Cancer Science
Japanese Cancer Association JCA
Open Access

Japanese genome-wide association study identifies a significant colorectal cancer susceptibility locus at chromosome 10p14

Yusuke Takahashi,[1,2] Keishi Sugimachi,[1,3] Ken Yamamoto,[4] Atsushi Niida,[5] Teppei Shimamura,[5,6] Tetsuya Sato,[7] Masahiko Watanabe,[8] Junichi Tanaka,[9] Shinei Kudo,[9] Kenichi Sugihara,[10] Kazuo Hase,[11] Masato Kusunoki,[12] Kazutaka Yamada,[13] Yasuhiro Shimada,[14] Yoshihiro Moriya,[14] Yutaka Suzuki,[15] Satoru Miyano,[5] Masaki Mori[2] and Koshi Mimori[1]

[1]Department of Surgery, Kyushu University Beppu Hospital, Beppu; [2]Department of Gastroenterological Surgery, Osaka University, Suita; [3]Department of Hepatobiliary-pancreatic surgery, National Kyushu Cancer Center, Fukuoka; [4]Department of Medical Biochemistry, Kurume University, Kurume; [5]Laboratory of DNA Information Analysis, Human Genome Center Institute of Medical Science, University of Tokyo, Tokyo; [6]Division of System Biology, Graduate School of Medicine, Nagoya University, Nagoya, Japan; [7]Division of Bioinformatics, Medical Institute of Bioregulation, Kyushu University, Fukuoka; [8]Department of Surgery, Kitazato University, Sagamihara; [9]Digestive Disease Center, Northern Yokohama Hospital, Showa University, Yokohama; [10]Department of Surgery, Tokyo Medical and Dental University, Tokyo; [11]Department of Surgery, National Defense University, Tokorozawa; [12]Department of Surgery, Mie University, Tsu; [13]Department of Surgery, Takano Hospital, Kumamoto; [14]Department of Surgery and Digestive Tract Medicine, National Cancer Center, Tokyo; [15]Department of Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Japan

Key words

Colorectal cancer, genome-wide association study, Japanese, single nucleotide polymorphism, 10p14

Correspondence

Koshi Mimori, Department of Surgery, Kyushu University Beppu Hospital, 4546 Tsurumihara, Beppu 874-0838, Japan.
Tel: +81-977-27-1650, Fax: +81-977-27-1652;
E-mail: kmimori@beppu.kyushu-u.ac.jp
and
Masaki Mori, Department of Gastroenterological Surgery, Graduate School of Medicine, Osaka University, Yamadaoka 2-2, Suita 565-0871, Japan.Tel: +81-6-6879-3251, Fax: +81-6-6879-3259;
E-mail: mmori@gesurg.med.osaka-u.ac.jp

Yusuke Takahashi and Keishi Sugimachi are equally contributed to the manuscript.

Masaki Mori and Koshi Mimori are also equally contributed to the manuscript.

Funding Information

This work was supported in part by the following grants and foundations: CREST, Japan Science and Technology Agency (JST); Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research, grant numbers 15H0912, 15H05707, 21591644, 21791295, 21791297, 215921014 and 21679006; the Funding Program for Next Generation World-Leading Researchers (LS094); New Energy and Industrial Technology Development Organization (NEDO) Technological Development for Chromosome Analysis; and Priority Issue on Post-K computer (hp170227 and hp160219).

Received June 15, 2017; Revised August 19, 2017; Accepted August 29, 2017

Cancer Sci **108** (2017) 2239–2247

doi: 10.1111/cas.13391

Genome-wide association studies are a powerful tool for searching for disease susceptibility loci. Several studies identifying single nucleotide polymorphisms (SNP) connected intimately to the onset of colorectal cancer (CRC) have been published, but there are few reports of genome-wide association studies in Japan. To identify genetic variants that modify the risk of CRC oncogenesis, especially in the Japanese population, we performed a multi-stage genome-wide association study using a large number of samples: 1846 CRC cases and 2675 controls. We identified 4 SNP (rs7912831, rs4749812, rs7898455 and rs10905453) in chromosome region 10p14 associated with CRC; however, there are no coding or non-coding genes within this region of fairly extensive linkage disequilibrium (a 500-kb block) on 10p14. Our study revealed that the 10p14 locus is significantly correlated with susceptibility to CRC in the Japanese population, in accordance with the results of multiple studies in other races.

The rates of mortality and morbidity from colorectal cancer (CRC) have been increasing exponentially in Japan, and CRC is considered a national problem that needs to be solved urgently. Identification of factors involved in the carcinogenesis and progression of CRC would help prevent the occurrence of CRC, as well as improve the clinical outcome of treatment

Table 1. Summary of previously reported single nucleotide polymorphisms (SNP) associated with colorectal cancer

