HHS Public Access

Author manuscript

Gut. Author manuscript; available in PMC 2016 October 01.

Published in final edited form as:

Gut. 2015 October ; 64(10): 1623–1636. doi:10.1136/gutjnl-2013-306705.

Genetic Architecture of Colorectal Cancer

Ulrike Peters[1,2], **Stephanie Bien**[1], and **Niha Zubair**[1]

[1]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

[2]Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, USA

Abstract

Colorectal cancer (CRC) is a complex disease that develops as a consequence of both genetic and environmental risk factors. A small proportion (3–5%) of cases arises from hereditary syndromes predisposing to early onset CRC as a result of mutations in over a dozen well-defined genes. In contrast, CRC is predominantly a late-onset "sporadic" disease, developing in individuals with no obvious hereditary syndrome. In recent years genome-wide association studies have discovered over 40 genetic regions to be associated with weak effects on sporadic CRC and it has been estimated that increasingly large genome-wide scans will identify many additional novel genetic regions. Subsequent experimental validations have identified the causally related variant(s) in a limited number of these genetic regions. Further biological insight could be obtained through ethnically diverse study populations, larger genetic sequencing studies, and development of higher-throughput functional experiments. Along with inherited variation, integration of the tumour genome may shed light on the carcinogenic processes in CRC. In addition to summarizing the genetic architecture of CRC, this review discusses genetic factors that modify environmental predictors of CRC, as well as examples of how genetic insight has improved clinical surveillance, prevention, and treatment strategies. In summary, substantial progress has been made in uncovering the genetic architecture of CRC and continued research efforts are expected to identify additional genetic risk factors that further our biological understanding of this disease.

Keywords

colorectal cancer; genetic association; genetic architecture; single nucleotide polymorphism; risk factors

Corresponding Authors: Ulrike Peters, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, M4-B402, PO Box 19024, Seattle, WA 98109, upeters@fhcrc.org, Phone (206) 667-2450, Fax (206) 667-7850. Stephanie Bien, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, M4-B402, PO Box 19024, Seattle, WA 98109, srosse@fredhutch.org, Phone (206) 667-7030, Fax (206) 667-4787.

CONTRIBUTORSHIP

Ulrike Peters, Stephanie Bien, and Niha Zubair were involved in manuscript writing.

COMPETING INTERESTS

The authors have no competing interests to disclose.

INTRODUCTION

It is estimated that in 2015 there will be 777,987 new cases and 352,589 deaths from CRC in developed countries.[1] The average lifetime risk in these populations varies across countries ranging from 4.3–5.3% for men and 2.7–4.9% for women.[2,3]

Inherited susceptibility is a major component of CRC predisposition with an estimated 12–35% of risk attributed to genetic factors.[4,5] Over the last two decades substantial progress has been made towards uncovering the genetic architecture of CRC, and yet there remains great opportunity for discovering additional variants. The genetic risk factors established thus far range between two extremes: (1) rare high-penetrance mutations, each conferring marked elevations in risk for hereditary syndromes to (2) common variants, also called polymorphisms, conferring weak effects on “sporadic” risk in individuals without family history of CRC (Figure 1). Revealing genetic factors underlying high-penetrance syndromes has led to more effective disease management for patients and their families. Further discovery of risk loci with weak effects could similarly improve clinical surveillance and prevention strategies. Although each common variant associates with weaker effects, collectively these variants enable a more accurate prediction of an individual’s risk given that the number of risk alleles carried by an individual varies substantially in a population. Moreover, common variants may modify the risk of CRC in individuals with hereditary syndromes.[6] In addition to personalized risk prediction, a deeper understanding of genetic etiology often implicates novel carcinogenic pathways and in turn new potential targets for therapeutics.

This review begins with a summary of what is known about the genetic architecture of both rare hereditary syndromes and more common sporadic development of CRC. In addition, new approaches to investigate rarer variation, as well as the studies that integrate the tumor genome will be reviewed. Given that risk prediction and biological insight are improved by the identification of functional variants within associated regions, this review will also describe the current state of laboratory follow-up studies. Next, several noteworthy gene-environment interactions with suggestive influences on CRC susceptibility are summarized. Lastly, the importance of translating genetic findings into clinical and public health practices is discussed.

GENETIC MUTATIONS PREDISPOSING TO HEREDITARY SYNDROMES

Hereditary syndromes resulting from high-penetrance germline mutations account for approximately 3–5% of all CRC.[7] Although rare, the mutations underlying these conditions were readily detected through relatively small linkage studies. To date, 14 genes underlying CRC syndromes have been identified (Table 1), beginning with the discovery of mutations in the *adenomatous polyposis coli (APC)* gene predisposing to familial adenomatous polyposis (FAP).[8] Later, the human homologs of the DNA mismatch repair (MMR) genes (*MLH1, MSH2, MSH6, PMS2*) were implicated in a non-polyposis familial condition now referred to as Lynch Syndrome.[9] Subsequently, mutations in the genes *STK11*,[10] *BMPR1A*,[11] *SMAD4*,[11] *PTEN*,[12] and *MUTYH*,[13] have been identified as additional genetic causes of polyposis syndromes. These genes highlight several

important molecular pathways, many of which are now thought to play a larger role in CRC pathogenesis supported by results from genome-wide association studies (GWAS) (Figure 1, Table 3). Previous genetic reviews of these syndromes [7,14,15] have discussed clinical management and screening strategies in greater detail. Here we describe the genetic etiologies and implicated pathways of hereditary syndromes.

Adenomatous Polyposis Coli (APC) and β–catenin

In 1987 the *APC* tumor suppressor gene was found to harbor causative germline mutations for the most severe polyposis syndrome, FAP.[8] Development of FAP requires the inheritance of a single mutated copy of *APC* and is characterized by the onset of hundreds to thousands of small adenomatous polyps throughout the entire length of the colon after the first decade of life. If left untreated, the risk of CRC by age 40 is nearly 100%. Decreased APC function is now understood to play an important role in colorectal tumorigenesis via aberrant regulation of intracellular levels of β–catenin (encoded by *CTNNB1*) within the wingless signal transduction (Wnt) pathway.[14,16] This pathway controls cell division, adhesion, and migration making it particularly important for cells with rapid turnover, such as intestinal epithelial that renew every 4–5 days.[17,18]

Classic FAP results from deleterious *APC* mutations that are typically positioned in or near functional domains that bind β–catenin. However, recent studies suggest that decreased transcription of *APC* through promoter mutations may also result in FAP predisposition.[19] Alternatively, mutations in the 5′ and 3′ ends of *APC* can result in a less profuse polyposis syndrome termed attenuated FAP (AFAP).[20,21] In comparison to the classic syndrome, AFAP has a delayed age of onset (mean 56 years) for CRC with reduced average lifetime risk of approximately 70%.[15,22,23]

MutY Homolog (MUTYH)

Unlike other hereditary syndromes, the inheritance of mutations in both copies of *MUTYH* (alias *MYH*) gene result in a recessive form of adenomatous polyposis, referred to as *MUTYH*-associated polyposis (MAP).[22] The *MUTYH* gene is involved in base excision repair of oxidative DNA damage. In some cases MAP is phenotypically indistinguishable from FAP or AFAP with an average lifetime CRC risk of about 80%.[7] The MAP carcinogenic pathway often involves a high frequency of somatically acquired *APC* mutations.[24]

Mismatch Repair Genes

Over the past two decades defective MMR genes have been discovered to play a role in subsets of both hereditary and sporadic CRC.[25] Specifically, Lynch syndrome (also known as hereditary nonpolyposis colorectal cancer, HNPCC), is the most common of the hereditary syndromes, accounting for 2–3% of all CRC cases [26]. Lynch syndrome is characterized by early-onset CRC (mean of 45 years) and an average lifetime risk of 66% for men and about 43% for women.[27] Lynch syndrome results from germline mutations in genes involved with DNA MMR (*MLH1*, *MSH2*, *MSH6*, *PMS2, and EPCAM*).[15,28] Loss of MMR activity leads to defective repair of single-base mismatches and insertion/deletions during DNA replication. The subsequent errors in replication results in the accumulation of

repetitive short nucleotide sequences referred to as microsatellites. Accordingly, high Microsatellite Instability (MSI) is a hallmark and critical component for the diagnosis of Lynch syndrome, although it should be noted that also approximately 12% of sporadic CRC are characterized by high MSI.[29]

DNA Polymerase Genes

DNA replication during cell division is inherently susceptible to errors that can be transmitted to daughter cells and incorporated as permanent mutations, which in turn can predispose to cancer. However, there are several conserved mechanisms to safeguard against replication error and subsequent somatic mutations. The role of such mechanisms in CRC risk have already been discussed in this review, such as mutations in base excision repair (MAP) and those in MMR genes (Lynch Syndrome). Recent discovery of germline mutations in the proofreading domains of two DNA polymerases (POLE and POLD1) now implicate a new highly penetrant hereditary syndrome referred to as 'polymerase proofreading-associated polyposis' (PAPP) leading to a large number of somatically acquired mutations (hypermutant tumors).[30–33] These discoveries reinforce the importance of mechanisms related to correct DNA replication, and show that reduced fidelity can result in a mutator phenotype increasing cancer susceptibility.

Serine/Threonine Protein Kinase 11 (STK11) and Phosphatase Tensin Homolog Deleted on Chromosome 10 (PTEN)

The harmatomatous polyposis syndromes (Peutz-Jeghers syndrome (PJS), Cowden Syndrome (CS), a subtype of CS (Bannayan-Riley-Ruvalcaba (BRR) syndrome) and Hereditary Mixed Polyposis Syndrome (HMPS) are very rare with frequencies of approximately 1 in 50,000 to 200,000 for PJS, and 1 in 200,000 to 250,000 births for CS, while the prevalence of BRR and HMPS remain unknown.[34]

STK11 and *PTEN* have been identified as genes underlying PJS, CS and BRR syndrome. *STK11* (alias *LKB1*) encodes a tumor suppressor that is activated in PJS and is related to cellular energy homeostasis and regulation of the mammalian target of rapamycin (mTOR) —a pathway that is central to metabolic signaling. In addition, STK11 governs whole body insulin sensitivity.[35] Similarly, *PTEN* is linked to CS, as well as BRR syndrome, and regulates metabolic signaling via the phosphatidylinositol 3-kinase (PI3K) and the v-akt murine thymoma viral oncogene homolog 1 (Akt1) pathway.[36] Moreover, the PI3k-Akt pathway is an important regulator of cell proliferation and is thought to mediate the effects of mTOR. As such, at least part of the activities of STK11 and PTEN are expected to converge through the mTOR pathway.

TGF-β superfamily

Other germline mutations linked to hereditary CRC syndromes, such as Juvenile Polyposis Syndrome (JPS) and Hereditary Mixed Polyposis Syndrome (HMPS) implicate genes that enhance growth and invasiveness, such as those belonging to the transforming growth factor (TGF)-β superfamily and the bone morphogenic protein (BMP) subfamilies (eg. *SMAD4,* [37,38] *BMPR1A*,[39,40] *BMP4*,[41] *GREM1*,[42] *ENG*[43,44]). The BMPs play a critical role in orchestrating proper tissue architecture through their interaction with specific surface

receptors referred to as BMP receptors (BMPRs), which mobilize the SMAD family proteins.[45] Development of JPS is linked to mutations in one of two known genes in the TFG-β/BMP pathway. Specifically, 20% of patients with JPS are linked to mutations in *SMAD4*, while a similar proportion of cases are attributable to *BMPR1A* mutations. To date, more than 40 mutations (both single nucleotide and larger structural) in these genes have been linked to CRC.[46]

Novel Approaches to Further Discovery of Higher-Penetrant Mutations

Given their high penetrance and apparent clustering in families the aforementioned mutations could be discovered through relatively small linkage studies. While many genes underlying these syndromes have been uncovered, it is expected that additional higher-penetrant genes exist but are more difficult to detect. For example, familial colorectal cancer type X (FCCTX) is a condition that meets the clinical criteria for Lynch Syndrome, with the caveat that FCCTX does not show mutations in any of the MMR genes. Although FCCTX is currently classified as a single condition, it is possible that FCCTX represents more than one underlying disorder resulting from various unknown genetic causes.[20,47] As such, additional research on familial CRC is likely to uncover additional rarer variants with moderate effects on risk. However, until recent advances in sequencing technology the investigation of such variation was prohibitively expensive.

