



OPEN ACCESS

EDITED BY

Sharon R Pine,
University of Colorado, United States

REVIEWED BY

Gianluca Tedaldi,
Scientific Institute of Romagna for the
Study and Treatment of Tumors (IRCCS),
Italy
Sergi Castellvi-Bel,
August Pi i Sunyer Biomedical Research
Institute (IDIBAPS), Spain

*CORRESPONDENCE

Raj Ramesar
✉ raj.ramesar@uct.ac.za

RECEIVED 06 July 2023

ACCEPTED 25 September 2023

PUBLISHED 27 October 2023

CITATION

Yildiz S, Musarurwa TN, Algar U,
Chambuso R, Rebello G, Goldberg PA and
Ramesar R (2023) Genetic insights: High
germline variant rate in an indigenous
African cohort with early-onset
colorectal cancer.
Front. Oncol. 13:1253867.
doi: 10.3389/fonc.2023.1253867

COPYRIGHT

© 2023 Yildiz, Musarurwa, Algar, Chambuso,
Rebello, Goldberg and Ramesar. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Genetic insights: High germline variant rate in an indigenous African cohort with early-onset colorectal cancer

Safiye Yildiz¹, Takudzwa N. Musarurwa¹, Ursula Algar²,
Ramadhani Chambuso¹, George Rebello¹,
Paul A. Goldberg² and Raj Ramesar^{1*}

¹UCT/MRC Genomic and Precision Medicine Research Unit, Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine, University of Cape Town and Affiliated Hospitals, Cape Town, South Africa, ²The Colorectal Unit of the Department of Surgery, Groote Schuur Hospital and the University of Cape Town, Cape Town, South Africa

Introduction: The increase in incidence of colorectal cancer in young patients of African ancestry coupled with increased aggressiveness has warranted investigation of the heritable nature of these cancers. Only a limited number of published reports of hereditary colorectal cancer in indigenous African populations have been reported and no systematic screening of these groups has been performed previously. We aimed to investigate causative germline variants and to establish the incidence of pathogenic/likely pathogenic germline variants in the known colorectal cancer genes in indigenous African colorectal cancer patients using a next-generation sequencing (NGS) multigene panel.

Materials and methods: Patients were selected from two hospitals in Cape Town and Johannesburg, South Africa. Patients with unresolved molecular diagnosis with an age of onset below or at 60 years were selected. Germline DNA samples were analyzed using a 14-gene NGS panel on the Ion Torrent platform. Variant calling and annotation were performed, and variants were classified according to the American College of Medical Genetics and Genomics guidelines. Observed variants were verified by Sanger sequencing and/or long-range PCR.

Results: Out of 107 patients, 25 (23.4%) presented with a pathogenic/likely pathogenic germline variant (PGV). Fourteen PGVs in at least one mismatch repair (MMR) gene were identified and verified in 12 patients (11.2%). Of these MMR gene variants, five were novel. The remaining 10 PGVs were in the APC, BMPR1A, MUTYH, POLD1, and TP53 genes.

Conclusion: The high incidence of PGVs associated with early-onset colorectal cancer in indigenous African patients has important implications for hereditary colorectal cancer risk management. These findings pave the way for personalized genetic screening programs and cascade testing in South Africa. The next step would involve further screening of the unresolved cases using tools to detect copy number variation, methylation, and whole exome sequencing.

KEYWORDS

hereditary colorectal cancer (CRC), next generation sequencing -(NGS), early-onset colorectal cancer, germline variants, genetic insight, indigenous african cohort, South African populations, multigene panel

1 Introduction

Colorectal cancer (CRC) is the third most common neoplastic disease worldwide (1). Approximately 20%–30% of CRC cases are due to inherited pathogenic mutations (2–4). Approximately 10% of all new CRC cases occur in young individuals, and CRC-related mortality has also increased in this group over the past decade (4, 5). It is estimated that approximately 25% of rectal cancers and 10% to 12% of colon cancers will be diagnosed worldwide in people younger than 50 years of age in the next 10 years (4). Notably, in sub-Saharan Africa, the prevalence of CRC is increasing predominantly in young individuals below the age of 40 years especially in indigenous African patients (6–9). Most importantly, comprehensive cancer genetics in early age-of-onset (eAOC) CRC has not yet been investigated in the indigenous African population despite their increasing prevalence and greater aggressiveness than in non-African populations (10). This presents a public-health problem in cancer genetics. Screening, treatment, and targeted prevention of CRC will remain limited and ineffective in indigenous Africans if this problem is not addressed.

Currently, it is difficult to establish a true picture of hereditary CRC burden in South Africa due to underreporting and underdiagnosis especially among indigenous Africans (11, 12). Consequently, systematic screening and genetic testing strategies for hereditary CRC in these populations have been suboptimal. Only a limited number of published reports of multigene panel testing for hereditary CRC in indigenous Africans have been reported, which remains a major gap in population-specific genetics (13, 14). As we have previously described, South Africa is located at the southern tip of Africa and in a geographic cul-de-sac for southward migration of indigenous Africans along the West and East coasts of the continent (15, 16). In this regard, and anthropologically, our setting in southern Africa provides us with an ideal platform to study the hereditary phenomenon of CRC in indigenous Africans (17, 18). In our ongoing CRC work in local populations in the Western and Northern Cape Provinces of South Africa, we have shown that genetic screening increased the rate of identification of at-risk individuals with inherited cancer-predisposing syndromes including Lynch syndrome (LS) with

pathogenic variants in the DNA mismatch repair (MMR) genes (19–21).

For CRC, multigene panel testing by next-generation sequencing (NGS) is now widely used to screen for hereditary CRC. The genetic syndromes associated with CRC are categorized into two distinct groups: non-polyposis and polyposis syndromes, where variants in the non-polyposis syndrome genes are found more prevalent (22). NGS panel testing strategies prioritize inclusion of genes with high prevalence associated with non-polyposis syndromes (23). The selection of genes to include in the NGS panel is based on (i) those most listed as high-risk genes according to the latest American College of Medical Genetics and Genomics (ACMG) guidelines for hereditary CRC, and (ii) genes known as key role players in hereditary CRC previously identified in other populations (24, 25). In addition to the four MMR genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*), the gene *EPCAM* is included in an NGS multigene panel because it has a well-established association with the most common hereditary CRC syndrome, LS, due to deletions in this gene leading to *MSH2* hypermethylation (26–28). The other genes included in the panel are associated with other types of hereditary CRC. These include *APC* for familial adenomatous polyposis (FAP); *BMPRIA* and *SMAD4* for Juvenile Polyposis Syndrome (JPS); *MUTYH* for *MUTYH*-associated polyposis; *PTEN* for *PTEN* tumor/hamartoma syndrome (or Cowden syndrome/Bannayan–Riley–Ruvalcaba syndrome); and *STK11* for Peutz–Jeghers syndrome. In addition, *TP53*, which is associated with familial Li–Fraumeni syndrome, and often seen in sporadic cancers, is included in the panel. Recently, studies have shown that *POLD1* and *POLE* play a role in some familial CRC and even in LS predisposition and are also included in our NGS 14-gene panel (29–31). Other genes such as *RPS20* (associated with MMR-proficient, non-polyposis CRC; *RPS20*-associated hereditary CRC) and *NTHL1* (associated with polyposis hereditary CRC; *NTHL1*-tumor syndrome) are less prevalent and may be added to CRC NGS panels. There are few other genes reported in the literature such as *RNF43* (associated with serrated polyposis CRC; *RNF43*-associated serrated polyposis syndrome) and *GREM1* (associated with mixed polyposis CRC; *GREM1*-associated mixed polyposis) in few affected individuals, but their prevalence is not well known and are less common in CRC multigene panel designs (22, 23).

We hypothesize that investigating the germline mutations in affected indigenous African patients with early-onset CRC and establishing the incidence of inherited mutations in the known CRC genes using the NGS 14-gene panel will provide more insights and inferential genetic information on the hereditary nature, treatment options, and targeted prevention of CRC. To investigate our hypothesis in indigenous African patients, we aimed to determine causative germline variants in CRC patients of indigenous African descent, and to establish the incidence of pathogenic or likely pathogenic germline variants in the known CRC genes in our indigenous African cohort, using an NGS multigene panel, in South Africa.

2 Materials and methods

2.1 Research ethics

All procedures were performed in accordance with guidelines of the Declaration of Helsinki. Ethical approval for the study was granted by the Human Research Ethics Committee of (i) the University of Cape Town (UCT), ethics approval number HREC:287/2020; (ii) all respective hospitals ethics departments; and (iii) the South African National Cancer Registry (NCR). Consent forms were previously provided to the study participants to be involved in genetic research before data collection and storage. The patients were counseled by a registered practitioner who explained the details of the research and answered all questions from the patients, in their home language, as part of the informed consent procedure.

2.2 Selection of patient cohort

Our study participant selection criteria were as follows: (i) diagnosed CRC patients of indigenous African origin, (ii) age of onset (or diagnosis age) below or equal to 60 years, (iii) available genomic DNA samples, and (iv) no known molecular diagnosis for CRC for a genetic/germline variant. Selection of these patients of indigenous African origin for our study cohort was based on self-identification criteria that patients note in their records in our hospital recruitment in South Africa, as also required by the NCR in the country¹ (where the cancer is a notifiable disorder). As part of a large ongoing hospital-based project on the genetics of CRC, a total of 107 DNA samples collected over the past 30 years were included in this study. In our cohort, we selected DNA samples from Groote Schuur Hospital in Cape Town whose records were stored in the Division of Human Genetics repository ($n = 64$), which contains clinical information and matching molecular test results. We also selected DNA samples from the Chris Hani Baragwanath Hospital in Johannesburg whose records were stored at the South African NCR database ($n = 43$), which also contains clinical information

and molecular test results. All CRC patients were previously admitted at these two hospitals in the last three decades.

2.3 Sample preparation

2.3.1 Initial quality control

DNA samples from peripheral blood were previously extracted using basic salting-out methodology according to adopted protocol in the Division of Human Genetics Laboratory at UCT and were subjected to an initial quality control (QC) to evaluate DNA dilutions before NGS screening (32). The QC involved stock DNA quantification through spectrophotometry using a Nanodrop ND 1000 spectrophotometer (Applied Biosystems, Thermo Fisher Scientific, Massachusetts, USA) and an integrity check using agarose gel electrophoresis. Samples with a concentration lower than 50 ng/ μ L, those that did not meet the DNA purity (e.g., A_{260}/A_{280} ratio), and degraded samples were excluded.