SNP	Chromosome	Gene	Odds ratio	*P*-value	Population	First author	Journal	Year	Reference number
rs4939827	18q21	*SMAD7*	1.15	$1.0 \times 10(-12)$	Caucasian	Broderick P	*Nat Genet*	2007	4
rs6983267	8q24	–	1.17	$3 \times 10(-11)$	European	Tomlinson IP	*Nat Genet*	2007	1
rs4444235	14q22.2	*BMP4*	1.11	$8.1 \times 10(-10)$	European	Houlston RS	*Nat Genet*	2008	6
rs9929218	16q22.1	*CDH1*	0.91	$1.2 \times 10(-8)$	European	Houlston RS	*Nat Genet*	2008	6
rs10411210	19q13.1	*RHPN2*	0.87	$4.6 \times 10(-9)$	European	Houlston RS	*Nat Genet*	2008	6
rs961253	20p12.3	–	1.12	$2.0 \times 10(-10)$	European	Houlston RS	*Nat Genet*	2008	6
rs4779584	15q13	*CRAC1*	1.26	$4.44 \times 10(-14)$	Ashkenazi Jews and Europeans	Jaeger E	*Nat Genet*	2008	7
rs3802842	11q23	–	1.1	$5.8 \times 10(-10)$	European, Japanese, Israeli	Tenesa A	*Nat Genet*	2008	5
rs10795668	10p14	–	1.12	$2.5 \times 10(-13)$	European	Tomlinson IP	*Nat Genet*	2008	2
rs16892766	8q23.3	*EIF3H*	1.26	$3.3 \times 10(-18)$	European	Tomlinson IP	*Nat Genet*	2008	2
rs7136702	12q13.13	–	1.06	$4.02 \times 10(-8)$	European	Houlston RS	*Nat Genet*	2010	8
rs6687758	1q41	–	1.09	$2.27 \times 10(-9)$	European	Houlston RS	*Nat Genet*	2010	8
rs6691170	1q41	–	1.06	$9.55 \times 10(-10)$	European	Houlston RS	*Nat Genet*	2010	8
rs4925386	20q13.33	*LAMA5*	0.93	$1.89 \times 10(-10)$	European	Houlston RS	*Nat Genet*	2010	8
rs10936599	3q26	*MYNN*	0.93	$3.39 \times 10(-8)$	European	Houlston RS	*Nat Genet*	2010	8
rs7758229	6q26	*SLC22A3*	1.28	$7.92 \times 10(-9)$	Japanese and Korean	Cui R	*Gut*	2011	9
rs1957636	14q22	*BMP4*	1.084	$3.93 \times 10(-10)$	Caucasian	Tomlinson IP	*Plos Genet*	2011	11
rs11632715	15p	*GREM1*	1.116	$2.30 \times 10(-10)$	Caucasian	Tomlinson IP	*Plos Genet*	2011	11
rs16969681	15p	*GREM1*	1.181	$5.33 \times 10(-8)$	Caucasian	Tomlinson IP	*Plos Genet*	2011	11
rs4813802	20p12	*BMP2*	1.093	$4.65 \times 10(-11)$	Caucasian	Tomlinson IP	*Plos Genet*	2011	11
rs3824999	11q13.4	*POLD3*	1.08	$3.65 \times 10(-10)$	European and Japanese	Dunlop MG	*Nat Genet*	2012	10
rs1321311	6p21	*CDKN1A*	1.1	$1.14 \times 10(-10)$	European and Japanese	Dunlop MG	*Nat Genet*	2012	10
rs5934683	Xp22.2	*SHROOM2*	1.07	$7.30 \times 10(-10)$	European and Japanese	Dunlop MG	*Nat Genet*	2012	10
rs4813802	20p12	*BMP2*	1.18	$7.3 \times 10(-5)$	Caucasian	Peters U	*Hum Genet*	2012	12
rs2853668	5p33.15	*TERT-CLPTM1L*	0.85	$1.9 \times 10(-4)$	Caucasian	Peters U	*Hum Genet*	2012	12
rs10774214	12p13.32	*CCND2*	1.17	$3.06 \times 10(-8)$	East Asian and European	Jia WH	*Nat Genet*	2013	13
rs2423279	20p12.3	*HAO1, PLCB1*	1.14	$6.64 \times 10(-9)$	East Asian and European	Jia WH	*Nat Genet*	2013	13
rs647161	5q31.1	*PITX1*	1.17	$1.22 \times 10(-10)$	East Asian and European	Jia WH	*Nat Genet*	2013	13
rs3987	4q26	*NDST3*	1.36	$4.02 \times 10(-8)$	Spanish	Real LM	*PLos One*	2014	14
rs35509282	4q32.2	*FSTL5*	1.53	$8.2 \times 10(-9)$	Ashkenazi Jews and Europeans	Schmit SL	*Carcinogenesis*	2014	15
rs12241008	10q25	*VTI1A*	1.19	$1.4 \times 10(-9)$	European, African and Japanese	Wang H	*Nat Commun*	2014	16
rs1035209	10p24.2	*MRP2*	1.13	$4.54 \times 10(-11)$	East Asians in our European	Whiffin N	*Hum Mol Genet*	2014	17
rs3217810	12p13.32	*CCND2*	1.19	$2.16 \times 10(-10)$	East Asians in our European	Whiffin N	*Hum Mol Genet*	2014	17
rs10911251	1q25.3	*LAMC1*	1.09	$1.75 \times 10(-8)$	East Asians in our European	Whiffin N	*Hum Mol Genet*	2014	17
rs7229639	18q21.1	*SMAD7*	1.22	$2.93 \times 10(-11)$	East Asians	Zhang B	*Int J Cancer*	2014	18
rs704017	10q22.3	*ZMIZ1-AS1*	1.1	$2.07 \times 10(-8)$	East Asians	Zhang B	*Nat Genet*	2014	18
rs11196172	10q25.2	*TCF7L2*	1.14	$1.04 \times 10(-12)$	East Asians	Zhang B	*Nat Genet*	2014	18
rs1535	11q12.2	*FADS2*	1.15	$8.21 \times 10(-20)$	East Asians	Zhang B	*Nat Genet*	2014	18
rs174537	11q12.2	*MYRF*	1.16	$9.22 \times 10(-21)$	East Asians	Zhang B	*Nat Genet*	2014	18
rs174550	11q12.2	*FADS1*	1.15	$1.58 \times 10(-19)$	East Asians	Zhang B	*Nat Genet*	2014	18

Table 1 (Continued)

