Chapter 2
Genetics of Colon Cancer Susceptibility

Graham Casey

Abstract Colorectal cancer (CRC) exhibits a strong familial risk with first-degree relatives of cases having a two to three times greater risk of developing CRC than the general population. An estimated 35% of CRC cases are due to genetic factors. Highly penetrant predisposing genes have been identified for several inherited CRC syndromes (e.g., FAP, Lynch syndrome, and juvenile polyposis) through genetic linkage studies. However, despite these considerable successes, mutations in these rare syndromes explain less than 6% of CRCs and only a small fraction of familial risk. While two recently described syndromes, MUTYH-associated polyposis, with its pattern of recessive inheritance, and familial CRC type X, account for additional genetic burden, they still account for only a small fraction of CRC risk. In the last few years, considerable effort has been directed toward the identification of common, low-penetrance mutations through the promising approach of genome-wide association studies (GWAS). With respect to CRC, 15 novel disease loci have been identified through GWAS including several genes involved in the TGFβ signaling pathway. The familial and population risks explained by these loci remain small, but it is expected that additional novel susceptibility markers will be identified as larger ongoing and pooled GWAS are completed. While the role of the majority of susceptibility genes identified through linkage studies and GWAS in energy balance remains unclear, a pattern is emerging of a possible link given that several TGFβ-related genes have been implicated in energy balance including susceptibility genes identified through linkage analyses or GWAS.

G. Casey, Ph.D. (✉)
Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
e-mail: gcasey@usc.edu
1 Familial Adenomatous Polyposis

Familial adenomatous polyposis (FAP) is an autosomal, dominantly inherited condition with the defining clinical feature of the development of hundreds to thousands of adenomatous polyps throughout the colon in childhood and adolescence [1, 2]. FAP exhibits nearly 100% penetrance [3] with equal gender distribution [4] and accounts for nearly 1% of all colorectal cancers (CRCs) [5]. FAP has a variable degree of clinical expression [6], including attenuated (10–100 polyps), sparse (100–500 polyps), and profuse (>2,000 polyps) forms. Attenuated FAP (AFAP) [7] shows a delayed onset of CRC, occurring on average 12 years later than classic/profuse FAP [8, 9].

Patients with FAP can develop a variety of extracolonic tumors including upper gastrointestinal tract malignancies and cancers of the thyroid, pancreas, biliary tree, brain, and hepatoblastomas [8]. A diagnosis of FAP that also includes medulloblastoma is termed Turcot’s syndrome [10], and the association of polyposis with osteomas and desmoid tumors has been referred to as Gardner’s syndrome. FAP patients can also develop a variety of extracolonic manifestations, including duodenal and fundic gland polyps or retinal epithelium abnormalities as seen in congenital hypertrophy of retinal pigment epithelium (CHRPE) [11]. Many of these extracolonic manifestations correlate with APC-specific mutations (see later in this section).

The gene responsible for FAP, the Adenomatous Polyposis Coli (APC) gene on chromosome 5q21, was cloned in 1991 following linkage analysis in families with FAP [12–15]. APC is a large gene that encodes a protein of 2,843 amino acids [16]. It functions as a tumor suppressor and has been implicated in a number of cell processes [16–18], but the best-characterized role for APC is as part of a scaffolding protein complex that negatively regulates Wingless/WNT signaling [16, 19, 20]. This pathway has been reviewed extensively elsewhere [17, 18] and is summarized here only briefly. APC and the transcription coregulator β-catenin play central roles in the WNT signaling pathway. In normal cells, in the absence of WNT signaling, APC, along with Axin, glycogen synthase kinase 3β (GSK3β) and casein kinase, recruit β-catenin into a destruction complex where it is phosphorylated by GSK3β, leading to β-catenin degradation by the ubiquitin-mediated proteosome pathway. This cellular process leads to the maintenance of low levels of free cytosolic β-catenin in the cytoplasm. When the WNT signaling pathway is activated the APC/Axin/GSK3β complex disassociates, allowing stabilization of cytosolic β-catenin. Accumulated β-catenin associates with T-cell factor (TCF) and lymphoid-enhancer factor (LEF) and the resulting complex enters the nucleus and activates transcription. Once it enters the nucleus the β-catenin/TCF/LEF proteins provide a potent transcriptional complex leading to transactivation of a number of critical genes including MYC and cyclin D1 [18, 21, 22]. Loss of control of this pathway through mutation and inactivation of APC leads to aberrant accumulation of β-catenin, and transcriptional activation of β-catenin/TCF/LEF complexes resulting in aberrant activation of target genes [16].

