Mutagenesis vol. 27 no. 2 pp. 143–151, 2012 doi:10.1093/mutage/ger059

COGENT (COlorectal cancer GENeTics) revisited

Richard S. Houlston* and members of COGENT†

Division of Genetics and Epidemiology, Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK.

†The members of COGENT are given under the appendix section.

*To whom correspondence should be addressed. Section of Cancer Genetics, Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey, SM2 5NG, UK. Tel: +44 (0) 208 722 4175; Fax: +44 (0) 208 722 4634; Email: richard.houlston@icr.ac.uk

Received on June 7, 2011; revised on August 4, 2011; accepted on August 5, 2011

Many colorectal cancers (CRCs) develop in genetically susceptible individuals most of whom are not carriers of germ line mismatch repair or *APC* gene mutations and much of the heritable risk of CRC appears to be attributable to the co-inheritance of multiple low-risk variants. The accumulated experience to date in identifying this class of susceptibility allele has highlighted the need to conduct statistically and methodologically rigorous studies and the need for the multi-centre collaboration. This has been the motivation for establishing the COGENT (COlorectal cancer GENeTics) consortium which now includes over 20 research groups in Europe, Australia, the Americas, China and Japan actively working on CRC genetics. Here, we review the rationale for identifying low-penetrance variants for CRC and the current and future challenges for COGENT.

Background

Many colorectal cancers (CRCs) develop in genetically susceptible individuals most of whom are not carriers of germ line mismatch repair or *APC* gene mutations and much of the heritable risk of CRC is thought to be the consequence of the co-inheritance of multiple low-risk variants (1–3). Recent genome-wide association (GWA) studies have vindicated this hypothesis identifying single-nucleotide polymorphisms (SNPs) localizing to multiple chromosomal regions which influence CRC risk (4–11). While the risk of CRC associated with each of the variants is individually modest, taken together, they could make a significant contribution to disease burden by virtue of their high frequencies in the population.

As well as establishing a role for genetic susceptibility in the development of CRC, these data provide novel insight into disease causation; notably a number of risk variants annotate genes encoding components of the transforming growth factor-beta superfamily signaling pathway (4,10,11). An important long-term outcome of such GWA studies of CRC is that the knowledge gained about the underlying molecular basis of CRC may lead to the development of innovative therapeutic and preventative measures.

It is apparent that the successful identification of low-risk variants for CRC is contingent on having access to large case–control sample sets, something only realistically achievable through multicentre collaboration. This has been the motivation for establishing the COGENT (COlorectal cancer GENeTics) consortium (12). Here, we review the current state of knowledge regarding low-penetrance susceptibility to CRC and the opportunities for COGENT researchers to identify novel CRC predisposition genes.

Characteristics of low-penetrance variants and implications for discovery

In recent years, the introduction of the Human Genome Project and other international initiatives has allowed detailed examination of the entire genetic code. This information led to the development of comprehensive sets of tagging SNPs that capture a high proportion of common genetic variation. This coupled with the advent of high-throughput analytical platforms capable of simultaneously genotyping hundreds of thousands of SNPs heralded the advent of GWA studies. The GWA approach is agnostic in that it does not depend upon prior knowledge of function or presumptive involvement of any gene in disease causation.

The GWA studies of CRC that have been performed so far have reported SNPs at 14 independent genetic loci conclusively associated with CRC risk: 1q41 (4), 3q26.2 (4), 8q23.1 (*EIF3H*) (6), 8q24.21 (5,7), 10p14 (6), 11q23 (8,13), 12q13.13 (4), 14q22.2 (*BMP4*) (9), 15q13.3 (*GREM1*) (10), 16q22.1 (*CDH1*) (9), 18q21.1 (*SMAD7*) (11), 19q13.1 *(RHPN2)* (9), 20p12.3 (9) and 20q13.33 (*LAMA5*) (4) (Table 1). Data from these GWA studies are proving to be highly informative regarding the allelic architecture of CRC susceptibility. Firstly, the CRC risks associated with each of the variants at each of these loci are at best modest (relative risks of 1.1–1.3). Secondly, while there is little evidence of interactive effects between loci, the distribution of risk alleles follows a normal distribution in both CRC cases and controls, with a shift towards a higher number of risk alleles in affected individuals, consistent with a polygenic model of disease predisposition (Figure 1). Hence, by acting in concert, a small proportion of the population which carry a large number of risk alleles can have ~3-fold increase in risk compared to those with the median number of risk alleles. Thirdly, few common variants account for >1% of inherited risk and only a small proportion of the heritability of CRC can be explained by the currently identified loci. Fourthly, multiple functional variants can localise to the same chromosomal region, including low frequency variants with significantly larger effects on CRC risk. Finally, most of the loci map to regions bereft of genes or protein-encoding transcripts. Hence, it is likely that much of the common variation in CRC risk is mediated through sequence changes influencing gene expression, perhaps in

Table I. The 14 loci associated with CRC risk identified from GWA studies

SNP	Chromosome position	Gene[a]	Major allele	Minor allele	Control MAF	Odds ratio
rs6691170	1q41	—	G	**T**	0.35	1.13
rs10936599	3q26.2	*MYNN*	**C**	T	0.24	1.07
rs16892766	8q23.3	*(EIF3H)*	A	**C**	0.08	1.28
rs6983267	8q24.21	*c-MYC*	**G**	T	0.48	1.16
rs10795668	10p14	—	**G**	A	0.32	1.14
rs3802842	11q23.1	*C11orf93*	A	**C**	0.29	1.17
rs11169552	12q13.13	—	**C**	T	0.28	1.12
rs4444235	14q22.2	*BMP4*	T	**C**	0.47	1.09
rs4779584	15q13.3	*GREM1/SCG5*	C	**T**	0.19	1.20
rs9929218	16q22.1	*CDH1*	**G**	A	0.30	1.10
rs4939827	18q21.1	*SMAD7*	**T**	C	0.48	1.16
rs10411210	19q13.11	*RHPN2*	**C**	T	0.10	1.14
rs961253	20p12.3	—	C	**A**	0.35	1.14
rs4925386	20q13.33	*LAMA5*	**C**	T	0.32	1.11