2.3.2 Extended quality control for NGS library preparation with Qubit HS and real-time PCR

A second QC step was implemented for specifically the NGS library preparation. The selected samples were diluted to 1 ng/ μ L for NGS library preparation. These dilutions were quantified using Qubit dsDNA High Sensitivity assay (Applied Biosystems, Thermo Fisher Scientific, Massachusetts, USA) followed by a real-time PCR using a TaqManTM RNase P Detection Reagents Kit (Applied Biosystems, Thermo Fisher Scientific, Massachusetts, USA) according to the manufacturer's instructions. Those with insufficient concentration and quality were excluded from NGS sequencing.

2.4 Next-generation sequencing panel screening

Library preparation of the NGS 14-gene panel for hereditary CRC (*APC*, *BMPRIA*, *EPCAM*, *MLH1*, *MSH2*, *MSH6*, *MUTYH*, *PMS2*, *POLD1*, *POLE*, *PTEN*, *SMAD4*, *STK11*, and *TP53*) was performed using the Ion AmpliSeq Kit for Chef DL8 according to the manufacturer's instructions (Applied Biosystems, Thermo Fisher Scientific, Massachusetts, USA). The panel was previously designed in our laboratory based on the existing literature and experiences in our CRC research. The list of the genes and their molecular genetic characteristics is provided in Table 1. Fifteen microliters of each diluted DNA sample (total DNA yield of 15 ng) of DNA was loaded into each well on the library DL8 plate on the Ion Chef instrument for automated Ion AmpliSeq library construction instrument (Thermo Fisher Scientific, Massachusetts, USA). The cycling conditions were set

¹ <https://www.nicd.ac.za/centres/national-cancer-registry/>

² <https://ionreporter.thermofisher.com/ir/>

³ <https://www.ncbi.nlm.nih.gov/clinvar/>

⁴ <https://www.ncbi.nlm.nih.gov/snp/>

TABLE 1 List of the 14 genes included in the next-generation sequencing panel and their reference sequence along with the chromosomal locations, risk levels, and the phenotypes.

Gene	RefSeq	Chr locus	Risk level ^a	Phenotype/Disease (ClinGen)
<i>APC: Adenomatous polyposis coli</i>	NM_000038.6	5q22.2	High	Familial adenomatous polyposis (FAP)/attenuated FAP Gastric adenocarcinoma and proximal polyposis of the stomach
<i>BMPRIA: Bone Morphogenetic Protein Receptor Type 1A</i>	NM_004329.3	10q23.2	High	Juvenile polyposis syndrome Hereditary mixed polyposis syndrome Pulmonary arterial hypertension
<i>EPCAM: Epithelial cellular adhesion molecule</i>	NM_002354.3	2p21	High	Lynch syndrome
<i>MLH1: MutL, E. coli, homolog of, 1</i>	NM_000249.4	3p21.3	High	Lynch syndrome, mismatch repair cancer syndrome 1 Hereditary breast carcinoma
<i>MSH2: MutS, E. coli, homolog of, 2</i>	NM_000251.3	2p22-p21	High	Lynch syndrome, mismatch repair cancer syndrome 1 Hereditary breast carcinoma
<i>MSH6: MutS, E. coli, homolog of, 6</i>	NM_000179.3	2p16	High	Lynch syndrome, mismatch repair cancer syndrome 1 Hereditary breast carcinoma
<i>MUTYH: MutY DNA glycosylase</i>	NM_001128425.2	1p34.1	High	MUTYH-associated polyposis FAP II
<i>PMS2: Postmeiotic segregation increased S. cerevisiae</i>	NM_000535.7	7p22	High	Lynch syndrome Mismatch repair cancer syndrome I Hereditary breast carcinoma
<i>POLD1: Polymerase (DNA directed), delta1, catalytic subunit</i>	NM_001256849.1	19q13.33	Risk	Polymerase proofreading-associated polyposis
<i>POLE: Polymerase, DNA, epsilon</i>	NM_006231.4	12q24.33	Risk	Polymerase proofreading-associated polyposis
<i>PTEN: Phosphatase and tensin homolog</i>	NM_000314.8	10q23.31	High	PTEN hamartoma tumor syndrome (or Cowden syndrome/Bannayan-Riley-Ruvalcaba syndrome)
<i>SMAD4: SMAD family member 4</i>	NM_005359.6	18q21.2	High	Juvenile polyposis syndrome (JPS); JPS/hereditary hemorrhagic telangiectasia syndrome
<i>STK11: Serine/threonine kinase 11</i>	NM_000455.5	19p13.3	High	Peutz-Jeghers syndrome
<i>TP53: Tumor protein p53</i>	NM_000546.5	17p13.1	Lower ^b	Li-Fraumeni syndrome

^aACMG-inherited CRC guidelines, 2021 (24). ^{ab}Recommendation: “Genes associated with hereditary cancer syndrome where CRC/polyposis are possible manifestations associated with defined hereditary cancer predisposition syndromes, which include CRC and/or polyposis as rare manifestations. As it is difficult to separate out specific CRC/polyposis risk, it is reasonable to include these genes routinely on multigene panels for hereditary CRC.”

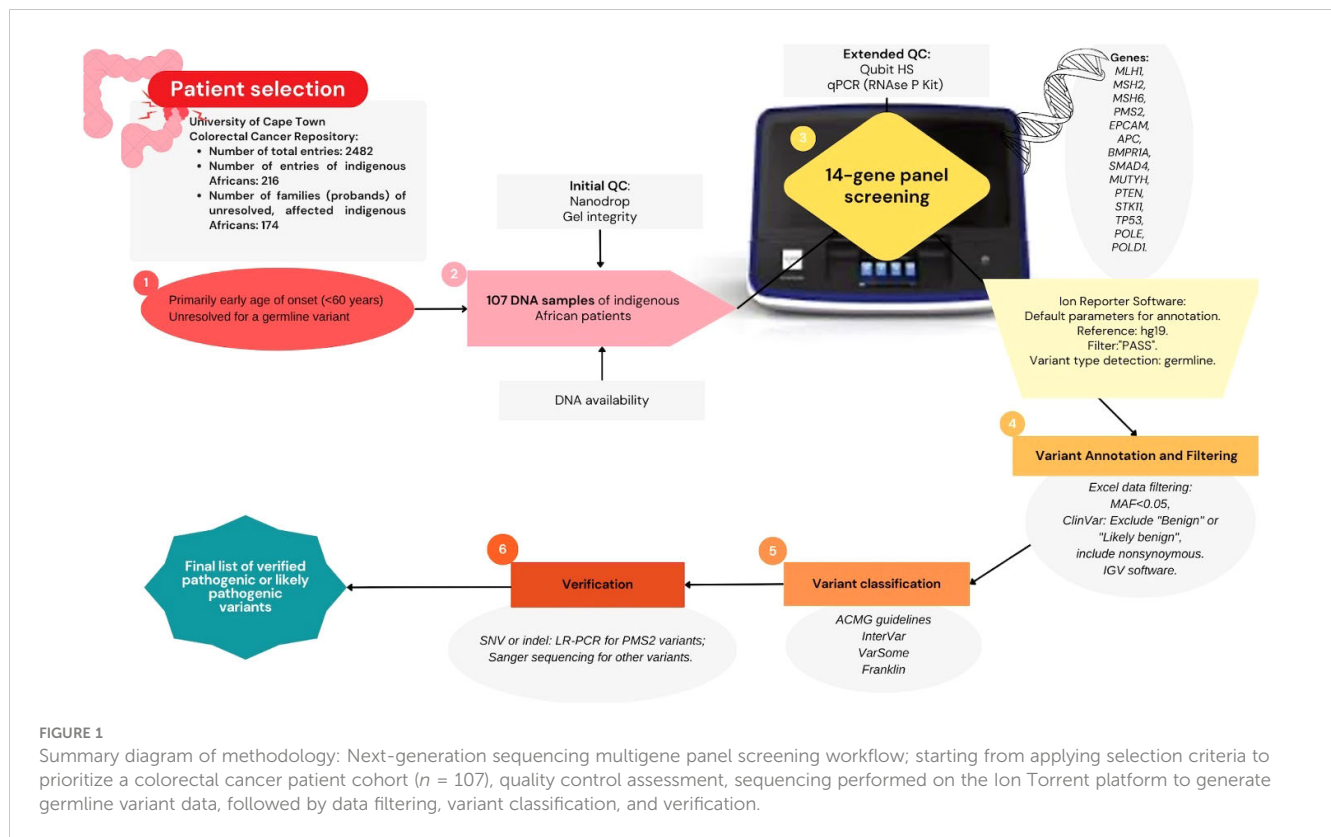
up as the following: the number of primer pairs per pool, 193–384; number of amplification cycles, 18; and anneal/extension time of 4 min. After library preparation, library quantitation was performed using an Ion Library TaqMan Quantitation Kit according to the manufacturer’s instructions (Thermo Fisher Scientific, Massachusetts, USA). The cycling conditions for library quantitation involved Uracil DNA glycosylase incubation at 50°C for 2 min, a polymerase activation at 95°C for 3 s and 40 cycles at 60°C for 30 s. The libraries were then diluted according to the quantification results after the qPCR, to a concentration ± 100 pM. Libraries were then pooled, and the pooled libraries were quantified by using the same qPCR protocol as for library quantitation used for each library. The pooled libraries were loaded onto the Ion 520™ Chips (Applied Biosystems, Thermo Fisher Scientific, Massachusetts, USA) for templating on the Ion Chef instrument. Finally, sequencing was performed on an S5 sequencer (Applied Biosystems, Thermo Fisher Scientific, Massachusetts, USA).

⁵ <https://franklin.genoox.com/>

2.5 Variant annotation and filtering

The sequencing data were processed through the Ion Server software and read alignment was done against the *Homo sapiens* reference genome hg19 to create BAM files. Variant calling and annotation were performed on the Ion Reporter software². The *tsv* output files generated for each patient were merged into a single Microsoft Excel file for variant filtering. The report included depth of coverage for each variant, minor allele frequency (MAF), location, function, four different pathogenicity prediction estimates (Sift, Polyphen, Grantham, and FATHMM), and information in some public archives of reports on variants and their phenotypes such as ClinVar³ and dbSNP⁴.