APC also participates in a number of other cellular processes related to cytoskeletal organization, in particular microtubule stability [22]. The genetic evidence of the importance of deregulation of the β-catenin signaling pathway in CRC strongly implicates a central role for the WNT/APC/β-catenin pathway in CRC development.
More than 800 different disease-causing APC germ-line mutations have been reported in FAP [23]. The majority of mutations occur between codons 1250 and 1464 in the 5’ region of exon 15, a region known as the mutation cluster region (MCR) [23]. Mutations at codons 1061 and 1309 (”hot spots”) account for approximately 11 and 17%, respectively, of all germ-line APC mutations [23]. The majority of the remaining mutations occur between codons 200 and 1600 with only a few mutations falling outside this region [16]. The majority of mutations are frameshift or nonsense mutations that lead to an inactive truncated protein product [16, 24]. Approximately 10–30% APC mutations are de novo [25]. A common missense mutation (I1307K) in APC has also been reported in the Ashkenazi Jewish population [26]. While this missense mutation does not appear to have any effect on APC function, carriers do have an increased risk of CRC but not polyposis or any other extra colonic manifestations of FAP [26].

As discussed earlier, there is marked variability in the clinical phenotype of FAP, with severity of disease often correlating with location of the APC mutation [27]. For example, mutations in codon 1250 to codon 1464 and particularly codon 1309 mutations correlate with profuse polyposis where symptoms usually occur 10 years earlier than milder forms [28–34]. Mutations at the extreme 5' and 3' ends of the APC gene are generally associated with AFAP where patients develop fewer than 100 colon polyps and cancer onset is delayed [6, 35–39].

The appearance of extracolonic manifestations also correlates with the location of APC mutation. For example, mutations between codons 1310 and 2011 are associated with the appearance of desmoid tumors [28], with the highest severity occurring between codons 1444/5 and 1580/1 [29, 40–42]. Mutations between codons 140 and 1309 are often associated with the occurrence of papillary thyroid cancer [43], whereas CHRPE is often associated with mutations in codons 457–1444 [12, 44]. Gardner’s syndrome involving severe desmoids, osteomas, epidermoid cysts, and upper gastrointestinal polyps is generally associated with APC mutations in codons 1403 and 1578 [44, 45]. While no consistent genotype correlation has been found for duodenal adenomas, FAP patients with APC mutations in codons 976–1067 have been reported to have a three- to fourfold increased risk [28].

Mouse models support a critical role for APC in the development of intestinal neoplasia. Although mice homozygous for inactivated Apc are embryonic lethal, mice heterozygous for Apc (the Multiple intestinal neoplasia or Min mouse) invariably develop multiple intestinal tumors [46]. While there are some differences in the tissue specificity and morphogenesis between Min mice and FAP, Min mice have proven an important model for intestinal tumorigenesis.

2 Hereditary Nonpolyposis Colon Cancer/Lynch Syndrome

Hereditary nonpolyposis colorectal cancer (HNPCC) (more commonly referred to as Lynch syndrome) is a clinically heterogeneous disease that has historically been diagnosed based on family history criteria (Amsterdam and Bethesda criteria) that are not very accurate [47–49]. Lynch syndrome is characterized by a high incidence
of CRC and endometrial cancer in families. The lifetime risk for colon cancer in Lynch syndrome subjects is approximately 50–60% [50]. There is increased incidence of extracolonic cancers in both males and females including those of the small bowel, stomach, pancreas, ovary, renal pelvis, ureter, bladder, brain, appendix, liver, bile duct, gall bladder, and skin [49, 51, 52]. Colon cancers arising in Lynch syndrome families have a propensity toward left sidedness with two-thirds arising in the proximal colon [51–53]. These tumors show a variety of common histologic features including tumor-infiltrating lymphocytes, mucinous or signet ring differentiation, and a medullary growth pattern [48, 49, 53, 54].

Like FAP, Lynch syndrome is an autosomal, dominantly inherited condition. However, Lynch syndrome is more challenging to diagnose than FAP because the clinical phenotype is far more varied and more genes are involved. The majority of Lynch syndrome cases are accounted for by mutations in one of four genes (MSH2, MLH1, MSH6, or PMS2) involved in DNA mismatch repair (MMR). Of those cases with defective MMR, approximately 80–90% have germ line mutations in one of these genes. The majority of cases are due to mutations in MSH2 and MLH1 that play central and critical roles in DNA MMR [55], with MSH2 forming a heterodimer with MSH6 and to a lesser extent MSH3, and MLH1 with PMS2.

