nature genetics

Common variation near *CDKN1A*, *POLD3* and *SHROOM2* influences colorectal cancer risk

Malcolm G Dunlop[1,*], Sara E Dobbins[2], Susan Mary Farrington[1], Angela M Jones[3], Claire Palles[3], Nicola Whiffin[2], Albert Tenesa[1], Sarah Spain[3], Peter Broderick[2], Li-Yin Ooi[1], Enric Domingo[3], Claire Smillie[1], Marc Henrion[2], Matthew Frampton[2], Lynn Martin[3], Graeme Grimes[1], Maggie Gorman[3], Colin Semple[1], Yusanne P Ma[2], Ella Barclay[3], James Prendergast[1], Jean-Baptiste Cazier[3], Bianca Olver[2], Steven Penegar[2], Steven Lubbe[2], Ian Chander[4], Luis G Carvajal-Carmona[3], Stephane Ballereau[1], Amy Lloyd[2], Jayaram Vijayakrishnan[2], Lina Zgaga[1,5], Igor Rudan[5], Evropi Theodoratou[5], The Colorectal Tumour Gene Identification (CORGI) Consortium[6], John M Starr[7], Ian Deary[7], Iva Kirac[8], Dujo Kovacević[9], Lauri A Aaltonen[10], Laura Renkonen-Sinisalo[11], Jukka-Pekka Mecklin[12], Koichi Matsuda[13], Yusuke Nakamura[13], Yukinori Okada[14], Steven Gallinger[15], David J Duggan[16], David Conti[17], Polly Newcomb[18], John Hopper[19], Mark A Jenkins[19], Fredrick Schumacher[17], Graham Casey[17], Douglas Easton[20,21], Mitul Shah[20], Paul Pharoah[20,21], Annika Lindblom[22], Tao Liu[22], The Swedish Low-Risk Colorectal Cancer Study Group[6], Christopher G Smith[23], Hannah West[23], Jeremy P Cheadle[23], The COIN Collaborative Group[6], Rachel Midgley[24], David J Kerr[24], Harry Campbell[1,5], Ian P Tomlinson[3,25] & Richard S Houlston[2]

We performed a meta-analysis of five genome-wide association studies to identify common variants influencing colorectal cancer (CRC) risk comprising 8,682 cases and 9,649 controls. Replication analysis was performed in case-control sets totaling 21,096 cases and 19,555 controls. We identified three new CRC risk loci at 6p21 (rs1321311, near *CDKN1A*; $P = 1.14 \times 10^{-10}$), 11q13.4 (rs3824999, intronic to *POLD3*; $P = 3.65 \times 10^{-10}$) and Xp22.2 (rs5934683, near *SHROOM2*; $P = 7.30 \times 10^{-10}$) This brings the number of independent loci associated with CRC risk to 20 and provides further insight into the genetic architecture of inherited susceptibility to CRC.

Many colorectal cancers develop in genetically susceptible individuals, most of whom are not carriers of germline mismatch-repair or *APC* mutations[1–3]. Genome-wide association studies (GWAS) have validated the hypothesis that part of the heritable risk of CRC is attributable to common, low-risk variants, identifying CRC susceptibility loci in 17 genomic regions[4–10]. The statistical power of individual GWAS is limited by the modest effect sizes of genetic variants and financial constraints on the numbers of variants that can be followed up. Meta-analysis of existing GWAS data offers the opportunity to discover additional disease loci, according to current projections for the number of independent regions harboring common variants associated with CRC risk[11]. In this study, we conducted a meta-analysis of GWAS data and validation in multiple independent case-control series, identifying three new susceptibility loci for CRC.

The discovery phase comprised five GWAS data sets from the UK population, totaling 8,682 cases and 9,649 controls (**Supplementary Table 1**). The Scotland1 GWAS consisted of genotyping 1,012 early-onset Scottish CRC cases and 1,012 controls using the Illumina HumanHap300 and HumanHap240S arrays (COGS Study). London phase 1 (UK1) was based on genotyping 940 cases with familial colorectal neoplasia and 965 controls ascertained through the Colorectal Tumour Gene Identification (CORGI) Consortium using Illumina HumanHap550 arrays. Scotland2 was based on an additional 2,057 cases and 2,111 controls (Scottish colorectal cancer study (SOCS)), and UK2 samples comprised an additional 2,873 CRC cases and 2,871 controls ascertained through the National Study of Colorectal Cancer Genetics (NSCCG). Scotland2 and UK2 samples were genotyped using Illumina Infinium-iSelect and GoldenGate arrays for a common set of 43,140 SNPs, including the 14,982 most strongly associated SNPs from UK1, the 14,972 most strongly associated SNPs from Scotland1 and the 13,186 SNPs showing the strongest association in a joint analysis of all CRC cases and controls from both phase 1 data sets. The VQ58 GWAS comprised 1,800 CRC cases from the UK-based VICTOR and QUASAR2 adjuvant chemotherapy clinical trials. Victor, Quasar, 1958 Birth Cohort (VQ58) cases were genotyped using the Illumina HumanHap300 and HumanHap370 arrays. The 2,690 controls, genotyped on the Illumina Human-1.2M-Duo Custom_v1 array, were from the UK population-based 1958 Birth Cohort.

Before undertaking meta-analysis of all GWAS data sets, we searched for potential biases in each case-control series (**Supplementary Fig. 1**). Comparison of the observed and expected χ^2 distributions showed little evidence for inflation of the test

*A full list of affiliations appears at the end of the paper.

Received 23 December 2011; accepted 30 April 2012; published online 27 May 2012; doi:10.1038/ng.2293

Table 1 Summary results for the SNPS rs1321311 (6p21), rs3824999 (11q13.4) and rs5934683 (Xp22.2) associated with CRC risk

SNP	Study	OR	95% CI	*P* value
rs1321311	Discovery	1.09	1.05–1.14	4.79×10^{-5}
	Replication	1.09	1.05–1.14	5.74×10^{-6}
	Japan	1.18	1.03–1.36	1.71×10^{-2}
	Combined	1.10	1.07–1.13	1.14×10^{-10} ($P_{het} = 0.55$, $I^2 = 0\%$)
rs3824999	Discovery	1.08	1.05–1.13	1.77×10^{-5}
	Replication	1.07	1.04–1.11	2.06×10^{-5}
	Japan	1.09	0.99–1.19	8.46×10^{-2}
	Combined	1.08	1.05–1.10	3.65×10^{-10} ($P_{het} = 0.05$, $I^2 = 41\%$)
rs5934683	Discovery	1.08	1.04–1.12	8.19×10^{-5}
	Replication	1.07	1.04–1.10	2.16×10^{-6}
	Japan	1.04	0.93–1.16	5.38×10^{-1}
	Combined	1.07	1.04–1.10	7.30×10^{-10} ($P_{het} = 0.31$, $I^2 = 13\%$)

statistics (**Supplementary Fig. 2**), thereby excluding the possibility of significant hidden population substructure, cryptic relatedness among subjects or differential genotype calling. Principal-component analysis showed that the cases and controls were genetically well matched (**Supplementary Fig. 3** and **Supplementary Note**). Any outliers or related individuals were excluded (Online Methods and **Supplementary Fig. 1**).

We also made use of data on 260 SNPs from 2,183 cases and 2,501 controls who had been genotyped as part of the COIN cases, National Blood Service controls (COINNBS) series. These SNPs had shown some evidence of association with CRC in a previous meta-analysis of the five GWAS data sets in which a smaller set of VQ cases were genotyped[8] (**Supplementary Table 1**).

Using data from the above six studies, we derived for each SNP joint odds ratios (ORs) and confidence intervals (CIs) under a fixed-effects model and determined the associated *P* values. We identified two SNPs, rs1321311 and rs3824999, showing good evidence of association ($P < 5.0 \times 10^{-5}$) and mapping to distinct loci not previously associated with CRC risk. The *P*-value threshold used does not exclude the possibility that other SNPs represent genuine association signals but was simply a pragmatic strategy for prioritizing replication.

