Review

Candidate Gene Discovery in Hereditary Colorectal Cancer and Polyposis Syndromes–Considerations for Future Studies

Iris B. A. W. te Paske [1], **Marjolijn J. L. Ligtenberg** [1,2], **Nicoline Hoogerbrugge** [1] and **Richarda M. de Voer** [1,*]

1   Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands; iris.tepaske@radboudumc.nl (I.B.A.W.t.P.); marjolijn.ligtenberg@radboudumc.nl (M.J.L.L.); nicoline.hoogerbrugge@radboudumc.nl (N.H.)

2   Department of Pathology, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands

*   Correspondence: richarda.devoer@radboudumc.nl; Tel.: +31-24-36-14107

check for updates

Abstract: To discover novel high-penetrant risk loci for hereditary colorectal cancer (hCRC) and polyposis syndromes many whole-exome and whole-genome sequencing (WES/WGS) studies have been performed. Remarkably, these studies resulted in only a few novel high-penetrant risk genes. Given this observation, the possibility and strategy to identify high-penetrant risk genes for hCRC and polyposis needs reconsideration. Therefore, we reviewed the study design of WES/WGS-based hCRC and polyposis gene discovery studies (*n* = 37) and provide recommendations to optimize discovery and validation strategies. The group of genetically unresolved patients is phenotypically heterogeneous, and likely composed of distinct molecular subtypes. This knowledge advocates for the screening of a homogeneous, stringently preselected discovery cohort and obtaining multi-level evidence for variant pathogenicity. This evidence can be collected by characterizing the molecular landscape of tumors from individuals with the same affected gene or by functional validation in cell-based models. Together, the combined approach of a phenotype-driven, tumor-based candidate gene search might elucidate the potential contribution of novel genetic predispositions in genetically unresolved hCRC and polyposis.

Keywords: colorectal tumors; genetic predisposition; missing heritability; molecular diagnosis; molecular biomarkers; rare variants

1. Introduction

Colorectal cancer (CRC) is the third most commonly occurring cancer in men and the second most commonly observed cancer in women in Western society [1]. While the majority of individuals with CRC are not genetically predisposed, a predisposition to CRC may be considered when an individual is diagnosed with CRC at a young age (<50 years), when close relatives of the proband are affected with (colorectal) cancer, or when an individual has developed multiple adenomas in the colorectum. The presence of multiple adenomas (cumulative number of >10) at a young age (<60 years) is clinically diagnosed as polyposis.

The exact contribution of heritable factors to CRC and polyposis is still not fully understood. In the early 2000s Nordic twin- and family studies indicated that 12–35% of CRCs are potentially linked to heritable factors [2,3]. Later, estimates for heritability of CRC decreased to about 15% of all CRC cases [4,5]. At present, it is estimated that about 5–10% of all CRC and polyposis cases are explained by rare pathogenic variants in high-penetrant risk genes [2,6–8]. Next to identification of rare

high-penetrant risk genes contributing to the heredity of CRC, it is estimated that common variants (minor allele frequency (MAF) in the general population >1%) may explain about ~12% of the relative risk for CRC [8–11].

The sum of these estimations covers largely the reported heritability in twin studies, but a substantial group of patients that fulfill the clinical criteria for hereditary CRC (hCRC) and/or polyposis are considered to remain without a genetic diagnosis [12–15]. It is important to resolve this missing heritability as a genetic diagnosis may favor a patient's disease prognosis, surveillance and counseling, but may also have large implications for family members in disease prevention. It was anticipated that the rapid developments of next-generation sequencing (NGS) would aid in reducing the missing heritability for CRC and polyposis. However, despite numerous NGS studies, little additional high-penetrant novel risk genes for hCRC and polyposis have been found in the past decade. Over the years, various hypotheses for the missing heritability in hCRC and polyposis have been proposed and include the contribution of environmental factors and epigenetics, but also technical and methodological limitations of previous research [13,14]. To have a better insight into the reasons why NGS studies have not been able to resolve more high-penetrant risk genes for CRC and/or polyposis, we reviewed the strategies that have led to the identification of currently well-established hCRC and polyposis predisposing genes and all whole exome- and genome-based NGS studies aimed at the discovery of novel hCRC and polyposis risk genes. Each study was reviewed for its experimental setup to identify its methodological strengths and limitations of candidate gene discovery. Based on our findings and with the latest knowledge of hereditary cancer, we have formulated improvements for future research aimed at unraveling the genetic predisposition of unresolved hCRC and/or polyposis syndromes.

2. High-Penetrant Risk Genes Discovery in hCRC and Polyposis

Most genes associated with hCRC and polyposis were discovered in the late 1900s, far before the NGS-era (Figure 1). In that time, linkage analysis of patients and families with early-onset CRC and/or polyp formation led to the discovery of Lynch syndrome, Familial Adenomatous Polyposis (FAP), and Hamartomatous Polyposis syndromes [16–25]. For these syndromes, genetic analysis of familial cases was supplemented with the analysis of cancer cell lines and (sporadic) tumors. Mismatch repair (MMR) associated genes *MSH2* and *MLH1* were discovered after the observation of linkage of microsatellite markers in multiple kindreds and instability of microsatellite repeats in tumors of affected relatives [20,21]. The observation that pathogenic variants in the MMR-genes *MSH2* and *MLH1* were the cause of familial CRC [22,23] led to the identification of *PMS2* as a cause for Lynch syndrome [24], followed by *MSH6* a couple of years later [25]. While the identification of these four genes explained the majority of individuals with MMR-deficient CRCs with a strong familial aggregation, several families remained without a genetics diagnosis, while their tumors revealed a deficiency of MSH2. It took almost another year before, in 2009, it was found that 3' deletions of *EPCAM*, located upstream of *MSH2*, were involved in Lynch syndrome as well [26,27].
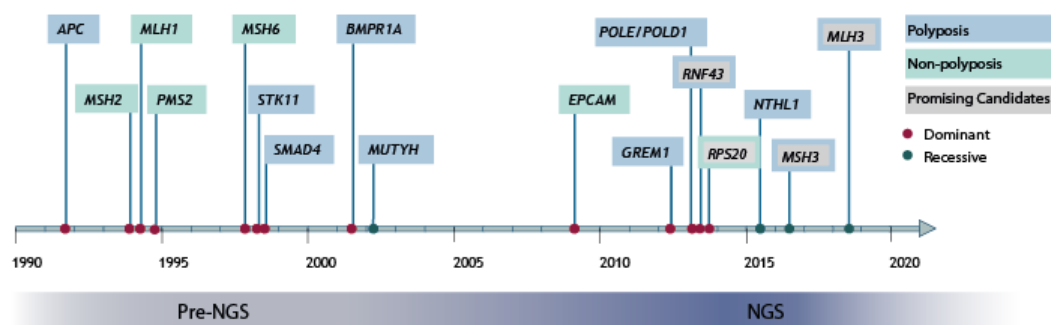

Figure 1. Genes associated with hereditary colorectal cancer (hCRC) and polyposis syndrome. Genes discovered over time for hCRC and polyposis. Established genes in full colors, promising risk genes with colored outline. NGS; Next-generation sequencing.

For discovery of FAP, linkage analysis of families in which polyposis affected multiple generations pointed toward the chromosome 5q21-22 region [17]. Additional analysis of tumors of sporadic CRC patients indicated that the 5q21 locus was frequently lost in tumor cells, pointing at a role for genes in this region in tumor development. Subsequently, germline pathogenic variants in *APC*, located in this 5q21 locus, were identified as the cause of the polyposis phenotype [19,28]. The observation of variable severity among FAP patients with regard to age-of-onset and polyp number led to the establishment of the 'attenuated FAP' (AFAP) phenotype. AFAP is linked to pathogenic variants outside the 3–15 exon regions of *APC* [18,29–31]. In the early 2000s, molecular phenotyping of tumors of three FAP-suspected siblings, but negative for germline pathogenic variants in *APC*, showed a shared increased rate of somatic C:G > A:T transversions in *APC* in their tumors. This observation led to the discovery of another adenomatous polyposis syndrome caused by biallelic pathogenic germline variants in the base-excision repair (BER) gene *MUTYH* [32], also known as *MUTYH*-associated polyposis (MAP). The discovery of more biallelic *MUTYH*-affected cases showed that MAP patients are equally associated with a classical FAP phenotype (>100 polyps) and an AFAP phenotype (<100 polyps) [33,34].

TGF-β signaling proteins *SMAD4* and *BMPR1A*, and AMPK-pathway activator *STK11*, were discovered by linkage and co-segregation studies in the late 1990s [35–37]. These genes are found to be associated with hamartomatous polyposis, characterized by juvenile polyps (for further reading; Zbuk and Eng 2007 [38]).

Subsequently, advances in sequencing techniques such as massive parallel sequencing allowed discovery approaches to change from targeted candidate gene sequencing to exome-wide and genome-wide sequencing of larger patient cohorts to identify disease-causing genes in a more hypothesis-free scenario. Over the past decade, multiple whole-exome and whole-genome sequencing-based studies have been performed (n = 37), all with the general aim to discover rare novel candidate risk genes for hCRC and polyposis (Table 1, Table S1). Thus far, these studies have resulted in over a hundred candidate genes for hCRC and/or polyposis, for which the majority have not been independently validated yet. Some genes are currently under debate as promising candidates for hCRC and polyposis, such as *RNF43*, *MSH3*, *RPS20* and *MLH3* (Figure 1) [39–42], but still await independent validation or additional functional evidence. However, it is remarkable that through all NGS efforts, only few novel high-penetrant risk genes for hCRC and polyposis have been established in the past decade, such as *POLE, POLD1* and *NTHL1* [43,44]. A detailed description of the established and most promising candidate genes for hCRC and polyposis identified through NGS is reviewed elsewhere [45,46].

Table 1. Whole exome-based and genome-based rare high-penetrant risk gene discovery studies for hereditary colorectal cancer and polyposis syndromes up to August 2020.