In newly replicated DNA, mismatches such as G>T [56] are recognized by MSH2–hMSH6 heterodimers (MutS alpha in yeast), whereas insertion–deletion loops are recognized primarily by MSH2–MSH6 heterodimers, but can also be mediated by the less abundant MSH2–MSH3 (MutS beta) heterodimeric protein complex that appears to function as a backup in the absence of MSH6. Loss of MSH2 therefore leads to the accumulation of aberrant length repeat sequences such as (A)n or (CA)n and high levels of Microsatellite Instability (MSI). Once the MSH2–MSH6 heterodimer recognizes DNA mismatches, this complex undergoes an ATP-dependent conformational change converting it to a sliding DNA clamp capable of moving away from the repair site [57, 58]. This is followed by the recruitment to the complex of MLH1–PMS2 heterodimers (MutL-alpha) [59]. This is then followed by exonuclease degradation of a few hundred bases of the newly synthesized mutant DNA strand followed by resynthesis of the complementary strand by DNA polymerase. As mutations in MSH2, MLH1, MSH6, and PMS2 do not appear to account for all MMR deficient cases it is possible that other MMR genes have yet to be identified [59]. A detailed description of the role of these proteins in DNA MMR and their specific roles in Lynch syndrome can be found in several reviews [60, 61].

Defective MMR repair was recognized as the underlying genetic basis for Lynch syndrome following the observation by three independent groups that MSI was a hallmark feature of tumors arising in Lynch syndrome family members [62–65]. MSI, also referred to as a replication error (RER) or “mutator” tumor phenotype [62, 63, 65], occurs as a result of failure to repair of errors in copying during DNA replication. Thousands of microsatellite short tandem repeat DNA sequences (mono-, di-, tri-, or tetranucleotides) exist throughout the human genome, and errors can occur during DNA replication when copying these sequences. Typically such misalignment errors would be repaired by the DNA MMR system. However, in cells with defective MMR repair, these errors are not repaired effectively, and tumor DNAs of
Lynch syndrome family members reveal a “stuttering” (loss or gain of one or more repeats) pattern of microsatellite markers when compared with DNA from normal cells from the same subject. Once it was recognized that the MSI phenotype was similar to the mutational spectrum seen in yeast caused by deletion or mutation of MMR genes, the *MSH2* and *MLH1* genes that account for the majority of Lynch syndrome cases were identified within a year [56, 66].

Germ-line mutations in *MLH1* and *MSH2* account for the majority of mutations found in families with Lynch syndrome with a smaller minority attributable to mutations in *MSH6* and *PMS2*. Germ line testing remains a challenge as mutations can occur throughout any of these relatively large genes and are not localized to any mutation hot spots as in the *APC* gene. *MSH2* consists of 935 amino acids over 16 exons, *MLH1* consists of 756 amino acids over 19 exons, *MSH6* consists of 1,360 amino acids over 10 exons, and *PMS2* consists of 862 amino acids over 15 exons. A wide range of types of mutations has been reported in these genes including missense, nonsense and splice site mutations. In addition, a number of large genomic deletions or rearrangements involving several exons have also been reported [67–73]. Testing for *PMS2* germ-line mutations is not straightforward as there are several highly homologous *PMS2* pseudogenes, the majority of which have homology with at least some of the ten exons at the 3’ end of the gene [74–77]. A comprehensive listing of MMR gene mutations can be found on the Mismatch Repair Genes Variant Database [78] and the MMR Gene Unclassified Variants database (http://www.mmrmissense.net/), which focuses more on functional assays and other types of data to support the interpretation of the unclassified variants in MMR genes.

Nearly 90% of Lynch syndrome colon tumors exhibit high levels of MSI [62, 65, 79], and there exists a strong correlation between MSI and loss of staining of MMR proteins using immunohistochemistry (IHC). As a result, IHC of the four MMR proteins along with an assessment of family history has been recommended as a starting point for diagnosing Lynch syndrome [79, 80]. However, it should be noted that the sensitivity of IHC staining is not as high as MSI analysis as not all MMR mutations lead to a loss of protein expression [81–83].

While defects in MMR are seen in nearly 15% of CRCs, tumors with MMR germ-line mutations account for less than 5% of all cases. This is because MMR defects are also seen in a subset of “sporadic” CRCs through somatic hypermethylation and inactivation of MLH1 [84]. “Sporadic” MSI-H tumors share many of the characteristics of those arising in MMR mutation carriers, including a tendency toward a proximal location in the colon and a mucinous phenotype, but they usually occur later in life. Although these cancers generally arise in the absence of a positive family history, a vertical transmission in some families has been reported [85–87].

