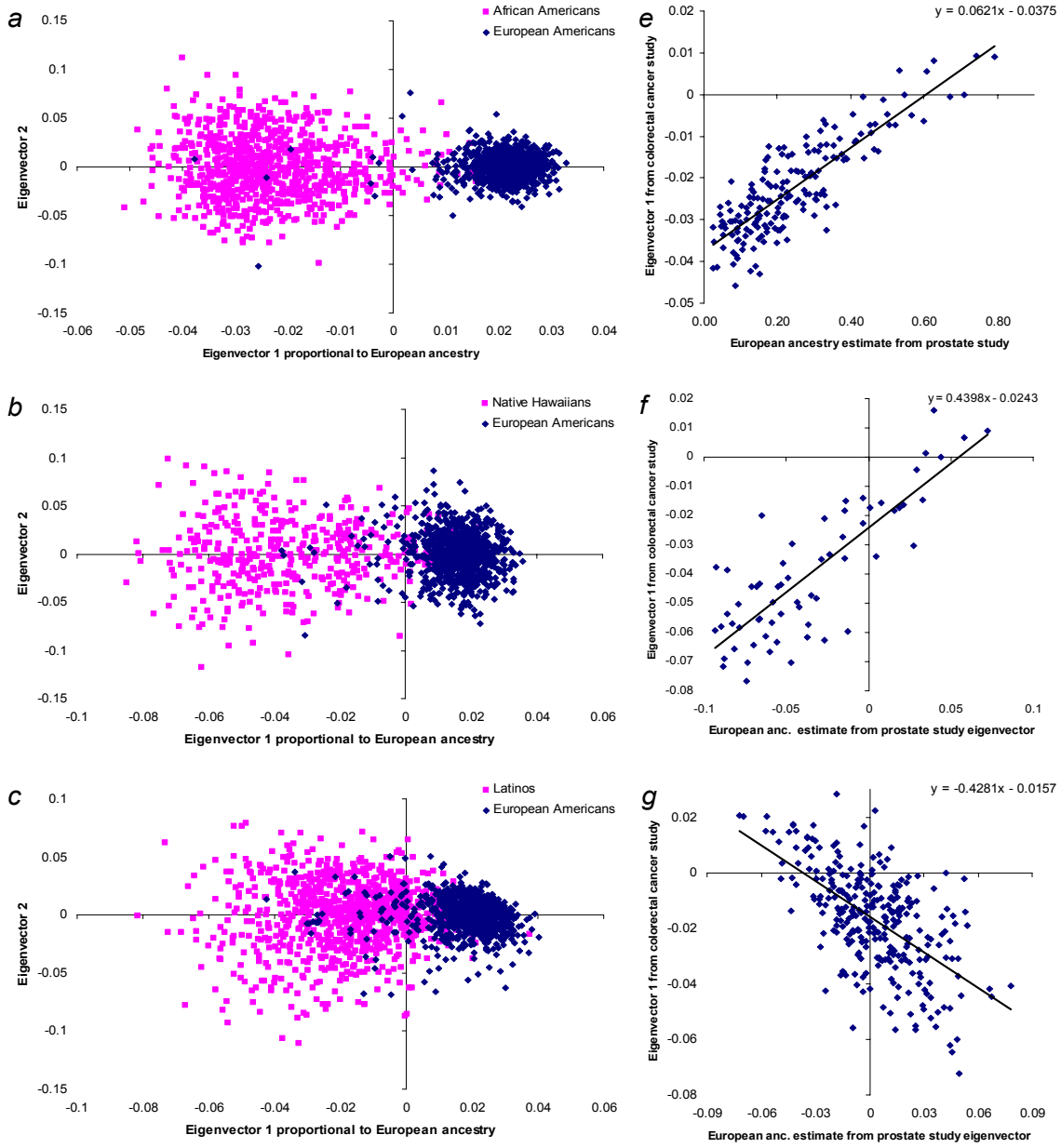
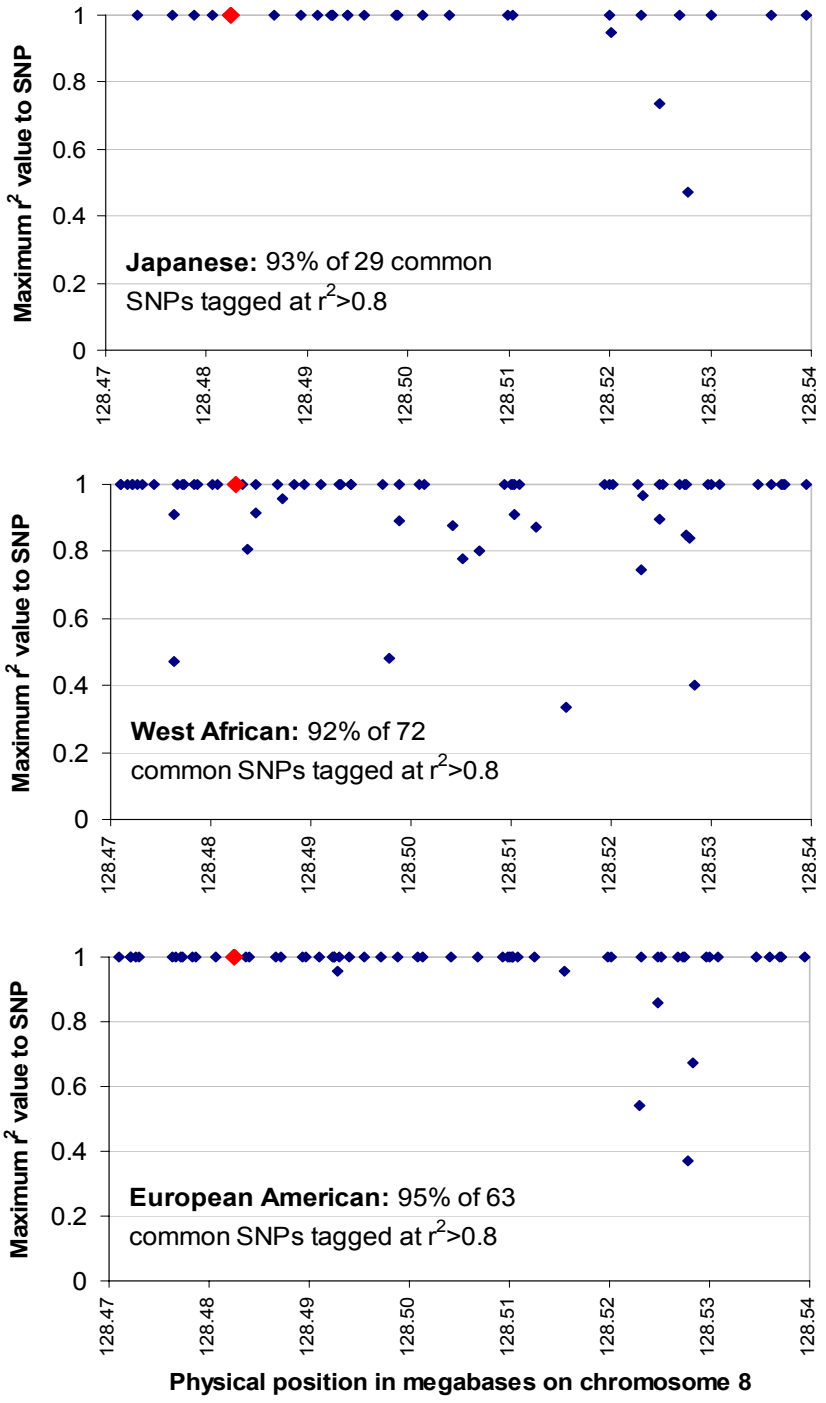
Supplementary Figure 1. Principal components analysis of European ancestry in the African American, Native Hawaiian and Latino populations.