It is expected that whole-exome or whole-genome sequencing of families at high risk with undetermined germline mutations will uncover higher-penetrance mutations within pathways or mechanisms not previously implicated in CRC. For instance, the discovery of POLE and POLD1 mutations resulted from such analyses.[45–48] In addition to discovering new genes it can be expected that sequencing studies will discover additional mutations in known CRC genes given that improved sequencing technologies more comprehensively captures entire genes and enables the study of more complex structural variation, such as copy number variations, translocations, and inversions.[49]

Whole-exome and whole-genome studies may also identify an increasing number of variants with unknown significance (VUS) in hereditary CRC genes.[48–50] Despite residing within genes known to harbor very rare and highly penetrant mutations, these uncharacterized VUS can range from benign to pathogenic and thus pose a challenging problem, particularly for clinical testing. For instance, unlike the well-characterized mutations linked to FAP and AFAP, more common variants in *APC* have also been observed near domains known to have functional importance. An example of this is the *APC*-I1307K variant (rs1801165) that is very rare in most populations but has a frequency of approximately 6% in Ashkenazim and a modest risk (odds ratio of 2.17) for colorectal cancer (Figure 1).[51] However, such effects (odds ratios near 2) can only be found with confidence if tested in sizable studies (several thousand participants). As the penetrance of mutations in syndromic genes varies by position within a gene and between genes, understanding the increasing discovery of variation within these genes through exome sequencing is a growing challenge.

To help catalog and better classify these variants, the International Society for Gastrointestinal Hereditary Tumours (InSiGHT) is curating a comprehensive database of observed variants in known syndromic genes for CRC. Such large-scale efforts in both

assembling locus specific databases and in expert review of variants will be important to establish consistent management of individuals suspected to have Lynch Syndrome,[50] as well as other hereditary syndromes.

DISCOVERIES OF COMMON LOW-PENETRANCE VARIANTS FROM GWAS OF SPORADIC CRC

Family-based linkage studies can successfully identify high-penetrance mutations; however, in a key work, Risch and Merikangas[52] argued that for many complex diseases such as CRC, linkage studies have limited power to detect more common variants with weaker effects. However, in combination, low penetrance mutations may contribute substantially to overall disease heritability given their higher prevalence in the population. As technologies were not available at the time to conduct comprehensive genome-wide scan, initial successes were limited to candidate gene approaches.[53–55] However, as genotyping technology rapidly evolved to allow simultaneous testing of more than a million single nucleotide variants, a growing number of lower-penetrance variants associated with sporadic CRC have been identified.[56–69]

To date GWAS have identified over 40 independent loci providing deeper insight into the underlying biology of CRC (Table 2). The risk allele frequency of these variants ranges from 0.06 to 0.9 and the genetic effect [odds ratio (OR) per risk allele] ranges from approximately 1.04 to 1.56 (Table 2). Importantly, most common susceptibility loci are positioned outside of coding regions many kilobases (kb) away from the nearest candidate gene. Although associated loci are often positioned in intergenic regions, their proximity to candidate target genes has implicated many known CRC-related genes and pathways. For example, GWAS have identified common variation in putative regulatory elements that impact the expression of genes within the *TGF-β/BMP* pathways (e.g. *BMP2, BMP4, SMAD7, CCND2, GREM1*)[70] and genes in the mitogen-activated protein kinases (MAPK) pathway (e.g. *DUSP10, MYO1B, MYC, CCND2, SH2B*).

In addition to genes within the TGF-β/BMP pathway, other GWAS findings have similarly demonstrated that genes linked to hereditary syndromes may also harbor common risk variants with weaker effects (Table 3). For instance, a truncating mutation *CDH1* was previously linked to early onset colorectal and gastric cancers[71] and more common independent variants in the same gene show weaker associations with sporadic CRC.[72] Like *APC* and *β–catenin* (*CTNNB1*), the gene product of *CDH1* is a component of adherens junctions and is involved with Wnt signaling, suggesting that aberrant regulation of *CDH1* expression could underlie the observed CRC association at the 16q22.1 locus near *CDH1* (rs9929218). [72,73] Similarly, GWAS have implicated *POLD3*[74] in CRC development. In light of the recent discovery of higher penetrance mutations in *POLE* and *POLD1* described earlier, these findings further suggest that DNA polymerase, as well as high- and low-penetrance variants in the same biological pathways, may play a role in CRC.

Although many GWAS-identified susceptibility loci are positioned in or near genes involved in established CRC-related pathways, many GWAS-identified regions do not harbor known candidate genes. This supports the utility of using agnostic genome-wide approaches, such

as GWAS to gain novel insight into the genetics of complex diseases. For instance, *CDKN1A, EIF3H, TPD52L3, ITIH2, LAMA5*, and *LAMC1* represent genes in pathways not previously linked to CRC. *CDKN1A* encodes p21 which impacts multiple tumor suppressor pathways and represses MYC-dependent transcription.[74] *TPD52L3* belongs to the family of tumor protein D52 genes, which have been implicated in cell proliferation and apoptosis and also serve as potential cancer biomarkers.[75] *ITIH* genes have been found to be down-regulated in multiple human solid tumors, including colon, breast, and lung. As such, *ITIH* genes may represent a family of understudied putative tumor suppressor genes.[76] *LAMA5* and *LAMC1* belong to the laminin gene family, which is involved in the maintenance of cell adhesion, migration, and signaling—suggesting that laminin genes may play an important role in the development of CRC.[77–79]

While we attempt to describe the most likely candidate gene linked to each of the newly identified genetic loci it is important to note that for most loci functional evaluations as described below are still missing and, hence, it is possible that for some loci the candidate genes may change as more functional evidence becomes available.

WHAT COMES AFTER DISCOVERY?

Fine-mapping and Functional Follow Up

While results for high-penetrance mutations usually point directly to the underlying casual variants, low-penetrance variants are typically correlated with (i.e., tag) the region that contains the underlying causal variant(s). For example, in Figure 2 any of the correlated single nucleotide polymorphisms (SNPs) with lower p-values could be the underlying causal variant in this GWAS locus (note the most significant variant is not necessarily the causal variant). To identify the causal variant(s) fine-mapping and laboratory (i.e., functional) follow-up studies are needed. Fine-mapping studies attempt to genotype or impute all genetic variants in a GWAS locus to test association with CRC risk and provide a comprehensive list of all potential candidate SNPs that could be the underlying causal variant.[80] Fine-mapping studies can further investigate the existence of multiple independent causal variants in the region by simultaneously including multiple variants in a single model (i.e., conditional analysis). This approach has successfully identified secondary independent CRC-related SNPs in 5 regions (*BMP4*-14q22.2, *BMP2*-20p12.3, *CCND2*-12p13.32, *SMAD7*-18q21, and *TCF7L2*-10q25.2, Table 2).[81] As has been shown for many other traits[82–88] fine-mapping is particularly powerful if conducted in multiple ethnicities with different haplotype structures to refine the number of possible causal variants. This is particularly true for participants of African descent given shorter haplotype blocks; unfortunately, there are currently limited numbers of available CRC studies in African decent populations.[80]

Once the list of potential causal variants has been refined, *in silico* functional follow-up can help to prioritize the most likely candidates for further evaluation in the laboratory to detect allele-specific effects and demonstrate likely target gene(s).[89] As stated earlier, most high-penetrance mutations associated with hereditary syndromes disrupt the coding of a protein resulting in qualitative differences in protein structure that can be more readily predicted based on our knowledge of the genetic code and biochemical properties of the encoded

amino acids. However, most low-penetrance variants are located in non-coding regions and are predicted to confer weak effects on the expression of an often unknown target gene. Our limited understanding of transcriptional regulation and the consequences of polymorphisms in putative functional elements make interpretation of non-coding loci far more challenging. To address this, Encyclopedia of non-coding DNA elements (ENCODE) and Roadmap epigenomics projects have created comprehensive catalogs of histone modifications and chromatin structure across many cell-types and tissues allowing researchers to prioritize variants that are most likely to disrupt regulatory elements. In addition, these resources allow researchers to form testable hypotheses about allelic effects on gene expression or chromatin structure for further laboratory follow-up.

When the first GWAS-identified CRC-related variant (rs6983267) was discovered in 2007, little was understood about its function. The locus was positioned in a gene desert, suggesting if the association was real, the effects must be exerted through some regulatory mechanism. However, the closest putative target gene, *MYC*, was positioned more than 300 kb away and rs6983267 was not associated with differences in *MYC* expression.[90] It took an additional 3 years for functional studies to demonstrate that the variant was located in transcriptional enhancer that differentially binds TCF7L2 (also known as TCF4) and physically interacts with the *MYC* promoter (Table 4).[91–94] Additional animal studies showed that knocking out the enhancer did not impact normal function. However, when crossed with APC^{min} mice, which have a mutated APC gene leading to multiple intestinal neoplasia (min), those mice with knocked out MYC enhancer had significantly reduced numbers of spontaneously developing colorectal tumors.[95] Accordingly, this extensive functional work was able to identify the causal variant and describe the mechanism by which it exerts long-range regulation of a gene. Notably, gene expression analysis was unable to reveal *MYC* as the target, demonstrating that even in the rare circumstance that the index association (or tagging SNP) is the causal variant, extensive functional follow-up is likely necessary to reveal the underlying biology driving CRC-associations in GWAS.

For many years researchers have focused on MYC as a candidate drug target,[96] but direct inhibition was difficult.[97] As such, the identification of the regulatory element through the GWAS variants rs6983267 positioned several hundred base pairs upstream of MYC and influencing MYC expression opens new avenues for drug development.[95] Given that most GWAS loci are not located in coding regions and may or may not impact the closest genes,[98] functional work is critical to fully understand the importance of GWAS loci and reveal potential drug targets. However, functional work requires very different expertise and tools than discovery of susceptibility loci—necessitating interdisciplinary collaboration, such as those funded by the GAME-ON Initiative.[99,100] Furthermore, it is critical to develop novel functional assays with higher throughput[101] that simultaneously evaluate the growing number of GWAS loci, which can bridge the expanding gap between the numerous susceptibility loci and the limited number with laboratory-validated function (Table 4).

FUTURE DIRECTION IN DISCOVERY

Identifying new risk loci through whole-genome and whole-exome sequencing

Common CRC susceptibility loci are expected to explain ~7–8% of the heritability of CRC,[102] however, the heritability explained currently by GWAS identified common susceptibility loci is only ~1–4%.[56–61,63,72,74,81,102–106] This gap suggests that many common variants remain to be discovered. Notably, a third of common CRC susceptibility loci were only discovered in the last year, highlighting that—through larger study populations as well as improved technologies and analytic approaches—additional GWAS discoveries will likely be made. This is consistent with other common cancers and complex diseases, such as breast[107] and prostate cancer[108] in which increasingly larger meta-analyses are discovering many novel genetic loci. Ongoing large-scale consortia, e.g. funded through the GAME-On Initiative of the National Cancer Institute and other non-US agencies are expected to further facilitate these discovery efforts.[109]

While GWAS discoveries are ongoing, next generation sequencing as well as denser genotyping arrays are increasingly being used to investigate less frequent (minor allele frequency, MAF=1–5%) and rare (MAF<1%) variants. Importantly, these variants contribute to the vast majority of the genetic variation in the genome (Figure 3) and, hence, likely account for part of the missing heritability of CRC. Progress in discovering these variants will depend on their effect size; initial data suggest that at least some of the less frequent and rare variants have stronger effects (OR>1.5).[110–113] This is consistent with the discovery of several independent signals in GWAS[114–116] and high-penetrance regions[117] with MAF<5% and ORs between 1.5 and 4.3. However, it is important to note that these initial findings likely overestimate the effect of less frequent variants as they represent the most easily detectable of these variants. Subsequent discoveries of less frequent and rare variants with weaker effects will require rigorous study designs at a much larger scale.

