

## A common genetic risk factor for colorectal and prostate cancer

Christopher A Haiman<sup>1</sup>, Loïc Le Marchand<sup>2</sup>, Jennifer Yamamoto<sup>2</sup>, Daniel O Stram<sup>1</sup>, Xin Sheng<sup>1</sup>, Laurence N Kolonel<sup>2</sup>, Anna H Wu<sup>1</sup>, David Reich<sup>3,4</sup> & Brian E Henderson<sup>1</sup>

**Variants on chromosome 8q24 contribute risk for prostate cancer; here, we tested whether they also modulate risk for colorectal cancer. We studied 1,807 affected individuals and 5,511 controls and found that one variant, rs6983267, is also significantly associated with colorectal cancer (odds ratio = 1.22;  $P = 4.4 \times 10^{-6}$ ) and that the apportionment of risk among the variants differs significantly between the two cancers. Comprehensive testing in the region uncovered variants capturing significant additional risk. Our results show that variants at 8q24 have different effects on cancer development that depend on the tissue type.**

We<sup>1,2</sup> and others<sup>3–5</sup> have recently identified multiple independent risk variants for prostate cancer spread over three unlinked regions of chromosome 8 between 128.14 and 128.62 Mb in the human reference sequence (build 35). However, there are no known genes in this span. Motivated by the observation that the region is a frequent site of somatic amplification not only in prostate cancer<sup>6,7</sup> but also in colorectal cancer (CRC)<sup>8,9</sup>, we set out to test whether the genetic variants also modulate risk for CRC. We explicitly considered three models. In the first model, the prostate cancer risk variants at 8q24 do not affect risk for CRC. In the second model, each of the 8q24 variants increases risk of cancer through the same biological mechanism, and the apportionment of risk among the variants is identical for the two cancers. In the third model, variants at 8q24 affect risk of both prostate cancer and CRC, but the relative contributions of the variants differ, indicating etiologic heterogeneity in how the variants contribute to cancer.

To distinguish among these models and to test whether the prostate cancer risk variants increase susceptibility to CRC as well, we genotyped six of the seven variants we previously identified<sup>2</sup> as capturing the risk for prostate cancer at 8q24 (excluding the microsatellite DG8S737-8 allele) in 1,124 individuals with invasive CRC and 4,573 controls. The samples were derived from five populations in the Multiethnic Cohort Study (MEC): African Americans, Japanese Americans, Native Hawaiians, Latinos and European Americans (**Supplementary Methods** and **Supplementary Fig. 1** online). We detected a highly significant association of CRC risk

with rs6983267 ( $P = 2.9 \times 10^{-5}$ ) and suggestive associations with rs10090154 ( $P = 0.059$ ) and rs7000448 ( $P = 0.068$ ) (**Table 1**). SNPs rs6983267 and rs7000448 were correlated (with an  $r^2$  range of 0.09–0.27 in the five ethnic groups), and the association with rs7000448 can be explained by its correlation with rs6983267. We did not observe any association with rs13254738, rs6983561 or Broad11934905 (**Table 1**).

To follow up on these results, we studied 683 additional individuals with CRC and 938 controls from two studies of Japanese Americans and European Americans (**Supplementary Methods**). The rs6983267 association was replicated ( $P = 0.03$ ) in one study but not in the other ( $P = 0.48$ ) (**Supplementary Table 1** online). In a pooled analysis of all studies (1,807 affected individuals and 5,511 controls; **Supplementary Table 2** online), the association with rs6983267 was highly significant (odds ratio (OR) per allele = 1.22 (95% confidence interval, 1.12–1.32);  $P = 4.4 \times 10^{-6}$ ). We also observed a weakly significant association with rs10090154 (OR = 1.16 (1.04–1.31);  $P = 0.01$ ), an association that should be considered provisional pending replication. We obtained informed consent from all participants. The three studies were approved by ethics review boards at the University of Southern California and the University of Hawaii.

We performed additional analyses focusing on rs6983267 to better understand the risk for CRC at this locus. We first tested whether the risk was consistent with an allele dosage effect. Notably, a nonmultiplicative risk model fit the data better ( $P = 7.7 \times 10^{-3}$ ) than a model in which the risk for homozygotes is the square of the risk for heterozygotes (OR for heterozygotes = 1.04 (0.90–1.20); OR for homozygotes = 1.47 (1.25–1.74)). The effect of rs6983267 on CRC risk was consistent across the five populations ( $P$  for heterogeneity = 0.63) and did not show any differences when stratified by sex, disease stage, tumor site, age of diagnosis and history of CRC cancer in a first-degree relative (**Supplementary Table 2**). In the MEC, we also did not detect any interaction with nongenetic CRC risk factors, body mass index, smoking, aspirin use, alcohol consumption or estrogen therapy in women ( $P > 0.05$ ).