There is some evidence that *MLH1* and *MSH2* mutation families exhibit different clinical expression. Several studies have been published, with overall findings of greater CRC risk, earlier CRC onset, and fewer extracolonic tumors in *MLH1* mutation carriers compared with *MSH2* mutation carriers [50, 88–95]. Clinically, identification of an MMR gene defect, whether occurring within the context of Lynch syndrome or sporadically, is important as it affects response to some chemotherapeutic agents and ultimately prognosis [96–99].
3 MUTYH-Associated Polyposis

Recent studies have identified germ-line mutations in the mutY homologue MUTYH (also called MYH) with a recessive mode of inheritance associated with high risk of multiple adenomatous polyps (10–1,000) and CRC in up to 50% of APC-negative polyposis cases [100–102]. MUTYH mutations account for nearly 1% of all CRC cases [103]. The majority of cases are associated with a relatively small number of common variants (around 0.2% population frequency in Caucasians) [104–106]. Biallelic carriers develop multiple polyps by 45–55 years, although this may be an overestimate as large population-based studies have not yet been conducted [103, 105, 107].

The MUTYH gene was implicated in CRC risk following the observation in tumors of APC mutation-negative multiple polyposis families that the APC gene harbored an excess of somatic G:T transversions [100]. Such mutations are hallmarks of oxidative DNA damage. This led Al-Tassan and coworkers to investigate a possible role for a constitutional defect in base excision repair (BER) and the subsequent identification of two germ-line variants (Y179C and G396D) in MUTYH that segregated with disease in family members [100]. The majority of MUTYH carriers are accounted for by these two common missense mutations (44 and 24%, respectively) with a number of additional rare MUTYH missense mutations including some truncating mutations accounting for a small fraction [101–106, 108–112]. The Y179C MUTYH variant correlates with a more severe phenotype than G396D, manifesting at an earlier age of onset of polyposis and a greater risk of developing CRC than the Y179C allele [104]. Some studies have suggested that monoallelic MUTYH mutations may be associated with an increased risk of CRC, but this remains controversial [102, 104–106, 111, 113–116].

MUTYH is involved in BER of DNA damage caused by reactive oxygen species (ROS) produced through cellular metabolism or exposure to ionizing radiation. Among the lesions caused by oxidative DNA damage is 8-oxoguanine (8-oxoG). 8-oxoG is stable and highly mutagenic product prone to post-DNA replication mispairing. MUTYH is a DNA glycosylase involved in the identification and removal of mismatched adenines incorporated opposite 8-oxoG during replication. Failure to correct 8-oxoG:A mispairing leads to characteristic G:C to T:A transversions in the next cycle of DNA replication [117, 118]. Two other enzymes, MTH1 and OGG1, also play critical roles in BER [119, 120], but to date no mutations in these genes have been linked convincingly to increased risk of either colorectal polyposis or CRC [121].

There are few discriminatory features to MUTYH-related CRC. While CRC can occur throughout the colon in MUTYH carriers [104, 105], there is an excess of proximal cancers [101–103, 109, 122]. There are no characteristic histopathology or clinicopathologic features [103–105, 123], and tumors are microsatellite stable [104, 105, 109, 124]. Gastroduodenal polyposis has been observed in nearly 20% of MUTYH biallelic carriers [125–127], but this is likely to be an overestimate as these studies were conducted in highly selected polyposis registry families. MUTYH variants have been implicated in a number of cancers including lung, breast, gastric, and endometrial cancers. However, there remains no definitive evidence for an elevated risk of such cancers.
4 Familial Colorectal Cancer Type X

Over the last few years, there has been growing recognition that many families that fulfill HNPCC Amsterdam 1 criteria do not harbor an inherited MMR mutation [93, 128]. Growing evidence suggests that this may reflect a separate syndrome.

In a large study using the resources of the Colon Cancer Family Registry [129], Lindor et al. compared 90 Amsterdam I families with MMR-deficient tumors with 71 Amsterdam I families with MMR-proficient tumors and showed that families with MMR-deficient tumors had a statistically significantly elevated risk of developing colorectal, endometrial, gastric, small intestine, and kidney cancers as expected for Lynch syndrome. In contrast, while there was a twofold increased risk of CRC in the families with MMR-proficient tumors, there was no increased risk of any other cancer site [130]. The average age at diagnosis of CRC was also later (61 years) in families with normal MMR compared to families with MMR deficiency (49 years). Based on these data, the authors concluded the normal MMR families that met Amsterdam I should not be considered Lynch syndrome families and coined the name “familial colorectal cancer type X” (FCCTX) [130].

A number of studies have now been published that support these findings and strongly imply that FCCTX should be regarded as a distinct syndrome(s) rather than a missed diagnosis of Lynch syndrome [131–133]. In support of this, FCCTX cases are more likely to be diagnosed at a later age than Lynch syndrome cases despite having a similar incidence of adenomas, are less likely to develop multiple primary tumors, and tumors are less likely to have Lynch syndrome characteristics such as a propensity toward right-sidedness, or a mucinous or tumor-infiltrating lymphocyte pathology [113, 134–136]. While the molecular phenotype of FCCTX tumors appears to differ from that of Lynch syndrome tumors, the phenotype does not appear to be distinct from that of sporadic CRC [137, 138].