SNP	Chromosome	Gene	Odds ratio	*P*-value	Population	First author	Journal	Year	Reference number
rs4246215	11q12.2	*FEN1*	1.15	$7.65 \times 10(-20)$	East Asians	Zhang B	*Nat Genet*	2014	18
rs10849432	12p13.31	*CD9*	1.14	$5.81 \times 10(-10)$	East Asians	Zhang B	*Nat Genet*	2014	18
rs12603526	17p13.3	*NXN*	1.1	$3.42 \times 10(-8)$	East Asians	Zhang B	*Nat Genet*	2014	18
rs1800469	19q13.2	*TGFB1*	1.09	$1.17 \times 10(-8)$	East Asians	Zhang B	*Nat Genet*	2014	18
rs2241714	19q13.2	*B9D2*	1.09	$1.36 \times 10(-8)$	East Asians	Zhang B	*Nat Genet*	2014	18
rs10904849	10p13	*CUBN*	1.14	$7.01 \times 10(-8)$	European	Al-Tassan NA	*Sci Rep*	2015	19
rs17836917	17q12	*MYO1D, CCL8, CCL13*	0.75	$4.55 \times 10(-8)$	Han Chinese	Jiang K	*Oncotarget*	2015	20
rs12522693	5q23.3	*HINT1*	1.31	$2.08 \times 10(-8)$	Han Chinese	Jiang K	*Oncotarget*	2015	20
rs17094983	14q23.1	–	0.87	$2.5 \times 10(-10)$	African	Lemire M	*Hum Genet*	2015	21
rs16941835	16q24.1	*RP11-58A18*	1.15	$5.06 \times 10(-8)$	African	Lemire M	*Hum Genet*	2015	21
rs72647484	1q36.2	*CDC42, WNT4*	1.21	$1.21 \times 10(-8)$	African	Lemire M	*Hum Genet*	2015	21
rs1119016410	10p24.1	*SLC25A28, ENTPD7, COX15, CUTC, ABCC2*	1.09	$4.0 \times 10(-8)$	European and Asian	Schumacher FR	*Nat Commun*	2015	22
rs318450412	12q24.12	*SH2B3*	1.09	$1.7 \times 10(-8)$	European and Asian	Schumacher FR	*Nat Commun*	2015	22
rs7320812012	12q24.22	*NOS1*	1.16	$2.8 \times 10(-8)$	European and Asian	Schumacher FR	*Nat Commun*	2015	22
rs606682520	20q13.13	*PREX1*	1.09	$4.4 \times 10(-9)$	European and Asian	Schumacher FR	*Nat Commun*	2015	22
rs8124813	3p14.1	*LRIG1*	1.09	$2.0 \times 10(-8)$	European and Asian	Schumacher FR	*Nat Commun*	2015	22
rs35360328	3p22.1	*CTNNB1*	1.14	$3.1 \times 10(-9)$	European and Asian	Schumacher FR	*Nat Commun*	2015	22
rs4711689	6p21.1	*TFEB*	1.11	$3.9 \times 10(-8)$	East Asian	Zeng C	*Gastroenterology*	2016	23
rs2450115	8q23.3	*EIF3H*	1.12	$1.2 \times 10(-12)$	East Asian	Zeng C	*Gastroenterology*	2016	23
rs6469656	8q23.3	*EIF3H*	1.11	$2.0 \times 10(-11)$	East Asian	Zeng C	*Gastroenterology*	2016	23
rs4919687	10q24.3	*CYP17A1*	1.14	$7.8 \times 10(-12)$	East Asian	Zeng C	*Gastroenterology*	2016	23
rs11064437	12p13.3	*SPSB2*	1.12	$4.5 \times 10(-11)$	East Asian	Zeng C	*Gastroenterology*	2016	23

of the disease. For the last ten years, several genome-wide association studies (GWAS) have identified single nucleotide polymorphisms (SNP) intimately connected to the onset of CRC. Tomlinson *et al.* identified rs6983267 at 8q24.21 as the SNP most strongly connected to the onset of CRC.[1,2] Zanke *et al.*[3] investigated 100k SNP in 7480 cases of CRC with double screening among different races and discovered SNP at 8q24 as well as one at 9q24. Brodelick *et al.*[4] report the SNP rs4938827 at 18q21 located within the gene SMAD7 from among 550k SNP in 7473 cases of CRC. Tenesa *et al.*[5] report the SNP rs3802842 at 11q23, rs7014346 at 8q24 and rs4939827 at 18q21. Table 1 shows the data derived from previous GWAS on CRC,[1,2,4–23] which reveal that risky genetic polymorphisms are different among various populations. For example, in the 8q23–24 region, rs6983267 is a risk factor for CRC among Caucasians, Asians and Africans, rs7014346 and rs10505477 are risky only among Caucasians, and rs16892766 is a risk factor for those with Caucasian and African ancestry.[24]

As for the Japanese population, several SNP-based GWAS have been performed for CRC. Matsuo *et al.*[25] performed a case-control study using 481 cases and 962 controls and report an association between CRC and rs6893267 at 8q24. Furthermore, Cui *et al.*[9] performed GWAS using 1583 Japanese CRC cases and 1898 controls and replication analyses using a total of 4809 CRC cases and 2973 controls, including 225 Korean subjects with distal colon cancer and 377 controls. They found an association between CRC and a known carcinogenic SNP at 8q24 and an association between distal CRC and rs7758229 in intron 5 of SLC22A3 at 6q26.[9] Zhang *et al.*[26] performed a case-control study and reported that microsomal glutathione S-transferase 1 (MGST1) gene polymorphisms had an association with CRC risk (OR = 1.682, $P = 0.004$) among the Han Chinese. However, the molecular biological mechanisms by which these SNP are involved in colorectal carcinogenesis remain unclear.