Author	Key Gene(s)	Inclusion Criteria Index Phenotype	Inclusion Criteria Age	Inclusion Criteria FH	Size Discovery Cohort
Palles et al., 2013 [43]	*POLD1, POLE*	≥10 colorectal tumors	<60 years	FDR or SDR with CRC	15 families (20 cases)
Smith et al., 2013 [47]	*FANCM, LAMB4, PTCHD3, LAMC3, REX2*	Advanced CRC	≤35 years (18 cases)	No, sporadic	50 cases
DeRycke et al., 2013 [48]	*CENPE, KIF23*	Familial CRC	NS	≥2 members affected	16 families (40 cases)
de Voer et al., 2013 [49]	*BUB1, BUB3*	Non-polyposis MMR-proficient CRC	≤40 years	NS	33 cases
Gylfe et al., 2013 [50]	*UACA, SFXN4, TWSG1, PSPH, NUDT7, ZNF490, PRSS37, CCDC18, PRADC1, MRPL3, ARK1C4*	Familial CRC	NS	≥1 FDR with CRC	96 cases
Gala et al., 2014 [42]	*RNF43*	Sessile serrated adenomas	NS	If <5 SSAs, ≥1 FDR with SSAs or CRC	20 cases
Rohlin et al., 2014 [51]	No novel genes found, *POLE* found	Hereditary CRC	n/a	n/a	1 family (3 affected, 1 unaffected)
Nieminen et al., 2014 [41]	*RPS20*	Amsterdam/Bethesda FCCTX	n/a	n/a	1 family (4 cases)
Schulz et al., 2014 [52]	*SEMA4A*	Amsterdam I FCCTX	n/a	n/a	1 family (4 cases)
Esteban-Jurado et al., 2015 [53]	*CDKN1B, XRCC4, EPHX1, NFKBIZ, SMARCA4, BARD1*	Familial CRC [1]	≥1 relative diagnosed <60 [1]	≥3 affected, ≥2 in consecutive generations	29 families (43 cases) [1]
Tanskanen et al., 2015 [54]	*ADAMTS4, CYTL1, SYNE1, MCTP2, ARHGAP12, ATM, DONSON, ROS1*	Non-syndromic early-onset CRC	<40 years	NS	22 cases
Wei et al., 2015 [55]	*HNRNPA0* and *WIF1*	Multiple early-onset cancer	n/a	n/a	1 family (4 affected, 4 unaffected)
Zhang et al., 2015 [56]	*EIF2AK4, LRP5, BUB1*	Familial CRC without polyposis	≤55 years	If ≥40 years, ≥1 FDR with CRC	21 families (23 cases)

Table 1. *Cont.*

Author	Key Gene(s)	Inclusion Criteria Index Phenotype	Inclusion Criteria Age	Inclusion Criteria FH	Size Discovery Cohort
Weren et al., 2015 [44]	*NTHL1*	Multiple adenomatous polyps	NS	NS	51 cases (48 families)
Segui et al., 2015 [57]	*FAN1*	Amsterdam I MMR-proficient CRC	n/a	n/a	1 family (3 cases)
Ngeow et al., 2015 [58]	*SMAD9*	HPS	n/a	n/a	1 family (1 case)
Arora et al., 2015 [59]	*ERCC6, WRN*	CRC or polyposis (≥10 polyps)	<50 years	≥1 relative with CRC	25 cases
Goldberg et al., 2015 [60]	*MCM9*	Multiple mixed polyposis and metastatic CRC	n/a	n/a	1 family (1 cases, 1 unaffected)
Rohlin et al., 2016 [61]	No novel genes found, *GREM1* and *POLE* found	AFAP/atypical polyposis	n/a	n/a	1 family (4 affected, 4 unaffected cases)
Spier et al., 2016 [62]	*DSC2, PIEZO1*	Colorectal adenomatous polyposis	NS	NS	7 cases
Thutkawkorapin et al., 2016 [63]	*DZIP1L, IGSF10, NOTCH1, SF3A1, GAL3ST1*	Familial rectal- and gastric cancer	n/a	n/a	1 family (3 cases)
de Voer et al., 2016 [13]	*PTPN12, LRP6*	non-polyposis MMR-proficient CRC	≤45 years	NS	55 cases
Esteban-Jurado et al., 2016 [64]	*BRCA2/FANCD1, BRIP1/FANCJ, FANCC, FANCE, REV3L/POLZ*	Familial CRC [1]	≥1 relative diagnosed <60 [1]	≥3 affected, ≥2 in consecutive generations	40 families (74 cases) [1]
Chubb et al., 2016 [65]	*POT1, POLE2, MRE11*	CRC	≤55 years	≥1 FDR with CRC	1006 cases
Adam et al., 2016 [40]	*MSH3*	≥20 synchronous or ≥40 metachronous colorectal adenomas	NS	NS	102 cases
Schubert et al., 2017, 2018 [66]	*MIA3*	Amsterdam I MMR stable familial CRC	n/a	n/a	1 family (3 cases WES, 2 cases WES/WGS, 1 cases WGS)
Martín-Morales et al., 2017 [67]	*SETD6*	Amsterdam I FCCTX	≥1 relative diagnosed <50	≥3 affected (≥1 FDR), ≥2 in consecutive generations	1 family (2 cases, 1 unaffected)
Bellido et al., 2018 [68]	*BRF1*	Amsterdam I hereditary CRC	n/a	n/a	1 family (3 CRC cases, 1 BC case)
Franch-Expósito et al., 2018 [69]	*TTF2, TMEM158*	Familial CRC [1]	≥1 relative diagnosed <60 [1]	≥3 affected, ≥2 in consecutive generations	WES: 38 families (71 cases), WGS: 1 case [1]

Table 1. *Cont.*

Author	Key Gene(s)	Inclusion Criteria Index Phenotype	Inclusion Criteria Age	Inclusion Criteria FH	Size Discovery Cohort
Yu et al., 2018 [70]	*DDX20, ZFYVE26, PIK3R3, SLC26A8, ZEB2, TP53INP1, SLC11A1, LRBA, CEBPZ, ETAA1, SEMA3G, IFRD2* and *FAT1*	Amsterdam I/II non-polyposis hereditary CRC	≥1 relative diagnosed <50	≥1 FDR & 2 generations affected	1 family (3 cases)
Olkinuora et al., 2018 [39]	*MLH3*	Adenomatous polyposis	NS	NS	40 cases
Thutkawkorapin et al., 2019 [71]	*BMPR1A, BRIP1, SRC, CLSPN, SEC24B, SSH2, ACACA, NR2C2, INPP4A, DIDO1, ATP10B, PKHD1, UGGT2, MYH13, TFF3*	Simplex early-onset CRC	<40 years	NS	51 cases
Diaz-Gay et al., 2019 [72]	*BRCA2, BLM, ERCC2, RECQL(=WRN), REV3L* and *RIF1*	Familial CRC [1]	≥1 relative diagnosed <60 [1]	≥3 affected, ≥2 in consecutive generations	18 cases [1]
Toma et al., 2019 [73]	*FBLN2*	Familial CRC/SPS	NS	≥2 affected in consecutive generations	16 families (39 cases)
Jansen et al., 2020 [74]	*NOTCH2, RAB25*	Familial CRC	NS	NS	5 families (9 cases)
Toma et al., 2020 [75]	*SMO*	Familial CRC [1]	≥1 relative diagnosed <60 [1]	≥3 affected, ≥2 in consecutive generations	18 families (47 cases) [1]
Bonjoch et al., 2020 [76]	*FAF1*	Familial CRC [1]	≥1 relative diagnosed <60 [1]	≥3 affected, ≥2 in consecutive generations	40 families (75 cases) [1]

Abbreviations: AFAP = Attenuated familial adenomatous polyposis, BC = Breast cancer, CRC = colorectal cancer, FCCTX = familial colorectal cancer type X, FDR = first degree relative, FH = family history, HPS = hamartomatous polyposis, MMR = mismatch repair, NS = not stated in article, n/a = not applicable, SDR = second degree relative, SPS = serrated polyposis syndrome, SSA = sessile serrated adenoma, WES = Whole-exome sequencing, WGS = Whole-genome sequencing. [1] Overlapping cohorts.

3. Strategies for Identification of Rare High-Penetrant Risk Genes

To discover rare but distinct monogenetic causes in a -now known as a- phenotypic heterogeneous group of hCRC and polyposis patients, robust discovery studies are needed. Up to August 2020, we retrospectively identified 37 whole-exome sequencing (WES) and whole-genome sequencing (WGS)-based studies that aimed to discover novel candidate risk genes for hCRC and/or polyposis (Table 1). We reviewed these studies following the general setup of candidate discovery studies, which cover cohort composition, variant discovery and prioritization, and variant validation (Table 1, Table S1). Six studies based discovery on the same cohort that was enlarged over time [53,64,69,72,75,76].

3.1. Discovery Cohort

For the discovery of high-penetrant risk genes for hCRC and polyposis, family history (FH) and inheritance patterns are key factors in variant discovery. We noted that among the 37 candidate gene discovery studies, FH-based inclusion criteria varied from study to study. Some studies used a relatively broad inclusion criterion such as "one first-degree relative or second-degree relative with CRC" while others applied more stringent criteria "the presence of at least three relatives with CRC, of which at least two in consecutive affected generations and at least one case diagnosed with CRC before the age of 60" (Table 1: Inclusion criteria FH) [42,43,48,50,53,56,59,64,69,70,72,75,76]. Furthermore, phenotypic characteristics strongly associated with hereditary CRC and polyposis syndromes, such as tumor types and age-of-onset strongly varied between, but also within cohorts (Table 1: Inclusion criteria index phenotype; Table 1: Inclusion criteria age). The phenotypes that were studied included either polyposis, familial colon and/or rectal cancer, or a mixture of the aforementioned phenotypes. Age-based inclusion criteria were applied in twelve out of the 32 unique discovery cohorts [13,43,47,49,53,54,56,59,64,65,67,69–72,75,76]. However, this age-based inclusion criterion was heterogeneous ranging from an age at diagnosis ≤35 years [47] to diagnosis <60 years [43], to at least one relative diagnosed <60 years [53,64,69,72,75,76]. The observed heterogeneity within these NGS study cohorts is in contrast with the discovery studies that were performed before the NGS-era, as discovery studies before the NGS-era were directed to families with multiple affected members and a strong phenotype of hCRC and/or polyposis. The elaborate inclusion of a range of phenotypes might have contributed to the limited number of high-penetrant risk genes discovered. Therefore, future candidate gene studies may benefit by composing clinical homogenous cohorts with respect to expected mode-of-inheritance and age-of-onset (Table 2).

