (0.13, CI95% 0.11-0.15) and cancer cases (0.14, CI95% 0.12-0.16) compared to controls (0.11, CI95% 0.10-0.12). Conditional logistic regression analysis, controlling for known risk factors, showed a positive association of African ancestry per 10% increase with both colorectal adenoma (Odds Ratio: 1.12 CI95% 0.97-1.30) and adenocarcinoma (Odds Ratio: 1.19 CI95% 1.05-1.35). In conclusion, increased African ancestry (or variants linked to it) contributes to the increased susceptibility of colorectal cancer in admixed Latin American population.

Keywords: colorectal neoplasm, Pedigree, Polymorphism, Genetic, Latin America, Colombia

Introduction

Colorectal cancer incidence in Latin-America has steadily increased during the last decades[1], however these rates are less than half of those observed in African Americans and Caucasian Americans in North America[2]. Colorectal cancer incidence rates largely vary across continents, showing the highest rates in those countries with mainly Caucasian populations[1]. Such differences, and even the recent rising trends observed in developing countries, have been attributed to the high and increasing prevalence of risk factors associated with a "westernized" lifestyle, such as obesity and physical inactivity[3]. The reasons behind the higher colorectal cancer risk among African Americans compared to non-Hispanic Whites (Caucasians) in the United States are still not clear. Several studies propose health disparities as the main reason behind these differences[4,5]. However, prospective evidence showed that such factors explain roughly 27% of the excess risk in, African Americans relative to Whites[6], suggests that variation in genetic susceptibility across populations may play an important role[7,8].

The observed increase in colorectal cancer incidence has possibly been accompanied, if not led, by an increase in the incidence of colorectal adenomas[9,10]. Adenomas are the main precursor lesions to most sporadic colorectal cancers and develop through the complex interactions of environmental and genetic risk factors[11]. Although recent reports have suggested that the risk of colorectal adenoma may be influenced by racial differences[12,13] in admixed populations, these findings are based on reported ethnicity rather than measured genetic ancestry. Self-reported measures of ethnicity in admixed populations are notoriously inaccurate regarding the individual ancestry[14]. This is especially important in populations of Latin-America where the admixture between people of at least three continents (Africa, Europe and America) has been widespread since the 17th century.

In this study we used a sparse set of Single Nucleotide Polymorphisms (SNPs) to evaluate the association of genetic ancestry with the risk of colorectal adenomas and adenocarcinomas in the Colombian population controlling for well known colorectal cancer and adenoma risk factors.

Material and methods

Study population and enrolment

Cases and controls were randomly extracted from a larger multi-center case-control study aimed at identifying the environmental and genetic risk factors of colorectal cancer in Colombia. After ethical approval from the ethics Board of The National Cancer Institute of Colombia, we recruited incident cases (diagnosed at enrolment) of colorectal adenoma and adenocarcinomas at major colonoscopy medical centers in six of the largest Colombian cities (Barranquilla, Bogota, Bucaramanga, Cali, Cartagena and Santa Marta) from January 2008 to February 2011. Colombia has not yet established a colorectal cancer screening program; therefore, most of the colonoscopy examinations were medically indicated. Cases were originally diagnosed after a complete and satisfactory colonoscopy examination, but only pathologically confirmed cases were finally enrolled in the study. Eligible cases were Colombians, residents in the city of enrollment, aged between 30 and 75 years at the time of colonoscopy, willing and mentally capable to participate, and did not have any personal history of colorectal cancer, ulcerative colitis or Crohn's disease. Controls were approached at the waiting room of primary care units, nearby or in the same hospital where the cases were recruited among individuals attending for medical conditions different from gastrointestinal discomfort and willing to participate; they were unrelated to cases and had no personal history of cancer or colorectal adenomas. Controls were matched by sex and age group (± 5 years) to the cases.

Participants gave written informed consent, donated a blood sample and answered a full epidemiological survey, a food frequency questionnaire (FFQ) designed for the study[15] and the short version of the IPAQ[16] (International Physic Activity Questionnaire), looking into the best known risk factors for colorectal adenoma and adenocarcinoma [17]. Buffy coats were kept in portable liquid nitrogen containers until transferred in dry ice to the National Cancer Institute Facilities in Bogotá for final storage at -80°C. Questionnaires were processed centrally. We used Teleform™, version 5.2 software package (Cardiff Software, Inc.) to increase the efficiency of data management and reduce typing error. By the end of the recruitment phase, we enrolled 506 controls, 322 adenocarcinomas and 239 colorectal adenomas. Due to funding constraints, we restricted our genetic analyses to a random subset sample of 264 controls, 206 adenocarcinomas and 126 adenomas. Adenomas included into the analysis were large (≥1 cm) without severe dysplasia, and with histopathological diagnosis of tubular, tubulovillous or villous adenoma (<20%, 20-80% and ≥80% of villous component, respectively).

SNP Genotyping

DNA was extracted from buffy coat samples, using the QIAamp DNA Blood Mini KIT (QIAGEN®), as recommended by the manufacturer and eluted in 100µl of Nuclease-free Water (Ambion®). Two hundred and fifty nanograms (250ng) of DNA were resuspended in 5µl of TE Buffer, denatured and bound to paramagnetic beads for high-throughput genotyping using the protocols described for the highly multiplexed GoldenGate assay[18] (Illumina Inc.). Briefly, two allele-specific oligonucleotide (ASO) probes, linked to universal primer sequences (labeled with either Cy3 or Cy5 for each allele), along with one locus-specific oligonucleotide (LSO) probe, also linked to a universal primer and an address sequence, are hybridized to the DNA. Extension of the ASO and ligation to the LSO is carried out and the product is amplified by PCR. The amplified product was hybridized to the

chips containing sequences complementary to each unique address sequence and the alleles were determined by the scanner according to the fluorescent emitted (Cy3, Cy5, or both). The SNP panel used for this study (Illumina Cancer Panel®) consist of 1,421 thoroughly screened and validated SNP loci, covering all chromosomes and tagging 408 genes chosen from the National Cancer Institute's (NCI) Cancer Genome Anatomy Project SNP500 Cancer Database[19]. According to the manufacturer, the mean Minor Allele Frequency (MAF) across all the SNPs in the genotyping panel was 0.25, 0.22 and 0.21 for Caucasians, Han Chinese/ Japanese and Yoruba Africans respectively.

