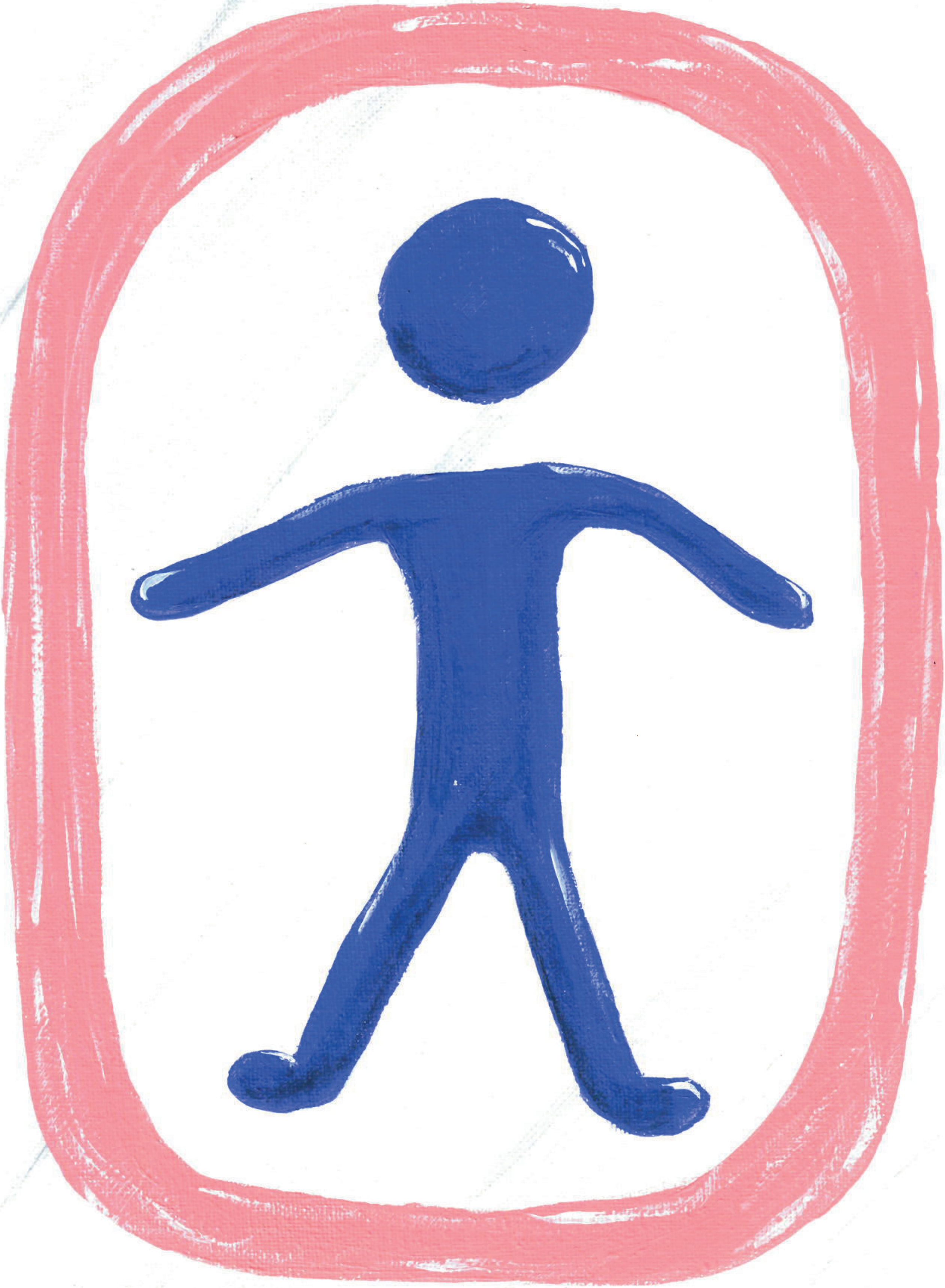
Identifying novel genetic causes of familial colorectal cancer

Schubert, S.A.

Citation

Schubert, S. A. (2025, December 9). *Identifying novel genetic causes of familial colorectal cancer*. Retrieved from https://hdl.handle.net/1887/4284678



Chapter 7

General discussion

General discussion

Familial and young-onset colorectal cancer still remains largely unexplained despite extensive efforts over the past decade to uncover novel genetic risk loci. Geneticists make a notable distinction between hereditary polyposis and non-polyposis predisposition syndromes, each seemingly rooted in distinct aetiologies. The hereditary non-polyposis colorectal cancers can be further divided into two main subgroups. The first subgroup is characterized by mismatch repair (MMR)-deficient and/or microsatellite unstable tumours, measured by the absence of MMR protein expression or the presence of the microsatellite instability-high (MSI-H) mutator phenotype. The second subgroup contains MMR-proficient or microsatellite stable (MSS) familial colorectal cancer, where severely affected families (fulfilling the Amsterdam Criteria) can also be referred to as Familial Colorectal Cancer type X.

This thesis focusses on the identification of underlying genetic causes in patients with unexplained MMR-proficient familial or young-onset colorectal cancers, as well as colorectal polyposis. The work presented in this thesis is based on the analysis of 100 whole-exome sequencing (WES) datasets from index patients and relatives, likely of carrying predisposing genetic variants to colorectal cancer or polyposis, but negative for causal variants identified in the diagnostic setting. We analysed 74 index patients: 25 index patients from MMR-proficient colorectal cancer families (plus 25 affected relatives and 1 unaffected spouse), 11 young-onset (<40 years) MMR-proficient colorectal cancer patients, 20 hyperplastic polyposis patients and 18 adenomatous polyposis patients. We performed initial filtering on variants affecting the coding sequence of genes and removed common variants in the general population (*i.e.,* with a population frequency above 1%), resulting in a total of 37,500 variants (unpublished data). Candidate variants were selected based on genes and pathways associated with cancer and, if multiple family members were included in the WES cohort, co-segregation of the variant with disease, resulting in just over 1,000 candidate coding or splice-site variants. Further prioritization was based on *in-silico* prediction tools, scientific literature and genotyping of family members as well as additional cohorts of local cases and controls.

Chapter 2 features a review article that provides a detailed overview of the known genetic causes of familial colorectal cancer, along with possible explanations for its missing heritability. The key findings from our WES analyses are presented in **Chapters 3 through 6**. In this chapter, we give an overview of the current challenges of genetic research using family-based approaches. We illustrate this with examples from our WES cohort, which is structured from most likely to explain the phenotype to more speculative candidate variants. Finally, we will discuss future perspectives at the end of this chapter.

1. Known causative genes for colorectal cancer and polyposis

Previously, germline screening of colorectal cancer and polyposis patients was performed in the diagnostic setting on leucocyte DNA using single genes or small gene panels containing selected colorectal cancer and polyposis genes. For our WES cohort, we selected colorectal cancer and polyposis patients who were reported to be negative for known pathogenic variants. Despite this, we identified known causative variants that may have been missed due to (i) atypical clinical presentation of the patients, which led to omission of the relevant gene from genetic screening, or (ii) detection limits of the sequencing method for certain regions within these known colorectal cancer genes. Furthermore, these germline screening methods do not adequately detect the presence of somatic mosaicism. Therefore, the source of DNA samples should also be carefully considered during genetic testing.

