Inherited predisposition to colorectal cancer: towards a more complete picture

Emma Short, Laura E Thomas, Joanna Hurley, Sian Jose, Julian R Sampson

ABSTRACT

Colorectal carcinoma (CRC) is the third most common cancer worldwide. Hereditary factors are important in 15%–35% of affected patients. This review provides an update on the genetic basis of inherited predisposition to CRC. Currently known genetic factors include a group of highly penetrant mutant genes associated with rare mendelian cancer syndromes and a group of common low-penetration alleles that have been identified through genetic association studies. Additional mechanisms, which may underlie a predisposition to CRC, will be outlined, for example, variants in intermediate penetrance alleles. Recent findings, including mutations in POLE, POLD1 and NTHL1, will be highlighted, and we identify gaps in present knowledge and consider how these may be addressed through current and emerging genomic approaches. It is expected that identification of the missing heritable component of CRC will be resolved through evermore comprehensive cataloguing and phenotypic annotation of CRC-associated variants identified through sequencing approaches. This will have important clinical implications, particularly in areas such as risk stratification, public health and CRC prevention.

BACKGROUND

Colorectal carcinoma (CRC) is the third most common cancer worldwide, with approximately 1 360 600 new cases diagnosed in 2012.1 The majority of CRCs occur sporadically, but in 15%–35% of patients, hereditary factors are important (reviewed in refs 2 3). Here, we provide an update on the genetic basis of inherited predisposition to CRC. Currently known genetic factors include a group of genes associated with rare mendelian cancer syndromes, and a group of common low-penetration alleles that have been identified through association studies in large cohorts of patients with CRC and controls. We highlight recent findings and identify gaps in present knowledge, and consider how these may be addressed through genomic approaches.

MENDELIAN COLORECTAL CANCER SYNDROMES

Up to 5% of CRC occurs in the context of highly penetrant dominantly inherited syndromes (reviewed in ref. 2), and a much smaller proportion occurs in recessive syndromes (see online supplementary table S1). These syndromic disorders include the adenomatous polyposis—familial adenomatous polyposis (FAP), MUTYH-associated polyposis (MAP), NTHL1-associated polyposis, polymerase proofreading-associated polyposis and constitutional mismatch repair (MMR) deficiency syndrome; the hamartomatous polyposis syndromes—juvenile polyposis syndrome, Peutz–Jeghers syndrome and the PTEN hamartoma syndrome; hereditary mixed polyposis syndrome, in which patients have multiple polyps of different morphologies and Lynch syndrome (LS) when colorectal cancer usually develops in the absence of obvious polyposis (figures 1 and 2).

In these disorders, the risk of CRC in the absence of preventive measures is very high (10%–100%). The associated genes have been identified, and testing for inherited mutations has become routine in patients and families where these disorders are suspected, including predictive testing in asymptomatic family members to facilitate planning of surveillance and preventive measures. Guidelines are well established for the management of patients with many of these diseases, for example, the Mallorca Group guidelines for the management of LS3 and FAP8 and the American College of Gastroenterology guidelines for the management of LS.7 However, there is still a paucity of large prospective studies to inform robust estimates of cancer risk and assess the efficacy of surveillance and interventions.

The recent identification of colorectal adenoma and carcinoma predisposing mutations in DNA polymerase genes and in the base excision repair gene NTHL1 serve as a reminder that there are still likely to be gaps in our knowledge of rare high-penetrance genes, which predispose to colorectal cancer. Palles et al8 identified heterozygous germ-line variants in POLE and POLD1 in individuals with a family history of multiple adenomas and CRC, but no detectable mutations in known polyposis-associated genes, including APC or MUTYH. The study employed whole-genome sequencing, linkage and association approaches. Two high-penetrance variants, POLE p.Leu424Val and POLD1 p.Ser478Asn, were localised to the proofreading (exonuclease) domains of these polymerases. POLE p.Leu424Val was subsequently also identified in 12 unrelated cases in a cohort of 3805 European individuals with colorectal cancer enriched for a family history of CRC, early onset disease and multiple adenomas. Ten individuals from two families were found to have the POLD1 p.Ser478Asn variant with a further individual identified in the validation phase.8 Tumours in these patients were microsatellite stable, but acquired base substitution mutations. More recently, the two germline mutations were studied in a further cohort of patients with unexplained familial and early onset CRC and/or polyposis. Vallee et al9 investigated 858 such patients: POLE p.Leu424Val was identified as a de novo mutation in a patient with sporadic polyposis and CRC. While POLD1