Current sequencing studies focus on either the 1 to 2% of the genome that encodes proteins (the “exome”) or on sequencing the “whole genome”. Exome sequencing studies are successful in identifying high-penetrance mutations; however, it remains unclear what fraction of lower penetrance variants will fall within the exome. Sequencing studies, when conducted at sufficient depth, allow for the investigation of more complex genetic variants, such as insertions or deletions (indels) or copy number variations (CNVs, also called structural variation).[118–124] Although the absolute number of these complex genetic variants is substantially lower than the number of single nucleotide variants, the fraction of the genome affected by CNVs is substantially larger.[125–132] The global assessment of complex variation has remained mostly elusive[122,133–136] and it is currently unknown to what extent these types of variations contribute to the heritability of CRC.

Given the tens of millions of rare variants currently discovered through whole genome sequencing studies, even large sequencing studies will have limited statistical power to detect these variants. For instance, to detect a low frequency (1% allele frequency) CRC risk variant with modest effect (odds ratio =1.5) among approximately five million tested variants would require 21,800 CRC cases and 21,800 controls as this variant would need to

reach a p-value of 1×10^{-8} (alpha threshold=0.05/5,000,000) to adequately account for the multiple comparisons. However, novel statistical methods that incorporate the growing body of functional data, such as ENCODE,[137] RoadMap,[138] GTEX,[139] and TCGA[140] are expected to improve prediction of variants likely to have functional importance, which will in turn enable more hypothesis-driven discovery of novel CRC susceptibility loci. Rigorous bioinformatics that combine data from various laboratory-based assays (e.g. RNA–seq, ChIP-seq, Dnase-seq, chromosome capture, and motif enrichment analysis) have improved substantially the resolution of predicted functional elements. These efforts have also enabled better prediction of long range interactions between regulatory regions and target genes. As such, functional information is beginning to reach sufficient fruition to help inform association testing of rare variants uncovered through sequencing efforts.

Integration of the cancer genome

Cancer is characterized by genetic and epigenetic alterations occurring in the tumor, also referred to as somatic mutations. In the past there has been limited exchange between cancer genetics research focusing on somatic mutations and germline genetics research focusing on inherited disease-related variants. However, high and low-penetrance germline variation for CRC are located in genes that often possess somatic mutations in CRC tumors, which are located in pathways thought to impact tumor development, such as Wnt, TGF-β, and MAPK.[141,142] This is somewhat unsurprising given that germline high-penetrance mutations in tumor suppressor genes (e.g. *APC, BMPR1A, PTEN)* only progress to cancer after a somatic mutation results in the dysfunction of the second copy of the gene. Further demonstrating the complexity of CRC, low-penetrance genetic variants are distinct because they are predominantly positioned outside of coding regions (yet often close to somatic driver genes) and confer more subtle effects leading to small increases in CRC risk. As stated previously, the link between somatic mutations and germline variants is an inherent feature of cancer development and, therefore, it is important for future studies to integrate both germline and tumor genetics to obtain a more comprehensive understanding of the carcinogenic processes in CRC.

GENE-ENVIRONMENT (GxE) INTERACTIONS

CRC has several established environmental risk factors, many of which are modifiable. The most consistently observed positive associations are seen with age, male sex, obesity, height, smoking, alcohol, and red and processed meat; protective associations are seen with physical activity, non-steroidal anti-inflammatory drug (NSAID) use, exogenous hormone use, calcium, vitamin D, folate, and, to a lesser extent, fruits, vegetables, and fiber.[143–148] Extensive methodological and applied research provides a strong rationale for examining GxE interactions.[149–153] GxE interaction analyses have identified interaction between a known GWAS locus, rs16892766 (8q23.3), and vegetable intake[154,155] and genome-wide approaches have found statistically significant interactions with several environmental risk factors, such as processed meat and postmenopausal hormone use.[156,157] However, it is important to note that investigation for GxE interactions is still at an early stage because sufficiently powered studies require well-characterized samples with environmental data harmonized across multiple ongoing studies. Statistical power is a particular challenge for

GxE analysis given that the discovery of GxE interaction requires approximately 4 times the sample size than the discovery of marginal effects of genetic variants.[158] However, these explorations are of particular interest to the public as they can help identify subpopulations for which modifiable environmental exposures are most influential. [159]

IMPACT OF GENETIC LOCI ON TREATMENT

As the discovery and functional follow-up of the many CRC-related genetic loci are ongoing it is important to consider the potential implications of these findings. Here, we present examples across multiple diseases that demonstrate the potential opportunities for genetic data. In Crohn's disease, for instance, GWAS loci implicated previously less appreciated physiologic processes, such as autophagy, innate immunity, and IL-23R signaling.[67,160,161] These discoveries have already lead to chemical screens for candidate therapeutic agents.[67,162,163] For age-related macular degeneration, GWAS identified several genes involved in inflammation, a link that was not previously established and has now opened up new treatment approaches and prevention strategies.[164,165] Identifying the genetic basis of several Mendelian disorders has led to the development of FDA-approved drugs.[166] Furthermore, genomic information can help improve clinical trial design (e.g., screening for subtypes and adverse drug reactions).[167,168] For the HIV antiviral drug, Abacavir, genetically guided prescription is now standard of care.[169] This is also true for *CYP2C19*/clopidogrel, *CYP2D6*/codeine, *TPMT*/azathioprine, and 27 other interactions now endorsed by the American Society of Health System Pharmacists.[170] These and other examples have led to new therapies[166,171–174] and improved medical practices,[175,176] demonstrating the potential of genetic findings.[68,177] However, as drug development takes years to establish efficacy and effectiveness in clinical settings,[178] it is likely that the full impact of the many recently discovered genetic findings is only beginning to be understood. In some instances, however, the identification of underlying genes will not readily translate into improved treatment. For example, the genes for cystic fibrosis and sickle cell anemia were identified more than twenty years ago;[67] although recent findings suggest that treatment options remain possible.[166,179] GWAS findings may also be useful for repositioning approved drugs.[180] For instance, GWAS findings revealed variants in *dopamine beta-hydroxylase* (*DBH*) impact smoking cessation and thereby opened the possibility for targeted use of Nepicastat (a drug targeting DHB and traditionally used to treat post-traumatic stress disorder). Many more examples for drug repositioning are shown here.[180] As these examples include both rare high-penetrance and common low-penetrance variants, it is clear that neither the frequency nor the effect size will determine which susceptibility variants will lead to new treatment strategies. Accordingly, coordinated efforts to screen the growing number of susceptibility loci for putative drug targets seems promising, particularly if combined functional studies to identify the underlying functional variant(s).

Impact of Genetic Loci on CRC Prevention Using Risk Prediction Modeling

Although common susceptibility variants have limited power in discriminating cancer outcomes and many have yet to be identified,[109,181–183] studies have begun to explore the potential clinical applications of polygenic risk profiling.[184] Such models could

potentially identify individuals at higher CRC risk for targeted screening and intervention. [182,184–186]

CRC remains the second leading cause of cancer death despite slight declines in CRC incidence. Paradoxically, it is among the most preventable and treatable of neoplastic diseases when detected early. For instance in the U.S., endoscopic screening, particularly colonoscopy, is the most commonly used strategy;[187–192] however, it is costly, invasive, and carries risk.[190,193–199] Although there have been improvements in screening uptake, 40–50% of eligible persons do not follow current screening recommendations.[193,199–203] Current recommendations are primarily based on age (over 50 years) and family history of CRC [204] despite the knowledge that incidence of CRC varies substantially in the population and most cases occur in those without positive family history.[205,206] Therefore, risk-prediction models that stratify the population into risk groups according to their risk profile could result in more effective screening. Furthermore, those at higher susceptibility may be more likely to follow recommendations once they have been made aware of their increased risk.[207–212] For instance, those with a positive family history of CRC are more likely to undergo endoscopy screening.[213]

We recently showed that a genetic risk score incorporating the first 27 known CRC GWAS findings in addition to family history improved the discriminatory accuracy.[213] Specifically, the AUC improved from 0.51 to 0.59 for men and 0.52 to 0.56 for women; these results were similar to that in a previous study.[214] Although the improvement in AUC risk prediction is modest, the genetic risk score could be used to develop age-specific guidance for screening more reflective of the individual's genetic risk. Current recommendations in the United States advise that screening should commence at an age of 50 years in those without a first-degree family history of CRC or age of 40 years in individuals with family history of CRC. However, the genetic risk score identified a large fraction of men and women without a positive family history of CRC that had a higher risk for CRC (comparable or higher than those with a positive family history of CRC) justifying an earlier screening age than 50 years in a subset of individuals without a positive family history.[206] These results demonstrate that the combination of multiple common susceptibility loci can lead to improved screening recommendations tailored to an individual's risk despite the weak effects of any of the individual locus. [215] In fact, an extensive GWAS of inflammatory bowel disease recently showed that including hundreds of variants with suggestive, but not genome-wide significant associations, improved substantially risk prediction compared to models limited to GWAS findings.[216] Accordingly, it can be expected that incorporating a large number of genetic variants from genome-wide scans can further improve risk-prediction models, particularly if based on large sample sizes.[216–220]

Risk models should provide equitable benefits to the public. Currently the majority of risk models have been evaluated among those of European descent. Evaluation of models across diverse ethnicities is critical because many GWAS findings are tagging a region, rather than the causal or disease-influencing variant, and this correlation structure can vary across ethnicities. Another concern with risk prediction models is the potential for over or under prediction of risk due to failure to validate in an independent study.[206,221,222] These

issues are no longer conceptual since genetic risk profiling has now entered the commercial world. For instance, close to half a million individuals have purchased genetic information from 23andMe, which provided customers with vastly incomplete CRC risk estimates based on only 4 GWAS loci. Although the FDA has now restricted 23andMe commercialization of genetic risk profiling, clinical practice will increasingly be confronted with ambiguous genetic information.[223] Accordingly, in the era of direct-to-consumer genetic testing, rigorous evaluation of screening models that include new genetic and non-genetic information is of critical importance and expected to reduce the burden of CRC and other common complex diseases.[224]

SUMMARY

In summary, substantial progress has been made to discover high-penetrance mutations and common variants, and we expect that discovery of many additional variants will occur. Discovery of CRC loci is driven by sample sizes and available technologies. Genetic variants can be discovered in well-defined CRC pathways, such as TGF-β or Wnt, but can also point to novel genes and pathways not previously implicated in CRC. It remains critical that we stay on the path to uncover the complete genetic architecture of CRC to more fully understand the etiology of this severe disease; these findings, in turn, can lead to improved treatment and prevention.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

FUNDING

Supported by grants from the National Cancer Institute (R01 CA059045, U01 CA137088, U01 CA164930, R01 CA120582)

Abbreviations

Note that genes often have the same name as the encoded protein. To distinguish between the two italics are used to refer to the gene.