The Ion Reporter software only generated output files of variants labeled as “PASS”, and these were subjected to filtering. All synonymous variants were excluded and only exonic variants and variants in the flanking regions with an MAF less than 0.05 were retained. The MAF <0.05 was considered as the cutoff for potential rare variants to ensure any population specific variant may not be missed. These variants were then prioritized for further



analysis. The data were not filtered based on depth of coverage, but each prioritized variant was manually inspected for the depth of coverage. Variants with $30\times$ coverage or less were excluded from the prioritized list. Variants that were classified as pathogenic, likely pathogenic, uncertain significance, conflicting interpretation, or with a blank record on ClinVar were selected and further filtered by manual inspection. The prioritized variants, after the filtering, were then manually inspected on the Integrative Genomics Viewer (IGV) software to assess depth of coverage and integrity of base calls (33). Some calls were found to be an artifact of random insertions or deletions around the region where the variant is located despite a high coverage, present in multiple samples, and these were excluded from the prioritized list. VarSome, Franklin,⁵ and InterVar were used to help in assigning variant classification, according to the ACMG guidelines (34, 35). Variants classified as pathogenic or likely pathogenic were selected for verification. The workflow for the NGS panel screening from patient selection to verification is summarized in Figure 1.

2.6 Verification of the pathogenic or likely pathogenic germline variants

2.6.1 Variant verification by Sanger sequencing

Primers were either used from available stock or newly designed for the variants that needed verification by Sanger sequencing. Basic PCR amplification for each variant was then performed using

⁶ <https://www.telethon.jp>.

GoTaq[®] Flexi DNA Polymerase kit according to the manufacturer's instructions (Promega Corporation, Wisconsin, USA). To purify the PCR products prior to the sequencing reaction, they were subjected to an enzymatic clean-up reaction by the FastAP (Thermosensitive alkaline phosphatase) and *ExoI* (Exonuclease I) enzymes according to the manufacturer's instructions (Promega Corporation, Wisconsin, USA). PCR products were incubated at 37°C for 1 h followed by 75°C for 15 min.

The presence of the prioritized pathogenic or likely pathogenic germline variants was then verified by Sanger sequencing using BigDye[™] Terminator v3.1 (BDT V3.1) Cycle Sequencing Kit (Applied Biosystems, Thermo Fisher Scientific, Massachusetts, USA) according to the manufacturer's instructions. Sequencing reactions were purified using ethanol precipitation prior to capillary electrophoresis, which was carried out on the ABI 3130xl and the SeqStudio genetic analyzers (Applied Biosystems, Thermo Fisher Scientific, Massachusetts, USA). Base calling was performed on the Sequencing Analysis v5.4 software (Applied Biosystems, Thermo Fisher Scientific, Massachusetts, USA). Sequences were visualized on MySequence application.⁶

2.6.2 Variant verification by long-range PCR

A protocol for long-range PCR (LR-PCR) with a "long and accurate" *Taq* polymerase (TaKaRa Bio Inc., Otsu, Shiga, Japan) was optimized, similar to other studies, and used to confirm the presence of the variants found in the regions of the *PMS2* gene that are in complete homology with the known pseudogene that

interferes with the coding *PMS2* variant interpretation (36–38). The observed *PMS2* variants were in the 3' end of *PMS2*, which is also contained in the pseudogene, *PMS2CL*. DNA samples were diluted to 50 ng/μL, and 2 μL (making a total DNA yield of 100 ng) was used in a total volume of a 25-μL LR-PCR reaction. Previously published primers were used in a final concentration of 0.2 μM (39, 40). The cycling involved an initial denaturation at 94°C for 1 min, followed by 35 cycles of 94°C for 15 s, 65°C for 30 s, and 68°C for 15 min. Final elongation was set at 72°C for 10 min. The LR products were electrophoresed on a 0.4% agarose gel for confirmation before proceeding with the second PCR. Lambda DNA/*HindIII* Marker (Applied Biosystems, Thermo Fisher Scientific, Massachusetts, USA) was used as a high-molecular-weight marker when visualizing the LR-PCR products. Each LR-PCR product was diluted with 1:10 ratio and the diluted LR-PCR products were used as a template for the second, exon-specific PCR. This was carried out using GoTaq® Flexi DNA Polymerase according to the manufacturer's instructions (Promega Corporation, USA) in a final reaction volume of 25 μL. Primers for the second PCR amplifications were designed according to the long-range PCR product's expected reference sequence. Finally, Sanger sequencing was performed using the second PCR products following the same protocol as in the Sanger sequencing verification above.

2.6.3 Variants of uncertain significance analysis in the NGS data

The total number of variants of uncertain significance (VUSs) was calculated from the unfiltered NGS panel data files of the 107 samples. All the *tsv* output files were compiled into a single *tsv* file using R Studio, which was then filtered in Excel. The VUSs were filtered on the "ClinVar" column in the Excel sheet (41). Further filtering criteria were applied following the same parameters as the filtering criteria for pathogenic/likely pathogenic germline variant (PGV) filtering; synonymous and non-exonic variants with a depth of coverage less than 30× and with MAF greater than 0.05 were excluded.

2.7 Statistical analysis

The number of patients with PGVs was divided by the total number of patients analyzed to get the descriptive statistics. Pearson's Chi-squared test or Fisher's exact test were used to compare categorical variables. PGVs were used as dependent variables for testing the variant pathogenicity variations between different predictor variables such as age of onset, gender, and tumor location or sidedness of each patient. A *p*-value less than 0.05 was considered significant. The analyses were conducted using R Studio (version 2021.9.1.372) (41).

3 Results

3.1 Patient cohort and characteristics

Overall patients' demographics and clinical-pathological characteristics are summarized in Table 2. Patients' clinical features were statistically compared to their mutation status (positive/negative) which did not show any statistical significance. The median age of the mutation-positive and the mutation-negative patients was similar: 35 and 37 years old, respectively (*p*-value = 0.630). The youngest and the oldest age at diagnosis were 26 and 58 years, respectively (median = 37). The number of male and female patients were nearly equal: 54 and 53 years, respectively. Tumor was predominantly localized in the colon with nearly 74% of the total cohort, and the ratio of the right-sided and left-sided tumors was 50% in the overall cohort. Only one and six patients lacked clear information about the tumor localization and sidedness, respectively. Available information on previous immunohistochemistry (IHC) testing result for patients in our database was also retrieved. Although these samples were only a handful of the total indigenous African cohort, it helped us make the comparison between the MMR PGVs versus patients' IHC status (23 out of 107) (Supplementary Table 1).

3.2 Variant filtering results

Variant detection was performed using the NGS multigene panel for the DNA samples that passed the QC. Mean depth of coverage of the complete targeted regions in the NGS data was 604×. The total number of variants detected, before any filtering, was 15,063 (100 to 200 per individual). After data filtering based on the location and the function of variants, this number decreased to 1,092 (10–15 variants per individual). Further filtering out of variants with a MAF (>0.05) then reduced the total number of retained variants to 360, and duplicates were removed (305 unique variants), which ranged from zero to eight variants per individual, for further inspection for variant classification. The final prioritized list of patients presented with a PGV, before verification, and their features are shown in Table 3 (see Supplementary Table 2 for detailed information).

3.3 Pathogenic/likely pathogenic germline variants identified after the NGS panel screening

The NGS multigene panel screening resulted in 24 germline PGVs in 27 patients, before verification. Fourteen PGVs in the MMR genes were identified: four pathogenic variants in *MLH1*,

TABLE 2 Patient demographics and clinical–pathological characteristics.^a

	All (N = 107)	Positive (N = 27)	Negative (N = 80)	p-value
Total:	107	25.23%	74.77%	
Diagnosis age (years)				
Mean (SD) ^b	35.64 (8.71)	35.33 (9.24)	35.76 (8.58)	
Median	37	35	37	0.630 ^c
Range	26 to 58	26 to 58	18 to 58	
Categorical age				
<50 years	104 (97.20%)	25 (92.59%)	79 (98.75%)	0.156 ^d
≥50 years	3 (2.80%)	2 (7.41%)	1 (1.25%)	
Gender				
Male	54 (50.94%)	15 (55.56%)	39 (48.75%)	0.697 ^e
Female	53 (49.06%)	12 (46.15%)	41 (50.62%)	
Tumor localization				
Colon	79 (73.83%)	18 (66.67%)	61 (76.25%)	0.069 ^d
Rectum	25 (23.36%)	7 (25.93%)	18 (22.50%)	
Both	2 (1.87%)	2 (7.41%)	0	
Not applicable	1 (0.93%)	0	1 (1.25%)	
Tumor sidedness				
Right sided	47 (43.93%)	12 (44.44%)	35 (43.75%)	0.144 ^d
Left sided	47 (43.93%)	10 (37.04%)	37 (46.25%)	
Both	7 (6.54%)	4 (14.81%)	3 (3.75%)	
Not applicable	6 (5.61%)	1 (3.70%)	5 (6.25%)	
Immunohistochemistry testing status				
Loss of a mismatch repair protein	14 (13.08%)	5 (18.52%)	9 (11.25%)	
Normal staining (no loss)	9 (8.41%)	3 (11.11%)	6 (7.50%)	
Not applicable	84 (78.50%)	19 (70.37%)	66 (82.50%)	

^aThis table includes all the patients presented with a pathogenic/likely pathogenic variant before the verification step. ^bStandard deviation. ^cWilcoxon rank-sum test. ^dFisher's exact test. ^ePearson's Chi-squared test.

seven pathogenic variants and one likely pathogenic variant in *MSH2*, two pathogenic variants in *PMS2*, and no PGV in *MSH6*. The remainder of the observed germline variants were in the *APC* (one likely pathogenic variant), *BMPRIA* (two pathogenic and two likely pathogenic variants), *MUTYH* (one likely pathogenic variant), *POLD1* (one pathogenic and one likely pathogenic variants), and *TP53* (two pathogenic variants) genes, shown in Figure 2. Ten of the observed PGVs were novel; 50% of these were MMR variants. The distribution of the variants per gene in the NGS 14-gene panel used in this study with variant types of pathogenic, likely pathogenic, and VUS is summarized in Figure 3 where *STK11* did not carry any of the three types of variants, and *EPCAM* and *PTEN* only contained VUSs.

3.3.1 Variants identified in MMR genes

An *MLH1* (*NM_000249.4*) missense pathogenic variant, c.2263A>G, was found in one individual, patient 2, whose age at

diagnosis was less than 30 years. Another pathogenic *MLH1* variant, located at canonical splice site (*MLH1*:c.117-1G>A), was identified in two individuals, patient 1 and patient 3, who were both diagnosed under 40 years of age. Another pathogenic variant in *MLH1*, c.793C>T, was carried by another individual, patient 4, diagnosed at under 50 years of age; this variant has been previously reported in literature in affected CRC patients (43–48). Importantly, manual inspection allowed us to identify two additional pathogenic MMR variants: one was a synonymous, but pathogenic splice site variant in *MLH1*, c.1731G>A in patient 20. This patient lacked detailed clinical information but was diagnosed under 50 years of age and the records indicated that the patient deceased only a few months after diagnosis. The other pathogenic variant was a two-base pair deletion, *MSH2*: c.387_388delTC in patient 17 (eAOO <30 years) who had a family history of cancer (Table 3; Supplementary Figure 1).