Using a case-control study on 1511 CRC cases and 2098 controls, we previously reported that the risk allele of rs6983267 in 8q24 is associated with CRC, especially in diabetes mellitus patients.[27] However, these SNPs, except for those in 8q24, have not been defined as the regulating polymorphisms of colorectal carcinogenesis beyond racial differences. To find the responsible host genetic factors, we designed an SNP-based GWAS to identify SNP associated with susceptibility to morbidity from CRC in a pure Japanese population.[28] We performed a multistage genome-wide association study in Japanese individuals, with a total of

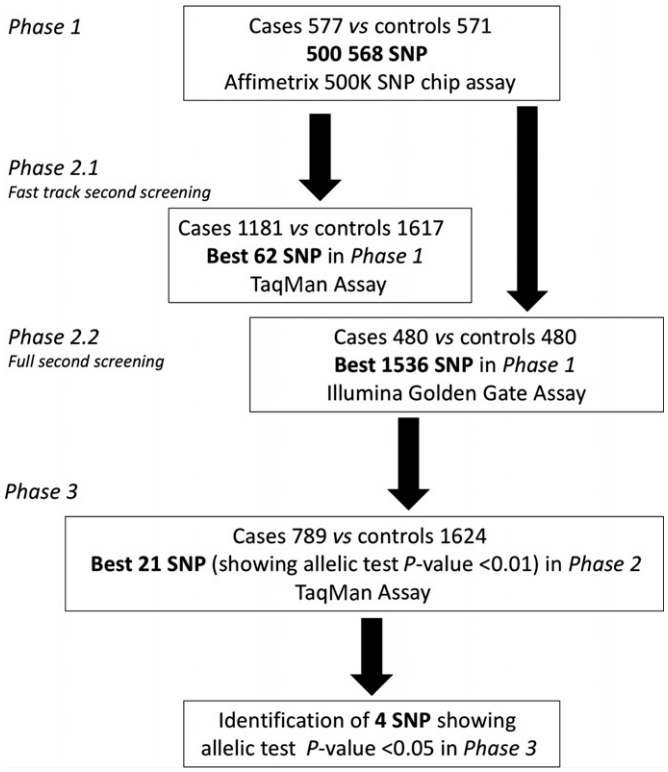

Fig. 1. The complete design of the genome-wide association study. In phase 1, 577 patients with colorectal cancer (CRC) and 571 controls were genotyped for 500 568 single nucleotide polymorphisms (SNP) with Affymetrix 500 K chip sets. Two additional rounds of screening using the Illumina GoldenGate Assay (1536 SNP for phase 2.2) and TaqMan Assay (21 SNP for phase 3) were performed to identify significant SNP.

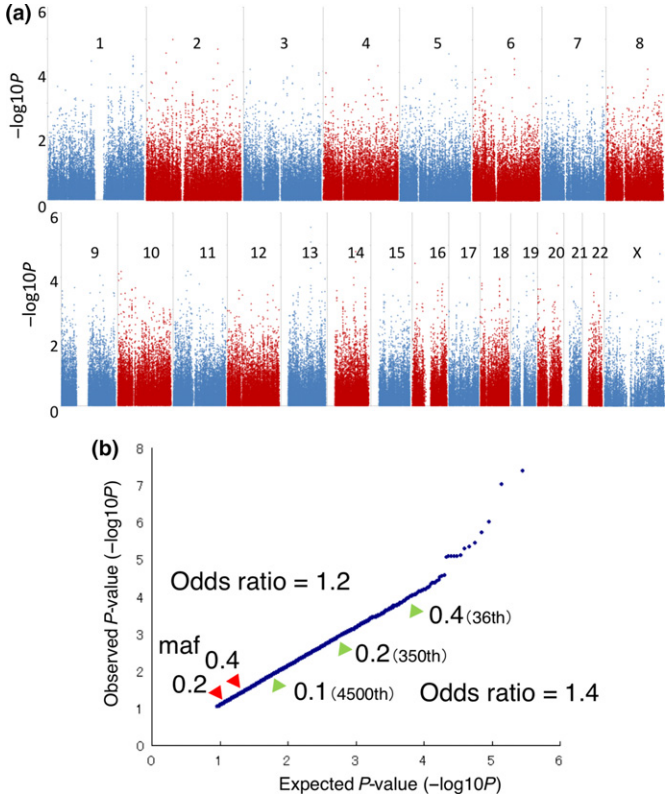

Fig. 2. (a) Manhattan plots from the phase 1 genome-wide association results. *P*-values (−log10 *P*, *y*-axis) are plotted against their respective chromosomal positions (*x*-axis). Each chromosome is depicted in alternating blue and red. (b) Log quantile-quantile *P*-values between the expected (theoretical) *P*-value and the observed *P*-value. The genomic inflation factor (based on median χ^2) is 1.10. If we set the odds ratio of the CRC-related genotype as 1.4 and the allelic frequency in the control as 0.2, it will be located at the 36th quantile by the *P*-value distribution. If we set the odds ratio of the CRC-related genotype as 1.4 and the allelic frequency in the control as 0.4, it will be located at the 350th. If we set the odds ratio of the CRC-related genotype as 1.2 and the allelic frequency in the control as 0.2, it will be located in greater than the 1000th, and the genotype will be difficult to determine.