Supplementary Figure 1: (a-c) To estimate proportion of European ancestry from each of the 3 admixed populations, we used a panel of ancestry informative markers chosen specifically for each population (43 for African Americans, 39 for Native Hawaiians and 41 for Latinos, as described in Supplementary Methods). Principal components analysis was performed combining the samples from the admixed population, with European Americans, to assist in the identification of an eigenvector proportional to ancestry. Eigenvector 1 (x axis) is proportional to European ancestry, whereas Eigenvector 2 is shown merely to help with visualization and is not significantly correlated with ancestry. **(d-f)** We also plot the estimate of European ancestry for our published prostate cancer study (x-axis), against the eigenvector-based estimate of European ancestry from the colorectal cancer candidate gene study (y-axis), for the samples from each of the three admixed populations that overlap between the two data sets. We used linear regression to relate the estimates in the two studies (shown in figure), allowing us to obtain a European ancestry estimate that is comparable across studies.

Supplementary Figure 2. Tag SNP coverage at 128.47-128.54 Mb in the Japanese, West African and European American HapMap populations.



Supplementary Figure 2: Maximum correlations (r^2) between the 82 SNPs genotyped from 128.47-128.54 Mb and all SNPs of >5% minor allele frequency in the Japanese, West Africans and European Americans in the HapMap database. Blue diamonds represents a SNP from HapMap (the red diamond is rs6983267). These tags capture 93%, 92% and 95% of the common SNPs in each population, respectively, with r^2>0.8.

Supplementary Methods

Study Populations

The Multiethnic Cohort Study: The initial testing of the six risk alleles was conducted in the Multiethnic Cohort Study (MEC). The MEC consists of over 215,000 men and women in Hawaii and Los Angeles (with additional African-Americans from elsewhere in California)[1]. The cohort is comprised predominantly of African Americans, Japanese Americans, Native Hawaiians, Latinos and European Americans who entered the study between 1993 and 1996 by completing a 26-page self-administered questionnaire that requested detailed information about dietary habits, demographic factors, personal behaviors, history of prior medical conditions, family history of common cancers, and for women, reproductive history and exogenous hormone use. The participants were between the ages 45 and 75 at enrollment. Incident cancers in the MEC are identified by cohort linkage to population-based cancer Surveillance, Epidemiology and End Results (SEER) registries covering Hawaii and Los Angeles County, and to the California State cancer registry covering all of California. From the registries, information about stage of disease and site of tumor (colon versus rectum) is available. Beginning in 1994, blood samples were collected from incident colorectal cancer cases and a random sample of MEC participants to serve as a control pool for genetic analyses in the cohort. Eligible cases in the colorectal cancer case-control study consisted of men and women with incident invasive colorectal cancer diagnosed after enrollment in the MEC through December, 2004. Controls were participants without colorectal, breast or prostate cancer prior to entry into the cohort and without a diagnosis up to December 2004 (the pool of controls was shared across our studies). The colorectal cancer case-control study in the MEC consisted of 1,140 invasive colorectal cancer cases and 4,607 controls (2,613 males were also included in our prostate cancer study[2]). We excluded 34 European Americans and 16 Japanese Americans due to discrepancies between self-reported and inferred ancestry (see below), leaving 1,124 colorectal cancer cases and 4,573 controls for analysis. Staging information was available for 97.7% (n=1,098) of cases and tumor location information was available on 100% of cases. This study was approved by the Institutional Review Boards at the University of Southern California and at the University of Hawaii.

The Hawaii Case-Control Study: SNPs rs6983267 and rs10090154 were also examined in a population-based case-control study of colorectal cancer that includes 327 invasive colorectal cancer cases and 525 controls. All the samples were Japanese Americans and European Americans; Native Hawaiians were part of the study, but were excluded from our analyses because they were not genotyped for ancestry informative markers, which were used as covariates in our analysis. This Hawaii Case-Control Study has been described in detail previously[3]. Cases were identified through the Hawaii SEER registry and consisted of Japanese American, European American and Native Hawaiian residents of Oahu, Hawaii, who were newly diagnosed with colon or rectal cancer between January 1994 and August 1998. Controls were selected from participants in an ongoing population-based health survey conducted by the Hawaii State Department of Health and from Health Care Financing Administration participants. Staging information was available on 98.8% (n=323) of cases and tumor location information was available on 100% of cases. A personal interview and a blood sample were obtained from each subject. This study was approved by the Institutional Review Board at the University of Hawaii.

The Los Angeles County Case-Control Study: SNPs rs6983267 and rs10090154 were also examined in a population-based case-control study of colorectal cancer that includes 356 invasive colorectal cancer cases and 413 controls who were European Americans. Non-European American case and control subjects were part of this study, but were not included in this analysis because they were not genotyped

for ancestry informative markers, which were used as covariates in our analysis. Cases were identified from the Los Angeles County Cancer Surveillance Program. Eligible cases included English-speaking women with a histologically confirmed incident colorectal cancer, diagnosed at ages 55 to 74 years, on or after January 1998 through December 2002 and were residents of Los Angeles County at the time of diagnosis. Controls were selected from the neighborhoods where cancer cases resided at the time of diagnosis using a well-established, standard algorithm to identify neighborhood controls used in numerous other case-control studies. A personal interview and a blood sample were obtained from each subject. Staging information was available on 99.7% (n=355) of cases and tumor location information was available on 100% of cases. This study was approved by the Institutional Review Board at the University of Southern California. The main results from this case-control study are described in a manuscript that is in preparation for submission[4].