Known polyposis genes and APC mosaicism

We found a likely pathogenic variant in four adenomatous polyposis patients (Table 1, unpublished data). The first patient was a homozygous carrier of a common pathogenic *MUTYH* variant in northern European populations[1]. With the detection of more than 100 polyps, the clinical presentation of this patient was more consistent with familial adenomatous polyposis (FAP; OMIM#175100) than with MUTYH-associated polyposis (MAP; OMIM#608456)[2]. This phenotype may have influenced the choice of genes included in genetic testing, or analysis might have been performed before the *MUTYH* gene was routinely incorporated into diagnostic panels (prior to 2004), which resulted in the common pathogenic *MUTYH* variants not being screened for in this patient. In the second patient, we identified a predicted disruptive splice-site variant in *BMPR1A*. This patient presented with coecum cancer and more than 30 adenomas at 30 years of age. *BMPR1A* variants are associated with hamartomatous polyposis (juvenile polyposis syndrome; OMIM#174900), however more recently these variants have also been associated with unexplained adenomatous polyposis without juvenile polyps[3]. The absence of hamartomatous polyps in this patient, likely resulted in the omission of *BMPR1A* in the genetic screening. In the third patient, we found a deletion-insertion resulting in an *APC* nonsense variant. The clinical presentation of this patient was consistent with FAP[2]. The selected sequencing method, *i.e.,* Sanger sequencing, might not have been able to detect this novel complex genetic variant, which we identified in the WES analysis. Finally, in the fourth patient a mosaic *APC* variant was identified in a parallel study from our group, described in Jansen *et al.*[4] (patient APC-18). In this study a sensitive DNA screening method, *i.e.,* KASPar assay, found this variant in a small fraction of the leucocyte DNA (3%). This variant was not detected in our WES analysis.

2. Novel regions in known cancer genes

In addition to gene selection, regions within genes should be carefully considered during genetic testing. Current clinical genetic screening approaches mainly focus on variants affecting the canonical protein isoform, excluding alternatively spliced exons or non-coding regions, because the clinical effect of these variants is uncertain and more difficult to predict. Expanding the analysis to include these regions will likely reveal additional pathogenic variants. Functional characterisation of these candidates is therefore essential, but requires extensive efforts and close collaborations between clinicians, translational researchers, and fundamental researchers to combine the expertise needed for accurate pathogenicity assessment.

Table 1. Explained cases

Patient ID	Phenotype	Inheritance	Gene[a]	Nucleotide variant	Protein variant
DDPOL6354	CRC33; >100 polyps (AD, some HP)	Sporadic	*MUTYH* (homozygous)	c.536A>G	p.Tyr179Cys
DDPOL6101	CRC30; 30 polyps (AD)	Sporadic	*BMPR1A*[b]	c.431-2A>C, r.431_449del	p.(Gly144Alafs*15)
DDPOL6356	CUP40; >100 polyps (AD; carpeted colon)	Sporadic	*APC*	c.394_399delins TGAGAAAGTAC	p.(Gly132*)
DDPOL6357	80 polyps (AD, some HP)	Sporadic	*APC*[c]	c.2493dupA	p.(Pro832Thrfs*12)

Abbreviations: AD, adenomas; CRC, colorectal cancer; CUP, cancer of unknown primary site; HP, hyperplastic polyps.

[a] Likely pathogenic variants were identified in *MUTYH* (NM_001128425.1), *BMPR1A* (NM_004329.2) and *APC* (NM_000038.5).

[b] The *BMPR1A* variant was predicted to disrupts the splice acceptor site (based dbscSNV[5] scores).

[c] The mosaic *APC* variant was identified with a 3% variant frequency in leucocyte DNA using KASPar Assay (described by Jansen *et al.*[4]).

*TP53*β-stop-lost

The complexity of studying genetic variants that only affect non-canonical protein isoforms is illustrated in **Chapter 4**. We identified a stop-lost variant (p.*342Serext*17) that specifically affects p53β isoforms of the *TP53* gene in four families suspected of an autosomal dominant cancer syndrome involving colorectal, breast and papillary thyroid cancers. The *TP53* gene encodes three distinct C-terminal protein isoforms (α, β and γ). The C-termini of the p53β and p53γ isoforms are encoded by two cryptic exons, located in intron-9, which contain premature stop codons[6]. Pathogenic variants in *TP53* that impact all isoforms or specifically alter the C-terminus of p53α isoforms can cause Li-Fraumeni syndrome (OMIM#151623), a rare hereditary disorder with an increased risk of a various young-onset cancers[7]. The precise function of wildtype p53β isoforms remains largely elusive, probably because its intrinsic activity is dependent on cell type, genetic background and cell signals[6]. Clinically, overexpression of p53β isoforms has been associated with a better prognosis in some cancers[8-10] and a poorer outcome in others[11-14], suggesting a modulatory role in tumorigenesis.

We characterize this stop-lost variant, which leads to a 17-amino acid extension of the p53β isoforms. We detected expression of multiple p53β-stop-lost isoforms at the protein level in a patient-derived cell line. Functional assays of the full-length isoform showed increased oligomerization with canonical p53α and dysregulation of p53 transcriptional targets in different cell line models. Together, our findings support a model of pathogenicity of this variant. Future research should investigate the exact biological consequences and functional properties of each of the p53β-stop-lost isoforms in different cell types and cell contexts to further elucidate the mechanisms leading to cancer predisposition in *TP53β*-stop-lost carriers. Additionally, extended clinical follow-up of the initial four families, along with the identification of additional carriers, will be essential to refine cancer risks estimates and guide appropriate clinical management.

3. Novel colorectal cancer candidate genes

One of the challenges of genetic research using family-based approaches is obtaining a sufficient number of cases to detect novel and extremely rare recurrently mutated genes. Due to national and international collaborations and the pooling of many unexplained familial and young-onset colorectal cancer cases, it is, however, possible to validate rare variants in novel genes. Additionally, as with the identification of novel cancer predisposing variants outside the canonical protein-coding sequence of known cancer genes, functional evidence is essential to confirm the involvement of variants in these new genes in cancer predisposition. Our WES cohort contributed to the discovery of two novel candidate genes associated with MMR-proficient non-polyposis colorectal cancer predisposition.

NPAT

In **Chapter 5** we describe two families for which loss-of-function *NPAT* variants could explain the colorectal cancer predisposition: one family from our WES cohort (NA41) and one Spanish family. The NPAT protein is involved in cell cycle regulation, more specifically in the G1/S phase transition[15]. In addition, NPAT contributes to the activation of the promoter of *ATM* and is therefore considered part of the DNA damage response pathway[16]. Loss of heterozygosity (LOH) at its locus has been found in DNA isolated from colorectal tumours, which also targets *ATM* in a malignant context[17]. In one family with familial Hodgkin lymphoma a loss-of-function variant in *NPAT* was previously described and a germline variant, resulting in an one amino acid deletion (p.Ser726del), was associated with an elevated risk of Hodgkin lymphoma (odds ratio = 4.11 [95% confidence interval, 1.27-13.35, P=0.018])[18].