Table 2. Fast tracked second screening (phase 2.1) of single nucleotide polymorphisms (SNP) related to colorectal cancer

Chromosome	SNP	Phase 1	Phase 2.1	First PCR	Second PCR	Total screening
1	rs325914	4.86E−05	8.08E−01			
1	rs1510310	1.24E−04	5.82E−01			
1	rs1442459	4.32E−05	8.54E−01			
1	rs6692131	1.24E−04	2.75E−01			
2	rs7586098	2.25E−04	4.17E−01			
2	rs10178331	2.63E−05	7.25E−01			
2	rs1036069	9.97E−06	7.52E−01			
2	rs12105972	2.13E−04	5.61E−01			
2	rs13000465	5.16E−05	5.09E−02	5.43E−01		
2	rs12624259	2.00E−05	3.16E−01			
3	rs3771021	2.84E−04	6.82E−01			
3	rs7597875	1.88E−04	3.58E−02	8.13E−01		
3	rs2881606	8.55E−05	4.95E−01			
3	rs6794054	1.45E−04	8.50E−01			
3	rs9823024	1.63E−04	9.13E−01			
4	rs12491172	6.77E−05	3.80E−01			
4	rs1436656	1.26E−04	2.34E−01			
4	rs11725389	1.13E−04	8.91E−01			
5	rs4512014	1.82E−04	2.90E−01			
5	rs13112145	6.74E−05	3.20E−01			
5	rs250222	2.15E−04	7.35E−01			
6	rs6595624	2.65E−05	1.69E−01			
6	rs2864246	1.73E−04	1.62E−01			
6	rs900402	1.70E−04	5.04E−01			
6	rs2523865	8.50E−05	1.67E−01			
7	rs220687	1.45E−04	1.71E+00			
7	rs2040432	3.83E−05	4.80E−01			
8	rs10242940	1.38E−04	6.46E−01			
9	rs6975879	9.73E−05	2.84E−01			
10	rs3107548	8.50E−05	9.40E−01			
10	rs7041802	8.59E−05	2.09E−01			
10	rs11251410	7.79E−05	1.81E−01			
11	rs7912831	6.57E−05	7.10E−02	1.25E−04	2.42E−05	9.31E−08
11	rs11239278	2.14E−04	5.73E−01			
11	rs5030317	8.75E−05	7.18E−01			
11	rs11032820	2.57E−04	4.90E−01			
12	rs7124825	8.12E−05	9.79E−01			
13	rs22576154	1.59E−04	6.54E−01			
13	rs11068349	1.79E−04	3.16E−01			
13	rs7327880	2.75E−06	8.87E−01			
13	rs8002855	5.18E−05	4.41E−01			
13	rs9544316	7.94E−06	8.24E−01			
13	rs1329338	1.80E−04	4.27E−01			
13	rs2803215	3.57E−05	7.87E−01			
13	rs9577345	1.12E−04	1.51E−01			
14	rs6573776	1.72E−04	8.25E−01			
14	rs234588	2.14E−04	2.17E−02	2.33E−01		
15	rs539901	5.63E−05	1.72E−01			
16	rs11860241	3.63E−05	1.44E−01			
16	rs150073	1.43E−04	7.16E−01			
16	rs8047051	2.17E−04	1.79E−01			
16	rs1110560	1.61E−04	6.57E−01			
17	rs7214294	1.04E−04	2.65E−01			
18	rs9303936	1.56E−04	2.23E−01			
18	rs9957443	1.19E−04	5.47E−01			
18	rs11082969	1.84E−04	8.56E−01			
18	rs2879526	1.13E−04	2.52E−01			
19	rs12609781	9.21E−05	9.59E−01			
19	rs326444	7.44E−05	4.48E−01			

Table 2 (Continued)

Chromosome	SNP	Phase 1	Phase 2.1	First PCR	Second PCR	Total screening
20	rs736232	4.30E−06	6.78E−01			
21	rs2825545	1.93E−04	7.94E−01			
22	rs4822015	7.76E−05	4.01E−01			

1846 cases and 2675 controls, to identify disease susceptibility SNP.

Materials and Methods

Study samples. We collected peripheral blood samples from nine collaborating institutes and hospitals for this project investigating the genetic risk factors of CRC cases. Newly diagnosed CRC cases were identified in eight hospitals (Kyushu University Beppu Hospital [Beppu, Japan], Kitazato University [Kanagawa, Japan], National Cancer Center Hospital [Tokyo, Japan], Northern Yokohama Hospital Showa University [Kanagawa, Japan], National Defense Medical College Hospital [Saitama, Japan], Tokyo Medical and Dental University Hospital [Tokyo, Japan], Mie University Hospital [Mie, Japan] and Takano Hospital [Kumamoto, Japan]) from 2000 to 2007. Controls without a prior history of CRC at the time of enrollment were also recruited from those hospitals. All controls were enrolled after having a colonoscopy to ensure that they had no disease. All participants provided documented informed content. The study protocol was reviewed and approved by each institute. We included 1846 cases and 2675 controls into the GWAS study. The average age of CRC patients was 61.9 ± 11.0 years, and that of controls was 59.8 ± 15.0 years. Age and gender details for each phase are shown in Table S1. All cases and controls were of East Asian ancestry and from Japan.

Extraction of genomic DNA and PCR of markers. Genomic DNA was extracted from samples using conventional methodologies and quantified using PicoGreen (Invitrogen, Carlsbad, CA, USA). PCR was done using GeneAmp PCR System 9700 (Applied Biosystems, Carlsbad, CA, USA). Genotyping was done using the ABI 3100 Genetic Analyzer (Applied Biosystems) and analyses and assignment of marker alleles were done with the GENOTYPER programs (Applied Biosystems). Information on SNP was obtained from the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/index.html).

Genotyping. In phase 1, the genotyping of the 577 CRC cases and 571 controls was carried out using the Affymetrix GeneChip Human Mapping 500K Array Set according to the manufacturer's protocols. The equal number of patient and control samples enabled us to analyze the genotype and phenotype independently.

In phase 2.1 (fast-track second screening), among the whole array of SNP, we focused on the highest-ranked 100 SNP. Among those 100 SNP, we excluded SNP with a minor allele frequency (MAF) <0.1 and selected TagSNP to avoiding overlapping and allelic imbalance, thus totaling 62 SNP that were confirmed by PCR in phase 2.1 by screening with a subset of 1181 cases and 1617 controls.