SNP Genotyping

The variants were genotyped in the MEC samples using Sequenom and TaqMan platforms at the Broad Institute and at the University of Southern California (USC), respectively. Genotyping of rs6983267 and rs10090154 in the Hawaii case-control study was conducted at the University of Hawaii. Genotyping of these two variants in the Los Angeles Case-Control Study was conducted at USC. Blinded duplicate samples (~2-5%) were included in all 96-well DNA plates. The discordance rate among duplicates was <0.5% in the MEC and 0% in the case-control studies.

Fine-Mapping of Region of Strong Linkage Disequilibrium Around rs6983267

To identify variants that could potentially capture additional risk for colorectal cancer, above and beyond rs6983267, we focused on the region from 128.47-128.54 Mb in strong linkage disequilibrium with this SNP. For this purpose, we mined genotyping data from the HapMap West African, European American and East Asian populations for 186 SNPs across the region: 109 of these SNPs have data in the HapMap database, while 77 SNPs were identified by new genotyping and SNP discovery in HapMap samples in our prostate cancer study (see Supp. Table 4 of ref. 2 for the non-HapMap SNPs). We selected 96 SNPs that captured at r^2>0.8 all variants of >5% minor allele frequency in any of the populations.

We used the Sequenom MassArray iPLEX Gold technology[5] at the Broad Institute to attempt to genotype 91 of these SNPS (the ones that we could successfully design with the technology) in 1,107 colorectal cancer cases and 1,844 controls from the MEC. After removing SNPs that failed quality control filters described previously (ref. 2), we were left with 82 SNPs with genotypes useable for our analysis. In **Supplementary Figure 2**, we present results on how SNPs of >5% minor allele frequency in HapMap are captured by this panel, in each of the West African, European American, and Japanese populations.

To prepare the data set for analysis, we removed 40 subjects (19 cases and 21 controls) with genotype call rates <60%. To assess genotyping quality, we used blinded duplicate samples (n=162) and HapMap trios (n=63). The concordance rate for these QC samples was >99.9%; there were only 13 discrepancies and no SNP had more that 2 discrepant genotype calls (<1% error). All SNPs conformed to Hardy-Weinberg Equilibrium in control samples for at least four of the five populations (P>0.01). The final data set thus included genotype data for 82 SNPs (including rs6983267 and rs7000448), genotyped in 1,088 cases and 1,823 controls (cases/controls: African Americans, 205/411; Japanese Americans, 374/410; Native Hawaiians, 60/239; Latinos, 243/426; European Americans, 206/337).

Statistical Analysis of the Six Variants at 8q24

Odds ratios (OR) and 95% confidence intervals were calculated using unconditional logistic regression. For each SNP, ethnic-specific and pooled odds ratios were estimated adjusting for gender, and, population and gender, respectively. In the African Americans, Native Hawaiians and Latinos we also controlled for the potential confounding effect of European genome-wide ancestry (estimated by principal components as described below). For analyses of rs6983267 and rs10090154 we also adjusted for study. Heterogeneity of effects by population or gender was examined by the inclusion of interaction terms in multivariate models. We examined the evidence for multiplicative allelic effects by comparing the fit between models that included genotype-specific covariates (separate indicator variables for homozygotes and heterozygotes) and genotype as a linear variable (0, 1 or 2 copies of the variant allele). We used a χ^2 test to evaluate significance. Interactions with age at diagnosis ($\leq$67 vs. >67 years, centered on the median age), family history of colorectal cancer in first-degree relatives (yes or no), body mass index (<25 vs. $\geq$25 kg/m^2), smoking history (ever vs. never), aspirin use (ever vs. never) alcohol consumption (<1 vs. $\geq$1 drink/day) and use of estrogen therapy among women (ever vs. never) were examined by a likelihood ratio test. Heterogeneity of effects by stage (localized vs. regional/distant) and site (colon vs. rectum) were examined by logistic regression in case-only analyses.

To calculate population attributable risk (PAR) we let k_j represent the number of copies of the risk allele at rs6983267 (j = 0,1,2). We also let P_i denote the proportion of controls in a given population with a given genotype, i, and let $R_i = \exp[\beta_1 k_1]$ denote the relative risk (odds-ratio) for each genotype. The PAR for each population is then $PAR = (\sum P_i R_i - 1) / \sum P_i R_i$.

To formally test whether the 6 SNPs differed from each other in terms of their ability to predict whether a case was of prostate versus colorectal cancer (which if rejected, would indicate a different mechanism associated with some SNPs), we carried out a case-only analysis (limited to colorectal and prostate cancer cases with complete data for all six SNPs: 1,799 prostate cases and 1,064 colorectal cases). In this analysis, we compared the likelihood of the data for a model in which a common odds ratio was fit to the allele dosage for all 6 SNPs (a maximum of 12 risk alleles), to the likelihood of the data in a model in which 6 separate OR parameters were estimated for each SNP separately. Twice the difference in the log likelihoods was compared to a chi-square distribution with 5 degrees of freedom to assess statistical significance. In this analysis adjustment was made for age at diagnosis, reported ethnicity, and estimated European ancestry in the admixed African American, Native Hawaiian and Latino populations.

The statistical analysis of the fine-mapping data was performed using logistic regression controlling for population, gender and European genome-wide ancestry proportion for the three admixed populations (African Americans, Native Hawaiians and Latinos). We performed both single SNP analyses and stepwise regression analyses. For the stepwise procedure, we started with the SNP of interest (rs6983267 or rs10808556), and then tested all other SNPs (n=81) in the model one at a time. This was also performed separately for each ethnicity.