Quality Control

We followed a standard quality control (QC) protocol for case-control genetic association studies[20] using the PLINK *software*[21]. SNPs were excluded from the analysis if they departed from Hardy-Weinberg equilibrium (HWE; $p<0.01$), there was a significant difference between missing genotype rates among cases and controls ($p <0.01$), the SNP overall call rate was <0.95, or the MAF was <0.04. Participants with call rates ≤0.95, or with heterozygosity rates >3 standard deviations from the sample mean, were also excluded. In addition, we excluded one individual of each pair featuring an Identity by Descendant value (IBD) >0.375 from the analysis, avoiding duplicated, related or contaminated samples. Gender could not be reliably estimated from the limited number of SNPs available on the X chromosome (N =13), and we relayed on our recorded gender. Eighteen percent of the controls (n=21), 10% of the adenomas (n=13), 7.7% of adenocarcinomas (n=16), and 15.8% of the SNPs (n=225) did not pass through the quality control, leaving 238 controls,113 adenomas and 190 adenocarcinomas for the analysis.

Inference of Ancestry Proportions

We used the Bayesian clustering algorithm STRUCTURE version 2.2[22] under an admixture model to estimate the proportions of European, African and Amerindian ancestry in each of our samples. We used a flat prior to ran a burn-in period of 5000 iterations and kept 1 in 5000 iterations. Under the admixture model, the genotype information of each individual is modeled assuming that they inherited a fraction of their genome from ancestors originating from one of the k^{th} populations of origin. We included a set of overlapping SNPs (N= 804) genotyped in three reference ancestral populations (k=3) from the HapMap3 project[23]: Utah residents with Northern and Western European ancestry (CEU), Luhya in Webuye, Kenya (LWK) and Han Chinese in Beijing, China (CHB). The well-established similarities of the allele frequencies between the latter population and Amerindians[24], made it a useful alternative to discriminate the Amerindian from African and European ancestry in the study sample. Six hundred seventy eight (678) autosomal SNPs remained for analysis after pruning for Linkage Disequilibrium (LD), excluding one of each pair with R^2 higher than 0.5, in a windows size of 50 SNPs and a window shift of 5 SNPs.

To verify the admixture estimations using the selected set of SNPs, we also estimated the ancestry fractions of individuals from two admixed populations included in the HapMap3 database: Mexican ancestry from Los Angeles (MEX) and African ancestry in Southwest USA (ASW). Finally, we calculated the informativeness for assignment measure (In) proposed by Rosenberg et al.[25] to estimate the ancestral information that each SNP included provides.

Statistical Analysis

We compared the mean ancestry fractions between cases (adenomas and carcinomas) and controls using one way ANOVA tests. To evaluate the association of genetic ancestry with adenoma and cancer separately, we used binary conditional logistic regression models controlling for potential confounding factors. Because ancestry fractions are dependant from each other they can not be handled as independent variables. To overcome this limitation, without leaving out from the analysis any of the ancestry fractions we include into the model two parameters: the arithmetic difference between European and Amerindian ancestry fractions (main genetic substitution in Latin American populations) and the estimated African ancestry (log transformed). The latter was log transformed as it was positively skewed in the study population (Figure 1). These two parameters were fitted alternatively as raw continuous (to evaluate the linear trend) and categorical variable to measure the variation in risk per 10% increase of African ancestry (from 0.01 to ≥0.30) and European replacement increase (from ≤-0.30 to ≥ 0.30). We avoid the pairwise comparison of the resulting categories in the logistic regression analysis, but we report their distribution for descriptive purposes.[26] The multivariate analysis, which had city of enrollment as conditional variable, included: gender, age, attained education level (none, elementary school, secondary school, technical studies (i.e. college), University or higher), family history of colorectal cancer in first-degree relatives, history of alcohol intake (no intake, <12.50 and ≥12.50 g/day); cigarette smoking (<0.5, 0.5-0.9 and ≥1 packs/year); red meat consumption (< 2 , 2-4 and ≥5 servings per week); physical activity (<10, 10-19, ≥20 hours per week), non-steroid anti-inflammatory drugs (at least one per week during the last 6 months, yes or no), dietary fiber and total energy intake (quartiles regarding the distribution among controls).

Results

Cases were slightly older than controls. While adenomas where positively associated with higher attained education (p=0.02), adenocancinomas showed the opposite, being positively

associated with lower educational level instead.($p<0.01$) Cancer cases also showed an inverse association with BMI at diagnosis, probably due to reverse causation (Table 1). No other differences were observed among the risk factors included into the analysis.

Among adenomas cases, European (mean 0.44 CI95% 0.42-0.46) and African (mean 0.13 CI95% 0.11-0.15) ancestry proportion were higher compared to controls (European mean 0.39 CI95% 0.38-0.41; African mean 0.11 CI95% 0.10-0.12), while the Amerindian mean proportion behave just inversely proportional to the European. In contrast, cancer cases showed a higher mean proportion of African ancestry compared to controls (0.14 vs 0.11 $p<0.01$) (Table 2). When comparing the categorical distribution of African ancestry and European ancestry substitution (European minus Amerindian ancestry fractuions) in the study population we observed similar results: an association of African ancestry with both adenomas ($p=0.07$) and cancer ($p=0.02$), while the European genetic substitution only was associated with adenoma ($p=0.001$) but not with cancer cases ($p=0.95$) (Table 3). Ancestry fractions estimated for the MEX and ASW population were very similar to those reported in the literature[27] despite the low In values featured by the SNP included in our analysis (max= 0.34, mean=0.03, SD=0.04), reassuring the reliability of the ancestry estimations.

After controlling for confounding, conditional logistic regression analysis results were consistent with the crude ones, showing a positive marginal association of increasing African ancestry with colorectal adenomas (risk variation per 10% increase [OR] 1.122 CI 95% [0.97, 1.30] p for linear trend=0.08) and statistically significant with adenocarcinomas (risk variation per 10% increase [OR] :1.19 CI95% [1.05 , 1.35], p for trend 0.003] (Table 4). In contrast, increasing European ancestry was positively associated only to adenoma (risk variation per 10% increase [OR] 1.25 CI95% [1.08,1.46] CI95% p for trend 0.001) but not to cancer (risk variation per 10% increase [OR] 1.02 CI95%[0.90- 1.16], p for trend 0.75). In addition, adenoma was associated with university or higher education compared to primary

school (the most prevalent category) (OR 3.81, 95%CI1.59 – 9.17), while colorectal cancer risk was marginally associated with no education attained (OR 2.5 95%CI 0.89 – 7.36). Adjusting by age and sex did not modify the results-significantly (Table 4). There was no evidence of heterogeneity in the mean differences of African ancestry when comparing adenomas or adenocarcinomas to controls across education strata (Figure 2). When exploring the ancestry association, stratified by distal and proximal colorectal neoplasms we did not observed any differences from the overall results (results not shown).

