Refinement of the basis and impact of common 11q23.1 variation to the risk of developing colorectal cancer

Alan M. Pittman¹, Emily Webb¹, Luis Carvajal-Carmona², Kimberley Howarth², Maria Chiara Di Bernardo¹, Peter Broderick¹, Sarah Spain², Axel Walther², Amy Price¹, Kate Sullivan¹, Philip Twiss¹, Sarah Fielding¹, Andrew Rowan², Emma Jaeger², Jayaram Vijayakrishnan¹, Ian Chandler¹, Steven Penegar¹, Mobshra Qureshi¹, Steven Lubbe², Enric Domingo², Zoe Kemp², Ella Barclay², Wendy Wood¹, Lynn Martin²,³,⁹,ⁱ⁰,¹¹,¹³, Maggie Gorman¹, Huw Thomas⁴, Julian Peto⁴,⁵,⁶, Timothy Bishop⁷, Richard Gray⁸, Eamonn R. Maher⁹,¹⁰, Anneke Lucassen¹¹, David Kerr¹², Gareth R. Evans¹³, The CORGI Consortium†, Tom van Wezel¹⁴, Hans Morreau¹⁴,¹⁵, Juul T. Wijnjen¹⁴,¹⁵, John L. Hopper¹⁶, Melissa C. Southey¹⁷, Graham G. Giles¹⁶,¹⁸, Gianluca Severi¹⁸, Sergi Castellví-Bel¹⁹, Clara Ruiz-Ponte²⁰, Angel Carracedo²⁰, Antoni Castells¹⁹, The EPICOLON Consortium², Asta Först¹²,²², Kari Hemminki²¹,²², Pavel Vodicka²³, Alessio Naccarati²³, Lara Lipton²⁴, Judy W.C. Ho²⁵, K.K. Cheng²⁵, Pak C. Sham²⁵, J. Luk²⁵, Jose A.G. Agüéndez²⁶, Jose M. Ladero²⁷, Miguel de la Hoya²⁸, Trinidad Caldeš²⁸, Iina Niittymäki²⁹, Sari Tuupanen²⁹, Auli Karhu²⁹, Lauri A. Aaltonen²⁹, Jean-Baptiste Cazier³⁰, Ian P.M. Tomlinson² and Richard S. Houlston¹,*

¹Section of Cancer Genetics, Institute of Cancer Research, Sutton SM2 5NG, UK, ²Molecular and Population Genetics Laboratory, London Research Institute, Cancer Research UK, London WC2A 3PX UK, ³Family Cancer Clinic, St Mark’s Hospital, Watford Road, Harrow HA1 3UJ, UK, ⁴Non-Communicable Disease Epidemiology Unit, London School of Hygiene and Tropical Medicine, London, London, UK, ⁵Institute of Cancer Research, Surrey, UK, ⁶Cancer Research UK Epidemiology and Genetics Unit, Institute of Cancer Research, London, UK, ⁷Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, University of Leeds, Leeds, UK, ⁸Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, UK, ⁹Department of Medical and Molecular Genetics, University of Birmingham School of Medicine, ¹⁰West Midlands Regional Genetics Service, Birmingham Women’s Hospital, Edgbaston, Birmingham, UK, ¹¹University of Southampton and Wessex Clinical Genetics Service, Mailpoint 105, Princess Anne Hospital, Oxford Road, Southampton SO16 5YA, UK, ¹²Department of Clinical Pharmacology, Oxford University, Radcliffe Infirmary, Oxford OX2 6HA, UK, ¹³Department of Medical Genetics, St Mary’s Hospital, Manchester M13 0JH, UK, ¹⁴Department of Pathology and ¹⁵Center for Human and Clinical Genetics, Leiden University Medical Center, Albinusdreef 2 2333ZA, Leiden, The Netherlands, ¹⁶Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, 723 Swanston street, Carlton VIC 3053, Australia, ¹⁷Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Grattan street, Parkville VIC 3052, Australia, ¹⁸Cancer Epidemiology Centre, The Cancer Council Victoria, 1 Rathdowne street, Carlton VIC 3053, Australia, ¹⁹Department of Gastroenterology, Institut de Malalties Digestives i Metabòliques, Hospital Clínic, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBER-EHD), IDIBAPS, Barcelona, Catalonia, Spain, ²⁰Genomic Medicine Group, Fundacion Publica Galega de Medicina Xenonica

†See Supplementary data.