FCCTX is likely to be a heterogenous group including families with a chance aggregation of CRC, families with an undiagnosed syndrome such as MUTYH-associated polyposis [113] or MSI-variable families [139], and families with an as yet to undiscovered syndrome.

5 Hamartomatous Polyposis and Other Rare Syndromes

Several familial syndromes have been described that are characterized by multiple hamartomatous polyps in the intestinal tract including Cowden disease, Peutz–Jeghers syndrome, and juvenile polyposis syndrome. Hamartoma refers to an excessive focal overgrowth and distorted architecture of cells and tissues native to the organ in which it occurs. These rare syndromes are all inherited in an autosomal dominant fashion, and specific genetic mutations have been identified. A more extensive review of these syndromes has recently been published [140].

Cowden disease is an autosomal dominant disease characterized by intestinal hamartomas, facial trichilemmomas, oral papillomas, goiter, and esophageal glycogenic
acanthosis [141–143] with an estimated incidence of 1 in 200,000. Breast and thyroid cancer risk is also pronounced in Cowden disease, with CRC developing in up to 10% of patients. Cowden disease and several related syndromes such as Bannayan–Ruvalcaba–Riley syndrome, proteus syndrome, and proteus-like syndrome are all associated with germ-line mutations in the PTEN (phosphatase and tensin homolog deleted on chromosome 10) gene. Clinical features include benign and malignant neoplasms of the thyroid, breast, uterus, and skin as well as hamartomatous intestinal polyps [144].

PTEN modulates G1 cell cycle progression through negatively regulating the survival signal mediated by the phosphatidylinositol 3-kinase (PI3K)/AKT pathway [145]. Inactivation of PTEN though mutation or deletion leads to the activation of AKT [146], increased cell proliferation and reduced apoptosis. Germ-line mutations in PTEN have been identified in approximately 80% of subjects diagnosed with Cowden syndrome. PTEN promoter mutations may account for at least another 10% of Cowden cases [147], and the remaining cases may arise from as yet undiscovered mutations in PTEN [148]. There appears to be a different pattern of mutation in Bannayan–Ruvalcaba–Riley syndrome cases. PTEN germ-line mutations account for 50–60% of patients, and large genomic deletions or rearrangements of exons of PTEN have been reported in Bannayan–Ruvalcaba–Riley syndrome patients but not Cowden syndrome patients. In addition, PTEN promoter mutations are uncommon in Bannayan–Ruvalcaba–Riley syndrome patients [143, 147, 149].

Peutz–Jeghers syndrome is a rare (approximately 1 in 200,000) autosomal dominant disorder characterized by the presence of pigmentation of the lips, buccal mucosa, hands, and feet; hamartomatous polyps throughout the gastrointestinal tract; and increased risk for gastrointestinal, breast, ovarian, and testicular cancers [150, 151]. The cumulative risk is around 30% for CRC and 50% for breast cancer [6].

Nearly half of Peutz–Jeghers cases are due to germ-line mutations in STK11/ LKB1 [152, 153]. STK11/LKB1 is a serine–threonine kinase that phosphorylates and activates AMP-activated protein kinase an essential positive regulator the mTOR pathway that is also implicated in the PTEN hamartomatous syndrome [146]. Genotype–phenotype correlation suggests that patients with Peutz–Jeghers, who have a truncation mutation in STK11/LKB1, have a significantly earlier age of onset than those who have a missense mutation or when no mutation is detected in STK11/LKB1 [154]. There are some families with Peutz–Jeghers syndrome that did not show linkage to the STK11/LKB1 chromosomal region suggesting genetic heterogeneity of this disease [155, 156].

Juvenile polyposis syndrome is a rare (1 in 100,000 births) autosomal dominant condition. It is characterized by juvenile polyps, which are distinctive hamartomas that have a smooth surface and are covered by normal colonic epithelium [157]. The polyps may affect not only the colon and rectum but also the proximal gastrointestinal tract. The clinical diagnosis consists of the following criteria: more than five juvenile polyps of the colorectum, or multiple juvenile polyps throughout the gastrointestinal tract, or any number of juvenile polyps and a family history of juvenile polyps [158]. The lifetime risk approaches 60% and patients are also at risk of developing cancers of the stomach and small intestine [159]. Germ-line mutations
in the TGFβ signaling genes SMAD4/MADH4 and BMPRIA account for around 20% of juvenile polyposis cases each [160–164]. More recently, mutations have been identified in a third gene, ENG, but the frequency remains unknown [165, 166]. Clinically, patients with an SMAD4/MADH4 mutation are more likely to develop large gastric polyps than those with a BMPRIA mutation and these patients usually have a family history of upper gastrointestinal polyposis [36, 167].

