The TERT variant rs2736100 is associated with colorectal cancer risk

Citation for published version:

Digital Object Identifier (DOI):
10.1038/bjc.2012.329

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
British Journal of Cancer

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
The TERT variant rs2736100 is associated with colorectal cancer risk

B Kinnersley¹, G Migliorini¹, P Broderick¹, N Whiffin¹, SE Dobbins¹, G Casey², J Hopper³, The Colon Cancer Family Registry², O Sieber⁴, L Lipton⁴, DJ Kerr⁵, MG Dunlop⁶, IPM Tomlinson⁷,⁸ and RS Houlston*,¹ The CORGI Consortium

¹Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, Surrey SM2 SNG, UK; ²Department of Preventive Medicine, University of Southern California, Los Angeles, CA, USA; ³Centre for Molecular, Environmental Genetic and Analytic Epidemiology, Melbourne School of Population Health, The University of Melbourne, Parkville, Victoria, Australia; ⁴Ludwig Institute for Cancer Research, Royal Melbourne Hospital, Western Hospital, PO Box 2008, Parkville 3050, Victoria, Australia; ⁵Department of Oncology, Oxford University, Radcliffe Infirmary, Old Road Campus Research Building, Headington, Oxford OX3 7DQ, UK; ⁶Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, University of Edinburgh and Medical Research Council Human Genetics Unit, Edinburgh EH4 2XU, UK; ⁷Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK; ⁸Oxford NIHR Comprehensive Biomedical Research Centre, Oxford, UK

BACKGROUND: Polymorphic variation at the 5p15.33 (TERT–CLPTM1L) locus is associated with the risk of many cancers but a relationship with colorectal cancer (CRC) risk has yet to be defined.

METHODS: We used data from six genome-wide association studies (GWAS) of CRC, linkage disequilibrium mapping and imputation, to examine the relationship between 73 single-nucleotide polymorphisms at 5p15.33 and CRC risk in detail.

RESULTS: rs2736100, which localises to intron 2 of TERT, provided the strongest evidence of an association with CRC ($P = 2.28 \times 10^{-6}$). The association was also shown in an independent series of 10,047 CRC cases and 6,918 controls ($P = 0.02$). A meta-analysis of all seven studies (totalling 16,039 cases, 16,430 controls) provided increased evidence of association ($P = 2.49 \times 10^{-5}$; per allele odds ratio = 1.07). The association of rs2736100 on CRC risk was shown to be independent of 15 low-penetrance variants previously identified.

CONCLUSION: The rs2736100 association demonstrates an influence of variation at 5p15.33 on CRC risk and further evidence that the 5p15.33 (TERT–CLPTM1L) locus has pleiotropic effects (reflecting generic or lineage-specific effects) on cancer risk.

Keywords: colorectal cancer; TERT–CLPTM1L; rs2736100

Many colorectal cancers (CRCs) develop in genetically susceptible individuals most of whom are not carriers of germ-line mismatch repair or APC mutations (Lichtenstein et al, 2000; Aaltonen et al, 2007). It is likely that much of the unexplained heritable risk is attributable to a combination of multiple low- to moderate-penetrance genetic variants, which are associated with relatively small effects on risk in the individual but contribute substantially to the overall risk in the population (Fletcher and Houlston, 2010).

Genome-wide association studies (GWAS), using large sets of cases and controls, have proven to be an effective strategy to identify common single-nucleotide polymorphisms (SNPs) associated with cancer risk without prior knowledge of position or function (Fletcher and Houlston, 2010). This approach has successfully identified novel loci for most of the common cancers including CRC (Fletcher and Houlston, 2010). The majority of SNP associations identified to date have been tumour specific, which is consistent with the epidemiological studies of familial cancer risks (Fletcher and Houlston, 2010). Evidence for pleiotropic effects, reflecting generic or lineage-specific effects, is provided by variation at 5p15.33 (TERT–CLPTM1L) that is associated with the risk of many tumours including breast, testicular, bladder and lung cancers (McKay et al, 2008; Wang et al, 2008; Rafnar et al, 2009; Shete et al, 2009; Van Dyke et al, 2009; Hsiung et al, 2010; Turnbull et al, 2010; Beesley et al, 2011; Gago-Dominguez et al, 2011; Haiman et al, 2011; Kratz et al, 2011; Law et al, 2011; Peters et al, 2012).