To validate our findings, we conducted a replication study of rs1321311 and rs3824999, genotyping samples from nine additional case-control series: the Colon Cancer Family Registry (CCFR1), the UK NSCCG (UK3), the UK CORGI (UK4), an Edinburgh study (Scotland3), a Cambridge study (Cambridge), a Croatian study (Croatia), the Finnish Colorectal Cancer Predisposition Study (Helsinki) and a Swedish study (Sweden), together with a Japanese study (Japan) (**Supplementary Table 1**). In the combined analysis, both rs1321311 ($P = 1.14 \times 10^{-10}$; $P_{het} = 0.55$, $I^2 = 0\%$) and rs3824999 ($P = 3.65 \times 10^{-10}$; $P_{het} = 0.05$, $I^2 = 41\%$) showed evidence for an association with CRC at genome-wide significance ($P < 5.0 \times 10^{-8}$) (**Table 1**, Online Methods and **Supplementary Table 2**).

rs3824999 maps to 11q13.4 at position 74,023,198 within intron 9 of the *POLD3* gene (encoding polymerase DNA-directed δ3; MIM 611415; **Fig. 1**). The POLD3 protein is a component of the DNA polymerase–δ complex that comprises proliferating cell nuclear antigen (PCNA), the multi-subunit replication factor C and the four-subunit polymerase complex. As well as being involved in suppression of homologous recombination, the DNA polymerase–δ complex participates in DNA mismatch and base-excision repair, key processes shown to be defective in Mendelian CRC susceptibility disorders[12].

rs1321311 maps to 6p21 at position 36,730,878 within a region of linkage disequilibrium (LD) that encompasses the *CDKN1A* gene (encoding cyclin-dependent kinase inhibitor 1A; MIM 116899; **Fig. 1**). Notably, rs1321311 has been shown to be associated with electrocardiographic QRS duration[13]. *CDKN1A* encodes $p21^{WAF1/Cip1}$, which mediates p53-dependent G1 growth arrest[14]. Moreover, p21 acts as a master effector of multiple tumor suppressor pathways that function independently of classical p53 tumor suppression. Also, by binding to PCNA, p21 interferes with PCNA-dependent DNA polymerase activity, thereby inhibiting DNA replication and modulating PCNA-dependent DNA repair[14]. Through binding to PCNA, p21 also competes for PCNA binding with DNA polymerase–δ and several other proteins involved in DNA synthesis, thus directly inhibiting DNA synthesis[14]. Similarly, p21 represses MYC-dependent transcription and, in turn, MYC disrupts the PCNA-p21 interaction, thus alleviating p21-dependent inhibition of PCNA and DNA synthesis[14]. Decreased p21 expression has been reported to be a feature of dysplastic aberrant crypt foci in colonic mucosa and adenomas. The finding that p21 downregulation inversely correlates with microsatellite instability (MSI) status in CRC, irrespective of p53 status, again invokes a relationship with defective DNA repair and genomic instability[14].

Including the two newly discovered SNPs, a total of 19 independent risk SNPs for CRC have been identified, all mapping to autosomal regions of the genome. The risk of sporadic CRC is higher for males in both economically developed and less-developed countries. Furthermore, males are at greater overall risk for CRC and have an earlier age of onset for Lynch syndrome[15–17]. It is possible that some of these differences in risk are attributable to sex chromosome genetic variation. To explore this hypothesis, we studied the relationship between SNPs mapping to the sex-specific region of the X chromosome and CRC risk. Genotypes were analyzed using an extension to the standard Cochran-Armitage test for trend[18] (Online Methods).

rs5934683 was the only SNP that showed strong evidence of association in the meta-analysis of the UK1, UK2, Scotland1, Scotland2 and VQ58 data sets. We also genotyped rs5934683 in the UK3, Scotland3, UK4, CCFR1, Cambridge, Croatia, Helsinki, Sweden and Japan studies (**Supplementary Table 1**). In combined analysis, rs5934683 showed evidence for an association with CRC at genome-wide significance ($P = 7.30 \times 10^{-10}$, $P_{het} = 0.31$, $I^2 = 13\%$; **Table 1** and **Supplementary Table 2**).

rs5934683 maps to Xp22.2 within a 43-kb region of LD (position 9,711,474; **Fig. 1**). Two genes map to this region, *GPR143* (encoding G protein–coupled receptor 143; MIM 300808), which is expressed by melanocytes and retinal pigment epithelium, and *SHROOM2* (encoding shroom family member 2; MIM 300103), a human homolog of the *Xenopus laevis APX* gene. rs5934683 is situated between *GPR143* and *SHROOM2* and seems to be within the distal promoter region of *SHROOM2*. There is also evidence of longer, less-abundant *GPR143* transcripts extending into the *SHROOM2* promoter. *SHROOM2* is known to have broad roles in cell morphogenesis during endothelial and epithelial tissue development[19]. Missense mutations in *SHROOM2* have been detected in large-scale screens for recurrent mutations in cancer cell lines[20]. Like *GPR143*, *SHROOM2* regulates melanosome biogenesis and localization in the retinal pigment epithelium[21]. Notably, abnormal retinal pigmentation, similar to the congenital hypertrophy of retinal pigment epithelium (CHRPE) lesions that are a component of the familial adenomatous polyposis syndrome, has previously been shown to be an extracolonic feature of non-FAP CRC[22,23]. To our knowledge, the relationship between Xp22.2 and CRC risk represents the first evidence for the role of X-chromosome variation in predisposition to a non–sex specific cancer.

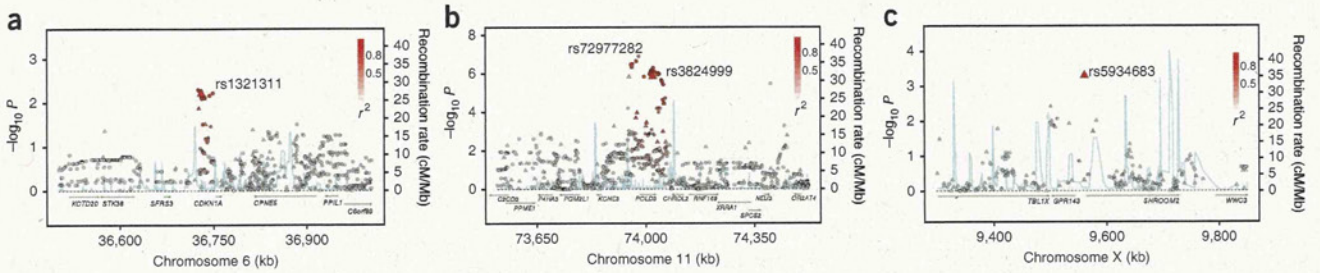

Figure 1 Regional plots of association results and recombination rates for the 6p21, 11q13.4 and Xp22.2 susceptibility loci. (**a**–**c**) Association results of both genotyped (triangles) and imputed (circles) SNPs in the GWAS samples and recombination rates within the loci at 6p21(**a**), 11q13.4 (**b**) and Xp22.2 (**c**). For each plot, $-\log_{10}$ *P* values (*y* axis) of the SNPs are shown according to their chromosomal positions (*x* axis). The top genotyped SNP in each combined analysis is shown as a large triangle and is labeled by its rsID. Color intensity of each symbol reflects the extent of LD with the top genotyped SNP. Genetic recombination rates, estimated using HapMap Utah residents of Western and Northern European ancestry (CEU) samples, are shown with a light blue line. Physical positions are based on NCBI Build 36 of the human genome. Also shown are the relative positions of genes and transcripts mapping to each region of association. Genes have been redrawn to show the relative positions; therefore, maps are not to physical scale.

Next, we assessed associations between clinicopathological variables (sex, age at diagnosis, family history of CRC and tumor site, stage and microsatellite instability) and genotype at rs1321311, rs3824999 and rs5934683 through case-only logistic regression (**Supplementary Table 3**). After adjusting for multiple testing, we did not find any significant association.

To analyze comprehensively the associations at 6p21, 11q13.4 and Xp22.2, we imputed genotypes in GWAS cases and controls using HapMap 3 and 1000 Genomes Project data for the autosomal regions and HapMap release 21 for Xp22.2 (**Fig. 1** and Online Methods). We did not find substantive evidence of stronger associations at the 6p21.2 and Xp22.2 risk loci. However, at the 11q13.4 locus, rs72977282, mapping 3,188 bp 5′ to *POLD3*, was more strongly associated with CRC than rs3824999 (**Fig. 1** and **Supplementary Table 4**). No nonsynonymous SNPs showing strong LD ($r^2 > 0.4$, $D' > 0.8$) with rs1321311, rs3824999 or rs5934683 at 6p21, 11q13.4 and Xp22.2 loci, respectively, were identified. These data indicate that it is likely that the associations between 6p21, 11q13.4 and Xp22.2, and CRC risk are mediated through changes that influence gene expression rather than protein sequence.

To examine whether any directly genotyped or imputed SNPs lie within or very close to a putative transcription factor–binding and/or enhancer element, we conducted a bioinformatics search using Transfac[24], ENCyclopedia of DNA Elements (ENCODE) CHIP-Seq and ENCODE UW DNAaseI Hypersensitivity data. These analyses did not provide evidence that rs1321311, rs3824999 or rs5934683, or any closely correlated SNP maps to a known or predicted region of transcriptional regulation (**Supplementary Table 4**).