An insertion in the first exon of *MSH2* (*NM_000251.3*), c.187_188insG, was present in patient 23, who was diagnosed at

TABLE 3 List of patients with germline pathogenic and/or likely pathogenic variants that were identified after the NGS multigene panel screening in this study.

Sample ID	Gene	Coding DNA	Protein	MAF ^a	ClinVar	ACMG ^b variant classification	Reported before?
1 ^c	<i>MLH1</i>	c.117-1G>A	p.?	NA	Pathogenic	Pathogenic	No (G>T only)
2 ^d	<i>MLH1</i>	c.2263A>G	p.Arg755Gly	NA	Pathogenic	Pathogenic	No
3	<i>MLH1</i>	c.117-1G>A	p.?	NA	Pathogenic	Pathogenic	No (G>T only)
4	<i>MLH1</i>	c.793C>T	p.Arg265Cys	NA	Pathogenic	Pathogenic	Yes
5 ^d	<i>BMPRI1A</i>	c.834C>A	p.Tyr278Ter	NA	Pathogenic	Pathogenic	No
6 ^d	<i>BMPRI1A</i>	c.717delA	p.Val240LeufsTer21	NA	No entry	Pathogenic	No
7	<i>MSH2</i>	c.2377C>T	p.Gln793Ter	NA	Pathogenic/likely pathogenic	Pathogenic	No
8 ^d	<i>POLD1</i>	c.1816C>A	p.Leu606Met	NA	Conflicting interpretation of pathogenicity	Likely pathogenic	Yes (<i>de novo</i> in one African)
9	<i>BMPRI1A</i>	c.185A>G	p.Tyr62Cys	NA	Uncertain significance	Likely pathogenic	No
10	<i>PMS2</i>	c.2192_2196delTAACT	p.Leu731CysfsTer3	0.0	Pathogenic	Pathogenic	Yes
11	<i>BMPRI1A</i>	c.717delA	p.Val240LeufsTer21	NA	No entry	Pathogenic	No
12 ^c	<i>MSH2</i>	c.187delG	p.Val63Ter	NA	Pathogenic	Pathogenic	Yes (breast cancer)
13	<i>TP53</i>	c.526T>C	p.Cys176Arg	NA	Conflicting interpretation of pathogenicity	Pathogenic	Yes (only somatic)
14	<i>POLD1</i>	c.1265_1266insG	p.Arg423ProfsTer212	NA	No entry	Pathogenic	No
15	<i>BMPRI1A</i>	c.355C>T	p.Arg119Cys	NA	Conflicting interpretation of pathogenicity	Pathogenic	Yes
16	<i>MSH2</i>	c.1705_1706delGA	p.Glu569IlefsTer2	NA	Pathogenic	Pathogenic	Yes
17 ^c	<i>MSH2</i>	c.387_388delTC	p.Gln130fs	NA	Pathogenic	Pathogenic	Yes
18	<i>PMS2</i>	c.1579_1580delAG	p.Arg527GlyfsTer14	0.0	Pathogenic	Pathogenic	Yes (breast/ovarian cancer)
19	<i>MSH2</i>	c.643C>A	p.Gln215Lys	NA	No entry	Likely pathogenic	No (C>T only)
20 ^d	<i>MLH1</i>	c.1731G>A	p.Ser577=	NA	Pathogenic	Pathogenic	Yes
21	<i>MSH2</i>	c.1923T>A	p.Cys641Ter	NA	No entry	Pathogenic	No
21	<i>PMS2</i>	c.2192_2196delTAACT	p.Leu731CysfsTer3	0.0	Pathogenic	Pathogenic	Yes
22	<i>PMS2</i>	c.2192_2196delTAACT	p.Leu731CysfsTer3	0.0	Pathogenic	Pathogenic	Yes
23	<i>MSH2</i>	c.187_188insG	p.Val63fs	NA	Pathogenic	Pathogenic	Yes
24	<i>MUTYH</i>	c.1476 + 2C>T	p.?	0.002	Conflicting interpretations of pathogenicity	Likely pathogenic	Yes
25	<i>MSH2</i>	c.1569delT	p.Arg524ValfsTer2	NA	Pathogenic	Pathogenic	Yes
26	<i>TP53</i>	c.742C>T	p.Arg248Trp	NA	Pathogenic	Pathogenic	Yes
27	<i>APC</i>	c.2868C>A	p.Tyr956Ter	NA	No entry	Likely pathogenic	No (C>G only)

^aMinor allele frequency, gnomAD. "NA" = "Not available". ^bAmerican College of Medical Genetics and Genomics (42). ^cFamily history of cancer. ^dDeceased after diagnosis.

under 40 years of age. A deletion was present at the same position; *MSH2*:c.187delG in another individual (patient 12); these both cause an early protein truncation. Both deletion and duplication at position 187 have previously been reported in LS, breast, endometrial, or ovarian cancer patients and affected family members (49–52). A frameshift pathogenic variant in *MSH2*; c.1705_1706delGA was present in one individual and has also

been reported in the literature before as a disease-causing variant associated with CRC (53). A novel pathogenic variant, *MSH2*: c.1569delT, was found in one patient, patient 25, with strong evidence for pathogenicity according to the ACMG guidelines (42).

One patient presented with two distinct pathogenic germline variants in the MMR genes *MSH2* and *PMS2* (*NM_000535.7*), after variant filtering and manual inspection. The *MSH2*:c.1923T>A

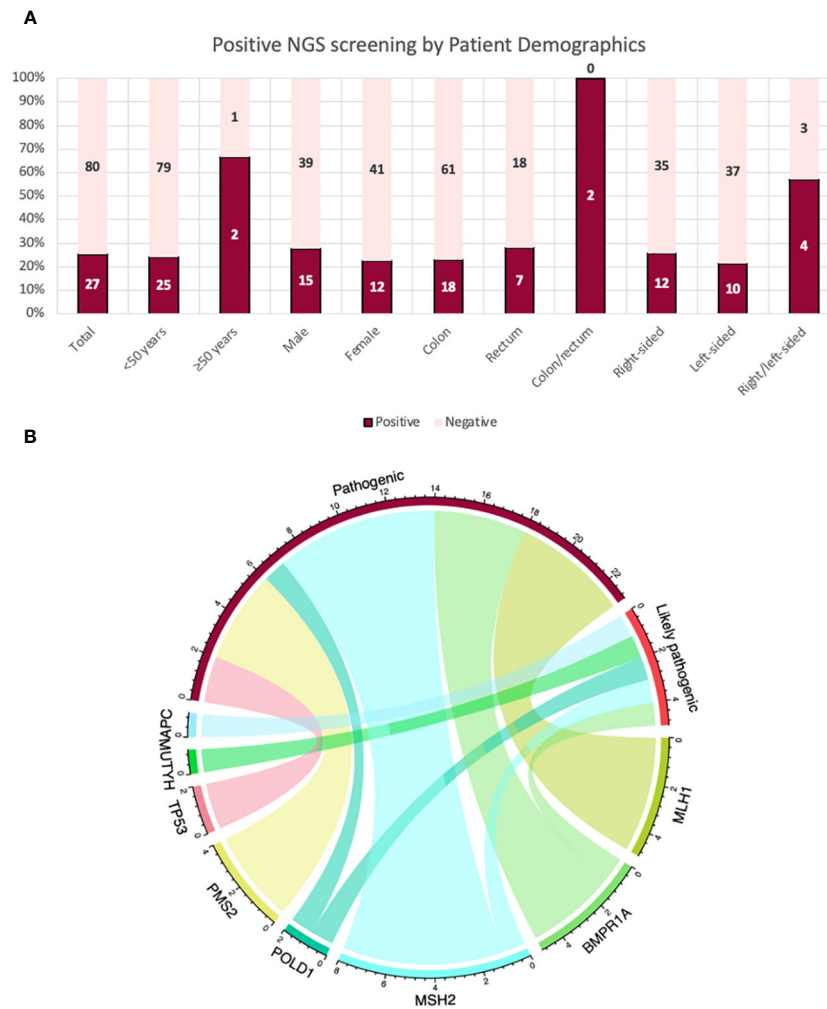


FIGURE 2 (A) Patients with a pathogenic/likely pathogenic variant post-NGS 14-gene panel analysis by their characteristics and demographics. (B) A chord diagram to demonstrate the relationship of the genes for pathogenic and likely pathogenic variants in the total cohort in this study: Each bar represents a variant in their corresponding genes; aligning with its corresponding type of variant as either pathogenic or likely pathogenic. The highest proportion of variants were in *MSH2* ($n = 8$); seven pathogenic variants and one likely pathogenic variant.

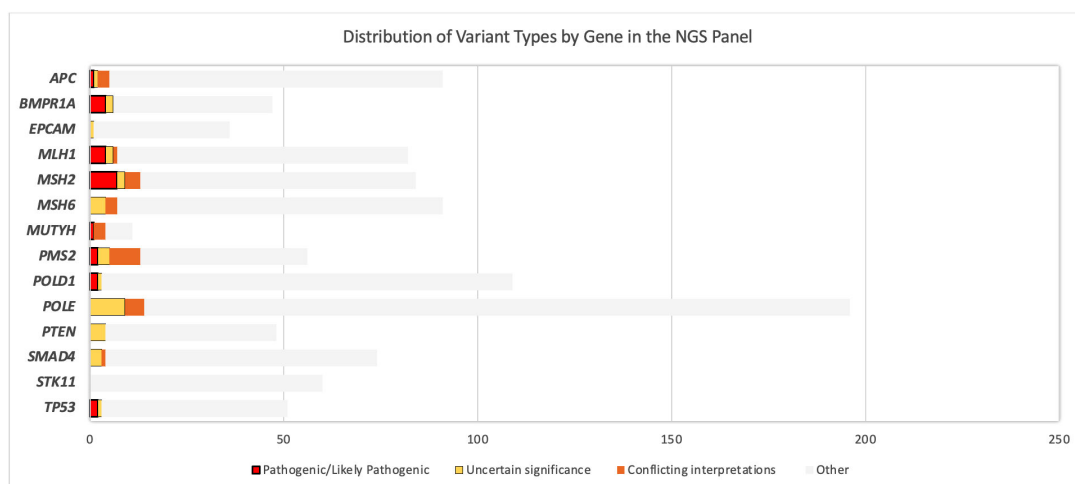


FIGURE 3 Distribution of the variants per gene in the next-generation sequencing 14-gene panel used in this study: with variant types of pathogenicity. *EPCAM*, *PTEN*, *SMAD4*, and *STK11* did not carry any pathogenic/likely pathogenic variant (PGV) or any variant of uncertain significance (VUS). "Other" includes variants besides PGV and VUS (e.g., benign/likely benign variants, variants not reported in ClinVar).