Although many cancer associations at 5p15.33 have been identified with rs2736100, which localises to intron 2 of TERT (McKay et al, 2008; Wang et al, 2008; Shete et al, 2009; Hsiung et al, 2010; Turnbull et al, 2010; Gago-Dominguez et al, 2011), the existence of other SNP associations within the region supports the existence of multiple risk loci with different tumour specificities. Recently, an association between the SNP rs2853668, which maps centromeric to TERT but is only weakly correlated with rs2736100, has been reported for CRC risk (Peters et al, 2012). Given the ubiquitous necessity for tumours to avoid replicative senescence through shortened telomere repeat length, a process that is often mediated through expression of telomerase (Hanahan and Weinberg, 2000), a variant at TERT associated with CRC would be biologically plausible.

Using data from six GWAS of CRC, linkage disequilibrium (LD) mapping and imputation, we have studied the relationship...
MATERIALS AND METHODS

Ethics

Collection of blood samples and clinico-pathological 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.

GWAS datasets

London 1 (LP1) comprised 940 cases with colorectal neoplasia (47\% male) ascertained through the Colorectal Tumour Gene Identification (CoRGI) consortium. All had at least one first-degree relative affected by CRC and one or more of the following phenotypes: CRC at age 75 years or less; any colorectal adenoma relative affected by CRC and one or more of the following: familial adenomatous polyposis (FAP), hereditary non-polyposis colorectal cancer (HNPCC), Lynch syndrome, or di-allelic MUTYH mutation carriers were excluded. All cases and controls had self-reported European ancestry. Both cases and controls were genotyped using Illumina HumanHap550 BeadChip Arrays (Teo et al., 2007).

Scotland1 (SP1) included 1012 CRC cases (51\% male; mean age at diagnosis 49.6 years, s.d.± 6.1) and 1012 cancer-free population controls (51\% male; mean age 51.0 years; s.d.± 5.9). Cases were selected for early age at onset (age ≤ 55 years). Known dominant polyposis syndromes, HNPCC/Lynch syndrome or bi-allelic MUTYH mutation carriers were excluded. Control subjects were sampled from the Scottish population NHS registers, matched by age (± 5 years), gender and area of residence within Scotland. Both cases and controls were genotyped using Illumina HumanHap300 and HumanHap240S arrays (Gunderson et al., 2006; Abraham et al., 2008).

VQ58 (VQ) comprised 1800 CRC cases (1099 males, mean age of diagnosis 62.5 years; s.d.± 10.9) from the VICTOR and QUASAR2 (http://www.octo-oxford.org.uk/alltrials/infolinkv2/q2.html) trials. Genotyping of cases was conducted using Illumina Hap300 and Hap370 arrays (Gunderson et al., 2006; Howarth et al., 2009). The 2690 controls, typed on the Illumina Human 1.2M-Duo Custom_v1 Array BeadChips, were from the UK population-based 1958 birth cohort, for which genotype data are publicly available from the Wellcome Trust Case-Control Consortium 2 (Power and Elliott, 2006; The Wellcome Trust Case-Control Consortium, 2007).

The Colon Cancer Family Registry (CFR1) data set comprised 2690 familial CRC cases and 1055 controls CFR (Colon-CFR) (http://cgi.epi.grants.cancer.gov/CFR/about_colon.html). The cases were recently diagnosed CRC cases reported to population complete cancer registries in the United States (Puget Sound, WA, USA) recruited by the Seattle Familial Colorectal Cancer Registry; in Canada (ON) recruited by the Ontario Familial Cancer Registry; and in Australia (Melbourne, VIC) recruited by the Australasian Colorectal Cancer Family Study. Controls were population-based and for this analysis were restricted to those without a family history of CRC (Newcomb et al., 2007). Cases and controls were genotyped using Illumina HumanHap550 and 1M and 1Mduo BeadChip Arrays.