To cite: Short E, Thomas LE, Hurley J, et al. J Med Genet Published Online First: [please include Day Month Year] doi:10.1136/jmedgenet-2015-103298

Copyright Article author (or their employer) 2015. Produced by BMJ Publishing Group Ltd under licence.
p.Ser478Asn was not observed, the group found a novel mutation, POLD1 c.1421T>C (p.Leu474Pro), in a family fulfilling the Amsterdam II criteria (that are used to identify families who may have LS), but whose members had MMR-proficient tumours. A similar study was carried out by Elsayed et al.\textsuperscript{10} They examined 1188 patients with familial CRC and polyposis for the two common mutations, and identified three patients (0.25\%) with the POLE p.Leu424Val variant.\textsuperscript{10} The tumours

**Figure 1** H&E microscopic images of colorectal polyps. (A) Tubular adenoma with low-grade dysplasia. (B) Hyperplastic polyp: note the presence of glands in the centre of the image with a serrated/tooth-like appearance. (C) A Peutz–Jeghers polyp. The glands are separated by bundles of smooth muscle. (D) A juvenile polyp. Cystically dilated glands are present in an oedematous and inflamed stroma.

**Figure 2** Diagram illustrating relationships between genes and signalling pathways involved in inherited colorectal cancer syndromes. FAP, familial adenomatous polyposis syndrome; HMPS, hereditary mixed polyposis syndrome; JPS, juvenile polyposis syndrome; LS, Lynch syndrome; MAP, MUTYH-associated polyposis; NTHL1, NTHL1-associated polyposis; PJS, Peutz–Jeghers syndrome; PPAP, polymerase proofreading-associated polyposis syndrome.
from these patients were microsatellite unstable, and immuno-
histochemistry showed MSH6/MSH2 deficiency, although there
were no germline mutations in the corresponding genes.10 Spier
et al11 carried out targeted sequencing of POLD1, POLD2, 
POLD3, POLE, POLE2, POLE3 and POLE4 in 266 unrelated
patients who had polyposis or who fulfilled Amsterdam criteria,
but in whom no germline mutation had been identified in LS or
other polyposis syndrome-associated genes. POLE p.Leu424Val
was detected in four patients, and the group identified a further
nine potentially pathogenic POL variants. Bellido et al12 studied
the exonuclease domains of POLE and POLD1 in 441 patients
with familial non-polyposis colorectal cancer and 88 patients
with polyposis. They used massively parallel sequencing to iden-
tify seven novel or rare genetic variants.13

Very recently, Weren et al13 applied whole exome sequencing
in 51 patients with multiple colorectal adenomas who did not
have identified mutations in APC or MUTYH. They found that
seven members of three families had homozygous germline
mutations in NTHL1. Tumours from affected individuals showed
a significant increase in the proportion of C:G>T:A transitions, consistent with the predicted effects of homozygous
loss-of-function mutations in NTHL1 on the accumulation of
somatic mutations. Further studies of larger patient cohorts are
required to characterise the newly recognised disorder
NTHL1-associated polyposis.