Estimation of European Ancestry in African Americans, Native Hawaiians and Latinos

To estimate the percentage of European ancestry in each of the individuals from 3 admixed populations from the MEC (African Americans, Native Hawaiians, and Latinos), we combined two different data sets. The first was collected for our study of prostate cancer susceptibility[2] (1,547 African Americans, 223 Native Hawaiians and 1,270 Latinos), and consisted of ancestry informative markers genotyped in each of the three populations. The second set was collected for a study of colorectal cancer susceptibility focusing

on 1,339 SNPs in candidate genes involved in DNA repair (850 African Americans, 415 Native Hawaiians, 842 Latinos, 921 European Americans, and 1,224 Japanese Americans; in preparation).

To combine the information from these two data sets to obtain a uniform estimate of ancestry, we used the fact that there were some overlapping samples between the two different studies (177 African American controls, 63 Native Hawaiian controls and 289 Latino controls). This allowed us to obtain equations relating the estimates of ancestry from the two data sets, leading to a single estimate proportional to European ancestry for 2,220 African Americans, 575 Native Hawaiians, and 1,923 Latinos. The equations we used are shown in **Supplementary Figure 1**.

Ancestry-Informative Panels of Markers Extracted from the DNA Repair Gene Data Set
From the 1,339 SNPs that had been genotyped for the DNA repair genes (all on chromosome 1-22), we were able to identify ancestry-informative sets of 43 SNPs in African Americans, 39 SNPs in Native Hawaiians, and 41 SNPs in Latinos.

To obtain these panels, we first estimated the frequency of the variant allele of each SNP in all 5 MEC populations (African Americans, Japanese Americans, Native Hawaiians, Latinos and European Americans). We then selected markers that were widely spaced in the genome (at least 0.5 centimorgans apart), and chosen to be maximally informative for estimating ancestry in each population of interest. For African Americans, we picked markers to be maximally differentiated in frequency between African Americans and European Americans. For Native Hawaiians, we picked markers to be maximally differentiated in frequency between Native Hawaiians and European Americans. For Latinos, we picked markers to be maximally differentiated in frequency between Latinos and European Americans.

Validation that the Markers are Ancestry Informative
In **Supplementary Figure 1**, we show that for each of the populations, the first principal component from application of Principal Components Analysis (PCA)[6] is highly informative for estimating European ancestry proportion in the admixed populations. To obtain these plots, we carried out PCA separately on the European Americans and each of the admixed populations, and plotted the first principal component against the second principal component. The first principal component clearly separates the self-identified European Americans from the admixed populations, and hence we used the value of this principal component for each individual as a number that is proportional to European ancestry.

As a second validation of the usefulness of these estimates, **Supplementary Figure 1** shows a plot of the ancestry estimate obtained from this study, with that used in our previous study of prostate cancer based on a different set of ancestry informative markers. The estimates are highly correlated for samples that overlapped between the two studies, as expected if they capture the true proportion of ancestry in the admixed populations.

Ancestry Estimates made Comparable across Colorectal and Prostate Cancer Data Sets
We carried out linear regressions to relate the European ancestry estimate from the prostate cancer study, to the European ancestry estimate from the colorectal cancer study, for the samples that had been genotyped in both panels of ancestry informative markers (177 African Americans, 63 Native Hawaiians, and 289 Latinos). The result of the regression analysis is shown by the trendline and equation in each of the three right panels in **Supplementary Figure 1** (one equation for each population).

We used the equation to convert the estimates of European ancestry for each population from the prostate cancer data set, to estimates proportional to those from the colorectal cancer data set. We thus obtained a uniform number proportional to European ancestry for 2,220 African Americans, 575 Native Hawaiians, and 1,923 Latinos. We used these numbers as covariates (proportional to genome-wide ancestry, allowing for stratification correction), in case-control analyses to detect SNP associations in each of the populations.

Identification of Outlier European American and Japanese American Samples
The first principal component plotted in the panels (a-c) of **Supplementary Figure 1** not only allows us to estimate the proportion of European ancestry in admixed populations, but also to identify European Americans that appear to have unusual ancestry compared with others of the same self-declared ancestry.

We compiled a list of 29 European American samples (out of the total of 921) that appeared to be outliers compared with others that were labeled with the same ancestry. We identified these as individuals with values <0.007 for the African American first principal component, <-0.02 for the Native American first principal component, and <-0.025 for the Latino first principal component.

We also identified additional outliers in the Japanese American samples, by repeating the analysis described above with a new panel of 39 markers chosen to be informative comparing Japanese American and European American ancestry (analysis not shown). This identified an additional 11 Japanese Americans (out of a total of 1,224 Japanese Americans genotyped in the colorectal cancer DNA repair gene data set) that appeared to be genetic outliers with respect to others of the same self-declared ancestry.

Finally, we combined the outlier analysis above, with the one from our previous study of prostate cancer, to obtain a combined list of ancestry outliers: 34 European Americans and 16 Japanese Americans. We removed these subjects from all subsequent analysis.

References

[1] Kolonel, L.N. et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol. 151, 346-357 (2000).

[2] Haiman, C.A. et al. Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet. 39, 638-644 (2007).

[3] Le Marchand, L. et al. Combined effects of well-done red meat, smoking, and rapid N-acetyltransferase 2 and CYP1A2 phenotypes in increasing colorectal cancer risk. Cancer Epidemiol Biomarkers Prev. 10, 1259-1266 (2001).

[4] Wu, A. et al. in preparation for publication.

[5] Tang, K. et al. Chip-based genotyping by mass spectrometry. Proc. Natl. Acad. Sci. USA 96, 10016–10020 (1999).

[6] Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 38, 904-909 (2006).

Supplementary Table 1: Association of rs6983267 and rs10090154 with colorectal cancer in the Hawaii and Los Angeles population-based case-control studies

	Hawaii Case-Control Study				Los Angeles Case-Control Study
Marker / Position (Mb)	**Japanese Americans (211 cases / 367 controls)**	**European Americans (116 cases / 158 controls)**	**P_{Het} [a]**	**Pooled OR (95% CI) [b]**	**European Americans (356 cases / 413 controls)**
rs6983267 128,482,487	1.32 (1.03-1.70) 32%	1.14 (0.81-1.60) 48%	0.51	1.26 (1.03-1.54) P=0.03	1.07 (0.88-1.31) P=0.48
rs10090154 128,601,319	1.07 (0.78-1.48) 16%	1.03 (0.57-1.85) 9%	0.91	1.06 (0.80-1.41) P=0.68	1.49 (1.06-2.10) 0.02

Each square gives odds ratios (and 95% confidence intervals) for allele dosage effects along with the risk allele frequency in controls. [a] P-value testing for heterogeneity of allelic effects across populations in the Hawaii study; [b] OR adjusted for ethnic population and gender. SNPs rs6983267 and rs10090154 were genotyped successfully for ≥98% of cases and controls in the Hawaii study and ≥97% of cases and controls in the Los Angeles study.