BMP4, bone morphogenetic protein 4; *C11orf93*, chromosome 11 open reading frame 93 (hypothetical gene); *CDH1*, E-cadherin; *c-MYC*, v-myc avian myelocytomatosis viral oncogene homolog; *EIF3H*, eukaryotic translation initiation factor 3, subunit H; *GREM1*, gremlin 1; *LAMA5*, Laminin, alpha 5; MAF, minor allele frequency; MYNN, myoneurin; OR, odds ratio; *RHPN2*, Rho GTPase binding protein 2; *SCG5*, secretogranin V.
[a]Brackets indicate gene annotated for non-intragenic markers. The risk allele for each SNP is highlighted in bold.

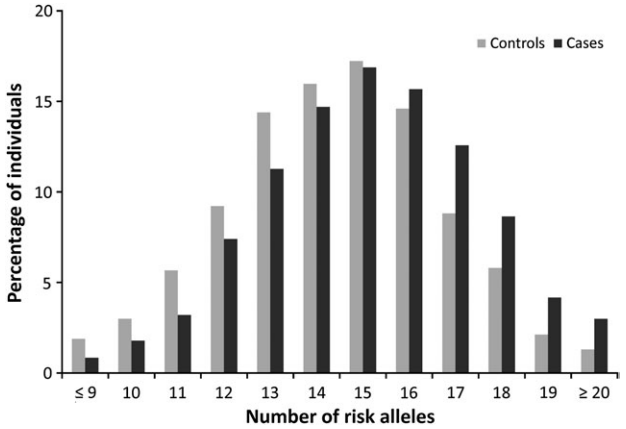

Fig. 1. Cumulative impact of the 14 variants on CRC risk. Distribution of risk alleles in controls (grey bars) and CRC cases (black bars) for the 14 loci (14).

a subtle fashion or through effects on pathway components mitigated by functional redundancy.

History of COGENT

Over a 10-year period, collaborations had been steadily developing between various researchers in the UK, Canada, the Americas, Holland, Germany, Finland, Spain and Australia that were engaged in studies of genetic susceptibility to CRC. What initially began from relatively loose affiliations centred around work on specific projects between individual groups begun to crystallise into a more formal collaborative network in 2007 with the advent of GWA studies of CRC. To continue and expand collaboration, a meeting was held at the University of Leiden, The Netherlands, in January 2009, to review the current state of ongoing association studies of CRC. There assembled an international team of researchers with expertise encompassing genetic epidemiology, statistical genetics, gene mapping, biology, molecular genetics, pathology and diagnosis and the clinical management of CRC. There was a consensus among participants that the challenges in this field of research could only be optimally addressed through international cooperative efforts and the group unanimously decided to establish the COGENT consortium (12). An invitation to join COGENT, subsequently extended to other groups known to be engaged in association studies of CRC, was well received. Subsequent to the Leiden meeting, a number of meetings have been held under the COGENT guise, notably in Barcelona and most recently in Edinburgh. These meetings have led to the consolidation of the group and presently, over 20 research groups are actively participating in COGENT led activities.

Future directions

Prospects for identifying novel risk variants through GWA-based analyses

The accumulated experience gained in conducting the GWA studies of CRC has served to highlight the difficulties in conducting statistically and methodologically rigorous association studies to identify novel CRC predisposition loci. The key issues are firstly, because of the large number of polymorphisms in the genome, false-positive associations are inevitably more frequent than true-positive associations when testing large numbers of markers even if studies are rigorously conducted; hence, associations need to attain a high level of statistical significance to be established beyond reasonable doubt. For this reason, in GWA studies, a *P*-value of 5.0×10^{-8} has been advocated as being appropriate threshold for defining genome-wide significance (15,16). Secondly, given that relative risks associated with variants are modest case–control studies involving just a few hundred cases and controls have very poor power to reliably identify genetic determinants conferring modest, but potentially important, risks. Thirdly, positive associations need to be replicated in independent case–control series to mitigate against Type 1 error. However, to increase power, the allelic architecture of the population from which these case–control series are ascertained needs to have similar ancestry and, ideally, the same linkage disequilibrium (LD) structure. Indeed, careful attention must be paid to population stratification as a source of confounding because cancer rates and allele frequencies vary with race/ethnicity. Wide comparisons between the population genetics of different ethnic groups have shown that SNP allele frequencies can vary greatly between ethnic groups, principally as a result of founder effects and genetic drift. Indeed, some SNPs may be informative in one population and not in another; consequently, some CRC modifier loci may exist in some populations but not in others. Failure to account for this is likely to be the reason for many of the associations reported over the years are spurious.