Discussion

To the best of our knowledge this is the first report on the association of genetic ancestry and sporadic colorectal adenomas and adenocarcinomas in an admixed population. Our findings add evidence to the hypothesis that genetic ancestry influences cancer risk in Latino populations. A similar positive association of genetic ancestry has been reported previously between European ancestry and breast cancer in the Mexican population[28]; In a clinical practice scenario genetic ancestry fractions has being proposed as a genome-wide biomarker useful to evaluate relapse in children undergoing therapy for acute lymphoblastic leukemia[29].

The association of African ancestry not only with adenocarcinoma but also marginally with adenoma supports our hypothesis of the role of genetic ancestry in early stages of colorectal carcinogenesis and may rely on differences in allele frequencies in polymorphism related to colorectal cancer risk[30]. We found that this association was not confounded by well-known risk factors. Education was chosen as proxy of socioeconomic status (SES) as it is attained early in life and does not change greatly after the third decade of life. SES has previously been associated independently to colorectal cancer worldwide and to genetic ancestry in Latin America. The nature of the association between SES and colorectal cancer

risk is discrepant across continents[31]. While studies in Europe, East Asia and Australia, in general, have found a positive association, in the US and Canada the association observed is inverse[31,32]. This discrepancy is not fully understood but it is partially explained by differences in screening coverage[31] and the way environmental factors (mediators) are interrelated with SES (determinants)[32,33]. Our results are contrary to a previous study reporting a positive association of colorectal cancer and SES in Colombia[34]. Here we found an inverse association of education level with adenocarcinoma., but also a positive association on higher education level with adenoma (Table 4). This finding may explain partially the association of European ancestry only with adenomas; more than a third of adenomas cases showed European ancestry fraction within its higher category (Table 3). A previous report has shown the positive association on European ancestry with higher education among Latinos. [35]

In contrast, the positive association of African ancestry to both adenoma and adenocarcinoma reported here is hard to explain due to differences in SES given that in our study: first, African Ancestry was not associated with education level (p for trend = 0.76) (Figure 3) and second, adenomas and adenocarcinomas showed opposite association to education level. It is worth mentioning that Afro-Colombians are an ethnic minority with large disparities compared with the overall population. In our study population the African ancestry was not high (interquartile range 0.6- 0.24) as African-Americans (80%) and its likely that within this range most of the individuals did not identify themselves as being of African descent.

Our study features several strengths. It includes both preneoplasic and neoplasic lesions confirmed by histopathology allowing us to evaluate the association of ancestry in the early events of colorectal carcinogenesis. We sampled cases and controls from the same population. As cases are mostly referred by general physicians this assures a better control for

selection bias. Our ancestry estimations are reliable as those estimated for the ASW and MEX individuals were similar to those published in the literature[27]. The differences in the anatomic distribution of adenomatous polyps and cancer showing a shift to the left for cancer cases was observed as previously described[36] and the similar age range in adenomas and adenocarcinomas does not suggest a selection bias[37].

There are some limitations of our study that should be considered when interpreting the results. It is likely that educational levels do not reflect, nor control, the entire variability of the socioeconomic status, thus residual confounding may exist. Nevertheless, here we assessed and included into the analysis the most relevant nutritional and lifestyle factors that may mediate the association of SES with colorectal adenoma and adenocarcinoma. There could be a differential access to colonoscopy, where wealthy people may have better access to these procedures. However, the chance of under representation of individuals with a higher level of education in the control group is not likely, as recent census data[38] in Colombia showed that only 9% of the population have university or a higher education degree, a similar value observed in our control sample (11%). Our sample size provides limited statistical power to detect small effects, especially regarding the observed effect in the association between African ancestry and adenomas. Nevertheless, we reduced multiple hypothesis testing to the minimum. In addition, the point estimates observed showed narrow confidence intervals and the crude and adjusted results were consistent. In addition, we would expect that if African ancestry is actually associated with colorectal cancer risk, it would also be associated with its main precursor lesion. However, the use of a common set of controls could also explain such association and therefore the results should be interpreted with caution. As an observational study, residual confounding cannot be ruled out. Replication of these results is warranted to validate our results.

The positive association of African genetic ancestry with adenoma and colorectal cancer is consistent with a recent publication reporting that colorectal cancer risk is likely to be mediated through genetic susceptibility to adenomas[39]. Early detection of adenomas is a key issue in colorectal cancer control. Newly published evidence supports that detecting adenomas and removing them, not only decreases the incidence, but also the mortality of colorectal cancer[40]. There is now promising evidence showing that genetic markers could discriminate population at increased risk of colorectal cancer. It has been shown, for example that adding information of SNPs associated with colorectal cancer to family history increases the absolute risk estimation of having the disease at population level[41].

Further research should address how these SNPs, discovered mainly in European Caucasian population, influence the genetic association here reported. Admixture mapping[42] could be the next step to further explore the mechanism behind this association. There is no admixture mapping analysis published so far on colorectal cancer despite the large number of GWAS on this cancer. A variant in chromosome 8q24 initially described by admixture mapping for prostate cancer in African Americans, showed also a positive association with colorectal cancer risk[30] and recently, a Case control study report an association of one 8q24 loci variant (rs380284) and adenoma risk in Caucasians[39]

In conclusion, we report for the first time, that African ancestry (or variants linked to it) contributes to the susceptibility of colorectal cancer in admixed Latin American population Our results are promising as may help get insights of colorectal carcinogenesis and even more to find biomarkers useful to stratify colorectal cancer risk in the Latino populations, where colorectal mortality rates are increasing, although not high enough to recommend mass screening programs[43].

Funding

National Cancer Institute of Colombia (Internal Grant No. 41030310-24);Colombian National Research Council,Colciencias (Grant No. 210145921617);Union International Cancer Control (UICC) (ACSBI Fellowship 2011 Grant (to GH)); NIGMS – P20GM103501 ;Louisiana Cancer Research Consortium ; Albert Tenesa acknowledges support from Cancer Research UK grant C12229/A13154, and the Biotechnology and Biological Sciences Research Council through the Roslin Institute Strategic Programme Grant.