Supplementary Table 2: Association of rs6983267 with colorectal cancer risk.

Population (cases/controls)	TT	GT	GG	Effect per G allele	P-value
African Americans (217 / 1,049)	Ref. 3%	1.11 (0.36-3.39) 25%	1.57 (0.52-4.71) 72%	1.37 (0.98-1.91)	0.065
Japanese Americans (592 / 1,564)	Ref. 46%	1.11 (0.90-1.36) 43%	1.49 (1.10-2.02) 11%	1.19 (1.03-1.37)	0.016
Native Hawaiians (61 / 347)	Ref. 48%	1.25 (0.65-2.40) 43%	2.95 (1.19-7.29) 9%	1.59 (1.02-2.47)	0.039
Latinos (251 / 1,007)	Ref. 16%	1.03 (0.67-1.59) 49%	1.45 (0.94-2.25) 35%	1.26 (1.02-1.55)	0.032
European Americans (686 / 1,544)	Ref. 25%	0.92 (0.73-1.17) 51%	1.32 (1.01-1.72) 24%	1.16 (1.02-1.33)	0.029
All populations (1,807 / 5,511)	Ref.	1.04 (0.90-1.20)	1.47 (1.25-1.74)	1.22 (1.12-1.32)	4.4×10^{-6}
Males (817 / 2,710)	Ref.	1.10 (0.90-1.35)	1.50 (1.18-1.91)	1.22 (1.08-1.38)	1.2×10^{-3}
Females (990 / 2,801)	Ref.	0.98 (0.82-1.22)	1.43(1.14-1.79)	1.20(1.07-1.35)	1.7×10^{-3}
Localized disease (902 / 5,511)	Ref.	1.13 (0.94-1.36)	1.53 (1.23-1.90)	1.24 (1.11-1.38)	1.3×10^{-4}
Regional or Distant disease (874 / 5,511)	Ref.	0.95 (0.78-1.15)	1.42 (1.14-1.76)	1.20 (1.07-1.34)	1.6×10^{-3}
Colon (1,359 / 5,511)	Ref.	1.03 (0.88-1.21)	1.48 (1.24-1.78)	1.22 (1.11-1.34)	2.4×10^{-5}
Rectum (448 / 5,511)	Ref.	1.08 (0.84-1.37)	1.44 (1.08-1.93)	1.20 (1.03-1.39)	0.016
Age ≤ 67 years (median) (850 / 2,880)	Ref.	1.09 (0.89-1.34)	1.32 (1.04-1.68)	1.15 (1.02-1.30)	0.024
Age > 67 years (957 / 2,631)	Ref.	1.01 (0.83-1.23)	1.65 (1.32-2.07)	1.28 (1.14-1.44)	2.5×10^{-5}
Family History of Colorectal Cancer in a First-Degree Relative (241 / 497)	Ref.	0.93 (0.62-1.38)	1.94 (1.19-3.14)	1.36 (1.07-1.74)	0.013
No Family History of Colorectal Cancer in a First-Degree Relative (1,397 / 4,451)	Ref.	1.04 (0.89-1.22)	1.41 (1.17-1.70)	1.19 (1.08-1.31)	2.9×10^{-4}

All the samples from the study are combined for this analysis. Each cell in the table gives odds ratios (and 95% confidence intervals) for allele dosage effects estimated using unconditional logistic regression, and frequency of each genotype class in controls for each ethnic group. ORs are adjusted for gender in ethnic-specific analysis, ethnicity in gender-specific analysis and both gender and ethnicity in pooled and stratified analyses. ORs estimated including the African Americans, Native Hawaiians and Latinos are also adjusted for genome-wide European ancestry. Genotyping was successful for 97.5% of cases and 94.8% of controls. P-value testing for heterogeneity of effects among cases and controls: ethnicity (P=0.63), gender (P=0.82), age (P=0.26) and family history (P=0.49). Information about first degree family history of colon cancer was available for 91% of cases and 90% controls. P-value testing for heterogeneity of effects in case-only analyses: localized vs regional/distant disease (P=0.70), and colon vs rectum (P=0.74).

Supplementary Table 3. Allele frequencies and associations of 82 SNPs in Region 3 (128.47-128.54 Mb) with colorectal cancer risk in the MEC populations (1,088 cases and 1,823 controls).

Cases/Controls: African Americans, 205/411; Japanese Americans, 374/410; Native Hawaiians, 60/239; Latinos, 243/426; European Whites, 206/337.

