TOPIC HIGHLIGHT

WJG 20th Anniversary Special Issues (5): Colorectal cancer

New genes emerging for colorectal cancer predisposition

Clara Esteban-Jurado, Pilar Garre, Maria Vila, Juan José Lozano, Anna Pristoupilova, Sergi Beltrán, Anna Abulí, Jenifer Muñoz, Francesc Balaguer, Teresa Ocaña, Antoni Castells, Josep M Piqué, Angel Carracedo, Clara Ruiz-Ponte, Xavier Bessa, Montserrat Andreu, Luis Bujanda, Trinidad Caldés, Sergi Castellví-Bel

Clara Esteban-Jurado, Anna Abulí, Jenifer Muñoz, Francesc Balaguer, Teresa Ocaña, Antoni Castells, Josep M Piqué, Sergi Castellví-Bel, Department of Gastroenterology, Hospital Clínic, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, 08036 Barcelona, Catalonia, Spain
Anna Abulí, Xavier Bessa, Montserrat Andreu, Department of Gastroenterology, Hospital del Mar-IMIM (Hospital del Mar Medical Research Centre), Pompeu Fabra University, 08003 Barcelona, Spain
Maria Vila, Juan José Lozano, Bioinformatics platform, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), 08036 Barcelona, Spain
Anna Pristoupilova, Sergi Beltrán, Centre Nacional d'Anàlisi Genòmica (CNAG), Parc Científic de Barcelona, 08028 Barcelona, Spain
Angel Carracedo, Clara Ruiz-Ponte, Galician Public Foundation of Genomic Medicine (FPGMX), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Genomics Medicine Group, Hospital Clínico, Santiago de Compostela, University of Santiago de Compostela, 15706 Galicia, Spain
Angel Carracedo, Center of Excellence in Genomic Medicine Research, King Abdulaziz University, 21589 Jeddah, Kingdom of Saudi Arabia
Luis Bujanda, Gastroenterology Department, Hospital Donostia, Networked Biomedical Research Centre for Hepatic and Digestive Diseases (CIBEREHD), Basque Country University, 20080 San Sebastián, Spain
Pilar Garre, Trinidad Caldés, Molecular Oncology Laboratory, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), 28040 Madrid, Spain

Abstract

Colorectal cancer (CRC) is one of the most frequent neoplasms and an important cause of mortality in the developed world. This cancer is caused by both genetic and environmental factors although 35% of the variation in CRC susceptibility involves inherited genetic differences. Mendelian syndromes account for about 5% of the total burden of CRC, with Lynch syndrome and familial adenomatous polyposis the most common forms. Excluding hereditary forms, there is an important fraction of CRC cases that present familial aggregation for the disease with an unknown germline genetic cause. CRC can be also considered as a complex disease taking into account the common disease-common variant hypothesis with a polygenic model of inheritance where the genetic components of common complex diseases correspond mostly to variants of low/moderate effect. So far, 30 common, low-penetrance susceptibility variants have been identified for CRC. Recently, new sequencing technologies including exome- and whole-genome sequencing have permitted to add

a new approach to facilitate the identification of new genes responsible for human disease predisposition. By using whole-genome sequencing, germline mutations in the *POLE* and *POLD1* genes have been found to be responsible for a new form of CRC genetic predisposition called polymerase proofreading-associated polyposis.

Core tip: Colorectal cancer (CRC) is caused by both genetic and environmental factors although 35% of the variation in CRC susceptibility involves inherited genetic differences. Mendelian syndromes account for about 5% of the total burden of CRC. Excluding hereditary forms, there is an important fraction of CRC cases that present familial aggregation for the disease with an unknown germline genetic cause. Recently, new sequencing technologies have permitted to add a new approach to identify new genes responsible for human disease predisposition. By doing so, germline mutations in the *POLE* and *POLD1* genes have been found to be responsible for a new form of CRC genetic predisposition.

INTRODUCTION

Colorectal cancer (CRC) is one of the most frequent neoplasms and an important cause of mortality in the developed world. Approximately 5% of the population develops CRC and this figure is expected to rise as life expectancy increases[1]. For 2015, approximately 473200 new cases are predicted and 233900 individuals will die from this disease in Europe[2]. When taking into account both genders together, it corresponds to the most frequent neoplasm in Spain. Although there has been recent progress in CRC clinical management and treatment that has permitted to reduce the number of cases in the developed countries, it is foreseen that its incidence will increase worldwide with developing nations bearing the brunt of the rise. The incidence of CRC varies widely between countries, depending on their degree of development and also on the quality of their cancer registries[3]. Around 60% of cases are diagnosed in the developed world[3]. The highest incidence rates are found in Australia and New Zealand, North America and Europe, whereas the lowest rates are registered in Africa and South-Central Asia[2] (Figure 1).

CRC survival depends on the stage of disease at diagnosis and typically ranges from a 90% 5-year survival rate for cancers detected at the localized stage to 10% for people diagnosed of a distant metastatic cancer[4]. The lifetime risk of CRC in the general population is about 5% in Western countries, but the likelihood of CRC diagnosis increases progressively with age, being more than 90% in individuals over age 50, and 70% of these over 65[4].

CRC is believed to develop from polyps, which have been traditionally classified as either hyperplastic or adenomatous. Until recently, according to the adenoma-carcinoma sequence proposed by Vogelstein *et al*[5] the adenoma was considered the exclusive precursor lesion while hypeplastic polyps were deemed to have no malignant potential. However, it is now recognized that lesions, formerly classified as hyperplastic, represent a heterogeneous group of polyps with a characteristic serrated morphology, some of which have a significant risk of malignant transformation through the serrated neoplasia pathway[4].