AFAP	attenuated form of familial adenomatous polyposis
APC	*adenomatous polyposis coli*
BMP	bone morphogenic protein
BMPRs	bone morphogenic protein receptors
BRR	Bannayan-Ruvalcaba-Riley
CRC	colorectal cancer
CS	Cowden Syndrome
DBH	*dopamine beta-hydroxylase*

FAP	familial adenomatous polyposis
FCCTX	familial colorectal cancer type X
GWAS	genome-wide association studies
GxE	Gene-environment
HMPS	Hereditary mixed polyposis syndrome
HNPCC	hereditary nonpolyposis colorectal cancer
InSiGHT	the International Society for Gastrointestinal Hereditary Tumours
JPS	Juvenile polyposis syndrome
kb	kilobases
MAF	minor allele frequency
MAP	*MUTYH*-associated polyposis
MAPK	mitogen-activated protein kinases
MMR	Mismatch Repair
MSI	Microsatellite Instability
NSAID	non-steroidal anti-inflammatory drug
PJS	Peutz-Jeghers syndrome
TGF	transforming growth factor
VUS	variants of unknown significance

Reference List

1. Ferlay J, Shin HR, Bray F, et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010; 127(12):2893–917. [PubMed: 21351269]
2. Howlader, N.; Noone, A.; Krapcho, M., et al. SEER Cancer Statistics Review, 1975–2010. Cronin, KA., editor. Bethesda, MD: National Cancer Institute; 2014. based on November 2012 SEER data submission, posted to the SEER web site, April 2013
3. Ferlay, J.; Soerjomataram, I.; Ervik, M., et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer; 2013. http://globocan.iarc.fr [accessed on 12/06/2014]
4. Lichtenstein P, Holm NV, Verkasalo PK, et al. 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(2):78–85. [PubMed: 10891514]
5. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9. 6 million individuals in the Swedish Family-Cancer Database. Int J Cancer. 2002; 99(2):260–6. [PubMed: 11979442]
6. Talseth-Palmer BA, Wijnen JT, Grice DM, et al. Genetic modifiers of cancer risk in Lynch syndrome: a review. Fam Cancer. 2013; 12(2):207–16. [PubMed: 23471748]
7. Jasperson KW, Tuohy TM, Neklason DW, et al. Hereditary and familial colon cancer. Gastroenterology. 2010; 138(6):2044–58. [PubMed: 20420945]
8. Bodmer WF, Bailey CJ, Bodmer J, et al. Localization of the gene for familial adenomatous polyposis on chromosome 5. Nature. 1987; 328(6131):614–6. [PubMed: 3039373]

9. Vasen HF. Clinical diagnosis and management of hereditary colorectal cancer syndromes. J Clin Oncol. 2000; 18(21 Suppl):81S–92S. [PubMed: 11060333]
10. Giardiello FM, Welsh SB, Hamilton SR, et al. Increased risk of cancer in the Peutz-Jeghers syndrome. N Engl J Med. 1987; 316(24):1511–4. [PubMed: 3587280]
11. Larsen, HJ.; Howe, JR. Juvenile Polyposis Syndrome. 1993.
12. Mallory SB. Cowden syndrome (multiple hamartoma syndrome). Dermatol Clin. 1995; 13(1):27–31. [PubMed: 7712647]
13. Al-Tassan N, Chmiel NH, Maynard J, et al. Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. Nat Genet. 2002; 30(2):227–32. [PubMed: 11818965]
14. Grady WM. Genetic testing for high-risk colon cancer patients. Gastroenterology. 2003; 124(6): 1574–94. [PubMed: 12761718]
15. Burt R, Neklason DW. Genetic testing for inherited colon cancer. Gastroenterology. 2005; 128(6): 1696–716. [PubMed: 15887160]
16. Goss KH, Groden J. Biology of the adenomatous polyposis coli tumor suppressor. J Clin Oncol. 2000; 18(9):1967–79. [PubMed: 10784639]
17. Yamulla RJ, Kane EG, Moody AE, et al. Testing Models of the APC Tumor Suppressor/beta-Catenin Interaction Reshapes Our View of the Destruction Complex in Wnt Signaling. Genetics. 2014; 197(4):1285–302. [PubMed: 24931405]
18. van der Flier LG, Clevers H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. Annu Rev Physiol. 2009; 71:241–60. [PubMed: 18808327]
19. Snow AK, Tuohy TM, Sargent NR, et al. APC promoter 1B deletion in seven American families with familial adenomatous polyposis. Clin Genet. 2014
20. Hegde M, Ferber M, Mao R, et al. ACMG technical standards and guidelines for genetic testing for inherited colorectal cancer (Lynch syndrome, familial adenomatous polyposis, and MYH-associated polyposis). Genet Med. 2014; 16(1):101–16. [PubMed: 24310308]
21. Nieuwenhuis MH, Vasen HF. Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): a review of the literature. Crit Rev Oncol Hematol. 2007; 61(2):153–61. [PubMed: 17064931]
22. Gammon A, Kohlmann W, Burt R. Can we identify the high-risk patients to be screened? A genetic approach Digestion. 2007; 76(1):7–19. [PubMed: 17947814]
23. Galiatsatos P, Foulkes WD. Familial adenomatous polyposis. Am J Gastroenterol. 2006; 101(2): 385–98. [PubMed: 16454848]
24. Lipton L, Halford SE, Johnson V, et al. Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway. Cancer Res. 2003; 63(22):7595–9. [PubMed: 14633673]
25. Al-Sohaily S, Biankin A, Leong R, et al. Molecular pathways in colorectal cancer. J Gastroenterol Hepatol. 2012; 27(9):1423–31. [PubMed: 22694276]
26. Desai TK, Barkel D. Syndromic colon cancer: lynch syndrome and familial adenomatous polyposis. Gastroenterol Clin North Am. 2008; 37(1):47–72. vi. [PubMed: 18313539]
27. Stoffel E, Mukherjee B, Raymond VM, et al. Calculation of risk of colorectal and endometrial cancer among patients with Lynch syndrome. Gastroenterology. 2009; 137(5):1621–7. [PubMed: 19622357]
28. Rowley PT. Inherited susceptibility to colorectal cancer. Annu Rev Med. 2005; 56:539–54. [PubMed: 15660526]
29. Perucho M. Correspondence re: C.R. Boland et al., A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res. 58: 5248–5257, 1998. Cancer Res. 1999; 59(1):249–56. [PubMed: 9892214]
30. Esteban-Jurado C, Garre P, Vila M, et al. New genes emerging for colorectal cancer predisposition. World J Gastroenterol. 2014; 20(8):1961–71. [PubMed: 24587672]
31. Seshagiri S. The burden of faulty proofreading in colon cancer. Nat Genet. 2013; 45(2):121–2. [PubMed: 23358219]

32. Briggs S, Tomlinson I. Germline and somatic polymerase epsilon and delta mutations define a new class of hypermutated colorectal and endometrial cancers. J Pathol. 2013; 230(2):148–53. [PubMed: 23447401]
33. Palles C, Cazier JB, Howarth KM, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet. 2013; 45(2): 136–44. [PubMed: 23263490]
34. Jasperson KW. Genetic testing by cancer site: colon (polyposis syndromes). Cancer J. 2012; 18(4): 328–33. [PubMed: 22846733]
35. Shimamoto Y, Yamaguchi M. HTLV-I induced extranodal lymphomas. Leuk Lymphoma. 1992; 7(1–2):37–45. [PubMed: 1472932]
36. Vivanco I, Sawyers CL. The phosphatidylinositol 3-Kinase AKT pathway in human cancer. Nat Rev Cancer. 2002; 2(7):489–501. [PubMed: 12094235]
37. Voorneveld PW, Kodach LL, Jacobs RJ, et al. Loss of SMAD4 alters BMP signaling to promote colorectal cancer cell metastasis via activation of Rho and ROCK. Gastroenterology. 2014; 147(1): 196–208. [PubMed: 24704720]
38. Howe JR, Bair JL, Sayed MG, et al. Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. Nat Genet. 2001; 28(2):184–7. [PubMed: 11381269]
39. Cao X, Eu KW, Kumarasinghe MP, et al. Mapping of hereditary mixed polyposis syndrome (HMPS) to chromosome 10q23 by genomewide high-density single nucleotide polymorphism (SNP) scan and identification of BMPR1A loss of function. J Med Genet. 2006; 43(3):e13. [PubMed: 16525031]
40. O'Riordan JM, O'Donoghue D, Green A, et al. Hereditary mixed polyposis syndrome due to a BMPR1A mutation. Colorectal Dis. 2010; 12(6):570–3. [PubMed: 19438883]
41. Lubbe SJ, Pittman AM, Matijssen C, et al. Evaluation of germline BMP4 mutation as a cause of colorectal cancer. Hum Mutat. 2011; 32(1):E1928–E1938. [PubMed: 20949628]
42. Jaeger E, Leedham S, Lewis A, et al. Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. Nat Genet. 2012; 44(6):699–703. [PubMed: 22561515]
43. Ngeow J, Heald B, Rybicki LA, et al. Prevalence of germline PTEN, BMPR1A, SMAD4, STK11, and ENG mutations in patients with moderate-load colorectal polyps. Gastroenterology. 2013; 144(7):1402–9. 1409. [PubMed: 23399955]
44. Sweet K, Willis J, Zhou XP, et al. Molecular classification of patients with unexplained hamartomatous and hyperplastic polyposis. JAMA. 2005; 294(19):2465–73. [PubMed: 16287957]
45. Bleuming SA, He XC, Kodach LL, et al. Bone morphogenetic protein signaling suppresses tumorigenesis at gastric epithelial transition zones in mice. Cancer Res. 2007; 67(17):8149–55. [PubMed: 17804727]
46. Cichy W, Klincewicz B, Plawski A. Juvenile polyposis syndrome. Arch Med Sci. 2014; 10(3): 570–7. [PubMed: 25097590]
47. Shiovitz S, Copeland WK, Passarelli MN, et al. Characterisation of Familial Colorectal Cancer Type X, Lynch syndrome, and non-familial colorectal cancer. Br J Cancer. 2014; 111(3):598–602. [PubMed: 24918813]
48. Dorschner MO, Amendola LM, Turner EH, et al. Actionable, pathogenic incidental findings in 1,000 participants' exomes. Am J Hum Genet. 2013; 93(4):631–40. [PubMed: 24055113]
49. Shirts BH, Jacobson A, Jarvik GP, et al. Large numbers of individuals are required to classify and define risk for rare variants in known cancer risk genes. Genet Med. 2013
50. Thompson BA, Spurdle AB, Plazzer JP, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet. 2014; 46(2):107–15. [PubMed: 24362816]
51. Liang J, Lin C, Hu F, et al. APC polymorphisms and the risk of colorectal neoplasia: a HuGE review and meta-analysis. Am J Epidemiol. 2013; 177(11):1169–79. [PubMed: 23576677]
52. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996; 273(5281):1516–7. [PubMed: 8801636]
53. Dong LM, Potter JD, White E, et al. Genetic susceptibility to cancer: the role of polymorphisms in candidate genes. JAMA. 2008; 299(20):2423–36. [PubMed: 18505952]

54. Ma X, Zhang B, Zheng W. Genetic variants associated with colorectal cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. Gut. 2014; 63(2):326–36. [PubMed: 23946381]
55. Theodoratou E, Montazeri Z, Hawken S, et al. Systematic meta-analyses and field synopsis of genetic association studies in colorectal cancer. J Natl Cancer Inst. 2012; 104(19):1433–57. [PubMed: 23019048]
56. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24. 21. Nat Genet. 2007; 39(8):984–8. [PubMed: 17618284]
57. Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet. 2007; 39(8):989–94. [PubMed: 17618283]
58. Broderick P, Carvajal-Carmona L, Pittman AM, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet. 2007; 39(11):1315–7. [PubMed: 17934461]
59. Tenesa A, Farrington SM, Prendergast JG, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet. 2008; 40(5):631–7. [PubMed: 18372901]
60. Jaeger E, Webb E, Howarth K, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13. 3 influence colorectal cancer risk. Nat Genet. 2008; 40(1):26–8. [PubMed: 18084292]
61. Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23. 3. Nat Genet. 2008; 40(5): 623–30. [PubMed: 18372905]
62. Houlston RS, Webb E, Broderick P, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nature Genetics. 2008; 40(12):1426–35. [PubMed: 19011631]
63. 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(11):973–7. [PubMed: 20972440]
64. Tomlinson IP, Carvajal-Carmona LG, Dobbins SE, et al. 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(6):e1002105. [PubMed: 21655089]
65. Carvajal-Carmona LG, Cazier JB, Jones AM, et al. Fine-mapping of colorectal cancer susceptibility loci at 8q23.3, 16q22.1 and 19q13.11: refinement of association signals and use of in silico analysis to suggest functional variation and unexpected candidate target genes. Hum Mol Genet. 2011
66. Peters U, Hutter CM, Hsu L, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. Hum Genet. 2011
67. Green ED, Guyer MS. Charting a course for genomic medicine from base pairs to bedside. Nature. 2011; 470(7333):204–13. [PubMed: 21307933]
68. Hunter DJ, Altshuler D, Rader DJ. From Darwin's finches to canaries in the coal mine--mining the genome for new biology. N Engl J Med. 2008; 358(26):2760–3. [PubMed: 18579810]
69. Manolio TA, Collins FS. The HapMap and genome-wide association studies in diagnosis and therapy. Annu Rev Med. 2009; 60:443–56. [PubMed: 19630580]
70. Tenesa A, Dunlop MG. New insights into the aetiology of colorectal cancer from genome-wide association studies. Nat Rev Genet. 2009; 10(6):353–8. [PubMed: 19434079]
71. Richards FM, McKee SA, Rajpar MH, et al. Germline E-cadherin gene (CDH1) mutations predispose to familial gastric cancer and colorectal cancer. Hum Mol Genet. 1999; 8(4):607–10. [PubMed: 10072428]
72. Houlston RS, Webb E, Broderick P, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet. 2008; 40(12):1426–35. [PubMed: 19011631]