To explore whether the rs1321311, rs3824999 and rs5934683 associations (or those of proxy SNPs) reflect *cis*-acting regulatory effects on *POLD3*, *CDKN1A*, *GPR143* or *SHROOM2*, we conducted expression studies using Illumina HT-12 arrays with RNA extracted from 42 samples of normal colonic epithelium (**Supplementary Table 5**). We also analyzed publicly available mRNA expression data from fibroblasts, lymphoblastoid cell lines (LCLs), T cells, adipose tissue and CRCs[25,26] (**Supplementary Table 5**). *In silico* analysis revealed a statistically significant relationship between the genotype at rs1321311 and expression of *CDKN1A*. However, this was observed only in LCL and T-cell data, with no evidence of an effect in the colon (**Supplementary Table 5**). We also found that the risk allele at rs5934683 was associated with a marked reduction in *SHROOM2* expression in both normal colonic epithelium and CRC tissue (**Supplementary Fig. 4**). The relationship between *SHROOM2* expression in normal colonic epithelium and rs5934683 genotype was very strong ($P = 1.3 \times 10^{-7}$) and was significant after accounting for all genes tested on the HT-12 array ($P = 9.0 \times 10^{-4}$). Indeed, rs5934683 genotype accounted for 55% of the variation in *SHROOM2* expression. Exploring the relationship between *SHROOM2* expression, rs5934683 risk genotype and CRC causation will be of considerable interest, not least because of the observations of an association between excess pigmented lesions in the retinal pigment epithelium and CRC[22,23]. There was no significant difference in the observed minor allele frequency (MAF) of rs5934683 between female and male cases, raising the possibility that skewed X-chromosome inactivation might underscore the associated CRC risk. Favored X-chromosome inactivation producing a normal phenotype has been documented in X-linked dominant disease[27], and skewed X-chromosome inactivation has been implicated as a risk factor for breast cancer[28]. The expression data were consistent with full dosage compensation, but, due to sample and effect sizes, we are currently unable to confirm or refute a dosage effect on risk. There was no detectable relationship between rs3824999 and *POLD3* expression in any of the expression studies. It should be noted that these exploratory analyses could only detect >5% differences in RNA expression by genotype with 80% power at a single time point, and, hence, we could not exclude any subtle effects of genotype on target tissues relevant to CRC.

By pooling GWAS data and conducting extensive replication analyses, we have identified three new loci influencing CRC susceptibility. The loci are of modest effect size, which is not unexpected, given that common alleles with a larger impact on CRC were likely to have been discovered in previous studies. Although additional analyses are required to determine the functional consequences that lead to CRC, our findings highlight the importance of variation in genes encoding components of the $p21^{WAF1/Cip1}$ signaling pathway in CRC. This pathway, elucidated through the extended interaction network of *CDKN1A*, incorporates not only *POLD3*, which was discovered as a CRC risk locus here, but also *MYC* and other genes (including SMADs and other transforming growth factor (TGF)-β pathway genes) that we have previously identified as risk factors for CRC.

URLs. R suite, http://www.r-project.org/; detailed information on the tag SNP panel, http://www.illumina.com; dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/; International HapMap Project, http://www.hapmap.org/; 1000 Genomes Project, http://www.1000genomes.org/; SNP Annotation and Proxy Search (SNAP),

http://www.broadinstitute.org/mpg/snap/; IMPUTE, https://mathgen.stats.ox.ac.uk/impute/impute.html; SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html; TRANSFAC Matrix Database, http://www.biobase-international.com/product/transcription-factor-binding-sites; Wellcome Trust Case Control Consortium, www.wtccc.org.uk/; Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/omim; SIFT: http://sift.jcvi.org/; PolyPhen, http://genetics.bwh.harvard.edu/pph/; GLOBOCAN, http://globocan.iarc.fr/; Cancer Genome Atlas, http://cancergenome.nih.gov/; The ENCyclopedia of DNA Elements (ENCODE) Project, http://genome.ucsc.edu/ENCODE/; GENe Expression VARiation (Genevar), http://www.sanger.ac.uk/resources/software/genevar/; Catalogue of Somatic Mutations in Cancer (COSMIC), http://www.sanger.ac.uk/genetics/CGP/cosmic/; Haploview, http://www.broad.mit.edu/mpg/haploview/.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

Cancer Research UK provided principal funding for this study individually to R.S.H. (C1298/A8362–Bobby Moore Fund for Cancer Research UK), I.P.T. and M.G.D. At the Institute of Cancer Research, additional funding was provided by a Centre Grant from Core as part of the Digestive Cancer Campaign, the National Cancer Research Network and the NHS via the Biological Research Centre of the NIHR at the Royal Marsden Hospital NHS Trust. S.L. received a PhD studentship from Cancer Research UK, I.C. received a Clinical Research Training Fellowship from St. George's Hospital Medical School, and N.W. received a PhD Studentship from the Institute of Cancer Research. M.H. was in receipt of a Post-Doctoral Training post from the Leukaemia Lymphoma Research Fund.

In Oxford, additional funding was provided by the Oxford Comprehensive Biomedical Research Centre (E.D., C.P. and I.P.T.) and the European Union Framework Programme 7 (FP7) CHIBCHA grant (A.M.J., L.G.C.-C. and I.P.T.). Core infrastructure support to the Wellcome Trust Centre for Human Genetics, Oxford, was provided by grant (090532/Z/09/Z).

We are grateful to many colleagues within the UK Clinical Genetics Departments (for CORGI) and to many collaborators who participated in the VICTOR and QUASAR2 trials. We also thank colleagues from the UK National Cancer Research Network (for NSCCG).

In Edinburgh, funding was provided by a Cancer Research UK Programme Grant (C348/A12076) and a Centre Grant from the CORE Charity. E.T. was funded by a Cancer Research UK Fellowship (C31250/A10107). L.-Y.O. is supported by a Cancer Research UK Research Training Fellowship (C10195/A12996). C. Smillie is supported by an MRC Research Studentship to the MRC Human Genetics Unit (HGU). We gratefully acknowledge the work of M. Walker and S. Reid in technical support, R. Wilson (SOCCS3 and COGS study coordinator), G. Barr for data entry in SOCCS studies, the research nurse recruitment teams, the Wellcome Trust Clinical Research Facility for sample preparation and all surgeons, oncologists and pathologists throughout Scotland at contributing centers. Lothian Birth Cohort Illumina genotyping was supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC). Phenotype collection in the Lothian Birth Cohort 1921 was supported by the BBSRC, The Royal Society and The Chief Scientist Office of the Scottish Government. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Research into Ageing (continues as part of Age UK's The Disconnected Mind project). The work on the Lothian Birth Cohorts was undertaken at the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross-council Lifelong Health and Wellbeing Initiative (G0700704/84698). Funding from the BBSRC, EPSRC, ESRC and MRC is gratefully acknowledged.

For the Cambridge study, we thank the SEARCH study team and all the participants in the study. SEARCH is funded by a grant from Cancer Research UK (C490/A10124).

In Cardiff, the work was supported by the Kidani Trust, Tenovus, Cancer Research Wales, The Bobby Moore Fund from CRUK (C10314/A4886), the Wales Assembly Government NISCHR Cancer Genetics BRU and the Wales Gene Park (all to J.P.C.). We acknowledge the use of DNA from the blood samples collected from COIN and COIN-B funded by Cancer Research UK and the MRC. We also acknowledge the use of DNA from the NBS (UKBS) collection, funded by the Wellcome Trust (076113/C/04/Z), by the Juvenile Diabetes Research Foundation (WT061858) and by the NIHR of England. The COIN-B Collaborative Group includes H. Wasan, T. Maughan, R. Adams, R. Wilson, A. Madi, E. Hodgkinson, M. Pope, P. Rogers and J. Cassidy.

Research was also funded by the European Union FP7 (FP7/207-2013) under grant 258236, FP7 collaborative project SYSCOL.

The Swedish sample and data resource were funded by the Swedish Cancer Society, the Swedish Scientific Research Council and the Stockholm Cancer Foundation. We acknowledge the contribution to recruitment and data collection of the Swedish Low-Risk Colorectal Cancer Study Group.

For the Helsinki study, the work was supported by grants from the Academy of Finland (Finnish Centre of Excellence Program; 2006-2011), the Finnish Cancer Society and the Sigrid Juselius Foundation.

This work of the Colon Cancer Family Registry (CFR) was supported by the US National Cancer Institute, National Institutes of Health (CA-95-011), and through cooperative agreements with members of the Colon CFR and Principal Investigators. Collaborating centers include the Australasian Colorectal Cancer Family Registry (U01 CA097735), the Familial Colorectal Neoplasia Collaborative Group (U01 CA074799), the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783) and the Seattle Colorectal Cancer Family Registry (U01 CA074794). The Colon CFR GWAS was supported by funding from the US National Cancer Institute, National Institutes of Health (U01CA122839 to G.C.).