TABLE 4 List of patients from our mutation-positive cohort whose previous immunohistochemistry (IHC) testing was available in our database.

Sample ID	IHC staining information	Variant identified	Age (years)
1	MLH1 negative	<i>MLH1</i> :c.117-1G>A	<40
8	All positive	<i>POLD1</i> :c.1816C>A	<30
14	All positive	<i>POLD1</i> : c.1265_1266insG	<30
16	MSH2 and MSH6 negative	<i>MSH2</i> : c.1705_1706delGA	<60
17	MSH2 negative	<i>MSH2</i> : c.387_388delTTC	<30
19	MLH1 negative	<i>MSH2</i> :c.643C>A	<30
20	MLH1 and PMS2 negative	<i>MLH1</i> :c.1731G>A	<50

variant identified in this individual (patient 21) is a novel, stop-gain variant in exon 12 (which has a total of 16 exons). The *PMS2*:c.2192_2196delTAACT variant is also a pathogenic, previously reported, frameshift variant. However, LR-PCR did not verify the presence of the *PMS2* variant in the patient.

3.3.2 Variants identified in the non-MMR genes in the NGS panel

Two pathogenic variants in *TP53* (*NM_000546.6*), *TP53*:c.526T>C and *TP53*:c.742C>T, were present in two unrelated patients, patient 13 and patient 26, respectively. These patients were both diagnosed at a relatively early age (less than 30 years). A pathogenic variant in *APC* (*NM_000038.6*), *APC*:c.2868C>A, was identified in another patient, patient 27, and this finding corresponded with the pathology report providing clinical

information associated with FAP. A frameshift *POLD1* (*NM_001256849.1*) pathogenic variant, c.1265_1266insG, which leads to protein truncation in exon 11 (total exon number = 27), was carried by patient 14 who was diagnosed at a relatively early age (younger than 25 years), also having rectal lesions. Insufficient clinical information was available on this patient to supply further information. A likely pathogenic variant was present in another non-MMR gene, *MUTYH* (*NM_001128425.2*); c.1476 + 2C>T, in patient 24, where the patient presented with polyps and right-sided tumor that spread to the liver, which is consistent with a polyposis (or *MUTYH*) associated CRC.

Five unrelated patients carried a PGV in *BMPRIA* (*NM_004329.3*). One of these patients (patient 5) carried a pathogenic variant, *BMPRIA*:c.834C>A, and this was consistent with the clinical diagnosis of the patient, i.e., the patient was clinically diagnosed with JPS, but no mutation had been identified to date. This patient was less than 40 years of age when diagnosed and died a year after diagnosis. Another pathogenic, frameshift variant, *BMPRIA*:c.717delA, was found in two unrelated individuals (patient 6 and patient 11) who were diagnosed at age younger than 30 and 40 years, respectively. Lastly, two likely pathogenic *BMPRIA* variants were also present in two unrelated individuals, *BMPRIA*:c.185A>G in patient 9 and *BMPRIA*:c.355C>T in patient 15. These patients, patient 9 and patient 15, were also diagnosed at relatively young ages, i.e., <40 and <30 years, respectively. Some of the patients who were recorded to have an MMR protein loss based on their previous IHC testing presented with either an MMR or a non-MMR variant in this study, and age range at diagnosis is also shown, for comparison purposes as in Table 4. Among these, only six of the IHC testing results corresponded with the NGS panel screening result. Interestingly, two individuals previously had a normal IHC staining of all the MMR genes, and the NGS panel screening resulted in a PGV in a non-MMR gene, *POLD1*, which supports the IHC finding. Another patient was IHC-negative for

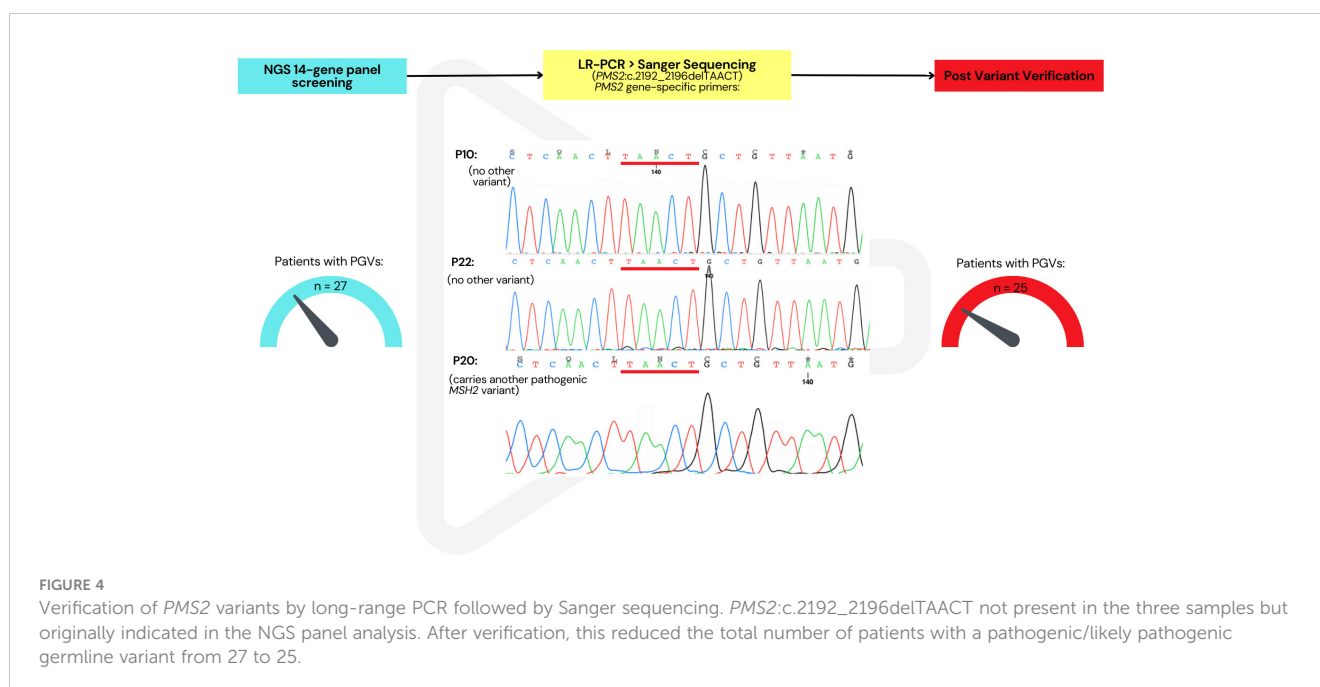


TABLE 5 List of prioritized variants of uncertain significance identified after the NGS multigene panel screening in this study.

Gene	Location	MAF ^a	Function	Coding	Protein
MMR ^b genes	<i>MLH1</i> :exonic:NM_000249.3	0.0	Missense	c.1013A>G	p.Asn338Ser
	<i>MLH1</i> :exonic:NM_000249.4	NA ^c	Missense	c.1772A>G	p.Asp591Gly
	<i>MSH2</i> :exonic:NM_000251.3	NA	Missense	c.508C>G	p.Gln170Glu
	<i>MSH2</i> :exonic:NM_000251.3	NA	Missense	c.157G>T	p.Ala53Ser
	<i>MSH6</i> :exonic:NM_000179.3	NA	Missense	c.560A>G	p.Lys187Arg
	<i>MSH6</i> :exonic:NM_000179.3	NA	Missense	c.2347T>A	p.Cys783Ser
	<i>MSH6</i> :exonic:NM_000179.3	NA	Missense	c.3489A>C	p.Glu1163Asp
	<i>PMS2</i> :exonic:NM_000535.7	0.0	Missense	c.924G>C	p.Glu308Asp
	<i>PMS2</i> :exonic:NM_000535.6	NA	Missense	c.1555T>C	p.Tyr519His
	<i>PMS2</i> :exonic:NM_000535.7	0.001	Frameshift	c.2186_2187delTC	p.Leu729GlnfsTer6
Non-MMR genes	<i>APC</i> :exonic:NM_000038.6	NA	Missense	c.5038C>G	p.Gln1680Glu
	<i>BMPRIA</i> :exonic:NM_004329.3	NA	Missense	c.185A>G	p.Tyr62Cys
	<i>POLE</i> :exonic:NM_006231.4	NA	Missense	c.73G>A	p.Ala25Thr
	<i>POLE</i> :exonic:NM_006231.4	NA	Missense	c.4759G>A	p.Val1587Ile
	<i>POLE</i> :exonic:NM_006231.4	NA	Missense	c.4144C>T	p.Arg1382Cys
	<i>POLE</i> :exonic:NM_006231.4	NA	Missense	c.3901G>A	p.Val1301Met
	<i>POLE</i> :exonic:NM_006231.4	NA	Missense	c.1004T>G	p.Phe335Cys
	<i>POLE</i> :exonic:NM_006231.3	NA	Missense	c.4124C>T	p.Ala1375Val
	<i>POLE</i> :exonic:NM_006231.3	NA	Missense	c.1337G>A	p.Arg446Gln
	<i>POLE</i> :exonic:NM_006231.4	NA	Missense	c.3311C>T	p.Thr1104Met
	<i>POLE</i> :exonic:NM_006231.4	NA	Missense	c.3970C>T	p.Arg1324Cys

^aMinor allele frequency, gnomAD. "NA" = "Not available." ^bMismatch repair.

MSH2 and *MSH6* and was found to carry a germline variant only in *MSH2*. *MSH2* deficiency was probably the first event followed by the *MSH6* deficiency afterwards. This was consistent with the IHC test result for *MSH2*, but *MSH6* needs to be further explored. Another patient previously presented with loss of *MLH1*; however, a likely pathogenic variant was present in *MSH2*, but no variant was detected in *MLH1*. Either *MSH2* is not the cause or IHC staining for *MLH1* might be incorrect, or we did not find the *MLH1* mutation yet (e.g., a copy number variant, an intronic change, or *MLH1* methylation). There were nine patients without any germline variant identified after the NGS panel screening although they previously presented with evidence for loss of at least one MMR gene through IHC testing.