Acknowledgements

We are most grateful to the participants in the study. Also we like to thank to all the in-field work team (Nurses, Gastroenterologists and Pathologists) in all cities who worked enrolling eligible individuals and to all the administrative support staff at the National Cancer Institute who collaborate with this study. Also we give thanks to all healthcare institutions that participate in this study: Bogota: Hospital San Rafael, Hospital Universitario Mederi, Clinica Colombia. Barranquilla: Clinica General de Norte, UGASEND. Bucaramanga: Hospital Universitario de Santander, (Cali), Fundación Oftalmológica de Santander. Cali: Hospital Univeristario del Valle, Fundación del Valle de Lili, Centro Medico Imbanaco. Cartagena: Hospital San Juan de Dios, Hospital Universitario del Caribe. Hospital Naval. Santa Marta: Laboratorio de patologia CITOPAT, Hospital Universitario Fernando Troconis, Unidad Endoscopica GASTROMAG. Finally the main author will like to thank the UICC for the ASCBI fellowship Grant which allow him to work on the analysis of data for this manuscript at the Roslin Institute

Conflict of Interest Statement*:*

The authors declare no conflict of interest.

References

1 Center MM, Jemal A, Smith RA, Ward E: Worldwide variations in colorectal cancer. *CA Cancer J Clin* 2009; **59**: 366-378.

2 Rim SH, Seeff L, Ahmed F, King JB, Coughlin SS: Colorectal cancer incidence in the United States, 1999-2004. *Cancer* 2009; **115**: 1967-1976.

3 Eheman C, Henley SJ, Ballard-Barbash R *et al*: Annual Report to the Nation on the status of cancer, 1975-2008, featuring cancers associated with excess weight and lack of sufficient physical activity. *Cancer* 2012; **118**: 2338-2366.

4 Agrawal S, Bhupinderjit A, Bhutani MS *et al*: Colorectal cancer in African Americans. *Am J Gastroenterol* 2005; **100**: 515-523; discussion 514.

5 Lansdorp-Vogelaar I, Kuntz KM, Knudsen AB, van Ballegooijen M, Zauber AG, Jemal A: Contribution of screening and survival differences to racial disparities in colorectal cancer rates. *Cancer Epidemiol Biomarkers Prev*; **21**: 728-736.

6 Ollberding NJ, Nomura AM, Wilkens LR, Henderson BE, Kolonel LN: Racial/ethnic differences in colorectal cancer risk: the multiethnic cohort study. *Int J Cancer*; **129**: 1899-1906.

7 Jovov B, Araujo-Perez F, Sigel CS *et al*: Differential gene expression between African American and European American colorectal cancer patients. *PLoS One*; **7**: e30168.

8 Ou J, Carbonero F, Zoetendal EG *et al*: Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. *Am J Clin Nutr*.

9 Chen HM, Weng YR, Jiang B *et al*: Epidemiological study of colorectal adenoma and cancer in symptomatic patients in China between 1990 and 2009. *J Dig Dis* 2011; **12**: 371-378.

10 Wilson LS, Lightwood J: Model of estimated rates of colorectal cancer from polyp growth by year of surveillance. *J Med Screen* 2001; **8**: 187-196.

11 Slattery ML, Fitzpatrick FA: Convergence of hormones, inflammation, and energy-related factors: a novel pathway of cancer etiology. *Cancer Prev Res (Phila)* 2009; **2**: 922-930.

12 Loffeld SM, Loffeld RJ: Colorectal cancer and adenomas are rare in individuals of Turkish descent living in the Zaanstreek region in the Netherlands. *J Cancer Res Clin Oncol* 2010; **136**: 1439-1443.

13 Wallace K, Ahnen D, Burke C *et al*: Race and Risk of Large Bowel Polyps in Younger and Older Patients. *Cancer Epidemiology Biomarkers & Prevention* 2012; **21**: 566.

14 Sucheston LE, Bensen JT, Xu Z *et al*: Genetic Ancestry, Self-Reported Race and Ethnicity in African Americans and European Americans in the PCaP Cohort. *PLoS ONE* 2012; **7**: e30950.

15 Herran OF, Ardila MF, Rojas MP, Hernandez GA: [Design of dietary questionnaires to study the relationships between diet and cancer prevalence in Colombia]. *Biomedica* 2010; **30**: 116-125.

16 Craig CL, Marshall AL, Sjostrom M *et al*: International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* 2003; **35**: 1381-1395.

17 Terry MB, Neugut AI, Bostick RM *et al*: Risk Factors for Advanced Colorectal Adenomas. *Cancer Epidemiology Biomarkers & Prevention* 2002; **11**: 622-629.

18 Cunningham JM, Sellers TA, Schildkraut JM *et al*: Performance of amplified DNA in an Illumina GoldenGate BeadArray assay. *Cancer Epidemiol Biomarkers Prev* 2008; **17**: 1781-1789.

19 Packer BR, Yeager M, Burdett L *et al*: SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. *Nucleic Acids Res* 2006; **34**: D617-621.

20 Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT: Data quality control in genetic case-control association studies. *Nat Protoc* 2010; **5**: 1564-1573.

21 Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559-575.

22 Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. *Genetics* 2000; **155**: 945-959.

23 Duan S, Zhang W, Cox NJ, Dolan ME: FstSNP-HapMap3: a database of SNPs with high population differentiation for HapMap3. *Bioinformation* 2008; **3**: 139-141.

24 Collins-Schramm HE, Chima B, Morii T *et al*: Mexican American ancestry-informative markers: examination of population structure and marker characteristics in European Americans, Mexican Americans, Amerindians and Asians. *Hum Genet* 2004; **114**: 263-271.

25 Rosenberg NA, Li LM, Ward R, Pritchard JK: Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet* 2003; **73**: 1402-1422.

26 Vickers AJ, Bennette C: Against Quantiles: categorization of continuous variables in epidemiologic research, and its discontents. *BMC Med Res Methodol* 2012; **12**: 21.

27 Seldin MF, Pasaniuc B, Price AL: New approaches to disease mapping in admixed populations. *Nat Rev Genet* 2011; **12**: 523-528.

28 Fejerman L, Romieu I, John EM *et al*: European ancestry is positively associated with breast cancer risk in Mexican women. *Cancer Epidemiol Biomarkers Prev* 2010; **19**: 1074-1082.

29 Yang JJ, Cheng C, Devidas M *et al*: Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. *Nat Genet* 2011; **43**: 237-241.

30 Price AL, Patterson N, Yu F *et al*: A genomewide admixture map for Latino populations. *Am J Hum Genet* 2007; **80**: 1024-1036.

31 Aarts MJ, Lemmens VE, Louwman MW, Kunst AE, Coebergh JW: Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome. *Eur J Cancer* 2012; **46**: 2681-2695.

32 Doubeni CA, Laiyemo AO, Major JM *et al*: Socioeconomic status and the risk of colorectal cancer. *Cancer* 2012; **118**: 3636-3644.

33 Leufkens AM, Van Duijnhoven FJ, Boshuizen HC *et al*: Educational level and risk of colorectal cancer in EPIC with specific reference to tumor location. *Int J Cancer* 2012; **130**: 622-630.