73. Pecina-Slaus N. Tumor suppressor gene E-cadherin and its role in normal and malignant cells. Cancer Cell Int. 2003; 3(1):17. [PubMed: 14613514]
74. Dunlop MG, Dobbins SE, Farrington SM, et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat Genet. 2012; 44(7):770–6. [PubMed: 22634755]
75. Byrne JA, Frost S, Chen Y, et al. Tumor protein D52 (TPD52) and cancer-oncogene understudy or understudied oncogene? Tumour Biol. 2014
76. Hamm A, Veeck J, Bektas N, et al. Frequent expression loss of Inter-alpha-trypsin inhibitor heavy chain (ITIH) genes in multiple human solid tumors: a systematic expression analysis. BMC Cancer. 2008; 8:25. [PubMed: 18226209]
77. Patarroyo M, Tryggvason K, Virtanen I. Laminin isoforms in tumor invasion, angiogenesis and metastasis. Semin Cancer Biol. 2002; 12(3):197–207. [PubMed: 12083850]
78. Pouliot N, Saunders NA, Kaur P. Laminin 10/11: an alternative adhesive ligand for epidermal keratinocytes with a functional role in promoting proliferation and migration. Exp Dermatol. 2002; 11(5):387–97. [PubMed: 12366691]
79. Turck N, Lefebvre O, Gross I, et al. Effect of laminin-1 on intestinal cell differentiation involves inhibition of nuclear nucleolin. J Cell Physiol. 2006; 206(2):545–55. [PubMed: 16245305]
80. Wang H, Haiman CA, Burnett T, et al. Fine-mapping of genome-wide association study-identified risk loci for colorectal cancer in African Americans. Hum Mol Genet. 2013; 22(24):5048–55. [PubMed: 23851122]
81. Tomlinson IP, Carvajal-Carmona LG, Dobbins SE, et al. Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. PLoS genetics. 2011; 7(6):e1002105. [PubMed: 21655089]
82. Peters U, North KE, Sethupathy P, et al. A systematic mapping approach of 16q12. 2/FTO and BMI in more than 20,000 African Americans narrows in on the underlying functional variation: results from the Population Architecture using *Genomics and Epidemiology (PAGE) study*. PLoS genetics. 2013; 9(1):e1003171. [PubMed: 23341774]
83. Williams SM. Identifying population differences in genes that affect body mass index. Genome Med. 2013; 5(11):102. [PubMed: 24286457]
84. Avery CL, Sethupathy P, Buyske S, et al. Fine-mapping and initial characterization of QT interval loci in African Americans. PLoS Genet. 2012; 8(8):e1002870. [PubMed: 22912591]
85. Kuo JZ, Sheu WH, Assimes TL, et al. Trans-ethnic fine mapping identifies a novel independent locus at the 3′ end of CDKAL1 and novel variants of several susceptibility loci for type 2 diabetes in a Han Chinese population. Diabetologia. 2013; 56(12):2619–28. [PubMed: 24013783]
86. Wu Y, Waite LL, Jackson AU, et al. Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. PLoS genetics. 2013; 9(3):e1003379. [PubMed: 23555291]
87. Du M, Auer PL, Jiao S, et al. Whole-exome imputation of sequence variants identified two novel alleles associated with adult body height in African Americans. Hum Mol Genet. 2014
88. Gong J, Schumacher F, Lim U, et al. Fine Mapping and Identification of BMI Loci in African Americans. Am J Hum Genet. 2013; 93(4):661–71. [PubMed: 24094743]
89. Freedman ML, Monteiro AN, Gayther SA, et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet. 2011; 43(6):513–8. [PubMed: 21614091]
90. Haiman CA, Le ML, Yamamato J, et al. A common genetic risk factor for colorectal and prostate cancer. Nat Genet. 2007; 39(8):954–6. [PubMed: 17618282]
91. Tuupanen S, Turunen M, Lehtonen R, et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nature Genetics. 2009; 41(8):885–U37. [PubMed: 19561604]
92. Pomerantz MM, Ahmadiyeh N, Jia L, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat Genet. 2009; 41(8):882–4. [PubMed: 19561607]
93. Wasserman NF, Aneas I, Nobrega MA. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. Genome Res. 2010; 20(9): 1191–7. [PubMed: 20627891]

94. Wright JB, Brown SJ, Cole MD. Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. Mol Cell Biol. 2010; 30(6):1411–20. [PubMed: 20065031]
95. Sur IK, Hallikas O, Vaharautio A, et al. Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. Science. 2012; 338(6112):1360–3. [PubMed: 23118011]
96. Soucek L, Whitfield J, Martins CP, et al. Modelling Myc inhibition as a cancer therapy. Nature. 2008; 455(7213):679–83. [PubMed: 18716624]
97. Evan G. Cancer. Taking a back door to target Myc. Science. 2012; 335(6066):293–4. [PubMed: 22267799]
98. Schaub MA, Boyle AP, Kundaje A, et al. Linking disease associations with regulatory information in the human genome. Genome Res. 2012; 22(9):1748–59. [PubMed: 22955986]
99. Kote-Jarai Z, Saunders EJ, Leongamornlert DA, et al. Fine-mapping identifies multiple prostate cancer risk loci at 5p15, one of which associates with TERT expression. Hum Mol Genet. 2013; 22(12):2520–8. [PubMed: 23535824]
100. Ogino S, Chan AT, Fuchs CS, et al. Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. Gut. 2011; 60(3):397–411. [PubMed: 21036793]
101. Zhao W, Pollack JL, Blagev DP, et al. Massively parallel functional annotation of 3′ untranslated regions. Nat Biotechnol. 2014; 32(4):387–91. [PubMed: 24633241]
102. Jiao S, Peters U, Berndt S, et al. Estimating the heritability of colorectal cancer. Hum Mol Genet. 2014; 23(14):3898–905. [PubMed: 24562164]
103. Peters U, Jiao S, Schumacher FR, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-Analysis. Gastroenterology. 2013; 144(4):799–807. [PubMed: 23266556]
104. Peters U, Hutter CM, Hsu L, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. Hum Genet. 2012; 131(2):217–34. [PubMed: 21761138]
105. Jia WH, Zhang B, Matsuo K, et al. Genome-wide association analyses in east Asians identify new susceptibility loci for colorectal cancer. Nat Genet. 2012; 45(2):191–6. [PubMed: 23263487]
106. Gong J, Hutter C, Baron JA, et al. A pooled analysis of smoking and colorectal cancer: timing of exposure and interactions with environmental factors. Cancer Epidemiol Biomarkers Prev. 2012; 21(11):1974–85. [PubMed: 23001243]
107. Zhang B, Beeghly-Fadiel A, Long J, et al. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. Lancet Oncol. 2011; 12(5):477–88. [PubMed: 21514219]
108. Eeles R, Goh C, Castro E, et al. The genetic epidemiology of prostate cancer and its clinical implications. Nat Rev Urol. 2014; 11(1):18–31. [PubMed: 24296704]
109. Park JH, Wacholder S, Gail MH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet. 2010; 42(7):570–5. [PubMed: 20562874]
110. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009; 106(23):9362–7. [PubMed: 19474294]
111. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747–53. [PubMed: 19812666]
112. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nature Genetics. 2008; 40(6):695–701. [PubMed: 18509313]
113. Gorlov IP, Gorlova OY, Sunyaev SR, et al. Shifting paradigm of association studies: Value of rare single-nucleotide polymorphisms. American Journal of Human Genetics. 2008; 82(1):100–12. [PubMed: 18179889]
114. 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(5925):387–9. [PubMed: 19264985]

115. 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
116. Sanna S, Li B, Mulas A, et al. Fine mapping of five Loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. PLoS Genet. 2011; 7(7):e1002198. [PubMed: 21829380]
117. Stacey SN, Sulem P, Jonasdottir A, et al. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. Nat Genet. 2011; 43(11):1098–103. [PubMed: 21946351]
118. Yoon S, Xuan Z, Makarov V, et al. Sensitive and accurate detection of copy number variants using read depth of coverage. Genome Res. 2009; 19(9):1586–92. [PubMed: 19657104]
119. Korbel JO, Abyzov A, Mu XJ, et al. PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. Genome Biol. 2009; 10(2):R23. [PubMed: 19236709]
120. Ye K, Schulz MH, Long Q, et al. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009; 25(21): 2865–71. [PubMed: 19561018]
121. Handsaker RE, Korn JM, Nemesh J, et al. Discovery and genotyping of genome structural polymorphism by sequencing on a population scale. Nat Genet. 2011; 43(3):269–76. [PubMed: 21317889]
122. Alkan C, Kidd JM, Marques-Bonet T, et al. Personalized copy number and segmental duplication maps using next-generation sequencing. Nat Genet. 2009; 41(10):1061–7. [PubMed: 19718026]
123. Hormozdiari F, Alkan C, Eichler EE, et al. Combinatorial algorithms for structural variation detection in high-throughput sequenced genomes. Genome Res. 2009; 19(7):1270–8. [PubMed: 19447966]
124. Li R, Zhu H, Ruan J, et al. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 2010; 20(2):265–72. [PubMed: 20019144]
125. Lupski JR. Genomic rearrangements and sporadic disease. Nat Genet. 2007; 39(7 Suppl):S43–S47. [PubMed: 17597781]
126. Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004; 36(9):949–51. [PubMed: 15286789]
127. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. Nature. 2006; 444(7118):444–54. [PubMed: 17122850]
128. Tuzun E, Sharp AJ, Bailey JA, et al. Fine-scale structural variation of the human genome. Nat Genet. 2005; 37(7):727–32. [PubMed: 15895083]
129. Kidd JM, Cooper GM, Donahue WF, et al. Mapping and sequencing of structural variation from eight human genomes. Nature. 2008; 453(7191):56–64. [PubMed: 18451855]
130. Conrad DF, Pinto D, Redon R, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2010; 464(7289):704–12. [PubMed: 19812545]
131. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. Science. 2004; 305(5683):525–8. [PubMed: 15273396]
132. Levy S, Sutton G, Ng PC, et al. The diploid genome sequence of an individual human. PLoS Biol. 2007; 5(10):e254. [PubMed: 17803354]
133. McCarroll SA. Copy number variation and human genome maps. Nat Genet. 2010; 42(5):365–6. [PubMed: 20428091]
134. Estivill X, Cheung J, Pujana MA, et al. Chromosomal regions containing high-density and ambiguously mapped putative single nucleotide polymorphisms (SNPs) correlate with segmental duplications in the human genome. Hum Mol Genet. 2002; 11(17):1987–95. [PubMed: 12165560]
135. Locke DP, Sharp AJ, McCarroll SA, et al. Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. Am J Hum Genet. 2006; 79(2):275–90. [PubMed: 16826518]
136. Cooper GM, Zerr T, Kidd JM, et al. Systematic assessment of copy number variant detection via genome-wide SNP genotyping. Nat Genet. 2008; 40(10):1199–203. [PubMed: 18776910]

137. Goldman M, Craft B, Swatloski T, et al. The UCSC Cancer Genomics Browser: update 2013. Nucleic Acids Res. 2013; 41(Database issue):D949–D954. [PubMed: 23109555]

138. Chadwick LH. The NIH Roadmap Epigenomics Program data resource. Epigenomics. 2012; 4(3): 317–24. [PubMed: 22690667]

139. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013; 45(6): 580–5. [PubMed: 23715323]

140. Forbes SA, Bindal N, Bamford S, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011; 39(Database issue):D945–D950. [PubMed: 20952405]

141. Garcia-Albeniz X, Nan H, Valeri L, et al. Phenotypic and tumor molecular characterization of colorectal cancer in relation to a susceptibility SMAD7 variant associated with survival. Carcinogenesis. 2013; 34(2):292–8. [PubMed: 23104301]

142. Campbell PT, Curtin K, Ulrich CM, et al. Mismatch repair polymorphisms and risk of colon cancer, tumour microsatellite instability and interactions with lifestyle factors. Gut. 2009; 58(5): 661–7. [PubMed: 18523027]

143. Friedenreich CM, Orenstein MR. Physical activity and cancer prevention: etiologic evidence and biological mechanisms. J Nutr. 2002; 132(11 Suppl):3456S–64S. [PubMed: 12421870]

144. Giovannucci E. Diet, body weight, and colorectal cancer: a summary of the epidemiologic evidence. J Womens Health (Larchmt). 2003; 12(2):173–82. [PubMed: 12737716]

145. Martinez ME. Primary prevention of colorectal cancer: lifestyle, nutrition, exercise. Recent Results Cancer Res. 2005; 166:177–211. [PubMed: 15648191]

146. Tomeo CA, Colditz GA, Willett WC, et al. Harvard Report on Cancer Prevention. Volume 3: prevention of colon cancer in the United States. Cancer Causes Control. 1999; 10(3):167–80. [PubMed: 10454062]

147. World Cancer Research Fund and American Institute for Cancer Research. Food, Nutrition, Physican Activity, and the Prevention of Cancer: a Global Perspective. Washington, DC: AICR; 2007.