Table 2. Summary of considerations for future candidate gene discovery studies.

Discovery Cohort Selection
• Clinical homogeneous cohorts based on the expected mode-of-inheritance and/or age-of-onset
Variant Prioritization
• Locus prioritization based on variant recurrence within the cohort • Set allele-frequency cut-offs based on expected mode-of-inheritance   - Dominant: MAF < 0.0005; Recessive MAF < 0.007 • Concordance of multiple in silico prediction tools • Co-segregation among family members
Variant Validation
• Molecular tumor analysis to determine molecular phenotype of tumor • Evidence based on cellular models • Case-control comparison to specify the genotype-phenotype correlation

MAF = Minor Allele Frequency.

3.2. Variant Prioritization

3.2.1. Locus Prioritization

To identify novel genetic predispositions, WES or WGS data of discovery cohorts need to undergo a filtering to prioritize potentially pathogenic variants. In general, we observed prioritization based on either complete screening of WES or WGS data or, more targeted, by focusing on specific genes or regions that are more likely to be involved in hereditary cancer. The main applied variant prioritization strategies included linkage (*n* = 8 studies [41,43,52,55,61,66,73,75]), variants shared among affected relatives or absence of the variant in unaffected relatives (*n* = 19 studies [41,43,48,51–53,55,57,60,61,63,64,66–70,74, 76]) and gene function (*n* = 14 studies [13,47,53,57–59,63–65,71,72,74–76]). Other approaches include the prioritization of recurrent variants (*n* = 7 studies [13,39,40,44,50,56,65]) and prioritization based on expected recessive or dominant mode-of-inheritance (*n* = 7 studies [52–54,60,62,67,71]). A minority of the studies primarily focused on deleteriousness of a variant (*n* = 6 studies [39,40,44,47,49,50]) or combined germline and tumor analysis (*n* = 1 studies [47]) (Table S1: Approach for discovery). Gene discovery studies that led to the identification of pathogenic variants in *POLE*, *POLD1*, and *NTHL1* were based on the prioritization of variants that were shared among affected family members or that were recurrent in the study population [43,44], suggesting that discovery cohort selection is most important for candidate gene discovery.

3.2.2. Allele Frequency Cut-Offs

Additional to these aforementioned strategies, a commonly used approach among discovery studies is the "rare disease rare variant hypothesis", meaning that rare phenotypes are caused by variants that are rare (i.e., have a (very) low MAF) in the general population. Within the 37 candidate gene discovery studies, the applied MAF cut-off ranged from 0.00–0.20 (Table S1: Applied MAF). Four studies explicitly adapted their MAF cut-off to presumed dominant or recessively inherited genetic predispositions in their study populations [40,44,62,71]. These four studies applied more stringent MAF cut-offs in dominant scenarios (MAF < 0.01–0.001) and looser MAF cut-offs for recessive inheritance patterns (MAF < 0.03–0.01) [40,44,62,71].

A prominent question in MAF-based filtering is 'How low can we go?'. In other words, what is the optimal MAF frequency to identify rare potentially pathogenic variants for follow up? Assuming a high-penetrant rare disease model, in theory MAFs could be set as low as the expected prevalence of disease in the general population for dominant inheritance scenarios. The most recently discovered dominantly inherited polyposis syndrome proof-reading associate polyposis (PPAP), caused by variants in the exonuclease domain of *POLE* and *POLD1*, is discovered only in about 0.5% of the familial early-onset CRC disease population [77]. ClinVar Class V variants located in the exonuclease domains of *POLE* (codons 268-471; NM_006231) and *POLD1* (codons 304-517; NM_002691), such as *POLE* p.(Leu424Val) and *POLD1* p.(Ser478Asn) are absent in the general population [78]. Based on these observations, we argue that novel dominant hCRC and polyposis syndromes are likely just as rare in the general population as PPAP and therefore identification of novel dominant high-penetrant risk genes will allow (very) low variant allele frequency cut-offs. Chubb et al. screened a selected population of familial early-onset CRC cases, and concluded that about 0.5 percent of this population carries a pathogenic or likely pathogenic variant in *POLE* or *POLD1*. Based on the assumption that this population of familial early-onset CRC cases could completely be explained by high-penetrant predispositions and that up to 10% of the CRCs can be explained by rare genetic predispositions [8,77], we anticipate that novel high-penetrant variants for hCRC and polyposis will have a MAF lower than 0.0005, or will be completely absent in the general population.

For recessive disease scenarios, including compound heterozygous variants, setting a MAF cut-off is much more difficult as heterozygous variants can be expected in the general population in the absence of a disease phenotype. For recessive disease genes such as *MUTYH* and *NTHL1*, combined minor allele frequencies up to 0.007 (0.7%) are observed, and population-specific allele frequencies

can be as high as 0.006 (North-Western European; *MUTYH* p.(Gly393Asp)) [44,79]. This observation implies that a MAF cut-off of 0.001, as often applied for dominant diseases, is too stringent. Based on these calculations, MAF cut-offs as low as 0.007 should be set for variants in recessive disease scenarios.

3.2.3. In Silico Pathogenicity

Not every selected rare variant will have pathogenic –and thus a cancer predisposing– potential. Therefore, after rare variant selection, the pathogenicity potential of a variant needs assessment. Twenty-nine out of 37 studies reported the use of one or more in silico tools to predict variant pathogenicity [24,39–41,43,44,48–50,52–59,61–64,67,68,71–76]. In general, combinations of different variant scoring tools such as SIFT, PolyPhen-2, and MutationTaster were used (Table S1). Specific criteria for pathogenic assertion were applied in 21 out of 29 studies, such as prediction for deleteriousness in a majority of the assessed tools and application of cut-offs for CADD and/or PhyloP scores that indicate the conservation of a position [80–82]. The performance and concordance of such in silico tools for pathogenicity greatly varies [80,83]. In line with the observation that most studies make use of multiple in silico tools, a previous comparison of 25 commonly used algorithms showed that prediction of five algorithms (SIFT, PolyPhen, CADD, PROVEAN, and MutationTaster) resulted in a higher concordance compared to other combinations [80]. For example, known pathogenic missense variants such as *MUTYH* (NM_012222.2; c.724C > T; p.(Arg242His)) and *POLE* (NM_006231.2; c.1270C > G; p.(Leu424Val)) are predicted to be deleterious in all five algorithms. However, a high concordance of these in silico tools is not a guarantee for the identification of a pathogenic variant, therefore in silico predictions should facilitate variant prioritization but should not serve as evidence in itself.

3.2.4. Co-Segregation

Next to germline-based filtering strategies and in silico pathogenicity predictions, additional prioritization methods are applied to select the most likely causative variant. One of the primary types of evidence is co-segregation of the variant with the affected status throughout a family. Co-segregation analysis was performed in 24 out of 37 studies (Table S1: Segregation [39–41,43,44,48–53,57,60,61,63,64, 66–69,73–76]). Within the studies, co-segregation of the candidate risk locus was not always concordant in affected vs. unaffected relatives. Gylfe et al. identified the *TWSG1* nonsense variant (c.121C > T; p.(Gln41 *)) in two families, which segregated with the affected status in one family, but not in the other [50]. Jansen et al. showed that that all variants that were found in the affected individual (with CRC at age 14) were also detected in either the unaffected father or the unaffected mother of the proband [74]. Co-segregation analysis is considered essential for decisions in variant follow up, especially for discovery of novel high-penetrant risk genes. To illustrate, incomplete co-segregation of the *RAD52* truncating variant (c.590_593dupAACC; p.(Ser199Thrfs*88)), in contrast to the complete genotype-phenotype segregation of *FAF1* missense variant (c.1111G > A; p.(Asp371Asn)) in a family with CRC, led to the decision to follow up the latter one [76]. Co-segregation analysis of both affected and unaffected family members will rapidly gain insight for variant follow up. The lack of segregation in healthy family members can indicate variant pathogenicity as well, taking into consideration the age of the person and the expected age-of-onset of the disease. Even when co-segregation analysis in affected family members cannot be performed, testing of unaffected family members may facilitate variant prioritization.

3.3. Variant Validation

3.3.1. Molecular Tumor Analysis

Somatic molecular events in colorectal tumors were studied in eleven discovery studies (Table S1: Molecular tumor characteristics [39,40,42–44,47,52,57,62,68,72]). Within these eleven studies, analysis of tumor mutational events varied from driver gene genotyping of *KRAS*, *BRAF*, and/or *NRAS* [39,42,68], to analysis of the genome-wide mutational spectrum of the tumor [43,44,57,72]. Driver gene analyses

are often applied in the context of therapy-stratification and evolutionary studies of the tumor. However, driver genes were also screened in several candidate gene discovery studies as well. Nine out of these eleven studies performed driver gene screening, but the vast majority of these nine studies did not find any predominant substitution in the analyzed driver genes [39,43,44,47,52,68,84]. Gala et al. found enrichment for *BRAF* p.(Val600Glu) in sessile serrated adenomas [42]. *BRAF* p.(Val600Glu) mutations in tumors are strongly associated with *MLH1* promoter methylation in sporadic CRC cases and thus are a predictor of negative MMR mutation status [85]. Summarizing these study results, driver gene screening in hCRC and polyposis gene discovery studies may have additional value when screening is performed on tumors of all discovery cases to further stratify hCRC and polyposis phenotypes, as complete screening could facilitate the identification of sporadic cases and may provide an extra tumor characteristic that could stratify patients for follow up screening. Nonetheless, known driver gene mutations seem not to be discriminative for specific germline predispositions for polyposis, at least not in adenomatous polyposis.