# The Author 2008. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org
The common single-nucleotide polymorphism (SNP) rs3802842 at 11q23.1 has recently been reported to be associated with risk of colorectal cancer (CRC). To examine this association in detail we genotyped rs3802842 in eight independent case–control series comprising a total of 10 638 cases and 10 457 healthy individuals. A significant association between the C allele of rs3802842 and CRC risk was found (per allele OR = 1.17; 95% confidence interval [CI]: 1.12–1.22; \( P = 1.08 \times 10^{-12} \)) with the risk allele more frequent in rectal than colon cancer (\( P = 0.02 \)). In combination with 8q21, 8q24, 10p14, 11q, 15q13.3 and 18q21 variants, the risk of CRC increases with an increasing numbers of variant alleles for the six loci (ORper allele = 1.19; 95% CI: 1.15–1.23; \( P_{\text{trend}} = 7.4 \times 10^{-24} \)). Using the data from our genome-wide association study of CRC, LD mapping and imputation, we were able to refine the location of the causal locus to a 60 kb region and screened for coding changes. The absence of exonic mutations in any of the transcripts (FLJ45503, LOC120376, C11orf53 and POU2AF1) mapping to this region makes the association likely to be a consequence of non-coding effects on gene expression.

INTRODUCTION

Inherited susceptibility is likely to play a role in the development of \(~35\%\) of colorectal cancer (CRC) (1). High-risk, germline mutations in \(APC, MLH1, MSH2\) and \(MYH\), however, account for \(<5\%\) of CRC (2), and it has been hypothesized that much of the remaining inherited predisposition is likely to be a consequence of multiple low-risk alleles.

Recent technological developments have allowed the search for common risk variants to be conducted by genome-wide association studies (GWASs). The SNP rs6983267 (chromosome 8q24.21) was the first to be identified as a common susceptibility variant for CRC through a GWAS and has subsequently been robustly validated in independent studies (3–5). A second susceptibility locus at 18q21.1 defined by \(SMAD7\) has subsequently been identified and independently validated (6,7). Most recently, we have identified three further CRC variants located at 8q23.3 (rs16892766), 10p14 (rs10795668) (8) and 15q13.3 (rs4779584) (9). At the same time, a further novel CRC variant rs3802842, mapping to 11q23, was identified (7).

To examine the 11q23 association in detail, we genotyped rs3802842 in eight independent case–control series comprising a total of 10 638 cases and 10 457 healthy individuals. Using the data from our GWAS of CRC, linkage disequilibrium (LD) mapping and imputation, we were able to refine the location of the causal locus to a 60 kb region and screened for coding changes in all transcripts mapping to this region. We also investigated the joint impact of 11q23 and all common low-risk variants identified to date.

RESULTS

A total of 10 374 patients with CRC (97.5%) and 10 248 controls (98.0%) from the eight were successfully genotyped for rs3802842 (Table 1). Call rates were consistently >90% for samples in the different study cohorts and there was no bias in the form of differential call rates between cases and controls in any of the series. Confirmation of genotypes through sequencing showed >99.99% concordance. In none of the eight series was there evidence of population stratification as evidenced by a departure of control genotypes from Hardy–Weinberg equilibrium (HWE). The frequency of the C allele among controls ranged between 0.24 and 0.32 in the different cohorts, similar to previously published frequencies in Caucasian populations.

In all eight series, the C allele was associated with an increased risk of CRC, statistically significant in five (Fig. 1). Pooling data from the eight series provided unequivocal evidence for an association between rs3802842 and risk of CRC (per allele OR = 1.17; 95% CI: 1.12–1.22; \( P = 1.08 \times 10^{-12} \); \( P_{\text{het}} = 0.25 \), \( P = 23\% \); Table 1); marginally greater than the OR of 1.11 (95% CI: 1.08–1.15) observed by
Tenesa et al. (7). The C allele was associated with an increased risk of CRC in a dose-dependent manner, most parsimonious with a multiplicative model. In the pooled analysis, the risks of CRC associated with CC homozygosity and AC heterozygosity were increased 1.35-fold (95% CI: 1.22–1.49; \( P = 6.25 \times 10^{-2} \)) and 1.18-fold (95% CI: 1.11–1.25; \( P = 3.80 \times 10^{-2} \)), respectively.