CFR2 comprised an additional 796 cases ascertained through the CFR (http://cgi.epi.grants.cancer.gov/CFR/about_colon.html). Cases were genotyped using 1M Omni-Express BeadChip Arrays.

Illumina HumanHap550 BeadChip data on 2304 individuals from the Cancer Genetic Markers of Susceptibility (CGEMS) studies served as control genotypes (Hunter et al., 2007; Yeager et al., 2007).

The Australian (AUS) study (Tie et al., 2010) comprised 591 patients treated for CRC at the Royal Melbourne, Western and St Francis Xavier Cabrini Hospitals in Melbourne from 1999 to 2009. The 2353 controls were derived from Queensland or Melbourne: for the former, the controls came from the Brisbane Twin Nvus Study (Duffy et al., 2010); for the latter, individuals were participants in the Genes in Myopia study (Baird et al., 2010). There was no overlap between the CFR and Australian datasets. Both cases and controls were genotyped using Illumina Human-Hap550 BeadChip Arrays.

Each of these six GWAS datasets was subjected to extensive quality control procedures. Specifically, the exclusion of samples and SNPs with call rates <95\%, non-European (CEU) ancestry, relatedness (duplicates or related within or between each case-control series) and sex discrepancy. Furthermore, there was no evidence of systematic inflation of the test statistic in any study, as assessed using the genomic overdispersion factor, λ_g, which ranged from 1.00 to 1.04.

Replication series

In total, 10,488 CRC cases, aged <80 years at diagnosis, were ascertained between March 2003 and October 2011 through the National Study of Colorectal Cancer Genetics (NSCCG) (Penegar et al., 2007) (n = 9268); the Study of the Genetic Epidemiology of Colorectal Cancer (n = 581) and the Royal Marsden Hospital National Health Services Trust (RMHNHST) family history DNA database (n = 639). Controls (n = 7137) were the spouses of cancer cases and were ascertained through the NSCCG (n = 3047); the Genetic Lung Cancer Predisposition Study (n = 1637); the Colorectal Adenoma Gene-Environment Interactions Study (n = 711); the Study of the Genetic Epidemiology of Colorectal Cancer (n = 344); and the RMHNHST family history DNA database (n = 1398). None of the controls had a personal history of malignancy at ascertainment. All subjects were British residents with self-reported European ethnicity and there were no obvious demographic differences between cases and controls.

Statistical and bioinformatic analysis

A P-value (two-sided) ≤ 0.05 was considered significant. We applied a Bonferroni correction to adjust for multiple testing. Statistical analyses were undertaken using SNPtest/META (The Wellcome Trust Case-Control Consortium, 2007), and STATA v.10 (StataCorp LP, College Station, TX, USA) software. The association between each SNP and risk of CRC was assessed by the Cochran–Armitage trend test. Odds ratios and associated 95\% confidence intervals (CIs) were calculated by unconditional logistic regression. Patterns of risk for associated SNPs were investigated by logistic regression, coding the SNP genotypes according to additive, dominant and recessive models. We then compared models by calculating the Akaike information criterion and Akaikie weights for each mode of inheritance. Interaction between SNP and genotypes was evaluated by likelihood ratio tests comparing an additive model to a model with an interaction term. Prediction of the non-genotyped SNPs within the 119.3-kb region of 5p15.33 (TERT–CLPTM1L) (1 243 475–1 362 793 bps, NCBI build b37) was carried out using IMPUTE2 based on the June 2011 release of 1000 Genomes Project data (Howie et al., 2009, 2011; The 1000 Genomes Project Consortium, 2010). Association testing of genotyped and imputed data were analysed using SNPTET v2 to account for uncertainties in SNP prediction. Imputed genotypes were only called if they had a probability >0.90. Association meta-analyses only included markers with proper info scores >0.9, imputed call rates/SNP >0.9 and Hardy–Weinberg >0.01. To condition by SNP,
MSI status was evaluated by case-only analysis. 