148. Potter JD. Colorectal cancer: molecules and populations. J Natl Cancer Inst. 1999; 91(11):916–32. [PubMed: 10359544]

149. Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet. 2005; 6(4):287–98. [PubMed: 15803198]

150. Khoury MJ, Wacholder S. Invited commentary: from genome-wide association studies to gene-environment-wide interaction studies--challenges and opportunities. Am J Epidemiol. 2009; 169(2):227–30. [PubMed: 19022826]

151. Manolio TA, Collins FS. Genes, environment, health, and disease: facing up to complexity. Hum Hered. 2007; 63(2):63–6. [PubMed: 17283435]

152. Thomas D. Gene-environment-wide association studies: emerging approaches. Nat Rev Genet. 2010; 11(4):259–72. [PubMed: 20212493]

153. Thomas D. Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. Annu Rev Public Health. 2010; 31:21–36. [PubMed: 20070199]

154. Kantor ED, Hutter CM, Minnier J, et al. Gene-Environment Interaction Involving Recently Identified Colorectal Cancer Susceptibility Loci. Cancer Epidemiol Biomarkers Prev. 2014

155. Hutter CM, Chang-Claude J, Slattery ML, et al. Characterization of gene-environment interactions for colorectal cancer susceptibility loci. Cancer Res. 2012; 72(8):2036–44. [PubMed: 22367214]

156. Garcia-Albeniz X, Rudolph A, Hutter C, et al. Novel CYP24A1 variant identified as a significant effect modifier of estrogen plus progesterone therapy and colorectal cancer risk among menopausal women. Endocrine-Related Cancer. 2014 Submitted.

157. Figueiredo JC, Hsu L, Hutter CM, et al. Genome-wide diet-gene interaction analyses for risk of colorectal cancer. PLoS genetics. 2014; 10(4):e1004228. [PubMed: 24743840]

158. Smith PG, Day NE. The design of case-control studies: the influence of confounding and interaction effects. Int J Epidemiol. 1984; 13(3):356–65. [PubMed: 6386716]

159. Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. Nat Rev Genet. 2006; 7(10):812–20. [PubMed: 16983377]
160. Van LJ, Wilson DC, Satsangi J. The genetics of Crohn's disease. Annu Rev Genomics Hum Genet. 2009; 10:89–116. [PubMed: 19453248]
161. Brest P, Corcelle EA, Cesaro A, et al. Autophagy and Crohn's disease: at the crossroads of infection, inflammation, immunity, and cancer. Curr Mol Med. 2010; 10(5):486–502. [PubMed: 20540703]
162. Rioux JD, Xavier RJ, Taylor KD, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat Genet. 2007; 39(5):596–604. [PubMed: 17435756]
163. Cadwell K, Liu JY, Brown SL, et al. A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. Nature. 2008; 456(7219):259–63. [PubMed: 18849966]
164. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. Science. 2005; 308(5720):385–9. [PubMed: 15761122]
165. Moshfeghi DM, Blumenkranz MS. Role of genetic factors and inflammation in age-related macular degeneration. Retina. 2007; 27(3):269–75. [PubMed: 17460581]
166. Dietz HC. New therapeutic approaches to mendelian disorders. N Engl J Med. 2010; 363(9):852–63. [PubMed: 20818846]
167. Simon R. The use of genomics in clinical trial design. Clin Cancer Res. 2008; 14(19):5984–93. [PubMed: 18829477]
168. Mok TS, Wu YL, Thongprasert S, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med. 2009; 361(10):947–57. [PubMed: 19692680]
169. Tozzi V. Pharmacogenetics of antiretrovirals. Antiviral Res. 2010; 85:190–200. [PubMed: 19744523]
170. Rieder MJ, Reiner AP, Gage BF, et al. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. N Engl J Med. 2005; 352(22):2285–93. [PubMed: 15930419]
171. Pao W, Miller V, Zakowski M, et al. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci U S A. 2004; 101(36):13306–11. [PubMed: 15329413]
172. Yan H, Parsons DW, Jin G, et al. IDH1 and IDH2 mutations in gliomas. N Engl J Med. 2009; 360(8):765–73. [PubMed: 19228619]
173. Brooke BS, Habashi JP, Judge DP, et al. Angiotensin II blockade and aortic-root dilation in Marfan's syndrome. N Engl J Med. 2008; 358(26):2787–95. [PubMed: 18579813]
174. d'Hulst C, Kooy RF. Fragile X syndrome: from molecular genetics to therapy. J Med Genet. 2009; 46(9):577–84. [PubMed: 19724010]
175. Miller DT, Adam MP, Aradhya S, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet. 2010; 86(5):749–64. [PubMed: 20466091]
176. Frueh FW, Amur S, Mummaneni P, et al. Pharmacogenomic biomarker information in drug labels approved by the United States food and drug administration: prevalence of related drug use. Pharmacotherapy. 2008; 28(8):992–8. [PubMed: 18657016]
177. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science. 2008; 322(5903): 881–8. [PubMed: 18988837]
178. Burke W, Laberge AM, Press N. Debating clinical utility. Public Health Genomics. 2010; 13(4): 215–23. [PubMed: 20395690]
179. Grasemann H, Ratjen F. Emerging therapies for cystic fibrosis lung disease. Expert Opin Emerg Drugs. 2010; 15(4):653–9. [PubMed: 20812885]
180. Sanseau P, Agarwal P, Barnes MR, et al. Use of genome-wide association studies for drug repositioning. Nat Biotechnol. 2012; 30(4):317–20. [PubMed: 22491277]
181. Gail MH. Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. J Natl Cancer Inst. 2008; 100(14):1037–41. [PubMed: 18612136]
182. Pharoah PD, Antoniou AC, Easton DF, et al. Polygenes, risk prediction, and targeted prevention of breast cancer. N Engl J Med. 2008; 358(26):2796–803. [PubMed: 18579814]

183. Wacholder S, Hartge P, Prentice R, et al. Performance of common genetic variants in breast-cancer risk models. N Engl J Med. 2010; 362(11):986–93. [PubMed: 20237344]

184. Pashayan N, Duffy SW, Chowdhury S, et al. Polygenic susceptibility to prostate and breast cancer: implications for personalised screening. Br J Cancer. 2011; 104(10):1656–63. [PubMed: 21468051]

185. Hingorani AD, Shah T, Kumari M, et al. Translating genomics into improved healthcare. BMJ. 2010; 341:c5945. [PubMed: 21056961]

186. Pashayan N, Pharoah P. Population-based screening in the era of genomics. Per Med. 2012; 9(4): 451–5. [PubMed: 22984365]

187. McQueen A, Bartholomew LK, Greisinger AJ, et al. Behind closed doors: physician-patient discussions about colorectal cancer screening. J Gen Intern Med. 2009; 24(11):1228–35. [PubMed: 19763699]

188. Zapka JM, Klabunde CN, Arora NK, et al. Physicians' colorectal cancer screening discussion and recommendation patterns. Cancer Epidemiol Biomarkers Prev. 2011; 20(3):509–21. [PubMed: 21239688]

189. Bandi P, Cokkinides V, Smith RA, et al. Trends in colorectal cancer screening with home-based fecal occult blood tests in adults ages 50 to 64 years, 2000–2008. Cancer. 2012; 118(20):5092–9. [PubMed: 22434529]

190. Smith RA, Cokkinides V, Brawley OW. Cancer screening in the United States, 2012: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin. 2012

191. Smith RA, Brooks D, Cokkinides V, et al. Cancer screening in the United States, 2013: a review of current American Cancer Society guidelines, current issues in cancer screening, and new guidance on cervical cancer screening and lung cancer screening. CA Cancer J Clin. 2013; 63(2): 88–105. [PubMed: 23378235]

192. Byers T. Examining stools for colon cancer prevention: what are we really looking for? Cancer Prev Res (Phila). 2011; 4(10):1531–3. [PubMed: 21972079]

193. Ling BS, Schoen RE, Trauth JM, et al. Physicians encouraging colorectal screening: a randomized controlled trial of enhanced office and patient management on compliance with colorectal cancer screening. Arch Intern Med. 2009; 169(1):47–55. [PubMed: 19139323]

194. DeBourcy AC, Lichtenberger S, Felton S, et al. Community-based preferences for stool cards versus colonoscopy in colorectal cancer screening. J Gen Intern Med. 2008; 23(2):169–74. [PubMed: 18157581]

195. Farmer MM, Bastani R, Kwan L, et al. Predictors of colorectal cancer screening from patients enrolled in a managed care health plan. Cancer. 2008; 112(6):1230–8. [PubMed: 18266204]

196. Doubeni CA, Laiyemo AO, Young AC, et al. Primary care, economic barriers to health care, and use of colorectal cancer screening tests among Medicare enrollees over time. Ann Fam Med. 2010; 8(4):299–307. [PubMed: 20644184]

197. Beydoun HA, Beydoun MA. Predictors of colorectal cancer screening behaviors among average-risk older adults in the United States. Cancer Causes Control. 2008; 19(4):339–59. [PubMed: 18085415]

198. O'Malley AS, Beaton E, Yabroff KR, et al. Patient and provider barriers to colorectal cancer screening in the primary care safety-net. Prev Med. 2004; 39(1):56–63. [PubMed: 15207986]

199. Mitka M. Colorectal cancer screening rates still fall far short of recommended levels. JAMA. 2008; 299(6):622. [PubMed: 18270348]

200. Use of colorectal cancer tests--United States, 2002, 2004, and 2006. MMWR Morb Mortal Wkly Rep. 2008; 57(10):253–8. [PubMed: 18340331]

201. Meissner HI, Breen N, Klabunde CN, et al. Patterns of colorectal cancer screening uptake among men and women in the United States. Cancer Epidemiol Biomarkers Prev. 2006; 15(2):389–94. [PubMed: 16492934]

202. Smith RA, Cokkinides V, Eyre HJ. Cancer screening in the United States, 2007: a review of current guidelines, practices, and prospects. CA Cancer J Clin. 2007; 57(2):90–104. [PubMed: 17392386]