Based on the established genes discovered for hCRC and polyposis, pathogenic variants in two main mechanisms can give rise to colorectal tumors. One is the altered activity of a tumorigenic process; i.e., inactivation of *APC* and activated WNT signaling, and the tumor suppressive roles of *SMAD4* and *BMPR1A* in TGF-β signaling. The other main mechanism that predisposes tumor development is a defect in DNA repair, such as MMR defects in Lynch syndrome, and disruptive base-excision repair (BER) in adenomatous polyposis. For DNA repair deficiencies in particular, it is known that germline defects in specific genes give rise to specific molecular tumor phenotypes. The most prominent molecular phenotype is the observation of microsatellite instability in the tumor due to (germline) MMR defects. Defects in DNA repair pathways may result in distinct mutational patterns in the genomes of tumors, now known as mutational signatures [86]. The observation of an increased rate of C:G > A:T transversions in *APC* in tumors led to the discovery of *MUTYH*-associated polyposis in the early 2000s. MUTYH deficiency, causing 8-OxoG BER pathway redundancy, is now linked to mutational signatures 18 and 36 [87,88]. Subsequently, in 2015, tumors of *NTHL1*-associated tumor syndrome (NATS) patients with germline nonsense mutations in the BER gene *NTHL1* showed an increased rate of C:G > T:A mutations in a unique mutation context, resulting in mutational signature 30 [89]. Polymerase proofreading defects give rise to mutational signatures 10a and 10b, and mismatch repair defects are associated with mutational signatures 6, 15, 20, and 26 [90,91]. These findings suggest that mutational patterns, rather than single driver gene events, may facilitate identification and validation of candidate genes for hCRC and polyposis syndromes (further reading: Grolleman et al., 2019 [92]).

3.3.2. Functional Characterization of the Variant

Co-segregation of the variant and mutational profiling give a strong indication for pathogenicity, but these two aspects do not directly confirm the causality of the germline variant to the disease phenotype. Therefore, additional evidence may include the expression pattern of the affected gene and functional characterization of the variant. In this review, we consider functional characterization as the use of in vitro or in vivo assays to determine whether: (i) genetic variants disrupt or enhance protein function, but more importantly (ii) how an altered protein function may give rise to a certain phenotype. A combination of these two is likely essential for full functional characterization. Gene and/or protein expression alone cannot be considered as a validation method, as the presence of gene product does not determine the pathogenicity of the variant and the effect on down-stream targets. Overall, functional characterization of variants was limited in the reviewed candidate studies. In total, 17 out of 37 studies used patient-derived material or used human cell lines as an in vitro model to test variant consequences and functional characterization (Table S1: Candidate gene transcription/protein expression and Functional characterization [13,39–42,44,49,52,57–60,66–69,76]). However, most studies provide mainly protein expression data and limited data on assessment of protein function and downstream interactors [13,40,66,69], while both expression data as well as functional data should

be in line with the role of the variant in tumorigenesis. For example, we previously showed that missense variants in *LRP6* did not affect expression and localization compared to wildtype *LRP6*. However, using a TOPflash assay we observed that these missense variants increased WNT-signaling activity [13]. In other studies, more in-depth functional analyses in tumorigenic processes included experiments to analyze the effect of variants on cell migration and proliferation, cell cycle progression, and apoptosis [52,68,76]. Schulz et al. showed increased activation of the PI3K/AKT and MAPK/ERK pathways for SEMA4A^{V78M} compared to SEMA4A^{wt} by flowcytometry and immunofluorescence in HCT116 cells, but not in 293T cells [52]. Bellido et al. showed a potential pathogenicity of *BRF1* missense variants as low viability was observed in a *BRF1*-dependent growth assay in yeast that harbored these missense variants [68]. Bonjoch et al. performed various assays including immunofluorescence and caspase-3 activity assay to show that *FAF1* missense variants lead to upregulation of β-catenin and reduced apoptosis in DLD1 cells [76]. In addition to two-dimensional models, colon epithelial organoid models may be an interesting alternative to study proliferation, survival and mutational processes for specific candidate predisposing genes in a three-dimensional setting. It was previously shown by Drost et al. that an organoid in which *NTHL1* was knocked out using CRISPR/Cas9 shows the same mutational pattern as tumors from NTHL1 deficient individuals [89,93]. Standardized methods have been developed for the in vitro culture of primary colon organoids, which may facilitate the use of this model over two-dimensional cultures [94]. Nevertheless, only a selected number of parameters can be studied with in vitro assays. Therefore, a combination of tests, including in vitro functional assays as well as tumor sequencing data and/or co-segregation analysis, should point towards a causal genotype-phenotype relationship (Table 2).

3.3.3. Case-Control Validation

Next to functional impact of a candidate variant, case-control validation is an alternative and complementary approach to validate the causal relationship between a germline variant and the hereditary tumor syndrome. The validation is based on a significantly higher recurrence rate of the hCRC or polyposis syndrome phenotype in cases than controls. During the identification process of candidate genes, a low MAF in population controls is already used as selection parameter, however in case-control validations other variants in the same gene are also taken into consideration. Moreover, case-control studies contribute to the description of the complete phenotypic spectrum of a candidate gene. Validation cohorts have been used in 23 out of 37 studies, and statistical testing for enrichment in cases vs. controls was performed in fifteen studies [13,42–44,52,55,56,59,64,65,68,72–75]. Despite efforts, the lack of sufficient power is frequently mentioned as reason for not finding significant differences in case-control analysis and not being able to validate newly found risk genes. To illustrate, Chubb et al. performed a screening of 1,006 cases and healthy 1,609 individuals. Even though the cohort was specifically targeted to dominant hCRC syndromes by the selection of CRC cases ≤55 years and with at least one first-degree relative with CRC, only the well-established genes for hCRC and polyposis (*APC*, *MLH1*, *MSH2*) reached significant enrichment in cases versus controls [65]. In hCRC and polyposis syndromes, the rareness of the newly found syndromes together with their population-specific allele frequency as e.g., noted for *NTHL1* [79], show that the low MAFs of rare variants make it almost impossible to find significant associations, even in studies with over a thousand cases. It is calculated that for genetic predispositions with moderate or high-penetrance (OR > 2), required sample sizes need to reach 10,000 to even 100,000 cases and controls [65,95]. Therefore, the purpose of case-control validation in future studies may shift from finding significant differences in cases versus controls, to specification of the phenotype associated with the risk gene. Validation cohorts may include a phenotypic range of genetic tumor risk syndrome patients to further determine the genotype-phenotype presentation in these rare disease patients (Table 2).

4. Missing Heritability Explained by Known or Common Risk Genes

In the era of massive parallel sequencing, whole-exome and whole-genome sequencing of suspected hCRC and polyposis patients resulted in relatively few widely established novel high-penetrant genes for hCRC and polyposis syndromes. In the above sections, we gave insights into the applied methods and discussed considerations for future candidate gene discovery studies. However, next to the inability to identify novel genetic predispositions possibly due to inconsistencies in study setup, other scenarios could also contribute to the observed missing heritability in hCRC and polyposis syndromes. These scenarios include that disease-causing variants might have been missed in known hCRC and polyposis syndrome genes due to technical limitations or unforeseen inheritance patterns, or that the hCRC and polyposis phenotypes are the result of multiple variants in low- and moderate penetrant risk genes.

4.1. Identification of Variants in Known hCRC and Polyposis Risk Genes by Whole-Genome Sequencing

Most routine diagnostic, but also research facilities, focus on the screening of coding regions of the genome for pathogenic variant identification, either by targeted- or whole-exome sequencing. However, targeted screening of both the coding and non-coding regions of the genes or WGS could be of great potential for reduction of missing heritability. Short-read WGS outperforms WES in detection of variants also in coding regions due to more homogeneous coverage with higher quality and better variant calling [96,97]. Moreover, WGS of patient cohorts might facilitate discovery of missed non-coding variants in known hCRC and polyposis genes. In the past, deep-intronic and promoter variants were described in tumor suppressor genes *APC* and *PTEN*, which makes sequencing of these non-coding regions of particular interest for unresolved hCRC and polyposis patients [98–103]. Long-read sequencing and optical mapping techniques might be valuable as well, as these techniques are specifically directed to the detection of complex and structural variants, and allow alignment and variant mapping in regions that used to be uncovered in the past due to their nucleotide composition (e.g., extreme GC-rich, and multiple short repeats) [104,105]. Complexities of these regions and the structural variants itself, make that these variants remain understudied in whole exome-based and whole genome-based techniques, and the inability to detect those might explain part of the observed missing heritability in hCRC and polyposis syndromes.

4.2. Mosaic and De Novo Variant in Known hCRC and Polyposis Syndrome Genes

Next to dominant and recessive inheritance patterns, other forms of predisposition might also explain a proportion of the missing heritability in hCRC and polyposis patients. A de novo onset of a constitutive genetic defect and mosaicism, caused by mutations arisen in (early) embryonic development, are likely overseen causes in genetically unresolved hCRC and polyposis patients, as these patients often lack a positive family history. Despite a negative family history, these patients may display severe polyposis and carcinomas at young age. In example, de novo and mosaicism rates among FAP patients with suspected sporadic disease range from 4% to 25% [84,106–109]. For identification of novel genes involved in de novo onset of disease, trio-studies could be extremely valuable, as sequencing both healthy parents and the proband increases the diagnostic yield in rare-diseases [110]. Trio-sequencing may be chosen in case of a severe polyposis phenotype or CRC at an exceptionally young age, but in the absence of familial aggregation. For detection of mosaicism, multiple clonal expansions like polyps should be evaluated to determine shared pathogenic variants, which subsequently can be evaluated at high sensitivity in leukocyte-derived DNA and normal tissues to render insight in variant distribution throughout the different tissues in the body.