We screened for loss-of-function and predicted deleterious *NPAT* variants in cohorts consisting of Spanish and Dutch familial/young-onset colorectal cancer patients as well as young-onset colorectal cancer patients in the CanVar database. We found enrichment of damaging and predicted damaging *NPAT* variants in the colorectal cancer patients compared to non-cancer individuals in the gnomAD database (odds ratio = 7.2 [95% confidence interval, 2.16-19.10, P=0.0011]). Based on these screening cohorts, we show that heterozygous loss-of-function and predicted deleterious variants could explain up to 0.32% of the familial and/or young-onset colorectal cancers. Previous studies in Drosophila models showed that *NPAT* abrogation results in chromosomal instability, increase of double strand breaks, and induction of tumour formation[19-23]. In line with these results, colorectal cancers with *NPAT* somatic variants, in the absence of additional DNA repair defects, have significantly higher aneuploidy levels than *NPAT*-wildtype colorectal cancers in the TCGA database. This would suggest that *NPAT* deficiency might lead to high chromosomal instability in tumours. Future clinical studies are, however, needed to show the pathogenicity of loss-of-function *NPAT* variants in colorectal cancer predisposition.

MCM9

Genetic variants in minichromosome maintance genes *MCM8* and *MCM9* may predispose carriers to infertility and cancer[24]. These proteins are implicated in multiple DNA-related processes, including homologous recombination, MMR and DNA replication[24]. Homozygous *MCM9* loss-of-function variants have been identified in female carriers with primary ovarian insufficiency (POI) and young-onset colorectal cancer or cervical cancer, in the presence and apparent absence of a mixed polyposis phenotype[25,26]. One male heterozygous carrier was diagnosed with young-onset MMR-proficient colorectal cancer, recurrent polyps and infertility, indicating a possible increased risk of infertility and cancer in heterozygous carriers.

Within our WES cohort, we found two family with carriers of compound heterozygous *MCM9* missense variants (Figure 1), as described in Golubicki *et al.*[27], which is not included as a separate chapter in this thesis. All four variants were classified as VUS in the Franklin by Genoox tool. In family 69294, a female carrier was diagnosed with POI, mixed polyposis and MMR-deficient colorectal cancer, due to a somatic *MSH2* mutation (c.2470C>T, VAF 38%). A male carrier was diagnosed with mixed polyposis and MMR-proficient colorectal cancer. The clinical presentation of the index patient with POI, colorectal cancer and mixed polyposis supports the pathogenicity of these variants. In family MMS13, we identified one female compound heterozygous carrier, who presented with MMR-proficient colorectal cancer, but without (mixed) polyposis or apparent fertility problems. Although, these missense variants were predicted to be likely deleterious (based on CADD scores >20), functional studies are needed to show their pathogenicity. Alternatively, somatic mutation analysis of tumours from carriers could provide supportive evidence of MCM9 deficiency by identifying signs of its impaired DNA repair capacity, such as MSI (specifically enrichment of tri- or tetranucleotide frameshifts) or mutational scares associated with homologous recombination deficiency[24].

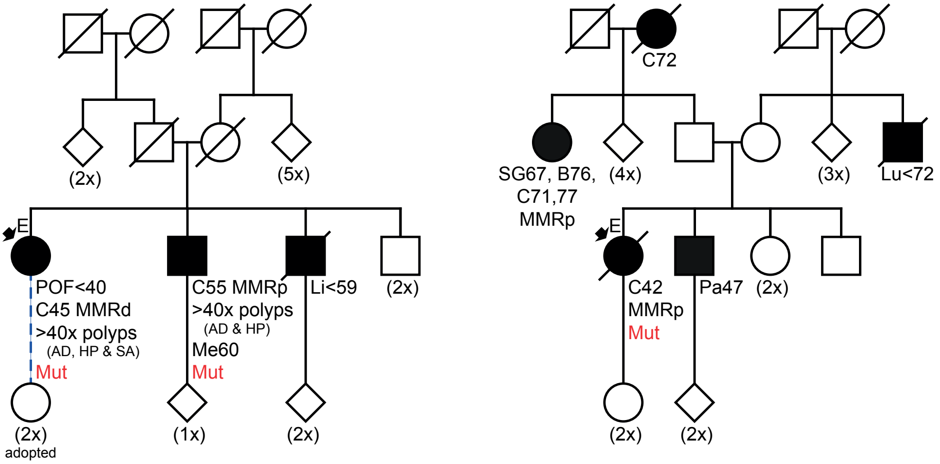


Figure 1. Pedigrees of biallelic carriers of MCM9 variants.
Compound heterozygous carriers of *MCM9* (NM_017696.2) variants c.152A>T,p.(Asn51Ile) and c.1642C>T,p.(Arg548Trp) were identified in family 69294 (left) and c.1640T>C,p.(Leu547Pro) and c.3425A>G,p.(Lys1142Arg) were identified in family MSS13 (right). All variants were classified as VUS in Franklin by Genoox (https://franklin.genoox.com, accessed on 18 April 2025). Index patients (arrow), patients selected for WES (E) and biallelic carriers (red Mut) are indicated. Abbreviations: AD, adenomas; C, colorectal cancer; B, breast cancer; HP, hyperplastic polyps; Li, liver cancer; Lu, lung cancer; Me, melanoma; MMRd, MMR-deficient; MMRp; MMR-proficient; Pa, pancreatic cancer; POF, premature ovarian failure; SA, serrated adenomas; SG, salivary gland carcinoma. Figure is adapted from Golubicki *et al.*[27]

4. Proposed colorectal cancer genes

In the past decade, extensive genetic screens have been conducted to discover novel genetic causes for colorectal cancer predisposition, leading to the identification of more than 100 candidate genes[28]. However, a problem with family-based genetic screening approaches is the reproducibility in validation cohorts. For example, Chubbs *et al.*[29] analysed over 1,500 young-onset familial colorectal cancer cases but was unable to validate any of the previously identified candidate genes. In addition, we combined adenomatous polyposis patients (>10 adenomas) from our WES cohort with similar cases collected in Nijmegen and Oxford in collaboration with Richarda de Voer, Claire Palles and colleagues, totalling 107 patients. Joined analysis of these patients was unable to identify recurrent variants in novel candidate genes (unpublished data). This analysis indicates that large study populations are crucial to identify novel high-risk susceptibility genes that are recurrently mutated across unrelated populations, even if patients have been selected for a homogenous phenotype. This project also illustrated challenges with combining datasets from multiple research groups, including different capture kits and analysis pipelines.