While mutations in genes that are already known to be asso-
ciated with polyposis syndromes can be identified in the major-
ity of patients with many tens or hundreds of polyps, the extent
to which unidentified changes in these genes account for the
remainder is unclear. Alternatively, further currently unknown
genets or other aetiologica factors may be involved in some
cases. Genes encoding components of pathways already implied
in CRC predisposition that regulate Wnt or BMP/TGF-B
signalling, cell adhesion, ubiquitin-mediated proteolysis and
DNA repair are further plausible functional candidates, as are
genets which have a role in other cancer predisposition syn-
dromes. Yurgelun et al14 have demonstrated that in a large
cohort of patients with early onset colorectal cancer, 1.3% (6/
457) had germline mutations in TP53, none of whom met the
clinical criteria for Li-Fraumeni syndrome. This frequency is
comparable with the prevalence of germline APC mutations in
colorectal cancer.14

UNUSUAL MUTATIONAL MECHANISMS IN MENDELIAN
COLORECTAL CANCER SYNDROMES

Historical approaches that were employed for molecular genetic
diagnosis such as Sanger sequencing and multiplex ligation-
dependent probe amplification have incomplete sensitivity for
the detection of mutations in coding sequences, and diagnostic
laboratories rarely test for point mutations in non-coding
sequences, except at intron–exon boundaries. In one study of
171 patients with multiple colorectal adenomas in whom APC
or MUTYH mutations had not been detected by Sanger sequen-
cing, the APC coding region was retested using high resolution
melt analysis leading to the detection of pathogenic heterozy-
gous APC mutations in 10 (6%) and mosaic APC mutations in
two (1%).15 Others have searched for and identified promoter
mutations, deep intronic mutations, complex genomic rearran-
gements and mosaic mutations in a proportion of polyposis
cases originally thought not to harbour mutations in APC or
MUTYH.16–21 The application of next-generation sequencing to
rapidly screen whole genomic loci at great depth overcomes
many of the limitations of other laboratory techniques, but
interpretation and functional characterisation of non-coding var-
ants remains problematic.

Constitutional epimutation, the epigenetic silencing of an
allele in normal somatic tissues, has been demonstrated to affect
MLH1 or MSH2 in a number of patients with LS, mostly spor-
adic cases, with MMR-deficient tumours.22–25 A variety of
sequence changes outside of the coding regions of the genes
itself has been shown to lead to promoter methylation and allele silencing. These include promoter mutations26 and
a relatively frequent deletion in the 3’ end of the EPCAM gene
that lies adjacent to MSH2. This leads to transcriptional read-
through from EPCAM into MSH2 and MSH2 promoter methy-
lation in the epithelial tissues in which EPCAM is expressed.27
(Constitutional epimutation has not been reported in other
CRC syndromes).

Mosaicism, the presence of genetically distinct cell lines in the
same individual, poses a particular challenge for mutation detec-
tion. Somatic mosaicism for APC mutations in which mutant
alleles are present at lower frequency than expected in heterozy-
gotes (ie, <50% and often much lower) appears to be relatively
common among patients with sporadic FAP.27 Segmental
somatic mosaicism in which the mutation is confined to specific
body segments requires analysis of tissue from the affected area,
and has been reported in patients with FAP who have had APC
mutations identified in the affected part of the colon, but
undetectable in other somatic tissues.28–30 The massively paral-
lel nature of next-generation sequencing makes this approach
very suitable for the detection of mutant alleles present at low
frequency,31 but does not overcome problems of segmental
mosaicism.

