

# Gene discovery in familial cancer syndromes by exome sequencing: prospects for the elucidation of familial colorectal cancer type X

Chee-Seng Ku<sup>1</sup>, David N Cooper<sup>2</sup>, Mengchu Wu<sup>1</sup>, Dimitrios H Roukos<sup>3,4</sup>, Yudi Pawitan<sup>5</sup>, Richie Soong<sup>1</sup> and Barry Iacopetta<sup>6</sup>

<sup>1</sup>Cancer Science Institute of Singapore, National University of Singapore, Singapore, Singapore; <sup>2</sup>Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, UK; <sup>3</sup>Personalized Cancer Genomic Medicine, Ioannina University, Ioannina, Greece; <sup>4</sup>Department of Surgery, Ioannina University School of Medicine, Ioannina, Greece; <sup>5</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden and <sup>6</sup>School of Surgery, University of Western Australia, Crawley, WA, Australia

Recent advances in genotyping and sequencing technologies have provided powerful tools with which to explore the genetic basis of both Mendelian (monogenic) and sporadic (polygenic) diseases. Several hundred genome-wide association studies have so far been performed to explore the genetics of various polygenic or complex diseases including those cancers with a genetic predisposition. Exome sequencing has also proven very successful in elucidating the etiology of a range of hitherto poorly understood Mendelian disorders caused by high-penetrance mutations. Despite such progress, the genetic etiology of several familial cancers, such as familial colorectal cancer type X, has remained elusive. Familial colorectal cancer type X and Lynch syndrome are similar in terms of their fulfilling certain clinical criteria, but the former group is not characterized by germline mutations in DNA mismatch-repair genes. On the other hand, the genetics of sporadic colorectal cancer have been investigated by genome-wide association studies, leading to the identification of multiple new susceptibility loci. In addition, there is increasing evidence to suggest that familial and sporadic cancers exhibit similarities in terms of their genetic etiologies. In this review, we have summarized our current knowledge of familial colorectal cancer type X, discussed current approaches to probing its genetic etiology through the application of new sequencing technologies and the recruitment of the results of colorectal cancer genome-wide association studies, and explore the challenges that remain to be overcome given the uncertainty of the current genetic model (ie, monogenic vs polygenic) of familial colorectal cancer type X.

*Modern Pathology* (2012) 25, 1055–1068; doi:10.1038/modpathol.2012.62; published online 20 April 2012

**Keywords:** exome sequencing; familial cancers; familial colorectal cancer type X; genome-wide association studies; Mendelian; polygenic

Recent developments in high-throughput sequence capture methods and next-generation sequencing technologies have made exome sequencing a viable approach to the identification of pathological muta-

tions, both from a technical standpoint and in terms of being cost-effective.<sup>1–4</sup> The advent of exome sequencing has already contributed significantly toward the identification of new causal mutations (and genes) for a number of previously unresolved Mendelian disorders such as Kabuki syndrome, Miller syndrome, Sensenbrenner syndrome, and Fowler syndrome to name just a few. Further, exome sequencing has proven to be an effective tool to interrogate the genetic basis of Mendelian disorders in samples derived from both families and unrelated individuals.<sup>5–8</sup> Since the inception of the idea of using exome sequencing as both a discovery<sup>9</sup> and a

Correspondence: Dr C-S Ku, PhD, Centre for Translational Medicine, Yong Loo Lin School of Medicine, Cancer Science Institute of Singapore, National University of Singapore, 14 Medical Drive, #12-01, MD6, Singapore 117599, Singapore or Professor DN Cooper, PhD, Institute of Medical Genetics, Cardiff University School of Medicine, Heath Park, Cardiff CF14 4XN, UK.  
E-mail: csku80@gmail.com or cooperDN@cardiff.ac.uk  
Received 30 December 2011; revised 17 February 2012; accepted 19 February 2012; published online 20 April 2012

diagnostic tool<sup>10</sup> for Mendelian disorders, this field has advanced very considerably.<sup>11</sup> Accompanied and aided by other technical advances such as the development of computational and statistical approaches to interrogate the myriad variants identified by exome sequencing,<sup>12,13</sup> including algorithms to detect copy number variants using exome sequencing data,<sup>14</sup> and the idea (and practical demonstration) of using single-nucleotide polymorphism genotypes extracted from exome sequencing data to perform accurate genetic linkage mapping to reduce the 'search space' for genetic variants,<sup>15</sup> exome sequencing has emerged as a mature analytical approach.

Although major progress has been made in understanding the genetic basis of Mendelian disorders over the past 3 years using exome sequencing, so far only limited studies have interrogated familial forms of cancer, ie, familial pancreatic cancer<sup>16</sup> and hereditary pheochromocytoma (a rare neural crest cell tumor).<sup>17</sup> By harnessing the latest technological advances, Jones *et al*<sup>16</sup> identified a germline truncating mutation in *PALB2* through exome sequencing a single patient with familial pancreatic cancer. That this patient might have a familial form of pancreatic cancer was suggested by the fact that his sister had also developed the disease. In similar manner, mutations in *MAX*, the MYC-associated factor X gene, were also identified through sequencing the exomes of three unrelated individuals with hereditary pheochromocytoma.<sup>17</sup>

Since 2005, >100 genome-wide association studies have been performed to interrogate the genetic basis of various sporadic or polygenic forms of cancer (such as colorectal, prostate, breast, and lung) for which numerous statistically robust and novel single-nucleotide polymorphisms or genetic loci have been identified.<sup>18,19</sup> In addition to their polygenic nature, these cancers are multifactorial, involving a complex interaction of multiple genetic and environmental factors. By contrast, little progress has so far been achieved in the context of 'familial' cancers (ie, cancers displaying a very evident family history with clustering of multiple affected family members). More specifically, familial forms of cancer typically occur in more individuals in a given family than would be expected by chance alone. Familial cancers are often characterized by their occurrence at a comparatively early age, thereby indicating the potential presence of a gene mutation that increases the risk of cancer. However, familial clustering of cases may also be a sign of a shared environment or lifestyle, or alternatively chance alone. By contrast, sporadic cancers lack any obvious family history of the disease.

The slow progress of research into familial cancer has been illustrated, for example, in hereditary diffuse gastric cancer. *CDH1* was the first causal gene identified for this cancer in 1998,<sup>20</sup> and it remains the only known gene underlying hereditary diffuse gastric cancer. However, germline mutations in this gene account for only a proportion of

hereditary diffuse gastric cancer cases,<sup>21</sup> suggesting that an as-yet-to-be identified gene(s) is likely to be responsible for the remaining cases unexplained by *CDH1*. Similarly, *BRCA1* and *BRCA2* are the only high-penetrance genes for familial breast cancer, although numerous novel single-nucleotide polymorphisms and genetic loci conferring low-to-moderate risk or effect size (odds ratio <1.5) have been identified by genome-wide association studies of polygenic breast cancer.<sup>22,23</sup> Some of these common alleles have been reported to modify risk in *BRCA1* and *BRCA2* mutations carriers.<sup>24</sup> However, so far the results from genome-wide association studies have limited value for individual risk prediction,<sup>25</sup> as compared with the high-penetrance inherited mutations in causal genes for familial breast cancer which can prompt drastic clinical intervention such as mastectomy. An analysis to evaluate the potential for individualized disease risk stratification based on common single-nucleotide polymorphisms identified by genome-wide association studies in breast cancer came to the conclusion that the clinical utility of single, common, low-penetrance genes for breast cancer risk prediction is currently quite limited.<sup>26</sup>

In the context of familial colorectal cancer, the genetic causes of familial adenomatous polyposis and Lynch syndrome have been well documented; in most instances, they are accounted for by germline mutations in the *APC* gene and DNA mismatch-repair genes (ie, *MSH2*, *MLH1*, *MSH6*, and *PMS2*), respectively. For example, ~90% of familial adenomatous polyposis cases are caused by germline mutations in the *APC* gene. The majority of these mutations introduce a premature stop codon resulting in a truncated protein. Similarly, the *MSH2* and *MLH1* genes harbor >90% of the germline mutations found in Lynch syndrome patients.<sup>27,28</sup> By contrast, the genetic etiology of familial colorectal cancer type X remains largely unknown.<sup>29</sup> It is widely anticipated that new insights generated from studies on familial colorectal cancer type X will lead to the molecular characterization of a novel form of familial colorectal cancer which will necessitate the reclassification of subsets of families with a strong history of colorectal cancer.

## Common approaches to interrogate the genetic basis of disease

Several genetic approaches are available to interrogate the genetic basis of disease. Hence, choosing the study design to best fit the research questions to be answered is critical. This is in turn influenced by the nature of the disease, or the postulated model of the disease. The nature of the disease refers to whether it is likely to be monogenic or polygenic at both extremes of a 'genetic spectrum'. At one extreme, monogenic disease is caused by the action of a high-penetrance mutation, whereas at the opposite extreme, polygenic disease is likely to result from the

complex interaction of multiple genetic variants with low penetrance (measured as an odds ratio in association studies, ie, odds ratio <1.5).<sup>30–37</sup> The penetrance estimates the proportion of individuals carrying a particular mutation or variant that also express an associated phenotype or disease. Thus, most carriers of high-penetrance mutations will eventually develop the clinical phenotype in question.

Some familial cancers such as familial adenomatous polyposis, Lynch syndrome, and hereditary diffuse gastric cancer are clear examples of monogenic diseases. However, the distinction between monogenic and polygenic diseases is frequently blurred, particularly in the context of cancer predisposition, and the nature of some disease states could lie somewhere on a continuum. Although monogenic disease is caused by a single complete penetrance or high-penetrance mutation, modifier mutations are frequently found that affect the clinical manifestations of the disease, impacting on, eg, disease severity or phenotypic heterogeneity.<sup>38–40</sup> Thus, variants on 8q23.3 and 11q23.1 have been found to modify colorectal cancer risk in Lynch syndrome.<sup>41,42</sup> On the other hand, some polygenic diseases such as age-related macular degeneration have been attributed to several genetic variants with considerably large effect sizes (eg, the variants in *CFH*; odds ratio=4.6), which taken together, account for 50% of the heritability of the disease. By contrast, 32 genetic variants identified as being associated with Crohn's disease are able to explain only 20% of the heritability.<sup>43</sup>

Genome-wide linkage studies followed by positional cloning have been very successful in identifying causal mutations for monogenic disorders because the mutations segregate perfectly with the disorders according to classical Mendelian inheritance patterns (ie, autosomal dominant, autosomal recessive, and X-linked). The perfect segregation patterns rely upon complete (or at least high) penetrance of the causal mutation.<sup>36</sup> For example, *APC* was identified back in 1991 through linkage analysis and positional cloning involving chromosome 5q21.<sup>44,45</sup> Homozygosity mapping, on the other hand, is a more powerful and effective approach to study recessive Mendelian disorders in consanguineous families. Such traditional approaches that narrow the search down to a specific genomic locus have now been coupled with next-generation sequencing to identify the underlying gene and causal mutation(s). However, extremely rare disorders or conditions characterized by *de-novo* mutations are not amenable to analysis using this type of study design. The advent of exome sequencing has however succeeded in identifying new causal mutations and genes for disorders previously unresolved by traditional approaches.<sup>7,8</sup>

Exome sequencing requires several sequence enrichment steps followed by massively parallel sequencing. The development of commercial

whole-exome enrichment kits by Agilent, Illumina, and Nimblegen has been largely responsible for making exome sequencing feasible in practical terms for the ordinary laboratory.<sup>1–4</sup> During the enrichment steps, the genomic regions of interest (ie, all exons) are captured, whereas the unwanted DNA sequences (ie, the non-coding regions) are removed before sequencing, leading to a significant reduction in the proportion of the genome that needs to be sequenced. As a consequence, a greater sequencing depth (ie, a larger number of sequence reads covering a given region) of the targeted regions can be achieved. This is particularly important in the context of diagnostic applications, whether to allow the accurate detection of inherited disease mutations or to confirm that an apparently negative result is correct.<sup>46</sup> Figure 1 displays a schematic illustration of library preparation using array-based and in-solution hybrid capture.

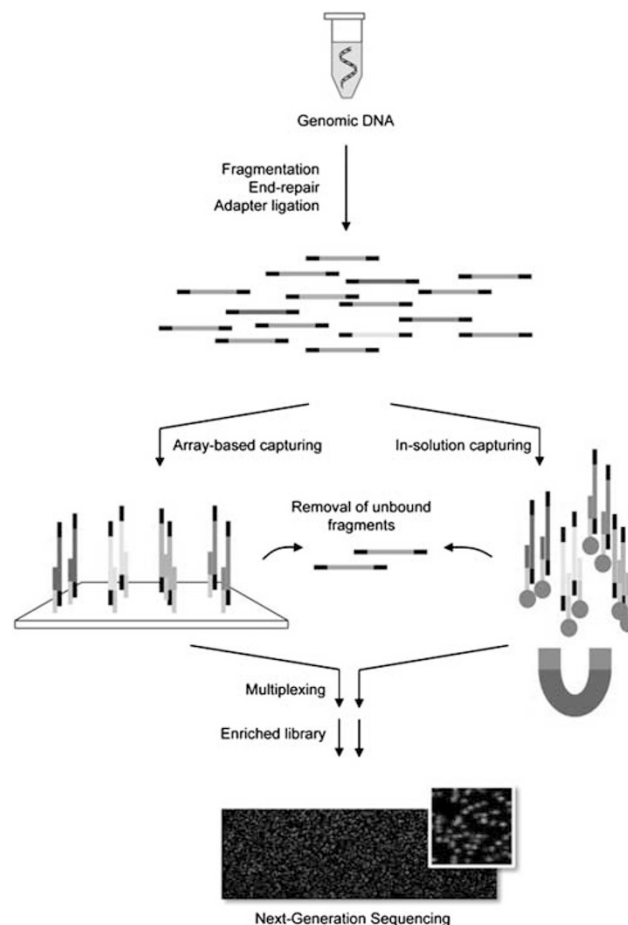
Exome sequencing is considered to be a more cost-effective tool, and analytically less challenging, than whole-genome sequencing. Despite the higher throughput and steadily decreasing cost of next-generation sequencing, performing whole-genome sequencing on hundred samples is still prohibitively expensive for most laboratories. This is particularly pertinent in the context of studying polygenic diseases where a large sample set is often required to achieve adequate statistical power to identify disease-associated variants that subsequently turn out to display low penetrance in genome-wide association studies. Although a commercial whole-genome sequencing service costs approximately US\$5000, the expense incurred for data storage and computational analysis is still substantial.<sup>47,48</sup> Further, because exome sequencing focuses on 1–2% of the human genome, greater sequencing depth of the coding regions is obtained. As a result, exome sequencing has become a more popular approach than whole-genome sequencing.

On the other hand, the genome-wide association studies approach, deployed as commercial genome-wide single-nucleotide polymorphism genotyping arrays, is being widely applied to interrogate polygenic diseases. The genotyping array is designed to investigate common single-nucleotide polymorphisms (with a minor allele frequency of >5%) in both coding and non-coding regions throughout the human genome.<sup>49</sup> The design of genome-wide association studies and single-nucleotide polymorphism selection in genotyping arrays has been largely driven by the common-disease common-variant hypothesis.<sup>50</sup> As a result, most of the risk alleles that have been identified by genome-wide association studies are common (allele frequency >5%) conferring small effect sizes (odds ratio <1.5). Owing to these small effect sizes, the identified single-nucleotide polymorphisms collectively only explain a small portion of the heritability of the diseases investigated by genome-wide association studies, leaving most of the heritability unexplained; this is commonly

known as the 'missing heritability'.<sup>43</sup> Thus, the validity of the common-disease common-variant model has now been challenged by the fact of this missing heritability and research efforts are already being directed toward the investigation of rare variants by sequencing the entire loci implicated by genome-wide association studies.<sup>51–55</sup> Results to date suggest that the two competing hypotheses (common-disease common variant and multiple rare variant) are not mutually exclusive, as studies have revealed the contributions of both common and rare variants to complex disease.<sup>56</sup> Existing data have already suggested that both the multiple rare variant and common-disease common-variant models can pertain at the same loci, eg, rare variants were identified in the loci implicated by genome-wide association studies through DNA sequencing, and these loci have been termed as 'pleomorphic risk loci'.<sup>57</sup> The development of newer genotyping arrays, such as the Illumina Omni arrays that leverage data from the 1000 Genomes Project, has increased the coverage of 'less common' single-nucleotide polymorphisms (minor allele frequency 1–5%). Exome sequencing has now further enhanced our ability to investigate rare single-nucleotide variants (<1%) within gene coding regions.<sup>52</sup>

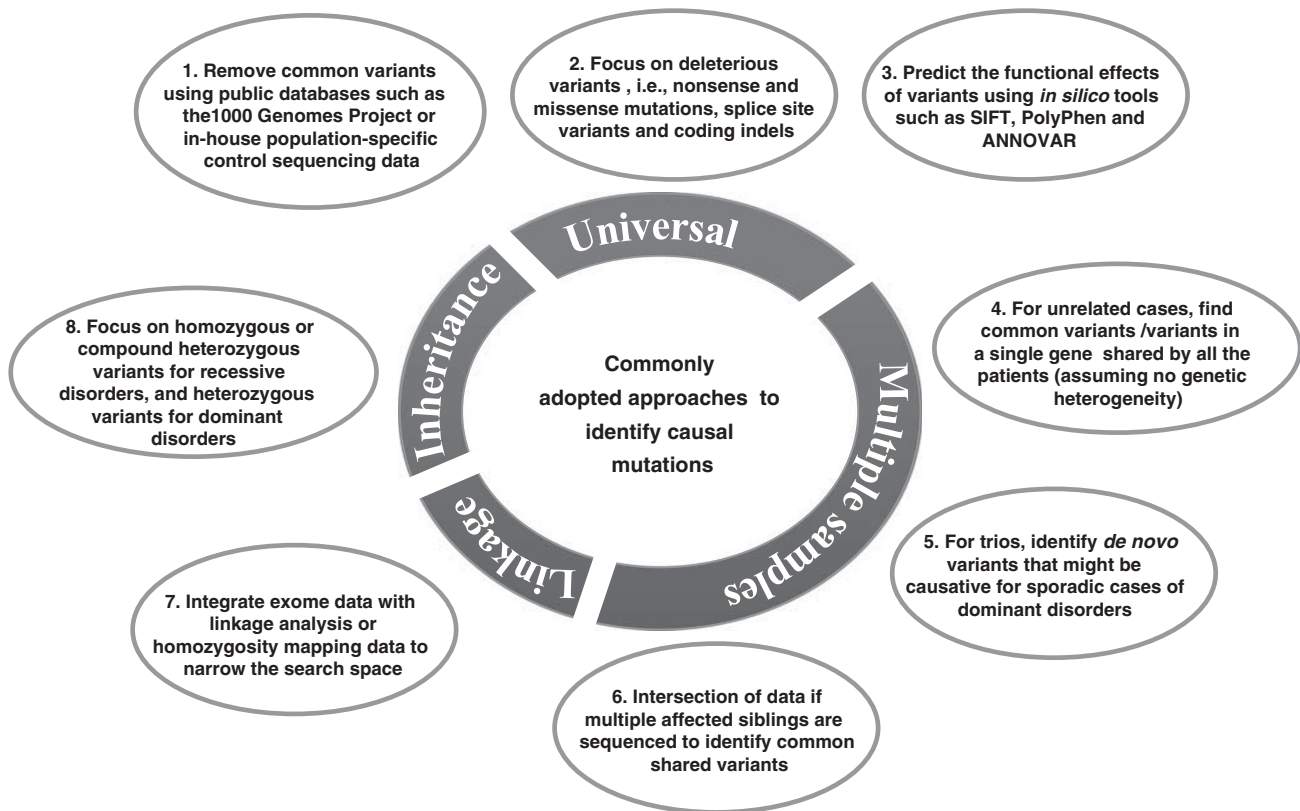
## Exome sequencing of familial cancers

Exome sequencing has increasingly been applied to the identification of somatic mutations in various sporadic cancers,<sup>58–62</sup> but so far few studies have identified deleterious germline mutations in familial cancers.<sup>16,17</sup> Despite the limited number of studies performed to date, the advent of exome sequencing has provided new opportunities to identify the germline mutations ultimately responsible for familial cancers. Thus, a germline truncating mutation in the *PALB2* gene was identified through exome sequencing of a patient with suspected familial pancreatic cancer.<sup>16</sup> A total of 15 461 germline variants were identified in patient's exome that comprised synonymous, missense, non-sense, and splice site mutations plus microdeletions and microinsertions. On the basis that the gene underlying a cancer with a hereditary predisposition may be expected to harbor no normal alleles (ie, one allele is inherited in mutant form, often truncating, whereas the other (wild-type) allele is inactivated by somatic mutation), the 'search space' was narrowed down to three genes (*SERPINB12*, *RAGE*, and *PALB2*). Of these, *PALB2* appeared to be the most promising causal gene candidate for the pancreatic cancer in question because the inherited stop codon mutations in the other two genes are relatively common in healthy individuals. This hypothesis was confirmed by finding that the patient harbored a germline frameshift deletion (4 bp) and had somatically acquired a second mutation at the canonical exon 10 splice site of *PALB2*. Further investigation



**Figure 1** Schematic illustration of library preparation using array-based and in-solution hybrid capture. The workflow involves multiple steps and is similar for the capture of both exome and custom or targeted regions. However, the difference between the two contexts lies in the different sets of probes required for the exome and the targeted analysis of specific genes. The difference between on-solid and in-solution capture methods is that the oligonucleotide probes are tethered on microarrays or are suspended in-solution (oligonucleotide probes attached to beads), respectively. The capture of the adapter-ligated DNA fragments is based on their complementarity with the oligonucleotide probe sequences. The coupling of these enrichment methods with NGS technologies has made exome sequencing more feasible technically as well as more cost-effective.

to determine whether *PALB2* mutations occur in additional putatively familial pancreatic cancers then identified truncating mutations in 3 of 96 patients screened.<sup>16</sup> In addition to familial pancreatic cancer, germline mutations in *PALB2* have previously been associated with familial breast cancer and Fanconi anemia.<sup>63,64</sup> Recent studies have also found evidence of inherited mutations in the *PALB2* gene underlying familial breast cancer.<sup>65–67</sup> For example, in a large-scale analysis of 1144 familial breast cancer patients, ~3.4% of patients were heterozygous for a non-sense, frameshift, or frameshift-associated splice mutation in *PALB2*, reinforcing the apparent general relevance of this gene in familial cancers.<sup>66</sup>



**Figure 2** Commonly adopted approaches to identifying causal mutations. The three ‘universal’ criteria used to filter the ‘less likely’ causal variants are removing common variants, focusing on deleterious variants, and predicting and retaining variants with functional effects (ie, criteria 1–3). The other criteria are dependent upon the study design, for example, whether unrelated or family samples are sequenced (criteria 4–6). The variant filtering or analysis also depends on whether linkage and homozygosity data from the families are available, since such data significantly reduce the search space for potentially causal variants (criterion 7). Finally, the mode of inheritance of the Mendelian disorder will determine whether the focus should be placed upon homozygous, compound heterozygous or heterozygous variants (criterion 8). Additional criteria may be needed, for example, restricting cases that are highly similar in terms of their phenotypic manifestations to minimize the clinical/phenotypic heterogeneity that may mask the identification of causal mutations. However, these criteria are likely to vary from study to study.

In similar vein, germline mutations in *MAX* were identified in three unrelated individuals with hereditary pheochromocytoma by exome sequencing.<sup>17</sup> This study studied heterozygous single-nucleotide variants and small indels, because it was postulated that it would be very unlikely that homozygous variants could act as founder mutations. Several common filtering criteria were applied, as with other studies Mendelian disorders (Figure 2), ie, selecting only those variants within coding regions which were predicted to yield amino-acid changes and which affected the same gene in all three samples. This resulted in the identification of a total of five single-nucleotide variants, located within two genes (*MAX* and *ADCY6*). The *ADCY6* gene was then excluded as the causal gene because one of the variants did not segregate with the disease through the family pedigree. By contrast, segregation of two variants in *MAX* with hereditary pheochromocytoma was observed in the two families from whom DNA from affected relatives was available. Additional evidence to support the causative role of the variants

in *MAX* came from their absence (or non-detection) in > 750 population-matched control chromosomes. Finally, additional screening for *MAX* mutations in 59 cases lacking germline mutations in known genes for hereditary pheochromocytoma identified two additional truncating mutations and three missense variants.<sup>17</sup> This remarkable discovery highlights the potential of exome sequencing to discover the genetic causes of those familial cancers that remain to be characterized.

Taken together, these studies have demonstrated the feasibility of exome sequencing as a powerful discovery tool for germline mutations in familial cancer. Thus, the number of successful examples of exome sequencing is expected to increase over the coming years until such a time as whole-genome sequencing becomes less challenging in terms of its ‘real’ cost (ie, not just the cost of the sequencing) and data analysis. Exome sequencing in familial cancers is a burgeoning field. By contrast, >100 genome-wide association studies have investigated the associations of germline variants (common single-nucleotide polymorphisms) with various polygenic

cancers. These approaches could also be applied to familial colorectal cancer type X. However, the choice of which approach to adopt depends upon both the question posed and the hypothesis to be investigated. Although exome sequencing has yet to be applied to familial colorectal cancer, >10 genome-wide association studies have so far been conducted in the context of polygenic colorectal cancer. The total number of polygenic colorectal cancer loci (identified by, eg, genome-wide association studies) has now exceeded 20 (A Catalog of Published Genome-Wide Association Studies, <http://www.genome.gov/gwastudies/>). As with other cancers, all the risk alleles so far identified have an odds ratio of <1.3, indicating that most of the heritability of polygenic colorectal cancer remains unaccounted for. However, it is currently unclear whether the results from these genome-wide association studies are also applicable to familial colorectal cancer type X.

### Familial colorectal cancer type X is distinct from lynch syndrome

Hereditary non-polyposis colorectal cancer is the most common familial form of colorectal cancer. The term hereditary non-polyposis colorectal cancer has been applied to heterogeneous groups of families that fulfill the Amsterdam Criteria.<sup>68</sup> The Amsterdam Criteria are summarized in Table 1. The clinical hallmarks of hereditary non-polyposis colorectal cancer syndrome resulted in a classification scheme designated as Amsterdam Criteria I, later modified as Amsterdam Criteria II to take into account the extra-colonic cancers. Hereditary non-polyposis colorectal cancer is commonly used interchangeably with Lynch syndrome; however, only a subset of hereditary non-polyposis colorectal cancer cases have Lynch syndrome (characterized by microsatellite instability because of the deficiency in mismatch repair).

It has now however become apparent that not all Amsterdam Criteria-positive families have Lynch syndrome.<sup>69</sup> The recognition that microsatellite instability is caused by loss of mismatch-repair activity led to the original discovery of the genes that underlie Lynch syndrome. It is noteworthy that hereditary non-polyposis colorectal cancer has a clinical diagnosis; by contrast, Lynch syndrome is defined by a molecular or genetic diagnosis of germline mutations in DNA mismatch-repair genes. A total of four mismatch-repair genes have been identified (*MSH2*, *MLH1*, *MSH6*, and *PMS2*), the first two of which (*MSH2* and *MLH1*) account for >90% of the germline mutations.<sup>28</sup> In addition to germline mutations, epimutations of the mismatch-repair genes have also been identified as causing Lynch syndrome.<sup>70–72</sup> *EPCAM* (also known as *TACSTD1*) deletion leading to *MSH2* promoter methylation is another common genetic aberration causing Lynch syndrome.<sup>73–75</sup> Therefore, it is evident that only a proportion of hereditary

**Table 1** Amsterdam Criteria

<i>Amsterdam I Criteria</i>	<i>Amsterdam II Criteria</i>
<b>At least three relatives with colorectal cancer and the following:</b>	<b>At least three relatives with hereditary non-polyposis colorectal cancer-related cancers (colorectal, endometrial, small bowel, ureter, or renal pelvis) and the following:</b>
One should be a first-degree relative of the other two	One should be a first-degree relative of the other two
At least two consecutive generations should be affected	At least two consecutive generations should be affected
At least one case of colorectal cancer should be before age 50	At least one case of hereditary non-polyposis colorectal cancer-related cancer should be before age 50
Familial adenomatous polyposis should be excluded in any cases of colorectal cancer	Familial adenomatous polyposis should be excluded in any cases of colorectal cancer
Verification of tumors' histopathology	Verification of tumors' histopathology

non-polyposis colorectal cancer cases which fulfill the Amsterdam Criteria have germline mutations in mismatch-repair genes, which by 'definition' constitutes Lynch syndrome. As a result, families with a strong family history of colorectal cancer that do not have Lynch syndrome have been grouped as 'familial colorectal cancer-type X'; alternatively, some studies have referred to them as mismatch-repair gene mutation-positive and -negative families for Lynch syndrome and familial colorectal cancer type X, respectively.<sup>29</sup> Similarly, patients with a family history of gastric cancer but lacking the *CDH1* mutation may constitute a distinct condition termed as the familial gastric cancer syndrome,<sup>76,77</sup> which may be regarded as being conceptually similar to familial colorectal cancer type X.

Familial colorectal cancer type X constitutes a heterogeneous group of cases with clear clinical differences as compared with Lynch syndrome. The term familial colorectal cancer type X was introduced in 2005 by Lindor *et al*<sup>29</sup> who studied 161 familial clusters of colorectal cancer, all of which met the Amsterdam Criteria. The investigators divided the families into those in whom microsatellite instability was present in the colorectal cancer cases and those in whom microsatellite instability was absent. It was held to be reasonable to suppose that these two groups might be characterized by different disease entities, and it was shown that this was indeed the case. The families with microsatellite instability had all the classic clinical features of Lynch syndrome, and the average age to develop cancer was 48.7 years. There was a six-fold increase in risk for colorectal cancer and significantly increased risks for

extra-colonic cancers such as cancers of the uterus, stomach, kidney, ovary, small intestine, and ureter. By contrast, families in the other group (microsatellite instability absent) who met the Amsterdam Criteria but did not harbor mutations in their mismatch-repair genes had a significant two-fold increase in risk for colorectal cancer but not for other cancers. Further, the average age for developing colorectal cancer in this group was 60.7 years, significantly older than the group with microsatellite instability. Thus, by both genetic (germline mutations in mismatch-repair genes) and clinical criteria, this study suggested that Lynch syndrome and familial colorectal cancer type X are two distinct disease entities.<sup>29</sup>

This new definition and classification has important clinical implications. Before the concept of 'two disease entities', it was unclear whether families with the clinical hereditary non-polyposis colorectal cancer syndrome but lacking microsatellite instability and germline mutation in one of the mismatch-repair genes, should be managed in the same way as those families with both of these genetic characteristics. Following the demonstration of the genetic and clinical differences between Lynch syndrome and familial colorectal cancer type X, it was suggested that the surveillance protocol of frequent colonoscopy and endometrial surveillance usually advised for Lynch syndrome families would not be recommended for the familial colorectal cancer type X families. Instead, a less stringent protocol including colonoscopy every 5 years and starting at a more advanced age was proposed.<sup>29</sup> Thus, the identification of the casual mutations and genes underlying familial colorectal cancer type X is also expected to help in the molecular genetic diagnosis of the disease, as has been the case in other types of familial colorectal cancer such as Lynch syndrome, familial adenomatous polyposis, and MUTYH-associated polyposis, and in developing guidelines to ensure appropriate screening for any high-risk group and the clinical management of patients.

In similar vein, significant clinical differences between these two different hereditary non-polyposis colorectal cancer groups have also been found in multiple studies.<sup>78-81</sup> In a study that investigated 25 families with truncating mutations in *MLH1* or *MSH2*, and 16 families that fulfilled the Amsterdam Criteria but lacked mutations in these genes, major clinical differences in age of onset, tumor spectrum, tumor localization, and tumor progression were found.<sup>81</sup> These differences included an earlier age of onset for colorectal cancer and other tumors, and more synchronous and metachronous colorectal and extra-colonic tumors in the group with mutations in *MLH1* or *MSH2*. Another striking difference between the two groups was the localization of colorectal cancer. Within the group with mutations in *MLH1* or *MSH2*, 68% of the cancers were found proximal to the splenic flexure, representing a distribution over the entire colorectum, whereas in the other group,

79% of cancers were located distal to the splenic flexure, all in the sigmoid colon and rectum. Another important difference was in the number of adenomas observed between the two groups. A higher colorectal adenoma/carcinoma ratio and a tendency toward more synchronous or metachronous adenomas were observed in the group without mutations, indicating a slower progression of adenomas to carcinomas in the absence of a mismatch-repair gene deficiency.

Chromosomal instability, microsatellite instability, and CpG island methylation phenotype are three major pathways that account for the majority of colorectal cancers. The autosomal dominant familial adenomatous polyposis (caused by mutations in the *APC* gene) and the autosomal recessive MUTYH-associated polyposis (caused by mutations in *MUTYH*) exerted their influence through the chromosomal instability pathway. On the other hand, the autosomal dominant Lynch syndrome (caused by mutations in the mismatch-repair genes) involved the microsatellite instability pathway. Finally, CpG island methylation phenotype is associated with hyperplastic polyposis and the development of colorectal cancer from serrated polyps, with the presence of *BRAF* somatic mutations and *MLH1* promoter methylation in the tumor.<sup>82-84</sup> The CpG island methylation phenotype is no longer grouped into positive (CpG island methylation phenotype high) vs negative, because an additional category known as CpG island methylation phenotype low has been identified which has some unique features including associations with *KRAS* mutations,<sup>85-87</sup> *MGMT* promoter methylation,<sup>88</sup> and low-level methylation at specific loci.<sup>89</sup> These unique features are distinct from the CpG island methylation phenotype high and CpG island methylation phenotype-negative categories which suggests that there are at least three different CpG island methylation phenotypic categories in human colorectal cancer. Therefore, it would be interesting to explore whether germline mutations in these known genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*, *APC*, and *MUTYH*) and pathways could be involved in familial colorectal cancer type X. However, it has been shown that the majority of families fulfilling the Amsterdam Criteria did not have an identifiable mutation in these six genes and none had evidence of methylator pathway dysfunction. This suggests that high-penetrance mutations in as-yet-to-be identified genes may cause familial colorectal cancer type X.<sup>90</sup> Taken together, a variety of genetic and clinical differences between Lynch syndrome and familial colorectal cancer type X have emerged.

Various molecular features (such as epigenetics) of familial colorectal cancer type X have been further characterized and these have helped to distinguish it from other types of colorectal cancer.<sup>91,92</sup> More specifically, colorectal cancers with microsatellite stable (ie, lacking mismatch-repair deficiency) and Amsterdam Criteria positive were found to display a

significantly lower degree of long interspersed element-1 methylation as compared with other patient groups, ie, Lynch syndrome with identified mismatch-repair mutations and sporadic colorectal cancers with or without microsatellite instability. Long interspersed element-1 comprises a substantial portion of the human genome and its methylation correlates with global methylation status. Genome-wide DNA hypomethylation has an important role in genomic instability and carcinogenesis.<sup>91</sup> Additionally, it was suggested that cancers characterized by long interspersed element-1 extreme hypomethylation constitute a previously unrecognized and distinct subtype of colorectal cancer.<sup>92</sup> In addition to young onset (consistent with familial predisposition), colorectal cancers characterized by long interspersed element-1 hypomethylation consistently exhibited a poor prognosis.<sup>93,94</sup> Taken together, these findings further suggest molecular heterogeneity of familial colorectal cancer type X, with several possible molecular subtypes.

### Monogenic vs polygenic models for familial cancers

The fulfillment of Amsterdam Criteria has reinforced our view that familial colorectal cancer type X genetics has a monogenic component, ie, it is likely to be caused by high-penetrance mutations, similarly to the germline mutations in mismatch-repair genes responsible for Lynch syndrome. However, the clinical characteristics such as the late onset of disease and the lower risk of colorectal cancer suggest that familial colorectal cancer type X might also have a polygenic component (ie, involving interaction of multiple low-penetrance genetic variants) similarly to sporadic colorectal cancer cases, which have been investigated by genome-wide association studies. Elucidating the nature of the disease is important as it will influence the approaches adopted to interrogate the genetic basis of familial colorectal cancer type X.

A similar scenario has been noted in the context of another cancer; familial nasopharyngeal carcinoma cases are defined as multiple affected members within a family. However, there is little evidence to suggest that a familial form of nasopharyngeal carcinoma exists with characteristics markedly distinct from sporadic cases.<sup>95</sup> This conclusion has been supported by a more recent analysis that investigated the fit of single gene, polygenic and multifactorial models to the observed pattern of transmission of nasopharyngeal carcinoma in a hospital-based family history study, which also suggested a multifactorial mode of inheritance for nasopharyngeal carcinoma.<sup>96</sup> So far, no high-penetrance causal mutation or gene has been identified for familial nasopharyngeal carcinoma, although several susceptibility loci have been tentatively identified by linkage studies.<sup>97–99</sup> Similarly, familial

testicular germ cell tumors are defined as testicular germ cell tumors that have been diagnosed in at least two blood relatives. Nevertheless, previous linkage studies of multiple familial testicular germ cell tumor families did not reveal any high-penetrance genes and it has therefore been concluded that the combined effects of multiple common genetic variants, each conferring a modest risk, might underlie the disease.<sup>100</sup> By contrast, familial adenomatous polyposis, Lynch syndrome, and hereditary diffuse gastric cancer represent clear examples of monogenic cancers where high-penetrance causal mutations and genes have been identified. Similarly, hereditary pheochromocytoma is known to be caused by germline mutations in one of the nine genes, namely *RET*, *VHL*, *SDHA*, *SDHB*, *SDHC*, *SDHD*, *SDHAF2*, *NF1*, and *TMEM127*.<sup>17</sup> Although it is often assumed that monogenic models underlie the genetic etiology of familial cancers whereas polygenic models account for sporadic cancers, this is not always the case.

### How to interrogate the genetics of familial colorectal cancer type X?

The nature of the disease determines the study design required to unravel the causal mutations or risk-predisposing variants for familial colorectal cancer type X. However, there is little evidence to show whether familial colorectal cancer type X is a monogenic or polygenic disease or whether it is somewhere in between. The evidence suggesting that familial colorectal cancer type X is a monogenic disease comes mainly from the fulfillment of Amsterdam Criteria. The Amsterdam Criteria state that at least three relatives must have colorectal cancer. However, the familial aggregation, with multiple affected family members in one family, could also be due to shared non-genetic factors, which would not therefore necessarily be compatible with the monogenic model. Such environmental factors would be expected to interact with multiple genetic risk factors causing colorectal cancer, a multifactorial disease model proposed for polygenic disease. This therefore raises the question as to whether the Amsterdam Criteria are sufficient to support a monogenic basis for familial colorectal cancer type X. Furthermore, some of the clinical features of familial colorectal cancer type X implied that it could have a polygenic basis. This uncertainty in the nature of the disease for familial colorectal cancer type X presents substantial challenges in terms of deciding upon an optimal approach to interrogate its genetic basis.

The targeted sequencing of causal genes, already applied in the context of other familial cancers (such as *CDH1* (hereditary diffuse gastric cancer), *BRCA1* and *BRCA2* (familial breast cancer), and the genes underlying hereditary pheochromocytoma), appears to be a worthwhile approach to identify deleterious



germline mutations for familial colorectal cancer type X. The rationale is that germline mutations in these genes could underlie different familial cancers, as for example in the case of the *PALB2* germline mutations that have been found in both familial pancreatic and breast cancers.<sup>16,63</sup> Another notable example is provided by the germline mutations in the *BRCA2* gene that not only increase the risk of breast and ovarian cancer, but also pancreatic cancer.<sup>101</sup> This targeted sequencing approach has been greatly aided by high-throughput enrichment methods and next-generation sequencing technologies to selectively enrich for regions of interest. Hundreds of genes can be sequenced efficiently, leveraging these technological advances compared with traditional PCR-based Sanger sequencing. The efficiency of this approach has been exemplified in a targeted sequencing study of germline mutations in 21 tumor suppressor genes for 360 women with inherited ovarian, peritoneal, or fallopian tube carcinoma.<sup>102</sup> This study harnessed the power of the Sure-Select enrichment system and the Illumina sequencing platform to sequence these genes; 24% of the patients were found to carry germline loss-of-function mutations in 12 genes, six of which had not previously been implicated in inherited ovarian carcinoma. Although this targeted approach has limited discovery value, as these genes had already been implicated in causing familial cancers, it could still have some novelty value by identifying germline mutations in known genes for cancers, which have not yet been linked to these genes.

This targeted approach can be expanded to include the entire set of exons in all genes in the human genome. Exome sequencing on its own or coupled with linkage analysis has already unravelled multiple new causal mutations and genes for Mendelian disorders.<sup>7,8</sup> Furthermore, these discoveries were made by exome sequencing fewer than 10 patient samples in most of the studies reported. As such, it is also widely anticipated that exome sequencing will represent a powerful tool to reveal the genetic causes of familial colorectal cancer type X by identifying rare and deleterious or high-penetrance mutations within gene coding regions. However, the appropriate selection of cases will have a key role in determining the success or otherwise of exome sequencing in this context. In addition to fulfilling the Amsterdam Criteria, and excluding germline mutations in mismatch-repair genes, selecting cases with a very early onset of disease, severe clinico-pathological manifestations or the 'extreme' familial colorectal cancer type X phenotypes are expected to enrich for the 'monogenic' component and hence enhance our chances of identifying high-penetrance mutations. Recurrent mutations (similar mutations in different samples) or genes harboring several different deleterious mutations (which include single-nucleotide variants and small indels) across multiple samples can then be prioritized for further studies using a larger sample of cases.

On the other hand, if we assume that familial colorectal cancer type X has a polygenic component, then genome-wide association studies would represent the ideal approach to identify common single-nucleotide polymorphisms associated with this disease. Further, whole-genome genotyping arrays would also allow copy number variants to be investigated to a certain extent for their associations with familial colorectal cancer type X within a single genome-wide association study. High-density genotyping arrays have been used to identify copy number variants in a cohort of 41 colorectal cancer patients who were below 40 years of age at diagnosis and/or who exhibited an overt family history.<sup>103</sup> Multiple copy number variants, encompassing genes such as *CDH18*, *GREM1*, and *BCR*, were identified in six patients as well as two deletions encompassing two microRNA genes, *hsa-mir-491/KIAA1797* and *hsa-mir-646/AK309218*. Interestingly, these copy number variants had not previously been reported in relation to colorectal cancer predisposition, nor had they been encountered in large control cohorts. This illustrates the potential power of copy number variant investigation to identify novel causal or susceptibility genes or genetic loci for both familial and sporadic colorectal cancers. Through another interesting observation, multiple genomic aberrations including copy number gains and losses in different chromosomes have also been detected in 30 mismatch repair-proficient familial colorectal cancers. In particular, the frequency of 20q gain is remarkably increased when compared with sporadic colorectal cancer, suggesting that the 20q gain is involved in the genetic etiology of these mismatch repair-proficient familial colorectal cancers.<sup>104</sup> The finding that most of these genomic aberrations were also observed in sporadic colorectal cancer further suggests that familial and sporadic colorectal cancers could share genetic predisposition to a certain extent.

It is however noteworthy that genome-wide association studies represent an indirect association study design, based on linkage disequilibrium, to detect the disease-causing variants, as compared with direct sequencing. To achieve the required statistical power and significance threshold to detect common single-nucleotide polymorphisms conferring small effect sizes (odds ratio <1.5), several thousands of cases and controls are required for the initial genome-wide genotyping and subsequent replication studies.<sup>105</sup> Although the cost of genotyping arrays is steadily becoming much cheaper, a hefty investment is still required to analyze thousands of samples. In addition to this cost, collecting the adequate sample size of patients to embark on a genome-wide association study is a considerable challenge if this is to be achieved without an international consortium (because of the rarity of familial colorectal cancer type X as compared with sporadic colorectal cancer cases). The polygenic basis of familial colorectal cancer type X is still a speculative issue. Bearing in mind

this uncertainty, an alternative is to leverage the results from genome-wide association studies of colorectal cancer by genotyping the robust single-nucleotide polymorphism associations in a familial colorectal cancer type X cohort. This approach might be more feasible in terms of cost-effectiveness and sample size (without the need of a stringent significance threshold to account for several hundred thousand single-nucleotide polymorphisms). The penalty of multiple testing imposed in genome-wide association studies should increase the attractiveness of this approach in the context of testing single-nucleotide polymorphisms identified by genome-wide association studies for familial colorectal cancer type X. One may speculate that if familial colorectal cancer type X has a polygenic component, some of these polymorphisms should also be associated with familial colorectal cancer type X, which would then warrant a comprehensive genome-wide association study for familial colorectal cancer type X in the future. This speculation appears reasonable because common shared single-nucleotide polymorphisms or genetic loci have been found in several different cancers. There have been several examples of the practical utility of genome-wide association study results in the context of familial cancers. These studies have provided evidence to suggest that low-penetrance variants may explain the increased cancer risk in familial colorectal cancer<sup>106–108</sup> and in familial testicular germ cell tumors.<sup>100</sup>

Finally, the genes or genetic loci implicated in colorectal cancer by genome-wide association studies can be captured and sequenced. This targeted sequencing approach is very cost-effective as up to 96 samples can be multiplexed through barcoding for massively parallel sequencing. This targeted sequencing approach will interrogate both rare variants and common single-nucleotide polymorphisms in the loci identified by genome-wide association studies. The promise of this approach in unravelling rare variants in loci implicated by genome-wide association studies has already been demonstrated.<sup>51,53–55</sup> For example, deep resequencing of such loci has identified independent rare variants associated with inflammatory bowel disease.<sup>55</sup>

## Perspectives and conclusions

The genetic and clinical differences between Lynch syndrome and familial colorectal cancer type X have been well documented. However, the genetic etiologies of familial colorectal cancer type X remain to be determined. There is also a paucity of evidence to indicate one way or the other whether familial colorectal cancer type X is a monogenic or a polygenic disease. On the other hand, the genetics of sporadic/polygenic colorectal cancer have been comprehensively investigated by >10 genome-wide association studies over the past few years. One

striking observation is the sharing of common single-nucleotide polymorphisms or genetic loci across different cancers. It is therefore reasonable to speculate that if familial colorectal cancer type X has a polygenic basis, some of the single-nucleotide polymorphisms identified by genome-wide association studies as conferring risk of colorectal cancer might be expected to show associations with familial colorectal cancer type X as well. Given the expense and logistic challenges involved in collecting a large number of familial colorectal cancer type X cases to embark on a genome-wide association study, together with the uncertainty of the disease model, we believe that the genotyping of genome-wide association study-identified single-nucleotide polymorphisms in familial colorectal cancer type X would be a more feasible first approach to explore the genetic etiology of this disease. However, given the low incidence of familial colorectal cancer type X (ie, only ~2–3% of colorectal cancer families meet Amsterdam Criteria and about half of these are Lynch syndrome cases), collecting an adequate large sample size is difficult and challenging especially for studying the association of single-nucleotide polymorphisms with modest effect sizes. Thus, National or International Consortia involving many centers are likely to be needed to recruit large numbers of patients. Alternatively, the genes or loci identified by genome-wide association studies could be investigated using a targeted sequencing approach to unravel rare variants of larger effect size.

One of the limitations of genome-wide association studies is that they are based upon an indirect association study design, which is reliant on linkage disequilibrium to identify the disease functional variants. As a result, the surrogate markers (ie, the associated single-nucleotide polymorphisms) identified by genome-wide association studies generally lack functional significance. Furthermore, to enhance the statistical power, genome-wide association studies have tended to lump all colorectal cancers in the disease group, even although it is well recognized that colorectal cancers are inherently heterogeneous. These challenges have led to the notion and conceptualization of ‘molecular pathological investigation’, which is a relatively new field of epidemiology based upon the molecular classification of cancer.<sup>109</sup> It is a multidisciplinary field involving the investigation of the interrelationship between exogenous and endogenous (eg, genetic) factors, tumoral molecular signatures, and tumor progression. Further, integrating genome-wide association studies with molecular pathological investigation allows examination of the relationship between susceptibility alleles identified by genome-wide association studies and specific molecular alterations/subtypes, which can help to elucidate the function of these alleles and provide insights into whether the detected susceptibility alleles are truly causal. Although there are challenges, molecular pathological epidemiology has unique

strengths, and can provide insights into the pathogenic process.

In addition, exome sequencing of multiple 'well-selected' cases could be performed, assuming a monogenic basis in which high-penetrance mutations are predicted to underlie the genetic etiology of familial colorectal cancer type X. Exome sequencing of families with multiple affected individuals also represents a promising study design. This family-based design has the advantage that it allows for the genetically heterogeneous nature of familial colorectal cancer type X. Comparing unrelated individuals or probands from different families to identify 'common/shared' putative pathological variants or genes harboring putative pathological variants might not be a successful strategy for genetically heterogeneous diseases. However, it still depends on the degree of genetic heterogeneity (ie, allelic heterogeneity versus locus heterogeneity) characterizing the disease and this remains unknown. Although the family design is robust with respect to genetic heterogeneity (comparing affected and unaffected members in a family), one must recognize that it could also be problematic because the penetrance of disease mutations for familial colorectal cancer type X is likely to be lower than that for Lynch syndrome.

Moving forward, it is arguable that whole-genome sequencing should probably be considered instead of exome sequencing, as the cost differential between the two approaches (given a small patient sample size) would not be substantial, and because the former approach will generate genetic data for the entire genome rather than just 1–2% as for exome sequencing. However, one should select the study design that best fits the hypothesis where rare deleterious mutations in coding regions underlie the genetic etiology of a Mendelian disorder or familial cancer. So far, all the discoveries made by whole-genome sequencing could also have been achieved using exome sequencing for Mendelian disorders.<sup>110,111</sup> Furthermore, the genetic variants in most of the non-coding regions revealed by whole-genome sequencing remain 'uninterpretable' biologically. In taking a practical (rather than theoretical) point of view, whole-genome sequencing still presents a very substantial technical challenge as well as a challenge in terms of analyzing and interpreting the sequence data generated.

The disease models underpinning multiple familial cancers such as familial nasopharyngeal carcinoma,<sup>112</sup> familial testicular germ cell tumor,<sup>113</sup> familial chronic lymphocytic leukemia,<sup>114</sup> and familial colorectal cancer (familial colorectal cancer type X)<sup>29</sup> remain contentious as the high-penetrance mutations are yet to be identified. By contrast, multiple low-penetrance variants that confer an effect size of odds ratio <1.5 have been revealed through genome-wide association studies for the sporadic cases of these cancers; interestingly, some of these single-nucleotide polymorphisms have also been found to be associated with the familial cases

(nasopharyngeal carcinoma,<sup>115</sup> testicular germ cell tumor,<sup>116,117</sup> chronic lymphocytic leukemia,<sup>118,119</sup> and colorectal cancer<sup>120</sup>). In the context of familial colorectal cancer type X, we believe that the disease model and its genetic basis are likely to become more apparent when the approaches that we have outlined and discussed are applied in practice. This should facilitate the iterative interrogation of the genetics of familial colorectal cancer type X and other familial cancers of similar nature before embarking on either a comprehensive genome-wide association studies or whole-genome sequencing approach.

## Disclosure/conflict of interest

The authors declare no conflict of interest.

**Author contributions:** CSK, DNC, and BI contributed to the conceptualization of this article. CSK and DNC contributed to the writing of the article. MW, DHR, YP, and RS were involved in the discussion and critical reading. CSK, DNC, and BI approved the final version and had final responsibility for this article.

## References

- 1 Asan NF, Xu Y, Jiang H, *et al*. Comprehensive comparison of three commercial human whole-exome capture platforms. *Genome Biol* 2011;12:R95.
- 2 Clark MJ, Chen R, Lam HY, *et al*. Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol* 2011;29:908–914.
- 3 Parla JS, Iossifov I, Grabill I, *et al*. A comparative analysis of exome capture. *Genome Biol* 2011;12:R97.
- 4 Sulonen AM, Ellonen P, Almusa H, *et al*. Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol* 2011;12:R94.
- 5 Bamshad MJ, Ng SB, Bigham AW, *et al*. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011;12:745–755.
- 6 Gilissen C, Hoischen A, Brunner HG, *et al*. Unlocking Mendelian disease using exome sequencing. *Genome Biol* 2011;12:228.
- 7 Ku CS, Naidoo N, Pawitan Y. Revisiting Mendelian disorders through exome sequencing. *Hum Genet* 2011;129:351–370.
- 8 Ng SB, Nickerson DA, Bamshad MJ, *et al*. Massively parallel sequencing and rare disease. *Hum Mol Genet* 2010;19:R119–R124.
- 9 Ng SB, Turner EH, Robertson PD, *et al*. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;461:272–276.
- 10 Choi M, Scholl UI, Ji W, *et al*. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA* 2009;106:19096–19101.
- 11 Ku CS, Cooper DN, Polychronakos C, *et al*. Exome sequencing: dual role as a discovery and diagnostic tool. *Ann Neurol* 2012;71:5–14.
- 12 Robinson PN, Krawitz P, Mundlos S. Strategies for exome and genome sequence data analysis in disease-gene discovery projects. *Clin Genet* 2011;80:127–132.

- 13 Stitzel NO, Kiezun A, Sunyaev S. Computational and statistical approaches to analyzing variants identified by exome sequencing. *Genome Biol* 2011;12:227.
- 14 Sathirapongsasuti JF, Lee H, Horst BA, *et al*. Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics* 2011;27:2648–2654.
- 15 Smith KR, Bromhead CJ, Hildebrand MS, *et al*. Reducing the exome search space for Mendelian diseases using genetic linkage analysis of exome genotypes. *Genome Biol* 2011;12:R85.
- 16 Jones S, Hruban RH, Kamiyama M, *et al*. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science* 2009;324:217.
- 17 Comino-Mendez I, Gracia-Aznarez FJ, Schiavi F, *et al*. Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nat Genet* 2011;43:663–667.
- 18 Chung CC, Chanock SJ. Current status of genome-wide association studies in cancer. *Hum Genet* 2011;130:59–78.
- 19 Varghese JS, Easton DF. Genome-wide association studies in common cancers—what have we learnt? *Curr Opin Genet Dev* 2010;20:201–209.
- 20 Guilford P, Hopkins J, Harraway J, *et al*. E-cadherin germline mutations in familial gastric cancer. *Nature* 1998;392:402–405.
- 21 Kaurah P, MacMillan A, Boyd N, *et al*. Founder and recurrent CDH1 mutations in families with hereditary diffuse gastric cancer. *JAMA* 2007;297:2360–2372.
- 22 Fanale D, Amodeo V, Corsini LR, *et al*. Breast cancer genome-wide association studies: there is strength in numbers. *Oncogene*; published online 26 September 2011 [e-pub ahead of print].
- 23 Ghossaini M, Fletcher O, Michailidou K, *et al*. Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet* 2012;44:312–318.
- 24 Antoniou AC, Kartsonaki C, Sinilnikova OM, *et al*. Common alleles at 6q25.1 and 1p11.2 are associated with breast cancer risk for BRCA1 and BRCA2 mutation carriers. *Hum Mol Genet* 2011;20:3304–3321.
- 25 Jostins L, Barrett JC. Genetic risk prediction in complex disease. *Hum Mol Genet* 2011;20:R182–R188.
- 26 Hartman M, Suo C, Lim WY, *et al*. Ability to predict breast cancer in Asian women using a polygenic susceptibility model. *Breast Cancer Res Treat* 2011;127:805–812.
- 27 Gala M, Chung DC. Hereditary colon cancer syndromes. *Semin Oncol* 2011;38:490–499.
- 28 Jaspersion KW, Tuohy TM, Neklason DW, *et al*. Hereditary and familial colon cancer. *Gastroenterology* 2010;138:2044–2058.
- 29 Lindor NM, Rabe K, Petersen GM, *et al*. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. *JAMA* 2005;293:1979–1985.
- 30 Antonarakis SE, Chakravarti A, Cohen JC, *et al*. Mendelian disorders and multifactorial traits: the big divide or one for all? *Nat Rev Genet* 2010;11:380–384.
- 31 Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 2010;11:415–425.
- 32 Carvajal-Carmona LG. Challenges in the identification and use of rare disease-associated predisposition variants. *Curr Opin Genet Dev* 2010;20:277–281.
- 33 Plomin R, Haworth CM, Davis OS. Common disorders are quantitative traits. *Nat Rev Genet* 2009;10:872–878.
- 34 McCarthy MI, Abecasis GR, Cardon LR, *et al*. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008;9:356–369.
- 35 Badano JL, Katsanis N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nat Rev Genet* 2002;3:779–789.
- 36 Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 2003;33(Suppl):228–237.
- 37 Dean M. Approaches to identify genes for complex human diseases: lessons from Mendelian disorders. *Hum Mutat* 2003;22:261–274.
- 38 Antoniou AC, Chenevix-Trench G. Common genetic variants and cancer risk in Mendelian cancer syndromes. *Curr Opin Genet Dev* 2010;20:299–307.
- 39 Cutting GR. Modifier genes in Mendelian disorders: the example of cystic fibrosis. *Ann NY Acad Sci* 2010;1214:57–69.
- 40 Milne RL, Antoniou AC. Genetic modifiers of cancer risk for BRCA1 and BRCA2 mutation carriers. *Ann Oncol* 2011;22(Suppl 1):i11–i17.
- 41 Talseth-Palmer BA, Brenne IS, Ashton KA, *et al*. Colorectal cancer susceptibility loci on chromosome 8q23.3 and 11q23.1 as modifiers for disease expression in Lynch syndrome. *J Med Genet* 2011;48:279–284.
- 42 Wijnen JT, Brohet RM, van Eijk R, *et al*. Chromosome 8q23.3 and 11q23.1 variants modify colorectal cancer risk in Lynch syndrome. *Gastroenterology* 2009;136:131–137.
- 43 Manolio TA, Collins FS, Cox NJ, *et al*. Finding the missing heritability of complex diseases. *Nature* 2009;461:747–753.
- 44 Kinzler KW, Nilbert MC, Su LK, *et al*. Identification of FAP locus genes from chromosome 5q21. *Science* 1991;253:661–665.
- 45 Nishisho I, Nakamura Y, Miyoshi Y, *et al*. Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science* 1991;253:665–669.
- 46 Walsh T, Lee MK, Casadei S, *et al*. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc Natl Acad Sci USA* 2010;107:12629–12633.
- 47 Mardis ER. The \$1000 genome, the \$100 000 analysis? *Genome Med* 2010;2:84.
- 48 Sboner A, Mu XJ, Greenbaum D, *et al*. The real cost of sequencing: higher than you think!. *Genome Biol* 2011;12:125.
- 49 Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005;6:95–108.
- 50 Schork NJ, Murray SS, Frazer KA, *et al*. Common vs rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* 2009;19:212–219.
- 51 Johansen CT, Wang J, Lanktree MB, *et al*. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* 2010;42:684–687.
- 52 Li Y, Vinckenbosch N, Tian G, *et al*. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet* 2010;42:969–972.

- 53 Momozawa Y, Mni M, Nakamura K, *et al*. Resequencing of positional candidates identifies low frequency IL23R coding variants protecting against inflammatory bowel disease. *Nat Genet* 2011;43:43–47.
- 54 Nejentsev S, Walker N, Riches D, *et al*. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 2009;324:387–389.
- 55 Rivas MA, Beaudoin M, Gardet A, *et al*. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* 2011;43:1066–1073.
- 56 Clarke AJ, Cooper DN. GWAS: heritability missing in action? *Eur J Hum Genet* 2010;18:859–861.
- 57 Singleton A, Hardy J. A generalizable hypothesis for the genetic architecture of disease: pleomorphic risk loci. *Hum Mol Genet* 2011;20:R158–R162.
- 58 Gui Y, Guo G, Huang Y, *et al*. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet* 2011;43:875–878.
- 59 Li M, Zhao H, Zhang X, *et al*. Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. *Nat Genet* 2011;43:828–829.
- 60 Varela I, Tarpey P, Raine K, *et al*. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* 2011;469:539–542.
- 61 Wei X, Walia V, Lin JC, *et al*. Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nat Genet* 2011;43:442–446.
- 62 Yan XJ, Xu J, Gu ZH, *et al*. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet* 2011;43:309–315.
- 63 Rahman N, Seal S, Thompson D, *et al*. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* 2007;39:165–167.
- 64 Reid S, Schindler D, Hanenberg H, *et al*. Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. *Nat Genet* 2007;39:162–164.
- 65 Balia C, Sensi E, Lombardi G, *et al*. PALB2: a novel inactivating mutation in a Italian breast cancer family. *Fam Cancer* 2010;9:531–536.
- 66 Casadei S, Norquist BM, Walsh T, *et al*. Contribution of inherited mutations in the BRCA2-interacting protein PALB2 to familial breast cancer. *Cancer Res* 2011;71:2222–2229.
- 67 Papi L, Putignano AL, Congregati C, *et al*. A PALB2 germline mutation associated with hereditary breast cancer in Italy. *Fam Cancer* 2010;9:181–185.
- 68 Jass JR. Hereditary non-polyposis colorectal cancer: the rise and fall of a confusing term. *World J Gastroenterol* 2006;12:4943–4950.
- 69 Iacopetta B, Grieu F, Amanuel B. Microsatellite instability in colorectal cancer. *Asia Pac J Clin Oncol* 2010;6:260–269.
- 70 Hesson LB, Hitchins MP, Ward RL. Epimutations and cancer predisposition: importance and mechanisms. *Curr Opin Genet Dev* 2010;20:290–298.
- 71 Hitchins MP, Ward RL. Constitutional (germline) MLH1 epimutation as an aetiological mechanism for hereditary non-polyposis colorectal cancer. *J Med Genet* 2009;46:793–802.
- 72 Hitchins MP, Wong JJ, Suthers G, *et al*. Inheritance of a cancer-associated MLH1 germ-line epimutation. *N Engl J Med* 2007;356:697–705.
- 73 Niessen RC, Hofstra RM, Westers H, *et al*. Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. *Genes Chromosomes Cancer* 2009;48:737–744.
- 74 Nagasaka T, Rhee J, Kloor M, *et al*. Somatic hypermethylation of MSH2 is a frequent event in Lynch Syndrome colorectal cancers. *Cancer Res* 2010;70:3098–3108.
- 75 Rumilla K, Schowalter KV, Lindor NM, *et al*. Frequency of deletions of EPCAM (TACSTD1) in MSH2-associated Lynch syndrome cases. *J Mol Diagn* 2011;13:93–99.
- 76 Oliveira C, Seruca R, Carneiro F. Hereditary gastric cancer. *Best Pract Res Clin Gastroenterol* 2009;23:147–157.
- 77 Pollock J, Welsh JS. Clinical cancer genetics: Part I: Gastrointestinal. *Am J Clin Oncol* 2011;34:332–336.
- 78 Francisco I, Albuquerque C, Lage P, *et al*. Familial colorectal cancer type X syndrome: two distinct molecular entities? *Fam Cancer* 2011;10:623–631.
- 79 Klarskov L, Holck S, Bernstein I, *et al*. Hereditary colorectal cancer diagnostics: morphological features of familial colorectal cancer type X *versus* Lynch syndrome. *J Clin Pathol* 2012;65:352–356.
- 80 Llor X, Pons E, Xicola RM, *et al*. Differential features of colorectal cancers fulfilling Amsterdam criteria without involvement of the mutator pathway. *Clin Cancer Res* 2005;11:7304–7310.
- 81 Mueller-Koch Y, Vogelsang H, Kopp R, *et al*. Hereditary non-polyposis colorectal cancer: clinical and molecular evidence for a new entity of hereditary colorectal cancer. *Gut* 2005;54:1733–1740.
- 82 Ogino S, Goel A. Molecular classification and correlates in colorectal cancer. *J Mol Diagn* 2008;10:13–27.
- 83 Curtin K, Slattery ML, Samowitz WS. CpG island methylation in colorectal cancer: past, present and future. *Patholog Res Int* 2011;2011:902674.
- 84 Hughes LA, Khalid-de Bakker CA, Smits KM, *et al*. The CpG island methylator phenotype in colorectal cancer: progress and problems. *Biochim Biophys Acta* 2012;1825:77–85.
- 85 Ogino S, Kawasaki T, Kirkner GJ, *et al*. CpG island methylator phenotype-low (CIMP-low) in colorectal cancer: possible associations with male sex and KRAS mutations. *J Mol Diagn* 2006;8:582–588.
- 86 Yagi K, Akagi K, Hayashi H, *et al*. Three DNA methylation epigenotypes in human colorectal cancer. *Clin Cancer Res* 2010;16:21–33.
- 87 Hinoue T, Weisenberger DJ, Lange CP, *et al*. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res* 2012;22:271–282.
- 88 Ogino S, Kawasaki T, Kirkner GJ, *et al*. Molecular correlates with MGMT promoter methylation and silencing support CpG island methylator phenotype-low (CIMP-low) in colorectal cancer. *Gut* 2007;56:1564–1571.
- 89 Kawasaki T, Ohnishi M, Noshio K, *et al*. CpG island methylator phenotype-low (CIMP-low) colorectal cancer shows not only few methylated CIMP-high-specific CpG islands, but also low-level methylation at individual loci. *Mod Pathol* 2008;21:245–255.
- 90 Woods MO, Younghusband HB, Parfrey PS, *et al*. The genetic basis of colorectal cancer in a population-based incident cohort with a high rate of familial disease. *Gut* 2010;59:1369–1377.
- 91 Goel A, Xicola RM, Nguyen TP, *et al*. Aberrant DNA methylation in hereditary nonpolyposis colorectal

- cancer without mismatch repair deficiency. *Gastroenterology* 2010;138:1854–1862.
- 92 Baba Y, Huttenhower C, Noshu K, *et al*. Epigenomic diversity of colorectal cancer indicated by LINE-1 methylation in a database of 869 tumors. *Mol Cancer* 2010;9:125.
  - 93 Ogino S, Noshu K, Kirkner GJ, *et al*. A cohort study of tumoral LINE-1 hypomethylation and prognosis in colon cancer. *J Natl Cancer Inst* 2008;100:1734–1738.
  - 94 Ahn JB, Chung WB, Maeda O, *et al*. DNA methylation predicts recurrence from resected stage III proximal colon cancer. *Cancer* 2011;117:1847–1854.
  - 95 Ung A, Chen CJ, Levine PH, *et al*. Familial and sporadic cases of nasopharyngeal carcinoma in Taiwan. *Anticancer Res* 1999;19:661–665.
  - 96 Jia WH, Collins A, Zeng YX, *et al*. Complex segregation analysis of nasopharyngeal carcinoma in Guangdong, China: evidence for a multifactorial mode of inheritance (complex segregation analysis of NPC in China). *Eur J Hum Genet* 2005;13:248–252.
  - 97 Feng BJ, Huang W, Shugart YY, *et al*. Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4. *Nat Genet* 2002;31:395–399.
  - 98 Xiong W, Zeng ZY, Xia JH, *et al*. A susceptibility locus at chromosome 3p21 linked to familial nasopharyngeal carcinoma. *Cancer Res* 2004;64:1972–1974.
  - 99 Hu LF, Qiu QH, Fu SM, *et al*. A genome-wide scan suggests a susceptibility locus on 5p 13 for nasopharyngeal carcinoma. *Eur J Hum Genet* 2008;16:343–349.
  - 100 Kratz CP, Han SS, Rosenberg PS, *et al*. Variants in or near KITLG, BAK1, DMRT1, and TERT-CLPTM1L predispose to familial testicular germ cell tumour. *J Med Genet* 2011;48:473–476.
  - 101 Hruban RH, Canto MI, Goggins M, *et al*. Update on familial pancreatic cancer. *Adv Surg* 2010;44:293–311.
  - 102 Walsh T, Casadei S, Lee MK, *et al*. Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc Natl Acad Sci USA* 2011;108:18032–18037.
  - 103 Venkatasubramanian R, Verwiel ET, Kamping EJ, *et al*. Identification of candidate predisposing copy number variants in familial and early-onset colorectal cancer patients. *Int J Cancer* 2011;129:1635–1642.
  - 104 Middeldorp A, van Eijk R, Oosting J, *et al*. Increased frequency of 20q gain and copy-neutral loss of heterozygosity in mismatch repair proficient familial colorectal carcinomas. *Int J Cancer* 2011;130:837–846.
  - 105 Ioannidis JP, Castaldi P, Evangelou E. A compendium of genome-wide associations for cancer: critical synopsis and reappraisal. *J Natl Cancer Inst* 2010;102:846–858.
  - 106 Middeldorp A, Jagmohan-Changur S, van Eijk R, *et al*. Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. *Cancer Epidemiol Biomarkers Prev* 2009;18:3062–3067.
  - 107 Middeldorp A, Jagmohan-Changur SC, van der Klift HM, *et al*. Comprehensive genetic analysis of seven large families with mismatch repair proficient colorectal cancer. *Genes Chromosomes Cancer* 2010;49:539–548.
  - 108 Niittymäki I, Kaasinen E, Tuupainen S, *et al*. Low-penetrance susceptibility variants in familial colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 2010;19:1478–1483.
  - 109 Ogino S, Chan AT, Fuchs CS, *et al*. Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. *Gut* 2011;60:397–411.
  - 110 Lupski JR, Reid JG, Gonzaga-Jauregui C, *et al*. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 2010;362:1181–1191.
  - 111 Rios J, Stein E, Shendure J, *et al*. Identification by whole-genome resequencing of gene defect responsible for severe hypercholesterolemia. *Hum Mol Genet* 2010;19:4313–4318.
  - 112 Zeng YX, Jia WH. Familial nasopharyngeal carcinoma. *Semin Cancer Biol* 2002;12:443–450.
  - 113 Greene MH, Kratz CP, Mai PL, *et al*. Familial testicular germ cell tumors in adults: 2010 summary of genetic risk factors and clinical phenotype. *Endocr Relat Cancer* 2010;17:R109–R121.
  - 114 Goldin LR, Slager SL, Caporaso NE. Familial chronic lymphocytic leukemia. *Curr Opin Hematol* 2010;17:350–355.
  - 115 Bei JX, Li Y, Jia WH, *et al*. A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. *Nat Genet* 2010;42:599–603.
  - 116 Rapley EA, Turnbull C, Al Olama AA, *et al*. A genome-wide association study of testicular germ cell tumor. *Nat Genet* 2009;41:807–810.
  - 117 Turnbull C, Rapley EA, Seal S, *et al*. Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. *Nat Genet* 2010;42:604–607.
  - 118 Crowther-Swanepoel D, Broderick P, Di Bernardo MC, *et al*. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. *Nat Genet* 2010;42:132–136.
  - 119 Di Bernardo MC, Crowther-Swanepoel D, Broderick P, *et al*. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nat Genet* 2008;40:1204–1210.
  - 120 Houlston RS, Cheadle J, Dobbins SE, *et al*. 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.