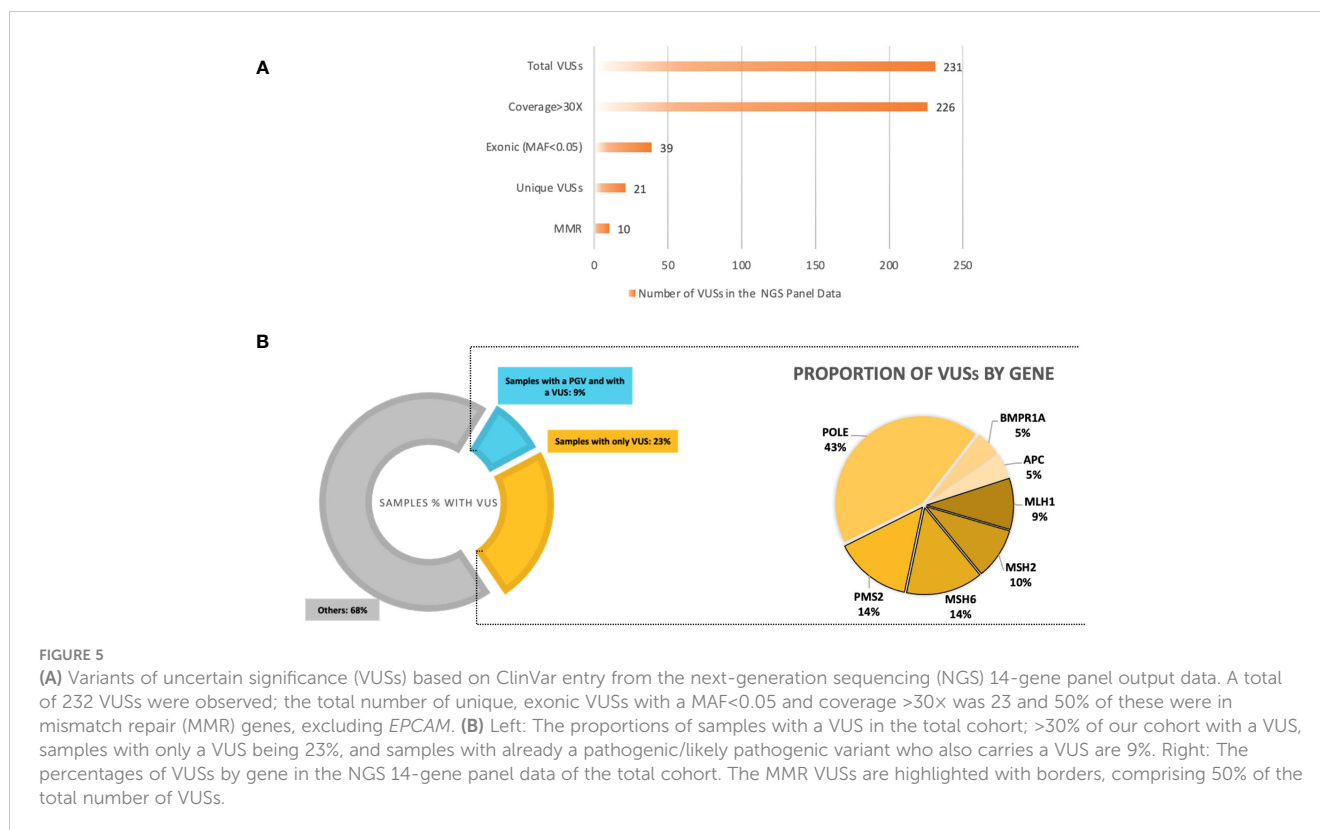
3.4 Verification of the observed germline variants

Sanger sequencing and/or LR-PCR were successful in verifying the presence of all the observed variants, except for one of the *PMS2* variants. LR-PCR was performed for two *PMS2* variants, followed by a second PCR and Sanger sequencing. The variant *PMS2*:c.1579_1580delAG was present in one sample, successfully

verified by LR-PCR (Supplementary Figure 2). However, LR-PCR verification of the *PMS2*:c.2192_2196delTAACT variant showed that this variant was not present in *PMS2* in three patients. The status of the patients who only presented with this "unverified" variant ($n = 2$) remained "unresolved" in our database after this study (Figure 4). Further methodologies are being investigated to finalize the search for disease-causing genetic changes of these cases. Consequently, the number of verified PGVs was 24, found in 25 patients (23.4%).

3.5 VUS identified after the NGS panel screening

The total number of VUSs called, based on the ClinVar entries on the Ion reporter, was 231. Filtering out the VUSs with a coverage less than 30× dropped this to 226. Out of these, 39 VUSs were in an exonic region; after removing duplicates, the final number of unique VUSs was 21 (MAF < 0.05 or not available), present in 35 patients (32.71%) (Table 5; Supplementary Table 3). Ten of these unique VUSs were in an MMR gene; two in *MLH1*, two in *MSH2*, three in *MSH6*, and three in *PMS2*, carried by 19 patients (19/107 = 17.8%).



Nine of the 35 patients who carried a VUS were already found with a PGV (Figure 5).

4 Discussion

We screened a cohort uniquely consisting of indigenous African patients using an NGS multigene panel for CRC as an advance approach relative to other studies among African continental populations. Twenty-five patients (23.4%) were confirmed to carry a PGV associated with the disease after successful verification (24 PGVs in eight different genes: *APC*, *BMPR1A*, *MLH1*, *MSH2*, *MUTYH*, *PMS2*, *POLD1*, and *TP53*), yielding a higher rate of PGVs than the average rate reported in the literature (54–57). These previous studies with similar approach to ours also made use of an NGS multigene panel in an early-onset CRC cohort. To our knowledge, on one hand, the lowest rate of PGVs was 10.1%, using a 32-gene panel (55). On the other hand, the highest rate of PGV identification previously reported was 24.5%, using an 18-gene panel (56). However, these were both in a CRC cohort from a Danish population/ethnic group. It is evident from these two studies that (i) the size of the gene panel for CRC testing does not necessarily correlate with the PGVs' detection rate; (ii) low penetrance genes may influence a low detection rate of PGVs associated with hereditary CRC; and (iii) the cutoff eAOO for inclusion in their cohort may be a factor for a high or low detection rate. Compared to these studies, the detection rate of PGVs in our population-specific CRC multigene panel in a unique and understudied population group, indigenous Africans, was

above the expected threshold. Altogether, these support that our multigene panel comprising 14 genes in a unique indigenous African cohort was ample and yielded in a high rate of PGVs associated with CRC.

More than half of the PGVs identified in our cohort would have been missed using traditional methodologies without the NGS multigene panel in patients admitted to the hospital in the last three decades. Inclusion of eAOO younger than 60 years was found to be an acceptable threshold in our cohort of indigenous African patients that helped us include the unresolved cases in our three-decade records. Selection criteria in the NGS multigene panel studies are generally based on LS-associated features such as family history or Amsterdam/revised Bethesda II criteria, which we did not strictly account for when selecting our cohort (58–60). Our cohort, to some extent, was unselected for most of the LS features except for onset age; we included any unresolved indigenous African patients with an eAOO < 60 years but not < 50 years in our CRC repository. The study by Perkins et al. exemplifies the advantage of an unselected cohort resolving those with inaccurate clinical information recorded who were possibly admitted to the hospital much later than the actual disease onset (58). In our study, patients with an onset of earlier or later than 50 years of age did not show any significant difference for their mutation positivity (Table 2). This was best supported by two of our patients whose disease onset was ≥ 50 years old but presented with a pathogenic, LS-associated MMR variant (Table 3), which could be due to limited access to healthcare resulting in late diagnosis. This study thus adds unique evidence to initiate a genetic testing strategy for this population group affected by CRC,

and to gauge the potential implications on value of clinical surveillance services and early intervention.

The existing literature stresses the need for more advanced genetic screening studies in indigenous Africans, to which our work contributed (61, 62). Our findings helped us confirm the feasibility of our current NGS panel, which could be interrogated as part of a comprehensive strategy for a systematic CRC genetic screening in South Africa. Our data have provided additional insight into and understanding of the hereditary nature of early-onset CRC in indigenous African patients. The high rate of actionable disease-causing variants in a unique indigenous African cohort of early-onset CRC demonstrates the need for extended genetic screening programs and research in African populations.

The prevalence of LS among study cohorts of different CRC populations varies below 10% (63–65). Our cohort's prevalence of LS was 13% (14 out of 107), which resolved the genetic diagnosis for our suspected LS patients. However, two of these patients (patient 16 and patient 18) presented with non-classical LS feature of having a cancer diagnosis after the age of 50 years old. It has been reported that indigenous Africans present with CRC at a young age especially in cases where they are found to carry a PGV associated with LS, and we therefore believe that these two LS patients' relatively older age at cancer diagnosis could be due to either underdiagnosis or, equally as important, other genomic variations in cancer including epigenetic modifiers in this study population (12, 14, 66). Interestingly, only five of the LS patients had a loss of MMR genes reported previously by IHC staining in our laboratories. Without conducting this study, 12 LS patients would have been missed due to lack of or inaccurate IHC testing outcome or insufficient pathological information. This may be due to the unique population group, which may present differently for LS classical features compared to LS in other studied populations such as Europeans and Asians (8, 61, 67, 68). These findings suggest that there may be differences in how LS presents in the same or different populations, and that accurate personalized genetic testing is essential for proper diagnosis. Further research is needed to better understand these differences and improve diagnosis and treatment for LS across all populations.

Identification of novel variants is crucial in terms of identifying founder mutations in especially understudied populations and was expected due to the nature of our cohort being population-specific to indigenous Africans. The NGS 14-gene panel screening consisted of not only novel MMR variants, but also a high proportion of novel non-MMR germline variants. Over 90% of the other PGVs within this cohort were reported only in other population groups including Asian and European sources and not in an African population. Our findings give strong evidence that an advanced genetic screening program in indigenous Africans is a necessity and there is yet much more to explore to provide appropriate personalized treatment and intervention to these understudied indigenous African patients of CRC in South Africa.

Interestingly, as an example of the novel variants, a splice site acceptor pathogenic variant, *MLH1*:c.117-1G>A, was present in two patients (patient 1 and patient 3, Table 3), with a strong evidence on pathogenicity: with the position being well conserved, and the "Human Splicing Finder" predicting a functional effect on

splicing (Supplementary Figure 3) (69). A different variant at this splice site has been reported in a Spanish family and has been proven to disrupt RNA splicing, confirmed with segregation analysis among the CRC-affected family members (70). At the time of the study, the variant at the same position c.117G>T was a VUS but present in the affected family members. RNA analysis by Ruiz et al. confirmed that the variant is disease-causing (70). Further screening of the novel variants in wider CRC patient cohorts will help determine whether the changes we have detected are founders in the indigenous Southern African population.