			Allele Frequency in Controls[b]					Chi Square						P-value	P-value	P-value
SNP ID[a]	Position	Call Rate	AA	JA	NH	LA	EA	AA	JA	NH	LA	EA	Pooled	Pooled	Adj. for rs6983267	Adj. for rs10808556
rs10505479	128470917	0.985	0.220	0.000	0.000	0.016	0.005	0.42			0.04	0.33	0.61	0.44	0.58	0.54
rs1562871	128470954	0.986	0.188	0.455	0.171	0.298	0.186	0.62	0.69	0.06	0.03	1.28	0.00	0.96	0.26	0.42
rs10505478	128471736	0.995	0.056	0.000	0.000	0.000	0.000	0.71					0.48	0.49	0.47	0.38
rs12549845	128472089	0.991	0.137	0.063	0.082	0.247	0.254	0.31	0.21	0.91	0.41	0.34	0.36	0.55	0.41	0.61
rs10441525	128472135	0.968	0.099	0.132	0.060	0.262	0.137	0.63	3.22	0.44	0.23	0.50	0.03	0.87	0.37	0.93
rs7844673	128472696	0.974	0.106	0.176	0.085	0.034	0.072	0.20	1.09	0.36	1.41	0.05	0.01	0.94	0.94	0.79
rs10956365	128473069	0.988	0.074	0.050	0.217	0.117	0.156	0.26	0.48	0.21	2.37	0.40	0.03	0.86	0.62	0.48
rs7830548	128473097	0.980	0.228	0.000	0.002	0.017	0.005	0.01		0.0002	0.003	0.0004	0.04	0.84	0.55	0.63
rs16902147	128474254	0.988	0.229	0.001	0.004	0.026	0.015	1.23	0.001	0.35	1.86	1.22	0.01	0.92	0.88	0.84
rs10505477	128476625	0.976	0.209	0.682	0.712	0.416	0.505	7.39	1.94	3.73	3.70	0.98	14.38	1.5x10-4	0.18	0.73
rs12334317	128477246	0.974	0.184	0.076	0.102	0.120	0.137	0.20	7.88	2.30	0.54	0.31	3.67	0.06	0.78	0.69
rs10505476	128477298	0.990	0.634	0.184	0.144	0.179	0.237	4.00	1.70	2.19	0.86	0.26	8.37	3.8x10-3	0.47	0.45
rs11985829	128478414	0.982	0.219	0.180	0.207	0.272	0.281	1.89	10.07	3.63	1.29	1.20	15.41	8.6x10-5	0.06	0.46
rs10808555	128478693	0.978	0.346	0.192	0.212	0.286	0.297	1.85	9.13	3.86	1.52	1.71	15.06	1.0x10-4	0.07	0.45
rs12682374	128480130	0.990	0.174	0.682	0.716	0.408	0.505	4.40	2.30	5.59	3.86	1.44	14.18	1.7x10-4	0.32	0.77
rs10505475	128480639	0.978	0.094	0.106	0.031	0.023	0.052	0.05	1.48	0.02	0.87	0.05	0.04	0.84	0.26	0.06
rs10808556	128482329	0.978	0.668	0.302	0.247	0.331	0.390	4.69	3.96	4.68	8.28	1.61	22.44	2.2x10-6	7.4x10-3	-
rs6983267	128482487	0.976	0.144	0.687	0.721	0.403	0.502	1.24	3.29	5.74	4.44	1.21	13.46	2.4x10-4	-	0.86
rs7013278	128484074	0.979	0.529	0.196	0.220	0.309	0.338	4.40	8.21	4.28	4.48	1.13	20.90	4.8x10-6	9.9x10-3	0.25
rs7014039	128484533	0.986	0.061	0.000	0.000	0.006	0.000	0.19			1.22		0.51	0.48	0.60	0.72
rs10505474	128486686	0.977	0.775	0.302	0.250	0.343	0.407	2.70	3.70	4.38	5.61	0.38	15.19	9.7x10-5	0.13	0.18
rs13273004	128488376	0.987	0.037	0.000	0.000	0.000	0.002	2.03				0.0003	2.24	0.13	0.14	0.22
rs11986916	128488689	0.994	0.001	0.118	0.032	0.004	0.002	1.92	0.05	0.42	0.0003	0.0003	0.02	0.88	0.61	0.42
rs2060776	128489299	0.986	0.704	0.302	0.252	0.332	0.392	4.40	3.54	4.28	6.32	1.21	19.18	1.2x10-5	2.0x10-2	0.31
rs13248944	128489740	0.986	0.015	0.106	0.030	0.016	0.051	0.06	1.51	0.10	0.77	0.03	0.35	0.56	0.13	3.5x10-2
rs4871788	128490967	0.973	0.486	0.299	0.243	0.317	0.391	1.92	3.55	4.77	4.77	0.71	15.01	1.1x10-4	0.09	0.37
rs10956369	128492999	0.979	0.644	0.297	0.250	0.323	0.391	2.60	4.36	4.35	5.46	0.49	16.31	5.4x10-5	0.08	0.34
rs7014346	128493974	0.990	0.410	0.192	0.219	0.300	0.335	2.24	8.96	4.21	4.79	0.78	19.03	1.3x10-5	1.6x10-2	0.31
Broad11938345	128496959	0.992	0.052	0.000	0.000	0.007	0.002	0.17			2.35	0.12	1.34	0.25	0.29	0.48
rs4871789	128497243	0.967	0.734	0.303	0.288	0.384	0.483	2.22	3.59	3.64	3.68	1.47	15.46	8.4x10-5	0.05	0.56
Broad11938400	128497903	0.982	0.034	0.000	0.000	0.001	0.002	0.89			0.12	0.0003	0.83	0.36	0.23	0.62
rs6470509	128498701	0.955	0.471	0.224	0.176	0.267	0.371	1.91	0.07	2.28	7.20	0.03	7.40	6.5x10-3	0.16	0.85
rs6470510	128498842	0.989	0.218	0.120	0.118	0.184	0.212	1.59	1.64	1.86	9.30	0.16	11.25	8.0x10-4	0.05	0.26
rs7842552	128500876	0.987	0.382	0.108	0.186	0.198	0.261	2.85	6.93	0.24	0.01	0.57	0.76	0.38	0.47	0.25
rs12375310	128501388	0.972	0.276	0.225	0.196	0.207	0.273	0.60	0.60	0.59	1.80	0.75	0.03	0.85	0.39	0.11
rs9297756	128509349	0.981	0.189	0.134	0.153	0.129	0.128	1.02	3.56	0.03	0.92	0.82	0.44	0.51	0.55	0.37
rs6995633	128509833	0.982	0.027	0.068	0.087	0.222	0.076	0.33	0.15	1.61	0.23	0.95	0.01	0.92	0.23	0.59
rs6999789	128510043	0.983	0.087	0.014	0.024	0.079	0.084	1.17	2.51	1.05	0.91	0.37	1.86	0.17	0.49	0.06
rs7000448	128510352	0.978	0.627	0.251	0.222	0.299	0.363	0.40	0.02	1.05	0.24	1.03	0.95	0.33	0.64	0.79
rs6982665	128510403	0.973	0.378	0.233	0.205	0.209	0.269	0.94	0.03	0.10	1.13	1.40	0.01	0.92	0.26	0.08