34 Haenszel W, Correa P, Cuello C: Social class differences among patients with large-bowel cancer in Cali, Colombia. *J Natl Cancer Inst* 1975; **54**: 1031-1035.

35 Florez JC, Price AL, Campbell D *et al*: Strong association of socioeconomic status with genetic ancestry in Latinos: implications for admixture studies of type 2 diabetes. *Diabetologia* 2009; **52**: 1528-1536.

36 Fenoglio L, Castagna E, Comino A *et al*: A shift from distal to proximal neoplasia in the colon: a decade of polyps and CRC in Italy. *BMC Gastroenterol* 2010; **10**: 139.

37 Ikeda Y, Mori M, Yoshizumi T, Sugimachi K: Cancer and adenomatous polyp distribution in the colorectum. *Am J Gastroenterol* 1999; **94**: 191-193.

38 Ministerio de Educacion, Colombia. http://www.mineducacion.gov.co/cvn/1665/article-99519.html, 2006.

39 Carvajal-Carmona LG, Zauber AG, Jones AM *et al*: Much of the genetic risk of colorectal cancer is likely to be mediated through susceptibility to adenomas. *Gastroenterology* 2013; **144**: 53-55.

40 Zauber AG, Winawer SJ, O'Brien MJ *et al*: Colonoscopic Polypectomy and Long-Term Prevention of Colorectal-Cancer Deaths. *New England Journal of Medicine* 2012; **366**: 687-696.

41 Dunlop MG, Tenesa A, Farrington SM *et al*. Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 420.103 individuals: Gut, 2012.

42 Chanock SJ: A twist on admixture mapping. *Nat Genet* 2011; **43**: 178-179.

43 Lambert R, Sauvaget C, Sankaranarayanan R: Mass screening for colorectal cancer is not justified in most developing countries. *International Journal of Cancer* 2009; **125**: 253-256.

LEGENDS TO FIGURES

Figure 1. Quantile-Normal plot per ancestry fraction in the study population

Figure 2. Forest plot for the Standardized Mean Differences (SMD) of African Ancestry fraction (log scale) in Adenomas and Adenocarcinomas stratified by Education level attained.

Figure 3. European and African Ancestry Fraction Box plot per Education level attained in the study population

Table 1. Main demographic characteristics and risk factors in the study population

Risk factor	Controls (n=243)		Ademoma (n=113)		*p*	Adenocarcinoma (n=190)		*p*
	n	%	n	%		n	%	
Sex								
Female	141	58.02	54	47.79		90	47.4	
Male	102	41.98	59	52.21	0.13	100	52.6	0.14
Age								
mean y, (sd)		54.2 (0.76)		57.4 (0.94)	0.01		57.9 (12.3)	0.01
Education Attained								
None	6	2.5	2	1.8		18	9.5	
Primary School	91	37.4	33	29.2		75	39.5	
Secondary	89	36.6	37	32.7		61	32.1	
College	28	11.5	11	9.7		11	5.8	
University	29	11.9	30	26.5	0.02	25	13.2	<0.01
Colorectal Cancer Family History			113					
No	231	95.0	104	92.0		176	92.6	
Yes	13	5.0	9	8.0	0.26	14	7.4	0.29
Alcohol Intake (gr/ day)								
No intake	135	55.6	72	63.7		107	56.3	
<12.5	70	28.8	22	19.5		49	25.8	
12.5	38	15.6	18	15.9	0.17	33	17.4	0.75
Smoking (packs/y)								
Non smoker	146	60.1	71	62.8		112	58.9	
less than 1	51	21.0	16	14.2		32	16.8	
1	46	18.9	26	23.0	0.35	45	23.7	0.27
Red Meat (servings p/w)								
less than 2	42	17.3	20	18.6		35	18.4	
2-5	121	49.8	51	44.2		80	42.1	
5	78	32.1	34	30.1	0.83	62	32.6	0.59
Energy Intake (Quartile)								
Q1	60	24.7	30	26.5		29	15.3	
Q2	61	25.1	29	25.7		41	21.6	
Q3	61	25.1	19	16.8		50	26.3	
Q4	60	24.7	25	22.1	0.53	56	29.5	0.13
BMI								
<20	6	2.5	5	4.4		35	18.4	
20-24	85	35.0	36	31.9		82	43.2	
25-29	111	45.7	47	41.6		47	24.7	
30	45	18.5	23	20.4		24	12.6	<0.01

NSAID								
No	184	75.7	87	77.0		148	77.9	
Yes	59	24.3	26	23.0	0.84	42	22.1	0.6
Dietary Fiber (Quartile)								
Q1	61	25.1	24	21.2		44	23.2	
Q2	61	25.1	23	20.4		39	20.5	
Q3	61	25.1	25	22.1		48	25.3	
Q4	59	24.3	33	29.2	0.63	46	24.2	0.88
Physical Activity (ipaq categories)								
Low	102	42.0	58	51.3		77	40.5	
Moderate	72	29.6	27	23.9		49	25.8	
High	69	28.4	28	24.8	0.17	64	33.7	0.6
City of enrolment								
Bucaramanga	69	28.4	35	31.0		40	21.1	
Bogota	30	12.3	10	8.8		23	12.1	
Cartagena	33	13.6	12	10.6		26	13.7	
Cali	33	13.6	10	8.8		33	17.4	
SantaMarta	26	10.7	20	17.7		27	14.2	
Barranquilla	52	21.4	26	23.0	0.51	41	21.6	0.56
Anatomic location								
Colon NOS			8	7.1		8	4.2	
Right Colon			40	35.4		48	25.3	
Left Colon			43	38.1		42	22.1	
Rectum			22	19.5		92	48.4	

Table2. Mean ancestry fraction and 95% Confidence intervals in controls, adenomas, adenocarcinomas and admixed population (MEX and ASW) included into the analysis.

	Ancestry Fraction		
Group / Population	European mean [95% CI]	Amerindian$^{\Omega}$ mean [95% CI]	African mean† [95% CI]
Controls (n=238)	0.39 [0.38-0.41]	0.45 [0.43-0.46]	0.11 [0.10-0.12]
Adenoma (n=115)	0,44*** [0.42-0.46]	0,39*** [0.37-0.41]	0,13* [0.11-0.15]
Cancer (n=190)	0.38 [0.37-0.40]	0.43 [0.41-0.45]	0,14** [0.12-0.16]
MEX (n=77)	0.48 [0.45-0.51]	0.48 [0.45-0.51]	0.04 [0.03-0.05]
ref value‡	**0.5**	**0.45**	**0.05**
ASW (n=83)	0.18 [0.15-0.20]	0.05 [0.04-0.07]	0.77 [0.75-0.79]
ref value‡	**0.2**	**-**	**0.8**

Ω Based on CHB reference population

†Back transformated from log

<0,001***, <0,01** and <0.1* p value for one-way ANOVA test

‡ (Seldin et al., 2011)

Table3. Categorical Distribution of African and European Ancestry Substitution * in controls, adenomas and Colorectal cancer cases, included into the analysis.