In phase 2.2 (full second screening), 480 CRC and 480 control samples were genotyped at the 1536 best SNP (allelic $P < 0.013$) using the Illumina GoldenGate Assay. When multiple SNP displayed strong linkage disequilibrium (LD) with each other (r2 > 0.8), the most closely associated SNP was chosen to avoid redundancy during the selection of the 1536 SNP. The samples with a genotype call rate <0.98 and SNP with a call rate <0.98, in Hardy–Weinberg disequilibrium ($P < 1.0 \times 10^{-4}$) in the controls, or with a MAF <0.05 were excluded from the association analysis.

For the 21 SNP that showed an allelic $P < 0.01$ in phase 2.2, genotyping with the TaqMan method in 789 CRC cases and 1624 controls was performed in phase 3.

Statistical analysis. Genotype data cleaning and pairwise identity-by-descent (IBD) analysis were carried out using the PLINK software (version 1.06). We used the Haploview software (v3.2) to establish the LD structure on chromosome 10p14.

Results

An overview of the current study design is shown in Figure 1. The genome-wide association study was carried out using the Affymetrix GeneChip Human Mapping 500K Array Set. In phase 1, we genotyped 500 568 tagSNP in 577 individuals with CRC and 571 controls using the Bayesian robust linear model with Mahalanobis (BRLMM) algorithm (http://media.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf). Samples with a genotype call rate <0.94 for either the NspI or StyI GeneChip SNP were removed from analysis. To detect duplicate samples, relatives and DNA-contaminated samples, pairwise IBD estimation was carried out. After applying the strict quality control criteria described above, genotype

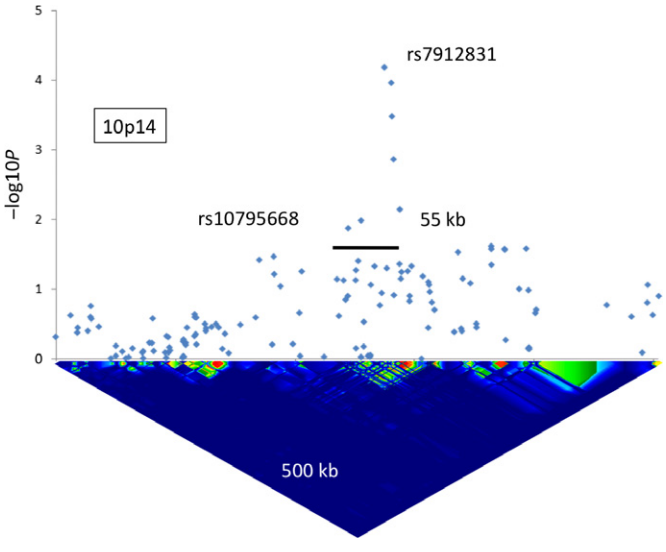

Fig. 3. Linkage disequilibrium (LD) structure at 10p14. The polymorphic site rs7912831 is depicted in an LD block of 500 kb where there are coding and non-coding genes, such as non-coding RNA and micro RNA. The SNP rs10795668, which has been previously reported by Tomlinson *et al.*, is located close to rs7912831. Data were analyzed using Haploview software (v3.2)

Table 3. Final Findings of genome-wide association studies in Japanese colorectal cancer cases

Chromosome	SNP	Position	Phase 1			Phase 2.2			Phase 3			Combined			Odds ratio (95% CI)
			Risk allele frequency		*P* (allelic)	Risk allele frequency		*P* (allelic)	Risk allele frequency		*P* (allelic)	Risk allele frequency		*P* (allelic)	
			Case	Control		Case	Control		Case	Control		Case	Control		
10p14	rs7912831	8771261	0.60	0.53	0.00045	0.62	0.55	0.0030	0.60	0.56	0.0035	0.61	0.55	0.0000000931	1.27 (1.16–1.39)
10p14	rs4749812	8777570	0.61	0.53	0.00020	0.62	0.55	0.0030	0.60	0.56	0.0046	0.61	0.55	0.0000000721	1.27 (1.17–1.39)
10p14	rs7898455	8778914	0.68	0.61	0.00045	068	0.62	0.0085	0.67	0.63	0.0027	0.68	0.62	0.000000155	1.28 (1.17–1.40)
10p14	rs10905453	8784027	0.25	0.20	0.00972	0.25	0.20	0.0082	0.25	0.21	0.0040	0.25	0.21	0.00000501	1.27 (1.15–1.41)

SNP, single nucleotide polymorphisms.

data for 529 cases and 521 controls were chosen for analysis. SNP were removed from the analysis if they had a call rate of less than 0.95, showed a difference in call rate of more than 0.03 between CRC and controls, displayed Hardy–Weinberg disequilibrium ($P < 1.0 \times 10^{-6}$) in the control group, or had a MAF <0.10. SNP that were not selected in the updated GeneChip SNP5.0 (Affymetrix) were also excluded. After these exclusions, 280 972 SNP remained in the first stage. Figure 2a shows a Manhattan plot of these data. The genomic inflation factor based on the median χ^2-value was 1.100 in this genome-wide association analysis (Fig. 2b), implying that there was no systematic increase of false positives owing to population stratification or any other form of bias. The strongest associations identified in phase 1 were found at the polymorphic sites rs13000465 ($P = 5.16 \times 10^{-5}$) on chromosome 2 and equally at rs7912831 ($P = 6.57 \times 10^{-5}$) on chromosome 10.