GENETIC AND ENVIRONMENTAL RISK FACTORS

As other complex diseases, CRC is caused by both genetic and environmental factors. The role of environmental factors on colorectal carcinogenesis is indicated by the increase in CRC incidence in parallel with economic development and adoption of Western diets and lifestyles, responsible for the high incidence of CRC in industrialized countries[7]. Although the majority of CRC occur mostly in industrialized countries, their incidence rates are rapidly rising in economically transitioning countries in the world[8]. These observations highlight the importance of environmental influences on CRC development and suggest that Western lifestyle risk factors play an important role in the etiology of the disease. However, although environmental causes such as smoking and diet are undoubtedly risk factors for CRC, twin studies have shown that 35% of the variation in CRC susceptibility involves inherited genetic differences[9,10]. In that sense, a minority of CRC cases (about 5%) show strong familial aggregation and belong to the well-known hereditary CRC forms mainly caused by germline mutations in *APC*, *MUTYH* and the DNA mismatch repair genes[11]. Approximately 30% of CRC cases show some family history of the disease but do not fit in the previous category and are regarded as familial CRC, whereas a majority of cases do not show any familial aggregation and correspond to sporadic CRC. For instance, familial CRC accounted for about 30% of all CRC cases in an
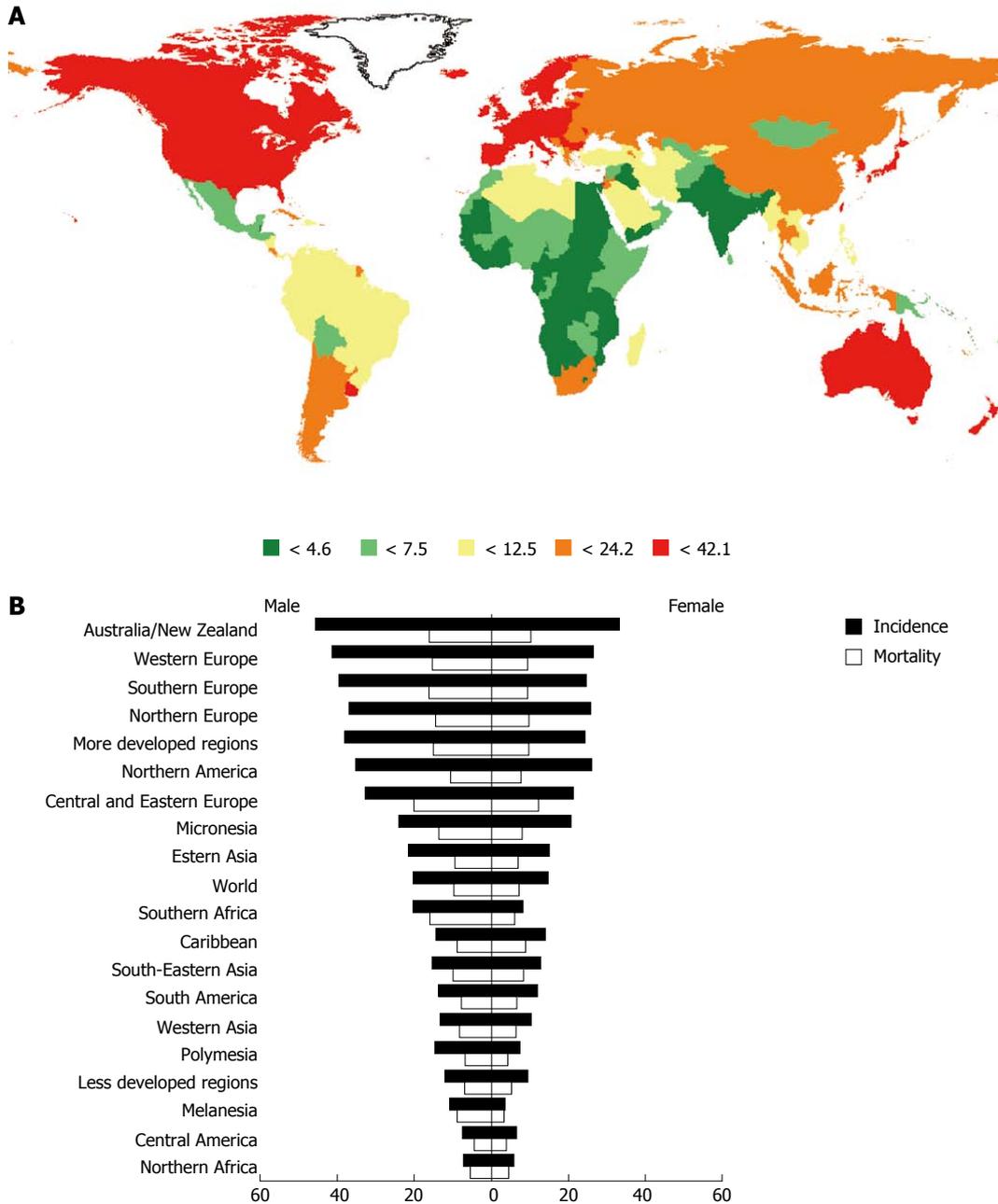
A

■ < 4.6	■ < 7.5	■ < 12.5	■ < 24.2	■ < 42.1

B

Figure 1 Colorectal cancer in the world. A: Estimated age-standarized incidence rate per 100000 individuals (both genders and all ages); B: Estimated age-standarized incidence and mortality rate per 100000 individuals by genders (data adapted from Ferlay *et al*[2]).

epidemiological study in the Spanish population[12].

HEREDITARY CRC

Mendelian cancer syndromes account for about 5% of the total burden of CRC[11]. The genetic components involved in these less frequent hereditary forms were successfully identified using linkage analysis in the past two decades and they correspond to rare highly penetrant alleles that predispose to CRC. Two major subgroups can be clinically divided on the presence or absence of colorectal polyposis. An overview of all CRC syndromes is provided in Table 1. The most frequent forms are hereditary nonpolyposis colorectal cancer and familial

polyposis syndrome, which are further described below.

Hereditary Nonpolyposis Colorectal Cancer (HNPCC; MIM No.120435), also known as Lynch syndrome, is the most common form of hereditary CRC accounting for at least 3% of all CRC. HPNCC is an autosomal dominant syndrome defined clinically by the Amsterdam criteria (Table 2), which are used in clinical practice to identify individuals at risk for this disease who require further evaluation and are based on strong familial aggregation and early onset. It is characterized by early-onset CRC (mean age at diagnosis, approximately 45 years), excess synchronous and metachronous colorectal neoplasms and right-sided predominance compared to sporadic neoplasms. In addition, there is an increased

Table 1 Hereditary colorectal cancer genes				
	Gene	Chromosome	Mendelian pattern	Function
Familial adenomatous polyposis	*APC*	5q	AD	Regulation of canonical Wnt signaling pathway
	MUTYH	1p	AR	Base-excision repair
Hereditary non-polyposis CRC (Lynch syndrome)	*MLH1*	3p	AD	Mismatch repair
	MSH2	2p	AD	Mismatch repair
	MSH6	2p	AD	Mismatch repair
	PMS2	7p	AD	Mismatch repair
Peutz-Jeghers	*LKB1*	19p	AD	Regulation of Wnt signaling pathway
Juvenile polyposis	*SMAD4*	18q	AD	TGFBR signaling pathway
	BMPR1A	10q	AD	TGFBR signaling pathway
Cowden's disease	*PTEN*	10q	AD	Negative regulation of PI3K signaling
	KLLN	10q	AD	Apoptotic process

CRC: Colorectal cancer; AD: Autosomal dominant; AR: Autosomal recessive; TGFBR: Transforming growth factor beta receptor; PI3K: Phosphatidylinositol 3-kinase.

Table 2 Amsterdam criteria in Lynch syndrome	
Amsterdam criteria I	**Amsterdam criteria II**
At least three relatives with CRC; all of the following must be met:	At least three relatives with colorectal, endometrial, small bowel, ureter, or renal pelvis cancer; all of the following must be met:
One affected individual is a first degree relative of the other two	One affected individual is a first degree relative of the other two
At least two successive generations affected	At least two successive generations affected
At least one CRC diagnosed before the age of 50 years	At least one tumor diagnosed before the age of 50 years
Familial adenomatous polyposis has been excluded	Familial adenomatous polyposis has been excluded

CRC: Colorectal cancer.

incidence of extracolonic neoplasms (endometrial, small bowel, gastric, upper urinary tract, ovarian, brain and pancreatic tumors) being endometrial cancer the most common malignancy associated with Lynch syndrome. Indeed, Lynch syndrome is responsible for approximately 2% of all endometrial cancers[13]. The lifetime risk for developing CRC in individuals affected with Lynch syndrome have been estimated in approximately 66% for men and about 43% for women. The cumulative risk of endometrial cancer is approximately 40% and the lifetime risk of endometrial cancer or CRC in women is approximately 73%[14]. Lynch syndrome tumors develop as a consequence of defective DNA mismatch repair (MMR) associated with germline mutations in the MMR genes, including *MSH2* on chromosome 2p16, *MLH1* on chromosome 3p21, *MSH6* on chromosme 2p16, and *PMS2* on chromosome 7q11. In addition, germline epigenetic inactivation of *MLH1*, by hypermethylation of its promoter, can also lead to Lynch syndrome[15]. Recently, germline deletions of the 3' region of *EPCAM* gene were found in a subset of families with Lynch syndrome. This deletion leads to promoter hypermetilation of *MSH2*, located upstream of the deleted gene[16]. The MMR system is necessary to maintaining genomic fidelity by correcting single-base mismatches and insertion-deletion loops during DNA replication. As a consequence, Lynch syndrome tumors accumulate errors in short repetitive sequences, a phenomenon called microsatellite instability (MSI), which is considered a landmark for this disease. It is noteworthy to mention that in sporadic MSI CRC cancers, loss of expression of MLH1 due to hypermethilation of its promoter is a frequent event, and it is linked with the somatic mutation V600E in the *BRAF* gene[17].