We estimate that the risk allele at rs6983267 has a large effect on population prevalence of CRC because of its high frequency (from 31% in Native Hawaiians to 85% in African Americans). If it were possible to reduce the risk to a similar level as found in individuals homozygous for the protective allele, we estimate that rates of CRC would decrease by 42% in African Americans, 11% in Japanese Americans, 28% in Native Hawaiians, 25% in Latinos and 14% in European Americans (**Supplementary Methods**). However, because of the modest ~1.22-fold increased risk per allele, the power to predict whether any particular individual is diagnosed with colorectal cancer is low; we calculate that it explains no more than 0.9% to 1.8% of the

<sup>1</sup>Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90089, USA. <sup>2</sup>Epidemiology Program, Cancer Research Center of Hawaii, University of Hawaii, Honolulu, Hawaii 96813, USA. <sup>3</sup>Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. <sup>4</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. Correspondence should be addressed to C.A.H. (haiman@usc.edu).

Received 28 March; accepted 8 June; published online 8 July 2007; doi:10.1038/ng2098

**Table 1 Association of SNPs on 8q24 with CRC and comparison with prostate cancer in the Multiethnic Cohort**

Marker, position (in Mb) and region	Colorectal cancer					Prostate cancer	
	African Americans <sup>a</sup> (217 cases, 1,049 controls)	Japanese Americans (381 cases, 1,197 controls)	Native Hawaiians <sup>a</sup> (61 cases, 347 controls)	Latinos <sup>a</sup> (251 cases, 1,007 controls)	European Americans (214 cases, 973 controls)	Pooled OR (95% c.i.) <sup>b</sup>	Pooled OR (95% c.i.) <sup>c</sup>
<b>rs13254738</b>	0.91	0.96	1.17	0.90	1.11	0.97	1.32
128173525 (region 2)	(0.73–1.12) 58%	(0.81–1.13) 63%	(0.77–1.76) 48%	(0.73–1.09) 51%	(0.89–1.38) 32%	(0.88–1.07)	(1.22–1.43)
<b>rs6983561</b>	0.96	0.91	1.43	1.99	1.06	1.03	1.53
128176062 (region 2)	(0.77–1.19) 41%	(0.73–1.14) 16%	(0.78–2.62) 13%	(1.31–3.03) 3%	(0.62–1.80) 4%	(0.90–1.18)	(1.38–1.70)
<b>Broad11934905<sup>d</sup></b>	1.08	–	–	–	–	1.08	2.22
128200991 (region 2)	(0.55–2.10) 2%	–	–	<1%	–	(0.55–2.10)	(1.45–3.40)
<b>rs6983267</b>	1.37	1.13	1.59	1.26	1.28	1.25	1.20
128482487 (region 3)	(0.98–1.91) 85%	(0.96–1.34) 32%	(1.02–2.47) 31%	(1.02–1.55) 60%	(1.03–1.58) 49%	(1.12–1.38)	(1.11–1.30)
<b>rs7000448</b>	1.17	1.04	1.31	1.01	1.16	1.10	1.26
128510352 (region 3)	(0.93–1.46) 62%	(0.86–1.26) 24%	(0.81–2.11) 22%	(0.81–1.25) 30%	(0.93–1.46) 36%	(0.99–1.22)	(1.15–1.38)
<b>rs10090154</b>	1.17	1.41	0.81	0.97	0.84	1.14	1.45
128601319 (region 1)	(0.89–1.53) 16%	(1.14–1.75) 15%	(0.45–1.46) 18%	(0.66–1.43) 7%	(0.57–1.24) 9%	(1.00–1.31)	(1.30–1.62)

Calculations were based on 1,124 affected individuals and 4,573 controls. Regions are defined based on ref. 2: region 1, 128.54–128.62 Mb; region 2, 128.14–128.28 Mb; region 3, 128.47–128.54 Mb. Each cell gives odds ratios (ORs) (and 95% confidence intervals (c.i.)) for allele dosage effects along with the risk allele frequency in controls. All ORs are adjusted for gender.