SERRATED POLYPOSIS SYNDROME

The aetiological basis of serrated polyposis syndrome (SPS) is
currently unknown. Most hyperplastic, metaplastic or serrated
polyps are small (<1 cm) sporadic lesions, but patients with SPS
develop multiple polyps that may be large (>1 cm). The condi-
tion was first described in 1980 by Williams et al.32 They
observed seven patients, with a mean age of 37.4 years, who
each had at least 50 colorectal polyps (then termed ‘metaplastic
polyposis’). At that time, the authors concluded that ‘it is impos-
tible to deduce whether or not “metaplastic polyposis” is a dis-
tinct entity. There is no good evidence that it is familial in this
small series, but the appearance of numerous metaplastic polyps
of an unusually large size and configuration, predominantly in
young males, might suggest a specific disease’. The identification
of further similar patients led Burt and Jass to propose a defini-
tion for ‘hyperplastic polyposis syndrome’ in the 2000 WHO
classification of tumours.33 This definition was modified in
2010, and the disease was renamed serrated polyposis syn-
drome.34 Current diagnostic criteria for SPS require a patient to
have:

1. at least five serrated (hyperplastic) polyps proximal to the
sigmoid colon, two of which are >10 mm in diameter or;
2. any number of serrated polyps proximal to the sigmoid
colon in an individual who has a first-degree relative with
serrated polyposis or;
3. more than 20 serrated polyps of any size distributed
throughout the colon.

It was thought that SPS affected 1 in 3000 individuals
between the ages of 55 and 64 years (reviewed in ref. 35) and as
many as 1 in 151 patients having a colonoscopy following a
positive faecal occult blood test (reviewed in ref. 36). SPS shows
no sex predilection, and the mean age of diagnosis is 55
years.\textsuperscript{37,38} In addition to hyperplastic lesions, there are often conventional adenomas present in the bowel.\textsuperscript{16}

Patients with SPS seem to fall into one of two ‘molecular groups’ (reviewed in refs. \textsuperscript{37,38}), one with relatively few large right-sided polyps, which harbour \textit{BRAF} mutations, the other with many small left-sided lesions, which often harbour \textit{KRAS} mutations (reviewed in refs. \textsuperscript{37,38}). The combined incidence of \textit{KRAS} and \textit{BRAF} mutations ranges from 64\% to 75\% (reviewed in refs. \textsuperscript{37,39}). Polyps with epithelial dysplasia have higher rates of these mutations (90\%) (reviewed in ref. \textsuperscript{37}). Some patients may have a combination of left-sided and right-sided lesions.\textsuperscript{18}

Patients with SPS are at an increased risk of developing CRC, which typically occurs between 50 and 60 years of age (reviewed in ref. \textsuperscript{34}). Malignancy is associated with a larger number of polyps, the presence of dysplasia (reviewed in refs. \textsuperscript{37,40}), and the existence of conventional adenomas in addition to hyperplastic polyps.\textsuperscript{35} The estimated incidence of CRC in patients with SPS varies from 14\% to 58\% (reviewed in refs. \textsuperscript{35–42}). Interestingly, a large proportion of CRCs seen in patients with SPS do not develop through the ‘serrated pathway of carcinogenesis’, driven by \textit{BRAF} mutation.\textsuperscript{35} The cancers show various molecular changes, including those more likely to be associated with the traditional adenoma–carcinoma pathway, for example, \textit{\beta}-catenin activation and/or overexpression of p53.\textsuperscript{35}

SPS may be a genetic disease, but if so, the mode of inheritance is unclear and recessive and dominant models have both been suggested (reviewed in ref. \textsuperscript{37}). SPS could be the result of a predisposition to promoter hypermethylation (reviewed in refs. \textsuperscript{37,43}), and some studies have reported hypermethylation in apparently normal colonic mucosa in patients with SPS (reviewed in ref. \textsuperscript{44}).