4.3. Polygenic Risk Scores

The influence of common, low-penetrant risk loci has been studied since the introduction of genome-wide studies. Genome-wide association studies use the genetic risk information from

the millions of discovered single nucleotide polymorphisms to determine an individual's genetic susceptibly for a specific, usually complex, trait. Using this information, the sum of all common, intermediate and rare variants that are thought to contribute to disease susceptibility. The interactions within and between these variants form a Polygenic risk score (PRS). PRSs have been studied in several complex traits as well as several cancers, including breast and prostate cancer [111,112]. Additionally for colorectal cancer, PRSs in combination with family history seem to be feasible for risk stratification [113]. However, little is known on the additional role of polygenicity in contribution to monogenic causes of hCRC and polyposis syndromes. A preliminary publication of Fahed et al. studied whether polygenic risk can account for variation among carriers for monogenic variants that are predisposed to Lynch syndrome and showed that the odds ratios for colorectal cancer increased with higher polygenic scores [114]. Research from Schlafly et al. shows a discovery approach using PRS to prioritize families for high-penetrant rare risk genes. Using this approach in 404 melanoma-prone families, they found that families carrying putative causal predisposition had a lower PRS [115]. It is too early for implementation, but both studies show the potential of PRSs as a tool to prioritize families for discovery cohort inclusion in hCRC and polyposis syndromes gene discovery studies.

5. Conclusions

A fair number of hCRC and polyposis patients are considered to remain genetically unexplained, which hampers risk assessments for patients in whom no genetically underlying cause is identified. The estimates of missing heritability are mainly based on twin and family studies, which may be biased by non-additive genetic effects or incorrect assumption about the shared environment. Nevertheless, the proportion of unresolved early-onset and/or familial CRC patients, urges the investigation of additional genetic causes. By collecting all whole exome-based and whole genome-based discovery studies and listing their study design, we aimed to provide knowledge on why the missing heritability is not (yet) reduced and provide improvements for future studies. These improvements cover the setup of high-quality discovery studies by including phenotypically well-defined early-onset CRC and/or polyposis syndrome patients into perhaps smaller, but more specific cohorts for candidate gene searches. In this approach, the availability of enough material and patient information should have the highest priority for inclusion to facilitate detailed characterization of both germline DNA and tumor material. Once a variant and/or candidate gene is selected, validation needs to be multi-leveled and elaborate to provide robust and unambiguous evidence for the casual role of the genetic variant.

In conclusion, novel high-penetrant risk genes for hCRC and polyposis syndromes will be rare in a disease group that is heterogeneous in nature. This heterogeneity needs to be taken into account in future discovery and validation strategies for the identification of novel genetic predispositions in hCRC and polyposis syndromes. A stringently selected study population and strict criteria for variant identification, together with appropriate functional validation, will contribute to a further delineation of the missing heritability. This complete analysis of this heterogeneous disease group will provide in-depth genotype-phenotype information, contributing to future diagnostics and lead to tumor- and patient-specific treatment and surveillance strategies.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AFAP	Attenuated familial adenomatous polyposis syndrome
BER	Base-excision repair pathway
CRC	Colorectal cancer
FAP	Familial adenomatous polyposis
FH	Family history
hCRC	Hereditary colorectal cancer
MAF	Minor allele frequency
MAP	MutyH-associated polyposis
MMR	Mismatch repair
NATS	*NTHL1*-associated tumor syndrome
NGS	Next-generation sequencing
PPAP	Polymerase proofreading associated polyposis
PRS	Polygenic risk score
WES	Whole exome sequencing
WGS	Whole genome sequencing

References

1. *Cancer Today*; International Agency for Research on Cancer: Lyon, France, 2020.
2. 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. [CrossRef] [PubMed]
3. 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*, 260–266. [CrossRef] [PubMed]
4. Mucci, L.A.; Hjelmborg, J.B.; Harris, J.R.; Czene, K.; Havelick, D.J.; Scheike, T.; Graff, R.E.; Holst, K.; Moller, S.; Unger, R.H.; et al. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA* **2016**, *315*, 68–76. [CrossRef] [PubMed]
5. Graff, R.E.; Moller, S.; Passarelli, M.N.; Witte, J.S.; Skytthe, A.; Christensen, K.; Tan, Q.; Adami, H.O.; Czene, K.; Harris, J.R.; et al. Familial Risk and Heritability of Colorectal Cancer in the Nordic Twin Study of Cancer. *Clin. Gastroenterol. Hepatol.* **2017**, *15*, 1256–1264. [CrossRef] [PubMed]
6. Yurgelun, M.B.; Kulke, M.H.; Fuchs, C.S.; Allen, B.A.; Uno, H.; Hornick, J.L.; Ukaegbu, C.I.; Brais, L.K.; McNamara, P.G.; Mayer, R.J.; et al. Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. *J. Clin. Oncol.* **2017**, *35*, 1086–1095. [CrossRef] [PubMed]
7. Lynch, H.T.; de la Chapelle, A. Hereditary colorectal cancer. *N. Engl. J. Med.* **2003**, *348*, 919–932. [CrossRef]
8. Huyghe, J.R.; Bien, S.A.; Harrison, T.A.; Kang, H.M.; Chen, S.; Schmit, S.L.; Conti, D.V.; Qu, C.; Jeon, J.; Edlund, C.K.; et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat. Genet.* **2019**, *51*, 76–87. [CrossRef]
9. Weigl, K.; Chang-Claude, J.; Hsu, L.; Hoffmeister, M.; Brenner, H. Establishing a valid approach for estimating familial risk of cancer explained by common genetic variants. *Int. J. Cancer* **2019**. [CrossRef]
10. Schmit, S.L.; Edlund, C.K.; Schumacher, F.R.; Gong, J.; Harrison, T.A.; Huyghe, J.R.; Qu, C.; Melas, M.; Van Den Berg, D.J.; Wang, H.; et al. Novel Common Genetic Susceptibility Loci for Colorectal Cancer. *J. Natl. Cancer Inst.* **2019**, *111*, 146–157. [CrossRef]
11. Law, P.J.; Timofeeva, M.; Fernandez-Rozadilla, C.; Broderick, P.; Studd, J.; Fernandez-Tajes, J.; Farrington, S.; Svinti, V.; Palles, C.; Orlando, G.; et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat. Commun.* **2019**, *10*, 2154. [CrossRef]
12. De Voer, R.M.; Hahn, M.M.; Weren, R.D.; Mensenkamp, A.R.; Gilissen, C.; van Zelst-Stams, W.A.; Spruijt, L.; Kets, C.M.; Zhang, J.; Venselaar, H.; et al. Identification of Novel Candidate Genes for Early-Onset Colorectal Cancer Susceptibility. *PLoS Genet.* **2016**, *12*, e1005880. [CrossRef] [PubMed]

13. Hahn, M.M.; de Voer, R.M.; Hoogerbrugge, N.; Ligtenberg, M.J.; Kuiper, R.P.; van Kessel, A.G. The genetic heterogeneity of colorectal cancer predisposition—Guidelines for gene discovery. *Cell. Oncol.* **2016**, *39*, 491–510. [CrossRef] [PubMed]

14. Schubert, S.A.; Morreau, H.; de Miranda, N.; van Wezel, T. The missing heritability of familial colorectal cancer. *Mutagenesis* **2019**. [CrossRef]

15. Risch, N. The genetic epidemiology of cancer: Interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol. Biomark. Prev.* **2001**, *10*, 733–741.

16. 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. [CrossRef]

17. Solomon, E.; Voss, R.; Hall, V.; Bodmer, W.F.; Jass, J.R.; Jeffreys, A.J.; Lucibello, F.C.; Patel, I.; Rider, S.H. Chromosome 5 allele loss in human colorectal carcinomas. *Nature* **1987**, *328*, 616–619. [CrossRef]

18. Leppert, M.; Burt, R.; Hughes, J.P.; Samowitz, W.; Nakamura, Y.; Woodward, S.; Gardner, E.; Lalouel, J.M.; White, R. Genetic analysis of an inherited predisposition to colon cancer in a family with a variable number of adenomatous polyps. *N. Engl. J. Med.* **1990**, *322*, 904–908. [CrossRef]

19. Nishisho, I.; Nakamura, Y.; Miyoshi, Y.; Miki, Y.; Ando, H.; Horii, A.; Koyama, K.; Utsunomiya, J.; Baba, S.; Hedge, P. Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science* **1991**, *253*, 665–669. [CrossRef]

20. 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. [CrossRef]

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. [CrossRef]

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. [CrossRef]

23. 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. [CrossRef]

24. 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. [CrossRef] [PubMed]

25. Papadopoulos, N.; Nicolaides, N.C.; Liu, B.; Parsons, R.; Lengauer, C.; Palombo, F.; D'Arrigo, A.; Markowitz, S.; Willson, J.K.; Kinzler, K.W.; et al. Mutations of GTBP in genetically unstable cells. *Science* **1995**, *268*, 1915–1917. [CrossRef] [PubMed]

26. Kovacs, M.E.; Papp, J.; Szentirmay, Z.; Otto, S.; Olah, E. Deletions removing the last exon of TACSTD1 constitute a distinct class of mutations predisposing to Lynch syndrome. *Hum. Mutat.* **2009**, *30*, 197–203. [CrossRef]

27. Ligtenberg, M.J.; Kuiper, R.P.; Chan, T.L.; Goossens, M.; Hebeda, K.M.; Voorendt, M.; Lee, T.Y.; Bodmer, D.; Hoenselaar, E.; Hendriks-Cornelissen, S.J.; et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3′ exons of TACSTD1. *Nat. Genet.* **2009**, *41*, 112–117. [CrossRef]