Fanconi anaemia pathway

Many of the causal genes associated with hereditary cancer and polyposis syndromes are involved in DNA repair pathways[28]. This has lead researchers, including our group, to examine these genes for potential candidate variants in non-polyposis colorectal cancer. One such pathway is the Fanconi anaemia (FA) pathway, which plays a central role in the repair of DNA double-stranded breaks and interstrand cross-links[30]. Fanconi anaemia is a rare autosomal or X-chromosomal recessive genetic disease, characterized by short stature, bone marrow failure and predisposition to cancer. The FA pathway includes 22 core FA proteins, as well as several FA-associated proteins, such FAAP20/24/100, MHF1/2, and interacting partners like FAN1[30]. Notably, heterozygous pathogenic variants in at least five FA genes, *BRCA1*, *BRCA2*, *PALB2*, *BRIP1* and *RAD51C,* are known hereditary breast cancer genes (OMIM#11448)[31]. In addition, pathogenic variants in *BRCA1*, *BRCA2* and *PALB2* have been linked to an increased risk for colorectal cancer, although the evidence remains contradictory across studies[32-34]. Moreover, candidate variants in *FAN1*, *BRIP1*, *FANCC*, *FANCD2* and *FANCE* have also been reported in familial colorectal cancer cases[35-39].

Within our WES cohort, we identified 54 variants of unknown significance (VUS) across 22 FA genes and 4 FA-associated genes (Figure 2). These included 45 missense variants, 3 nonsense variants, 3 frameshifts, 1 in-frame deletion and 2 splice site variants (unpublished data). Of note, we observed 3 missense variants in *BRCA1*, 3 missense variants and 1 nonsense variant in *BRCA2,* and 2 missense variants in *FAN1*. Although these findings include potentially interesting candidate variants for cancer predisposition, this gene list represents a highly selective subset, some of which lack prior evidence linking them to colorectal cancer risk. Further functional and clinical validation is needed to determine whether these variants warrant consideration in the clinical management of carriers.

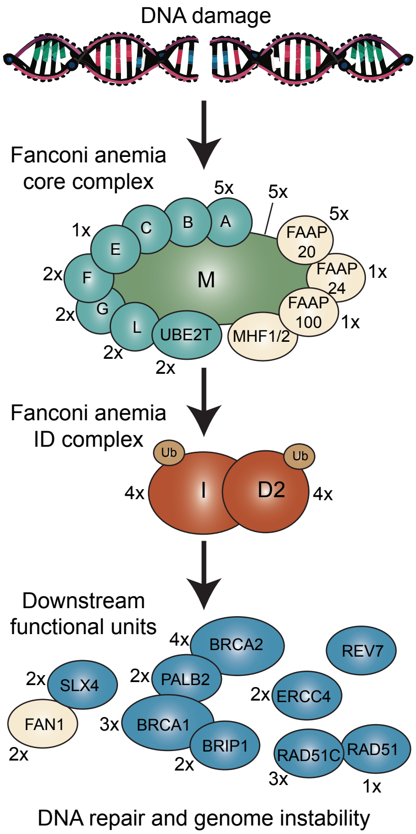


Figure 2. 54 variants of unknown significance in FA genes. Rare protein-coding or splice site variants were investigated 22 FA proteins and 6 FA-associated genes in our WES cohort containing familial colorectal cancer patients, young-onset colorectal cancer patients and polyposis patients. The figure was adapted from Che *et al.*[30]

5. Alternative inheritance patterns

An alternative mode of inheritance to the classical dominant or recessive diseases could also result in apparent familial aggregation of disease. With the increased application of whole-genome sequencing (WGS), WES and large multigene panel testing in clinical genetic counselling, it is likely that more individuals will be identified who carry multiple (likely) pathogenic variants in different cancer-associated genes. The interaction of these variants can lead to the appearance of a unique cancer predisposition syndromes.

Digenic inheritance

Digenic inheritance, where the genotype at two loci is required to explains the phenotype, has been proposed in several hereditary conditions, most notably retinitis pigmentosa, a progressive retinal degenerative disease[40]. For colorectal cancer predisposition, digenic inheritance of monoallelic *MSH6* VUSs and *MUTYH* variants has been suggested[41,42]. Co-segregation with disease in the familial setting was established in one colorectal cancer family from our WES cohort (NA15), described in **Chapter 3**. The involvement of both MSH6 and MUTYH in oxidative DNA damage repair and their physical interaction enhancing MUTYH's repair activity, substantiates the association of variants in these genes and cancer risk[43]. Functional evidence supporting the pathogenicity of combined defects in these oxidative DNA damage repair genes is still lacking, however future research could measure accumulation of oxidised DNA lesions (specifically 8-oxoG) in research models or patient tissues and investigate the physical interaction of the two mutant proteins by co-immunoprecipitation (Co-IP).

Digenic inheritance was further proposed in a young-onset CRC patient carrying heterozygous germline variants in two genes involved in the repair of oxidised DNA lesions (*MUTYH* and *OGG1*)[44]. In addition, digenic carriers of nonsense variants in *BRCA1* and *RNF43* were identified in a large cancer-prone family affected by multiple cancer, including colorectal cancer, metastatic cancer of liver, breast, ovarian, prostate and serrated polyposis.[45] Heterozygous pathogenic mutations in the *RNF43* gene (OMIM#617108) have been attributed to serrated polyposis syndrome and CRC in very rare cases.[46-48]. Taken together, digenic inheritance should be considered in familial and young-onset colorectal cancer patients.

Complex patterns of inheritance

Another example of a family suggested of a complex pattern of inheritance from our WES cohort is family V (unpublished data). The pedigree of the family exhibits a diverse array of tumours across multiple branches of the ancestry (Figure 3). WES identified three potential candidate missense variants in *STAT3*, *REV1* and *WRAP53*. STAT3 is a transcription factor and a multifunctional regulator in tumorigenesis[49]. Loss of STAT3 has been shown to promote adenoma-to-carcinoma transition in colorectal cancer mouse models[50]. REV1 is a translesion synthesis and FA-associated protein, and functions as a scaffolding protein, but also has its own DNA polymerase activity. *In vitro* assays have shown that the REV1 variant identified in family V has reduced catalytic function on damaged DNA, including G-quadruplex (G4) motifs, which are caused by chemical and viral carcinogens[51]. *C. elegans* defective in human *REV1* showed a decrease in mutational signatures resulting from error-prone DNA repair of induced DNA damage[52]. WRAP53 is an essential component of the telomerase holoenzyme complex, and its mRNA also functions as a p53 antisense transcript[53,54]. Single nucleotide polymorphisms in this gene have been linked to an increased risk of breast and ovarian cancer[55-57].

The two siblings affected by young-onset colorectal cancer (IV-2) and young-onset breast cancer (IV-3), shared only the *STAT3* variant. In addition, patient IV-3 carriers the WRAP53 variant. On the other hand, the *REV1* variant was shared by IV-2 and III-4, the latter was affected by colorectal cancer at age 49 years. To further investigate the *REV1* variant as a potential colorectal cancer risk factor, we performed tumour analysis of the cancer of IV-2, using WES and bulk RNA-sequencing. No second hit in *REV1* was identified, and mutational signature analysis did not reveal changes in mutational signatures related to REV1. Future research could utilize whole-genome sequencing of tumour DNA to elucidate the pathogenicity of the *REV1* variant, as this approach allows for better investigation of chromosomal instability and *de novo* mutational signature discovery than WES. Potential DNA repair deficiencies could be validated by reproduction in genetically engineered cells. With the introduction of CRISPR-Cas9 gene editing tools, it has become easier to generate cells and reproduce mutational signatures in model systems, as previously illustrated for POLE- and NTHL1-deficiency[58,59].