The Japanese study was conducted as part of the BioBank Japan Project that was supported by the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government.

This study made use of genotyping data from the 1958 Birth Cohort and NBS samples, kindly made available by the Wellcome Trust Case Control Consortium 2. A full list of the investigators who contributed to the generation of the data is available (see URLs). Finally, we would like to thank all individuals who participated in the study.

AUTHOR CONTRIBUTIONS

The study was designed and financial support was obtained by R.S.H., I.P.T. and M.G.D. The manuscript was drafted by R.S.H., I.P.T. and M.G.D. Statistical and bioinformatics analyses were conducted by S.E.D. and N.W., with contributions from Y.P.M., M.H., M.F., C.S., G.G., R.S.H., P.B., S.S. and I.P.T.

Institute of Cancer Research and local collaborators: Subject recruitment and sample acquisition to NSCCG were undertaken by S.P. The coordination of sample preparation and genotyping was performed by P.B. Sample preparation and genotyping were performed by J.V., A. Lloyd B.O. and N.W. Tumor pathology analyses were performed by I.C. and S.L.

Oxford and local collaborators: Subject recruitment and sample acquisition were performed by E.B., M.G., L.M. and members of the CORGI Consortium and by R.M. and D.J.K. Genotyping was performed and coordinated by L.G.C.-C., A.M.J. and E.D.

Colon Cancer Genetics Group, Edinburgh, and local collaborators: Subject recruitment and sample acquisition were performed by S.F., J.M.S., I.D., H.C. and M.G.D., as well as by members of the SOCCS and COGS recruitment teams. Sample preparation was coordinated by S.M.F. Genotyping was performed and coordinated by S.M.F. and M.G.D. Data curation and analysis in Edinburgh were conducted by A.T., S.M.F., E.T., L.Z., J.P. and S.B. Recruitment sample preparation, wet lab expression analysis and genotyping were performed by L.-Y.O., and C. Smillie, G.G. and C. Semple performed the bioinformatics analyses.

The following authors from collaborating groups conceived the local or national study, undertook assembly of case-control series in their respective regions, collected data and samples and variously undertook genotyping and analysis: C.G.S., H.W. and J.P.C. in Cardiff; L.R.S., J.P.M. and L.A.A. in Finland; and P.P. in Cambridge. All other authors undertook sample collection and phenotype data collection and collation in their respective centers.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2293.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **343**, 78–85 (2000).
2. Aaltonen, L., Johns, L., Jarvinen, H., Mecklin, J.P. & Houlston, R. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. *Clin. Cancer Res.* **13**, 356–361 (2007).

3. Lubbe, S.J., Webb, E.L., Chandler, I.P. & Houlston, R.S. Implications of familial colorectal cancer risk profiles and microsatellite instability status. *J. Clin. Oncol.* **27**, 2238–2244 (2009).
4. Tomlinson, I.P. *et al.* A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.* **40**, 623–630 (2008).
5. Tomlinson, I.P. *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.* **7**, e1002105 (2011).
6. Tenesa, A. *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat. Genet.* **40**, 631–637 (2008).
7. Houlston, R.S. *et al.* Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat. Genet.* **40**, 1426–1435 (2008).
8. Houlston, R.S. *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.* **42**, 973–977 (2010).
9. Broderick, P. *et al.* A genome-wide association study shows that common alleles of *SMAD7* influence colorectal cancer risk. *Nat. Genet.* **39**, 1315–1317 (2007).
10. Jaeger, E. *et al.* Common genetic variants at the *CRAC1* (*HMPS*) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat. Genet.* **40**, 26–28 (2008).
11. Tenesa, A. & Dunlop, M.G. New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nat. Rev. Genet.* **10**, 353–358 (2009).
12. Miquel, C. *et al.* Frequent alteration of DNA damage signalling and repair pathways in human colorectal cancers with microsatellite instability. *Oncogene* **26**, 5919–5926 (2007).
13. Holm, H. *et al.* Several common variants modulate heart rate, PR interval and QRS duration. *Nat. Genet.* **42**, 117–122 (2010).
14. Abbas, T. & Dutta, A. p21 in cancer: intricate networks and multiple activities. *Nat. Rev. Cancer* **9**, 400–414 (2009).
15. Dunlop, M.G. *et al.* Cancer risk associated with germline DNA mismatch repair gene mutations. *Hum. Mol. Genet.* **6**, 105–110 (1997).
16. Quehenberger, F., Vasen, H.F. & van Houwelingen, H.C. Risk of colorectal and endometrial cancer for carriers of mutations of the hMLH1 and hMSH2 gene: correction for ascertainment. *J. Med. Genet.* **42**, 491–496 (2005).
17. Baglietto, L. *et al.* Risks of Lynch syndrome cancers for *MSH6* mutation carriers. *J. Natl. Cancer Inst.* **102**, 193–201 (2010).
18. Clayton, D. Testing for association on the X chromosome. *Biostatistics* **9**, 593–600 (2008).
19. Farber, M.J., Rizaldy, R. & Hildebrand, J.D. Shroom2 regulates contractility to control endothelial morphogenesis. *Mol. Biol. Cell* **22**, 795–805 (2011).
20. Forbes, S.A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
21. Fairbank, P.D. *et al.* Shroom2 (APXL) regulates melanosome biogenesis and localization in the retinal pigment epithelium. *Development* **133**, 4109–4118 (2006).
22. Houlston, R.S. *et al.* Congenital hypertrophy of retinal pigment epithelium in patients with colonic polyps associated with cancer family syndrome. *Clin. Genet.* **42**, 16–18 (1992).
23. Dunlop, M.G. *et al.* Extracolonic features of familial adenomatous polyposis in patients with sporadic colorectal cancer. *Br. J. Cancer* **74**, 1789–1795 (1996).
24. Matys, V. *et al.* TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34**, D108–D110 (2006).
25. Dimas, A.S. *et al.* Common regulatory variation impacts gene expression in a cell type–dependent manner. *Science* **325**, 1246–1250 (2009).
26. Nica, A.C. *et al.* The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* **7**, e1002003 (2011).
27. Levin, J.H. & Kaler, S.G. Non-random maternal X-chromosome inactivation associated with PHACES. *Clin. Genet.* **72**, 345–350 (2007).
28. Kristiansen, M. *et al.* High incidence of skewed X chromosome inactivation in young patients with familial non-*BRCA1*/*BRCA2* breast cancer. *J. Med. Genet.* **42**, 877–880 (2005).

CORGI Consortium: Huw Thomas[26], Eamonn Maher[27], Gareth Evans[28], Lisa Walker[29], Dorothy Halliday[29], Anneke Lucassen[30], Joan Paterson[31], Shirley Hodgson[32], Tessa Homfray[32], Lucy Side[33], Louise Izatt[34], Alan Donaldson[35], Susan Tomkins[35], Patrick Morrison[36], Carole Brewer[37], Alex Henderson[38], Rosemarie Davidson[39], Victoria Murday[39], Jaqueline Cook[40], Neva Haites[41], Timothy Bishop[42], Eamonn Sheridan[42], Andrew Green[43], Christopher Marks[44], Sue Carpenter[44], Mary Broughton[44], Lynn Greenhalge[45] & Mohnish Suri[46]

The Swedish Low-Risk Colorectal Cancer Study Group: David Edler[20], Claes Lenander[47], Johan Dalén[48], Fredrik Hjern[49], Nils Lundqvist[50], Ulrik Lindforss[51], Lars Påhlman[52], Kennet Smedh[53], Anders Törnqvist[54], Jörn Holm[55], Martin Janson[20], Magnus Andersson[56], Susanne Ekelund[57] & Louise Olsson[58]

The COIN Collaborative Group:

Investigators: Graham MacDonald[59], Leslie M Samuel[59], Athar Ahmad[60], Philippa Corrie[60], Duncan Jodrell[60], Cheryl Palmer[60], Charles Wilson[60], Julie O'Hagan[61], David Smith[61], Ray McDermott[62], Janice Walshe[62], James Cassidy[63], Alec McDonald[63], Nazia Mohammed[63], Jeff White[63], Hosney Yosef[63], Oscar Breathnach[64], Liam Grogan[64], Robert Thomas[65], Martin Eatock[66], Paul Henry[66], Russell Houston[66], Patrick Johnston[66], Richard Wilson[66], Ian Geh[67], Falalu Danwata[68], Andrew Hindley[68], Shabbir Susnerwala[68], Chris Bradley[69], Andrew Conn[69], Anne Raine[69], Chris Twelves[69], Stephen Falk[70], Kirstin Hopkins[70], Saad Tahir[71], Amandeep Dhadda[72], Anthony Maraveyas[72], Joseph Sgouros[72], Mark Teo[72], Rizvana Ahmad[73], Susan Cleator[73], Antonia Creak[73], Charles Lowdell[73], Philippa Riddle[73,74], Kim Benstead[75], David Farrugia[75], Nicholas Reed[75], Stephen Shepherd[75], Edward Levine[76], Saifee Mullamitha[76], Mark Saunders[77], Juan Valle[77], Gregory Wilson[76], Adam Jones[78], Andrew Weaver[78], Peter I Clark[79], Brain Haylock[79], Mussawar I Iqbal[79], Arthur S Myint[79], David Smith[79], Sharon Beesley[80], Timothy Sevitt[80], Jonathan Nicoll[81], Francis Daniel[82], Victoria Ford[82], Toby Talbot[82], Mohammad Butt[83], Abdel Hamid[83], Paul Mack[83], Raj Roy[83], Richard Osborne[84], Fiona McKinna[85], Heidar Alsab[86], Devyanee Basu[86], Phillip Murray[86], Bruce Sizer[86], Faisal A Azam[87], Rekha Neupane[87], Ashita Waterston[63], John Glaholm[88], Claire Blesing[89], Sarah Lowndes[89], Aruna Medisetti[89], Andrew Gaya[90], Martin Leslie[90], Nicholas Maisey[90], Paul Ross[90], Grainne Dunn[91], Omar Al-Salihi[92], Harpreet S Wasan[92], Cheryl Palmer[93], Li Tee Tan[93], Jo Dent[94], Uschi Hofmann[94], Johnathan K Joffe[94], Elizabeth Sherwin[95], Rabinder Soomal[95], Amitabha Chakrabarti[96], Shibu Joseph[96], Johannes Van der Voet[96], Nicholas J Wadd[96], David Wilson[96], Salim Anjarwalia[97], Julia Hall[97], Rob Hughes[98], Andrew Polychronis[98],

John H Scarffe[98], Mark Hill[99], Rodger D James[99], Riyaz Shah[99], Jeffrey Summers[99], Andrew Hartley[100], Desmond Carney[101], John McCaffrey[101], Brano Bystricky[102], Seamus O'Reilly[102], Rajnish Gupta[103], Talib Al-Mishlab[104], Fiona Gidden[104], Richard O'Hara[104], Jill Stewart[104], Richard Ashford[105], Robert Glynne-Jones[105], Mark Harrison[105], Suzannah Mawdsley[106], Hilary Barlow[107], Mary Tighe[107], Julie Walther[107], Janet Neal[108], Charlotte Rees[108], John Bridgewater[109], Stephen Karp[109], Ursula McGovern[109], Philip J Atherton[110], Hany El-deeb[111], Craig Macmillan[111], Kinnari Patel[111], Eric Michael Bessell[112], Peter D Dickinson[112], Vanessa Potter[112], Catherine Jephcott[113], Karen McAdam[113], Jessica Wrigley[113], Richard Osborne[114], Sethupathi Muthuramalingam[115], Anne O'Callaghan[115], John Bridgewater[116], Lucinda Melcher[116], Chiara Braconi[117], Ju Ian Geh[117], Dan Palmer[117], Priya Narayana[117], Neil Steven[117], Andres Gaya[118], Nicholas Maisey[118], Sarah Rudman[119], Prabir Chakraborti[120], Kay Kelly[120], Carol Macgregor[120], David Whillis[121], Alice Freebairn[122], James Gildersleve[122], Sahid Sharif[122], George Astras[123], Tamas Hickish[123], Darren Beech[124], Richard Ellis[124], Rajendra Kulkarni[125], K Shankland[125], Richard Begent[126], Astrid Mayer[126], Timothy Meyer[126], Sandra Strauss[126], Viriginia Hall[127], Sanjay Raj[127], Ian Chau[128], David Cunningham[128], Alison Birtle[129], Ashoke Biswas[129], Shabbir Susnerwala[129], Marcus Wise[129], Sebastian Cummins[130], Sharadah Essapen[130], Gary Middleton[130], Claire Topham[130], Ruth Langley[131], Fiona McKinna[131], Andrew Webb[131], Marie Wilkins[132], Timothy John Iveson[132], Amandeep Dhadda[133], Abdel Hamid[134], Colin Askill[135], John Wagstaff[135], Ashraf Azzabi[136], Adrian Bateman[137], Jan Prejbisz[138], David Tsang[138], Nasim Ali[139], Amanda Jones[139], Paul O'Neill[139], Christopher Cottrill[140], David Propper[140], Fiona Jane Lofts[141], James Kennedy[142], David A Anthoney[143], Rachel Cooper[143], Adrian Crellin[143], Alan Melcher[143], Matthew Seymour[143], Christopher Baughan[144], Emma Alexander[145], Susan Cleator[145], John Crown[146], David Fennelly[146], Fawzi Adab[147], Selvaraj Giridharan[147], Ian Pedley[148], Kathryn Wright[148], Peter Bliss[149], Geoffrey Cogill[149], Nangi Lo[149], Elizabeth Toy[149], John Bridgewater[150], Daniel Hochhauser[150], Jonathan Ledermann[150], Alison Brewster[151], Timothy Maughan[151,152], Diana Mort[151], Somnath Mukherjee[151], Werner Dobrowsky[153], Paula Calvert[154], Gregory Leonard[154], Rizvana Ahmad[73], Hugo Ford[155], Anne Margeret Moody[155], Sonia Goriah[156], Margeret Wilkins[156], Sally Clive[157], Lesley Dawson[157], Catriona McLean[157], Hamish Andrew Phillips[157], Kakumudi Gopi[158], Marjorie Tomlinson[158], Susan Clenton[159], Debra Furniss[159], Joanne Hornbuckle[159], Simon Pledge[159], Jonathan Wadsley[159], Mohammed Abbas[160], Ernest Marshall[160], Catherine Harper-Wynne[161], Anne Barnes[162], Satish Kumar[162], Vallipuram Vigneswaran[162], David Farrugia[163], Andrew Webb[164], Simon Gollins[86], Stephen Falk[165], M Genton[165], Geoff Sparrow[165], Catherine Bale[166], C Fuller[166], Anna Mullard[166], Nicholas Stuart[166], Rachel Williams[166] & Maccon Keane[167]

Trial Management Group: Tim Maughan[152,168], Matthew Seymour[143], Richard Wilson[65], Harpreet Wasan[91], Richard Adams[151], Ayman Madi[151], James Cassidy[62], James Kennedy[142], Elizabeth Hodgkinson[169], Penny Rogers[170] & Malcolm Pope[151]

MRC Clinical Trials Unit: Rick Kaplan[171], Angela Meade[171], Mahesh Parmar[172], Sarah Kenny[171], David Fisher[171], Lynda Harper[171], Jena Mitchell[171], Laura Nichols[171], Benjamin Sydes[171], Louise Clement[171], Edward Kay[171], Cheryl Courtney[171], Madalena Gallagher[171], Claire Murphy[171], Lindsay Thompson[171], Sarah Beall[171], Sharma Hassan[171], Richard Gracie[171] & Gareth Griffiths[152]

Trial Steering Committee: Malcolm Mason[151], Chris Parker[172], Robin Rudd[173], Patrick Johnson[174] & Jeremy Whelan[175]

Independent Data Monitoring Committee: John Northover[176], Julia Brown[177], Matti Aapro[178] & Ronald Stout[75]