28. Nakamura, Y.; Nishisho, I.; Kinzler, K.W.; Vogelstein, B.; Miyoshi, Y.; Miki, Y.; Ando, H.; Horii, A.; Nagase, H. Mutations of the adenomatous polyposis coli gene in familial polyposis coli patients and sporadic colorectal tumors. *Princess Takamatsu Symp.* **1991**, *22*, 285–292. [CrossRef]

29. Lynch, H.T.; Smyrk, T.C.; Lanspa, S.J.; Lynch, P.M.; Watson, P.; Strayhorn, P.C.; Bronson, E.K.; Lynch, J.F.; Priluck, I.A.; Appelman, H.D. Phenotypic variation in colorectal adenoma/cancer expression in two families. Hereditary flat adenoma syndrome. *Cancer* **1990**, *66*, 909–915. [CrossRef]

30. Knudsen, A.L.; Bisgaard, M.L.; Bulow, S. Attenuated familial adenomatous polyposis (AFAP). A review of the literature. *Fam. Cancer* **2003**, *2*, 43–55. [CrossRef]

31. Spirio, L.; Olschwang, S.; Groden, J.; Robertson, M.; Samowitz, W.; Joslyn, G.; Gelbert, L.; Thliveris, A.; Carlson, M.; Otterud, B.; et al. Alleles of the APC gene: An attenuated form of familial polyposis. *Cell* **1993**, *75*, 951–957. [CrossRef]

32. 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. [CrossRef] [PubMed]

33. Nielsen, M.; Franken, P.F.; Reinards, T.H.C.M.; Weiss, M.M.; Wagner, A.; van der Klift, H.; Kloosterman, S.; Houwing-Duistermaat, J.J.; Aalfs, C.M.; Ausems, M.G.E.M.; et al. Multiplicity in polyp count and extracolonic manifestations in 40 Dutch patients with MYH associated polyposis coli (MAP). *J. Med. Genet.* **2005**, *42*, e54. [CrossRef] [PubMed]

34. Wang, L.; Baudhuin, L.M.; Boardman, L.A.; Steenblock, K.J.; Petersen, G.M.; Halling, K.C.; French, A.J.; Johnson, R.A.; Burgart, L.J.; Rabe, K.; et al. MYH mutations in patients with attenuated and classic polyposis and with young-onset colorectal cancer without polyps. *Gastroenterology* **2004**, *127*, 9–16. [CrossRef] [PubMed]

35. Jenne, D.E.; Reimann, H.; Nezu, J.; Friedel, W.; Loff, S.; Jeschke, R.; Muller, O.; Back, W.; Zimmer, M. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nat. Genet.* **1998**, *18*, 38–43. [CrossRef]

36. 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. [CrossRef]

37. 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. [CrossRef]

38. Zbuk, K.M.; Eng, C. Hamartomatous polyposis syndromes. *Nat. Clin. Pract. Gastroenterol. Hepatol.* **2007**, *4*, 492–502. [CrossRef]

39. Olkinuora, A.; Nieminen, T.T.; Martensson, E.; Rohlin, A.; Ristimaki, A.; Koskenvuo, L.; Lepisto, A.; Swedish Extended Genetic Analysis of Colorectal Neoplasia Study Group; Gebre-Medhin, S.; Nordling, M.; et al. Biallelic germline nonsense variant of MLH3 underlies polyposis predisposition. *Genet. Med.* **2018**. [CrossRef]

40. Adam, R.; Spier, I.; Zhao, B.; Kloth, M.; Marquez, J.; Hinrichsen, I.; Kirfel, J.; Tafazzoli, A.; Horpaopan, S.; Uhlhaas, S.; et al. Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. *Am. J. Hum. Genet.* **2016**, *99*, 337–351. [CrossRef]

41. Nieminen, T.T.; O'Donohue, M.F.; Wu, Y.; Lohi, H.; Scherer, S.W.; Paterson, A.D.; Ellonen, P.; Abdel-Rahman, W.M.; Valo, S.; Mecklin, J.P.; et al. Germline mutation of RPS20, encoding a ribosomal protein, causes predisposition to hereditary nonpolyposis colorectal carcinoma without DNA mismatch repair deficiency. *Gastroenterology* **2014**, *147*, 595–598.e5. [CrossRef]

42. Gala, M.K.; Mizukami, Y.; Le, L.P.; Moriichi, K.; Austin, T.; Yamamoto, M.; Lauwers, G.Y.; Bardeesy, N.; Chung, D.C. Germline mutations in oncogene-induced senescence pathways are associated with multiple sessile serrated adenomas. *Gastroenterology* **2014**, *146*, 520–529. [CrossRef] [PubMed]

43. 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. [CrossRef] [PubMed]

44. Weren, R.D.; Ligtenberg, M.J.; Kets, C.M.; de Voer, R.M.; Verwiel, E.T.; Spruijt, L.; van Zelst-Stams, W.A.; Jongmans, M.C.; Gilissen, C.; Hehir-Kwa, J.Y.; et al. A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. *Nat. Genet.* **2015**, *47*, 668–671. [CrossRef] [PubMed]

45. Valle, L. Recent Discoveries in the Genetics of Familial Colorectal Cancer and Polyposis. *Clin. Gastroenterol. Hepatol.* **2017**, *15*, 809–819. [CrossRef] [PubMed]

46. Valle, L.; de Voer, R.M.; Goldberg, Y.; Sjursen, W.; Forsti, A.; Ruiz-Ponte, C.; Caldes, T.; Garre, P.; Olsen, M.F.; Nordling, M.; et al. Update on genetic predisposition to colorectal cancer and polyposis. *Mol. Aspects Med.* **2019**. [CrossRef] [PubMed]

47. Smith, C.G.; Naven, M.; Harris, R.; Colley, J.; West, H.; Li, N.; Liu, Y.; Adams, R.; Maughan, T.S.; Nichols, L.; et al. Exome resequencing identifies potential tumor-suppressor genes that predispose to colorectal cancer. *Hum. Mutat.* **2013**, *34*, 1026–1034. [CrossRef] [PubMed]

48. DeRycke, M.S.; Gunawardena, S.R.; Middha, S.; Asmann, Y.W.; Schaid, D.J.; McDonnell, S.K.; Riska, S.M.; Eckloff, B.W.; Cunningham, J.M.; Fridley, B.L.; et al. Identification of novel variants in colorectal cancer families by high-throughput exome sequencing. *Cancer Epidemiol. Biomark. Prev.* **2013**, *22*, 1239–1251. [CrossRef]

49. De Voer, R.M.; Geurts van Kessel, A.; Weren, R.D.; Ligtenberg, M.J.; Smeets, D.; Fu, L.; Vreede, L.; Kamping, E.J.; Verwiel, E.T.; Hahn, M.M.; et al. Germline mutations in the spindle assembly checkpoint genes BUB1 and BUB3 are risk factors for colorectal cancer. *Gastroenterology* **2013**, *145*, 544–547. [CrossRef]

50. 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. [CrossRef]

51. Rohlin, A.; Zagoras, T.; Nilsson, S.; Lundstam, U.; Wahlstrom, J.; Hulten, L.; Martinsson, T.; Karlsson, G.B.; Nordling, M. A mutation in POLE predisposing to a multi-tumour phenotype. *Int. J. Oncol.* **2014**, *45*, 77–81. [CrossRef]

52. Schulz, E.; Klampfl, P.; Holzapfel, S.; Janecke, A.R.; Ulz, P.; Renner, W.; Kashofer, K.; Nojima, S.; Leitner, A.; Zebisch, A.; et al. Germline variants in the SEMA4A gene predispose to familial colorectal cancer type X. *Nat. Commun.* **2014**, *5*, 5191. [CrossRef] [PubMed]

53. Esteban-Jurado, C.; Vila-Casadesus, M.; Garre, P.; Lozano, J.J.; Pristoupilova, A.; Beltran, S.; Munoz, J.; Ocana, T.; Balaguer, F.; Lopez-Ceron, M.; et al. Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer. *Genet. Med.* **2015**, *17*, 131–142. [CrossRef] [PubMed]

54. Tanskanen, T.; Gylfe, A.E.; Katainen, R.; Taipale, M.; Renkonen-Sinisalo, L.; Jarvinen, H.; Mecklin, J.P.; Bohm, J.; Kilpivaara, O.; Pitkanen, E.; et al. Systematic search for rare variants in Finnish early-onset colorectal cancer patients. *Cancer Genet.* **2015**, *208*, 35–40. [CrossRef] [PubMed]

55. Wei, C.; Peng, B.; Han, Y.; Chen, W.V.; Rother, J.; Tomlinson, G.E.; Boland, C.R.; Chaussabel, D.; Frazier, M.L.; Amos, C.I. Mutations of HNRNPA0 and WIF1 predispose members of a large family to multiple cancers. *Fam. Cancer* **2015**, *14*, 297–306. [CrossRef] [PubMed]

56. Zhang, J.X.; Fu, L.; de Voer, R.M.; Hahn, M.M.; Jin, P.; Lv, C.X.; Verwiel, E.T.; Ligtenberg, M.J.; Hoogerbrugge, N.; Kuiper, R.P.; et al. Candidate colorectal cancer predisposing gene variants in Chinese early-onset and familial cases. *World J. Gastroenterol.* **2015**, *21*, 4136–4149. [CrossRef] [PubMed]

57. Segui, N.; Mina, L.B.; Lazaro, C.; Sanz-Pamplona, R.; Pons, T.; Navarro, M.; Bellido, F.; Lopez-Doriga, A.; Valdes-Mas, R.; Pineda, M.; et al. Germline Mutations in FAN1 Cause Hereditary Colorectal Cancer by Impairing DNA Repair. *Gastroenterology* **2015**, *149*, 563–566. [CrossRef] [PubMed]