	Control n %	Adenomas n %	Cancer n %
African Ancestry			
0.01 - 0.09	110 45.27	39 34.51	58 30.53
0.10- 0.19	58 23.87	38 33.63	64 33.68
0.20 - 0.29	39 16.05	24 21.24	34 17.89
0.30 - 0.39	36 14.81	12 10.62	34 17.89
p value+		0.07	0.02
European Ancestry*			
≤ -0.30	55 22.63	12 10.62	46 24.21
-0.29 - -0.20	40 16.46	17 15.04	34 17.89
-0.19 - -0.10	49 20.16	20 17.7	39 20.53
0 - 0,09	58 23.87	22 19.47	36 18.95
0.10- 0.19	41 16.87	42 37.17	35 18.42
p value+		0.001	0.87

*European Ancestry corrected for Amerindian Ancestry (European Fraction *minus* Amerindian Ancestry)

+ p value for chi squared test.

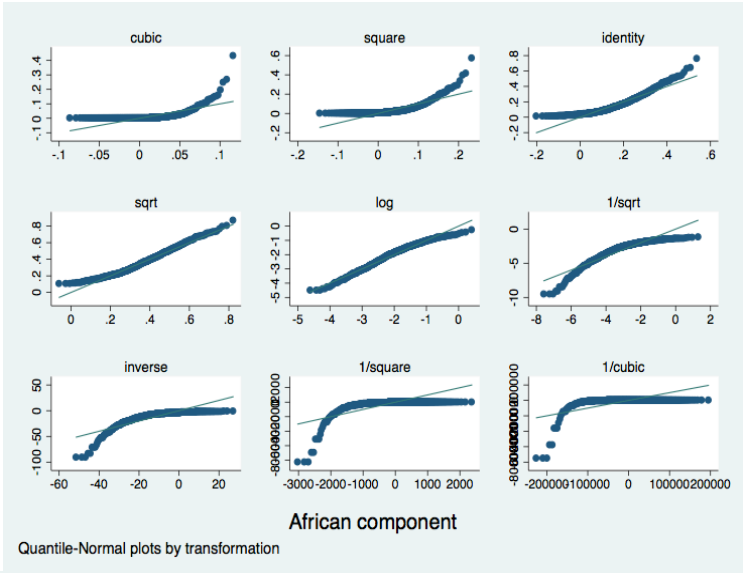
Table 4. Odds Ratio (OR) and 95% Confidence intervals (CI) of Age and Sex Adjusted and fully adjusted conditional regression models for Genetic ancestry fractions and known risk factor of colorectal adenoma and adenocarcinoma.

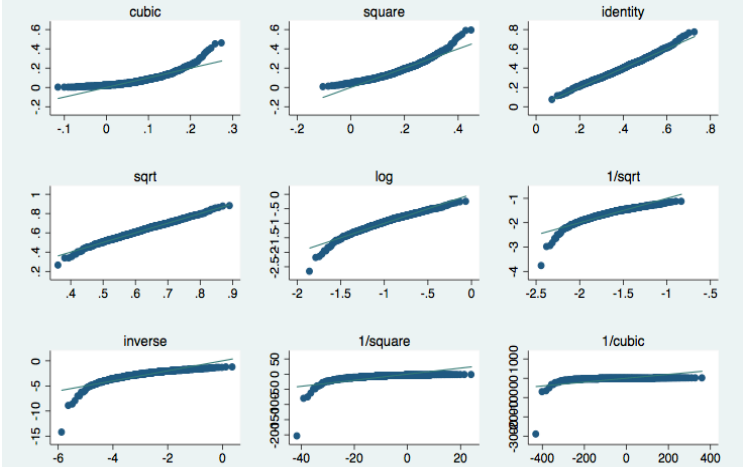
Characteristic		Adenoma			Adenocarcinoma		
		OR	95% CI	p	OR	95% CI	p
Age-Sex adjusted model							
African Ancestry							
risk variation per 10% increase		1.13	[0.99,1.27]	0.06	1.18	[1.06,1.31]	0.001
p for trend			0.09			0.002	
European Ancestry							
risk variation per 10% increase		1.25	[1.08,1.46]	0.001	0.98	[0.88,1.10]	0.79
p for trend			0.0004			0.91	
Full Adjusted model							
African Ancestry		ref	-		ref	-	
risk variation per 10% increase		1.12	[0.97,1.30]	0.15	1.19	[1.05,1.35]	0.006
p for trend			0.08			0.003	
European Ancestry							
risk variation per 10% increase		1.25	[1.08,1.46]	0.007	1.02	[0.90,1.16]	0.75
p for trend			0.001			0.68	
Education Attained	None	0.77	[0.11,5.24]	0.79	2.56	[0.89,7.36]	0.08
	Primary School	ref	-		ref	-	
	Secundary	1.5	[0.74,3.06]	0.26	0.77	[0.45,1.30]	0.4
	College	2.11	[0.78,5.75]	0.14	0.55	[0.23,1.31]	0.23
	University	3.82	[1.59,9.17]	0.01	1.03	[0.60,2.52]	0.95
Colorectal Cancer Family History*	yes	1.73	[0,60,5.02]	0.83	1.72	[0.67, 4.40]	0.74
Alcohol Intake (gr/ day)*	<12.5	0.65	[0.30,1.44]	0.29	1.08	[0.56,2.07]	0.82
	12.5	0.86	[0.40,1.83]	0.69	0.97	[0.48,1.94]	0.92
Smoking (packs/y)	<0.5	ref	-		ref	-	
	0.5-0.9	0.65	[0.30,1.44]	0.29	1.08	[0.56,2.07]	0.82
	1	0.86	[0.40,1.83]	0.69	0.97	[0.48,1.94]	0.92
Read Meat (times p/w)	<2	ref	-		ref	-	
	2-4	0.87	[0.41,1.84]	0.71	1.34	[0.67,2.67]	0.41
	5	0.84	[0.37,1.88]	0.67	1.66	[0.80,3.42]	0.17
Energy Intake (Quartile)	Q1	ref	-		ref	-	
	Q2	0.81	[0.35,1.88]	0.62	0.64	[0.31,1.33]	0.23
	Q3	1.21	[0.49,2.99]	0.69	0.68	[0.30,1.52]	0.34
	Q4	1.36	[0.51,3.65]	0.54	0.44	[0.18,1.08]	0.09
NSAID	Yes	0.6	[0.29,1.21]	0.15	0.93	[0.52,1.66]	0.82
Dietary Fiber (Quartile)	Q1		1				
	Q2	0.81	[0.35,1.88]	0.62	0.64	[0.31,1.33]	0.23
	Q3	1.21	[0.49,2.99]	0.69	0.68	[0.30,1.52]	0.34
	Q4	1.36	[0.51,3.65]	0.54	0.44	[0.18,1.08]	0.07
Physical Activity (ipaq categories)	Low						
	Moderate	0.67	[0.34,1.32]	0.24	0.77	[0.43,1.37]	0.37
	High	0.76	[0.37,1.54]	0.44	0.89	[0.50,1.60]	0.7