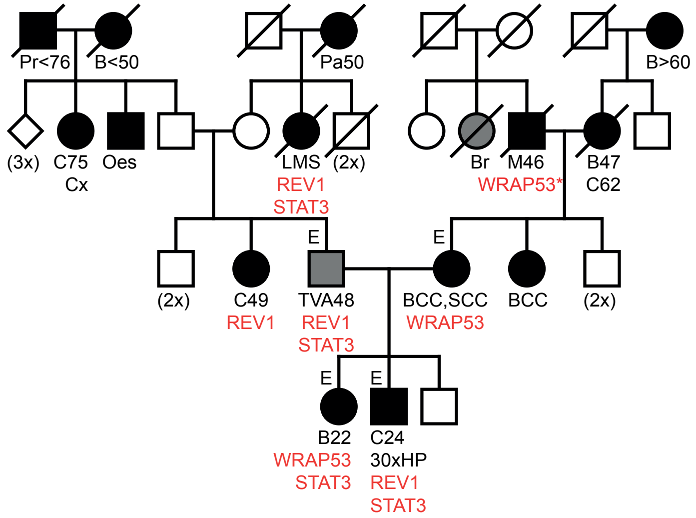


Figure 3. Pedigree of family likely showing complex pattern of inheritance.
Family V is highly affected by multiple cancer types (black fill) and benign tumours (grey fill). Segregation of three candidate genes is shown. WES identified three potential candidate variants: *STAT3* NM_139276: c.2221G>C,p.(Gly741Arg), *REV1* NM_001037872/NM_016316: c.1301G>A,p.(Arg434Gln) and *WRAP53* NM_018081.2: c.920G>A,p.(Arg307Gln). Patient selected for WES (E) and obligate carriers (asterisks) are indicated. Cancers in the first and second generation are anamnestically reported. Abbreviations: B, breast cancer; BBC, basal cell carcinoma; Br, benign brain cancer; C, colorectal cancer; Cx, cervical cancer; HP, hyperplastsic polyps; LMS, leiomyosarcoma; M, melanoma; Oes, oesophageal cancer; Pr, prostate cancer; Pa, pancreatic cancer; TVA, tubullovilleus adenoma.

Genetic modifiers

Thus far, the combined inheritance of multiple common low-risk alleles has been suggested to attribute to up to 15% of the familial risk of colorectal cancer[60]. Also, the enrichment of common low-risk loci in familial colorectal cancer cases has previously been shown by our group[61]. However, less research has been performed on the identification and co-inheritance of novel genetic modifiers using family-based approaches. For example, in one family from our WES cohort, the Dutch NPAT family (NA41) described in **Chapter 5**, co-inheritance was observed between a loss-of-function *NPAT* variant and a rare missense variant, encoding PPP2R1B p.(R410H), both located on chromosome 11q22.3-q23.1 (unpublished data). Previous genetic linkage analysis in this family had not revealed any linkage regions (LOD>2.0), nor enrichment of common risk variants[62]. While the *NPAT* nonsense variant remains the most promising candidate, a modifying effect of the *PPP2R1B* variant on the phenotype cannot be excluded based on co-segregation alone. *PPP2R1B* encodes a subunit of the protein phosphatase 2A (PP2A) complex, a central protein in the cells' dephosphorylation pathways, and is frequently mutated (15%) in colorectal cancers[63]. *In-silico* prediction tools provide conflicting assessments of the *PPP2R1B* p.(R410H) variant: while its REVEL score falls below the threshold for predicted pathogenicity (0.098; REVEL > 0.35), it is predicted to be likely deleterious based on its CADD score (25.6; threshold >20). Functional validation, such as protein stability or binding assays[64], will be necessary to assess the pathogenic potential of this missense variant, as well as its potential interaction with the *NPAT* variant in modulating cancer risk within this family.

6. Technical limitations of whole-exome sequencing approaches

Various next-generation sequencing strategies are being employed to analyse germline DNA in the search for genetic variants associated with disease. In this thesis we have focussed on WES, which targets approximately 1% of the genome, encompassing the protein-coding regions and their flanking sequences. Pathogenic germline variants likely remain undetected in many familial colorectal cancer cases using WES, as this method only captures a small fraction of the genome and does not assess genetic aberrations affecting promoter regions, microRNAs and epigenetic regulation, such as DNA methylation. As with missed causative genetic variants in know cancer genes (as discussed in section **1. Known causative genes for colorectal cancer and polyposis**), technical limitations of sequencing can also hamper the identification of causative variants in novel genes. Specifically, capture and sequencing technologies can have a profound effect on the exome coverage, hindering the identification of variants in protein-coding regions of certain genes.

Limitations in DNA capture

An example from our WES cohort where capture and sequencing technologies had a profound effect on the identification of potential causative variants in novel genes is family NA96 (unpublished data). For this family we performed WES on leucocyte DNA from three individuals (Figure 4): an index patient diagnosed MMR-proficient colorectal cancer at age 25 years (V-1), a relative diagnosed with multiple polyps at age 44 (IV-4) and an unaffected spouse, selected to infer the genotype of a patient diagnosed with colorectal cancer at age 48 and 68 (III-3). WES and data processing (*i.e.,* quality control, alignment, and variant calling) were performed by BGI sequencing science (BGI-Shenzhen, Shenzhen, China) using the BGI capture kit followed by sequencing on the Complete Genomics' Sequencing Platform[65]. We identified seven promising heterozygous candidate variants (in *BLMH*, *MTAP*, *SH3GLB1*, *NPAT*, *ERCC3*, *MCM9* and *RNF43*) and performed co-segregation analysis in a 6th-degree relative of the index patient, who was also affected by colorectal cancer at age 57 (III-1). Genotyping, however, did not show this relative to carry any of these variants, which weakened their association with disease in this family.

Subsequently, sequencing efforts within the diagnostic setting of this family, selecting individuals V-1 and III-1 for WES, identified the *RPL22L1* p.(Glu119Valfs*61) frameshift variant in both affected relatives. This gene encodes a 123 amino acids protein paralog of ribosomal protein L22E, a protein subunit of the 60S ribosomal subunit. Overexpression of this protein in colorectal cancers has been associated with poor prognosis and chemotherapy resistance[66]. In addition, the ribosomal gene *RPS20*, which encoding a component of the 40S ribosomal subunit, is currently the only gene uniquely linked to MMR-proficient non-polyposis familial colorectal cancer[67,68], suggesting that other ribosomal genes may also play a role in predisposition to this disease. This *RPL22L1* frameshift variant has also been reported in a 21-year-old patient with colorectal in the COSMIC database (v99; 1/155 tested samples), even though it is very rare in the general population, with a population frequency of 0.002% in gnomAD (v.4.0.0; 17/802,420 individuals). To investigate co-segregation of this variant with cancer predisposition within family NA96, we re-examined the BGI WES data. Visual inspection of this regions indicated the presence of this variant in V-1 and IV-4, and absence in the unaffected spouse of III-3 (Figure 4). This variant, however, did not pass BGI's quality control threshold for variant calling, likely because of low coverage of this region (~20X). Next, we undertook multiple efforts to perform genotyping of additional family members, nevertheless all Sanger sequencing attempts to amplify this region failed due to a notable number of AT-rich repeats and low-complexity regions in the surrounding DNA sequence. Taken together, and despite the need for further research to show the functional effect of the 61 amino acids extension, it can be speculated that this *RPL22L1* frameshift variant severely effects its protein function and, therefore, could potentially predispose to colorectal cancer.