203. National Health Interview Survey Public Use Data File 2008. National Center for Health Statistics, Centers for Disease Control and Prevention; 2009.
204. Levin B, Lieberman DA, McFarland B, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. Gastroenterology. 2008; 134(5):1570–95. [PubMed: 18384785]
205. Jackson-Thompson J, Ahmed F, German RR, et al. Descriptive epidemiology of colorectal cancer in the United States, 1998–2001. Cancer. 2006; 107(5 Suppl):1103–11. [PubMed: 16835911]
206. Win AK, Macinnis RJ, Hopper JL, et al. Risk prediction models for colorectal cancer: a review. Cancer Epidemiol Biomarkers Prev. 2012; 21(3):398–410. [PubMed: 22169185]
207. Chan AT, Giovannucci EL. Primary prevention of colorectal cancer. Gastroenterology. 2010; 138(6):2029–43. [PubMed: 20420944]
208. Longacre AV, Cramer LD, Gross CP. Screening colonoscopy use among individuals at higher colorectal cancer risk. J Clin Gastroenterol. 2006; 40(6):490–6. [PubMed: 16825930]
209. Cairns SR, Scholefield JH, Steele RJ, et al. Guidelines for colorectal cancer screening and surveillance in moderate and high risk groups (update from 2002). Gut. 2010; 59(5):666–89. [PubMed: 20427401]
210. Ramsey S, Blough D, McDermott C, et al. Will knowledge of gene-based colorectal cancer disease risk influence quality of life and screening behavior? Findings from a population-based study. Public Health Genomics. 2010; 13(1):1–12. [PubMed: 20160979]
211. Graves KD, Leventhal KG, Nusbaum R, et al. Behavioral and psychosocial responses to genomic testing for colorectal cancer risk. Genomics. 2013
212. McBride CM, Koehly LM, Sanderson SC, et al. The behavioral response to personalized genetic information: will genetic risk profiles motivate individuals and families to choose more healthful behaviors? Annu Rev Public Health. 2010; 31:89–103. [PubMed: 20070198]
213. Hsu L, Jeon J, Brenner H, et al. A predictive model of colorectal cancer risk using common genetic susceptibility loci. Gastroenterology. 2014 Submitted.
214. Dunlop MG, Tenesa A, Farrington SM, et al. Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42 103 individuals. Gut. 2012; 62(6):871–81. [PubMed: 22490517]
215. Guttmacher AE, McGuire AL, Ponder B, et al. Personalized genomic information: preparing for the future of genetic medicine. Nat Rev Genet. 2010; 11(2):161–5. [PubMed: 20065954]
216. Wei Z, Wang W, Bradfield J, et al. Large Sample Size, Wide Variant Spectrum, and Advanced Machine-Learning Technique Boost Risk Prediction for Inflammatory Bowel Disease. Am J Hum Genet. 2013; 92:1–5.
217. Chatterjee N, Wheeler B, Sampson J, et al. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nat Genet. 2013; 45(4):400–3. [PubMed: 23455638]
218. Park JH, Gail MH, Greene MH, et al. Potential usefulness of single nucleotide polymorphisms to identify persons at high cancer risk: an evaluation of seven common cancers. J Clin Oncol. 2012; 30(17):2157–62. [PubMed: 22585702]
219. Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One. 2008; 3(10):e3395. [PubMed: 18852893]
220. Wei Z, Wang K, Qu HQ, et al. From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes. PLoS genetics. 2009; 5(10):e1000678. [PubMed: 19816555]
221. Lloyd-Jones DM. Cardiovascular risk prediction: basic concepts, current status, and future directions. Circulation. 2010; 121(15):1768–77. [PubMed: 20404268]
222. Vickers AJ. Prediction models: revolutionary in principle, but do they do more good than harm? J Clin Oncol. 2011; 29(22):2951–2. [PubMed: 21690474]
223. Feero WG, Manolio TA, Khoury MJ. Translational research is a key to nongeneticist physicians' genomics education. Genet Med. 2014
224. Khoury MJ, Janssens AC, Ransohoff DF. How can polygenic inheritance be used in population screening for common diseases? Genet Med. 2013; 15(6):437–43. [PubMed: 23412608]

225. Whiffin N, Hosking FJ, Farrington SM, et al. Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. Hum Mol Genet. 2014; 23(17):4729–37. [PubMed: 24737748]

226. Schumacher FR, Stenzel SL, Jiao S, et al. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. Nat Communications. 2014 In Press.

227. Cui R, Okada Y, Jang SG, et al. Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. Gut. 2011; 60(6):799–805. [PubMed: 21242260]

228. Hutter CM, Slattery ML, Duggan DJ, et al. Characterization of the association between 8q24 and colon cancer: gene-environment exploration and meta-analysis. BMC Cancer. 2010; 10(1):670. [PubMed: 21129217]

229. Kocarnik JD, Hutter CM, Slattery ML, et al. Characterization of 9p24 risk locus and colorectal adenoma and cancer: gene-environment interaction and meta-analysis. Cancer Epidemiol Biomarkers Prev. 2010; 19(12):3131–9. [PubMed: 20978172]

230. Zhang B, Jia WH, Matsuda K, et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. Nat Genet. 2014; 46(6):533–42. [PubMed: 24836286]

231. Wang H, Burnett T, Kono S, et al. Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in VTI1A. Nat Commun. 2014; 5:4613. [PubMed: 25105248]

232. Zhang B, Jia WH, Matsuo K, et al. Genome-wide association study identifies a new SMAD7 risk variant associated with colorectal cancer risk in East Asians. Int J Cancer. 2014; 135(4):948–55. [PubMed: 24448986]

233. Pittman AM, Naranjo S, Jalava SE, et al. Allelic variation at the 8q23. 3 colorectal cancer risk locus functions as a cis-acting regulator of EIF3H. PLoS Genet. 2010; 6(9):e1001126. [PubMed: 20862326]

234. Tuupanen S, Turunen M, Lehtonen R, et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nat Genet. 2009; 41(8):885–90. [PubMed: 19561604]

235. Biancolella M, Fortini BK, Tring S, et al. Identification and characterization of functional risk variants for colorectal cancer mapping to chromosome 11q23. 1. Hum Mol Genet. 2014; 23(8): 2198–209. [PubMed: 24256810]

236. Lubbe SJ, Pittman AM, Olver B, et al. The 14q22.2 colorectal cancer variant rs4444235 shows cis-acting regulation of BMP4. Oncogene. 2012; 31(33):3777–84. [PubMed: 22158048]

237. Pittman AM, Naranjo S, Webb E, et al. The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. Genome Res. 2009; 19(6):987–93. [PubMed: 19395656]

238. Gorlov IP, Gorlova OY, Frazier ML, et al. Evolutionary evidence of the effect of rare variants on disease etiology. Clin Genet. 2011; 79(3):199–206. [PubMed: 20831747]

239. NHLBI GO Exome Sequencing Project (ESP) webpage. Sep 8. 2014 https://esp.gs.washington.edu/drupal/

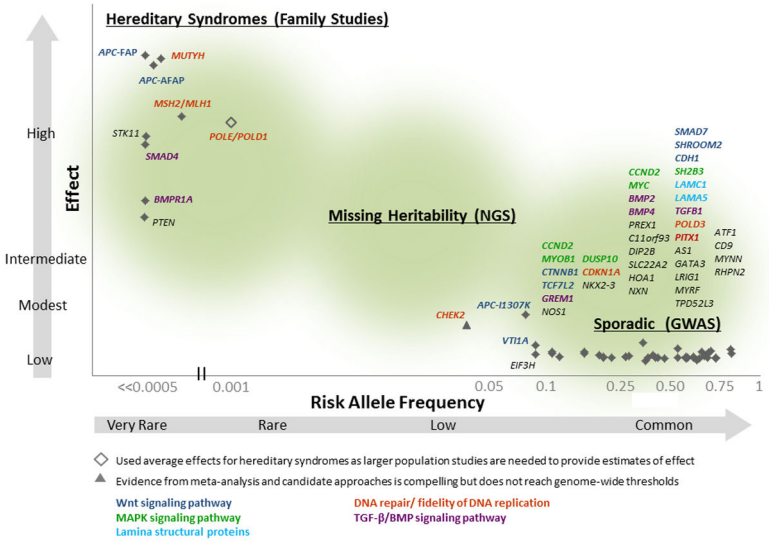

Figure 1. Genetic architecture of known CRC genetic susceptibility loci

Allele frequency shown for the risk allele frequency of the ethnicity in which the locus was discovered; except for variants with a recessive effect (MUTYH), for which the frequency of the homozygote rare allele is shown. Supplemental Table 1 provides details on each genetic variant presented in this Figure 1.

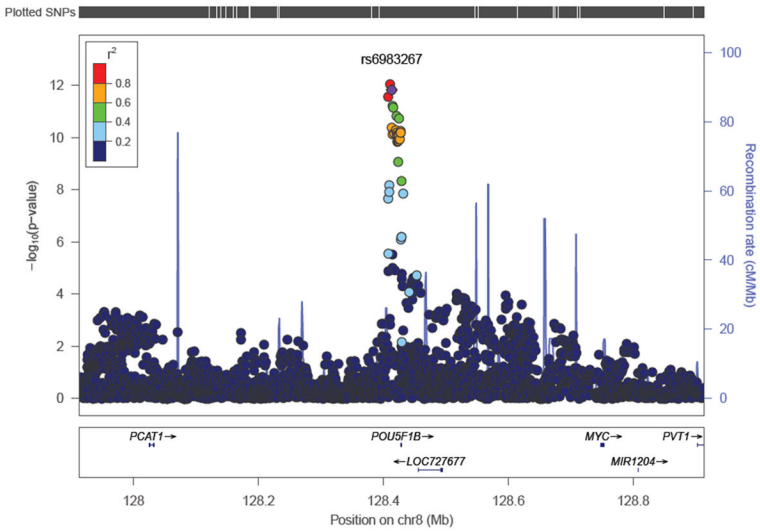

Figure 2. Fine-mapping of GWAS findings. Association results (p-values) and correlation structure for all SNPs in the 8q24 risk locus

The top part of the Figure has physical position along the x axis, and the $-\log_{10}$ of the SNP-CRC association p-value on the y-axis. Each dot on the plot represents the p-value of the association for one SNP with risk of CRC. The most significant SNP (rs6983267) is marked as a purple diamond. The color scheme represents the pairwise correlation (r^2) for the SNPs across the 8q24 region with the most significant SNP (rs6983267) based on the European descent participants from the 1000 Genomes Project data. Gray indicates that correlation was missing for this p-value because the variant had no r2 estimation due to low MAF or because the SNP is not in older versions of the 1000 Genomes data. The bottom half of the Figure shows the position of the genes across the region.

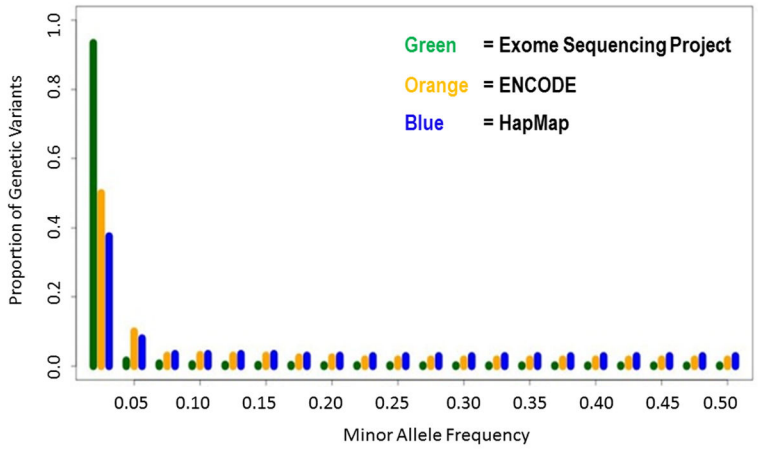

Figure 3. Most genetic variants are rare - Distribution of genetic variants by minor allele frequency (MAF)

Sources: Gorlov et al. Clin Genet 2011;[238] https://esp.gs.washingtonedu/drupal[239]

Table 1

Genes with predisposing mutations to inherited colorectal cancer syndromes

Gene	Hereditary Syndrome	Age of Onset (years)	Pathway/Biological function*
APC	FAP, AFAP	34–43	Wnt signaling pathway
MUTYH	MAP		
MLH1, MSH2,MSH6, PMS2,EPCAM	Lynch Syndrome	44–56	Mismatch repair
PTEN	Cowden syndrome (includes Bannayan-Ruvalcaba-Riley (BRR) syndrome)	<50 (BRR pediatric onset)	Negative regulator of metabolic signaling
STK11	Peutz-Jeghers Syndrome (PJS)	65	Tumour suppressor responding to changes in cellular energy balance
GREM1,15q13 locus	Hereditary mixed polyposis syndrome (HMPS)	48	TGF-β/BMP signaling pathway
BMPR1A	HMPS, juvenile polyposis syndrome	48, 42	TGF-β/BMP signaling pathway
MADH4/SMAD4	Juvenile polyposis syndrome	42	TGF-β/BMP signaling pathway
POLE, POLD1	Oligopolyposis or Polymerase proofreading-associated polyposis	23–80	DNA repair