rs7357486	128510805	0.972	0.310	0.233	0.202	0.206	0.275	1.08	0.13	0.28	2.32	0.47	0.15	0.69	0.15	3.9x10-2
Broad11938551	128511857	0.981	0.051	0.000	0.000	0.001	0.000	0.05			2.12		0.03	0.86	0.99	0.91
Broad11938559	128512126	0.984	0.254	0.009	0.019	0.090	0.084	2.62	2.21	1.57	0.68	0.65	4.35	3.7x10-2	0.24	1.3x10-2
rs11506132	128518885	0.971	0.345	0.211	0.228	0.288	0.359	0.66	1.36	0.85	0.19	1.41	3.00	0.08	0.34	0.22
rs13280578	128519269	0.996	0.152	0.001	0.002	0.007	0.002	1.67	0.01	0.0001	0.37	0.0003	1.79	0.18	0.31	0.36
rs6470512	128519904	0.963	0.324	0.210	0.226	0.276	0.355	0.66	1.03	1.12	0.08	1.94	3.41	0.06	0.28	0.21
Broad11938876	128523204	0.981	0.022	0.001	0.002	0.004	0.000	0.01	0.0007	0.0002	0.01		0.02	0.88	0.95	0.85
rs6981397	128524836	0.977	0.146	0.704	0.621	0.374	0.309	0.76	1.09	0.33	0.28	1.39	1.36	0.24	0.75	0.83
Broad11938930	128524921	0.989	0.054	0.000	0.002	0.005	0.000	2.97		0.0002	0.54		4.92	2.7x10-2	3.0x10-3	3.9x10-2
rs10101741	128525201	0.971	0.573	0.206	0.240	0.298	0.363	1.00	0.98	0.75	0.001	2.79	5.32	2.1x10-2	0.23	0.14
rs28623168	128525530	0.973	0.643	0.208	0.244	0.298	0.373	0.70	1.09	0.28	0.10	1.25	4.34	3.7x10-2	0.30	0.21
rs7012462	128526872	0.982	0.288	0.264	0.289	0.438	0.414	0.02	2.54	0.02	0.85	0.01	0.01	0.92	0.35	0.33
Broad11939006	128526977	0.969	0.074	0.006	0.033	0.060	0.072	0.02	0.32	0.17	5.92	1.21	3.74	0.05	0.21	6.8x10-3
Broad11939015	128527247	0.975	0.380	0.016	0.082	0.163	0.216	1.98	0.04	0.40	0.78	1.10	3.86	0.05	0.10	1.8x10-2
rs10109622	128527333	0.977	0.540	0.020	0.080	0.180	0.219	0.92	0.47	0.11	0.78	1.94	3.52	0.06	0.20	0.05
rs6996874	128527491	0.986	0.438	0.021	0.082	0.160	0.228	0.90	0.74	0.03	2.70	0.81	3.87	4.9x10-2	0.08	1.4x10-2
Broad11939081	128528520	0.993	0.062	0.000	0.000	0.002	0.002	0.08			1.04	0.0003	0.41	0.52	0.63	0.45
rs6470517	128529586	0.976	0.173	0.011	0.048	0.099	0.163	0.03	0.002	0.15	0.16	0.28	0.20	0.65	0.46	0.67
rs7841228	128530060	0.990	0.174	0.083	0.129	0.300	0.264	0.62	0.28	0.002	0.37	0.22	1.12	0.29	0.17	0.27
rs10094059	128530789	0.976	0.344	0.077	0.135	0.306	0.268	0.23	0.86	0.04	0.03	0.11	0.00	0.94	0.80	0.95
rs10719294	128531078	0.990	0.222	0.080	0.151	0.341	0.341	0.24	0.50	0.57	0.24	0.22	0.25	0.62	0.33	0.36
Broad11939263	128531943	0.979	0.091	0.009	0.031	0.074	0.075	0.05	2.15	0.23	1.52	0.43	0.71	0.40	0.74	0.12
Broad11939350	128533803	0.973	0.424	0.258	0.341	0.504	0.558	0.16	2.15	0.05	0.07	0.02	1.01	0.32	0.53	0.33
Broad11939378	128534131	0.929	0.455	0.591	0.554	0.322	0.178	2.60	0.28	0.15	1.31	0.12	0.00	0.95	0.59	0.67
rs28451337	128534369	0.981	0.095	0.148	0.112	0.168	0.267	1.52	0.77	0.71	1.18	0.004	0.18	0.67	0.48	0.28
rs9643221	128534669	0.983	0.462	0.593	0.535	0.329	0.177	1.56	0.59	0.33	0.83	0.01	0.03	0.87	0.60	0.65
rs7836345	128534862	0.973	0.511	0.628	0.553	0.388	0.273	0.99	0.09	0.94	0.86	0.07	0.05	0.82	0.41	0.58
Broad11939524	128535073	0.982	0.152	0.000	0.000	0.012	0.002	0.68			0.01	0.11	0.52	0.47	0.22	0.24
Broad11939636	128535734	0.981	0.067	0.001	0.000	0.004	0.000	0.05	0.0007		0.0003	0.0003	0.08	0.78	0.93	0.95
rs10107830	128535806	0.971	0.247	0.165	0.185	0.399	0.450	0.00	1.14	1.08	0.09	1.85	0.00	0.96	0.75	0.77
rs7841264	128535996	0.976	0.169	0.595	0.542	0.298	0.165	0.94	0.50	0.11	1.26	0.29	0.17	0.68	0.70	0.83
rs7017671	128536936	0.990	0.530	0.437	0.302	0.534	0.536	2.06	2.04	0.63	1.65	1.58	1.78	0.18	0.07	0.06
rs10099905	128537116	0.989	0.243	0.128	0.108	0.281	0.225	0.96	0.10	2.61	0.01	2.26	1.03	0.31	0.17	0.25
rs10100179	128537265	0.981	0.028	0.000	0.000	0.001	0.003	0.06			0.0003	0.0006	0.28	0.60	0.57	0.50
Broad11939763	128537526	0.979	0.219	0.014	0.048	0.072	0.109	1.68	0.71	0.74	0.11	0.02	1.21	0.27	0.31	0.41
rs3999771	128538680	0.974	0.282	0.003	0.049	0.140	0.237	1.04	1.52	0.78	0.17	1.90	0.15	0.70	0.54	0.74
rs6990480	128538890	0.979	0.192	0.130	0.195	0.079	0.089	1.88	4.26	1.19	0.08	6.33	0.74	0.39	0.57	0.60
rs10956372	128539438	0.975	0.758	0.382	0.408	0.305	0.315	0.70	0.42	0.19	0.13	0.05	0.01	0.91	0.89	0.77
Broad11939966	128539443	0.993	0.076	0.000	0.002	0.005	0.002	0.0001		0.0001	0.17	0.0003	0.03	0.87	0.91	0.90
Broad11939976	128539576	0.984	0.043	0.000	0.000	0.004	0.000	1.34			0.0004		1.23	0.27	0.26	0.35
Broad11939980	128539731	0.970	0.043	0.000	0.000	0.002	0.000	0.02			1.13		0.22	0.64	0.68	0.78
rs7825928	128539790	0.976	0.200	0.503	0.559	0.599	0.628	0.27	0.69	0.83	0.000004	0.23	0.18	0.67	0.97	0.99