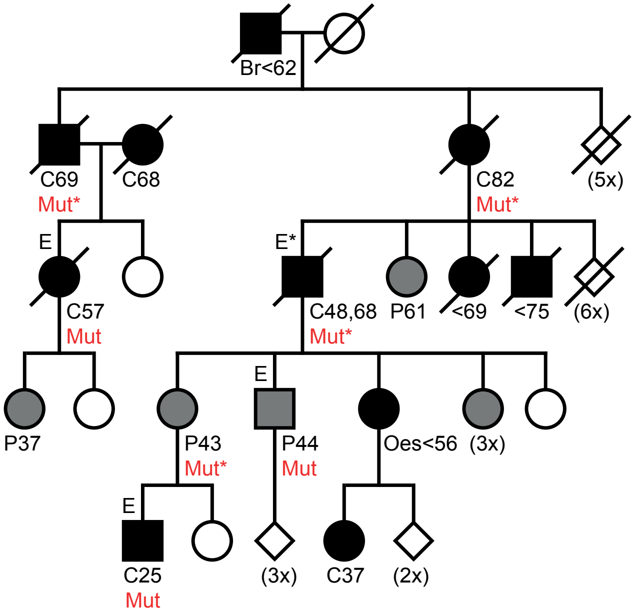


Figure 4. Pedigree of family NA96.
Family is highly affected by cancers (black fill) and colorectal polyps (grey fill). Patient selected for WES (E), *RPL22L1* frameshift variant (NM_001099645.1:c356_357delAG,p.(Glu119Valfs*61)) carriers (Mut) and obligate carriers (asterisks) are indicated. Two relatives (III-5,III-6) have cancer of unknown origin. Abbreviations: Br, brain cancer; C, colorectal cancer; Oes, oesophageal cancer; P, polyps.

Rare intergenic or structural variants

In one of the families from our WES cohort, family NA68 discussed in **Chapter 6**, no likely candidate variants were found using a combination of WES and re-analysis of genetic linkage data from a previous study by our group[62]. To investigate the inheritance of non-coding variants WGS was performed on three individuals. Filtering of genetic variants within the linkage peak region at the chromosome 1q32.2-q42.2 locus produced a list of over 450 rare, heterozygous variants. Based on our current knowledge of this region, none could be linked to a functional effect on microRNAs or other regulatory sequences and thus could not be associated with an increased risk of (colorectal) cancer. In addition, no promising rare structural variants were identified. Hopefully, with the increase application of WGS in germline and tumour genetics more complete reference datasets will be available and improved analysis tools will be developed. This will aid the identification of rare intergenic and structural variants associated with familial colorectal cancer in future studies. Although no candidate variants were identified in family NA68, the study did point towards *MIA3* (*TANGO*) as a player in colorectal cancer tumorigenesis, showing that even these “negative” studies can contribute to our understanding of cancer aetiology.

7. Future perspectives

Around the world, the incidence of young-onset colorectal cancer is rising[69]. For most patients, underlying genetic causes remain largely unknown, however, studies have highlighted the importance of various established lifestyle and environmental risk factors[69]. Early and accurate identification of individuals at risk for colorectal cancer is essential for both prevention and early detection. Notably, 6-30% of the young-onset patients meet criteria for familial colorectal cancer, and in more than 50% of these cases, the disease could potentially have been prevented or diagnosed earlier if appropriate surveillance had been in place[70]. A key step towards addressing this is the improved detection of pathogenic variants in cancer patients and subsequent inclusion of at-risk relatives in surveillance programs.

Several possible explanations for the missing heritability of familial colorectal cancer are discussed in our review article in **Chapter 2**, including contribution from common colorectal cancer low-risk loci and non-genetic risk factors. The following section introduces two research lines that may further expand our understanding of colorectal cancer aetiology and help explain additional unresolved cases.

Tumour-based multi-omics approaches

With the increasing implementation of WES and WGS in the molecular diagnostics of tumours, tumour-focussed germline predisposition screening strategies are identifying a growing number of carriers of known pathogenic cancer variants[71]. This approach is expected to broaden the tumour spectra of known hereditary cancer syndromes, which had previously been constrained by germline screening using multigene panels selected based on personal or family medical history of the index patient.

To further elucidate the unexplained heritability of colorectal cancer, integrated multi-omics analyses of tumours will likely be required. By combining genomics, epigenomics, transcriptomics and proteomics, these approaches can provide a more comprehensive overview of the molecular landscape of colorectal cancers. Techniques such as tumour mutational signature analysis, RNA-sequencing, ATAC-seq (assay for transposase-accessible chromatin using sequencing), and metabolomics profiling hold great promise in uncovering novel biomarkers and mechanisms of tumorigenesis. Such insights may enable improved risk stratification, personalized treatment plans, and a deeper understanding of cancer susceptibility in general.

Gene-environment interactions

Not all cancer susceptibility observed in the families from our WES cohort appear to be fully explained by the most promising candidate variant. For example, as described in

Chapter 4, seven cancer patients did not carry the identified *TP53*β-stop-lost variant. In addition to the potential involvement of other inherited cancer-risk alleles (as discussed in section **5. Alternative inheritance patterns**), shared environmental risk factors may also contribute to the familial cancer risk. Although, large-scale studies are beginning to explore the interaction between common genetic low-risk loci and environmental risk factors in colorectal cancer[72], more research using family-based approaches is needed. Thus far, the microbiome has been implicated in tumorigenesis in hereditary colorectal cancer patients, including carriers of biallelic *NTHL1* and *MUTYH* variants[73]. Furthermore, in Lynch Syndrome a few studies have investigated environmental risk factors, such as microbiome[74] and aspirin use[75]. Notably, no lifestyle and environmental data was collected for the patients in our WES cohort. We therefore believe that future research should place greater emphasis on the interplay between genetic and environmental risk factors in familial colorectal cancer predisposition.

Concluding remarks

The work presented in this thesis illustrated that WES approaches on leucocyte DNA can help elucidate the genetic cause of hereditary disease. Of the in total 37,500 rare variants identified in our WES cohort over 1,000 candidate coding or splice-site variants were prioritized based on our current understanding of tumorigenesis. Many of these variants, however, still remain classified as VUS and need to be investigated in future functional and clinical studies before they can be considered in patient care. It is also important to note that individuals in this study were referred for clinical genetic counselling, making it more likely that our cohort includes severely affected patients and families, while mildly affected or unaffected relatives are underrepresented. As a result, the risk estimates presented may suffer from bias, potentially leading to an overestimation of the true cancer risks.