It is worth noting that identification and verification of *PMS2* variants in CRC NGS multigene panel studies are challenging and require extra effort (36, 71, 72). Most of the studies of NGS panel testing do not include specific methodology for *PMS2* variants to avoid pseudogene interference (5, 57). Our study involved a specific technique for these variants, and we were able to show whether the *PMS2* variants were gene-specific, which gave strength in the verification process. Patients who presented with a *PMS2* PGV and verified by LR-PCR were resolved. In the patients where the gene-specific LR-PCR showed that the variants were absent were marked as unresolved in their molecular diagnosis of CRC in our database.

Our NGS analysis observed some unexpected but critical findings. Of these, one was the pathogenic *MLH1*:c.1731G>A variant identified in patient 20, which was initially filtered out of the analyses as it was categorized as a synonymous variant (by the Ion reporter annotation). However, during a second manual inspection, it was found to be a splice acceptor site variant. This variant has been reported in the literature in multiple affected CRC individuals in unrelated studies, albeit of European (e.g., Germany, Italy, France, and Poland) ancestry (73–76). Identification of this variant highlights the importance of the annotation and filtering procedure and may suggest that an optimized strategy for annotating splice sites variants more accurately is necessary. Secondly, after variant verification, patient 21 was found to carry two pathogenic MMR variants: *MSH2*:c.1923T>A (novel) and *PMS2*:c.2192_2196delTAACT. One of the studies that reported this variant also performed a gene-specific verification technique to avoid pseudogene interference and was able to show the presence of the variant in the actual gene (38). However, our verification by LR-PCR showed that the variant was absent in the actual *PMS2* gene. This is consistent with this patient only having one LS mutation in *MSH2*, which is a more likely/expected event. Importantly, Chong et al. reported an "African" *PMS2* pathogenic variant to be changed to a *PMS2CL* variant, making it a "pseudo" variant (77). Our *PMS2* variants found in this study that failed verification for the presence of the variant in the actual gene are likely to be pseudogene variants rather than a PGV associated with CRC in this population. Additional work is undertaken to finalize these findings.

Another unexpected finding was a non-MMR, but novel pathogenic variant, *POLD1*:c.1265_1266insG found in patient 14 (eAOO <30 years). This finding gives support to the previous published studies that variants in the proofreading activity domain genes may be included in testing for familial CRC, as

candidate variants are being identified in affected individuals (29–31). It serves as a good example of novel *POLD1* pathogenic variants associated with hereditary CRC and may be a population-specific candidate variant.

The number of VUSs identified in our study was higher than expected in the general population, yet is expected in such understudied populations as the indigenous African population in South Africa due to lack of data reported for this population group (18, 78). We also found a high proportion of VUSs in the MMR genes (47%) among the VUSs that suggest pathogenicity at some level, together with changes in other genes, e.g., *POLE*, *APC*, and *BMPRIA*. These variants could be better classified if further testing was available such as tumor sequencing. For instance, patient 24 carries a single PGV in *MUTYH* and, thus, has a small increased risk for CRC, which is known to lead to late cancer progression (23, 79). However, the patient was already diagnosed with CRC at an early age, which progressed in a short time, spreading to the liver. It is worth noting that this patient carries three VUSs in *POLE*. One of these variants, as an example, *POLE:c.3901G>A*, was evaluated, and it could have an effect on pathogenicity when combined with the heterozygous *MUTYH* variant. Based on the ACMG criteria, *POLE:c.3901G>A* is classified as likely pathogenic when it is present with another heterozygous PGV. To confirm this, further investigation and additional information especially involving patient's tumor sequencing would be necessary.

Our study had some limitations. We had difficulties in finding available biological material for DNA extraction for the samples that were either depleted, failed QC, or missing. This led to the exclusion of those patients from the NGS panel screening, and we may have missed resolving these probands and their families. Also, we were only able to include the tumor localization and sidedness but no other pathological information or family history that would help us clearly classify each patient with specific types of hereditary CRC. For instance, only three of the mutation-positive patients in our cohort had a family history (Table 3). Additionally, there may be VUSs unique to the population group in the study to be reclassified as PGV, which require additional evidence and additional studies before their reclassification could possibly be made. The panel testing produced no PGVs, but three VUSs in *MSH6* were carried by five patients (4.7%); this, however, is in line with reports that *MSH6* variants are less prevalent (54). Owing to the insufficient number of studies in African populations, a high percentage of VUSs is reported in literature and is a current drawback in the field (78). Our work adds value to the literature to support that a high proportion of VUSs are often observed in African populations and extensive studies are required to overcome this uncertainty.

Our study suggests some critical future work to overcome the limitations and resolve the remaining cases yet with an unknown molecular diagnosis, which are already in place in our laboratory. These include as follows:

- (i) An expanded variant calling and annotation tool is being implemented to improve the pickup rate of PGVs.

- (ii) The classification of the VUSs could be explored, especially for patients in whom we did not find any germline findings as expected based on previous testing results (e.g., IHC).
- (iii) Those with a very strong clinical background suggesting inheritance yet with no PGV after the NGS screening are being analyzed for copy number variants and methylation analysis and will then undergo whole exome sequencing (WES). Families in which DNA samples (or blood/saliva) of at least two family members (one from each generation) are available will be prioritized for WES.

In conclusion, our study demonstrates that a genetic screening program is an urgent necessity in an African setting for affected CRC patients. In our cohort, almost one in four patients (1:4) had an actionable disease-causing variant and a high percentage of the patients carried a novel PGV that could only be detected by a CRC NGS 14-multigene panel analysis. This study supports the evidence for our hypothesis that investigating the germline variants in affected indigenous African individuals with early-onset CRC can provide more insights and inferential information on the genetic nature of CRC in this population. As part of our ongoing patient support program to translate research into diagnostics, our clinical team will convey these test results to the patients for possible discussion on the intervention and possibilities for cascade testing for their family members. This work adds to the literature and supports the overdue genetic research in indigenous African patients.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1005297>. Accession number: PRJNA1005297.

Ethics statement

The studies involving humans were approved by The Human Research Ethics Committee (HREC), Faculty of Health Sciences, University of Cape Town (UCT). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

SY: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. TM: Data curation, Formal Analysis, Methodology, Writing – review & editing. UA: Data curation, Methodology, Resources, Writing – review & editing. RC: Conceptualization, Investigation, Supervision, Writing – review & editing. GR: Conceptualization, Data curation, Investigation,

Methodology, Project administration, Supervision, Writing – review & editing. PG: Data curation, Resources, Writing – review & editing. RR: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Funding

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant Number 143503). The study is supported by the South African Medical Research Council with funds received from the South African Department of Science and Innovation.

Acknowledgments

We thank the South African National Cancer Registry and the former head, the late Dr. Elvira Singh, for the collaboration, and for providing the DNA samples of patients from the Chris Hani Baragwanath Hospital, Johannesburg.

References

- Wild C, Weiderpass E, Stewart B. *World Cancer Report: Cancer Research for Cancer Prevention*. France: Lyon (2020).
- Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. *Transl Oncol* (2021) 14(10):101174. doi: 10.1016/j.tranon.2021.101174
- Jaspersion KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. *Gastroenterology* (2010) 138(6):2044–58. doi: 10.1053/j.gastro.2010.01.054
- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
- Stoffel EM, Murphy CC. Epidemiology and mechanisms of the increasing incidence of colon and rectal cancers in young adults. *Gastroenterology* (2020) 158(2):341–53. doi: 10.1053/j.gastro.2019.07.055
- Katsidzira L, Gangaidzo I, Thomson S, Rusakaniko S, Matenga J, Ramesar R. The shifting epidemiology of colorectal cancer in sub-Saharan Africa. *Lancet Gastroenterol Hepatol* (2017) 2(5):377–83. doi: 10.1016/S2468-1253(16)30183-2
- Angelo N, Dreyer L. Colorectal carcinoma—a new threat to black patients? A retrospective analysis of colorectal carcinoma received by the Institute for Pathology, University of Pretoria. *S Afr Med J* (2001) 91(8):689–93.
- Bebington B, Singh E, Fabian J, Jan Kruger C, Prodehl L, Surridge D, et al. Design and methodology of a study on colorectal cancer in Johannesburg, South Africa. *JGH Open* (2018) 2(4):139–43. doi: 10.1002/jgh3.12061
- Cronje L, Paterson AC, Becker PJ. Colorectal cancer in South Africa: a heritable cause suspected in many young black patients. *S Afr Med J* (2009) 99(2):103–6.
- Aldera AP, Pillay K, Robertson B, Boutall A, Ramesar R. Genomic landscape of colorectal carcinoma in sub-Saharan Africa. *J Clin Pathol* (2023) 76(1):5–10. doi: 10.1136/jcp-2022-208482
- Katsidzira L, Chokunonga E, Gangaidzo IT, Rusakaniko S, Borok M, Matsena-Zingoni Z, et al. The incidence and histo-pathological characteristics of colorectal cancer in a population based cancer registry in Zimbabwe. *Cancer Epidemiol* (2016) 44:96–100. doi: 10.1016/j.canep.2016.08.001
- Singh E, Underwood JM, Nattey C, Babb C, Sengayi M, Kellett P. South African National Cancer Registry: Effect of withheld data from private health systems on cancer incidence estimates. *S Afr Med J* (2015) 105(2):107–9. doi: 10.7196/samj.8858
- Holla R, Vorster A, Lockett M, De Haas M, Oke OA, Govender D, et al. Immunohistochemical determination of mismatch repair gene product in colorectal carcinomas in a young indigenous African cohort. *S Afr J Surg* (2022) 60(1):28–33. doi: 10.17159/2078-5151/2022/v60n1a3687
- Katsidzira L, Vorster A, Gangaidzo IT, Makunike-Mutasa R, Govender D, Rusakaniko S, et al. Investigation on the hereditary basis of colorectal cancers in an

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2023.1253867/full#supplementary-material>