Analysis of LD was performed using the Broad Institute SNP Annotation and Proxy (SNAP) Search utilising 1000 Genomes Project data. Transcription factor-binding prediction was performed with TFSearch. Cross-species evolutionary conservation was assessed with the deCode ECR browser. The UCSC genome browser was used to examine H3K4Me1, H3K4Me3 and DNase-I hypersensitivity in publicly available cell line data.

Association between rs2736100 and tumour site (colon-ICD International Classification of Diseases 9th revision (ICD9)-154), Dukes stage (A–D), grade (poorly; moderate/well differentiated), sex, age at diagnosis (rectal cancer: ICD9-154), and family history of CRC in a first-degree relative and MSI status was evaluated by case-only analysis.

**Table 1**

<table>
<thead>
<tr>
<th>Study</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUS</td>
<td>376 (379)</td>
<td>1702 (1703)</td>
</tr>
<tr>
<td>CFR1</td>
<td>1175 (1175)</td>
<td>999 (999)</td>
</tr>
<tr>
<td>CFR2</td>
<td>796 (796)</td>
<td>2232 (2234)</td>
</tr>
<tr>
<td>LP1</td>
<td>890 (890)</td>
<td>900 (900)</td>
</tr>
<tr>
<td>SPI</td>
<td>962 (973)</td>
<td>998 (998)</td>
</tr>
<tr>
<td>VQ</td>
<td>1793 (1794)</td>
<td>2681 (2686)</td>
</tr>
<tr>
<td>Replication</td>
<td>10 047 (10 488)</td>
<td>6918 (7137)</td>
</tr>
</tbody>
</table>

**Table 2**

<table>
<thead>
<tr>
<th>Descriptive data</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG</td>
<td>590</td>
<td>2623</td>
</tr>
<tr>
<td>GT</td>
<td>247</td>
<td>4173</td>
</tr>
<tr>
<td>TT</td>
<td>214</td>
<td>1171</td>
</tr>
<tr>
<td>RAF</td>
<td>690</td>
<td>1274</td>
</tr>
</tbody>
</table>

Abbr: AUS = Australian; CFR1 = Colon Cancer Family Registry; LP1 = London 1; RAF = risk allele frequency; VQ = VQ58.

**RESULTS**

**Descriptive data**

Table 1 provides summary information on the clinico-pathological characteristics and demographic information on each of the six GWAS datasets and the replication case–control series.

**LD structure of the 5p15.33 region**

The six GWAS studies of CRC provided genotype data for 12–45 SNPs (depending on study) mapping to the 5p15.33 region in a total of 6007 cases and 9520 controls (Figure 1). To further investigate the relationship between genetic variation at this region and CRC risk, using 1000 genomes data, we imputed the genotypes of 22 SNPs not directly genotyped in one or more studies and 45 SNPs not genotyped in any study. Including 6 SNPs directly genotyped in all studies, a total of 73 SNPs were imputed (Figure 1). In a combined analysis of these data, the strongest association was shown by the directly typed SNP, rs2736100 ($P = 2.28 \times 10^{-4}$; Figure 1). The per allele OR of CRC associated

the SNP was added as a covariate. Meta-analysis was performed using a fixed-effects model, estimating Cochran’s Q statistic to test for heterogeneity and the $I^2$ statistic to quantify the proportion of the total variation between studies.

**Figure 1**

Regional plot of SNP association with CRC across the 5p15.33 locus. Association results of both genotyped (triangles) and imputed (circles) SNPs in the GWAS samples and recombinations. $-\log_{10}$ $P$-values ($y$-axis) of the SNPs are shown according to their chromosomal positions ($x$-axis). rs2736100 is represented by a large triangle. The grayscale intensity of each symbol reflects the extent of LD with rs2736100: white ($r^2 = 0$) through to dark grey ($r^2 = 1.0$). Genetic recombination rates (cM/Mb), estimated using HapMap CEU samples, are shown with a light grey line. Physical positions are based on NCBI build 37 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. For SNPs where $r^2$ data were unavailable, these values were set to 0. The colour reproduction of this figure is available at British Journal of Cancer online.