[1]Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, University of Edinburgh and Medical Research Council (MRC) Human Genetics Unit, Edinburgh, UK. [2]Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, UK. [3]Wellcome Trust Centre for Human Genetics, Oxford, UK. [4]Department of Histopathology, Royal Devon and Exeter Hospital, Exeter, UK. [5]Public Health Sciences, University of Edinburgh, Edinburgh, UK. [6]A full list of members and affiliations appears at the end of this paper. [7]University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, Edinburgh, UK. [8]Department of Surgical Oncology, University Hospital for Tumors, University Hospital Center Sestre Milosrdnice, Zagreb, Croatia. [9]Department of Surgery, University Hospital Center Sestre Milosrdnice, Zagreb, Croatia. [10]Department of Medical Genetics, Genome-Scale Biology Research Program, Biomedicum Helsinki, University of Helsinki, Helsinki, Finland. [11]Department of Surgery, Helsinki University Central Hospital, Helsinki, Finland. [12]Professor of General Surgery, Jyväskylä Central Hospital, University of Eastern Finland, Jyväskylä, Finland. [13]Laboratory of Molecular Medicine, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [14]Laboratory for Statistical Analysis, Center for Genomic Medicine, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan. [15]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. [16]Translational Genomics Research Institute (TGen), Phoenix, Arizona, USA. [17]Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [18]Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA. [19]Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, The University

of Melbourne, Melbourne, Victoria, Australia. [20]Department of Oncology, University of Cambridge, Cambridge, UK. [21]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [22]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. [23]Institute of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff, UK. [24]Department of Oncology, Oxford University, Radcliffe Infirmary, Headington, Oxford, UK. [25]Oxford National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre, Oxford, UK. [26]Family Cancer Clinic, St Mark's Hospital, Harrow and Imperial College, London, UK. [27]Department of Clinical Genetics, University of Birmingham, Birmingham, UK. [28]Department of Clinical Genetics, University of Manchester, Manchester, UK. [29]Oxford Regional Genetics Service, Churchill Hospital, Oxford, UK. [30]Wessex Regional Genetics Service, Princess Anne Hospital, Southampton, UK. [31]Anglia Regional Genetics Service, Addenbrooke's Hospital, Cambridge, UK. [32]South-West Thames Regional Genetics Service, St George's Hospital, London, UK. [33]North-East Thames Regional Genetics Service, Great Ormond Street Hospital, London, UK. [34]South-East Thames Regional Genetics Service, Guy's Hospital, London, UK. [35]South-West Regional Genetics Service, Bristol, UK. [36]Northern Ireland Regional Genetics Service, City Hospital, Belfast, UK. [37]South-West Regional Genetics Service, Royal Devon and Exeter Hospital, Exeter, UK. [38]Northern Regional Genetics Service, International Centre for Life, Newcastle, UK. [39]West of Scotland Regional Genetics Service, Yorkhill Hospital, Glasgow, UK. [40]Sheffield Regional Genetics Service, Children's Hospital, Sheffield, UK. [41]North of Scotland Regional Genetics Service, Foresterhill Hospital, Aberdeen, UK. [42]Yorkshire Regional Genetics Service, St James's Hospital, Leeds, UK. [43]Republic of Ireland Genetics Service, Our Lady's Hospital for Sick Children, Dublin, Ireland. [44]The Royal Surrey County Hospital, Guildford, UK. [45]Department of Clinical Genetics, Royal Liverpool Children's Hospital, Alderhay, Liverpool, UK. [46]Department of Clinical Genetics, City Hospital, Nottingham, UK. [47]Mag-Tarm-Centrum, Ersta Sjukhus, Stockholm, Sweden. [48]St Görans Sjukhus, Stockholm, Sweden. [49]Institutionen för Kliniska Vetenskaper, Danderyds Sjukhus, Danderyd, Sweden. [50]Norrtälje Sjukhus, Norrtälje, Sweden. [51]Södertälje Sjukhus, Södertälje, Sweden. [52]Akademiska Sjukhuset, Uppsala, Sweden. [53]Centrallasarettet, Västerås, Sweden. [54]Centralsjukhuset, Karlstad, Sweden. [55]Länssjukhuset Gävle-Sandviken, Gävle, Sweden. [56]Universitetssjukhuset, Örebro, Sweden. [57]Södersjukhuset, Stockholm, Sweden. [58]Mälarsjukhuset, Eskilstuna, Sweden. [59]Aberdeen Royal Infirmary, Foresterhill, Aberdeen, UK. [60]Department of Oncology, Addenbrooke's Hospital, UK. [61]Aintree, University Hospital, UK. [62]Adelaide and Meath Hospital, Dublin, Ireland. [63]Beatson Oncology Centre, Glasgow, UK. [64]Beaumont Hospital, Dublin, Ireland. [65]Bedford Hospital NHS Trust, Bedford, UK. [66]Belfast City Hospital, Belfast, UK. [67]Birmingham Heartlands Hospital, Birmingham, UK. [68]Blackpool Victoria Hospital, Blackpool, UK. [69]Bradford Royal Infirmary, Bradford, UK. [70]Bristol Haematology and Oncology Centre, Bristol, UK. [71]Broomfield Hospital, Essex, UK. [72]Castle Hill Hospital, Hull, UK. [73]Charing Cross Hospital, London, UK. [74]West Middlesex Hospital, Isleworth, Middlesex, UK. [75]Cheltenham General Hospital, Cheltenham, UK. [76]Christie Hospital, Manchester, UK. [77]Clinical and Translational Research Team, Christie Hospital, Manchester, UK. [78]Churchill Hospital, Oxford, UK. [79]Clatterbridge Centre for Oncology, Wirral, UK. [80]Conquest Hospital, St Leonards-on-Sea, UK. [81]Cumberland Infirmary, Carlisle, UK. [82]Derriford Hospital, Plymouth, UK. [83]Diana, Princess of Wales Hospital, Grimsby, UK. [84]Dorset County Hospital, Dorchester, UK. [85]Eastbourne District General, Eastbourne, UK. [86]Essex County Hospital, Colchester, UK. [87]Glan Clwyd Hospital, Rhyl, UK. [88]Good Hope Hospital, Birmingham, UK. [89]Great Western Hospital, Swindon, UK. [90]Guy's & St Thomas' NHS Foundation Trust, London, UK. [91]Hairmyres Hospital, East Kilbride, UK. [92]Hammersmith Hospital, London, UK. [93]Hinchingbrooke Hospital, Huntingdon, UK. [94]Huddersfield Royal Infirmary, Huddersfield, UK. [95]Ipswich Hospital NHS Trust, Ipswich, UK. [96]James Cook University Hospital, Middlesborough, UK. [97]Kent and Canterbury Hospital, Canterbury, UK. [98]Lister Hospital, London, UK. [99]Maidstone Hospital, Maidstone, UK. [100]Manor Hospital, Walsall, UK. [101]Mater Private Hospital, Dublin, Ireland. [102]Mercy Hospital, Cork, Ireland. [103]Midwestern Regional Hospital, Limerick, Ireland. [104]Milton Keynes General Hospital, Milton Keynes, UK. [105]Cancer Services, Mount Vernon Hospital, Northwood, UK. [106]Mount Vernon Hospital, Northwood, UK. [107]Musgrove Park Hospital, formerly Taunton and Somerset, Taunton, UK. [108]North Hampshire Hospital, Basingstoke, UK. [109]North Middlesex Hospital, London, UK. [110]North Tyneside General Hospital, North Shields, UK. [111]Northampton General, Northampton, UK. [112]Nottingham City Hospital, Nottingham, UK. [113]Peterborough District, Peterborough, UK. [114]Poole Hospital, Poole, UK. [115]Portsmouth Oncology Centre, Queen Alexandra Hospital, Portsmouth, UK. [116]Princess Alexandra Hospital, Harlow, UK. [117]Queen Elizabeth, University Hospital, Birmingham, UK. [118]Queen Elizabeth Hospital, Woolwich, UK. [119]Cancer Studies, School of Medicine, Kings College London, London, UK. [120]London Road Community Hospital, Derby, UK. [121]Raigmore Hospital, Inverness, UK. [122]Cancer Services, Royal Berkshire, Reading, UK. [123]Royal Bournemouth Hospital, Bournemouth, UK. [124]Royal Cornwall, Truro, UK. [125]Royal Derby Hospital, Derby, UK. [126]Royal Free Hospital, London, UK. [127]Royal Hampshire County Hospital, Winchester, UK. [128]Royal Marsden NHS Foundation Trust, Sutton, UK. [129]Royal Preston Hospital, Lancashire, UK. [130]Royal Surrey County Hospital, St Luke's, UK. [131]Royal Sussex County Hospital, Brighton, UK. [132]Salisbury District, Salisbury, UK. [133]Scarborough Hospital, Scarborough, UK. [134]Scunthorpe General, Scunthorpe, UK. [135]Singleton Hospital, Swansea, UK. [136]South Tyneside District Hospital, South Shields, UK. [137]Southampton Hospital, Southampton, UK. [138]Southend Hospital, Westcliff-on-Sea, UK. [139]Southport & Formby District General Hospital, Southport, UK. [140]St Bartholomew's Hospital, City of London, London, UK. [141]St George's Hospital, London, UK. [142]St James' Hospital, Dublin, UK. [143]St James' University Hospital, Leeds, UK. [144]St Mary's Hospital Newport, Sunderland, UK. [145]St Mary's Hospital, London, UK. [146]St Vincent's Hospital, Dublin, Ireland. [147]University Hospital North Staffordshire, Stoke-on-Trent, UK. [148]Sunderland Royal Hospital, UK. [149]Torbay Hospital, Torquay, Devon, UK. [150]University College Hospital, London, UK. [151]Velindre Hospital, Cardiff, UK. [152]Wales Cancer Trials Unit, University Hospital of Wales, Cardiff, UK. [153]Wansbeck General Hospital, Ashington, UK. [154]Waterford Regional Hospital, Waterford, Ireland. [155]West Suffolk Hospital, Suffolk, UK. [156]West Wales General Hospital, Carmarthen, UK. [157]Western General Hospital, Edinburgh, UK. [158]Western General Hospital, Weston Super Mare, UK. [159]Weston Park Hospital, Sheffield, UK. [160]Whiston Hospital, Merseyside, UK. [161]William Harvey Hospital, Ashford, UK. [162]Withybush General Hospital, Pembrokeshire, UK. [163]Worcestershire Royal Hospital, Worcester, UK. [164]Worthing Hospital, Worthing, UK. [165]Yeovil District Hospital, Yeovil, UK. [166]Ysbyty Gwynedd, Bangor, UK. [167]University College Hospital, Galway, Ireland. [168]Gray Institute for Radiation Oncology and Biology, University of Oxford, Oxford, UK. [169]Pharmacy Department, Weston Park Hospital, Sheffield, UK. [170]Radiotherapy Department, Charing Cross Hospital, London, UK. [171]MRC Clinical Trials Unit, London, UK. [172]Section of Academic Radiotherapy, Institute of Cancer Research, Sutton, UK. [173]Barts Mesothelioma Research, London, UK. [174]School of Medicine, Dentistry and Biomedical Sciences, Queen's University, Belfast, UK. [175]Department of Oncology, University College Hospitals NHS Foundation Trust, London, UK. [176]The London Clinic, London, UK. [177]Section of Clinical Trials Research, University of Leeds, UK. [178]Institut Multidisciplinaire d'Oncologie, Clinique de Genolier, Switzerland. Correspondence should be addressed to M.G.D. (malcolm.dunlop@hgu.mrc.ac.uk), I.P.T. (iant@well.ox.ac.uk) or R.S.H. (richard.houlston@icr.ac.uk).

ONLINE METHODS

Ethics statement. Collection of blood samples and clinicopathological information from subjects was undertaken with informed consent and ethical review board approval at all sites in accordance with the tenets of the Declaration of Helsinki.

Data sets, sample preparation and genotyping. Full details of each data set are provided in the **Supplementary Note**.

DNA was extracted from samples using conventional methods and quantified using PicoGreen (Invitrogen). The VQ, UK1 and Scotland1 GWAS cohorts were genotyped using Illumina Hap300, Hap240S, Hap370 or Hap550 arrays. The 1958BC and NBS samples were genotyped as part of the WTCCC2 study on Hap1.2M-Duo Custom arrays. The CCFR1 samples were genotyped using Illumina Hap1M or Hap1M-Duo arrays. In UK2 and Scotland2, genotyping was conducted using custom Illumina Infinium arrays, according to the manufacturer's protocols. Some SNPs from COIN samples were genotyped on custom Illumina GoldenGate arrays. To ensure quality of genotyping, a series of duplicate samples was genotyped, resulting in 99.9% concordant calls in all cases. Other genotyping was conducted using competitive allele-specific PCR KASPar chemistry (KBiosciences), TaqMan (Life Sciences) or MassARRAY (Sequenom). Details of all primers, probes and conditions used are available upon request. Genotyping quality control was tested using duplicate DNA samples within studies and SNP assays, together with direct sequencing of subsets of samples to confirm genotyping accuracy. For all SNPs, >99% concordant results were obtained.

Quality control and sample exclusion. We excluded SNPs from analysis if they failed one or more of the following tests: GenCall scores of <0.25; overall call rates of <95%; MAF of <0.01; departure from Hardy-Weinberg equilibrium (HWE) in controls at $P < 1 \times 10^{-4}$ or in cases at $P < 1 \times 10^{-6}$; outliers in terms of signal intensity or X:Y ratio; showed discordance between duplicate samples; or, for SNPs with evidence of association, poor clustering on inspection of X:Y plots. We excluded individuals from analysis if they failed one or more of the following tests: duplication or cryptic relatedness to estimated identity by descent (IBD) of >6.25%; overall successfully genotyped SNPs of <95%; mismatch between predicted and reported gender; outliers in a plot of heterozygosity versus missingness; and evidence of non-European ancestry in principal-components analysis–based testing in comparison with HapMap samples. Details of all sample exclusions are provided (**Supplementary Fig. 1**).

To identify individuals who might have non-northern European ancestry, we merged our case and control data from all sample sets with the 60 European (CEU), 60 Nigerian (YRI) and 90 Japanese (JPT) and 90 Han Chinese (CHB) individuals from the International HapMap Project. For each pair of individuals, we calculated genome-wide identity-by-state distances on the basis of markers shared between HapMap 2 and our SNP panel and used these as dissimilarity measures upon which to perform principal-components analysis. Principal-components analysis was performed in R using CEU, YRI and HCB HapMap samples as references. The first two principal components for each individual were plotted, and any individual not present in the main CEU cluster (>5% of the principal-component distance from the HapMap CEU cluster centroid) was excluded from subsequent analyses (**Supplementary Fig. 3**).

We had previously shown the adequacy of the case-control matching and possibility of differential genotyping of cases and controls using quantile-quantile plots of test statistics. The inflation factor λ_{GC} was calculated by dividing the mean of the lower 90% of the test statistics by the mean of the lower 90% of the expected values from a χ^2 distribution with 1 degree of freedom. Deviation of the genotype frequencies in the controls from those expected under HWE was assessed by χ^2 test (1 degree of freedom) or by Fisher's exact test, where the expected cell count was <5.

Statistical and bioinformatic analysis. Main analyses were undertaken using R (v2.6), STATA v.11 and PLINK (v1.06) software[29]. The association between each SNP and risk of CRC was assessed by the Cochran-Armitage trend test. ORs and associated 95% CIs were calculated by unconditional logistic regression. Meta-analysis was conducted using standard methods[30]. Cochran's Q statistic to test for heterogeneity[30] and the I^2 statistic to quantify the proportion of the total variation due to heterogeneity were calculated[31]. I^2 values of ≥75% are considered characteristic of large heterogeneity[31,32]. Associations by sex, age and clinicopathological phenotypes were examined by logistic regression in case-only analyses.

For SNPs on the non-pseudoautosomal region of the X chromosome, males carry only one copy, and, in females, most loci are subject to X inactivation[33]. To test for X-chromosome associations, we used an extension to the standard 1-degree-of-freedom Cochran-Armitage test for trend[18] whereby males can be regarded as homozygous females. This 1-degree-of-freedom trend test adjusts for the different variances for males and females.

Prediction of the ungenotyped SNPs was carried out using IMPUTEv2, based on HapMap Phase 3 haplotypes, release 2 (HapMap Data Release 27, phase 3, February 2009, on NCBI Build 36 assembly, dbSNP26) and 1000 Genomes Project data. Imputation of the X-chromosome loci was only possible using IMPUTEv1 with HapMap Data Release 21 on NCBI Build 35. Imputed data were analyzed using SNPTEST v2 to account for uncertainties in SNP prediction. An imputation info score of 0.95 was used to remove SNPs with poor imputation quality. LD metrics between HapMap SNPs were based on Data Release 27, phase 3 (February 2009), on NCBI Build 36 assembly, dbSNP26, viewed using Haploview software (v4.2) and plotted using SNAP. LD blocks were defined on the basis of HapMap recombination rate, as defined using the Oxford recombination hotspots[34] and on the basis of distribution of confidence intervals as defined[35]. To annotate potential regulatory sequences within disease loci, we implemented *in silico* searches using Transfac Matrix Database v7.29 (ref. 24) and PReMod10 (ref. 36) software. We used the *in silico* algorithms SIFT and PolyPhen to predict the impact of amino-acid substitutions.

Relationship between SNP genotype and mRNA expression. *Expression studies in colonic epithelium.* To examine for a relationship between SNP genotype and mRNA expression in colonic epithelium, 42 samples were collected fresh immediately after surgical resection of specimens for colorectal cancer ($n = 34$), solitary adenoma ($n = 5$) or benign conditions (not inflammatory bowel disease) ($n = 3$). For 2 of the 42 subjects, 3 samples of mucosa were harvested from different locations of the fresh resected bowel. Normal epithelium was dissected from muscularis propria, and samples were snap frozen and placed in RNAlater (Applied Biosystems) and kept at 4 °C overnight before storage at −80 °C. Tissue was disrupted and homogenized using TissueLyser LT (Qiagen), and RNA was extracted using the Ribopure kit (Applied Biosystems). RNA integrity and concentration were assessed on an Agilent Bioanalyzer, and RNA purity (A260/A280 and A260/A230) was assessed by Nanodrop. RT-PCR products were analyzed on HumanHT-12 Expression BeadChips, which were scanned using Illumina HiScan. Array data processing and analysis were performed using Illumina GenomeStudio software (version 2011.1). Microarray data were exported from Illumina BeadStudio software, processed and normalized using the R Bioconductor beadarray and limma packages[37,38]. Before normalization, probes that were not detected (detection P value > 0.01) on the microarrays were removed. Microarrays were quantile normalized to remove technical variation. Three mucosal samples were available for 2 of the 42 subjects, and, for these, we used the average signal of the replicates in the analysis. The limma package was used to find differentially expressed genes, using the functions lmFit, eBayes and topTable. To test all associations between SNPs and expression, a linear model was fitted to the expression level of each probe, using this genotype value as effect. For SNP associations with gene expression on the X chromosome, gender was added to the model. Significant associations were considered as <0.05, using P values adjusted for multiple testing with the Benjamini-Hochberg method from R's p.adjust function.

In silico analysis of publicly available expression data. We analyzed expression data generated from (i) fibroblasts, lymphoblastoid cell lines (LCLs) and T cells derived from the umbilical cords of 75 Geneva GenCord individuals[25]; (ii) 166 adipose, 156 LCL and 160 skin samples derived from a subset of healthy female twins of the MuTHER resource[26] using Sentrix Human-6 Expression BeadChips (Illumina)[39,40]; and (iii) AgilentG4502A_07_3 custom gene expression data on 154 CRCs obtained as part of the Cancer Genome Atlas project. Power of assays to establish a relationship between genotype and expression was evaluated using STATA software.

Assignment of microsatellite instability in colorectal cancers. Tumor MSI status in CRCs was determined using the mononucleotide microsatellite loci BAT25 and BAT26, which are highly sensitive MSI markers. Briefly, 10-mm sections were cut from formalin-fixed paraffin-embedded CRC tumors and lightly stained with toluidine blue, and regions containing at least 60% tumor were microdissected. Tumor DNA was extracted using the QIAamp DNA Mini kit (Qiagen), according to the manufacturer's instructions, and genotyped for the BAT25 and BAT26 loci using either [^{32}P]-labeled or fluorescently labeled oligonucleotide primers (in the UK2-UK3 and COINNBS studies, respectively). Samples showing at least five novel alleles compared to normal DNA at either or both markers were assigned as MSI-high (MSI-H)[41].

29. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
30. Pettiti, D. *Meta-analysis Decision Analysis and Cost-effectiveness Analysis* (Oxford University Press, 1994).
31. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
32. Ioannidis, J.P., Ntzani, E.E. & Trikalinos, T.A. 'Racial' differences in genetic effects for complex diseases. *Nat. Genet.* **36**, 1312–1318 (2004).
33. Chow, J.C., Yen, Z., Ziesche, S.M. & Brown, C.J. Silencing of the mammalian X chromosome. *Annu. Rev. Genomics Hum. Genet.* **6**, 69–92 (2005).
34. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005).
35. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
36. Ferretti, V. *et al.* PReMod: a database of genome-wide mammalian *cis*-regulatory module predictions. *Nucleic Acids Res.* **35**, D122–D126 (2007).
37. Dunning, M.J., Smith, M.L., Ritchie, M.E. & Tavare, S. beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics* **23**, 2183–2184 (2007).
38. Smyth, G.K. Limma: linear models for microarray data. in *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (eds. Gentleman, R., Carey, V., Dudoit, S., Irizarry, R. & Huber, W.) 397–420 (Springer, New York, 2005).
39. Stranger, B.E. *et al.* Genome-wide associations of gene expression variation in humans. *PLoS Genet.* **1**, e78 (2005).
40. Stranger, B.E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
41. Boland, C.R. *et al.* A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* **58**, 5248–5257 (1998).

doi:10.1038/ng.2293

uicc
global cancer control

IJC
International Journal of Cancer

Serum level of adiponectin and the risk of liver cancer development in chronic Hepatitis C patients

Toru Arano[1*], Hayato Nakagawa[1,2*], Ryosuke Tateishi[1], Hitoshi Ikeda[2], Koji Uchino[1], Kenichiro Enooku[1], Eriko Goto[1], Ryota Masuzaki[1], Yoshinari Asaoka[1], Yuji Kondo[1], Tadashi Goto[1], Shuichiro Shiina[1], Masao Omata[3], Haruhiko Yoshida[1] and Kazuhiko Koike[1]

[1] Department of Gastroenterology, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655
[2] Department of Clinical Laboratory Medicine, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655
[3] Yamanashi Prefectural Hospital Organization, 1-1-1, Fujimi, Kofu, Yamanashi 400-8506

Obesity and metabolic syndrome are recognized risk factors for development of hepatocellular carcinoma (HCC) in patients with chronic hepatitis C (CHC). Dysregulation of adipokines, particularly the decreased secretion of adiponectin, appears to play a key role. To investigate the association between adiponectin and hepatocarcinogenesis, we conducted a large-scale retrospective cohort study. We enrolled 325 patients with CHC (146 men, 179 women; mean age 58.0 ± 10.3 years) whose serum samples were collected between January 1994 and December 2002. Subjects were divided into two groups according to their serum adiponectin levels. We evaluated the association between adiponectin level and the risk of subsequent HCC development using univariate and multivariate Cox proportional hazard regression. Because average serum adiponectin level was higher in females than males, each gender was analyzed separately. Patients with CHC had significantly higher adiponectin levels than healthy controls. During the follow-up period (mean: 9.0 years), HCC developed in 122 subjects. Unexpectedly, subjects with higher serum adiponectin levels had a higher incidence of HCC (males: $p = 0.032$; females: $p = 0.01$; log-rank test). Multivariate analysis revealed that a high serum adiponectin level was independently associated with HCC development (hazard ratio [HR] = 2.07; $p = 0.031$ in females and HR = 1.82; $p = 0.05$ in males). Isoform analysis revealed that middle- and low-molecular-weight isoforms contributed to the risk of HCC. In conclusion, Patients who had CHC with high serum adiponectin levels had a higher risk of liver cancer development. Adiponectin may thus be tumorigenic or indicate a liver disease state independently of other clinical parameters.

Epidemiology

Hepatocellular carcinoma (HCC) is the fifth most common cancer worldwide, with an increasing incidence globally.[1,2] Recently, obesity and metabolic syndrome were shown in several epidemiologic studies to increase the risk of HCC.[3–5] Because the prevalence of obesity and metabolic syndrome has been increasing in both Japan and Western nations, a possible association between obesity and hepatocarcinogenesis has attracted considerable attention in recent years.

The mechanism by which obesity and metabolic syndrome promote hepatocarcinogenesis remains not fully understood. However, obesity-induced dysregulation of adipokines, cytokines secreted by adipose tissue, is considered to play a key role.[6,7] Adipose tissue controls the functions of other organs through the secretion of various adipokines such as leptin, adiponectin, tumor necrosis factor α (TNFα), interleukin-6 (IL-6), and resistin. Obesity with visceral fat accumulation increases the levels of leptin, TNFα, IL-6, and resistin, and decreases adiponectin levels.[6,7] These adipokines flow directly into the liver through the portal vein and exert a variety of effects on liver diseases.[8]

Adiponectin, one of the major adipokines, possesses anti-inflammatory and insulin-sensitizing properties, and levels typically decline with increasing body weight.[9] Hypoadiponectinemia has been implicated in the development of obesity-related morbidities such as dyslipidemia and cerebrovascular disease.[10–12] In addition, hypoadiponectinemia has been reported to enhance hepatic steatosis, inflammation, fibrosis,

Key words: hepatocellular carcinoma, carcinogenesis, chronic hepatitis C, adiponectin

Abbreviations: AFP: alpha fetoprotein; ALT: alanine aminotransferase; AST: aspartate aminotransferase; BMI: body mass index; CHC: chronic hepatitis C; CI: confidence interval; HCC: hepatocellular carcinoma; HCV: hepatitis C virus; HR: hazard ratio; IL-6: interleukin-6

Additional Supporting Information may be found in the online version of this article.

*T.A. and H.N. contributed equally to this work

Grant sponsor: Ministry of Health Labour and Welfare of Japan (H20-kannenn-008)

DOI: 10.1002/ijc.25861

History: Received 16 Aug 2010; Accepted 30 Nov 2010; Online 17 Dec 2010

Correspondence to: Hayato Nakagawa, Department of Gastroenterology, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Tel.: +81-3-3815-5411, Fax: +81-3-3814-0021, E-mail: n-hayato@yf7.so-net.ne.jp