LOW-PENETRANCE ALLELES AND POLYGENIC PREDISPOSITION

Genome-wide association studies (GWAS) have so far identified approximately 40 CRC susceptibility loci, each of which is associated with a small increase in CRC risk (see online supplementary table S2). Variants at these loci are not strictly ‘mutations’, and may alter gene expression or function or, more usually, be in linkage disequilibrium with the mechanistic variant. On the basis of an additive model, the 10 CRC loci initially identified were thought to collectively account for approximately 6\% of excess familial risk.\textsuperscript{45} However, it is clear that such an additive model cannot adequately explain increases in disease risk,\textsuperscript{46} and it appears that the combined effect of such variants follows a multiplicative (log additive) risk model (reviewed in ref. \textsuperscript{47}). The risks associated with each of the variants have been much too small for translation to testing in clinical practice, but the development of algorithms estimating cumulative risks associated with carriage of multiple alleles are expected to lead to clinical application, as has already been proposed in the context of breast cancer.\textsuperscript{45} Dunlop \textit{et al.}\textsuperscript{45} sought to assess the feasibility of CRC risk prediction using common genetic variant data combined with other risk factors. They used binary logistic regression to assess the combined effects of age, gender, family history and genotypes at 10 susceptibility loci that confer only small CRC risks. They found that while genotype data provided additional information that complemented age, gender and family history as risk factors, clinically useful individualised genetic risk prediction was not currently feasible. However, it was proposed that such risk prediction models might help to identify high-risk groups that could benefit from public health measures to prevent CRC.

INTERMEDIATE PENETRANCE ALLELES

While 15\%–35\% of colorectal cancers are believed to occur in the context of significant inherited predisposition (reviewed in ref. \textsuperscript{2,3}), currently identified genetic factors can probably only account for up to 10\% of CRCs. One type of currently missing heritability is likely to be mutations (variants) of intermediate penetrance, that is, those which confer a relative cancer risk from 1.5 to 5.\textsuperscript{20} In general, the allele frequencies of such mutations will be too low to have been identified in GWAS and penetrance too low to have been identified in traditional family linkage studies. However, as sequencing-based approaches have been applied, such mutations have already been identified in patients and families with breast cancer, in genes such as \textit{CHEK2}, \textit{ATM}, \textit{PALB2} and \textit{BRIP1},\textsuperscript{30} and it is anticipated that sequencing approaches in patients and families with CRC will identify comparable mutations that predispose to CRC. For example, a very recent report suggests that mutations in the DNA repair gene \textit{FAN1} may predispose to colorectal cancer.\textsuperscript{31}

CONCLUSION

CRC is the third most common cancer worldwide,\textsuperscript{1} and inherited factors are important in a significant proportion of cases. Previous genetic and genomic studies have identified the basis of only a small part of this heritability. Clinical application of testing for highly penetrant and therefore high-risk mutations has become an established part of managing CRC, and is enshrined in international guidelines.\textsuperscript{3–7} By contrast, the incomplete picture of the genomic architecture of inherited predisposition to CRC has so far precluded the clinical application of testing for low-penetrance variants. It is expected that identification of the missing heritable component of CRC will be resolved through evermore comprehensive cataloguing and phenotypic annotation of CRC-associated variants identified through sequencing approaches. Clinical applications in public health and CRC prevention may follow.

Acknowledgements The authors would like to thank Professor Geraint Williams for his help in identifying suitable microscopic images of colorectal polyps and the Wales Gene Park and Health and Care Research Wales (previously NISCHR) for support for the authors’ work.

Contributors All authors have been involved with writing the review article and with draft revision. All authors give their approval of the final version, and are accountable for its contents.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

new sporadic Lynch syndrome cases, with derivation of the epimutation on the paternal allele in one. *Int J Cancer* 2011;128:869–79.


Inherited predisposition to colorectal cancer: towards a more complete picture

Emma Short, Laura E Thomas, Joanna Hurley, Sian Jose and Julian R Sampson

J Med Genet  published online August 21, 2015

Updated information and services can be found at:
http://jmg.bmj.com/content/early/2015/08/21/jmedgenet-2015-103298

These include:

Supplementary Material
Supplementary material can be found at:
http://jmg.bmj.com/content/suppl/2015/08/21/jmedgenet-2015-103298.DC1.html

References
This article cites 45 articles, 7 of which you can access for free at:
http://jmg.bmj.com/content/early/2015/08/21/jmedgenet-2015-103298#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes