

Review

Architecture of Inherited Susceptibility to Colorectal Cancer: A Voyage of Discovery

Nicola Whiffin * and Richard S. Houlston

Molecular and Population Genetics Team, Genetics and Epidemiology, The Institute of Cancer Research, Sutton, SM2 5NG, UK; E-Mail: richard.houlston@icr.ac.uk

* Author to whom correspondence should be addressed; E-Mail: nicola.whiffin@icr.ac.uk.

Received: 25 December 2013; in revised form: 7 March 2014 / Accepted: 10 March 2014 /

Published: 27 March 2014

Abstract: This review looks back at five decades of research into genetic susceptibility to colorectal cancer (CRC) and the insights these studies have provided. Initial evidence of a genetic basis of CRC stems from epidemiological studies in the 1950s and is further provided by the existence of multiple dominant predisposition syndromes. Genetic linkage and positional cloning studies identified the first high-penetrance genes for CRC in the 1980s and 1990s. More recent genome-wide association studies have identified common low-penetrance susceptibility loci and provide support for a polygenic model of disease susceptibility. These observations suggest a high proportion of CRC may arise in a group of susceptible individuals as a consequence of the combined effects of common low-penetrance risk alleles and rare variants conferring moderate CRC risks. Despite these advances, however, currently identified loci explain only a small fraction of the estimated heritability to CRC. It is hoped that a new generation of sequencing projects will help explain this missing heritability.

Keywords: colorectal cancer, genetics, susceptibility

1. Introduction

Colorectal cancer (CRC) is the third most common cancer worldwide with half a million new individuals diagnosed annually [1]. In the UK CRC affects ~40,000 individuals and is responsible for ~16,000 deaths each year (Cancer Research UK, 2013) amounting to a life-time risk of 5%–6%. It is now an established fact that inherited susceptibility has an important role in predisposition to CRC.

The earliest evidence for this came from epidemiological studies in the 1950s which showed a two- to three-fold increased risk of CRC in first degree relatives of patients [2]. Subsequent studies have identified a number of CRC susceptibility genes. These discoveries have greatly increased our understanding of the mechanisms underlying CRC biology and have provided promising targets for therapeutic intervention. Moreover, the ability to identify individuals at increased risk of CRC is of important clinical relevance.

2. Early Models of Genetic Susceptibility

Large families with multiple individuals affected by CRC have long been reported in the clinic. It was not until the late 1950s, however, that epidemiological studies attempted to quantify this familial clustering by comparing the incidence of CRC in first degree relatives (FDRs) of cases to control groups [3–5]. Recent analysis estimates an approximate two-fold increase in risk in FDRs [2]. This risk increases further to four-fold when the relative is diagnosed with early-onset CRC (<45 years of age), indicative of colorectal tumours developing in genetically susceptible individuals at an earlier age.

In 1969 Ashley [6] proposed that colonic cancer development could be ascribed to a series of carcinogenic “hits” on normal intestinal mucosa cells. He further noted that the number of necessary “hits” was lower for genetically susceptible individuals with the Mendelian predisposition syndrome familial adenomatous polyposis (FAP). In the same year, DeMars [7] suggested that FAP, along with other apparently autosomal dominant syndromes, is autosomally recessive at the cellular level; individuals with a germline mutation in one allele of a tumour suppressor gene develop cancer as a result of subsequent somatic mutations in the other gene copy.

Anderson [8], in 1974, made the first argument for a polygenic mechanism to cancer susceptibility based on the increased risk in FDRs of cancer patients being limited to two- to three-fold. He stated these results were “not indicative of strong genetic effects” and rather suggested a mechanism involving many genes with small effects acting in concert with environmental factors with larger effects. Whilst there is growing evidence to suggest that his conclusion is, at least in part, correct, the reasoning behind this statement is flawed as the relative risks associated with FDRs are likely to be underestimated. This is because calculations of relative risks typically include both sporadic and genetically susceptible cases that are then compared to the general population which, to compound the problem, also contains individuals that are genetically susceptible to CRC.

3. Identification of Rare High-Penetrance Susceptibility Alleles

Evidence for Mendelian transmission of CRC was first provided by reports of large families with CRC segregating in a dominant fashion. Perhaps the most notable case report is “family G” first described in 1913 by Warthin and subsequently revisited by Lynch and Krush in 1971 [9]. This family of over 650 blood relatives provided scientists with one of the longest, most detailed cancer genealogies in the world and was instrumental in establishing the syndrome of hereditary non-polyposis colorectal cancer (HNPCC or Lynch syndrome).

Family based genetic linkage and positional cloning studies in the late 1980s and early 1990s led to the identification of numerous CRC susceptibility genes (Table 1). The *APC* gene on chromosome 5 was the first gene to be shown to be associated with CRC susceptibility when it was identified as

mutated in FAP patients [10–14]. Subsequently, mutations in genes of the mismatch repair (MMR) pathway, particularly *MSH2*, *MLH1* and *MSH6*, the TGF- β signalling pathway genes, *SMAD4* and *BMPR1A*, and the serine/threonine kinase gene *STK11*, were revealed as the causes of HNPCC [15–26], Juvenile Polyposis syndrome (JPS) [27,28] and Peutz-Jeghers syndrome (PJS) [29] respectively. Risk alleles in these genes are rare (<0.1%) and confer a >10 fold increase in risk of CRC. These genes are tumour suppressors conforming to DeMars’ “two-hit” model of cancer susceptibility through an apparently dominant mode of inheritance. The clinical utility of testing for high penetrance mutations in these genes has long been established and identification of individuals with such mutations has been shown reduce CRC incidence through prevention strategies and screening [30,31].

Table 1. Colorectal cancer predisposition syndromes and associated high-penetrance mutations.

Gene(s)	Syndrome	Risk in mutation carriers	Mode of inheritance	References
APC	FAP	90% by age 45	Dominant	[10–14]
Mismatch repair (<i>MLH1/MSH2/MSH6/PMS2</i>)	HNPCC/Lynch syndrome	40%–80% by age 75	Dominant	[15–26,32,33]
<i>SMAD4/BMPR1A</i>	JPS	17%–68% by age 60	Dominant	[26,27]
<i>STK11</i>	PJS	39% by age 70	Dominant	[28]
<i>MUTYH</i>	MYH-associated polyposis	35%–53%	Recessive	[34,35]
<i>POLD1/POLE</i>	Oligopolyposis		Dominant	[36]

4. More Recent Models of Genetic Susceptibility

Studies examining the difference in CRC development between monozygotic and dizygotic twins estimated that ~35% of CRC could be ascribed to a genetic predisposition [37]. However, <10% of all CRC can be accounted for by germline mutations in *APC* and the MMR genes and crucially ~70% of the familial risk of CRC remains unexplained [38].

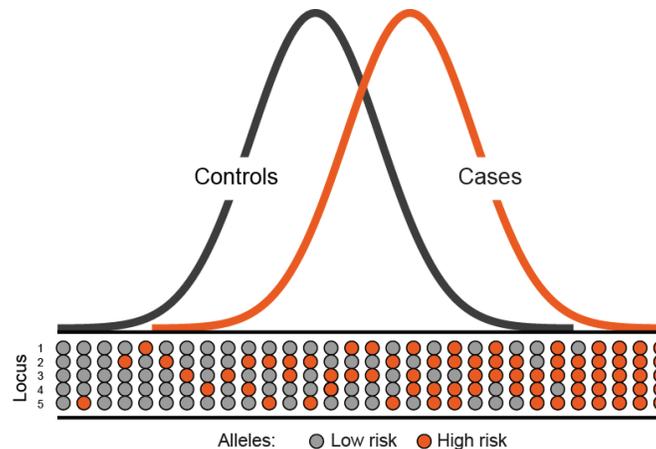
Over the past 20 years, extensive efforts to identify additional, highly penetrant cancer susceptibility genes for CRC through conventional linkage scans have met with limited success [39,40]. This strongly implies that any additional high penetrance CRC gene will individually account for only a small proportion of the familial risk. Statistical modelling of the pattern of familial occurrence of CRC after exclusion of known high-risk genes suggests that much of the inherited susceptibility is likely to be polygenic with the co-inheritance of multiple genetic variants, each with a modest individual effect, causing a wide range of risk in the population (Figure 1).

Over the past two decades candidate gene studies have identified rare moderately-penetrant risk alleles (minor allele frequency (MAF) < 2%; relative risks (RRs) > 2.0) and more recent genome-wide association studies (GWAS) have identified common, low-penetrance alleles (MAF > 10%; OR < 1.5). In reality, these variants are likely to occur as a continuum and the separate classes of risk alleles merely reflect the subgroups detectable using current methodologies.

5. Rare, Moderately-Penetrant Disease-Causing Variants

The “rare-variant” hypothesis suggests that much of the remaining heritability could be due to the combined effect of rare, moderately-penetrant risk alleles [41]. These variants are hypothesised to act independently and to confer modest, but detectable, increases in risk.

Figure 1. Polygenic model of disease susceptibility. The distribution of risk alleles in both cases and controls follows a normal distribution. However, cases have a shift towards a higher number of high risk alleles.



Attempts to identify this class of disease allele have mainly been through resequencing candidate genes in affected families, the success of which has been hampered by our limited knowledge of tumour biology. The identification of the missense variant, APC I1307K, carried by ~6% of Ashkenazi Jews and conferring around a two-fold increase in risk of CRC [42] and the more recent discovery of the functional promoter variant -93G>A of MLH1, shown to predispose to microsatellite unstable CRC [43], represent rare successes of this approach.

A priori rare disease-causing alleles are likely to act in a dominant fashion; however, functional variants in the base-excision repair gene *MUTYH* provide an example of a recessive model of inheritance [34]. Biallelic or compound heterozygosity of the G396D and Y179C mutations in *MUTYH*, which are carried by around 1%–2% of the UK population, confers a CRC risk comparable to that seen in carriers of germline MMR mutations [35].

Mechanistically these variants are likely to be directly causal. For the variants in *MUTYH*, insights into biological basis of susceptibility came from the method in which they were discovered; a FAP family with no apparent *APC* mutation was found to possess a mutator phenotype reflective of defective base excision, resulting in somatic mutation of *APC* and other genes [34,44]. In contrast, the *APC* I1307K T>A variant appears to increase replication errors in *APC* through generation of a run of eight adenines [42].

6. Identification of Common Low-Penetrance Alleles

The “common-variant, common-disease” hypothesis states that a substantial proportion of the remaining risk is likely to be accounted for by the summation of numerous low-penetrant genetic variants, each with a relatively high frequency in the population [45]. These variants have more subtle effects on gene regulation and predominantly reside within non-coding regions of the genome. Each individual variant is associated with only a modest increase in risk; however, collectively they may confer a substantial increase in disease susceptibility. These alleles rarely cause multiple cases in families and therefore cannot be detected through genetic linkage studies [46]. Initial attempts to

identify this class of allele through candidate gene association studies were based on small case-control series and had little success; any proposed variants were not successfully validated in subsequent studies.

Genome-wide association studies (GWAS), typically based on genotyping of 300,000 to over 1 million SNPs, have proved to be a powerful approach in identifying common, low penetrance susceptibility loci for CRC without prior knowledge of location and function.

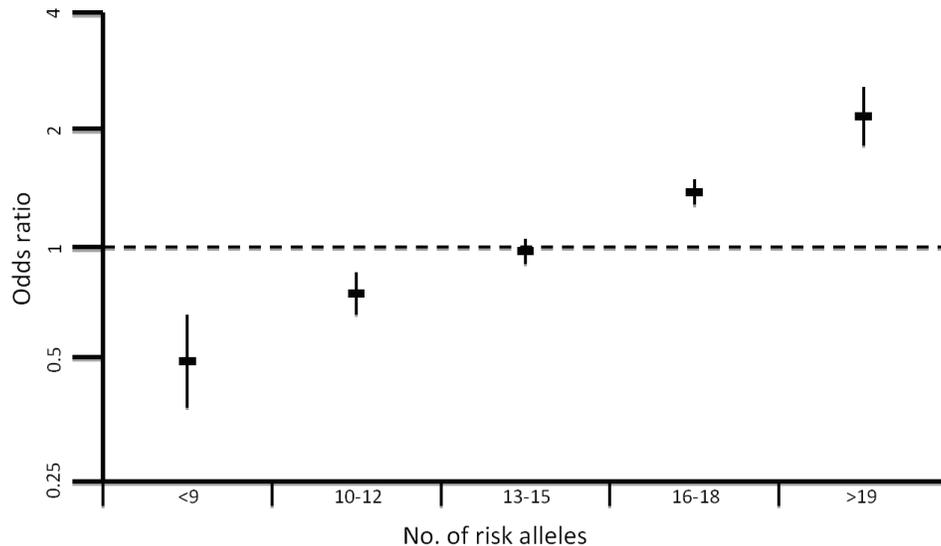
Since the first CRC GWAS in 2007, 18 CRC susceptibility loci have been identified in European populations (Table 2) [47–53]. While each individual risk allele confers only a small relative risk ($1.06 < OR < 1.26$), the SNPs are common ($MAF > 10\%$) and hence contribute significantly to the overall incidence of CRC. Moreover, by acting in concert they can impact significantly on an individual's risk of developing CRC (Figure 2) [54]. The design of association studies is advantageous as large numbers of unrelated case and control samples may be readily obtained, providing adequate power to detect loci with relatively small effects. This is in contrast to the difficulty in recruiting the extensive pedigrees required for linkage studies.

Importantly, few genes implicated in GWAS were previously evaluated in candidate gene studies, highlighting the importance of such an agnostic approach for gene discovery and understanding of CRC aetiology. None of the currently identified loci, for example, are involved in DNA repair, which is the principle pathway underscoring high-penetrance CRC susceptibility. Interestingly, five of the loci discovered to date are within or near to genes involved in the TGF- β signalling pathway [47,50]. This pathway has already been implicated in pathogenesis of CRC, as the dominant CRC predisposition syndrome JPS is caused by high penetrance mutations in the TGF- β family genes *SMAD4* and *BMPRIA* [13,14].

Table 2. Loci identified as associated with colorectal cancer through genome-wide association studies and meta-analyses.

Locus	Nearest Gene(s)	GWAS tagSNP	Location	Risk Allele	Alt Allele	RAF
1q41	<i>DUSP10</i>	rs6691170	222,045,446	T	G	0.40
3q26.2	<i>TERC, MYNN</i>	rs10936599	169,492,101	C	T	0.75
6p21.2	<i>CDKN1A</i>	rs1321311	36,622,900	T	G	0.21
8q23.3	<i>EIF3H</i>	rs16892766	117,630,683	C	A	0.09
8q24.21	<i>MYC</i>	rs6983267	128,413,305	G	T	0.52
10p14	<i>GATA3</i>	rs10795668	8,701,219	G	A	0.67
11q13.4	<i>POLD3</i>	rs3824999	74,345,550	C	A	0.47
11q23.1	<i>FLJ45803</i>	rs3802842	111,171,709	C	A	0.27
12q13	<i>DIP2B, ATF1</i>	rs11169552	51,155,663	C	T	0.75
14q22.2	<i>BMP4</i>	rs4444235	54,410,919	C	T	0.48
15q13.3	<i>SCG5, GREM1</i>	rs4779584	32,994,756	T	C	0.19
16q22.1	<i>CDH1</i>	rs9929218	68,820,946	G	A	0.71
18q21.2	<i>SMAD7</i>	rs4939827	46,453,463	T	C	0.53
19q13.11	<i>RHPN2, GPATCH1</i>	rs10411210	33,532,300	C	T	0.90
20p12.3	<i>BMP2</i>	rs961253	6,404,281	A	C	0.37
		rs4813802	6,699,595	G	T	0.34
20q13.33	<i>LAMA5</i>	rs4925386	60,921,044	C	T	0.68
Xp22.2	<i>SHROOM2</i>	rs5934683	9,751,474	T	C	0.56

Figure 2. Plot showing the increase in odds ratio for colorectal cancer with an increasing number of risk alleles.



7. Functional Effects of GWAS Loci

Elucidating the basis of association at common low-penetrance loci represents a significant challenge. The tagging SNPs (tagSNPs) used in GWAS are not necessarily strong candidates for being causal and were instead chosen to capture variation across large genomic regions. Hence, establishing which of a set of highly correlated variants in linkage disequilibrium (LD) with the tagSNP is the true causal variant is a challenge. In addition, many GWAS loci map to non-coding regions or gene deserts, suggesting the true cause of the association at these regions is through subtle effects on gene expression rather than changes to protein coding sequence.

Fine mapping of CRC loci is still in its infancy with most studies attempting only to narrow down the location of a likely functional variant through imputation or re-sequencing [55–57]. These studies suggest candidate variants but very few functional studies have been carried out to test these assertions. To date, in only four regions has a SNP been proposed as the likely functional candidate and a mechanism of action suggested (8q23.3 [58], 8q24.21 [59,60], 11q23.1 [61], 18q21.1 [62]). The most intriguing of these regions is 8q24.21, which has pleiotropic effects on cancer susceptibility, also harbouring risk loci for breast [57], ovarian [63], bladder [64], CLL [65] and multiple independent loci for prostate cancer [66–69]. This is in contrast to most associations found to date which appear to be disease specific. The rs6983267 risk SNP is associated with both prostate and colorectal cancers and lies within an evolutionarily conserved region. The two alleles of rs6983267 show differential binding of the TCF4 transcription factor [59] to an enhancer element that has been shown to physically interact with the *MYC* gene promoter [60]. *MYC* is amplified or over-expressed in multiple cancer types leading to up-regulation of many genes controlling cell proliferation, so it is predicted that variation at this locus acts through subtle effects on *MYC* gene expression. The risk allele of rs6983267 has also been suggested as a marker of worse prognosis in CRC patients [70]. The strongest CRC GWAS association is at 18q21.1 (RR = 1.26) within the *SMAD7* gene, which acts as an antagonist of the TGF- β signalling pathway. Resequencing of the 17 kb region of linkage disequilibrium surrounding

the GWAS tagSNP rs4939827 identified a novel variant, termed Novel1 (rs58920878), which was shown to affect *SMAD7* expression [62].

8. Impact of Common Variation on CRC Risk

Collectively, the currently identified risk loci explain only ~6% of the overall familial risk of CRC. This estimate is likely conservative as the effect of the causal variant at each locus is expected to be greater than the association detected through a GWAS tagSNP. In addition, as evinced by the 14q22 (*BMP4*) [52] association multiple risk variants may exist at each locus, including low-frequency variants with significantly larger effects. Moreover, epistatic interactions may exist between common risk loci, which could make the contribution of low-penetrance susceptibility alleles much higher. Such interactions remain difficult to detect due to substantial multiple testing penalties, and existing studies suggest the effects of most common low-penetrance alleles seem to be independent. In addition, interactions of these alleles with epigenetic regulation or environmental factors may lead to a greater increase in disease risk. For example the *MLH1*-93G>A polymorphism has been shown to be associated with increased methylation of the *MLH1* promoter [71]. Another important consideration is the possible modification of the effect of a low-penetrance allele by the presence of a high-penetrance mutation. The only evidence to support this assertion in CRC is a small study implicating the 8q23.3 and 11q23.1 CRC SNPs as modifiers of CRC risk in MMR mutation carriers [72]. Although there was initial hope for the use of low penetrance variants in the clinic, the small proportion of familial risk explained and the apparent lack of epistatic interactions between the variants leads to them being of low predictive power. The increased risk associated with having a high number of risk alleles (Figure 2) has the prospect of identifying individuals in the general population who might benefit from earlier screening [54].

9. Identifying Novel CRC Susceptibility Loci

It is unlikely that there are any common CRC risk loci with appreciable MAF (>10%) and with relative risks >1.1 that remain to be uncovered given the size of existing GWAS studies. Small variant effect sizes combined with stringent thresholds for establishing statistical significance and financial constraints on the number of variants which can be followed up constrain GWAS study power. However, many GWAS have long tails of association with alleles of increasingly small effect, suggesting much of the remaining susceptibility may be embodied in a multitude of common risk alleles. Larger GWAS studies combining multiple phases and tens of thousands of cases may identify many more of these susceptibility loci, although the effect of these on CRC risk is likely to be minimal. Such studies have been conducted in both breast [73] and prostate [74] cancer identifying 41 and 23 novel risk loci respectively.

Commercial arrays used for GWAS capture a large proportion of common SNPs with minor allele frequencies (MAF) > 10%. However, a much lower fraction of less common (5%–10% MAF) and rare (MAF < 5%) SNPs are captured by these arrays. The power of GWAS to detect variants with MAF < 10% is therefore limited. Additionally, GWAS arrays are not optimally formatted to capture indels and copy number variants, both of which are likely to have roles in disease susceptibility. New

“exome chips” have recently been released that aim to address some of these limitations. However the success of these new arrays remains to be evaluated.

Since the completion of the human genome project in 2003, the utilisation of massively parallel sequencing technologies to identify variants has become feasible. Moreover, these methods can be used to detect small indels, substitutions and structural variants. Although in their infancy, such studies are beginning to identify additional variants. For example, highly-penetrant mutations have recently been identified in the proof reading domains of *POLD1* (S478N) and *POLE* (L424V) in CRC families [36]. Another recent study identified 11 candidate CRC susceptibility genes with truncating mutations in two or more of 96 familial CRC cases [75]. To maximally utilise sequencing data, new bioinformatic techniques are required to remove sequencing errors and prioritise variants. Additionally, due to financial considerations such studies remain small and not powered to detect less common variants with moderate effects risk of disease. Using cases enriched for genetic susceptibility evidenced by a strong family history or early disease onset is a useful technique to increase the efficiency of these studies. In addition, utilising whole exome sequencing, as performed by Palles *et al.* [36], can dramatically reduce the costs associated with such projects. Coding variants are also much easier to interpret than those in non-coding regions. However, with every individual’s genome harboring 250–300 putative loss of function variants [76] and many missense variants being of unknown effect, identification of the disease causing variant presents a significant challenge. Recent studies are working to interpret this class of variants [77] and algorithms such as SIFT [78], PolyPhen2 [79] and CONDEL [80] aim to guide researchers by predicting the functional effects of coding mutations. Work to develop similar methods to deal with non-coding regions is still in its infancy [81,82], however, recent evidence [83,84] suggests that these regions are also *a priori* likely to contain variants with a large effect on CRC risk.

10. Conclusions

Much has been achieved in the study of genetic susceptibility to CRC in the last five decades. The architecture underlying this susceptibility is now recognised to be defined by a spectrum of predisposition alleles with different effect sizes and frequencies in the population. GWAS has proved a successful approach for identification of novel low-penetrance CRC risk alleles, improving our understanding of disease aetiology and providing novel therapeutic targets. Determining the biological processes underlying the associations at these loci presents a significant challenge and will likely require large collaborations between genetic researchers and functional biologists. Despite these advances, a large proportion of the heritability to CRC remains unaccounted for. It is hoped that the new generation of sequencing projects will help to uncover this missing heritability.

Acknowledgements

Principle funding was provided to RSH by Cancer Research UK (C1298/A8362-Bobby Moore Fund for Cancer Research UK). NW is in receipt of a PhD studentship from the Institute of Cancer Research.

Author Contributions

Manuscript was conceived and written by NW and RSH.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Jemal, A.; Siegel, R.; Ward, E.; Murray, T.; Xu, J.; Smigal, C.; Thun, M.J. Cancer statistics, 2006. *CA Cancer J. Clin.* **2006**, *56*, 106–130.
2. Johns, L.E.; Houlston, R.S. A systematic review and meta-analysis of familial colorectal cancer risk. *Am. J. Gastroenterol.* **2001**, *96*, 2992–3003.
3. Woolf, C.M. A genetic study of carcinoma of the large intestine. *Am. J. Hum. Genet.* **1958**, *10*, 42–47.
4. Macklin, M.T. Inheritance of cancer of the stomach and large intestine in man. *J. Natl. Canc. Inst.* **1960**, *24*, 551–571.
5. Lovett, E. Familial factors in the etiology of carcinoma of the large bowel. *Proc. R. Soc. Med.* **1974**, *67*, 751–752.
6. Ashley, D.J. Colonic cancer arising in polyposis coli. *J. Med. Genet.* **1969**, *6*, 376–378.
7. DeMars, R. *23rd Annual Symposium on Fundamental Cancer Research*; Williams and Wilkins: Baltimore, MD, USA, 1969; pp. 105–106.
8. Anderson, D.E. Genetic study of breast cancer: Identification of a high risk group. *Cancer* **1974**, *34*, 1090–1097.
9. Lynch, H.T.; Krush, A.J. Cancer family “G” Revisited: 1895–1970. *Cancer* **1971**, *27*, 1505–1511.
10. Bodmer, W.F.; Bailey, C.J.; Bodmer, J.; Bussey, H.J.; Ellis, A.; Gorman, P.; Lucibello, F.C.; Murday, V.A.; Rider, S.H.; Scambler, P.; *et al.* Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature* **1987**, *328*, 614–616.
11. Leppert, M.; Dobbs, M.; Scambler, P.; O’Connell, P.; Nakamura, Y.; Stauffer, D.; Woodward, S.; Burt, R.; Hughes, J.; Gardner, E.; *et al.* The gene for familial polyposis coli maps to the long arm of chromosome 5. *Science* **1987**, *238*, 1411–1413.
12. Herrera, L.; Kakati, S.; Gibas, L.; Pietrzak, E.; Sandberg, A.A. Gardner syndrome in a man with an interstitial deletion of 5q. *Am. J. Med. Genet.* **1986**, *25*, 473–476.
13. Groden, J.; Thliveris, A.; Samowitz, W.; Carlson, M.; Gelbert, L.; Albertsen, H.; Joslyn, G.; Stevens, J.; Spirio, L.; Robertson, M.; *et al.* Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* **1991**, *66*, 589–600.
14. Kinzler, K.W.; Nilbert, M.C.; Su, L.K.; Vogelstein, B.; Bryan, T.M.; Levy, D.B.; Smith, K.J.; Preisinger, A.C.; Hedge, P.; McKechnie, D.; *et al.* Identification of fap locus genes from chromosome 5q21. *Science* **1991**, *253*, 661–665.
15. Miyaki, M.; Konishi, M.; Tanaka, K.; Kikuchi-Yanoshita, R.; Muraoka, M.; Yasuno, M.; Igari, T.; Koike, M.; Chiba, M.; Mori, T. Germline mutation of msh6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat. Genet.* **1997**, *17*, 271–272.

16. Papadopoulos, N.; Nicolaides, N.C.; Wei, Y.F.; Ruben, S.M.; Carter, K.C.; Rosen, C.A.; Haseltine, W.A.; Fleischmann, R.D.; Fraser, C.M.; Adams, M.D.; *et al.* Mutation of a mutl homolog in hereditary colon cancer. *Science* **1994**, *263*, 1625–1629.
17. Bronner, C.E.; Baker, S.M.; Morrison, P.T.; Warren, G.; Smith, L.G.; Lescoe, M.K.; Kane, M.; Earabino, C.; Lipford, J.; Lindblom, A.; *et al.* Mutation in the DNA mismatch repair gene homologue hmlh1 is associated with hereditary non-polyposis colon cancer. *Nature* **1994**, *368*, 258–261.
18. Strand, M.; Prolla, T.A.; Liskay, R.M.; Petes, T.D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **1993**, *365*, 274–276.
19. Peltomaki, P.; Aaltonen, L.A.; Sistonen, P.; Pylkkanen, L.; Mecklin, J.P.; Jarvinen, H.; Green, J.S.; Jass, J.R.; Weber, J.L.; Leach, F.S.; *et al.* Genetic mapping of a locus predisposing to human colorectal cancer. *Science* **1993**, *260*, 810–812.
20. Aaltonen, L.A.; Peltomaki, P.; Leach, F.S.; Sistonen, P.; Pylkkanen, L.; Mecklin, J.P.; Jarvinen, H.; Powell, S.M.; Jen, J.; Hamilton, S.R.; *et al.* Clues to the pathogenesis of familial colorectal cancer. *Science* **1993**, *260*, 812–816.
21. Lindblom, A.; Tannergard, P.; Werelius, B.; Nordenskjold, M. Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nat. Genet.* **1993**, *5*, 279–282.
22. Leach, F.S.; Nicolaides, N.C.; Papadopoulos, N.; Liu, B.; Jen, J.; Parsons, R.; Peltomaki, P.; Sistonen, P.; Aaltonen, L.A.; Nystrom-Lahti, M.; *et al.* Mutations of a muts homolog in hereditary nonpolyposis colorectal cancer. *Cell* **1993**, *75*, 1215–1225.
23. Drummond, J.T.; Li, G.M.; Longley, M.J.; Modrich, P. Isolation of an hms2-p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science* **1995**, *268*, 1909–1912.
24. Palombo, F.; Gallinari, P.; Iaccarino, I.; Lettieri, T.; Hughes, M.; D'Arrigo, A.; Truong, O.; Hsuan, J.J.; Jiricny, J. Gtbp, a 160-kilodalton protein essential for mismatch-binding activity in human cells. *Science* **1995**, *268*, 1912–1914.
25. Akiyama, Y.; Sato, H.; Yamada, T.; Nagasaki, H.; Tsuchiya, A.; Abe, R.; Yuasa, Y. Germ-line mutation of the hms6/gtbp gene in an atypical hereditary nonpolyposis colorectal cancer kindred. *Canc. Res.* **1997**, *57*, 3920–3923.
26. Shibata, D.; Peinado, M.A.; Ionov, Y.; Malkhosyan, S.; Perucho, M. Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. *Nat. Genet.* **1994**, *6*, 273–281.
27. Howe, J.R.; Roth, S.; Ringold, J.C.; Summers, R.W.; Jarvinen, H.J.; Sistonen, P.; Tomlinson, I.P.; Houlston, R.S.; Bevan, S.; Mitros, F.A.; *et al.* Mutations in the smad4/dpc4 gene in juvenile polyposis. *Science* **1998**, *280*, 1086–1088.
28. Howe, J.R.; Bair, J.L.; Sayed, M.G.; Anderson, M.E.; Mitros, F.A.; Petersen, G.M.; Velculescu, V.E.; Traverso, G.; Vogelstein, B. Germline mutations of the gene encoding bone morphogenetic protein receptor 1a in juvenile polyposis. *Nat. Genet.* **2001**, *28*, 184–187.
29. Hemminki, A.; Markie, D.; Tomlinson, I.; Avizienyte, E.; Roth, S.; Loukola, A.; Bignell, G.; Warren, W.; Aminoff, M.; Hoglund, P.; *et al.* A serine/threonine kinase gene defective in peutz-jeghers syndrome. *Nature* **1998**, *391*, 184–187.

30. Jarvinen, H.J.; Aarnio, M.; Mustonen, H.; Aktan-Collan, K.; Aaltonen, L.A.; Peltomaki, P.; de La Chapelle, A.; Mecklin, J.P. Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology* **2000**, *118*, 829–834.
31. Burn, J.; Gerdes, A.M.; Macrae, F.; Mecklin, J.P.; Moeslein, G.; Olschwang, S.; Eccles, D.; Evans, D.G.; Maher, E.R.; Bertario, L.; *et al.* Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: An analysis from the capp2 randomised controlled trial. *Lancet* **2011**, *378*, 2081–2087.
32. Nicolaides, N.C.; Papadopoulos, N.; Liu, B.; Wei, Y.F.; Carter, K.C.; Ruben, S.M.; Rosen, C.A.; Haseltine, W.A.; Fleischmann, R.D.; Fraser, C.M.; *et al.* Mutations of two pms homologues in hereditary nonpolyposis colon cancer. *Nature* **1994**, *371*, 75–80.
33. Hendriks, Y.M.; Jagmohan-Changur, S.; van der Klift, H.M.; Morreau, H.; van Puijenbroek, M.; Tops, C.; van Os, T.; Wagner, A.; Ausems, M.G.; Gomez, E.; *et al.* Heterozygous mutations in pms2 cause hereditary nonpolyposis colorectal carcinoma (lynch syndrome). *Gastroenterology* **2006**, *130*, 312–322.
34. Al-Tassan, N.; Chmiel, N.H.; Maynard, J.; Fleming, N.; Livingston, A.L.; Williams, G.T.; Hodges, A.K.; Davies, D.R.; David, S.S.; Sampson, J.R.; *et al.* Inherited variants of myh associated with somatic g:C-->t:A mutations in colorectal tumors. *Nat. Genet.* **2002**, *30*, 227–232.
35. Lubbe, S.J.; di Bernardo, M.C.; Chandler, I.P.; Houlston, R.S. Clinical implications of the colorectal cancer risk associated with mutyh mutation. *J. Clin. Oncol.* **2009**, *27*, 3975–3980.
36. Palles, C.; Cazier, J.B.; Howarth, K.M.; Domingo, E.; Jones, A.M.; Broderick, P.; Kemp, Z.; Spain, S.L.; Guarino, E.; Salguero, I.; *et al.* Germline mutations affecting the proofreading domains of pole and pold1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.* **2013**, *45*, 136–144.
37. Lichtenstein, P.; Holm, N.V.; Verkasalo, P.K.; Iliadou, A.; Kaprio, J.; Koskenvuo, M.; Pukkala, E.; Skytthe, A.; Hemminki, K. Environmental and heritable factors in the causation of cancer—Analyses of cohorts of twins from sweden, denmark, and finland. *N. Engl. J. Med.* **2000**, *343*, 78–85.
38. Lubbe, S.J.; Webb, E.L.; Chandler, I.P.; Houlston, R.S. Implications of familial colorectal cancer risk profiles and microsatellite instability status. *J. Clin. Oncol.* **2009**, *27*, 2238–2244.
39. Papaemmanuil, E.; Carvajal-Carmona, L.; Sellick, G.S.; Kemp, Z.; Webb, E.; Spain, S.; Sullivan, K.; Barclay, E.; Lubbe, S.; Jaeger, E.; *et al.* Deciphering the genetics of hereditary non-syndromic colorectal cancer. *Eur. J. Hum. Genet.* **2008**, *16*, 1477–1486.
40. Wiesner, G.L.; Daley, D.; Lewis, S.; Ticknor, C.; Platzer, P.; Lutterbaugh, J.; MacMillen, M.; Baliner, B.; Willis, J.; Elston, R.C.; *et al.* A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2–31.2. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12961–12965.
41. Bodmer, W.; Bonilla, C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.* **2008**, *40*, 695–701.
42. Laken, S.J.; Petersen, G.M.; Gruber, S.B.; Oddoux, C.; Ostrer, H.; Giardiello, F.M.; Hamilton, S.R.; Hampel, H.; Markowitz, A.; Klimstra, D.; *et al.* Familial colorectal cancer in ashkenazim due to a hypermutable tract in apc. *Nat. Genet.* **1997**, *17*, 79–83.
43. Whiffin, N.; Broderick, P.; Lubbe, S.J.; Pittman, A.M.; Penegar, S.; Chandler, I.; Houlston, R.S. Mlh1-93G>A is a risk factor for msi colorectal cancer. *Carcinogenesis* **2011**, *32*, 1157–1161.

44. Lipton, L.; Halford, S.E.; Johnson, V.; Novelli, M.R.; Jones, A.; Cummings, C.; Barclay, E.; Sieber, O.; Sadat, A.; Bisgaard, M.L.; *et al.* Carcinogenesis in myh-associated polyposis follows a distinct genetic pathway. *Canc. Res.* **2003**, *63*, 7595–7599.
45. Reich, D.E.; Lander, E.S. On the allelic spectrum of human disease. *Trends Genet.* **2001**, *17*, 502–510.
46. Risch, N.; Merikangas, K. The future of genetic studies of complex human diseases. *Science* **1996**, *273*, 1516–1517.
47. Broderick, P.; Carvajal-Carmona, L.; Pittman, A.M.; Webb, E.; Howarth, K.; Rowan, A.; Lubbe, S.; Spain, S.; Sullivan, K.; Fielding, S.; *et al.* A genome-wide association study shows that common alleles of smad7 influence colorectal cancer risk. *Nat. Genet.* **2007**, *39*, 1315–1317.
48. Dunlop, M.G.; Dobbins, S.E.; Farrington, S.M.; Jones, A.M.; Palles, C.; Whiffin, N.; Tenesa, A.; Spain, S.; Broderick, P.; Ooi, L.Y.; *et al.* Common variation near cdkn1a, pold3 and shroom2 influences colorectal cancer risk. *Nat. Genet.* **2012**, *44*, 770–776.
49. Houlston, R.S.; Cheadle, J.; Dobbins, S.E.; Tenesa, A.; Jones, A.M.; Howarth, K.; Spain, S.L.; Broderick, P.; Domingo, E.; Farrington, S.; *et al.* Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat. Genet.* **2010**, *42*, 973–977.
50. Houlston, R.S.; Webb, E.; Broderick, P.; Pittman, A.M.; di Bernardo, M.C.; Lubbe, S.; Chandler, I.; Vijayakrishnan, J.; Sullivan, K.; Penegar, S.; *et al.* Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat. Genet.* **2008**, *40*, 1426–1435.
51. Tenesa, A.; Farrington, S.M.; Prendergast, J.G.; Porteous, M.E.; Walker, M.; Haq, N.; Barnetson, R.A.; Theodoratou, E.; Cetnarskyj, R.; Cartwright, N.; *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*, 631–637.
52. Tomlinson, I.P.; Carvajal-Carmona, L.G.; Dobbins, S.E.; Tenesa, A.; Jones, A.M.; Howarth, K.; Palles, C.; Broderick, P.; Jaeger, E.E.; Farrington, S.; *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*, e1002105.
53. Tomlinson, I.P.; Webb, E.; Carvajal-Carmona, L.; Broderick, P.; Howarth, K.; Pittman, A.M.; Spain, S.; Lubbe, S.; Walther, A.; Sullivan, K.; *et al.* A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.* **2008**, *40*, 623–630.
54. Lubbe, S.J.; di Bernardo, M.C.; Broderick, P.; Chandler, I.; Houlston, R.S. Comprehensive evaluation of the impact of 14 genetic variants on colorectal cancer phenotype and risk. *Am. J. Epidemiol.* **2012**, *175*, 1–10.
55. Whiffin, N.; Dobbins, S.E.; Hosking, F.J.; Palles, C.; Tenesa, A.; Wang, Y.; Farrington, S.M.; Jones, A.M.; Broderick, P.; Campbell, H.; *et al.* Deciphering the genetic architecture of low-penetrance susceptibility to colorectal cancer. *Hum. Mol. Genet.* **2013**, *22*, 5075–5082.

56. Carvajal-Carmona, L.G.; Cazier, J.B.; Jones, A.M.; Howarth, K.; Broderick, P.; Pittman, A.; Dobbins, S.; Tenesa, A.; Farrington, S.; Prendergast, J.; *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**, *20*, 2879–2888.
57. Easton, D.F.; Pooley, K.A.; Dunning, A.M.; Pharoah, P.D.; Thompson, D.; Ballinger, D.G.; Struwing, J.P.; Morrison, J.; Field, H.; Luben, R.; *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **2007**, *447*, 1087–1093.
58. Pittman, A.M.; Naranjo, S.; Jalava, S.E.; Twiss, P.; Ma, Y.; Olver, B.; Lloyd, A.; Vijaykrishnan, J.; Qureshi, M.; Broderick, P.; *et al.* Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a cis-acting regulator of eif3h. *PLoS Genet.* **2010**, *6*, e1001126.
59. Tuupainen, S.; Turunen, M.; Lehtonen, R.; Hallikas, O.; Vanharanta, S.; Kivioja, T.; Bjorklund, M.; Wei, G.; Yan, J.; Niittymaki, I.; *et al.* The common colorectal cancer predisposition snp rs6983267 at chromosome 8q24 confers potential to enhanced wnt signaling. *Nat. Genet.* **2009**, *41*, 885–890.
60. Pomerantz, M.M.; Ahmadiyeh, N.; Jia, L.; Herman, P.; Verzi, M.P.; Doddapaneni, H.; Beckwith, C.A.; Chan, J.A.; Hills, A.; Davis, M.; *et al.* The 8q24 cancer risk variant rs6983267 shows long-range interaction with myc in colorectal cancer. *Nat. Genet.* **2009**, *41*, 882–884.
61. Pittman, A.M.; Webb, E.; Carvajal-Carmona, L.; Howarth, K.; di Bernardo, M.C.; Broderick, P.; Spain, S.; Walther, A.; Price, A.; Sullivan, K.; *et al.* Refinement of the basis and impact of common 11q23.1 variation to the risk of developing colorectal cancer. *Hum. Mol. Genet.* **2008**, *17*, 3720–3727.
62. Pittman, A.M.; Naranjo, S.; Webb, E.; Broderick, P.; Lips, E.H.; van Wezel, T.; Morreau, H.; Sullivan, K.; Fielding, S.; Twiss, P.; *et al.* The colorectal cancer risk at 18q21 is caused by a novel variant altering smad7 expression. *Genome Res.* **2009**, *19*, 987–993.
63. Ghousaini, M.; Song, H.; Koessler, T.; Al Olama, A.A.; Kote-Jarai, Z.; Driver, K.E.; Pooley, K.A.; Ramus, S.J.; Kjaer, S.K.; Hogdall, E.; *et al.* Multiple loci with different cancer specificities within the 8q24 gene desert. *J. Natl. Canc. Inst.* **2008**, *100*, 962–966.
64. Kiemeny, L.A.; Thorlacius, S.; Sulem, P.; Geller, F.; Aben, K.K.; Stacey, S.N.; Gudmundsson, J.; Jakobsdottir, M.; Bergthorsson, J.T.; Sigurdsson, A.; *et al.* Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat. Genet.* **2008**, *40*, 1307–1312.
65. Crowther-Swanepoel, D.; Broderick, P.; di Bernardo, M.C.; Dobbins, S.E.; Torres, M.; Mansouri, M.; Ruiz-Ponte, C.; Enjuanes, A.; Rosenquist, R.; Carracedo, A.; *et al.* Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. *Nat. Genet.* **2010**, *42*, 132–136.
66. Al Olama, A.A.; Kote-Jarai, Z.; Giles, G.G.; Guy, M.; Morrison, J.; Severi, G.; Leongamornlert, D.A.; Tymrakiewicz, M.; Jhavar, S.; Saunders, E.; *et al.* Multiple loci on 8q24 associated with prostate cancer susceptibility. *Nat. Genet.* **2009**, *41*, 1058–1060.
67. Amundadottir, L.T.; Sulem, P.; Gudmundsson, J.; Helgason, A.; Baker, A.; Agnarsson, B.A.; Sigurdsson, A.; Benediksdottir, K.R.; Cazier, J.B.; Sainz, J.; *et al.* A common variant associated with prostate cancer in european and african populations. *Nat. Genet.* **2006**, *38*, 652–658.

68. Yeager, M.; Orr, N.; Hayes, R.B.; Jacobs, K.B.; Kraft, P.; Wacholder, S.; Minichiello, M.J.; Fearnhead, P.; Yu, K.; Chatterjee, N.; *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* **2007**, *39*, 645–649.
69. Gudmundsson, J.; Sulem, P.; Manolescu, A.; Amundadottir, L.T.; Gudbjartsson, D.; Helgason, A.; Rafnar, T.; Bergthorsson, J.T.; Agnarsson, B.A.; Baker, A.; *et al.* Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat. Genet.* **2007**, *39*, 631–637.
70. Takatsuno, Y.; Mimori, K.; Yamamoto, K.; Sato, T.; Niida, A.; Inoue, H.; Imoto, S.; Kawano, S.; Yamaguchi, R.; Toh, H.; *et al.* The rs6983267 snp is associated with myc transcription efficiency, which promotes progression and worsens prognosis of colorectal cancer. *Ann. Surg. Oncol.* **2013**, *20*, 1395–1402.
71. Mrkonjic, M.; Roslin, N.M.; Greenwood, C.M.; Raptis, S.; Pollett, A.; Laird, P.W.; Pethe, V.V.; Chiang, T.; Daftary, D.; Dicks, E.; *et al.* Specific variants in the mlh1 gene region may drive DNA methylation, loss of protein expression, and msi-h colorectal cancer. *PLoS One* **2010**, *5*, e13314.
72. Wijnen, J.T.; Brohet, R.M.; van Eijk, R.; Jagmohan-Changur, S.; Middeldorp, A.; Tops, C.M.; van Puijenbroek, M.; Ausems, M.G.; Gomez Garcia, E.; Hes, F.J.; *et al.* Chromosome 8q23.3 and 11q23.1 variants modify colorectal cancer risk in lynch syndrome. *Gastroenterology* **2009**, *136*, 131–137.
73. Michailidou, K.; Hall, P.; Gonzalez-Neira, A.; Ghoussaini, M.; Dennis, J.; Milne, R.L.; Schmidt, M.K.; Chang-Claude, J.; Bojesen, S.E.; Bolla, M.K.; *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **2013**, *45*, 353–361, 361e1–361e2.
74. Eeles, R.A.; Olama, A.A.; Benlloch, S.; Saunders, E.J.; Leongamornlert, D.A.; Tymrakiewicz, M.; Ghoussaini, M.; Luccarini, C.; Dennis, J.; Jugurnauth-Little, S.; *et al.* Identification of 23 new prostate cancer susceptibility loci using the icogs custom genotyping array. *Nat. Genet.* **2013**, *45*, 385–391, 391e1–391e2.
75. Gylfe, A.E.; Katainen, R.; Kondelin, J.; Tanskanen, T.; Cajuso, T.; Hanninen, U.; Taipale, J.; Taipale, M.; Renkonen-Sinisalo, L.; Jarvinen, H.; *et al.* Eleven candidate susceptibility genes for common familial colorectal cancer. *PLoS Genet.* **2013**, *9*, e1003876.
76. Abecasis, G.R.; Altshuler, D.; Auton, A.; Brooks, L.D.; Durbin, R.M.; Gibbs, R.A.; Hurles, M.E.; McVean, G.A. A map of human genome variation from population-scale sequencing. *Nature* **2010**, *467*, 1061–1073.
77. Thompson, B.A.; Spurdle, A.B.; Plazzer, J.-P.; Greenblatt, M.S.; Akagi, K.; Al-Mulla, F.; Bapat, B.; Bernstein, I.; Capellá, G.; den Dunnen, J.T.; *et al.* Application of a 5-tiered scheme for standardized classification of 2360 unique mismatch repair gene variants in the insight locus-specific database. *Nat. Genet.* **2014**, *46*, 107–115.
78. Kumar, P.; Henikoff, S.; Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the sift algorithm. *Nat. Protoc.* **2009**, *4*, 1073–1081.
79. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A method and server for predicting damaging missense mutations. *Nat. Methods* **2010**, *7*, 248–249.
80. Gonzalez-Perez, A.; Lopez-Bigas, N. Improving the assessment of the outcome of nonsynonymous snvs with a consensus deleteriousness score, condel. *Am. J. Hum. Genet.* **2011**, *88*, 440–449.

81. Khurana, E.; Fu, Y.; Colonna, V.; Mu, X.J.; Kang, H.M.; Lappalainen, T.; Sboner, A.; Lochovsky, L.; Chen, J.; Harmanci, A.; *et al.* Integrative annotation of variants from 1092 humans: Application to cancer genomics. *Science* **2013**, *342*, 1235587.
82. Kircher, M.; Witten, D.M.; Jain, P.; O’Roak, B.J.; Cooper, G.M.; Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **2014**, *46*, 310–315.
83. Weedon, M.N.; Cebola, I.; Patch, A.M.; Flanagan, S.E.; de Franco, E.; Caswell, R.; Rodriguez-Segui, S.A.; Shaw-Smith, C.; Cho, C.H.; Lango Allen, H.; *et al.* Recessive mutations in a distal *ptf1a* enhancer cause isolated pancreatic agenesis. *Nat. Genet.* **2014**, *46*, 61–64.
84. Horn, S.; Figl, A.; Rachakonda, P.S.; Fischer, C.; Sucker, A.; Gast, A.; Kadel, S.; Moll, I.; Nagore, E.; Hemminki, K.; *et al.* Tert promoter mutations in familial and sporadic melanoma. *Science* **2013**, *339*, 959–961.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).