© 2012 Cancer Research UK

British Journal of Cancer (2012) 107(6), 1001 – 1008

Genetics and Genomics
with rs2736100-T genotype was 1.10 (95% CI:1.05–1.15). The association between rs2853668, a SNP previously reported to be associated with CRC(Peters et al, 2012), and CRC risk was non-significant (P = 0.06). To explore the possibility of secondary associations with CRC, we conducted pairwise conditional analyses of the eight SNPs showing the best evidence for an association with risk. rs2736100 genotype was shown to be sufficient to capture the 5p15.33 association with CRC risk (Table 2).

Replication of the rs2736100 association

To provide further independent replication of the rs2736100 association with CRC risk, we genotyped an additional 10 488 cases and 7137 controls. Genotypes were obtained for 96% of cases (n = 10 047) and 97% of controls (n = 6918) (Table 1). There was no evidence of population stratification in controls as the genotype distribution satisfied Hardy–Weinberg equilibrium (P = 0.76). As with the GWAS data there was a significant over-representation of the rs2736100-T genotype in CRC cases (P = 0.019; Supplementary Table 1). The impact of rs2736100 genotype on CRC was thus comparable to that shown in the combined analysis of the six GWAS datasets. Although the pattern of risk for CRC was most parsimonious with a dominant model, a multiplicative model was equally favoured (P = 0.27). To enhance our ability to demonstrate a relationship between 5p15.33 variation and CRC risk, we conducted a combined analysis of all datasets (Figure 2). In this meta-analysis, the per allele OR was 1.07 (95% CI: 1.04–1.11; P = 2.49 × 10⁻³; Bonferroni-adjusted P-value was 1.82 × 10⁻³) and there was no evidence for between-study heterogeneity (Phet = 0.7, I² = 0%).

Relationship between rs2736100 genotype and phenotype

To explore the relationship between rs2736100 genotype and CRC phenotype, we performed a case-only analysis using the replication series. This analysis provided no statistically significant evidence that the CRC association was modified by age, sex, or family history of CRC (Supplementary Table 1). For 3200 of the cases with known MSI status, rs2736100 genotypes were successfully generated on 2981 (93%), allowing us to calculate CRC risks stratified by MSI status. This analysis did not provide evidence for a relationship between SNP genotype and MSI status (Supplementary Table 1).

Interaction between rs2736100 and other common CRC risk variants

Using logistic regression analysis, we tested for an interaction between rs2736100 and each of 15 SNPs shown previously to be associated with CRC, namely, rs6691170, rs10936599, rs16892766, rs693267, rs10795668, rs3802842, rs11169552, rs4444235, rs4779584, rs9929215, rs4925386 and rs11632715. No evidence of statistical interaction between any of the 15 SNPs and rs2736100 was shown (Supplementary Table 2).

Table 2 Conditional analysis on the SNPs most significantly associated with CRC in TERT–CLPTM1L after imputation

<table>
<thead>
<tr>
<th>SNP</th>
<th>Main effect P-value</th>
<th>r² With rs2736100</th>
<th>Conditioned on rs2736100 P-value</th>
<th>r² With rs2853668</th>
<th>Conditioned on rs2853668 P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2736100</td>
<td>0.0002</td>
<td>1 (ref)</td>
<td>1 (ref)</td>
<td>0.15</td>
<td>0.001</td>
</tr>
<tr>
<td>rs2736122</td>
<td>0.001</td>
<td>0.20</td>
<td>0.01</td>
<td>0.00</td>
<td>0.003</td>
</tr>
<tr>
<td>rs4583925</td>
<td>0.002</td>
<td>0.03</td>
<td>0.01</td>
<td>0.00</td>
<td>0.004</td>
</tr>
<tr>
<td>rs4507531</td>
<td>0.003</td>
<td>0.03</td>
<td>0.02</td>
<td>0.00</td>
<td>0.004</td>
</tr>
<tr>
<td>rs79323805</td>
<td>0.003</td>
<td>0.03</td>
<td>0.02</td>
<td>0.00</td>
<td>0.005</td>
</tr>
<tr>
<td>rs4075202</td>
<td>0.005</td>
<td>0.03</td>
<td>0.03</td>
<td>0.00</td>
<td>0.008</td>
</tr>
<tr>
<td>rs35595862</td>
<td>0.03</td>
<td>0.00</td>
<td>0.01</td>
<td>0.05</td>
<td>0.007</td>
</tr>
<tr>
<td>rs2853668</td>
<td>0.06</td>
<td>0.15</td>
<td>0.50</td>
<td>1 (ref)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CRC = colorectal cancer; ref = reference; SNP = single-nucleotide polymorphism. The eight most significant SNPs for association using the Mantel–Haenszel method for combining study data are shown, with P-values for an initial unconditioned additive test, and then after conditioning on rs2736100 and rs2853668, respectively.

Figure 2 Forest plot of allelic odds ratio associated with rs2736100 genotype and CRC in the six genome-wide association studies and in the replication series. Horizontal lines represent 95% CIs. Each box represents the allelic OR point estimate, with the area being proportional to the weight of the study. The diamond (and broken line) denotes the overall summary estimate, with CIs given by its width. The unbroken vertical line is at the null value (OR = 1.0).
rs2736100 is associated with colorectal cancer risk

B Kinnersley et al

Figure 3  Evolutionary conservation and chromatin status of a 14.8-kb region of 5p15.33 encompassing rs2736100. Regions of DNase-I hypersensitivity are shown by greyscale bars. Also shown are regions prone to H3K4Me1 or H3K4Me3 modifications, often found near regulatory elements and promoters respectively. Evolutionarily conserved regions (ECRs) between humans and chimpanzee (Pan troglodytes), mouse (Mus musculus) and fugu fish (Takifugu rubripes) are denoted by light grey horizontal lines relative to TERT (darkest grey), coding exons; lightest grey, untranslated regions; darker grey, intergenic regions; light grey, intronic regions; medium grey, transposons and simple repeats of the gene). The colour reproduction of this figure is available at British Journal of Cancer online.

DISCUSSION

Here we have demonstrated a statistically significant association between rs2736100 genotype and risk of CRC. We have also been able to show that variation at 5p15.33 influences CRC risk independently of other previously identified common risk variants. This is consistent with a model in which rs2736100 is acting additively with other common risk variants in mediating CRC susceptibility. As the risk allele of rs2736100 is common, the variant is likely to underscore ~7% of all CRC in European populations.

Intriguingly rs2736100-T is associated with elevated risk of testicular cancer (Turnbull et al, 2010), but reduced risk of glioma (Shete et al, 2009), lung adenocarcinoma (McKay et al, 2008; Wang et al, 2008) and bladder cancer (Gago-Dominguez et al, 2011). These differential effects of genotype are likely to be reflective of tumour- and lineage-specific effect.

Although rs2736100 localises to intron 2 of TERT, this does not exclude the possibility of long-range effects as the functional basis for the 5p15.33 cancer association. However, although the 5p15.33 locus includes the TERT and CLPTML genes, these essentially map to two distinct regions of LD, making it likely rs2736100 impacts either directly or indirectly on TERT.

A recent study reported an association between the TERT SNP rs2853668 and CRC risk (Peters et al, 2012). In our study, rs2736100 provided superior evidence for an association with CRC than rs2853668. As rs2736100 and rs2853668 are correlated, albeit weakly ($r^2 = 0.15$, $D' = 0.69$), it is likely that the association reported by Peters et al, (2012) reflects the impact of rs2736100 genotype, or a hitherto unidentified correlated variant on CRC risk.

Functional variants in TERT have been shown to affect telomerase expression through modulating promoter activity (Beesley et al, 2011), and such a mechanism offers a possible explanation for how a putative functional variant at TERT affects CRC risk (Aisner, 2002). Although sequence conservation in non-coding regions has been shown to be a good predictor of cis-regulatory sequences (Gomez-Skarmeta et al, 2006), there is little evidence for high conservation directly at rs2736100 (Figure 3). ENCODE project data does not show evidence for DNAse hypersensitivity sites (indicating open chromatin), or histone H3KMe1/H3KMe3 methylation (often near regulatory elements) at rs2736100. Although these data do not support rs2736100 being directly functional, in an analysis of putative binding sites at rs2736100 using TFSearch, SRY and Hhf-2 sites are only predicted for rs2736100-T and not rs2736100-G. SRY (sex determining region Y) is a male-expressed gene involved in sex determination (Wallis et al, 2008). Although speculative, rs2736100-T-mediated SRY recruitment to TERT might lead to increased expression of telomerase in germ cells, thereby providing an explanation for the increased risk of testicular cancer associated with rs2736100-T (Turnbull et al, 2010). It is unknown whether Hhf-2 (also known as FOXD3) regulates telomerase expression, however, this forkhead transcription factor has a role in early cell development, thus suggesting another biological basis for the 5p15.33 association (Guo et al, 2002). Such speculations are predicated on the assumption that rs2736100 underscores the 5p15.33 association. Although our imputation provided no evidence for a stronger signal at 5p15.33 than that afforded by rs2736100, it is possible the association is mediated through one or more rare disease-causing variants, which are not adequately catalogued by the 1000 Genomes project data. High-depth coverage sequencing of a large series of CRC cases for 5p15.33 variation would allow this possibility to be explored.

In conclusion, our data demonstrate that polymorphic variation at 5p15.33 is a determinant of CRC risk. It has recently been shown that polymorphisms in TERC (telomerase RNA component) are associated with CRC risk and increased telomere length (Codd et al, 2010; Houlston et al, 2010; Jones et al, 2012); collectively these data extend the role of genetic variation in telomere elongation mechanisms in defining cancer risk per se.

ACKNOWLEDGEMENTS

Cancer Research UK provided principal funding for this study individually to RSH. (C1298/A8362 – Bobby Moore Fund for Cancer Research UK), IPMT and MGD. At the Institute of Cancer Research additional funding was provided 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 National Institute for Health Research at the Royal Marsden Hospital NHS Trust. BK and NW are in receipt of PhD Studentships from the Institute of Cancer Research, with BK receiving sponsorship from the Sir John Fisher Foundation. OS was supported by the Hilton Ludwig Cancer Metastasis Initiative. In Oxford, additional funding was provided by the Oxford Comprehensive Biomedical Research Centre. Core infrastructure support to the Wellcome Trust Centre for Human Genetics, Oxford, was provided by grant (090532/Z/09/Z). In Edinburgh funding was provided by a Cancer Research UK Programme Grant...
REFERENCES


IMPUTE: http://mathgen.stats.ox.ac.uk/impute/

1000genomes: http://www.1000genomes.org/

deCode ECR Browser: http://ecrbrowser.dcode.org/

UCSC genome browser: http://genome.ucsc.edu/

Web addresses

1000genomes: http://www.1000genomes.org/

IMPUTE: http://mathgen.stats.ox.ac.uk/impute/

SNPTEST: https://mathgen.stats.ox.ac.uk/genetics_software/snptest/

META: http://www.mathgen.stats.ox.ac.uk/~jdl/beta.html

SNAP http://www.broadinstitute.org/mpg/snap/

R project: http://www.r-project.org/


TFSearch: http://www.cbc.jp/research/db/TFSEARCH.html

decode ECR Browser: http://ecrbrowser.dcode.org/

UCSC genome browser: http://genome.ucsc.edu/

Supplementary Information accompanies the paper on British Journal of Cancer website (http://www.nature.com/bjc)


Consortium (2011) Common variant at the TERT-CLPTM1L locus is associated with longer telomeres. *Gut* 60: 874–878


Mai PL, Pfeiffer R, Greene MH (2011) Variants in or near KITLG, BAK1, DMRT1, and TERT-CLPTM1L predispose to familial testicular germ cell tumour. *J Med Genet* 48: 473–476


This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.