58. Ngeow, J.; Yu, W.; Yehia, L.; Niazi, F.; Chen, J.; Tang, X.; Heald, B.; Lei, J.; Romigh, T.; Tucker-Kellogg, L.; et al. Exome Sequencing Reveals Germline SMAD9 Mutation That Reduces Phosphatase and Tensin Homolog Expression and Is Associated With Hamartomatous Polyposis and Gastrointestinal Ganglioneuromas. *Gastroenterology* **2015**, *149*, 886–889.e5. [CrossRef]

59. Arora, S.; Yan, H.; Cho, I.; Fan, H.Y.; Luo, B.; Gai, X.; Bodian, D.L.; Vockley, J.G.; Zhou, Y.; Handorf, E.A.; et al. Genetic Variants That Predispose to DNA Double-Strand Breaks in Lymphocytes From a Subset of Patients With Familial Colorectal Carcinomas. *Gastroenterology* **2015**, *149*, 1872–1883.e9. [CrossRef]

60. Goldberg, Y.; Halpern, N.; Hubert, A.; Adler, S.N.; Cohen, S.; Plesser-Duvdevani, M.; Pappo, O.; Shaag, A.; Meiner, V. Mutated MCM9 is associated with predisposition to hereditary mixed polyposis and colorectal cancer in addition to primary ovarian failure. *Cancer Genet.* **2015**, *208*, 621–624. [CrossRef]

61. Rohlin, A.; Eiengard, F.; Lundstam, U.; Zagoras, T.; Nilsson, S.; Edsjo, A.; Pedersen, J.; Svensson, J.; Skullman, S.; Karlsson, B.G.; et al. GREM1 and POLE variants in hereditary colorectal cancer syndromes. *Genes, chromosomes & cancer* **2016**, *55*, 95–106. [CrossRef]

62. Spier, I.; Kerick, M.; Drichel, D.; Horpaopan, S.; Altmuller, J.; Laner, A.; Holzapfel, S.; Peters, S.; Adam, R.; Zhao, B.; et al. Exome sequencing identifies potential novel candidate genes in patients with unexplained colorectal adenomatous polyposis. *Fam. Cancer* **2016**, *15*, 281–288. [CrossRef] [PubMed]

63. Thutkawkorapin, J.; Picelli, S.; Kontham, V.; Liu, T.; Nilsson, D.; Lindblom, A. Exome sequencing in one family with gastric- and rectal cancer. *BMC Genet.* **2016**, *17*, 41. [CrossRef] [PubMed]

64. Esteban-Jurado, C.; Franch-Exposito, S.; Munoz, J.; Ocana, T.; Carballal, S.; Lopez-Ceron, M.; Cuatrecasas, M.; Vila-Casadesus, M.; Lozano, J.J.; Serra, E.; et al. The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer. *Eur. J. Hum. Genet.* **2016**, *24*, 1501–1505. [CrossRef] [PubMed]

65. Chubb, D.; Broderick, P.; Dobbins, S.E.; Frampton, M.; Kinnersley, B.; Penegar, S.; Price, A.; Ma, Y.P.; Sherborne, A.L.; Palles, C.; et al. Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. *Nat. Commun.* **2016**, *7*, 11883. [CrossRef]

66. Schubert, S.A.; Ruano, D.; Elsayed, F.A.; Boot, A.; Crobach, S.; Sarasqueta, A.F.; Wolffenbuttel, B.; van der Klauw, M.M.; Oosting, J.; Tops, C.M.; et al. Evidence for genetic association between chromosome 1q loci and predisposition to colorectal neoplasia. *Br. J. Cancer* **2017**, *117*, 1215–1223. [CrossRef]

67. Martin-Morales, L.; Feldman, M.; Vershinin, Z.; Garre, P.; Caldes, T.; Levy, D. SETD6 dominant negative mutation in familial colorectal cancer type X. *Hum. Mol. Genet.* **2017**, *26*, 4481–4493. [CrossRef]

68. Bellido, F.; Sowada, N.; Mur, P.; Lazaro, C.; Pons, T.; Valdes-Mas, R.; Pineda, M.; Aiza, G.; Iglesias, S.; Soto, J.L.; et al. Association Between Germline Mutations in BRF1, a Subunit of the RNA Polymerase III Transcription Complex, and Hereditary Colorectal Cancer. *Gastroenterology* **2018**, *154*, 181–194.e20. [CrossRef]

69. Franch-Exposito, S.; Esteban-Jurado, C.; Garre, P.; Quintanilla, I.; Duran-Sanchon, S.; Diaz-Gay, M.; Bonjoch, L.; Cuatrecasas, M.; Samper, E.; Munoz, J.; et al. Rare germline copy number variants in colorectal cancer predisposition characterized by exome sequencing analysis. *J. Genet. Genom.* **2018**, *45*, 41–45. [CrossRef]

70. Yu, L.; Yin, B.; Qu, K.; Li, J.; Jin, Q.; Liu, L.; Liu, C.; Zhu, Y.; Wang, Q.; Peng, X.; et al. Screening for susceptibility genes in hereditary non-polyposis colorectal cancer. *Oncol. Lett.* **2018**, *15*, 9413–9419. [CrossRef]

71. Thutkawkorapin, J.; Lindblom, A.; Tham, E. Exome sequencing in 51 early onset non-familial CRC cases. *Mol. Genet. Genom. Med.* **2019**, *7*, e605. [CrossRef]

72. Diaz-Gay, M.; Franch-Exposito, S.; Arnau-Collell, C.; Park, S.; Supek, F.; Munoz, J.; Bonjoch, L.; Gratacos-Mulleras, A.; Sanchez-Rojas, P.A.; Esteban-Jurado, C.; et al. Integrated Analysis of Germline and Tumor DNA Identifies New Candidate Genes Involved in Familial Colorectal Cancer. *Cancers* **2019**, *11*, 362. [CrossRef] [PubMed]

73. Toma, C.; Diaz-Gay, M.; Soares de Lima, Y.; Arnau-Collell, C.; Franch-Exposito, S.; Munoz, J.; Overs, B.; Bonjoch, L.; Carballal, S.; Ocana, T.; et al. Identification of a Novel Candidate Gene for Serrated Polyposis Syndrome Germline Predisposition by Performing Linkage Analysis Combined With Whole-Exome Sequencing. *Clin. Transl. Gastroenterol.* **2019**. [CrossRef] [PubMed]

74. Jansen, A.M.L.; Ghosh, P.; Dakal, T.C.; Slavin, T.P.; Boland, C.R.; Goel, A. Novel candidates in early-onset familial colorectal cancer. *Fam. Cancer* **2020**, *19*, 1–10. [CrossRef] [PubMed]

75. Toma, C.; Diaz-Gay, M.; Franch-Exposito, S.; Arnau-Collell, C.; Overs, B.; Munoz, J.; Bonjoch, L.; Soares de Lima, Y.; Ocana, T.; Cuatrecasas, M.; et al. Using linkage studies combined with whole-exome sequencing to identify novel candidate genes for familial colorectal cancer. *Int. J. Cancer* **2020**, *146*, 1568–1577. [CrossRef] [PubMed]

76. Bonjoch, L.; Franch-Exposito, S.; Garre, P.; Belhadj, S.; Munoz, J.; Arnau-Collell, C.; Diaz-Gay, M.; Gratacos-Mulleras, A.; Raimondi, G.; Esteban-Jurado, C.; et al. Germline mutations in FAF1 are associated with hereditary colorectal cancer. *Gastroenterology* **2020**. [CrossRef]

77. Chubb, D.; Broderick, P.; Frampton, M.; Kinnersley, B.; Sherborne, A.; Penegar, S.; Lloyd, A.; Ma, Y.P.; Dobbins, S.E.; Houlston, R.S. Genetic diagnosis of high-penetrance susceptibility for colorectal cancer (CRC) is achievable for a high proportion of familial CRC by exome sequencing. *J. Clin. Oncol.* **2015**, *33*, 426–432. [CrossRef]

78. Karczewski, K.J.; Francioli, L.C.; Tiao, G.; Cummings, B.B.; Alfoldi, J.; Wang, Q.; Collins, R.L.; Laricchia, K.M.; Ganna, A.; Birnbaum, D.P.; et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **2020**, *581*, 434–443. [CrossRef]

79. Weren, R.D.; Ligtenberg, M.J.; Geurts van Kessel, A.; De Voer, R.M.; Hoogerbrugge, N.; Kuiper, R.P. NTHL1 and MUTYH polyposis syndromes: Two sides of the same coin? *J. Pathol.* **2018**, *244*, 135–142. [CrossRef]

80. Ghosh, R.; Oak, N.; Plon, S.E. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol.* **2017**, *18*, 225. [CrossRef]

81. 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. [CrossRef]

82. J. Craig Venter Institute—PROVEAN. Available online: http://provean.jcvi.org/about.php (accessed on 12 August 2020).

83. Mahmood, K.; Jung, C.H.; Philip, G.; Georgeson, P.; Chung, J.; Pope, B.J.; Park, D.J. Variant effect prediction tools assessed using independent, functional assay-based datasets: Implications for discovery and diagnostics. *Hum. Genom.* **2017**, *11*, 10. [CrossRef] [PubMed]

84. Spier, I.; Drichel, D.; Kerick, M.; Kirfel, J.; Horpaopan, S.; Laner, A.; Holzapfel, S.; Peters, S.; Adam, R.; Zhao, B.; et al. Low-level APC mutational mosaicism is the underlying cause in a substantial fraction of unexplained colorectal adenomatous polyposis cases. *J. Med. Genet.* **2016**, *53*, 172–179. [CrossRef] [PubMed]

85. Parsons, M.T.; Buchanan, D.D.; Thompson, B.; Young, J.P.; Spurdle, A.B. Correlation of tumour BRAF mutations and MLH1 methylation with germline mismatch repair (MMR) gene mutation status: A literature review assessing utility of tumour features for MMR variant classification. *J. Med. Genet.* **2012**, *49*, 151–157. [CrossRef] [PubMed]