Of the 74 index patients in our WES cohort affected by unexplained colorectal cancer and polyposis, 5% could be attributed to known mechanisms of colorectal cancer predisposition. This comprised of pathogenic variants in known polyposis genes and mosaic APC mutations (as discussed in section **1. Known causative genes for colorectal cancer and polyposis**). In addition, novel candidate variants likely associated with disease were identified in 8% of the index patients (as discussed in sections **2. Novel regions in known cancer genes** and **3. Novel colorectal cancer candidate genes**). We are optimistic that, as our understanding of the aetiology of colorectal cancer advances and additional genetic and environmental risk factors are uncovered, more patients from our WES cohort can be attributed to specific underlying causes. Moreover, the remaining unexplained cases represent an invaluable resource for validating novel candidate variants linked to hereditary colorectal cancer. Altogether, this work sought to identify new genetic variants involved in colorectal cancer predisposition and provides a foundation for future research into the underlying tumorigenic mechanisms and refined risk assessment, with the ultimate goal of improving prevention strategies for variant carriers.

References

1. Al-Tassan, N. et al. Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. Nat Genet 30, 227-32 (2002).

2. Jasperson, K.W., Tuohy, T.M., Neklason, D.W. & Burt, R.W. Hereditary and familial colon cancer. Gastroenterology 138, 2044-58 (2010).

3. Rosner, G. et al. Adenomatous Polyposis Phenotype in BMPR1A and SMAD4 Variant Carriers. Clin Transl Gastroenterol 13, e00527 (2022).

4. Jansen, A.M. et al. Distinct Patterns of Somatic Mosaicism in the APC Gene in Neoplasms From Patients With Unexplained Adenomatous Polyposis. Gastroenterology 152, 546-549.e3 (2017).

5. Jian, X., Boerwinkle, E. & Liu, X. In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res 42, 13534-44 (2014).

6. Joruiz, S.M. & Bourdon, J.C. p53 Isoforms: Key Regulators of the Cell Fate Decision. Cold Spring Harb Perspect Med 6(2016).

7. Bougeard, G. et al. Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers. J Clin Oncol 33, 2345-52 (2015).

8. Anensen, N. et al. Correlation analysis of p53 protein isoforms with NPM1/FLT3 mutations and therapy response in acute myeloid leukemia. Oncogene 31, 1533-45 (2012).

9. Avery-Kiejda, K.A., Morten, B., Wong-Brown, M.W., Mathe, A. & Scott, R.J. The relative mRNA expression of p53 isoforms in breast cancer is associated with clinical features and outcome. Carcinogenesis 35, 586-96 (2014).

10. Zhang, H. et al. p53beta: a new prognostic marker for patients with clear-cell renal cell carcinoma from 5.3 years of median follow-up. Carcinogenesis 39, 368-374 (2018).

11. Rojas, E.A. et al. Expression of p53 protein isoforms predicts survival in patients with multiple myeloma. Am J Hematol 97, 700-710 (2022).

12. Groen, K., Steffens Reinhardt, L., Bourdon, J.C. & Avery-Kiejda, K.A. It is not all about the alpha: elevated expression of p53β variants is associated with lower probability of survival in a retrospective melanoma cohort. Cancer Cell Int 23, 228 (2023).

13. Gadea, G. et al. TP53 drives invasion through expression of its Delta133p53beta variant. Elife 5(2016).

14. Kazantseva, M. et al. Elevation of the TP53 isoform Delta133p53beta in glioblastomas: an alternative to mutant p53 in promoting tumor development. J Pathol 246, 77-88 (2018).

15. Miele, A. et al. HiNF-P directly links the cyclin E/CDK2/p220NPAT pathway to histone H4 gene regulation at the G1/S phase cell cycle transition. Mol Cell Biol 25, 6140-53 (2005).

16. Medina, R. et al. The HiNF-P/p220NPAT cell cycle signaling pathway controls nonhistone target genes. Cancer Res 67, 10334-42 (2007).

17. Lee, A.S. et al. Detailed deletion mapping at chromosome 11q23 in colorectal carcinoma. Br J Cancer 83, 750-5 (2000).

18. Saarinen, S. et al. Exome sequencing reveals germline NPAT mutation as a candidate risk factor for Hodgkin lymphoma. Blood 118, 493-8 (2011).

19. Tanabe, K., Awane, R., Shoda, T., Yamazoe, K. & Inoue, Y.H. Mutations in mxc Tumor-Suppressor Gene Induce Chromosome Instability in Drosophila Male Meiosis. Cell Struct Funct 44, 121-135 (2019).

20. Sang, R. et al. Mxc, a Drosophila homolog of mental retardation-associated gene NPAT, maintains neural stem cell fate. Cell Biosci 12, 78 (2022).

21. Santamaría, P. & Randsholt, N.B. Characterization of a region of the X chromosome of Drosophila including multi sex combs (mxc), a Polycomb group gene which also functions as a tumour suppressor. Mol Gen Genet 246, 282-90 (1995).

22. Saget, O., Forquignon, F., Santamaria, P. & Randsholt, N.B. Needs and targets for the multi sex combs gene product in Drosophila melanogaster. Genetics 149, 1823-38 (1998).

23. Kurihara, M., Komatsu, K., Awane, R. & Inoue, Y.H. Loss of Histone Locus Bodies in the Mature Hemocytes of Larval Lymph Gland Result in Hyperplasia of the Tissue in mxc Mutants of Drosophila. Int J Mol Sci 21(2020).

24. Helderman, N.C. et al. Molecular functions of MCM8 and MCM9 and their associated pathologies. iScience 26, 106737 (2023).

25. Goldberg, Y. et al. Mutated MCM9 is associated with predisposition to hereditary mixed polyposis and colorectal cancer in addition to primary ovarian failure. Cancer Genet 208, 621-4 (2015).

26. Goldberg, Y. et al. MCM9 is associated with germline predisposition to early-onset cancer-clinical evidence. NPJ Genom Med 6, 78 (2021).

27. Golubicki, M. et al. Germline biallelic Mcm8 variants are associated with early-onset Lynch-like syndrome. JCI Insight 5(2020).

28. Valle, L. et al. Update on genetic predisposition to colorectal cancer and polyposis. Mol Aspects Med (2019).

29. Chubb, D. et al. Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. Nat Commun 7, 11883 (2016).

30. Che, R., Zhang, J., Nepal, M., Han, B. & Fei, P. Multifaceted Fanconi Anemia Signaling. Trends Genet 34, 171-183 (2018).

31. Fang, C.B., Wu, H.T., Zhang, M.L., Liu, J. & Zhang, G.J. Fanconi Anemia Pathway: Mechanisms of Breast Cancer Predisposition Development and Potential Therapeutic Targets. Front Cell Dev Biol 8, 160 (2020).

32. Cullinane, C.M. et al. Risk of colorectal cancer associated with BRCA1 and/or BRCA2 mutation carriers: systematic review and meta-analysis. Br J Surg 107, 951-959 (2020).

33. Oh, M. et al. BRCA1 and BRCA2 Gene Mutations and Colorectal Cancer Risk: Systematic Review and Meta-analysis. J Natl Cancer Inst 110, 1178-1189 (2018).