In phase 2.1, we focused on the 100 highest-ranked SNP among the whole set of SNP. Among these 100 SNP, 62 SNP were confirmed by PCR after screening with a subset of 1181 cases and 1617 controls (Table 2). We confirmed the susceptibility of 4 SNP (rs13000465, rs7597875, rs7912831 and rs234588) to morbidity from CRC. Eventually, we identified the colorectal cancer susceptibility locus rs7912831 on chromosome 10p14 ($P = 9.31E{-}08$) (Table 2). Then, we proceeded to the phase 2.2 full screening and genotyped 480 CRC cases and 480 controls using the Illumina GoldenGate Assay for the top-ranking 1536 SNP. A total of 21 SNP were identified in phase 2.2, including rs7912831 on chromosome 10p14.

For the 21 SNP that showed an allelic $P < 0.01$ in phase 2, genotyping with the TaqMan method in 789 CRC cases and 1624 controls was performed in phase 3. The average SNP call rate of these 21 SNP was 0.998. We identified 4 SNP (rs7912831, rs4749812, rs7898455 and rs10905453) with an allelic $P < 0.05$ on chromosome 10p14. As depicted in Figure 3, these SNP are within a region of fairly extensive LD consisting of a 500-kb block on chromosome 10p14, which maps within 55 kb to the SNP rs10795668 reported by Tomlinson *et al.*[2] This chromosomal region has already been reported to contain polymorphic sites; however, the current study is the first to report that variants at these specific polymorphic sites are associated with susceptibility to CRC. The final findings of our GWAS in Japanese CRC cases are shown in Table 3.

Discussion

In this study, we identified 4 SNP that are significantly associated with morbidity from CRC at the 10p14 locus. These novel SNP are close to the variants on 10p14 described by Tomlinson *et al.*[1] These independent whole genome-wide association studies both found loci on 10p14 to be commonly associated with CRC. We consider this 10p14 locus to be a significant region of CRC susceptibility because it was identified in multiple studies in more than one race, including European and Japanese populations.[29,30] Recently, a significant interaction between an SNP at 10p14 (rs4143094) and processed meat consumption was also reported in CRC patients.[31]

We performed data mining for genes that exist in this susceptibility locus in order to investigate the mechanism by which these SNP are connected to the causes of CRC. No genes, including noncoding RNA, were found in the susceptibility locus identified in the current study using the online human genome database (UCSC Genome Bioinformatics [http://genome.ucsc.edu/]) (Fig. S1). The genes nearest to our

susceptibility locus at 10p14 are GATA3 and CUGBP2. However, no significant correlations between our SNP and the expression of these genes were found by the combined SNP and cDNA expression arrays. We also sought to find new transcripts related to colorectal carcinogenesis at the 10p14 locus. First, we performed a RACE assay for each region where the SNP exist using total RNA extracted from CRC cell lines (Fig. S2), but we could not find any transcripts around all 4 SNP. Second, we found a sequence homologous to mmu-mir-1981, which is a murine micro RNA, just beside rs10905453. However, no transcript was found with northern blotting (Fig. S3). Moreover, we also did not find new protein-coding or RNA genes in the susceptibility region using mRNA whole-transcriptome analysis of 25 CRC cancer tissues. We could not find any functional genes around carcinogenic SNP in 10p14 in this study. Few reports have demonstrated the mechanism by which an SNP regulates the expression of genes or noncoding RNA to promote carcinogenesis and the development of CRC, but further study is warranted.

The mechanisms of CRC carcinogenesis caused by carcinogenic SNP, including the most common SNP rs6983267 at 8q24, are not well known because these SNP do not exist in coding regions. Tuupanen *et al.*[32] report that the risk allele of rs6983267 is associated with microsatellite-stable cancer, and propose that the underlying germ line genetic defect in 8q24 was a target in the somatic evolution of CRC. Interestingly, they also report that the risk allele G shows a copy number increase during CRC development and that rs6983267 affected a binding site for the Wnt-regulated transcription factor TCF4/LEF1, which leads to the enhancement of *MYC* transcription *in vitro* and *in vivo*.[33,34] The abundant expression of *MYC* contributes to carcinogenesis and progression of CRC.[27,34–36]

We consider that further analysis of such loci will enable us to understand the unknown mechanisms of colorectal carcinogenesis, including discovering new genes or noncoding RNA. For example, we reported that an SNP in miR-146a targeting EGFR and IRAK1 is associated with the prognosis of gastric cancer patients.[37] As the underlying basis of the association identified at 10p14 is presently unclear, there are no clues to explain how this region is involved in morbidity from CRC. Our data reveal that 10p14 is a colorectal cancer susceptible region for more than one racial subgroup, but further studies are warranted to find the mechanistic relationship between 10p14 and colorectal carcinogenesis.

In conclusion, using a multistage GWAS in Japanese individuals, we identified a significant genome-wide level of association for 4 SNP on chromosome 10p14 associated with the onset of CRC.

Acknowledgements

We thank Ms Oda, Ms Kasagi and Ms Kawano for their technical assistance.

Disclosure Statement

The authors have no conflicts of interest to declare.