Familial Adenomatous Polyposis (FAP; MIM No.175100) is the most common polyposis syndrome, classically characterized by the development of hundreds to thousands of adenomatous polyps in the rectum and colon. FAP is an autosomal dominant disease and accounts for approximately 1% of all CRC cases. In the majority of patients polyps begin to develop during the second decade of life and nearly 100% of untreated patients will have malignancy by ages 40-50 years. Individuals with FAP can also develop a variety of extracolonic manifestations, including cutaneous lesions such as fibromas, lipomas, sebaceous and epidermoid cysts, facial osteomas, congenital hypertrophy of the retinal pigment epithelium, desmoid tumours and extracolonic cancers (tyroid, liver, biliary tract and central nervous system)[18]. Duodenal cancer is the second most common malignancy in FAP, with a lifetime risk of approximately 4%-12%. *Adenomatous polyps* are also found in the stomach and duodenum, especially the *periampullary area* and can develop into adenocarcinomas. After colectomy, periampullary carcinoma is the most common malignancy, occurring in approximately 5%-6% of the patients[19]. Some lesions such as skull and mandible osteomas, dental abnormalities and fibromas are indicative of the Gardner syndrome, a clinical variant of FAP where the extracolonic features are prominent. FAP is caused by germline mutations in the *APC* gene on

chromosome 5q22, which encodes a tumor suppressor protein that plays an important role in the *Wnt* signaling pathway. Most patients have a family history of colorectal polyps and cancer, but de novo *APC* mutations are responsible for approximately 25% of cases[11].

APPROACHES TO IDENTIFY GENETIC VARIANTS FOR CRC RISK

Among CRC cases of unknown inherited cause, there are large families with a clear positive family history of CRC, which are likely caused by highly penetrant risk loci. In the last few years, it has been described that approximately 40%-50% of CRC families that fulfill the Amsterdam Criteria for Lynch syndrome do not show evidence of MMR deficiency. Studying relatives in such families showed that CRC risk is lower than in those families with Lynch syndrome, that CRC diagnosis is in average 10 years later and that there is no increased incidence of extracolonic malignancies[20,21]. The designation of Familial CRC type X was proposed to describe this type of CRC clustering[20]. Meanwhile, genes responsible for this new entity are unknown, and most patients are included in the heterogeneous group of non-syndromic familial CRC.

Recently, there have been several efforts to identify additional genetic factors that predispose to CRC with uneven success. Linkage analysis in affected families were able to pinpoint chromosomal regions of interest such as 9q22 and 3q22 but no clear CRC predisposition genes were identified after screening for interesting candidates within these areas[22,23].

Since the known high-risk syndromes only account for a small minority of CRC cases, there has been an intensified search for low-penetrance genetic variants that probably underlie part of the hereditary predisposition and together with environmental interactions are responsible for CRC as a complex disease. Therefore, the common disease-common variant hypothesis has been also considered, being a polygenic model of inheritance where the genetic components of common complex diseases correspond mostly to variants of low/moderate effect (typically < 1.5-fold increased risk) that appeared at an elevated frequency in the population (> 5%), each exerting a small influence on disease risk. In this regard, case-control genome-wide association studies (GWAS) have been more successful by discovering up to now 31 common, low-penetrance genetic variants involved in CRC susceptibility[24-32] (Table 3).

OVERVIEW OF NEW SEQUENCING TECHNOLOGIES

Until recently, the Sanger method was the dominant approach and gold standard for DNA sequencing[33]. Next generation sequencing (NGS), also called massive

parallel sequencing, is based in sequencing millions of DNA fragments at the same time[34]. It consists in a mix of techniques of DNA shearing, PCR amplification and sequencing through modified nucleotides attached to a reversible terminator and a fluorophore, which permits fluorescent detection with an imaging system. Once the fragments are sequenced, they are assembled de novo or aligned with a reference genome by bioinformatics tools and positions that differ are designated as variants. Variants are annotated assigning their position in a gene, retrieving frequency information from genetic variation databases and categorizing them by their functional class (nonsense, missense, synonymous, frameshift, splicing, intronic, untranslated regions, regulatory).

The advantage of NGS comparing with conventional Sanger sequencing is that millions of DNA fragments are sequenced at the same time which permits to have an entire human genome sequenced in few days, and the cost is greatly reduced. However, data analysis that includes filtering of the false positives and prioritization of the candidate variants for the studied phenotypic condition is the main bottleneck of NGS, being time consuming and requiring different strategies that will be discussed later. Another disadvantage of NGS is that PCR amplification and sequencing reaction steps systematically introduce mistakes, producing base-calling errors and shorter sequenced fragments that difficult the mapability to the reference sequence. Due to recent technology and variant calling algorithm improvements, NGS is probably nowadays more accurate than conventional Sanger sequencing[35]. However, although there is a very small error rate associated with NGS, a huge amount of false positives are still detected since millions of variants are sequenced per genome[36]. Thus, after data analysis and selection of the candidate variants it is necessary to validate them using a technology with a different systematic error associated, such as conventional Sanger sequencing, which increases the costs and time of the analysis.

In order to detect genomic sequence variation by NGS, it is possible to sequence the entire genome (whole-genome sequencing, WGS) or capture and sequence only specific regions of interest (targeted enrichment). The most commonly used application for NGS target enrichment in the human genome is whole-exome sequencing (WES) that captures and amplifies the entire protein coding sequence (1% genome), flanking intronic regions and some noncoding RNAs[37]. It is a cost effective approach for detecting rare high penetrance variants based on the fact that for Mendelian disorders over the 85% of causative mutations are in coding regions. One advantage of WES is that is about much cheaper than WGS, which allows sequencing a larger number of samples with better accuracy or coverage. The term coverage corresponds to the read depth or depth and it is the average number of times that a nucleotide has been sequenced in a different sequencing read. Also, the data analysis pipelines are simpler in WES than for

SNP	Region	Gene	Sample size (cases/controls)	Effect size OR (95%CI)
rs693267	8q24.21	*MYC*	8264/6206	1.21 (1.15-1.27)
rs4939827	18q21.1	*SMAD7*	8413/6949	1.18 (1.12-1.23)
rs16892766	8q23.3	*EIF3H*	18831/18540	1.25 (1.19-1.32)
rs3802842	11q23.1	*?*	14500/13294	1.12 (1.07-1.17)
rs4779584	15q13.3	*GREM1*	7922/6741	1.26 (1.19-1.34)
rs10795668	10p14	*?*	18831/18540	1.12 (1.10-1.16)
rs4444235	14q22.2	*BMP4*	20288/20971	1.11 (1.08-1.15)
rs9929218	16q22.1	*CDH1*	20288/20971	1.10 (1.06-1.12)
rs10411210	19q13	*RHNP2*	20288/20971	1.15 (1.10-1.20)
rs961253	20p12.3	*BMP2*	20288/20971	1.12 (1.08-1.16)
rs6691170	1q41	*DUSP10*	18185/20197	1.06 (1.03-1.09)
rs10936599	3q26.2	*TERC*	18185/20197	0.93 (0.91-0.96)
rs11169552	12q13.3	*?*	18185/20197	0.92 (0.90-0.95)
rs4925386	20q13.33	*LAMA5*	18,185/20,197	0.93 (0.91-0.95)
rs1957636	14q22.2	*BMP4*	24910/26275	1.08 (1.05-1.11)
rs4813802	20p12.3	*BMP2*	24910/26275	1.09 (1.06-1.12)
rs2736100	5p15.33	*TERT*	16039/16430	1.07 (1.04-1.10)
rs1321311	6p21	*CDKN1A*	21096/19555	1.10 (1.07-1.13)
rs3824999	11q13.4	*POLD3*	21096/19555	1.10 (1.07-1.13)
rs5934683	Xp22.2	*SHROOM2*	21096/19555	1.07 (1.04-1.10)
rs12080929	1p33	*SLC5A9*	2317/2447	0.86 (0.78-0.95)
rs11987193	8p12	*DUSP4*	2317/2447	0.78 (0.70-0.87)
rs10774214	12p13.32	*CCND2*	11870/14190	1.04 (1.00-1.09)
rs647161	5q31.1	*PITX1*	11870/14190	1.07 (1.02-1.11)
rs2423279	20p12.3	*HAQ1*	11870/14190	1.07 (1.03-1.12)
rs11903757	2q32.3	*NABP1*	15752/21771	1.16 (1.10-1.22)
rs10911251	1q25.3	*LAMC1*	15752/21771	1.09 (1.06-1.13)
rs3217810	12p13.32	*CCND2*	13654/16022	1.20 (1.12-1.28)
rs3217901	12p13.32	*CCND2*	15752/21771	1.10 (1.06-1.14)
rs59336	12q24.21	*TBX3*	15752/21771	1.09 (1.06-1.13)

Table 3 Genetic variants associated with colorectal cancer susceptibility identified by genome-wide association studies (as for September 2013)

WGS. However, WES need for larger amounts of DNA sample and only covering coding variants are among the shortcomings for this technique. It is noteworthy mentioning that NGS target enrichment can also be used to sequence a panel of known genes for clinical diagnosis[33] or regions of linkage disequilibrium for a disease.

The election of individuals to sequence is a critical process to take into account for further analysis and will depend of the disease phenotype and pattern of genetic inheritance. Also, it should be noted that is possible to obtain good results with NGS when using carefully selected patients in contrast to GWAS, where number of cases and controls that are compared needs to be much higher in order to obtain statistically significant findings. For diseases with genetic heterogeneity as human cancers, different strategies can be used including the selection of families with strong disease aggregation or sequencing sporadic cases with early onset for the disease. Both situations are suggestive of the involvement of a germline predisposition. When focusing in families with several affected members, sequencing can be performed in several cases in each family and only those shared variants will be taken into account. On the other hand, if sporadic early-onset cases are chosen, genes with variants in different individuals can be selected. Sequencing non affected individuals of the same family can be useful to discard the variants shared with patients, as long as the disease has complete penetrance or it is quite likely that the non affected individuals will not express the disease in their lifetime.

Data filtering and prioritization in NGS
Based on several recently sequenced individual genomes a pattern has been recognized that, in general, approximately 3-4 million variants are expected to be found in a human genome by WGS[38] and 20000 single nucleotide variants are to be found in a human exome by WES[39], so it is necessary to do a filtering strategy in order to eliminate as many false positives as possible. The first filter to apply is for those variants that do not pass a coverage threshold (typically 5-10x).

The second filtering process is based on the kind of inheritance, penetrance and frequency of the disease. Regarding the inheritance, for monogenic diseases where unrelated affected individuals have been sequenced, it is necessary to select only the genes that have variants in all of them. If a disease with genetic heterogeneity is studied, variants shared between the affected members of the same family and not shared by the unaffected ones will be chosen. Also, if dominant inheritance is present heterozygous mutations will be expected, whereas homozygous or compound heterozygous mutations will be selected in the case of recessive inheritance. However, variants in the non pseudoautosomal regions of X chromosome for dominant inheritance have to take into account also. In men, they will be annotated as

homozygous and it is necessary to select these variants too and not filter them out. Regarding variant effect on protein, it is assumed that high penetrance mutations are causative of Mendelian disorders with a large effect on protein function. Therefore, a positive selection for variants with a strong effect on the protein is advised including those affecting canonical splice sites, as well as frameshift, nonsense and missense mutations.

Proportionally, more deleterious than polymorphic variants are expected to be rare so a causative mutation is not expected to be present at a high frequency in the general population[40]. Thus, variants present at high frequency at reference genetic variation databases can be removed as potential candidates to be causative mutations.

However, many variants can still remain for each individual as putative causative mutations for the disease after filtering. A logical approach to reduce the number of candidate variants is to prioritize the mutations in genes previously implicated with the studied disease. Also, since the protein products of genes responsible for the same disorder tend to physically interact with each other so as to carry out certain biological functions, another approach for the prioritization strategy will be to include genes interacting with those previously implicated with the studied disease[41]. Finally, knowledge of the pathways implicated in a disease can be helpful also to prioritize those genes related with those pathways. After filtering and prioritization, a list of candidate variants will be available.

Sequencing validation by Sanger sequencing or any other PCR technology designed to detect a specific nucleotide change is necessary after NGS to confirm the prioritized variants and exclude sequencing artifacts. Also, segregation analysis in families permits to check if a candidate variant segregates correctly with the disease. Therefore, affected members need to be carriers and non-affected individuals old enough to be expressing the disease should be non-carriers in order to find correct segregation of the candidate variant with the studied disease. Additionally in the case of hereditary cancer, when heterozygous candidate variants with correct segregation are identified, it is necessary to confirm if there is loss of the second allele in the tumor DNA in order to establish the candidate gene as a tumor suppressor gene. Case-control screening studies can also be performed in order to identify additional carriers of the candidate variants in ample disease cohorts and further demonstrate its absence in controls. Finally, functional assessment of the candidate variant and affected gene will be also necessary to further confirm the negative effect of the variant in the protein and prove its involvement in disease development by *in vitro* studies and animal models.

NEW GENES IDENTIFIED FOR CRC GENETIC PREDISPOSITION

New sequencing technologies made available recently including exome- and whole-genome sequencing have permitted to add a new approach to facilitate the identification of new genes responsible for human disease predisposition. Indeed, some seminal efforts have been already completed very recently for CRC. However, before these high-throughput technologies have yielded results in CRC families, some previous low-throughput sequencing studies reported directed screening of some plausible gene candidates for various reasons. Most studies have not been replicated in additional cohorts and, therefore, there is a strong need to further validate them before considering these genes as hereditary CRC genes perse.

A truncating mutation was found in the *CDH1* gene in a family with predisposition to CRC and gastric cancer, suggesting that germline mutations in this gene could contribute to early onset CRC and gastric cancer[42]. Later on, the *AXIN2* gene, a component of the Wnt signaling, was found to be mutated in a Finnish family with severe permanent tooth agenesis and CRC[43]. In a subsequent study in patients with unexplained hamartomatous or hyperplastic/mixed polyposis, two early-onset disease patients were found to have germline mutations in *ENG*, encoding endoglin, previously associated only with hereditary hemorrhagic telangiectasia[44]. This study suggested *ENG* as a new predisposition gene for juvenile polyposis, however this gene was found to be mutated in an additional study only in patients with ≥ 5 cumulative lifetime gastrointestinal polyps but not in juvenile polyposis[45]. *EPHB2* was also evaluated as a candidate tumor suppressor gene for CRC and found mutated in 3 out 116 population-based familial CRC cases, suggesting this gene may contribute to a small fraction of hereditary CRC[46]. In 2009, the *GALNT12* gene was also found mutated in the germline of 6 CRC patients[47]. This gene encodes one of the proteins involved in mucin type O-linked glycosylation and it is located in chromosomal region 9q22, previously involved in familial CRC. A more recent study detected additional deleterious variants in this gene reinforcing its role as a new candidate gene for hereditary CRC[48]. Also, an inherited duplication affecting the protein tyrosine phosphatase *PTPRJ* and causing epigenetic silencing of this gene was detected in a CRC family without polyposis and MMR alteration, being indicative of its contribution to a fraction of hereditary CRC with unknown basis[49]. Afterwards, *BMP4*, a gene close to 2 of the CRC genetic susceptibility variants identified by GWAS, was also screened in 504 genetically enriched CRC and 3 pathogenic mutations were identified[50]. Then, it could be plausible that some genes identified by CRC GWAS could be also involved in hereditary CRC. In 2011, the *BMPR1A* gene, previously involved in juveline polyposis and mixed polyposis germline predisposition, was also found mutated in familial CRC type X cases, expanding its phenotype also to this CRC hereditary form[51]. Finally, Cowden syndrome individuals without germline *PTEN* mutations were found to carry germline mutations in *PIK3CA* and *AKT1*, expanding the genetic spec-

trum of this hereditary CRC condition[52].

Regarding NGS studies to identify new CRC predisposition genes, Palles *et al*[53] reported very recently the identification of germline mutations in the *POLE* (polymerase (DNA directed), epsilon, catalytic subunit) and *POLD1* (polymerase (DNA directed), delta 1, catalytic subunit) genes in individuals with multiple colorectal adenomas, carcinoma or both, using whole-genome sequencing[54]. *POLE* and *POLD1* encode the catalytic and proofreading activities of the leading-strand DNA polymerase ε and the lagging-strand polymerase δ. The proofreading capacity of the exonuclease domain is essential for the maintenance of replication fidelity and may act not only on newly misincorporated bases but also on mismatches produced by non-proofreading polymerases. They identified a heterozygous p.Leu424Val missense variant in *POLE* DNA polymerase in a family affected with adenomas and CRC and a p.Ser478Asn missense variant in *POLD1* in a second family with CRC. The same *POLD1* p.Ser478Asn variant was also identified in the affected members of an independent family. These findings were further validated in a screen of 3,085 individuals with CRC, enriched for a family history of colorectal tumors, in which they detected 12 individuals with the p.Leu424Val variant in *POLE* and one additional individual with the pSer478Asn in *POLD1*. Functional assessment supported the importance of these mutations in *POLE* and *POLD1*. Mutagenesis studies of Polδ and Pol3 in yeast showed that the mutation of the equivalent residue produces a mutator phenotype and loss of the proofreading activity of the protein[53,55,56]. Also, mice expressing proofreading-impaired Pole and Pold1 in a homozygous state developed spontaneous intestinal adenocarcinomas or a spectrum of cancers[57]. Thus, germline variants in *POLE* and *POLD1* predispose to individuals to either a multiple colorectal adenoma phenotype similar to that observed in MUTYH-associated polyposis or a HNPCC phenotype, in which carriers develop early-onset CRC. Although additional studies will be needed to evaluate these rare germline variants in *POLD1* and *POLE* and their associated phenotypes, the authors suggest that screening for these variants should be considered in patients with an unexplained personal or family history of multiple adenomas, early onset CRC or both. On the other hand, carriers are potential candidates for regular and frequent colonoscopic surveillance starting at an early age.

Two additional reports using exome sequencing have also been published very recently but their results are not as solid as those for the polymerase genes previously mentioned. A cohort of 50 sporadic CRC patients was sequenced including 18 early-onset cases with a relatively low coverage in the first study[58]. Variants were biased selected when found in a list of 1,138 genes likely to play a role in CRC. Further selection to include only those genes undergoing bialleic inactivation yielded *FANCM*, *LAMB4*, *PTCHD3*, *LAMC3* and *TREX2* as potential tumor suppressor candidates. In the second study, exome sequencing was completed for 40 familial cases from 16 families by selecting distant relatives to decrease the number of shared, non-predisposition variants[59]. Data was analyzed firstly by an agnostic search for CRC predisposition genes not taking into account a biased list of candidates, and secondly by selecting genes previously involved in CRC predisposition or within CRC linkage regions. Two missense variants in the *CENPE* and *KIF23* genes that complied with family segregation and belong to regions on chromosomes 1 and 15 formerly linked to CRC were considered the more plausible candidates for CRC predisposition but additional studies are needed to further elucidate their role.

CONCLUSION

CRC is one of the most frequent neoplasms and an important cause of mortality in the developed world. CRC is caused by both genetic and environmental factors although 35% of the variation in CRC susceptibility involves inherited genetic differences. Mendelian cancer syndromes account for about 5% of the total burden of CRC, being Lynch syndrome and familial adenomatous polyposis the most common forms. Familial CRC type X is an example of CRC with unknown inherited cause. A clear positive family history of CRC is present (Amsterdam criteria for Lynch syndrome are fulfilled) although MMR is proficient. When considering CRC as a complex disease, low-penetrance genetic variants probably underlie part of the hereditary predisposition together with environmental interactions. So far, 30 susceptibility variants have been identified for CRC. New sequencing technologies made available recently including exome- and whole-genome sequencing have permitted to add a new approach to facilitate the identification of new genes responsible for human disease predisposition. Germline mutations in the *POLE* and *POLD1* genes are responsible for a new form of CRC genetic predisposition called polymerase proofreading-associated polyposis.

ACKNOWLEDGMENTS

REFERENCES

1   **Cohen AM**, Shank B, Friedman MA. Colorectal cancer. In: De Vita V, Hellman S, Rosenberg S (eds): Cancer: Principles and Practice of Oncology. Philadelphia: Lippincott, 1989: 895

2   **Ferlay J**, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010; **127**: 2893-2917 [PMID: 21351269 DOI: 10.1002/ijc.25516]

3   **Jemal A**, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011; **61**: 69-90 [PMID: 21296855 DOI: 10.3322/caac.20107]

4   **Parker SL**, Tong T, Bolden S, Wingo PA. Cancer statistics, 1996. *CA Cancer J Clin* 1996; **46**: 5-27 [PMID: 8548526]

5   **Vogelstein B**, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM, Bos JL. Genetic alterations during colorectal-tumor development. *N Engl J Med* 1988; **319**: 525-532 [PMID: 2841597]

6   **Leggett B**, Whitehall V. Role of the serrated pathway in colorectal cancer pathogenesis. *Gastroenterology* 2010; **138**: 2088-2100 [PMID: 20420948 DOI: 10.1053/j.gastro.2009.12.066]

7   **Gingras D**, Béliveau R. Colorectal cancer prevention through dietary and lifestyle modifications. *Cancer Microenviron* 2011; **4**: 133-139 [PMID: 21909875 DOI: 10.1007/s12307-010-0060-5]

8   **O'Callaghan T**. Introduction: The prevention agenda. *Nature* 2011; **471**: S2-S4 [PMID: 21430716 DOI: 10.1038/471S2a]

9   **Lichtenstein P**, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skythe A, Hemminki K. Environmental and heritable factors in the causation of cancer-analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000; **343**: 78-85 [PMID: 10891514]

10   **Hemminki K**, Chen B. Familial risk for colorectal cancers are mainly due to heritable causes. *Cancer Epidemiol Biomarkers Prev* 2004; **13**: 1253-1256 [PMID: 15247139]

11   **Jasperson KW**, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. *Gastroenterology* 2010; **138**: 2044-2058 [PMID: 20420945 DOI: 10.1053/j.gastro.2010.01.054]

12   **Piñol V**, Castells A, Andreu M, Castellví-Bel S, Alenda C, Llor X, Xicola RM, Rodríguez-Moranta F, Payá A, Jover R, Bessa X. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. *JAMA* 2005; **293**: 1986-1994 [PMID: 15855432]

13   **Hampel H**, Frankel W, Panescu J, Lockman J, Sotamaa K, Fix D, Comeras I, La Jeunesse J, Nakagawa H, Westman JA, Prior TW, Clendenning M, Penzone P, Lombardi J, Dunn P, Cohn DE, Copeland L, Eaton L, Fowler J, Lewandowski G, Vaccarello L, Bell J, Reid G, de la Chapelle A. Screening for Lynch syndrome (hereditary nonpolyposis colorectal cancer) among endometrial cancer patients. *Cancer Res* 2006; **66**: 7810-7817 [PMID: 16885385]

14   **Stoffel E**, Mukherjee B, Raymond VM, Tayob N, Kastrinos F, Sparr J, Wang F, Bandipalliam P, Syngal S, Gruber SB. Calculation of risk of colorectal and endometrial cancer among patients with Lynch syndrome. *Gastroenterology* 2009; **137**: 1621-1627 [PMID: 19622357 DOI: 10.1053/j.gastro.2009.07.039]

15   **Suter CM**, Martin DI, Ward RL. Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet* 2004; **36**: 497-501 [PMID: 15064764]

16   **Ligtenberg MJ**, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, Lee TY, Bodmer D, Hoenselaar E, Hendriks-Cornelissen SJ, Tsui WY, Kong CK, Brunner HG, van Kessel AG, Yuen ST, van Krieken JH, Leung SY, Hoogerbrugge N. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat Genet* 2009; **41**: 112-117 [PMID: 19098912 DOI: 10.1038/ng.283]

17   **McGivern A**, Wynter CV, Whitehall VL, Kambara T, Spring KJ, Walsh MD, Barker MA, Arnold S, Simms LA, Leggett BA, Young J, Jass JR. Promoter hypermethylation frequency and BRAF mutations distinguish hereditary non-polyposis colon cancer from sporadic MSI-H colon cancer. *Fam Cancer* 2004; **3**: 101-107 [PMID: 15340260]

18   **Half E**, Bercovich D, Rozen P. Familial adenomatous polyposis. *Orphanet J Rare Dis* 2009; **4**: 22 [PMID: 19822006 DOI: 10.1186/1750-1172-4-22]

19   **Bülow S**, Björk J, Christensen IJ, Fausa O, Järvinen H, Moesgaard F, Vasen HF. Duodenal adenomatosis in familial adenomatous polyposis. *Gut* 2004; **53**: 381-386 [PMID: 14960520]

20   **Lindor NM**, Rabe K, Petersen GM, Haile R, Casey G, Baron J, Gallinger S, Bapat B, Aronson M, Hopper J, Jass J, LeMarchand L, Grove J, Potter J, Newcomb P, Terdiman JP, Conrad P, Moslein G, Goldberg R, Ziogas A, Anton-Culver H, de Andrade M, Siegmund K, Thibodeau SN, Boardman LA, Seminara D. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. *JAMA* 2005; **293**: 1979-1985 [PMID: 15855431]

21   **Mueller-Koch Y**, Vogelsang H, Kopp R, Lohse P, Keller G, Aust D, Muders M, Gross M, Daum J, Schiemann U, Grabowski M, Scholz M, Kerker B, Becker I, Henke G, Holinski-Feder E. Hereditary non-polyposis colorectal cancer: clinical and molecular evidence for a new entity of hereditary colorectal cancer. *Gut* 2005; **54**: 1733-1740 [PMID: 15955785]

22   **Wiesner GL**, Daley D, Lewis S, Ticknor C, Platzer P, Lutterbaugh J, MacMillen M, Baliner B, Willis J, Elston RC, Markowitz SD. A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. *Proc Natl Acad Sci USA* 2003; **100**: 12961-12965 [PMID: 14566058]

23   **Kemp Z**, Carvajal-Carmona L, Spain S, Barclay E, Gorman M, Martin L, Jaeger E, Brooks N, Bishop DT, Thomas H, Tomlinson I, Papaemmanuil E, Webb E, Sellick GS, Wood W, Evans G, Lucassen A, Maher ER, Houlston RS. Evidence for a colorectal cancer susceptibility locus on chromosome 3q21-q24 from a high-density SNP genome-wide linkage scan. *Hum Mol Genet* 2006; **15**: 2903-2910 [PMID: 16923799]

24   **Tenesa A**, Dunlop MG. New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nat Rev Genet* 2009; **10**: 353-358 [PMID: 19434079 DOI: 10.1038/nrg2574]

25   **Fernandez-Rozadilla C**, Cazier JB, Tomlinson IP, Carvajal-Carmona LG, Palles C, Lamas MJ, Baiget M, López-Fernández LA, Brea-Fernández A, Abulí A, Bujanda L, Clofent J, Gonzalez D, Xicola R, Andreu M, Bessa X, Jover R, Llor X, Moreno V, Castells A, Carracedo Á, Castellvi-Bel S, Ruiz-Ponte C. A colorectal cancer genome-wide association study in a Spanish cohort identifies two variants associated with colorectal cancer risk at 1p33 and 8p12. *BMC Genomics* 2013; **14**: 55 [PMID: 23350875 DOI: 10.1186/1471-2164-14-55]

26   **Jia WH**, Zhang B, Matsuo K, Shin A, Xiang YB, Jee SH, Kim DH, Ren Z, Cai Q, Long J, Shi J, Wen W, Yang G, Delahanty RJ, Ji BT, Pan ZZ, Matsuda F, Gao YT, Oh JH, Ahn YO, Park EJ, Li HL, Park JW, Jo J, Jeong JY, Hosono S, Casey G, Peters U, Shu XO, Zeng YX, Zheng W. Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. *Nat Genet* 2013; **45**: 191-196 [PMID: 23263487 DOI: 10.1038/ng.2505]

27   **Peters U**, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, Berndt SI, Bézieau S, Brenner H, Butterbach K, Caan BJ, Campbell PT, Carlson CS, Casey G, Chan AT, Chang-Claude J, Chanock SJ, Chen LS, Coetzee GA, Coetzee SG, Conti DV, Curtis KR, Duggan D, Edwards T, Fuchs CS, Gallinger S, Giovannucci EL, Gogarten SM, Gruber SB, Haile RW, Harrison TA, Hayes RB, Henderson BE, Hoffmeister M, Hopper JL, Hudson TJ, Hunter DJ, Jackson RD, Jee SH, Jenkins MA, Jia WH, Kolonel LN, Kooperberg C, Küry S, Lacroix AZ, Laurie CC, Laurie CA, Le Marchand L, Lemire M, Levine D, Lindor NM, Liu Y, Ma J, Makar KW, Matsuo K, Newcomb PA, Potter JD, Prentice RL, Qu C, Rohan T, Rosse SA, Schoen RE, Seminara D, Shrubsole M, Shu XO, Slattery ML, Taverna D, Thibodeau SN, Ulrich CM, White E, Xiang Y, Zanke BW, Zeng YX, Zhang B, Zheng W, Hsu L. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. *Gastroenterology* 2013; **144**: 799-807.e24 [PMID: 23266556 DOI: 10.1053/j.gastro.2012.12.020]

28   **Houlston RS**, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, Chandler I, Vijayakrishnan J, Sullivan K, Penegar S, Carvajal-Carmona L, Howarth K, Jaeger E, Spain SL, Walther A, Barclay E, Martin L, Gorman M, Domingo E, Teixeira AS, Kerr D, Cazier JB, Niittymäki I, Tuupanen S, Karhu A, Aaltonen LA, Tomlinson IP, Farrington SM, Tene-

sa A, Prendergast JG, Barnetson RA, Cetnarskyj R, Porteous ME, Pharoah PD, Koessler T, Hampe J, Buch S, Schafmayer C, Tepel J, Schreiber S, Völzke H, Chang-Claude J, Hoffmeister M, Brenner H, Zanke BW, Montpetit A, Hudson TJ, Gallinger S, Campbell H, Dunlop MG. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet* 2008; **40**: 1426-1435 [PMID: 19011631 DOI: 10.1038/ng.262]

29   Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K, Spain SL, Broderick P, Domingo E, Farrington S, Prendergast JG, Pittman AM, Theodoratou E, Smith CG, Olver B, Walther A, Barnetson RA, Churchman M, Jaeger EE, Penegar S, Barclay E, Martin L, Gorman M, Mager R, Johnstone E, Midgley R, Niittymäki I, Tuupanen S, Colley J, Idziaszczyk S, Thomas HJ, Lucassen AM, Evans DG, Maher ER, Maughan T, Dimas A, Dermitzakis E, Cazier JB, Aaltonen LA, Pharoah P, Kerr DJ, Carvajal-Carmona LG, Campbell H, Dunlop MG, Tomlinson IP. 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* 2010; **42**: 973-977 [PMID: 20972440 DOI: 10.1038/ng.670]

30   Tomlinson IP, Carvajal-Carmona LG, Dobbins SE, Tenesa A, Jones AM, Howarth K, Palles C, Broderick P, Jaeger EE, Farrington S, Lewis A, Prendergast JG, Pittman AM, Theodoratou E, Olver B, Walker M, Penegar S, Barclay E, Whiffin N, Martin L, Ballereau S, Lloyd A, Gorman M, Lubbe S, Howie B, Marchini J, Ruiz-Ponte C, Fernandez-Rozadilla C, Castells A, Carracedo A, Castellvi-Bel S, Duggan D, Conti D, Cazier JB, Campbell H, Sieber O, Lipton L, Gibbs P, Martin NG, Montgomery GW, Young J, Baird PN, Gallinger S, Newcomb P, Hopper J, Jenkins MA, Aaltonen LA, Kerr DJ, Cheadle J, Pharoah P, Casey G, Houlston RS, Dunlop MG. Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet* 2011; **7**: e1002105 [PMID: 21655089 DOI: 10.1371/journal.pgen.1002105]

31   Dunlop MG, Dobbins SE, Farrington SM, Jones AM, Palles C, Whiffin N, Tenesa A, Spain S, Broderick P, Ooi LY, Domingo E, Smillie C, Henrion M, Frampton M, Martin L, Grimes G, Gorman M, Semple C, Ma YP, Barclay E, Prendergast J, Cazier JB, Olver B, Penegar S, Lubbe S, Chander I, Carvajal-Carmona LG, Ballereau S, Lloyd A, Vijayakrishnan J, Zgaga L, Rudan I, Theodoratou E, Starr JM, Deary I, Kirac I, Kovacević D, Aaltonen LA, Renkonen-Sinisalo L, Mecklin JP, Matsuda K, Nakamura Y, Okada Y, Gallinger S, Duggan DJ, Conti D, Newcomb P, Hopper J, Jenkins MA, Schumacher F, Casey G, Easton D, Shah M, Pharoah P, Lindblom A, Liu T, Smith CG, West H, Cheadle JP, Midgley R, Kerr DJ, Campbell H, Tomlinson IP, Houlston RS. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat Genet* 2012; **44**: 770-776 [PMID: 22634755 DOI: 10.1038/ng.2293]

32   Kinnersley B, Migliorini G, Broderick P, Whiffin N, Dobbins SE, Casey G, Hopper J, Sieber O, Lipton L, Kerr DJ, Dunlop MG, Tomlinson IP, Houlston RS. The TERT variant rs2736100 is associated with colorectal cancer risk. *Br J Cancer* 2012; **107**: 1001-1008 [PMID: 22878375 DOI: 10.1038/bjc.2012.329]

33   Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977; **74**: 5463-5467 [PMID: 271968]

34   Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 2008; **452**: 872-876 [PMID: 18421352 DOI: 10.1038/nature06884]

35   Ulahannan D, Kovac MB, Mulholland PJ, Cazier JB, Tomlinson I. Technical and implementation issues in using next-generation sequencing of cancers in clinical practice. *Br J Cancer* 2013; **109**: 827-835 [PMID: 23887607 DOI: 10.1038/bjc.2013.416]

36   Koboldt DC, Ding L, Mardis ER, Wilson RK. Challenges of sequencing human genomes. *Brief Bioinform* 2010; **11**: 484-498 [PMID: 20519329 DOI: 10.1093/bib/bbq016]

37   Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010; **42**: 30-35 [PMID: 19915526 DOI: 10.1038/ng.499]

38   Desai AN, Jere A. Next-generation sequencing: ready for the clinics? *Clin Genet* 2012; **81**: 503-510 [PMID: 22375550 DOI: 10.1111/j.1399-0004.2012.01865.x]

39   DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**: 491-498 [PMID: 21478889 DOI: 10.1038/ng.806]

40   Kimura M. The Neutral Theory of Molecular Evolution. New York: Cambridge Press, 1983

41   Lim J, Hao T, Shaw C, Patel AJ, Szabó G, Rual JF, Fisk CJ, Li N, Smolyar A, Hill DE, Barabási AL, Vidal M, Zoghbi HY. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* 2006; **125**: 801-814 [PMID: 16713569]

42   Richards FM, McKee SA, Rajpar MH, Cole TR, Evans DG, Jankowski JA, McKeown C, Sanders DS, Maher ER. Germline E-cadherin gene (CDH1) mutations predispose to familial gastric cancer and colorectal cancer. *Hum Mol Genet* 1999; **8**: 607-610 [PMID: 10072428]

43   Lammi L, Arte S, Somer M, Jarvinen H, Lahermo P, Thesleff I, Pirinen S, Nieminen P. Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer. *Am J Hum Genet* 2004; **74**: 1043-1050 [PMID: 15042511]

44   Sweet K, Willis J, Zhou XP, Gallione C, Sawada T, Alhopuro P, Khoo SK, Patocs A, Martin C, Bridgeman S, Heinz J, Pilarski R, Lehtonen R, Prior TW, Frebourg T, Teh BT, Marchuk DA, Aaltonen LA, Eng C. Molecular classification of patients with unexplained hamartomatous and hyperplastic polyposis. *JAMA* 2005; **294**: 2465-2473 [PMID: 16287957]

45   Ngeow J, Heald B, Rybicki LA, Orloff MS, Chen JL, Liu X, Yerian L, Willis J, Lehtonen HJ, Lehtonen R, Mester JL, Moline J, Burke CA, Church J, Aaltonen LA, Eng C. Prevalence of germline PTEN, BMPR1A, SMAD4, STK11, and ENG mutations in patients with moderate-load colorectal polyps. *Gastroenterology* 2013; **144**: 1402-1409, 1409.e1-5 [PMID: 23399955 DOI: 10.1053/j.gastro.2013.02.001]

46   Zogopoulos G, Jorgensen C, Bacani J, Montpetit A, Lepage P, Ferretti V, Chad L, Selvarajah S, Zanke B, Hudson TJ, Pawson T, Gallinger S. Germline EPHB2 receptor variants in familial colorectal cancer. *PLoS One* 2008; **3**: e2885 [PMID: 18682749 DOI: 10.1371/journal.pone.0002885]

47   Guda K, Moinova H, He J, Jamison O, Ravi L, Natale L, Lutterbaugh J, Lawrence E, Lewis S, Willson JK, Lowe JB, Wiesner GL, Parmigiani G, Barnholtz-Sloan J, Dawson DW, Velculescu VE, Kinzler KW, Papadopoulos N, Vogelstein B, Willis J, Gerken TA, Markowitz SD. Inactivating germ-line and somatic mutations in polypeptide N-acetylgalactosaminyltransferase 12 in human colon cancers. *Proc Natl Acad Sci U S A* 2009; **106**: 12921-12925 [PMID: 19617566 DOI: 10.1073/pnas.0901454106]

48   Clarke E, Green RC, Green JS, Mahoney K, Parfrey PS, Younghusband HB, Woods MO. Inherited deleterious variants in GALNT12 are associated with CRC susceptibility. *Hum Mutat* 2012; **33**: 1056-1058 [PMID: 22461326 DOI:

10.1002/humu.22088]

49   **Venkatachalam R**, Ligtenberg MJ, Hoogerbrugge N, Schackert HK, Görgens H, Hahn MM, Kamping EJ, Vreede L, Hoenselaar E, van der Looij E, Goossens M, Churchman M, Carvajal-Carmona L, Tomlinson IP, de Bruijn DR, Van Kessel AG, Kuiper RP. Germline epigenetic silencing of the tumor suppressor gene PTPRJ in early-onset familial colorectal cancer. *Gastroenterology* 2010; **139**: 2221-2224 [PMID: 21036128 DOI: 10.1053/j.gastro.2010.08.063]

50   **Lubbe SJ**, Pittman AM, Matijssen C, Twiss P, Olver B, Lloyd A, Qureshi M, Brown N, Nye E, Stamp G, Blagg J, Houlston RS. Evaluation of germline BMP4 mutation as a cause of colorectal cancer. *Hum Mutat* 2011; **32**: E1928-E1938 [PMID: 20949628 DOI: 10.1002/humu.21376]

51   **Nieminen TT**, Abdel-Rahman WM, Ristimäki A, Lappalainen M, Lahermo P, Mecklin JP, Järvinen HJ, Peltomäki P. BMPR1A mutations in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. *Gastroenterology* 2011; **141**: e23-e26 [PMID: 21640116 DOI: 10.1053/j.gastro.2011.03.063]

52   **Orloff MS**, He X, Peterson C, Chen F, Chen JL, Mester JL, Eng C. Germline PIK3CA and AKT1 mutations in Cowden and Cowden-like syndromes. *Am J Hum Genet* 2013; **92**: 76-80 [PMID: 23246288 DOI: 10.1016/j.ajhg.2012.10.021].]

53   **Palles C**, Cazier JB, Howarth KM, Domingo E, Jones AM, Broderick P, Kemp Z, Spain SL, Guarino E, Salguero I, Sherborne A, Chubb D, Carvajal-Carmona LG, Ma Y, Kaur K, Dobbins S, Barclay E, Gorman M, Martin L, Kovac MB, Humphray S, Lucassen A, Holmes CC, Bentley D, Donnelly P, Taylor J, Petridis C, Roylance R, Sawyer EJ, Kerr DJ, Clark S, Grimes J, Kearsey SE, Thomas HJ, McVean G, Houlston RS, Tomlinson I. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* 2013; **45**: 136-144 [PMID: 23263490 DOI: 10.1038/ng.2503]

54   **Seshagiri S**. The burden of faulty proofreading in colon cancer. *Nat Genet* 2013; **45**: 121-122 [PMID: 23358219 DOI: 10.1038/ng.2540]

55   **Jin YH**, Ayyagari R, Resnick MA, Gordenin DA, Burgers PM. Okazaki fragment maturation in yeast. II. Cooperation between the polymerase and 3'-5'-exonuclease activities of Pol delta in the creation of a ligatable nick. *J Biol Chem* 2003; **278**: 1626-1633 [PMID: 12424237]

56   **Murphy K**, Darmawan H, Schultz A, Fidalgo da Silva E, Reha-Krantz LJ. A method to select for mutator DNA polymerase deltas in Saccharomyces cerevisiae. *Genome* 2006; **49**: 403-410 [PMID: 16699561]

57   **Albertson TM**, Ogawa M, Bugni JM, Hays LE, Chen Y, Wang Y, Treuting PM, Heddle JA, Goldsby RE, Preston BD. DNA polymerase epsilon and delta proofreading suppress discrete mutator and cancer phenotypes in mice. *Proc Natl Acad Sci USA* 2009; **106**: 17101-17104 [PMID: 19805137 DOI: 10.1073/pnas.0907147106]

58   **Smith CG**, Naven M, Harris R, Colley J, West H, Li N, Liu Y, Adams R, Maughan TS, Nichols L, Kaplan R, Wagner MJ, McLeod HL, Cheadle JP. Exome resequencing identifies potential tumor-suppressor genes that predispose to colorectal cancer. *Hum Mutat* 2013; **34**: 1026-1034 [PMID: 23585368 DOI: 10.1002/humu.22333]

59   **DeRycke MS**, Gunawardena SR, Middha S, Asmann YW, Schaid DJ, McDonnell SK, Riska SM, Eckloff BW, Cunningham JM, Fridley BL, Serie DJ, Bamlet WR, Cicek MS, Jenkins MA, Duggan DJ, Buchanan D, Clendenning M, Haile RW, Woods MO, Gallinger SN, Casey G, Potter JD, Newcomb PA, Le Marchand L, Lindor NM, Thibodeau SN, Goode EL. Identification of novel variants in colorectal cancer families by high-throughput exome sequencing. *Cancer Epidemiol Biomarkers Prev* 2013; **22**: 1239-1251 [PMID: 23637064 DOI: 10.1158/1055-9965.EPI-12-1226]

Baishideng®

ISSN 1007-9327

0 8>

9 771007 932045