†‡ European ancestry substitution effect (European - Amerindian ancestry) adjusted only for age, sex and city of enrollment

† European ancestry substitution effect (European - Amerindian ancestry) adjusted for age, sex, city of enrollment and all other factors listed

*Absence of the exposure considered as reference value


cubic
square
identity
sqrt
log
1/sqrt
inverse
1/square
1/cubic
African component
Quantile-Normal plots by transformation


cubic
square
identity
sqrt
log
1/sqrt
inverse
1/square
1/cubic
European component
Quantile-Normal plots by transformation

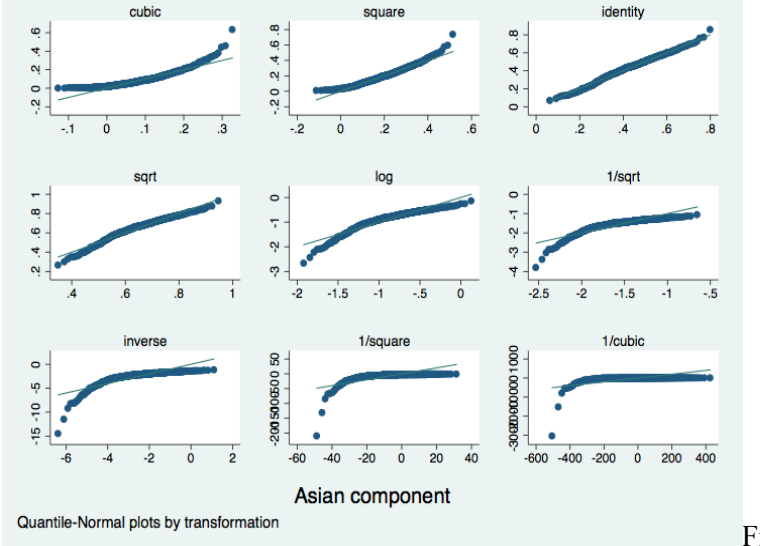

cubic
square
identity
sqrt
log
1/sqrt
inverse
1/square
1/cubic
Asian component
Quantile-Normal plots by transformation