Hereditary mixed polyposis syndrome (HMPS) is characterized by colonic polyps of mixed hyperplastic, adenomatous, and occasional juvenile types that eventually lead to the development of CRC [168]. The syndrome is similar to FAP in that it is an autosomal dominantly inherited condition. However, unlike the excessive number of adenomas seen in FAP, the polyps in HMPS are fewer in number, of mixed histology, and appear to be confined to the large bowel. Using a linkage approach, the BMPRIA gene was identified and an 11-bp deletion in the BMPRIA gene found in one family [168]. BMPRIA mutations were later confirmed in other families [169, 170].

Germ-line mutations in BMPRIA have been previously associated with a subset of juvenile polyposis syndrome patients [36, 161, 162]. However, the phenotypic features of the two families in this study differ from JPS. Just as germ-line mutations in APC can cause diverse phenotypic manifestations including those of Turcot and Gardner syndromes, it is perhaps not surprising that mutations in BMPRIA could be responsible for two different syndromes.

6 Genome-Wide Association Studies and Low-Penetrance Mutations

Over the last 5 years, genome-wide association studies (GWAS) have provided a powerful new approach for identifying susceptibility loci. Rather than focusing on the highly penetrant rare mutations described above, GWAS focus on the identification of common, low-penetrance mutations. As with linkage studies, GWAS represents an agnostic survey of the genome, but unlike linkage analyses that use a relatively small number of markers to screen cancer-dense families, GWAS employs SNP arrays containing hundreds of thousands of SNPs to screen relatively large populations. GWAS have only become possible in recent years due to major technological advances in the development of genotyping platforms that allow cost-effective high throughput genotyping of large sample sets. This approach has begun to reveal novel findings that are improving our understanding of the contribution of common alleles to risk of many complex genetic disorders including CRC.

GWAS have met with unprecedented success for a range of complex diseases [171]. As of the second quarter of 2011, there have been 1449 published genome-wide associations (at \( p < 5 \times 10^{-8} \)) for 237 traits [172] and this number is expected to increase substantially over the next few years. With respect to CRC, as of January 2011, 15 novel disease loci have been identified in European populations [173–180]. Table 2.1 summarizes the
<table>
<thead>
<tr>
<th>Study reference</th>
<th>Initial sample</th>
<th>Genotyping platform [SNPs]</th>
<th>SNP</th>
<th>Gene/region</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Houlston, Nat Genet [180]</td>
<td>3,334/4,638</td>
<td>Illumina [various]</td>
<td>rs6691170</td>
<td>1q41</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs6687758</td>
<td>1q41</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs10936599</td>
<td>3q26.2</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs11169552</td>
<td>12q13.13</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs7136702</td>
<td>12q13.13</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs4925386</td>
<td>20q13.33</td>
<td>0.93</td>
</tr>
<tr>
<td>Houlston, Nat Genet [179]</td>
<td>1,902/1,929</td>
<td>Illumina [up to 548,586]</td>
<td>rs961253-A</td>
<td>20p12.3</td>
<td>1.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs4444235-C</td>
<td>14q22.2</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs10411210-C</td>
<td>19q13.11</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs9929218-A</td>
<td>16q22.1</td>
<td>1.10</td>
</tr>
<tr>
<td>Tenesa, Nat Genet [181]</td>
<td>981/1,002</td>
<td>Illumina [541,628]</td>
<td>rs4939827-T</td>
<td>18q21 SMAD7</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs7014346-A</td>
<td>8q24</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs3802842-C</td>
<td>11q23</td>
<td>1.11</td>
</tr>
<tr>
<td>Tomlinson, Nat Genet [177]</td>
<td>922/927</td>
<td>Illumina [547,647]</td>
<td>rs10795668-A</td>
<td>10p14</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs16892766-A</td>
<td>8q23.3 EIF3H</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs6983267</td>
<td>8q24.21</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs4779584</td>
<td>15q13.3</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs4939827</td>
<td>18q21.1</td>
<td>1.18</td>
</tr>
<tr>
<td>Broderick, Nat Gen [174]</td>
<td>940/965</td>
<td>Affymetrix [547,647]</td>
<td>rs4939827-T</td>
<td>18q21 SMAD7</td>
<td>1.16</td>
</tr>
<tr>
<td>Tomlinson, Nat Genet [175]</td>
<td>930/960</td>
<td>Illumina [547,647]</td>
<td>rs6983267-G</td>
<td>8q24</td>
<td>1.21</td>
</tr>
<tr>
<td>Zanke, Nat Genet [173]</td>
<td>1,257/1,336</td>
<td>Illumina and Affymetrix [99,632]</td>
<td>rs10505477-A</td>
<td>8q24</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs719725</td>
<td>9p24</td>
<td>1.14</td>
</tr>
</tbody>
</table>

*aNHGRI GWAS database (http://www.genome.gov). All studies in Caucasians

bCases/controls
published findings from these studies, and without exception the risks conferred have been low with odds ratios between 1.1 and 1.3 [173–180]. To date these studies have been limited to individuals of European ancestry.