[a] SNPs labeled "BroadXXXXXXXX" were identified by resequencing as described in ref. 2.

[b] Minor allele based on all populations combined; AA, African Americans; JA, Japanese Americans; NH, Native Hawaiians; LA, Latinos; EA, European Whites.

Supplementary Table 4: Association of rs10808556 and rs7013278 with colorectal cancer risk in the MEC (1,088 cases, 1,823 controls).

Marker / Position (Mb)	African Americans[a] (205 cases / 411 controls)	Japanese Americans (374 cases / 410 controls)	Native Hawaiians[a] (60 cases / 239 controls)	Latinos[a] (243 cases / 426 controls)	European Americans (206 cases / 337 controls)	Colorectal Cancer Pooled OR[b] (95% CI)
rs10808556 128,482,329	1.37 (1.03-1.82) 67%	1.23 (1.00-1.51) 30%	1.63 (1.03-2.57) 25%	1.44 (1.14-1.84) 33%	1.18 (0.91-1.53) 39%	1.32 (1.18-1.49) $P=2.2x10^{-6}$
rs7013278 128,484,074	1.32 (1.02-1.72) 53%	1.41 (1.11-1.77) 20%	1.62 (1.02-2.58) 22%	1.33 (1.04-1.69) 31%	1.16 (0.89-1.51) 34%	1.32 (1.17-1.49) $P=4.8x10^{-6}$

Each cell gives odds ratios (and 95% confidence intervals) for allele dosage effects along with the risk allele frequency in controls. All ORs are adjusted for gender. [a] OR also adjusted for genome-wide European ancestry (African Americans, Native Hawaiians and Latinos). [b] OR adjusted for population and genome-wide European ancestry (African Americans, Native Hawaiians and Latinos). P-value testing for heterogeneity of effects: rs10808556, P=0.53; rs7013278, P=0.73. SNPs rs10808556 and rs7013278 were genotyped successfully for ≥98% of cases and ≥97% of controls.

Supplementary Table 5: Linkage disequilibrium between the 9 risk variants evaluated in this study at 8q24 (r^2 upper, |D'| lower diagonal)

Polymorphism (Position) Region	Population	rs13254738	rs6983561	Broad11934905	rs13281615	rs10808556	rs6983267	rs7013278	rs7000448	rs10090154
rs13254738 (128,173,525) Region 2	AA		0.19	0.02	0	0	0.02	0	0	0
	JA		0.11	-	0	0	0	0	0	0.01
	NH	NA	0.11	-	0.02	0.02	0.02	0.2	0	0
	LA		0	-	0	0	0	0	0.01	0
	EA		0	-	0	0	0.1	0	0.02	0
rs6983561 (128,176,062) Region 2	AA	0.60		0.04	0	0.01	0.05	0.01	0.01	0
	JA	1.0		-	0	0	0	0	0	0
	NH	0.87	NA	-	0.01	0	0	0	0	0.02
	LA	0.04		-	0	0.01	0	0	0.02	0
	EA	0.58		-	0	0	0	0	0	0
Broad11934905 (128,200,991) Region 2	AA	1.0	1.0		0	0	0	0	0	0
	JA	-	-		-	-	-	-	-	-
	NH	-	-	NA	-	-	-	-	-	-
	LA	-	-		-	-	-	-	-	-
	EA	-	-		-	-	-	-	-	-
rs13281615 (128,424,800) (breast cancer)	AA	0.03	0.03	0.42		0	0	0	0	0
	JA	0.05	0	-		0	0	0	0	0
	NH	0.15	0.33	-	NA	0.04	0.05	0.03	0	0.01
	LA	0.08	0.05	-		0.02	0	0	0.06	0
	EA	0.06	0.16	-		0	0	0	0	0
rs10808556 (128,482,329) Region 3	AA	0.09	0.20	0.13	0.06		0.33	0.55	0.02	0
	JA	0.03	0.05	-	0.18		0.94	0.60	0.25	0
	NH	0.24	0.02	-	0.25	NA	0.83	0.84	0.12	0.01
	LA	0.10	0.43	-	0.13		0.34	0.87	0.04	0
	EA	0.07	0.09	-	0		0.63	0.80	0.02	0
rs6983267 (128,482,487) Region 3	AA	0.30	0.69	0.31	0.15	0.99		0.18	0.11	0.01
	JA	0.02	0.07	-	0.16	0.99		0.57	0.25	0
	NH	0.19	0.18	-	0.25	1.0	NA	0.70	0.20	0.01
	LA	0.02	0.60	-	0.09	0.99		0.30	0.06	0
	EA	0.16	0.19	-	0.09	0.99		0.51	0.11	0
rs7013278 (128,484,074) Region 3	AA	0.05	0.13	0.57	0.06	1.0	0.98		0	0
	JA	0	0	-	0.14	1.0	1.0		0.04	0
	NH	0.26	0.04	-	0.24	1.0	1.0	NA	0.07	0
	LA	0.09	0.29	-	0.10	1.0	0.99		0.02	0
	EA	0.06	0.04	-	0.04	1.0	1.0		0	0
rs7000448 (128,510,352) Region 3	AA	0.09	0.18	0.26	0.03	0.16	0.66	0.09		0
	JA	0.12	0.13	-	0.05	0.60	0.62	0.22		0
	NH	0.11	0	-	0.07	0.39	0.55	0.27	NA	0
	LA	0.18	0.49	-	0.30	0.24	0.48	0.15		0
	EA	0.15	0.01	-	0.13	0.13	0.45	0.03		0.03
rs10090154 (128,601,319) Region 1	AA	0.02	0.15	0.15	0.18	0.25	0.67	0.16	0.06	
	JA	0.18	0.19	-	0.05	0.04	0.01	0.09	0.24	
	NH	0.03	0.15	-	0.31	0.35	0.38	0.34	0.02	NA
	LA	0.13	0.08	-	0.11	0	0.12	0	0.32	
	EA	0.44	0	-	0.13	0.04	0	0.13	0.77	

r^2 and D' values estimated among cases and controls combined using the Haploview program (http://www.broad.mit.edu/mpg/haploview/). LD between the 4 variants in the region fine-mapped (128.47-128.54 Mb) are shown with a heavy black line. AA, African American; JA, Japanese American; NH, Native Hawaiian; LA, Latino; EA, European American