Figure 1

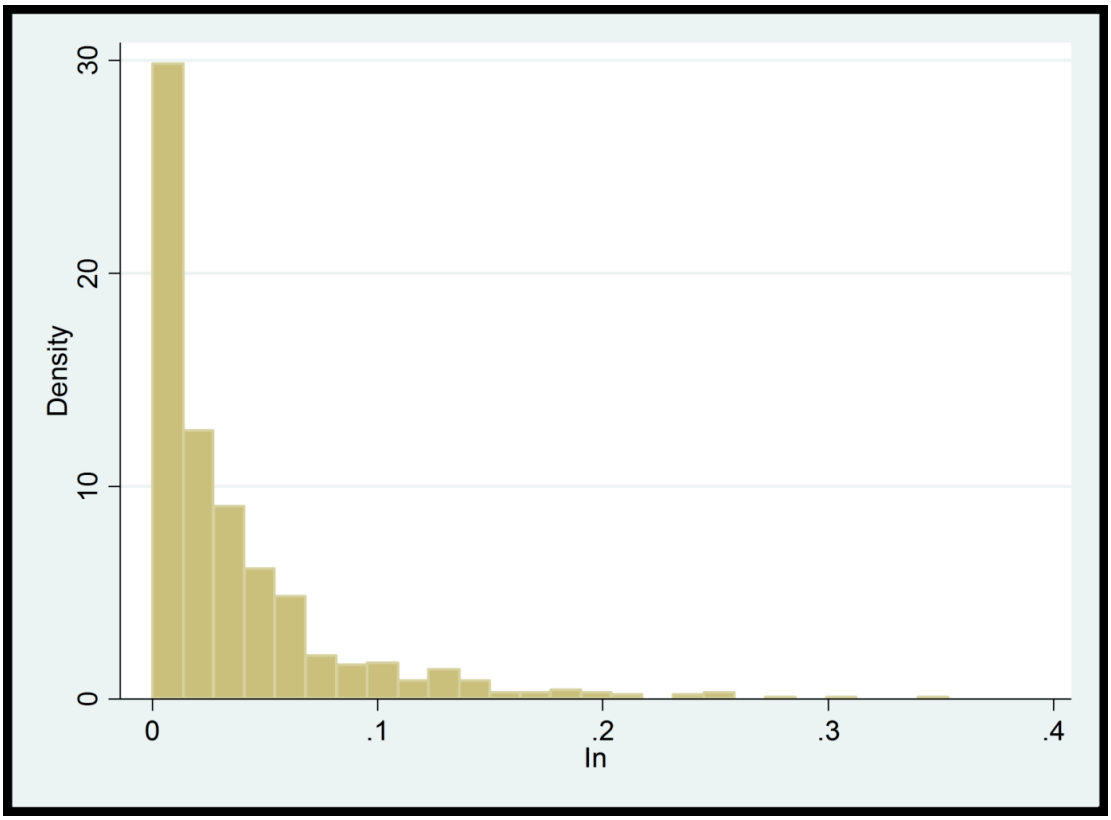


Figure 2

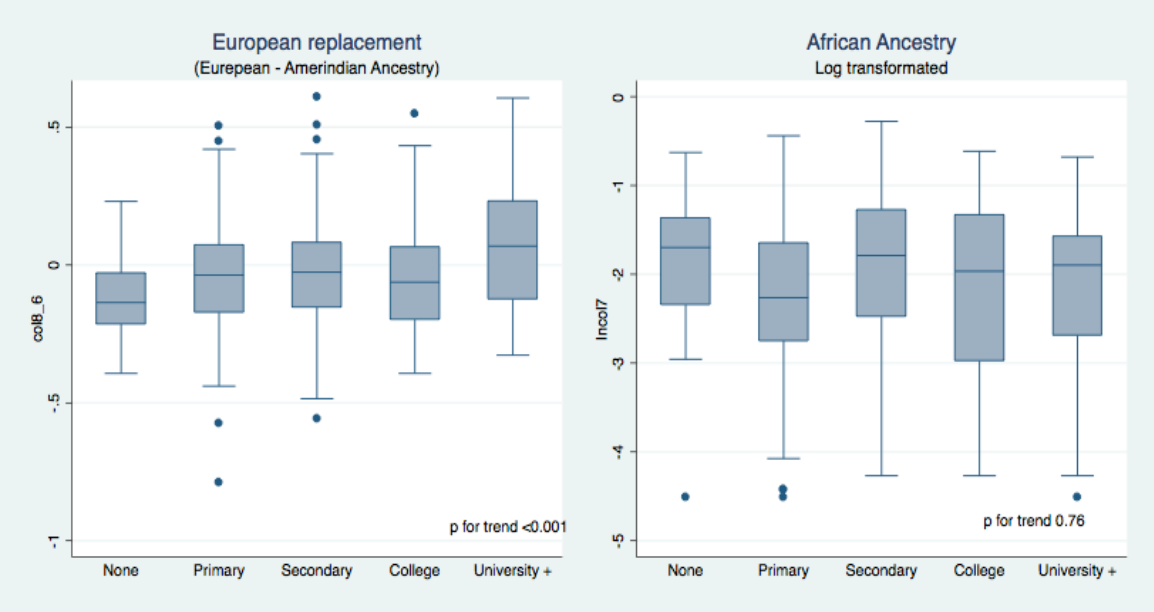

European replacement
(Eurepean - Amerindian Ancestry)
col8_6
.5
0
-.5
-1
p for trend <0.001
None
Primary
Secondary
College
University +
African Ancestry
Log transformated
lncol7
0
-1
-2
-3
-4
-5
p for trend 0.76
None
Primary
Secondary
College
University +

Figure 3

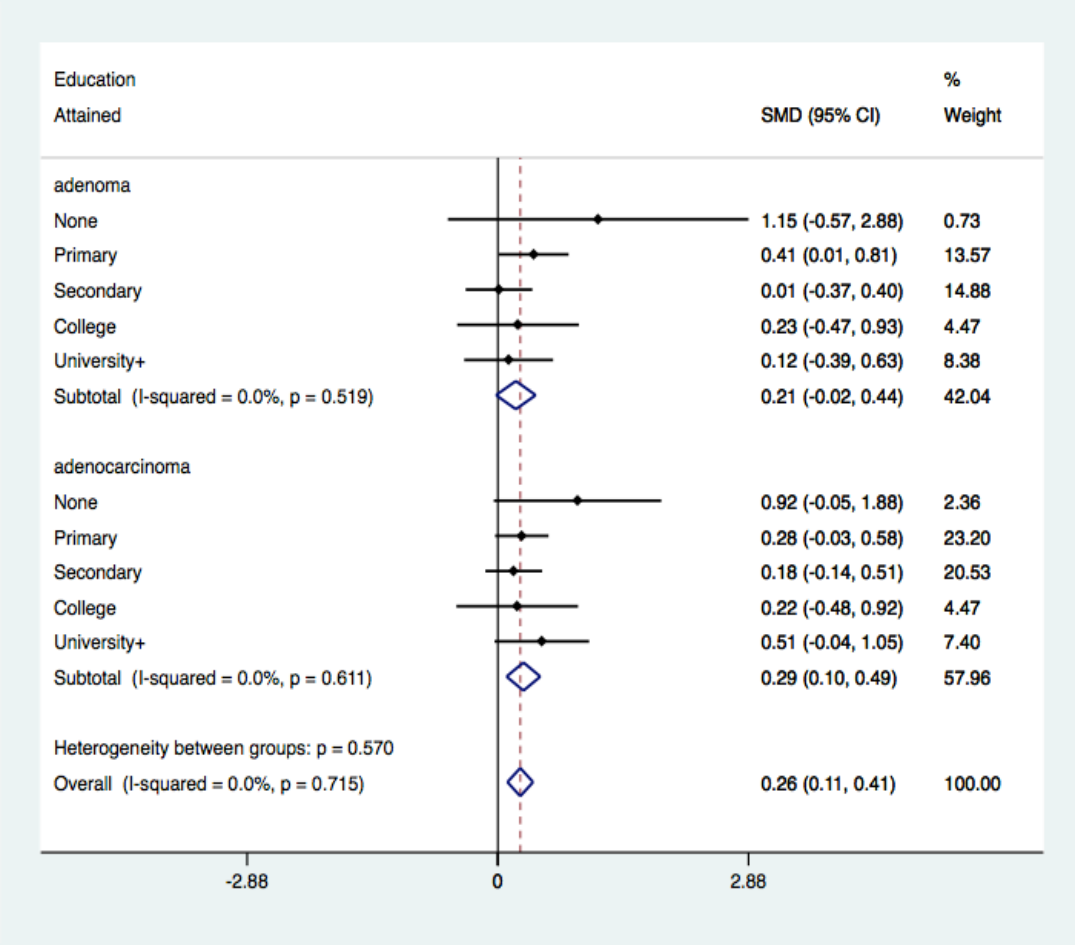

Education
Attained
SMD (95% CI)
%
Weight
adenoma
None
1.15 (-0.57, 2.88)
0.73
Primary
0.41 (0.01, 0.81)
13.57
Secondary
0.01 (-0.37, 0.40)
14.88
College
0.23 (-0.47, 0.93)
4.47
University+
0.12 (-0.39, 0.63)
8.38
Subtotal (I-squared = 0.0%, p = 0.519)
0.21 (-0.02, 0.44)
42.04
adenocarcinoma
None
0.92 (-0.05, 1.88)
2.36
Primary
0.28 (-0.03, 0.58)
23.20
Secondary
0.18 (-0.14, 0.51)
20.53
College
0.22 (-0.48, 0.92)
4.47
University+
0.51 (-0.04, 1.05)
7.40
Subtotal (I-squared = 0.0%, p = 0.611)
0.29 (0.10, 0.49)
57.96
Heterogeneity between groups: p = 0.570
Overall (I-squared = 0.0%, p = 0.715)
0.26 (0.11, 0.41)
100.00
-2.88
0
2.88

Figure 4