34. AlDubayan, S.H. et al. Inherited DNA-Repair Defects in Colorectal Cancer. Am J Hum Genet 102, 401-414 (2018).

35. Esteban-Jurado, C. et al. The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer. Eur J Hum Genet 24, 1501-5 (2016).

36. Esteban-Jurado, C. et al. Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer. Genet Med 17, 131-42 (2015).

37. Martin-Morales, L. et al. Novel genetic mutations detected by multigene panel are associated with hereditary colorectal cancer predisposition. PLoS One 13, e0203885 (2018).

38. Segui, N. et al. Germline Mutations in FAN1 Cause Hereditary Colorectal Cancer by Impairing DNA Repair. Gastroenterology 149, 563-6 (2015).

39. Singh, A.K. et al. Detection of germline variants with pathogenic potential in 48 patients with familial colorectal cancer by using whole exome sequencing. BMC Med Genomics 16, 126 (2023).

40. Schaffer, A.A. Digenic inheritance in medical genetics. J Med Genet 50, 641-52 (2013).

41. Niessen, R.C. et al. Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. Genes Chromosomes Cancer 48, 737-44 (2009).

42. Giraldez, M.D. et al. Association of MUTYH and MSH6

germline mutations in colorectal cancer patients. Fam Cancer 8, 525-31 (2009).

43. Gu, Y. et al. Human MutY homolog, a DNA glycosylase involved in base excision repair, physically and functionally interacts with mismatch repair proteins human MutS homolog 2/human MutS homolog 6. J Biol Chem 277, 11135-42 (2002).

44. Morak, M., Massdorf, T., Sykora, H., Kerscher, M. & Holinski-Feder, E. First evidence for digenic inheritance in hereditary colorectal cancer by mutations in the base excision repair genes. Eur J Cancer 47, 1046-55 (2011).

45. Chan, J.M. et al. Inherited BRCA1 and RNF43 pathogenic variants in a familial colorectal cancer type X family. Fam Cancer (2023).

46. Quintana, I. et al. Evidence suggests that germline RNF43 mutations are a rare cause of serrated polyposis. Gut 67, 2230-2232 (2018).

47. Gala, M.K. et al. Germline mutations in oncogene-induced senescence pathways are associated with multiple sessile serrated adenomas. Gastroenterology 146, 520-9 (2014).

48. Brinch, H.H. et al. Germline pathogenic variants in RNF43 in patients with and without serrated polyposis syndrome. Fam Cancer 24, 3 (2024).

49. Wang, H.Q. et al. STAT3 pathway in cancers: Past, present, and future. MedComm (2020) 3, e124 (2022).

50. Lee, J. et al. Signal transducer and activator of transcription 3 (STAT3) protein suppresses adenoma-to-carcinoma transition in Apcmin/+ mice via regulation of Snail-1 (SNAI) protein stability. J Biol Chem 287, 18182-9 (2012).

51. Yeom, M. et al. Effects of Twelve Germline Missense Variations on DNA Lesion and G-Quadruplex Bypass Activities of Human DNA Polymerase REV1. Chem Res Toxicol 29, 367-79 (2016).

52. Volkova, N.V. et al. Mutational signatures are jointly shaped by DNA damage and repair. Nat Commun 11, 2169 (2020).

53. Mahmoudi, S. et al. Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. Mol Cell 33, 462-71 (2009).

54. Venteicher, A.S. et al. A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. Science 323, 644-8 (2009).

55. Cao, H.Y., Wang, S., Zhang, Z.Y. & Lou, J.Y. Association between the WRAP53 gene rs2287499 C>G polymorphism and cancer risk: A meta-analysis. Genet Mol Res 15(2016).

56. Garcia-Closas, M. et al. Common genetic variation in TP53 and its flanking genes, WDR79 and ATP1B2, and susceptibility to breast cancer. Int J Cancer 121, 2532-8 (2007).

57. Mędrek, K. et al. Association of common WRAP 53 variant with ovarian cancer risk in the Polish population. Mol Biol Rep 40, 2145-7 (2013).

58. Drost, J. et al. Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. Science 358, 234-238 (2017).

59. Van Gool, I.C. et al. Adjuvant Treatment for POLE Proofreading Domain-Mutant Cancers: Sensitivity to Radiotherapy, Chemotherapy, and Nucleoside Analogues. Clin Cancer Res 24, 3197-3203 (2018).

60. Schmit, S.L. et al. Novel Common Genetic Susceptibility Loci for Colorectal Cancer. J Natl Cancer Inst 111, 146-157 (2019).

61. Middeldorp, A. et al. Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. Cancer Epidemiol Biomarkers Prev 18, 3062-7 (2009).

62. Middeldorp, A. et al. Comprehensive genetic analysis of seven large families with mismatch repair proficient colorectal cancer. Genes Chromosomes Cancer 49, 539-48 (2010).

63. Kiely, M. & Kiely, P.A. PP2A: The Wolf in Sheep's Clothing? Cancers (Basel) 7, 648-69 (2015).

64. Sablina, A.A. et al. The tumor suppressor PP2A Abeta regulates the RalA GTPase. Cell 129, 969-82 (2007).

65. Carnevali, P. et al. Computational techniques for human genome resequencing using mated gapped reads. J Comput Biol 19, 279-92 (2012).

66. Rao, S. et al. RPL22L1 induction in colorectal cancer is associated with poor prognosis and 5-FU resistance. PLoS One 14, e0222392 (2019).

67. Broderick, P. et al. Validation of Recently Proposed Colorectal Cancer Susceptibility Gene Variants in an Analysis of Families and Patients-a Systematic Review. Gastroenterology 152, 75-77.e4 (2017).

68. Nieminen, T.T. et al. Germline mutation of RPS20, encoding a ribosomal protein, causes predisposition to hereditary nonpolyposis colorectal carcinoma without DNA mismatch repair deficiency. Gastroenterology 147, 595-598.e5 (2014).

69. Sung, H. et al. Colorectal cancer incidence trends in younger versus older adults: an analysis of population-based cancer registry data. Lancet Oncol 26, 51-63 (2025).

70. Daca-Alvarez, M. et al. Familial component of early-onset colorectal cancer: opportunity for prevention. Br J Surg 109, 1319-1325 (2022).

71. Koster, R. et al. Impact of genetic counseling strategy on diagnostic yield and workload for genome-sequencing-based tumor diagnostics. Genet Med 26, 101032 (2023).

72. Xin, J. et al. Risk assessment for colorectal cancer via polygenic risk score and lifestyle exposure: a large-scale association study of East Asian and European populations. Genome Med 15, 4 (2023).

73. Terlouw, D. et al. Colibactin mutational signatures in NTHL1 tumor syndrome and MUTYH associated polyposis patients. Genes Chromosomes Cancer 63, e23208 (2024).

74. Naddaf, R. et al. Gut microbial signatures are associated with Lynch syndrome (LS) and cancer history in Druze communities in Israel. Sci Rep 13, 20677 (2023).

75. Burn, J. et al. Cancer prevention with aspirin in hereditary colorectal cancer (Lynch syndrome), 10-year follow-up and registry-based 20-year data in the CAPP2 study: a double-blind, randomised, placebo-controlled trial. Lancet 395, 1855-1863 (2020).