86. Alexandrov, L.B.; Kim, J.; Haradhvala, N.J.; Huang, M.N.; Tian Ng, A.W.; Wu, Y.; Boot, A.; Covington, K.R.; Gordenin, D.A.; Bergstrom, E.N.; et al. The repertoire of mutational signatures in human cancer. *Nature* **2020**, *578*, 94–101. [CrossRef] [PubMed]

87. Viel, A.; Bruselles, A.; Meccia, E.; Fornasarig, M.; Quaia, M.; Canzonieri, V.; Policicchio, E.; Urso, E.D.; Agostini, M.; Genuardi, M.; et al. A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer. *EBioMedicine* **2017**, *20*, 39–49. [CrossRef]

88. Pilati, C.; Shinde, J.; Alexandrov, L.B.; Assie, G.; Andre, T.; Helias-Rodzewicz, Z.; Ducoudray, R.; Le Corre, D.; Zucman-Rossi, J.; Emile, J.F.; et al. Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. *J. Pathol.* **2017**, *242*, 10–15. [CrossRef]

89. Grolleman, J.E.; de Voer, R.M.; Elsayed, F.A.; Nielsen, M.; Weren, R.D.A.; Palles, C.; Ligtenberg, M.J.L.; Vos, J.R.; Ten Broeke, S.W.; de Miranda, N.; et al. Mutational Signature Analysis Reveals NTHL1 Deficiency to Cause a Multi-tumor Phenotype. *Cancer Cell* **2019**, *35*, 256–266.e5. [CrossRef]

90. Cancer Genome Atlas Research, N.; Kandoth, C.; Schultz, N.; Cherniack, A.D.; Akbani, R.; Liu, Y.; Shen, H.; Robertson, A.G.; Pashtan, I.; Shen, R.; et al. Integrated genomic characterization of endometrial carcinoma. *Nature* **2013**, *497*, 67–73. [CrossRef]

91. Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **2012**, *487*, 330–337. [CrossRef]

92. Grolleman, J.E.; Diaz-Gay, M.; Franch-Exposito, S.; Castellvi-Bel, S.; de Voer, R.M. Somatic mutational signatures in polyposis and colorectal cancer. *Mol. Aspects Med.* **2019**, *69*, 62–72. [CrossRef]

93. Drost, J.; van Boxtel, R.; Blokzijl, F.; Mizutani, T.; Sasaki, N.; Sasselli, V.; de Ligt, J.; Behjati, S.; Grolleman, J.E.; van Wezel, T.; et al. Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science* **2017**, *358*, 234–238. [CrossRef] [PubMed]

94. Sato, T.; Stange, D.E.; Ferrante, M.; Vries, R.G.; Van Es, J.H.; Van den Brink, S.; Van Houdt, W.J.; Pronk, A.; Van Gorp, J.; Siersema, P.D.; et al. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **2011**, *141*, 1762–1772. [CrossRef] [PubMed]

95. Mur, P.; Sanchez-Cuartielles, E.; Ausso, S.; Aiza, G.; Valdes-Mas, R.; Pineda, M.; Navarro, M.; Brunet, J.; Urioste, M.; Lazaro, C.; et al. Scarce evidence of the causal role of germline mutations in UNC5C in hereditary colorectal cancer and polyposis. *Sci. Rep.* **2016**, *6*, 20697. [CrossRef] [PubMed]

96. Lelieveld, S.H.; Spielmann, M.; Mundlos, S.; Veltman, J.A.; Gilissen, C. Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. *Hum. Mutat.* **2015**, *36*, 815–822. [CrossRef]

97. Belkadi, A.; Bolze, A.; Itan, Y.; Cobat, A.; Vincent, Q.B.; Antipenko, A.; Shang, L.; Boisson, B.; Casanova, J.L.; Abel, L. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5473–5478. [CrossRef]

98. Lin, Y.; Lin, S.; Baxter, M.D.; Lin, L.; Kennedy, S.M.; Zhang, Z.; Goodfellow, P.J.; Chapman, W.C.; Davidson, N.O. Novel APC promoter and exon 1B deletion and allelic silencing in three mutation-negative classic familial adenomatous polyposis families. *Genome Med.* **2015**, *7*, 42. [CrossRef]

99. Teresi, R.E.; Zbuk, K.M.; Pezzolesi, M.G.; Waite, K.A.; Eng, C. Cowden syndrome-affected patients with PTEN promoter mutations demonstrate abnormal protein translation. *Am. J. Hum. Genet.* **2007**, *81*, 756–767. [CrossRef]

100. Snow, A.K.; Tuohy, T.M.; Sargent, N.R.; Smith, L.J.; Burt, R.W.; Neklason, D.W. APC promoter 1B deletion in seven American families with familial adenomatous polyposis. *Clin. Genet.* **2015**, *88*, 360–365. [CrossRef]

101. Pavicic, W.; Nieminen, T.T.; Gylling, A.; Pursiheimo, J.P.; Laiho, A.; Gyenesei, A.; Jarvinen, H.J.; Peltomaki, P. Promoter-specific alterations of APC are a rare cause for mutation-negative familial adenomatous polyposis. *Genes Chromosomes Cancer* **2014**, *53*, 857–864. [CrossRef]

102. Nieminen, T.T.; Pavicic, W.; Porkka, N.; Kankainen, M.; Jarvinen, H.J.; Lepisto, A.; Peltomaki, P. Pseudoexons provide a mechanism for allele-specific expression of APC in familial adenomatous polyposis. *Oncotarget* **2016**, *7*, 70685–70698. [CrossRef]

103. Spier, I.; Horpaopan, S.; Vogt, S.; Uhlhaas, S.; Morak, M.; Stienen, D.; Draaken, M.; Ludwig, M.; Holinski-Feder, E.; Nothen, M.M.; et al. Deep intronic APC mutations explain a substantial proportion of patients with familial or early-onset adenomatous polyposis. *Hum. Mutat.* **2012**, *33*, 1045–1050. [CrossRef] [PubMed]

104. Mantere, T.; Kersten, S.; Hoischen, A. Long-Read Sequencing Emerging in Medical Genetics. *Front. Genet.* **2019**, *10*, 426. [CrossRef] [PubMed]

105. Chan, S.; Lam, E.; Saghbini, M.; Bocklandt, S.; Hastie, A.; Cao, H.; Holmlin, E.; Borodkin, M. Structural Variation Detection and Analysis Using Bionano Optical Mapping. *Methods Mol. Biol.* **2018**, *1833*, 193–203. [CrossRef] [PubMed]

106. Aretz, S.; Stienen, D.; Friedrichs, N.; Stemmler, S.; Uhlhaas, S.; Rahner, N.; Propping, P.; Friedl, W. Somatic APC mosaicism: A frequent cause of familial adenomatous polyposis (FAP). *Hum. Mutat.* **2007**, *28*, 985–992. [CrossRef] [PubMed]

107. Hes, F.J.; Nielsen, M.; Bik, E.C.; Konvalinka, D.; Wijnen, J.T.; Bakker, E.; Vasen, H.F.; Breuning, M.H.; Tops, C.M. Somatic APC mosaicism: An underestimated cause of polyposis coli. *Gut* **2008**, *57*, 71–76. [CrossRef]

108. Ciavarella, M.; Miccoli, S.; Prossomariti, A.; Pippucci, T.; Bonora, E.; Buscherini, F.; Palombo, F.; Zuntini, R.; Balbi, T.; Ceccarelli, C.; et al. Somatic APC mosaicism and oligogenic inheritance in genetically unsolved colorectal adenomatous polyposis patients. *Eur. J. Hum. Genet.* **2018**, *26*, 387–395. [CrossRef]

109. Kim, B.; Won, D.; Jang, M.; Kim, H.; Choi, J.R.; Kim, T.I.; Lee, S.T. Next-generation sequencing with comprehensive bioinformatics analysis facilitates somatic mosaic APC gene mutation detection in patients with familial adenomatous polyposis. *BMC Med. Genom.* **2019**, *12*, 103. [CrossRef]

110. Acuna-Hidalgo, R.; Veltman, J.A.; Hoischen, A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biol.* **2016**, *17*, 241. [CrossRef]

111. Mavaddat, N.; Michailidou, K.; Dennis, J.; Lush, M.; Fachal, L.; Lee, A.; Tyrer, J.P.; Chen, T.H.; Wang, Q.; Bolla, M.K.; et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am. J. Hum. Genet.* **2019**, *104*, 21–34. [CrossRef]

112. Fritsche, L.G.; Gruber, S.B.; Wu, Z.; Schmidt, E.M.; Zawistowski, M.; Moser, S.E.; Blanc, V.M.; Brummett, C.M.; Kheterpal, S.; Abecasis, G.R.; et al. Association of Polygenic Risk Scores for Multiple Cancers in a Phenome-wide Study: Results from The Michigan Genomics Initiative. *Am. J. Hum. Genet.* **2018**, *102*, 1048–1061. [CrossRef]

113. Weigl, K.; Chang-Claude, J.; Knebel, P.; Hsu, L.; Hoffmeister, M.; Brenner, H. Strongly enhanced colorectal cancer risk stratification by combining family history and genetic risk score. *Clin. Epidemiol.* **2018**, *10*, 143–152. [CrossRef] [PubMed]

114. Fahed, A.C.; Wang, M.; Homburger, J.R.; Patel, A.P.; Bick, A.G.; Neben, C.L.; Lai, C.; Brockman, D.; Philippakis, A.; Ellinor, P.T.; et al. Polygenic background modifies penetrance of monogenic variants conferring risk for coronary artery disease, breast cancer, or colorectal cancer. *medRxiv* **2019**. [CrossRef]

115. Schlafly, A.; Pfeiffer, R.M.; Nagore, E.; Puig, S.; Calista, D.; Ghiorzo, P.; Menin, C.; Fargnoli, M.C.; Peris, K.; Song, L.; et al. Contribution of Common Genetic Variants to Familial Aggregation of Disease and Implications for Sequencing Studies. *PLoS Genet.* **2019**, *15*, e1008490. [CrossRef] [PubMed]