<sup>a</sup>OR also adjusted for genome-wide European ancestry (African Americans, Native Hawaiians and Latinos). <sup>b</sup>OR adjusted for population and genome-wide European ancestry (African Americans, Native Hawaiians and Latinos). <sup>c</sup>The ORs for prostate cancer are from Supplementary Table 6 of ref. 2 and are adjusted for population and genome-wide European ancestry (African Americans, Native Hawaiians and Latinos). <sup>d</sup>Data for Broad11934905 is presented for African Americans only, as this is the only population with an appreciable frequency of the risk allele.

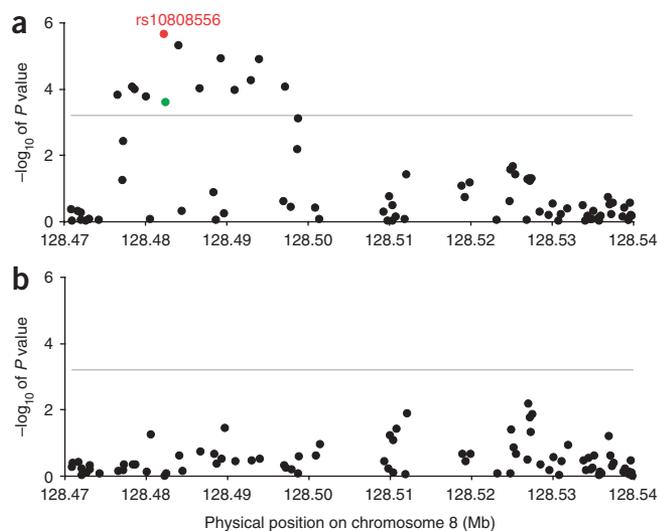
approximately 1.5-fold increased risk to siblings that is empirically observed in these studies.

Motivated by the discovery of an association with CRC at 8q24, we carried out fine mapping across the region in linkage disequilibrium (LD) with rs6983267 (from 128.47–128.54 Mb). We genotyped 82 SNPs capturing  $\geq 92\%$  of genetic variants of  $> 5\%$  frequency in the three HapMap populations (Supplementary Fig. 2 online) in 1,088 individuals with CRC and 1,823 controls. We identified two SNPs, rs10808556 and rs7013278 (158 and 1,587 bp from rs6983267, respectively), that are nominally more associated with risk (Supplementary Tables 3 and 4 online). These SNPs were strongly correlated with each other across all populations ( $r^2 = 0.55\text{--}0.87$ ), as well as with rs6983267, except among African Americans and Latinos ( $r^2 \leq 0.34$ ; Supplementary Table 5 online). The implication is that rs6983267 may not be sufficient to capture all risk at the locus. When this is formally tested, SNPs rs10808556 and rs7013278 capture significant additional risk for CRC after controlling for rs6983267 ( $P = 7.4 \times 10^{-3}$  and  $P = 9.9 \times 10^{-3}$ ). In these analyses, rs6983267 is no longer significant ( $P > 0.27$ ). No other SNPs were significant after conditioning on rs10808556 and correcting for 82 hypotheses tested (Fig. 1 and Supplementary Table 3).

The association of rs6983267 and closely linked variants with CRC and prostate cancer signifies a common biological mechanism for cancer risk at 8q24. However, the failure to detect CRC associations

with the five other variants (Table 1), is notable. Under the simple model in which variants at 8q24 contribute risk equally for prostate and CRC, we would expect the apportionment of risk among the alleles to be the same, even if the overall penetrance is different for the two cancers. However, we can reject this hypothesis with high statistical significance ( $P = 6.3 \times 10^{-6}$ ; Supplementary Methods). Two other lines of evidence support biological differences among the risk alleles at 8q24. First, at SNP rs13281615 at 128.42 Mb, recently associated with breast cancer risk<sup>10</sup>, we did not observe any association with either CRC ( $P = 0.59$ ) or prostate cancer ( $P = 0.46$ )<sup>2</sup>. Second, variants in the 128.14–128.28 Mb region, but not elsewhere in 8q24, have been associated with age of prostate cancer diagnosis in two independent large-scale studies<sup>2,4</sup>. This again suggests heterogeneity in how 8q24 alleles contribute to risk in different tissue types.