So what candidate genes have been identified through CRC GWAS? Of the 15 SNPs identified to date, 6 map to regions that include TGFβ signaling pathway genes, a pathway that previously has been implicated in CRC. These include SMAD7 [174, 179], GREM1 [176], RHPN2, the bone morphogenetic protein genes BMP2 and BMP4 [179], and most recently LAMA5 that is required for the production of noggin, a secreted BMP antagonist [180]. TGFβ proteins play critical roles in proliferation, differentiation, cell migration, adhesion, and extracellular matrix (ECM) production [182, 183], and also energy balance (see below), and several lines of evidence support a key role for the TGFβ pathway in CRC susceptibility. For example rare, high-penetrance variants in other TGFβ-related genes (SMAD4 and the BMP receptor BMPR1A/ALK3) have been reported for juvenile polyposis [36, 161, 162] and for HMPS [169, 170]. In addition, somatic mutations of SMAD4 and the TGFβ receptor TGFBR2 have been identified in CRC tumors. The cancer initiation properties of TGFβ seem to be distinct from those of progression, as activation of the TGFβ signaling pathway leads to enhanced tumor growth and increased metastatic potential [184].

In addition to TGFβ signaling candidates, there are intriguing findings for some of the other CRC loci identified through GWAS. For example, 8q24 is a gene desert region that has been identified as a risk locus for several different cancers including CRC [173, 175, 185–188]. While no known genes map to this region, the MYC oncogene maps within 300–400 kb of several independently associated SNPs. Replication, sequencing, and fine-mapping studies of this locus have identified rs6983267 as the most promising variant for functional assessment in CRC and other cancers [189]. This SNP lies in a sequence that is highly conserved across vertebrates and is predicted to have regulatory function [189]. Its relative proximity to MYC makes it plausible that it may disrupt a putative enhancer. However, while MYC is often amplified in CRC, this variant has not been found to correlate with MYC expression in CRC tumors or lymphoblastoid cell lines [190], although tissuespecific long-range chromatin loops between putative enhancer elements in this region and MYC have been shown [191]. Many of the other associated loci (e.g., 9p24, 10p14, 11q23.1, 18q23, and 20p12.3) also lie in intergenic or gene desert regions with no known biological relevance.

It is important to note that any candidate genes identified through GWAS, including those belonging to the TGFβ signaling pathway, have not yet been confirmed as causal, and there is growing emphasis on dissecting the functional consequences of GWAS findings [192]. One of the challenges for GWAS is that they rarely identify the causal variant or gene, as the SNPs that are included on commercial SNP arrays are chosen to capture regions of linkage disequilibrium (LD) identified through the HapMap project [193] rather than for any functional or putative functional role. As a result, the nearest gene mapping adjacent to an associated SNP may not be the causal gene. Considerable work is needed before functionality can be assigned to any susceptibility SNPs. This is not a trivial task as most effect sizes
are relatively small and the functional effect of any causal SNP is likely to be subtle. In addition, the majority of disease-associated SNPs identified through GWAS map to intergenic regions or gene “deserts” such as the 8q24 region [194] described above, suggesting that they affect regulatory elements such as enhancers, posing even greater challenges. Undoubtedly, a large amount of work will be needed to clarify the biological implications of these associations.

Only limited data are available with regards to the epidemiological characteristics of GWAS associations. Rs3802842 at 11q23 and rs4939827 (SMAD7) have been reported to be more strongly associated with rectal cancer than colon cancer [178]. No differences in risk have been reported by tumor molecular subtypes for the published variants, with the exception of rs4444235 (BMP4) for which the association was found to be significantly stronger for MMR proficient than deficient tumors [179]. Low-penetrance susceptibility alleles may function as modifier genes contributing to the severity of CRC in high-risk subjects. In two large studies of Lynch syndrome, two GWAS hits (rs16892766 on 8q23.3 and rs3802842 on 11q23.1) were significantly associated with an increased CRC risk in these patients [195, 196].