The need to improve power to identify novel low-risk variants for CRC

For reasons of cost efficiency, GWA studies are performed adopting a staged design, whereby the best ranked SNPs from one stage are genotyped in progressively larger datasets to attain requisite genome-wide statistical significance for associations. Figure 2 illustrates the power of a two-stage GWA study stipulating a statistical threshold of 10^{-5} over the first two stages. SNPs which are truly associated at this threshold can generally be shown to be associated at 5.0×10^{-8} in subsequent large case–control series. Hence, power of the

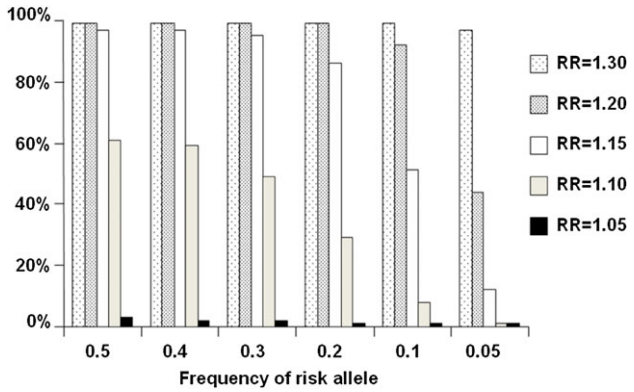

Fig. 2. Power to identify risk loci for CRC over a range of minor allele frequencies and relative risks. Illustrative study based on two-stage design—3000 cases and 3000 controls typed for 500 000 tagging SNPs in Stage 1 and 7000 cases and 7000 controls typed for 50 000 'top' ranking SNPs in Stage 2. *P*-value in combined analysis thresholded at 10^{-5}.

GWA studies that have been conducted to identify common alleles conferring relative risks of $\geq$1.2 (such as the 8q24 variant) is thus high (Figure 2). Therefore, there are unlikely to be many additional CRC SNPs with similar effects for alleles with frequencies of >20% in populations of European ancestry. The GWA studies have, however, had limited power to detect alleles with smaller effects and/or risk allele frequencies of <10%. By implication, variants with such profiles are likely to collectively confer substantial risk because of their multiplicity or sub-maximal LD with tagging SNPs. Hence, the current GWA-based strategies are not configured optimally to identify low frequency variants with potentially stronger effects or identify recessively acting alleles. Nor are current arrays formatted ideally to capture certain structural variants such as small scale insertions or deletions, which may impact on CRC risk. It is therefore likely that additional low-risk variants for CRC remain to be identified by GWA studies. Ongoing GWA studies of CRC being conducted by different research groups will inevitably generate SNP data from different array platforms with different SNP representation and coverage. Using statistical methodology whereby imputation of untyped SNPs can be generated in datasets allows for harmonisation and pooled analyses to be conducted (17). COGENT is investing heavily in efforts to expand the scale of GWA meta-analyses both in terms of sample size and SNP coverage and the number of SNPs taken forward for replication. Collectively, over 60 000 CRC cases and 57 000 controls have so far been accrued by COGENT researchers (Table II). COGENT is extremely well equipped to meet the requirement for large-scale replication analyses necessary to identify novel risk loci.

Strategies for identifying causal variants

Data has shown that the rs69783267 (8q24) association is a direct consequence of the SNP which differentially affects TCF4 binding and has cis-regulatory effects on myc promoter activity (18,19). For most associations, the SNPs directly typed in GWA studies are, however, unlikely to be directly functional but are correlated with a functional variant. This is exemplified by the 8q23 and 18q24 associations which have recently shown that the SNPs are correlated with variants having allele-specific cis-effects on *EIF3H* and *SMAD7* expression, respectively (20,21). Although blocks of LD allow the efficient survey of the genome, they hamper fine mapping of the disease-associated region, hence identifying the functional basis of associations is challenging. Different ethnic groups can have different LD block patterns which can be used to refine the location of a disease susceptibility locus prior to fine mapping genotyping and functional analyses. One recent development that greatly facilitates fine-mapping is the use of imputation of untyped SNPs from reference population panels which have been extensively genotyped. HapMap 2 has until recently been a major source of such reference data providing genotypes for ~8 M SNPs in four main ethnicities. Initiatives such as the 1000Genomes project have recently greatly improved polymorphic annotation of the human genome harvesting ~30 M SNPs thereby increasing the value of imputation.

Interactions between susceptibility alleles

It is entirely conceivable that epistatic interactions between common risk variants may exist. To the extent that they have been examined, the effects of the currently identified common low-penetrance alleles for CRC appear to be essentially independent. Such interactions are, however, difficult to detect unless marginal effects are significant, and there is also a large statistical penalty from large-scale multiple testing. It has recently been proposed that for most plausible interaction effects, a two-stage analysis has been shown to dramatically increase the power to identify interactions compared to a single-stage analysis (22).

One important implication of common susceptibility is the modification of the risks associated with high-penetrance susceptibility. Data are currently limited, but two recent studies have independently shown that 8q23.3 and 11q23.1 genotype modify CRC risk in Lynch syndrome (23,24).

Search for rare disease-causing variants

Much of the current thinking regarding polygenic susceptibility to CRC has been dominated by the 'common-disease common-variant' model. In addition to this class of disease susceptibility, it is essential to consider alternative models of CRC predisposition, based on rare disease causing variants. Such rare germ line variants may confer more profound effects on risk and hence have greater significance for individuals, though the population-attributable risk may be low. Examples of this class of susceptibility allele are provided by the *ATM*, *BRIP1*, *PALB2* and *CHEK2* variants which are detectable in 0.5–1% of many European populations and confer a 2- to 3-fold increased risk of breast cancer (25–28). In addition, as seen with *MUTYH* variants, alleles may act recessively and confer substantive risks of CRC (29). While some low frequency risk alleles may be harvestable through exploitation of pre-existing GWA data, for most, the sequencing of large numbers of CRC cases will be required for their identification. Advances in sequencing technology are making an increasing feasible proposition.

Sub-group analyses

To date, most searches for CRC risk variants have been largely predicted on the assumption of CRC being a homogeneous disease. CRC, however, displays considerable heterogeneity as evidenced by differences in the spectrum of somatic mutation seen in colon and rectal cancers which is likely to reflect differences in aetiological risk factors, both environmental and inherited (30,31). Moreover, the molecular profiles of mismatch repair (MMR) deficient and competent cancers are distinctive and are likely to be influenced by different risk factors (31).