*Many of these pathways interact at multiple levels and as such are not necessarily independent biological mechanisms

Table 2

Discovered loci with weak effects on CRC susceptibility

Locus [reference]	Independent Index SNP[a]	Risk Allele	RAF[b]	Odds Ratio (95%Confidence Interval)	Smallest P-value	Genomic Position	Closest Gene	Putative Gene Target[c]
Loci discovered by genome-wide association studies with independent replication and p-values <5×10^{-8}								
1q25.3[103,225]	rs10911251	A	0.54	1.09 (1.06–1.12)	2 × 10^{-8}	intron	*LAMC1*	*LAMC1*
1q41[63]	rs6687758	G	0.22	1.09 (1.06–1.12)	2 × 10^{-9}	intergenic	6.6kb 3′ of *RP11*	*DUSP10*
2q32.3[103]	rs11903757	C	0.16	1.16 (1.10–1.22)	4 × 10^{-8}	intergenic	34kb 3′ of *NABP1*	*MYO1B*
3p14.1[226]	rs812481	G	0.58	1.09 (1.05–1.11)	2 × 10^{-8}	intron	*LRIG1*	*LRIG1*
3p22.1[226]	rs35360328	A	0.16	1.14 (1.09–1.19)	3 × 10^{-9}	intergenic	*21kb 3′of RP11*	*CTNNB1*
3p22.2[54]	rs1800734	A	0.26	1.51 (1.34–1.69)	7×10^{-12}	5′ UTR	*MLH1*	*MLH1*
3q26.2[63]	rs10936599	C	0.75	1.04 (1.04–1.10)	3 × 10^{-8}	synonymous	*MYNN*	*TERC*
5q31.1[105]	rs647161	A	0.66	1.06 (1.02–1.11)	4 × 10^{-10}	intergenic	*PITX1*	*PITX1*
6p21[74]	rs1321311	A	0.20	1.10 (1.07–1.13)	1 × 10^{-10}	intergenic	*CDKN1A*	*CDKN1A*
6q25.3[227]	rs7758229	T	0.32	1.28 (1.18–1.39)	8 × 10^{-9}	intron	*SLC22A3*	*SLC22A2*
8q23.3[61]	rs16892766	C	0.09	1.27 (1.20–1.34)	3 × 10^{-18}	intergenic	24kb 3′ of *EIF3H*	*EIF3H*
8q24.21[57,59,61,90,103, 227,228]	rs6983267	G	0.48	1.19 (1.15–1.23)	9 × 10^{-26}	intergenic		*MYC*
9p24[57,229]	rs719725	A	0.60	1.14 (1.05–1.15)	1 × 10^{-5}	intergenic	*34kb 3′of TPD52L3*	*TPD52L3* or *UHRF2*
10p14[61]	rs10795668	G	0.67	1.12 (1.10–1.16)	3 × 10^{-13}	intergenic	*2.4kb 5′of RN5S299*	*GATA3*
10q22.3[230]	rs704017	G	0.58	1.10 (1.06–1.13)	2 × 10^{-8}	intron	*RP11.1*	*ZMIZ1 or AS1*

Locus [reference]	Independent Index SNP[a]	Risk Allele	RAF[b]	Odds Ratio (95%Confidence Interval)	Smallest P-value	Genomic Position	Closest Gene	Putative Gene Target[c]
10q24.2[225]	rs1035209	T	0.20	1.12 (1.08–1.16)	4×10^{-11}	intergenic	*16kb 5′of snoU13*	*NKX2-3*
10q25.2[231]	rs12241008	C	0.09	1.19 (1.12–1.26)	3×10^{-8}	intron	*VTI1A*	*VTI1A*
10q25.2[230]	rs11196172	A	0.11	1.14 (1.10–1.18)	1×10^{-12}	intron	*TCF7L2*	*TCF7L2*
11q12.2[230]	rs1535	A	0.64	1.09 (1.04–1.13)	8×10^{-20}	intron	*FADS1*	*FEN1*
11q13.4[74]	rs3824999	C	0.53	1.08 (1.05–1.10)	4×10^{-10}	intron	*POLD3*	*POLD3*
11q23[59]	rs3802842	C	0.27	1.11 (1.08–1.15)	6×10^{-10}	intron	*COLCA2*	*COLCA2*
12p13.31[230]	rs10849432	T	0.91	1.14 (1.09–1.18)	3×10^{-10}	intergenic	34kb 5′ of *PLEKHG6*	*CD9*
12p13.32[103,225]	rs3217810	T	0.12	1.10 (1.06–1.14)	2×10^{-10}	intron	*CCND2*	*CCND2*
12p13.32[105]	rs10774214	T	0.38	1.17 (1.11–1.23)	5×10^{-10}	intron	*CCND2*	*CCND2*
12q13.12[63]	rs11169552	C	0.75	1.09 (1.05–1.11)	2×10^{-10}	intergenic	1.8kb 5′ of *ATF1*	*ATF1*
12q13.13[63]	rs7136702	T	0.33	1.06 (1.04–1.08)	4×10	intergenic	6.4kb 3′ of *LARP4*	*DIP2B*
12q24.12[226]	rs3184504	C	0.53	1.09 (1.06–1.12)	2×10^{-8}	missense	*SH2B3*	*SH2B3*
12q24.22[226]	rs73208120	G	0.11	1.16 (1.11–1.23)	3×10^{-8}	intron	*NOS1*	*NOS1*
14q22.2[81]	rs1957636	A	0.41	1.08 (1.06–1.11)	1×10^{-9}	intergenic	135kb 5′ of *BMP4*	*BMP4*
14q22.2[61,72]	rs4444235	C	0.48	1.11 (1.08–1.15)	8×10^{-10}	intergenic	4.2kb 3′ of *MIR5580*	*BMP4*
15q13[60,61,103,104]	rs4779584	T	0.19	1.35 (1.14–1.6)	4×10^{-14}	intergenic	5.5kb 3′ of *SCG5*	*GREM1*
16q22.1[72]	rs9929218	G	0.71	1.10 (1.06–1.12)	1×10^{-8}	intron	*CDH1*	*CDH1*

Locus [reference]	Independent Index SNP[a]	Risk Allele	RAF[b]	Odds Ratio (95%Confidence Interval)	Smallest P-value	Genomic Position	Closest Gene	Putative Gene Target[c]
17p13.3[230]	rs12603526	C	0.26	1.10 (1.06–1.14)	3×10^{-8}	intron	*NXN*	*NXN*
18q21[58,59]	rs4939827	T	0.53	1.20 (1.16–1.24)	8×10^{-28}	intron	*SMAD7*	*SMAD7*
18q21[232]	rs7229639	A	0.16	1.22 (1.15–1.29)	3×10^{-11}	intron	*SMAD7*	*SMAD7*
19q13.1[72]	rs10411210	C	0.90	1.15 (1.10–1.20)	5×10^{-9}	intron	*RHPN2*	*RHPN2*
19q13.2[230]	rs1800469	G	0.69	1.09 (1.06–1.12)	1×10^{-8}	intron	*TMEM91*	*TGFB1*
20p12.3[81,105]	rs2423279	C	0.26	1.10 (1.06–1.14)	7×10^{-9}	intergenic	51kb 3′ of *HAO1*	*BMP2*
20p12.3[81,104]	rs4813802	G	0.34	1.10 (1.06–1.12)	7×10^{-11}	intergenic	12kb 3′ of *RP5-859D4.3*	*BMP2*
20p12.3	rs961253	A	0.37	1.12 (1.08–1.16)	2×10^{-10}	intergenic	23kb 5′ of *RP11-199O14.1*	*BMP2*
20q13.1[226]	rs6066825	G	0.36	1.09 (1.06–1.12)	4×10^{-9}	intron	*PREX1*	*PREX1*
20q13.33[63,104]	rs4925386	C	0.68	1.08 (1.05–1.10)	2×10^{-10}	intron	*LAMA5*	*LAMA5*
Xp22.2[74]	rs5934683	C	0.62	1.07 (1.04–1.10)	7×10^{-10}	intergenic	*GPR143*	*SHROOM2*
Loci discovered in candidate gene studies with cumulative evidence from subsequent meta-analysis considered to be strong								
22q12.1[54]	rs17879961	C	0.04	1.56 (1.32–1.84)	1×10^{-7}*	missense	*CHEK2*	*CHEK2*

[a] When more than one correlated variants (D′<0.5) is identified the most significant SNP across 21 GWAS studies is reported

[b] Risk allele frequency (RAF) from 1000 Genomes Project June 2011 release (European frequency is based on CEU+FIN+GBR+IBS+TSI and Asian frequency is based on CHB+CHS+JPT) when frequency was not reported from study control population.

[c] For loci with hypothesized regulatory potential the target gene was predicted using publicly available data and review of literature for: 1. allelic expression assays such as luciferase, 2. electromobility shift assays, 3. Roadmap activity assays in colorectal mucosa tissues, 4. chromatin capture in CRC cell lines, and 5. pathway analysis.

* False-positive report probability (FPRP) and cumulative evidence of association from meta-analysis is strong.

Table 3

Biological mechanisms marked by common risk loci

Implicated Genes[†]	Biological mechanism
	Well-established CRC pathways
CTNNB1, VTl1A, TCF7L2, CDH1, SMAD7, SHROOM2	Cell adhesion, differentiation, and migration (Wnt signaling pathway)
BMP4, BMP2, GREM1, TGFB1	Cell adhesion, differentiation, and migration (TGF-β/BMP signaling pathway)
DUSP10, MYO1B, MYC, CCND2, SH2B	Cell adhesion, differentiation, and migration (MAPK signaling pathway)
PITX, POLD3, CDKN1A, FEN1, MLH1, CHEK2	DNA repair, fidelity of DNA replication
	Novel CRC pathways
LAMC1, LAMA5	Cell adhesion (extracellular matrix structural constituent)
EIF3H	Translational initiation
CD9	Cell adhesion, differentiation, migration, and signaling (cell surface glycoprotein)
PLCB1	Intracellular transduction of extracellular signals
RHPN2	Organization of the actin cytoskeleton
PREX1	Intracellular signaling (guanine nucleotide exchange factor)
LRIG1	Intestinal tumor suppressor
TERC	Telomerase
PITX	Hormone Regulation
DNMT3B	DNA Methyltransferase involved with epigenetic modification
SLC22A2	Steroid binding
DIP2B	DNA Methylation
NOS1, ITIH	Tumor Suppressor

[†]Given that for most loci the functional variant(s) marked by the association have yet to be tested through laboratory expression assays the listed genes are based on predicted function and most likely candidate gene

Table 4

Functional evidence for variants in common risk loci

Locus	Independent Index SNP†	Predicted functional SNP (D′)	Findings from Laboratory Follow-up	Reference
8q23.3/*EIF3H*	rs16892766	rs16888589	Allele specific differential luciferase and chromosomal interaction through chromosome conformation capture (3C)	Pitmman et al. PLoS Genet. 2010[233]
8q24.21/*MYC*	rs6983267	rs6983267	Allele specific differential luciferase	Tuupanen et al. Nat Genet. 2009[92,234]
11q23/*COLCA2*	rs3802842	rs7130173	Allele specific differential luciferase and reduced binding affinity through electromobility shift assay	Biancolella et al. Hum Mol Genet. 2014[235]
14q22.2/*BMP4*	rs4444235	rs4444235	Allele specific differential luciferase	Lubbe et al. Oncogene. 2012[236]
18q21/*SMAD7*	rs4939827	rs58920878	Allele specific differential in transgenic Xenopus model and differential electromobility shift assay	Pittman et al. Genome Res. 2009[237]