Cases in two of the series were enriched for familial CRC (CORGI and DFCCS); hence, the OR may be biased away from unity. We therefore also computed pooled ORs with analysis restricted to data from the six series unselected for family history. Odds ratios were marginally closer to unity; per allele OR = 1.15 (95% CI: 1.09–1.20), ORhet = 1.16 (95% CI: 1.09–1.24) and OR hom = 1.29 (95% CI: 1.15–1.43).

The detailed patient data available from the case–control series enabled us to study the possible association of rs3802842 with clinical characteristics of CRC. The results for gender were based on all case series (complete data from 10 364 cases); results on age at diagnosis were based on all case series except CORGI (9635 cases); results for site were based on data from FCCPS, MCCS, NSCCG1, NSCCG2 and VCQ (8292 cases); results for family history status were based on data from EPICOLON, FCCPS, NSCCG1 and NSCCG2 (8083 cases); and results for MSI status were based on data from NSCCG1 and NSCCG2 (1839 cases).

There was no evidence that the association between CRC risk and rs3802842 genotype was modified by gender (\( P = 0.18 \)), family history of CRC (\( P = 0.21 \)), age at diagnosis (\( P = 0.71 \)) or MSI status (\( P = 1.00 \)) (Table 2). However, the risk allele C was more frequent in patients with rectal rather than colonic disease (\( P = 0.02 \); Table 2); so that the risk of rectal cancer associated with rs3802842 was thus significantly higher (per allele OR = 1.20 [95% CI: 1.12–1.27] than the risk of colonic cancer (OR = 1.10 [95% CI: 1.04–1.16]).

There was limited evidence for between-study heterogeneity in the risk of colonic cancer (\( P_{het} = 0.16, I^2 = 40\% \)) and no evidence of heterogeneity for the risk of rectal cancer (\( P_{het} = 0.75, I^2 = 0\% \)). These results are concordant with the observations of Tenesa et al. (7).

On the basis of GWAS data from CORGI and NSCCG1, rs3802842 maps to a 60 kb LD block (~110.64–110.69 Mb) within 11q23.1 consistent with imputed SNP data (Fig. 2). The region encompassing three predicted open-reading frames (ORFs): C11orf53, FLJ45803 and LOC120376. POU2AF1 lies close by (Fig. 2). To determine whether any variant in the coding regions of these ORFs explained the association of rs3802842 with CRC risk, we re-sequenced the transcribed regions and intron–exon boundaries of the

---

### Table 1. Risk of colorectal cancer associated with rs3802842 in each of the eight case–control study series

<table>
<thead>
<tr>
<th>Series</th>
<th>Cases</th>
<th>Controls</th>
<th>( P_{allele} )</th>
<th>OR_{allele}</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CORGI</td>
<td>85</td>
<td>280</td>
<td>249</td>
<td>0.37</td>
<td>80</td>
</tr>
<tr>
<td>DFCCS</td>
<td>68</td>
<td>341</td>
<td>351</td>
<td>0.31</td>
<td>39</td>
</tr>
<tr>
<td>EPICOLON</td>
<td>52</td>
<td>220</td>
<td>233</td>
<td>0.32</td>
<td>38</td>
</tr>
<tr>
<td>FCCPS</td>
<td>62</td>
<td>412</td>
<td>496</td>
<td>0.28</td>
<td>50</td>
</tr>
<tr>
<td>MCCS</td>
<td>32</td>
<td>199</td>
<td>237</td>
<td>0.28</td>
<td>55</td>
</tr>
<tr>
<td>NSCCG1</td>
<td>323</td>
<td>1239</td>
<td>1291</td>
<td>0.33</td>
<td>257</td>
</tr>
<tr>
<td>NSCCG2</td>
<td>274</td>
<td>1335</td>
<td>1386</td>
<td>0.31</td>
<td>244</td>
</tr>
<tr>
<td>VCQ</td>
<td>130</td>
<td>502</td>
<td>577</td>
<td>0.32</td>
<td>64</td>
</tr>
</tbody>
</table>