Table II. Number of CRC cases and controls currently established by COGENT consortium members

	Study name	General setting	Number of subjects	
			Cases	Controls
European				
Institute of Cancer Research, UK	NSCCG (National Study of Colorectal Cancer) (33).	Population-based UK study. Spouse controls from NSCCG (33) and GELCAPS (Genetic Lung Cancer Predisposition Study) (14).	18 800	8500
Edinburgh University, UK	COGS (Colorectal Cancer Genetics Susceptibly Study).	Population-based incident case series aged <55 years at diagnosis. Population-based controls.	1012	1012
	SOCCS (Scottish Colorectal Cancer Study).	Population-based incident case series; Scotland, UK.	3000	3000
Oxford University, UK	CORGI (Colorectal Tumour Gene Identification Consortium).	Cases with family history of CRC ascertained through clinical genetics centres in the UK. Spouse controls with no personal or family history of CRC	940	965
	VICTOR—Post-treatment stage of a Phase III, randomised double blind, placebo-controlled study of rofecoxib (VIOXX) in colorectal cancer patients following potentially curative therapy.	Samples from a closed clinical trial.	910	—
	QUAZAR2—Multicentre international study of capecitbine +/− bevacizumab as adjuvant treatment of CRC.	UK blood donors.	139	376
Cambridge University, UK	SEARCH (Studies of Epidemiology and Risk Factors in Cancer Hereditary).	Population based case–control study; Cambridge, UK.	3000	3000
Cardiff University in collaboration with the MRC Clinical Trials Unit, London, UK	COIN and COIN-B –MRC-funded trials comparing either COntinuous chemotherapy plus cetuximab of INtermittent chemotherapy with the standard palliative combination chemotherapy with oxaliplatin and a fluropyrimidine in first line treatment of patients with metastatic colorectal cancer.	Trial-based cohorts of advanced CRC	2200	
WTCCC UKBS controls				3200
Barcelona and Santiago, Spain	EPICOLON Consortium.	Population based case–control study; Spain.	2000	2000
Barcelona, Spain	ENTERICOS (Disinfection by-products and other Environmental, genetic and molecular determinants of colorectal cancer - Subproductos de la desinfección y otros determinantes ambientales, genéticos y moleculares del cáncer colorectal en España).	Case–control study of CRC to evaluate the increased risk associated with chronic DBP exposure through ingestion, inhalation and dermal absorption.	500	500
	Bellvitge case–control study.		370	325
University of Helsinki, Finland	FCCPS (Finnish Colorectal Cancer Predisposition Study).	Population based study; South-eastern Finland.	1440	2000
Karolinska Institute, Swede	The Swedish Low Risk Colorectal Cancer Study Group	Unselected cases ascertained through 12 hospitals serving the Stockholm-Gotland and Uppsala-Örebro health-care regions in Sweden. Blood donor controls.	3000	3000

Table II. *Continued*

	Study name	General setting	Number of subjects	
			Cases	Controls
German Cancer Research Center (DKFZ): on behalf of German HNPCC consortium	German HNPCC consortium.	Familial non-HNPCC cases recruited through German HNPCC consortium, principally through 6 hospitals of Bochum, Bonn, Dresden, Düsseldorf, Heidelberg and Munich/Regensburg. Controls: unrelated and ethnicity- and age-matched blood donors recruited by the Institute of Transfusion Medicine and Immunology, Faculty of Mannheim, Germany.	1000	1000
University of Keil and Greifswald, Germany	POPGEN (Population Genetic Cohort) from Schleswig-Holstein, north Germany. SHIP (Survey of Health in Pommerania) from east and north-east Germany.	Population-based biobank projects.	2720	2720
German Cancer Research Center	ESTHER (Epidemiologische Studie zu Chancen der Verhütung, Früherkennung und optimierten Therapie chronischer Erkrankungen in der älteren Bevölkerung).	Population-based biobank project.	670	670
Institute of Experimental Medicine, Academy of Science, Czech Republic	—	Unselected CRC cases mainly recruited from 10 oncological departments across the Czech Republic. Controls hospital patient and blood donors (45,46).	1300	2600
University of Groningen, The Netherlands	SCOPE study.	Unselected CRC cases, hospital patient controls from the Netherlands.	774	1000
University of Leiden, The Netherlands		Unselected CRC cases. Controls ascertained through genetic testing programmes for non-cancer related conditions.	1500	1500
Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy		Unselected CRC cases, population controls.	1000	1200
Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy	INT (Istituto Nazionale Tumori (study on genetics of sporadic colorectal cancer.	Unselected CRC cases with detailed clinical information, population controls	800	2000
UHC 'Sestre milosrdnice', University of Zegreb, Croatia		Unselected CRC cases, population-based controls	700	700
Australia				
Ludwig Institute for Cancer Research, Melbourne	Victorian Cancer Biobank.	Population-based biobank project.	1000	500
The University of Newcastle, New South Wales	Hunter Family Cancer Service.	Population based collection of cases and controls from the Hunter Region of New South Wales.	600	3000
The Americas				
Ibague, Colombia. Universidad del Tolima		Unselected CRC cases, population-based controls.	500	700
Toronto, Canada	OFCCR (Ontario Familial Colorectal Cancer Registry).	Population-based case–control study; Ontario.	1257	1336
Case Western Reserve University, USA	Kentucky Colon Cancer Genetic Epidemiology Study.	Population-based case–control study.	1267	1771

Table II. *Continued*

	Study name	General setting	Number of subjects	
			Cases	Controls
Asia				
University Hong Kong Medical Centre, China	UHKMC series.	Unselected CRC cases, hospital patient controls.	3000	3000
University of Tokyo, Japan	Biobank Japan.	Population-based biobank project.	6000	6000
TOTALS			61 399	57 575

Evidence that common risk alleles can have subtype effects on CRC risk is provided by the 11q23 association which appears highly specific for rectal disease (8,13). While stratified analyses provide a means of teasing out important subtype specific effects, the numbers of cases in many subgroups will inevitably constraint study power. This fact further underscores the value of bringing together independent case–control series for validation analyses through initiatives such as COGENT.