The familial and population risks explained by CRC GWAS loci remain small accounting for less than 10% of overall inherited risk and less than 1% of familial risk [179], and as a result they are not yet useful for risk prediction. However, it is expected that risk prediction will improve as additional susceptibility alleles are identified once ongoing, larger and pooled GWAS analyses as well as studies in other ethnic populations are completed [173–180]. In terms of risk, studies suggest that around 100 SNPs would be required to achieve 80% accuracy of prediction of CRC genetic risk [181], accounting for ~17% of the phenotypic variance providing useful predictive value. This does not take into account the contribution of rare or private variants and their effect on risk are unknown. It will take several years to more fully comprehend the impact of rare variants on CRC risk as these types of studies can only be accomplished through next generation sequencing GWAS that are just being contemplated.

It is clear that CRC etiology has a very strong environmental component [197, 198] and there are several ongoing studies examining the relationship between lifestyle risk factors for CRC and interactions with the risk alleles identified through GWAS (gene × environment interactions). Pooling of GWAS data through collaborative efforts should improve power to detect both gene × environment and gene × gene interactions [199].

7 CRC Susceptibility Genes and Energy Balance

As discussed above, while promising progress has been made in identifying CRC susceptibility genes through linkage analyses and GWAS, the susceptibility alleles identified to date still only account for a small fraction of CRC risk. Despite this, a growing understanding of the genetic etiology of CRC is beginning to emerge as
a significant number of susceptibility genes or candidate susceptibility genes belong to the TGFβ/BMP superfamily, including SMAD4, BMPRIA/ALK3, SMAD7, GREM1, RHPN2, BMP2, BMP4, and LAMA5. The TGFβ family of proteins is a well-known key regulator of many biological processes, and several lines of evidence implicate TGFβ1 signaling in energy balance. A review of the role of TGFβ in regulating adiposity and energy expenditure was recently published [200].

TGFβ is a negative regulator of adipogenesis, promoting preadipocyte proliferation while simultaneously inhibiting differentiation [201], a process augmented by SMAD7 (and SMAD6), a negative regulator of TGFβ signaling. TGFβ may also influence adipogenesis indirectly through upregulation of WNT signaling, a cascade that also inhibits adipocyte differentiation [202]. That APC mutations lead to the activation of WNT signaling may also implicate APC in energy balance. TGFβ1 expression also correlates with body mass index and visceral fat obesity, which along with insulin resistance, plays a central role in metabolic syndrome [203–207], and elevated serum TGFβ1 levels are associated with incident type 2 diabetes [208]. These findings are supported by observations in genetically engineered mice [209].

Several lines of evidence also support a role for BMPs in adipogenesis [210]. BMPs appear to play dual roles in this process. The candidate CRC susceptibility gene BMP4 is best recognized for its role in the earliest stages of white adipocyte differentiation [211, 212]. BMP4 promotes the formation of white adipocytes in a dose-dependent manner in mouse embryonic stem cells [211, 213] a finding supported by mouse studies [214]. Several lines of evidence suggest that BMP4 is an important risk factor for metabolic syndrome [215, 216]. BMP4 was associated with increased adiposity [217], recognized as being essential for energy balance [218], and white fat differentiation [212, 214, 219]. Serum BMP4 levels also correlated with body mass index, waist circumference, waist-to-hip ratio, triglycerides, HDL cholesterol, and fasting plasma insulin [216]. BMP4 mRNA expression has also been shown to correlate with obesity in ob/ob transgenic mice [219].

The CRC candidate susceptibility gene BMP2 has also been implicated in adipogenesis both as a pro- and anti-adipogenic protein. BMP2 has been shown to promote osteoblast differentiation while suppressing adipocyte development [220]. In contrast, BMP2 can also stimulate adipocyte differentiation [221–223].

The cellular response to BMP2 and BMP4 is mediated by ligand binding to cell surface receptors including BMPRIA [224, 225], a gene that has been implicated in both HMP5 and JPS patients. BMPRIA has been shown to be involved in adipocyte differentiation in vitro [105]. BMPRIA has been strongly implicated in obesity, where BMPRIA mRNA expression was elevated in patients with obesity, type 2 diabetes, and components of metabolic syndrome including body mass index, body mass, and waist-to-hip ratio [216]. Furthermore, BMPRIA mRNA levels were elevated in adipose tissues of obese and overweight adults and three SNP variants in the BMPRIA gene were associated with increased body mass index [225].

A pattern is, therefore, emerging of a possible link between some CRC susceptibility genes and energy balance that warrants further investigation. Based on growing evidence of a link between TGFβ-related genes, CRC susceptibility and the
development of features of metabolic syndrome, modulation of TGFβ signaling may represent a valuable therapeutic approach in at-risk individuals.

References

121. Dallosso AR, Dolwani S, Jones N et al (2008) Inherited predisposition to colorectal adenomas caused by multiple rare alleles of MUTYH but not OGG1, NUDT1, NTH1 or NEIL 1, 2 or 3. Gut 57:1252–1255