---

**Figure 1.** Forest plots of odds ratios (ORs) of colorectal cancer associated with rs3802842. (A) OR per allele; (B) heterozygous OR; (C) homozygous OR. Boxes denote OR point estimates, their areas being proportional to the inverse variance weight of the estimate. Horizontal lines represent 95% confidence intervals. Diamonds (and solid line) represent the summary OR, with 95% confidence interval given by its width. The broken vertical line is at the null value (OR = 1.0).
four genes in 92 unrelated individuals who had been genotyped for rs3802842. All of the variants found are detailed in Supplementary Material, Table S1. In summary, no common non-synonymous sequence changes in the putative protein-encoding regions of C11orf53, FLJ45803, LOC120376 or POU2AF1 were identified. We did, however, identify three SNPs in high LD ($r^2 > 0.8$) with rs3802842, rs3087967, located in the 3' - UTR of C11orf53; rs10891246,
a synonymous change located in the putative exon 1 of LOC1230376; and rs7105857, in an intron of LOC120376.

We investigated the possibility that rs3802842 might have cis-regulatory effects on neighboring genes by interrogation of publicly available data on expression in 90 and 400 lymphoblastoid cell lines. No evidence for relationship between rs3802842 and expression of FLJ45803, LOC120376 or C11orf53, POU2AF1 (defined by rs3802842 and expression of FLJ45803, LOC120376 or C11orf53, POU2AF1 (defined by P < 0.05) was observed in either data set.

Cases (n = 2852) and controls (n = 2818) in NSCCG1 had previously been genotyped for all risk variants at 8q24.21 (rs6983267), 8q23.3 (rs16892766), 10p14 (rs10795668), 11q23 (rs3802842), 15q13.3 (rs4779584) and 18q24.1 (rs12953717) as part of previously published work (5,6,8,9). Together with rs3802842, we modeled pair-wise combinations of all six SNPs. This analysis provided no evidence of interactive effects between any of the loci so far identified (P > 0.09; Supplementary Material, Table S5), suggesting that each locus has an independent role in defining the risk of developing CRC. The risk of CRC, however, increases with an increasing numbers of variant alleles for the six loci (ORper allele = 1.19; 95% CI: 1.15–1.23; Pmend = 7.4 × 10⁻²⁴) and for the 2% of the population who carry seven or more risk alleles the risk of disease is increased ∼5-fold (Fig. 3).

**DISCUSSION**

We have provided strong evidence that genetic variation defined by rs3802842 affects an individual’s risk of developing CRC. The biological basis of the association is currently unclear because rs3802842 does not reside in the coding sequence of a gene. We have excluded a coding change in all four ORFs in the region as the basis of the association.

Accepting the caveat that gene expression in lymphoblastoid cell lines may not reflect colonic tissue we found no evidence that rs3802842 exerts cis-regulatory effects on the ORFs mapping to 11q23.1. It is possible that the effect is mediated through LD with a hitherto uncharacterized gene or microRNA within the 60 kb region of LD. Alternatively, although the loss of heteroegosity at 11q23 has been reported to be frequent in CRC (10), suggesting a role for the region in tumor development, the underlying genomic sequence change defined by rs3802842 might exert cis- or trans-regulatory effects on gene expression of genes mapping outside 11q23.1.

It will probably be challenging to identify the mechanism by which rs3802842 affects CRC development, although determining the causal basis may prove highly informative, endorsing etiological hypotheses or suggesting new ones that merit testing through gene/environment-specific studies. In this respect, it is intriguing that we have demonstrated a small but significant difference in the pattern of site-specific CRC risk associated with rs3802842 as there are differences in the biology of colonic and rectal cancer both in terms of environmental risk factors and mutational spectra (11,12).

Many cancer predisposition genes influence the risk of more than one tumor type, and pleiotropic effects are a feature of 8q24 cancer-associated variants such as rs6983267, which affects the risk of CRC and prostate carcinoma. It is therefore plausible that rs3802842 will influence the other cancers. While GWAS data from Cancer Genetic Markers of Susceptibility (CGEMS) for breast and prostate cancer provides no evidence that this variant influences the risk of either tumor it does not preclude a role in the development of other common malignancies.

Irrespective of the nature of the causal variant responsible for the 11q23.1 association, a high proportion of the population are carriers of the at risk allele and hence the variant is likely to play an important role in CRC. On the basis of allele frequencies and genotypic risks, the locus is likely to be involved in ∼15% of CRC in European populations and account for ∼1% of the excess familial CRC risk. Although this risk is modest, as has been shown the locus has the potential, by acting in concert additively or multiplicatively with other similar variants to produce much larger risks in carriers of multiple risk alleles thereby potentially having direct clinical relevance in terms of defining screening requirements or entry into chemoprevention trials.

We estimate the six loci that we have identified to date through our GWAS account for ∼3% of the excess familial risk of CRC. It is acknowledged that the present data provide only crude estimates of the overall effect on susceptibility attributable to variation at these loci. The effect of the actual common causal variants responsible for these associations, once identified, will typically be larger, and many of the loci may carry additional causal variants, potentially including low-frequency variants with larger influence on CRC risk.

**MATERIALS AND METHODS**

**Study participants**

The study was based on eight independent case-control series:

CORGI: 619 CRC cases (279 males, 340 females) ascertained through the Colorectal Tumour Gene Identification (CoRGI) consortium. All had at least one first-degree relative affected by CRC. Controls (422 males, 510 females) were spouses or partners unaffected by cancer and without a family history...
of colorectal neoplasia. All cases and controls were of white UK ethnic origin.

DFCCS: 783 familial CRC cases (370 males, 413 females; mean age at diagnosis 53.4 ± 13.4 years) and 664 controls (251 males, 413 females; mean age 51.1 ± 11.3 years) ascertained at a clinically based genetic reference centre, Leiden, the Netherlands.

EPICOLON: 515 incident CRC cases (305 males, 210 females; mean age at diagnosis 70.6 ± 11.3 years) and 515 controls (290 males, 225 females; mean age 69.8 ± 11.7 years) ascertained through the EPICOLON initiative, a study of familial CRC.

FCCPS: 1001 CRC cases (509 males, 492 females; mean age at diagnosis 70.6 ± 11.8 years) and 1034 controls (randomly selected anonymous Finnish blood donors) from south-eastern Finland.

MCCS: 515 CRC cases (270 males, 245 females; mean age at diagnosis 66.2 ± 7.7 years) and 709 controls (352 males, 357 females; mean age 57.9 ± 7.0 years) from Melbourne, Australia. A random sample selected from the MCCS (Melbourne Collaborative Cohort study) cohort.

NSCCG1: 2863 CRC cases (1196 males, 1667 females; mean age at diagnosis 59.3 ± 8.7 years) ascertained through two ongoing initiatives at the Institute of Cancer Research/Royal Marsden Hospital NHS Trust (RMHNIST) from 1999 onwards—The National Study of Colorectal Cancer Genetics (NSCCG) and the Royal Marsden Hospital Trust/Institute of Cancer Research Family History and DNA Registry. A total of 2838 healthy individuals were recruited as part of ongoing National Cancer Research Network genetic epidemiological studies, NSCCG (1219), the Genetic Lung Cancer Pre-disposition Study (GELCAPS) (1999–2004; n = 911) and the Royal Marsden Hospital Trust/Institute of Cancer Research Family History and DNA Registry (1999–2004; n = 708). These controls (1136 males, 1702 females; mean age 59.8 ± 10.8 years) were the spouses/friends of cancer patients. None had a personal history of malignancy. All cases and controls were British and of European Decent.

NSCCG2: 3036 CRC cases (1629 males, 1407 females; mean age at diagnosis 59.4 ± 8.2 years) and 2944 healthy individuals (1183 males, 1753 females; mean age 55.2 ± 12.3 years) ascertained through the NSCCG post-2005.

VCQ: 202 individuals with CRC from the CORGI study; 910 patients from the VICTOR study, a randomized trial of VIOXX in patients with stages B and C CRC and 139 patients from the QUASAR2 trial comparing capecitabine against capcitabine plus bevacizumab. The controls were 250 unaffected spouses or partners from the CORGI study, 376 human random controls from ECACC and 173 blood donors. Overall, 53% of the cases and 58% of the controls were female. All cases and controls were British and were of European Decent.

Colorectal cancer was defined according to the ninth revision of the International Classification of Diseases by codes 153–154, and all cases had pathologically proved adenocarcinoma.

Collection of blood samples and clinico-pathologic information from patients and controls was undertaken with informed consent and ethical review board approval in accordance with the tenets of the Declaration of Helsinki.

Genotyping and sequencing

DNA was extracted from samples using conventional methodologies and quantified using PicoGreen (Invitrogen). CORGI samples were genotyped in the first phase of a GWAS using the Illumina 550K array. NSCCG1 samples were genotyped in the second phase of the GWAS using Illumina Infinium custom arrays (Illumina Inc., San Diego, USA). For all other series, genotyping of rs3802842 was conducted by competitive allele-specific PCR KASPar chemistry (KBiosciences Ltd, Hertfordshire, UK), Sequenom iPLEX (San Diego, USA), High Resolution Melt (HRM) Curve analysis or Taqman Applied Biosystems, Foster City, USA) according to the manufacturer’s protocols. Details of PCR primers and probes used are available on request. Genotyping quality control was tested using duplicate samples within studies and SNP assays, together with direct sequencing of subsets of samples to confirm the genotyping accuracy.

The putative coding regions and the intron–exon boundaries of the four genes C11orf53, FLJ45803, LOC120376 and POU2AF1 were re-sequenced in 92 unrelated individuals (controls from NSCCG1) genotyped for rs3802842. PCR and sequencing primers were designed by Primer3 software, primers sequences are available on request. Amplicons were sequenced by ABI chemistry (BigDye v3.1) and sequences analyzed using Mutation Surveyor software.

Microsatellite instability (MSI) in CRCs was determined as described previously (5). Samples showing novel alleles at either BAT26 or BAT25 or both markers were assigned as MSI (corresponding to a high level of instability, MSI-H) (13).

Statistical analysis

Statistical analyses were undertaken in Stata Version 8 (Station College, TX, USA) or R software. Deviation of the genotype frequencies in the controls from those expected under HWE was assessed by $\chi^2$ test. The risk of CRC associated with rs3802842 was estimated by allelic, heterozygous and homozygous OR using logistic regression. Meta-analyses were based on the Mantel–Haenszel method; Cochran’s Q statistic to test for heterogeneity and the $I^2$ statistic (14) to quantify the proportion of the total variation due to heterogeneity were calculated. Patterns of risk associated with rs3802842 were investigated by logistic regression, coding the SNP genotypes according to additive, dominant and recessive models and comparing Akaike information criterion and Akaike weights for each mode of inheritance. Associations by site (colon/rectum), MSI status, family history status (at least one first-degree relative with CRC), gender and age at diagnosis were examined by logistic regression in case-only analyses. The OR and trend test for increasing numbers of deleterious alleles was performed by counting two for a homozygote and one for a heterozygote.

The population attributable fraction was estimated by $(x - 1)/x$, where $x$ is defined by $(1 - p)^2 + 2p(1 - p) OR_1 + p^2 OR_2$, $p$ is the population allele frequency, and OR1 and OR2 are the ORs.
associated with hetero- and homozygosity, respectively. The sibling relative risk attributable to a given SNP was calculated using the formula (15):

$$
\lambda^* = \frac{p(pr_2 + qr_1)^2 + q(pr_1 + q)^2}{[p^2r_2 + 2pqr_1 + q^2]^2}
$$

where \( p \) is the population frequency of the minor allele, \( q = 1 - p \), and \( r_1 \) and \( r_2 \) are the relative risks (estimated as OR) for heterozygotes and rare homozygotes. Assuming a multiplicative interaction, the proportion of the familial risk attributable to an SNP was calculated as \( \log(\lambda^*)/\log(\lambda_0) \), where \( \lambda_0 \) is the overall familial relative risk estimated from epidemiological studies, assumed to be 2.2 (16). A naive estimation of the contribution of all of the loci identified to the excess familial risk of CRC under an additive model was calculated using the formula:

$$
\text{OR per allele} - 1
$$

\( 2(\lambda_0 - 1) \).

Linkage disequilibrium statistics were calculated using Haploview software (v4.0). We used genotype data from the GWAS study based on CORGI and NSCCG1 (8) to further investigate the region of association. Prediction of the un-typed SNP in the case–control data sets of phases II and I was carried out using MACH1.0 (17) on HapMap (HapMap Data Rel 21a/phase II Jan07 on NCBI B35 assembly, dbSNPb125) phase II data. In total, 236 HapMap SNPs were successfully imputed in the interval between 110.555 and 110.75 Mb at 11q23 using available SNP genotype data from NSCCG1 (eight SNPs) and CORGI (49 SNPs). Imputed data integrity was verified where possible, by cross-checking the concordance of imputed genotypes with that of available Illumina SNP genotype data.

Relationship between SNP genotypes and expression levels

To examine for a relationship between SNP genotype and gene expression in lymphocytes, we made use of publicly available expression data. We calculated expression in lymphocytes, we made use of publicly available Illumina SNP genotype data. To examine for a relationship between SNP genotype and gene expression in lymphocytes, we made use of publicly available Illumina SNP genotype data.

**WEB ADDRESSES**

- HapMap: http://www.hapmap.org/
- KBioscience: http://kbioscience.co.uk/
- Haploview: http://www.broad.mit.edu/personal/jcbarrett/haploview/
- QUASAR: http://www.octo-oxford.org.uk/alltrials/trials/q2. html
- Cancer Genetic Markers of Susceptibility (CGEMS): http:// egems.cancer.gov/
- Primer 3: Software: http://frodo.wi.mit.edu/
- MACH1: http://www.sph.umich.edu/csg/abecasis/MACH/
- Illumina: http://www.illumina.com/

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG Online.

**FUNDING**

Cancer Research UK provided principal funding for this study. Institute of Cancer Research: Additional funding was provided by the European Union (CPRB LSHC-CT-2004-503465), the Bobby Moore Fund, CORE and the Thomas Falknor Fund. I.C. was in receipt of a clinical training fellowship from St. George’s Hospital Medical School. London Institute: Additional funding was provided by CORE and the Bobby Moore Fund. Barcelona: We are sincerely grateful to all patients participating in this study that were recruited in 25 Spanish hospitals as part of the EPICOLON project. This work was supported by grants from the Fondo de Investigación Sanitaria (03/0070, 05/0071 and 05/ 2031), from the Ministerio de Educación y Ciencia (SAF 04-07190 and 07-64873), the Asociación Española contra el Cáncer, from Merck, Co, from the Xunta de Galicia (PGI- DIT07PXB190101209PR), from Fundacion Olga Torres (S.C.-B.), and from Fundación de Investigación Médica Mutua Madrileña (C.R.-P.). CIBERERH and CIBERER are funded by the Instituto de Salud Carlos III. S.C.-B. is supported by a contract from the Fondo de Investigación Sanitaria (CP 03-0070, Ministerio de Sanidad). Extremadura: Work was supported by grants FIS 051056 and RD07/0064/ 0016 from Instituto de Salud Carlos III, Madrid, Spain. Finland: This work was supported by grants from Academy of Finland (Finnish Centre of Excellence Program 2006–2011), the Finnish Cancer Society, the Sigrid Juselius Foundation and the European Commission 9LSHG- CT-2004-512142), Heidelberg: Supported by Deutsche Krebshilfe and the Swedish Cancer Society. Kiel: This study was supported by the German Ministry of Education and Research through the National Genome Research Network through the POPGEN biobank project (01GS0426, 01GR0468) and the Medical Faculty, Kiel. The SHIP recruitment project is funded by the Federal Ministry of Education and Research (ZZ9603), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Leiden: DFCCS was supported by Dutch Cancer Society grant UL2005-3247 and approved by the local Medical Ethical Committee (protocol P01.019); samples were handled according to Code Proper Secondary Use of Human Tissue by the Dutch Federation of Medical Sciences (www.federa.org). Madrid: Work was supported by the Fondo Investigacion Sanitaria (PI070316 and RD06/0020/0021). Melbourne: The Melbourne Collaborative Cohort Study is supported by National Health and Medical Research Council (NHMRC) grants 209057, 251533 and 396414 and receives core funding and infrastructure support from the The Cancer Council Victoria. J.L.H. is a NHMRC Australia Fellow and M.C.S. is a NHMRC Senior Research Fellow. We would like to acknowledge Mr Fabrice Odefrey for performing the genotyping. Prague: Supported by the grant GACR 310/07/1430.
ACKNOWLEDGEMENTS

We would like to thank all individuals that participated in this study. We are grateful to colleagues at UK Clinical Genetics Centres and the UK National Cancer Research Network.

Conflict of Interest statement. None declared.

REFERENCES