Incorporating non-genetic risk factors into risk models

CRC risk is undoubtedly determined by complex interactions between genetic and lifestyle/dietary risk factors. Epidemiological studies have established several dietary risk factors for colorectal neoplasia; these include low vegetable and high red meat consumption and micronutrient deficiency and excessive alcohol intake. There is a weaker association between CRC, smoking and lack of physical activity. Common genetic variants are thus likely to interact with these environmental lifestyle risk factors to modify risk. Furthermore, common variants have the potential to determine the effectiveness of chemoprevention agents such as non-steroidal anti-inflammatory drugs, hormone replacement therapy and micronutrient supplementation.

In assessing the interplay between inherited and non-genetic risk factors, analyses using different population cohorts should be in theory, highly informative. At least in principle and probably in practice, some variants may have stronger or weaker effects on disease depending on environment or general genetic background as observed in inbred lines of mice. Hence, while consortia effectively permits for an increase in sample size, phenotype heterogeneity across studies represents a major obstacle potentially offsetting successful identification of gene by environment (G $\times$ E) interactions.

Even accepting such considerations, it has, however, been questioned whether any G $\times$ E studies are, in general, even possible, let alone worthwhile, especially given that the smaller the odds ratio, the more likely it will be that environmental factors will predominate (32). Despite such cautionary reservations, there is considerable interest in looking for G $\times$ E interactions, even when no main effects exist for either which is likely to generate a plethora of Type 1 errors.

Irrespective of such cautionary notes, incorporating environmental risk factor data into models of predisposition is likely to be a serious challenge. While ethnicity can be defined through genotype, environmental background is harder to standardise across studies and data harmonisation is likely to be a major obstacle to the generation of meaningful data.

Inherited prognostic and predictive variants

In addition to influencing the risk of developing CRC, inherited genetic factors may play a role in determining the natural course of the disease and its response to therapies. As a prognostic factor, the concept of germ line variation imparting inter-individual variability in tumor progression is currently receiving increasing attention for a number of cancers. Clearly, in CRC, there is a precedent for this notion as MMR status in tumours, which can be a consequence of germ line mutation, impacts on patient prognosis (33). Furthermore, there is some evidence that SNP genotypes may be preferentially associated with specific CRC histology (24). However, to date, there are no reliable examples of common variants influencing patient outcome from CRC. It is probably that a genetic variant affecting inter-individual disease expression will impact on later stages of clonal development rather than early events associated with an inherited susceptibility. For example, variants in growth factors or immune surveillance signalling pathways might not impact on risk of initiation but could have a substantial effect on progression or outcome of established disease. Chemotherapy response and toxicity may also be related to germ line genotype. As with conventional association studies, it is essential to impose appropriate statistical thresholds and conduct replication analyses to avoid the reporting of false positives. Linking GWA data to patient outcome provides an attractive strategy for identifying prognostic markers of outcome from CRC. It should be noted, however, that in a two-stage design for GWA studies on CRC patient' prognosis, the discovery and the validation series should be similar for phenotypes affecting prognosis. Ideally such analyses should be based on data derived from clinical trials as these patients are in receipt of standardised treatment, which may not be the case with other types of patient samples.

Concluding remarks

COGENT represents a major international collaborative study seeking to comprehensively understand the impact of inherited susceptibility to CRC and to describe the genetic landscape of the disease. The close cooperation between research groups which has been fostered will undoubtedly allow us to meet future challenges of identifying and characterising novel CRC risk alleles. The immediate goal is to work together collaboratively to study polymorphisms that have been shown to be associated with risk and to plan for future high quality biological and epidemiological studies as longer term aims. The consortium is keen to engage and have the involvement of other interested researchers and can be contacted through any of the COGENT members.

Funding

The work of UK groups is supported by grants from Cancer Research UK and the European Union. In Wales, work is

supported by grants from Cancer Research Wales to J.P.C. and Tenovus to J.P.C. We acknowledge the use of DNA from the UKBS, funded by the Wellcome Trust (grant 06113/C/04/Z); by the Juvenile Diabetes Research Foundation grant WT061858 and by the National Institute of Health Research of England. In Spain, work is supported by grants from the Instituto de Salud Carlos III, grants FIS 05/1006, 08/1359, 08/1635 and European Commission FP6 Food-CT-2006-036224 to V.M.. In Canada, work is supported by grants from the Ontario Research Fund from the Ontario Ministry of Research and Innovation, The Canadian Cancer Society Research Institute, the Canadian Institutes of Health Research and the Ontario Institute for Cancer Research. In Finland, work is supported by grants from Academy of Finland (Finnish Centre of Excellence Program 2006–2011), the Sigrid Juselius Foundation, the Finnish Cancer Societies and the European Union.

The work of the DFKS is supported by the German Genome Network (NGFNplus). In the Czech Republic, work is supported by GACR 310/07/1430 and P304/10/1286. In Sweden, work is supported by grants from the Swedish Cancer Foundation and the Swedish Research Council. In Barcelona and Santiago (Epicolon), work is supported by grants from Fondo de Investigación Sanitaria/FEDR (08/0024, 08/1276, PS09/02368, 11/00219, 11/0068), Xunta de Galicia (PGIDIT07PXIB9101209PR), Ministerio de Educación y Ciencia (SAF2010-19273), Asociación Española contra el Cáncer (Fundación Cientifica y Junta de Barcelona) Fundación Olga Torres to C.R.P. and Acción Transversal contra el Cáncer (Instituto de Salud Carlos III) and FP& CHIBCHA Constrotium to S.C.B. and Angel Carracedo CIBERER and CIBEREHD 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). The work was carried out (in part) at the Esteher Koplowtiz Centre, Barcelona. In the Netherlands, work of R.M.W.H is supported by the Dutch Cancer Society, the European Community, and the work of J.T.W., T.vW., H.M. and PD by the Dutch Cancer Society (UL2005-3247) and Fonds NutsOhra. In Hong Kong, work is supported by grants from the Michael and Betty Kadoorie Cancer Genetics Research Programme II; and the Bobby Moore Fund of Cancer Research UK. Work in Australia is conducted under the auspices of the Hilton Ludwig Cancer Metastasis Initiative and supported by a grant from the NHMRC (Project Grant 489418). In Japan, this work was conducted as a part of the BioBank Japan Project that was supported by the Ministry of Education, Culture, Sports, Sciences and Technology of Japanese government. The ESTHER and VERDI studies are supported by grants from the Baden Württemberg Ministry of Research, Science and Arts and the German Cancer Aid (Deutsche Krebshilfe), grant M24/95/BR I. The Kiel cohort (POPGEN) is funded through the German Ministry of Education and Reserach through the POPGEN Biobank grant and the Colon Cancer Network of the German National Genome Research Network. SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research, the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG. M.M.E., A.V. and L.C.-C. have received funding from Cancer Research UK and Universidad del Tolima. This study is supported by grants from Associazione/Fondazione Italiana per la Ricerca sul Cancro (AIRC/FIRC). The Kentucky Colon Cancer Genetic Epidemiology Study is supported by grants from the National Cancer Institute (CA136726) and the Damon-Runyon Cancer Research Foundation (CI-8).

Acknowledgements

We acknowledge the use of DNA from the UKBS. The Epicolon consortium is sincerely grateful to all patients participating who were recruited in 25 (EPICOLON 1) and 14 (EPICOLON 2) Spanish hospitals Provision of genotyping is gratefully provided by Santiago de Compostela and Barcelona nodes of the Spanish National Genotyping Center (CeGen). In Italy, collection of samples and clinical data of individuals recruited at Fondazione IRCCS, Istituto Nazionale Tumori was possible thanks to efforts by Silvia Veneroni, Laura Galastri, Fernando Ravagnani, Tiziana Camerini, Maria Gaetana Di Mauro, Alberto Vannelli, Tommaso A. Dragani, Maria Grazia Daidone, Marco A. Pierotti. URLs: Epidemiologische Studie zu Chancen der Verhütung, Früherkennung und optimierten Therapie chronischer Erkrankungen in der älteren Bevölkerung (ESTHER)—www.esther.dkfz.org/esther/.

Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH)

QUASAR

ENTERICOS—http://www.hiwate.eu/news/entericos-case-control-study-colon-cancer-spain-receives-extra-funding-spanish-ministy-health-c.

Biobank Japan—http://www.src.riken.go.jp/english/project/person/index.html.

Victorian cancer biobank—http://www.viccancerbiobank.org.au/.

National Study of Colorectal Cancer Genetics (NSCCG)—http://www.icr.ac.uk/research/research_sections/cancer_genetics/cancer_genetics_teams/molecular_and_population_genetics/nsccg/index.shtml.

References

1. Aaltonen, L., Johns, L., Jarvinen, H., Mecklin, J. P. and Houlston, R. (2007) Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. *Clin. Cancer Res.*, **13**, 356–361.
2. Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. and Hemminki, K. (2000) 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.
3. Lubbe, S. J., Webb, E. L., Chandler, I. P. and Houlston, R. S. (2009) Implications of familial colorectal cancer risk profiles and microsatellite instability status. *J. Clin. Oncol.*, **27**, 2238–2244.
4. Houlston, R. S., Cheadle, J., Dobbins, S. E. *et al.* (2010) 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.
5. Tomlinson, I., Webb, E., Carvajal-Carmona, L. *et al.* (2007) A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.*, **39**, 984–988.
6. Tomlinson, I. P., Webb, E., Carvajal-Carmona, L. *et al.* (2008) A genome-wide association study identifies colorectal cancer susceptibilit/y loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.*, **40**, 623–630.
7. Zanke, B. W., Greenwood, C. M., Rangrej, J. *et al.* (2007) Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.*, **39**, 989–994.
8. Tenesa, A., Farrington, S. M., Prendergast, J. G. *et al.* (2008) 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.
9. Houlston, R. S., Webb, E., Broderick, P. *et al.* (2008) Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat. Genet.*, **40**, 1426–1435.
10. Jaeger, E., Webb, E., Howarth, K. *et al.* (2008) Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat. Genet.*, **40**, 26–28.
11. Broderick, P., Carvajal-Carmona, L., Pittman, A. M. *et al.* (2007) A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat. Genet.*, **39**, 1315–1317.

12. Tomlinson, I. P., Dunlop, M., Campbell, H. *et al.* (2010) COGENT (COlorectal cancer GENeTics): an international consortium to study the role of polymorphic variation on the risk of colorectal cancer. *Br. J. Cancer*, **102**, 447–454.
13. Pittman, A. M., Webb, E., Carvajal-Carmona, L. *et al.* (2008) Refinement of the basis and impact of common 11q23.1 variation to the risk of developing colorectal cancer. *Hum. Mol. Genet.*, **17**, 3720–3727.
14. Lubbe, S. J., Di Bernardo, M. C., Broderick, P., Chandler, I. and Houlston, R. S. (2011) Comprehensive. evaluation of the impact of 14 genetic variants on colorectal cancer phenotype and risk. *Am. J. Epidemiol.*, (in press).
15. Dudbridge, F. and Gusnanto, A. (2008) Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.*, **32**, 227–234.
16. International HapMap Consortium. A haplotype map of the human genome. *Nature*, **437**, 1299–1320.
17. Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
18. Pomerantz, M. M., Ahmadiyeh, N., Jia, L. *et al.* (2009) The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat. Genet.*, **41**, 882–884.
19. Tuupanen, S., Turunen, M., Lehtonen, R. *et al.* (2009) The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat. Genet.*, **41**, 885–890.
20. Pittman, A. M., Naranjo, S., Jalava, S. E. *et al.* (2011) Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a cis-acting regulator of EIF3H. *PLoS Genet*, **7**, e1002105.
21. Pittman, A. M., Naranjo, S., Webb, E. *et al.* (2009) The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. *Genome Res.*, **19**, 987–993.
22. Kooperberg, C. and Leblanc, M. (2008) Increasing the power of identifying gene x gene interactions in genome-wide association studies. *Genet. Epidemiol.*, **32**, 255–263.
23. Wijnen, J. T., Brohet, R. M., van Eijk, R. *et al.* (2009) Chromosome 8q23.3 and 11q23.1 variants modify colorectal cancer risk in Lynch syndrome. *Gastroenterology*, **136**, 131–137.
24. Talseth-Palmer, B. A., Brenne, I. S., Ashton, K. A. *et al.* (2010) Colorectal cancer susceptibility loci on chromosome 8q23.3 and 11q23.1 as modifiers for disease expression in lynch syndrome. *J. Med. Genet.*, **48**, 279–284.
25. Rahman, N., Seal, S., Thompson, D. *et al.* (2007) PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat. Genet.*, **39**, 165–167.
26. Renwick, A., Thompson, D., Seal, S. *et al.* (2006) ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat. Genet.*, **38**, 873–875.
27. Meijers-Heijboer, H., van den Ouweland, A., Klijn, J. *et al.* (2002) Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat. Genet.*, **31**, 55–59.
28. Seal, S., Thompson, D., Renwick, A. *et al.* (2006) Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat. Genet.*, **38**, 1239–1241.
29. Al-Tassan, N., Chmiel, N. H., Maynard, J. *et al.* (2002) Inherited variants of MYH associated with somatic G: C–>T: a mutations in colorectal tumors. *Nat. Genet.*, **30**, 227–232.
30. Wei, E. K., Giovannucci, E., Wu, K., Rosner, B., Fuchs, C. S., Willett, W. C. and Colditz, G. A. (2004) Comparison of risk factors for colon and rectal cancer. *Int. J. Cancer*, **108**, 433–442.
31. Iacopetta, B. (2002) Are there two sides to colorectal cancer? *Int. J. Cancer*, **101**, 403–408.
32. Bodmer, W. and Bonilla, C. (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.*, **40**, 695–701.
33. Popat, S., Hubner, R. and Houlston, R. S. (2005) Systematic review of microsatellite instability and colorectal cancer prognosis. *J. Clin. Oncol.*, **23**, 609–618.

Appendix 1

Lauri A. Aaltonen: Department of Medical Genetics, Genome-Scale Biology Research Program, Biomedicum 9, Helsinki, University of Helsinki, Helsinki, Finland.

Hermann Brenner: Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.

Stephan Buch: Department of General Internal Medicine, University Hospital, Schleswig-Holstein, Campus Kiel, Schittenhelmstraße 12, 24105 Kiel, Germany.

Harry Campbell: Public Health Sciences, University of Edinburgh, Edinburgh EH89AG, UK.

Angel Carracedo: Fundacion Publica Galega de Medicina Xenomica (FPGMX), CIBERER, Genomic Medicine Group-University of Santiago de Compostela, Hospital Clinico, Santiago de Compostela, Galicia, Spain.

Luis Carvajal-Carmona: Molecular and Population Genetics, Nuffield Department of Medicine, University of Oxford, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK; Oxford Comprehensive Biomedical Research Centre; Departamento de Biología, Universidad del Tolima, Barrio Altos de Santa Helena, Ibague, Tolima, Colombia.

Antoni Castells: Department of Gastroenterology, Institut de Malalties Digestives i Metabòliques, Hospital Clínic, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), IDIBAPS, University of Barcelona, Barcelona, Catalonia, Spain.

Sergi Castellví-Bel: Department of Gastroenterology, Institut de Malalties Digestives i Metabòliques, Hospital Clínic, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), IDIBAPS, University of Barcelona, Barcelona, Catalonia, Spain.

Jeremy P. Cheadle: School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4XN, UK.

Peter Devilee: Department of Human and Department of Clinical Genetics, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands

Malcolm Dunlop: Institute of Genetics and Molecular Medicine, University of Edinburgh, MRC-HGU, Western General Hospital, Crewe Road South, Edinburgh EH4 2XU, UK.

Magdalena M. Echeverry: Departamento de Biología, Universidad del Tolima, Barrio Altos de Santa Helena, Ibague, Tolima, Colombia.

Steven Gallinger: Cancer Care Ontario, 620 University Avenue Toronto, Ontario, Canada M5G 2L7; Samuel Lunenfeld Research Institute, Mount Sinai Hospital and University of Toronto, 600 University Avenue Toronto, Ontario, Canada M5G 1X5.

Antonella Galvan: Fondazione IFOM, Istituto FIRC di Oncologia Molecolare, Milan, Italy.

Jochen Hampe: Department of General Internal Medicine, University Hospital, Schleswig-Holstein, Campus Kiel, Schittenhelmstraße 12, 24105 Kiel, Germany.

Kari Hemminki: German Cancer Research Center, Heidelberg, Germany.

Judy W. C. Ho: The University of Hong Kong, Pokfulam Road, Hong Kong, China.

Robert M. W. Hofstra: Department of Genetics, University Medical Center Groningen, University of Groningen, PO Box 30.0001, 9700 RB Groningen, The Netherlands.

Tom J. Hudson: The Ontario Institute for Cancer Research, The MaRS Center, 101 College Street, Suite 800, Toronto, Ontario, Canada M5G 1L7.

Iva Kirac: Department of Surgery, University Hospital Center Sestre milosrdnice', School of Medicine, University of Zagreb, Croatia.

Markus M. Lerch: Klinik für Innere Medizin A University Hospital Greifswald, Friedrich-Loeffler-Strasse 23a, 17487 Greifswald, Germany.

Li Li: Department of Family Medicine, Case Western Reserve University/ Case Comprehensive Cancer Center, Cleveland, OH, USA.

Annika Lindblom: Department of Molecular Medicine and Surgery, Karolinska Institutet, CMM02, S17176 Stockholm, Sweden.

Lara Lipton: LCCI Biomarker Laboratory, Ludwig Institute for Cancer Research, PO Box 2008, Royal Melbourne Hospital, Victoria 3050, Australia.

Koichi Matsuda: Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Japan.

Timothy S. Maughan: School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4XN, UK.

Victor Moreno: IDIBELL-Catalan Institute of Oncology and University of Barcelona, Avenida Gran Via 199, L'Hospitalet, 08907 Barcelona, Spain.

Hans Morreau: Department of Pathology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands.

Alessio Naccarati: Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, Videnska 1083, 142 00 Prague 4, Czech Republic.

Yusuke Nakamura: Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Japan.

Paolo Peterlongo: Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy.

Paul D. Pharoah: Department of Oncology, University of Cambridge, UK

Oliver Sieber: LCCI Biomarker Laboratory, Ludwig Institute for Cancer Research, PO Box 2008, Royal Melbourne Hospital, Victoria 3050, Australia.

Paolo Radice: Fondazione IFOM, Istituto FIRC di Oncologia Molecolare, Milan, Italy; Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy.

Clara Ruiz-Ponte: Fundacion Publica Galega de Medicina Xenomica (FPGMX), CIBERER, Genomic Medicine Group-University of Santiago de Compostela, Hospital Clinico, Santiago de Compostela, Galicia, Spain.

Clemens Schafmayer: POPGEN Biobank, University Hospital Schleswig-Holstein, Campus Kiel, Schittenhelmstrasse 12, 24105 Kiel, GermanyDepartment of General and Thoracic Surgery, University Hospital Schleswig-Holstein, Campus Kiel, Arnold-Heller-Strasse 3, 24105 Kiel, Germany.

Christian A. Schmidt: Klinik für Innere Medizin C, University Hospital Greifswald, Ferdinand-Sauerbruch-Strasse, 17487 Greifswald, Germany.

Witigo von Schönfels: Department of General and Thoracic Surgery, University Hospital Schleswig-Holstein, Campus Kiel, Arnold-Heller-Strasse 3, 24105 Kiel, Germany.

Stefan Schreiber: Department of General Internal Medicine, University Hospital, Schleswig-Holstein, Campus Kiel, Schittenhelmstraße 12, 24105 Kiel, Germany.

Rodney Scott: School of Biomedical Sciences, Faculty of Health, University of Newcastle, New South Wales, Australia.

Pak Sham: The University of Hong Kong, Pokfulam Road, Hong Kong, China.

Pavel Soucek: Toxicogenomics Department, Center for Toxicogenomics and Safety, National Institute of Public Health, Prague 10, Czech Republic.

Albert Tenesa: Institute of Genetics and Molecular Medicine, University of Edinburgh, MRC-HGU, Western General Hospital, Crewe Road South, Edinburgh EH4 2XU, UK; The Roslin Institute, The University of Edinburgh, Easter Bush, Roslin EH25 9RG, UK.

Ian P. M. Tomlinson: Molecular and Population Genetics, Nuffield Department of Medicine, University of Oxford, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK; Oxford Comprehensive Biomedical Research Centre

Alejandro Velez: Departamento de Patología, Hospital Pablo Tobon Uribe, Calle 78 B No. 69-240 Medellín, Colombia.

Cristna M. Villanueva: Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy

Pavel Vodicka: Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, Videnska 1083, 142 00 Prague 4, Czech Republic.

Henry Völzke: Institut für Community Medicine, University Hospital Greifswald, Walther-Rathenau-Strasse 48, 17487 Greifswald, Germany.

Tom van Wezel: Department of Pathology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands.

Juul T. Wijnen: Department of Human and Department of Clinical Genetics, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands.

Brent Zanke: Cancer Care Ontario, 620 University Avenue Toronto, Ontario, Canada M5G 2L7; The Ontario Institute for Cancer Research, The MaRS Center, 101 College Street, Suite 800, Toronto, Ontario, Canada M5G 1L7; The University of Ottawa Faculty of Medicine, 101 Smythe Road, Ottawa, Ontario, Canada K1H 8L6.

Centre for Research in Environmental Epidemiology (CREAL), Municipal Institute of Medical Research (IMIM-Hospital del Mar) and CIBER Epidemiología y Salud Pública (CIBERESP), Doctor Aiguader, 88 E-08003 Barcelona, Spain.