References

1 Tomlinson I, Webb E, Carvajal-Carmona L *et al.* A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 2007; **39**: 984–8.
2 Tomlinson IP, Webb E, Carvajal-Carmona L *et al.* A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet* 2008; **40**: 623–30.
3 Zanke BW, Greenwood CM, Rangrej J *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 2007; **39**: 989–94.
4 Broderick P, Carvajal-Carmona L, Pittman AM *et al.* A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet* 2007; **39**: 1315–7.
5 Tenesa A, Farrington SM, Prendergast JG *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet* 2008; **40**: 631–7.
6 COGENT Study, Houlston RS, Webb E *et al.* Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet* 2008; **40**: 1426–35.
7 Jaeger E, Webb E, Howarth K *et al.* Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat Genet* 2008; **40**: 26–8.
8 Houlston RS, Cheadle J, Dobbins SE *et al.* Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat Genet* 2010; **42**: 973–7.
9 Cui R, Okada Y, Jang SG *et al.* Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. *Gut* 2011; **60**: 799–805.
10 Dunlop MG, Dobbins SE, Farrington SM *et al.* Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat Genet* 2012; **44**: 770–6.
11 Tomlinson IP, Carvajal-Carmona LG, Dobbins SE *et al.* Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet* 2011; **7**: e1002105.
12 Peters U, Hutter CM, Hsu L *et al.* Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum Genet* 2012; **131**: 217–34.
13 Jia WH, Zhang B, Matsuo K *et al.* Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. *Nat Genet* 2013; **45**: 191–6.
14 Real LM, Ruiz A, Gayan J *et al.* A colorectal cancer susceptibility new variant at 4q26 in the Spanish population identified by genome-wide association analysis. *PLoS One* 2014; **9**: e101178.
15 Schmit SL, Schumacher FR, Edlund CK *et al.* A novel colorectal cancer risk locus at 4q32.2 identified from an international genome-wide association study. *Carcinogenesis* 2014; **35**: 2512–9.
16 Wang H, Burnett T, Kono S *et al.* Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in VTI1A. *Nat Commun* 2014; **5**: 4613.
17 Whiffin N, Hosking FJ, Farrington SM *et al.* Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Hum Mol Genet* 2014; **23**: 4729–37.
18 Zhang B, Jia WH, Matsuda K *et al.* Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nat Genet* 2014; **46**: 533–42.
19 Al-Tassan NA, Whiffin N, Hosking FJ *et al.* A new GWAS and meta-analysis with 1000 Genomes imputation identifies novel risk variants for colorectal cancer. *Sci Rep* 2015; **5**: 10442.
20 Jiang K, Sun Y, Wang C *et al.* Genome-wide association study identifies two new susceptibility loci for colorectal cancer at 5q23.3 and 17q12 in Han Chinese. *Oncotarget* 2015; **6**: 40327–36.
21 Lemire M, Qu C, Loo LW *et al.* A genome-wide association study for colorectal cancer identifies a risk locus in 14q23.1. *Hum Genet* 2015; **134**: 1249–62.
22 Schumacher FR, Schmit SL, Jiao S *et al.* Corrigendum: Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nat Commun* 2015; **6**: 8739.
23 Zeng C, Matsuda K, Jia WH *et al.* Identification of susceptibility loci and genes for colorectal cancer risk. *Gastroenterology* 2016; **150**: 1633–45.
24 Li L, Lv L, Liang Y *et al.* Association of 8q23-24 region (8q23.3 loci and 8q24.21 loci) with susceptibility to colorectal cancer: a systematic and updated meta-analysis. *Int J Clin Exp Med* 2015; **8**: 21001–13.
25 Matsuo K, Suzuki T, Ito H *et al.* Association between an 8q24 locus and the risk of colorectal cancer in Japanese. *BMC Cancer* 2009; **9**: 379.
26 Zhang H, Liao LH, Liu SM *et al.* Microsomal glutathione S-transferase gene polymorphisms and colorectal cancer risk in a Han Chinese population. *Int J Colorectal Dis* 2007; **22**: 1185–94.
27 Ishimaru S, Mimori K, Yamamoto K *et al.* Increased risk for CRC in diabetic patients with the nonrisk allele of SNPs at 8q24. *Ann Surg Oncol* 2012; **19**: 2853–8.

28 Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39**: 906–13.
29 Qin Q, Liu L, Zhong R *et al.* The genetic variant on chromosome 10p14 is associated with risk of colorectal cancer: results from a case-control study and a meta-analysis. *PLoS One* 2013; **8**: e64310.
30 Hong Y, Wu G, Li W, Liu D, He K. A comprehensive meta-analysis of genetic associations between five key SNPs and colorectal cancer risk. *Oncotarget* 2016; **7**: 73945–59.
31 Figueiredo JC, Hsu L, Hutter CM *et al.* Genome-wide diet-gene interaction analyses for risk of colorectal cancer. *PLoS Genet* 2014; **10**: e1004228.
32 Tuupanen S, Niittymaki I, Nousiainen K *et al.* Allelic imbalance at rs6983267 suggests selection of the risk allele in somatic colorectal tumor evolution. *Cancer Res* 2008; **68**: 14–7.
33 Tuupanen S, Turunen M, Lehtonen R *et al.* The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet* 2009; **41**: 885–90.
34 Takatsuno Y, Mimori K, Yamamoto K *et al.* The rs6983267 SNP Is Associated with MYC Transcription Efficiency, Which Promotes Progression and Worsens Prognosis of Colorectal Cancer. *Ann Surg Oncol* 2013; **20**: 1395–402.
35 Rochlitz CF, Herrmann R, de Kant E. Overexpression and amplification of c-myc during progression of human colorectal cancer. *Oncology* 1996; **53**: 448–54.
36 Sugimachi K, Niida A, Yamamoto K *et al.* Allelic imbalance at an 8q24 oncogenic SNP is involved in activating MYC in human colorectal cancer. *Ann Surg Oncol* 2014; **21**(Suppl 4): S515–21.
37 Kogo R, Mimori K, Tanaka F, Komune S, Mori M. Clinical significance of miR-146a in gastric cancer cases. *Clin Cancer Res* 2011; **17**: 4277–84.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Demographic data of patients and controls in this study.

Fig. S1. Data mining for genes which exist in oncogenic susceptibility locus using online database.

Fig. S2. Four primer sets, sandwiching 4 single nucleotide polymorphisms at 10p14 were designed for qRT-PCR assay.

Fig. S3. Exploration of transcripts in 10p14 region with northern blotting.