The risk variants we have identified for CRC in this region of 8q24 are all located between 128.47 and 128.54 Mb, a region containing no known genes. The interval does, however, contain highly conserved segments of DNA and a processed pseudogene, *POU5F1P1*. This gene is a retrotransposed copy of the POU-domain transcription factor Oct4 (ref. 11), which is a key regulator of stem cell pluripotency<sup>12</sup>. Many Oct4 pseudogenes, including *POU5F1P1*, are expressed in cancerous tissues, including colon<sup>13</sup>, although their mechanistic role in cancer is unknown. This region also contains two human endogenous retrovirus (HERV-H) sequences, which occur hundreds of



**Figure 1** Fine mapping in the region 128.47–128.54 Mb of 8q24. We present data for 82 SNPs that we genotyped in 1,088 individuals with CRC and 1,823 controls. These SNPs capture  $\geq 92\%$  of SNPs of  $> 5\%$  minor allele frequency at  $r^2 > 0.8$  in HapMap populations (**Supplementary Fig. 2**). (a) The most strongly associated SNP is rs10808556 (red),  $P = 2.2 \times 10^{-6}$ , which is substantially more associated than rs6983267 (green), the SNP that initially led us to the region ( $P = 2.4 \times 10^{-4}$ ). (b) Controlling for rs10808556 in a logistic regression analysis, no SNP is significant at  $P < 0.05$  after correcting for 82 hypotheses tested (threshold shown by a light gray line).

the Multiethnic Cohort Study was supported by US National Institutes of Health (NIH) grants CA63464 and CA54281. The Hawaii-based case-control study was supported by NIH grants CA60987 and CA72520 from the National Cancer Institute, United States Department of Health and Human Services. The Los Angeles-based case-control study was supported by grant P01 CA17054 from the National Cancer Institute. D.R. is supported by a Burroughs-Wellcome Center Development Award in the Biomedical Sciences.

#### AUTHOR CONTRIBUTIONS

The studies were initiated by B.E.H., L.N.K., L.L.M. and A.H.W. The genotyping was performed under the direction of C.A.H., L.L.M. and D.R. Covariate datasets were generated by J.Y. under the direction of L.L.M. The statistical analysis was performed by C.A.H. with the assistance of D.R., X.S. and D.O.S. The manuscript was written by C.A.H. with the assistance of D.R. and all co-authors.

#### COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at <http://www.nature.com/naturegenetics>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

times in the genome, including on either side of the *MYC* gene at 8q24, and may serve as a site for homologous recombination in somatic tissue. Based on the proximity to *MYC* ( $\sim 330$  kb), it is reasonable to speculate that one or more genetic variant(s) at 8q24 might alter expression of *MYC* either through modifying regulatory sequences or by increasing the propensity for somatic amplification in this region<sup>6–9</sup>. Discovery and association testing of all common and rare variation at 8q24 for association with these and other cancers, as well as functional studies of *MYC* and other nearby genes, will be required to understand the full genotype-phenotype correlation at 8q24 and the mechanism(s) whereby these changes result in the observed cancer-specific risks.

Note: Supplementary information is available on the Nature Genetics website.

#### ACKNOWLEDGMENTS

We thank the men and women who participated in these studies. We are grateful to L. Pooler, D. Wong, J. Neubauer, C. Schirmer and A. Waliszewska for assistance with genotyping, and to D. Altshuler, B. Berman, S. Greenway, M. Freedman, S. Myers, N. Patterson, J. Seidman for discussions and comments on the manuscript. The collection and genotyping of samples from

1. Freedman, M.L. *et al. Proc. Natl. Acad. Sci. USA* **103**, 14068–14073 (2006).
2. Haiman, C.A. *et al. Nat. Genet.* **39**, 638–644 (2007).
3. Amundadottir, L.T. *et al. Nat. Genet.* **38**, 652–658 (2006).
4. Gudmundsson, J. *et al. Nat. Genet.* **39**, 631–637 (2007).
5. Yeager, M. *et al. Nat. Genet.* **39**, 645–649 (2007).
6. Visakorpi, T. *et al. Cancer Res.* **55**, 342–347 (1995).
7. Sun, J. *et al. Prostate* **67**, 692–700 (2007).
8. Ried, T. *et al. Genes Chromosom. Cancer* **15**, 234–245 (1996).
9. Douglas, E.J. *et al. Cancer Res.* **64**, 4817–4825 (2004).
10. Easton, D.F. *et al. Nature* **447**, 661–678 (2007).
11. Scholer, H.R. *et al. Nature* **344**, 435–439 (1990).
12. Nichols, J. *et al. Cell* **95**, 379–391 (1998).
13. Suo, G. *et al. Biochem. Biophys. Res. Commun.* **337**, 1047–1051 (2005).