- African population with frequent early onset cases. *PLoS One* (2019) 14(10):e0224023. doi: 10.1371/journal.pone.0224023
- Choudhury A, Ramsay M, Hazelhurst S, Aron S, Barden S, Botha G, et al. Whole-genome sequencing for an enhanced understanding of genetic variation among South Africans. *Nat Commun* (2017) 8(1):2062. doi: 10.1038/s41467-017-00663-9
 - Ramesar R. African dawn. *Nature* (2015) 517(7534):276–7. doi: 10.1038/nature14077
 - Chambuso R, Robertson B, Ramesar R. A scoring model and protocol to adapt universal screening for lynch syndrome to identify germline pathogenic variants by next generation sequencing from colorectal cancer patients and cascade screening. *Cancers (Basel)* (2022) 14(12):1–11. doi: 10.3390/cancers14122901
 - Ramesar RC ed. Next generation DNA sequencing technologies and implementation science in reducing the impact of familial colorectal cancers in developing countries. In: *IST-Africa Conference (IST-Africa)*. Ireland: Enstitute of Electrical and Electronics Engineers. (2022) 16–20 May 2022.
 - Blokhuys MM, Pietersen GE, Goldberg PA, Algar U, van der Merwe L, Mbatani N, et al. Lynch syndrome: the influence of environmental factors on extracolonic cancer risk in hMLH1 c.C1528T mutation carriers and their mutation-negative sisters. *Fam Cancer* (2010) 9(3):357–63. doi: 10.1007/s10689-010-9334-9
 - Lamola L. *Genomics of Lynch syndrome and Constitutional mismatch repair deficiency syndrome*. Cape Town, South Africa: University of Cape Town (2018).
 - Wentink MQ, Rakers M, Stupart DA, Algar U, Ramesar R, Goldberg PA. Incidence and histological features of colorectal cancer in the Northern Cape Province, South Africa. *S Afr J Surg* (2010) 48(4):109–13.
 - Hampel H, Kalady MF, Pearlman R, Stanich PP. Hereditary colorectal cancer. *Hematol Oncol Clin North Am* (2022) 36(3):429–47. doi: 10.1016/j.hoc.2022.02.002
 - Rebuzzi F, Ulivi P, Tedaldi G. Genetic predisposition to colorectal cancer: how many and which genes to test? *Int J Mol Sci* (2023) 24(3):1–23. doi: 10.3390/ijms24032137
 - Mao R, Krautscheid P, Graham RP, Ganguly A, Shankar S, Ferber M, et al. Genetic testing for inherited colorectal cancer and polyposis, 2021 revision: a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genet Med* (2021) 23(10):1807–17. doi: 10.1038/s41436-021-01207-9
 - Weiss JM, Gupta S, Burke CA, Axell L, Chen LM, Chung DC, et al. NCCN guidelines(R) insights: genetic/familial high-risk assessment: colorectal, version 1.2021. *J Natl Compr Canc Netw* (2021) 19(10):1122–32. doi: 10.1164/jncn.2021.0048
 - Ligtenberg MJ, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, 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(1):112–7. doi: 10.1038/ng.283

27. Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin Genet* (2009) 76(1):1–18. doi: 10.1111/j.1399-0004.2009.01230.x
28. Niessen RC, Hofstra RM, Westers H, Ligtenberg MJ, Kooi K, Jager PO, et al. Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. *Genes Chromosomes Cancer* (2009) 48(8):737–44. doi: 10.1002/gcc.20678
29. Bellido F, Pineda M, Aiza G, Valdes-Mas R, Navarro M, Puente DA, et al. POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. *Genet Med* (2016) 18(4):325–32. doi: 10.1038/gim.2015.75
30. Buchanan DD, Stewart JR, Clendenning M, Rosty C, Mahmood K, Pope BJ, et al. Risk of colorectal cancer for carriers of a germ-line mutation in POLE or POLD1. *Genet Med* (2018) 20(8):890–5. doi: 10.1038/gim.2017.185
31. Jansen AML, Ghosh P, Dakal TC, Slavin TP, Boland CR, Goel A. Novel candidates in early-onset familial colorectal cancer. *Fam Cancer* (2020) 19(1):1–10. doi: 10.1007/s10689-019-00145-5
32. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* (1988) 16(3):1215. doi: 10.1093/nar/16.3.1215
33. Robinson JT, Thorvaldsdottir H, Turner D, Mesirov JP. igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics* (2023) 39(1):1–2. doi: 10.1093/bioinformatics/btac830
34. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, et al. VarSome: the human genomic variant search engine. *Bioinformatics* (2019) 35(11):1978–80. doi: 10.1093/bioinformatics/bty897
35. Li Q, Wang K. InterVar: clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. *Am J Hum Genet* (2017) 100(2):267–80. doi: 10.1016/j.ajhg.2017.01.004
36. Nakagawa H, Lockman JC, Frankel WL, Hampel H, Steenblock K, Burgart LJ, et al. Mismatch repair gene PMS2: disease-causing germline mutations are frequent in patients whose tumors stain negative for PMS2 protein, but paralogous genes obscure mutation detection and interpretation. *Cancer Res* (2004) 64(14):4721–7. doi: 10.1158/0008-5472.CAN-03-2879
37. Senter L, Clendenning M, Sotamaa K, Hampel H, Green J, Potter JD, et al. The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. *Gastroenterology* (2008) 135(2):419–28. doi: 10.1053/j.gastro.2008.04.026
38. van der Klift HM, Tops CM, Bik EC, Boogaard MW, Borgstein AM, Hansson KB, et al. Quantification of sequence exchange events between PMS2 and PMS2CL provides a basis for improved mutation scanning of Lynch syndrome patients. *Hum Mutat* (2010) 31(5):578–87. doi: 10.1002/humu.21229
39. Clendenning M, Hampel H, LaJeunesse J, Lindblom A, Lockman J, Nilbert M, et al. Long-range PCR facilitates the identification of PMS2-specific mutations. *Hum Mutat* (2006) 27(5):490–5. doi: 10.1002/humu.20318
40. Vaughn CP, Robles J, Swensen JJ, Miller CE, Lyon E, Mao R, et al. Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes. *Hum Mutat* (2010) 31(5):588–93. doi: 10.1002/humu.21230
41. Team R. *RStudio: Integrated Development Environment for R*. PBC, Boston, MA: RStudio (2021).
42. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* (2015) 17(5):405–24. doi: 10.1038/gim.2015.30
43. Fan Y, Chen J, Wang W, Wu P, Zhi W, Xue B, et al. Influence of eight unclassified missense variants of the MLH1 gene on Lynch syndrome susceptibility. *Biochem Genet* (2012) 50(1–2):84–93. doi: 10.1007/s10528-011-9467-z
44. Hardt K, Heick SB, Betz B, Goecke T, Yazdanparast H, Kupperts R, et al. Missense variants in hMLH1 identified in patients from the German HNPCC consortium and functional studies. *Fam Cancer* (2011) 10(2):273–84. doi: 10.1007/s10689-011-9431-4
45. Tang R, Hsiung C, Wang JY, Lai CH, Chien HT, Chiu LL, et al. Germ line MLH1 and MSH2 mutations in Taiwanese Lynch syndrome families: characterization of a founder genomic mutation in the MLH1 gene. *Clin Genet* (2009) 75(4):334–45. doi: 10.1111/j.1399-0004.2009.01162.x
46. Trojan J, Zeuzem S, Randolph A, Hemmerle C, Brieger A, Raedle J, et al. Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. *Gastroenterology* (2002) 122(1):211–9. doi: 10.1053/gast.2002.30296
47. Wahlberg S, Liu T, Lindblom P, Lindblom A. Various mutation screening techniques in the DNA mismatch repair genes hMSH2 and hMLH1. *Genet Test* (1999) 3(3):259–64. doi: 10.1089/109065799316563
48. Wolf B, Henglmüller S, Janschek E, Ilencikova D, Ludwig-Papst C, Bergmann M, et al. Spectrum of germ-line MLH1 and MSH2 mutations in Austrian patients with hereditary nonpolyposis colorectal cancer. *Wien Klin Wochenschr* (2005) 117(7–8):269–77. doi: 10.1007/s00508-005-0337-8
49. Huang RL, Chao CF, Ding DC, Yu CP, Chang CC, Lai HC, et al. Multiple epithelial and nonepithelial tumors in hereditary nonpolyposis colorectal cancer: characterization of germline and somatic mutations of the MSH2 gene and heterogeneity of replication error phenotypes. *Cancer Genet Cytogenet* (2004) 153(2):108–14. doi: 10.1016/j.cancergencyto.2004.01.003
50. Lotsari JE, Gylling A, Abdel-Rahman WM, Nieminen TT, Aittomäki K, Friman M, et al. Breast carcinoma and Lynch syndrome: molecular analysis of tumors arising in mutation carriers, non-carriers, and sporadic cases. *Breast Cancer Res* (2012) 14(3):R90. doi: 10.1186/bcr3205
51. Mangold E, Pagenstecher C, Friedl W, Fischer HP, Merkelbach-Bruse S, Ohlendorf M, et al. Tumours from MSH2 mutation carriers show loss of MSH2 expression but many tumours from MLH1 mutation carriers exhibit weak positive MLH1 staining. *J Pathol* (2005) 207(4):385–95. doi: 10.1002/path.1858
52. Rossi BM, Palmero EI, Lopez-Kostner F, Sarroca C, Vaccaro CA, Spirandelli F, et al. A survey of the clinicopathological and molecular characteristics of patients with suspected Lynch syndrome in Latin America. *BMC Cancer* (2017) 17(1):623. doi: 10.1186/s12885-017-3599-4
53. Jiang W, Li L, Ke CF, Wang W, Xiao BY, Kong LH, et al. Universal germline testing among patients with colorectal cancer: clinical actionability and optimised panel. *J Med Genet* (2022) 59(4):370–6. doi: 10.1136/jmedgenet-2020-107230
54. Coughlin SE, Heald B, Clark DF, Nielsen SM, Hatchell KE, Esplin ED, et al. Multigene panel testing yields high rates of clinically actionable variants among patients with colorectal cancer. *JCO Precis Oncol* (2022) 6:e2200517. doi: 10.1200/PO.22.00517
55. Djursby M, Madsen MB, Frederiksen JH, Berchtold LA, Therkildsen C, Willemoe GL, et al. New pathogenic germline variants in very early onset and familial colorectal cancer patients. *Front Genet* (2020) 11:566266. doi: 10.3389/fgene.2020.566266
56. Frostberg E, Petersen AH, Bojesen A, Rahr HB, Lindebjerg J, Ronlund K. The prevalence of pathogenic or likely pathogenic germline variants in a nationwide cohort of young colorectal cancer patients using a panel of 18 genes associated with colorectal cancer. *Cancers (Basel)* (2021) 13(20):1–12. doi: 10.3390/cancers13205094
57. Uson PLS Jr., Riegert-Johnson D, Boardman L, Kiesel J, Mountjoy L, Patel N, et al. Germline cancer susceptibility gene testing in unselected patients with colorectal adenocarcinoma: A multicenter prospective study. *Clin Gastroenterol Hepatol* (2022) 20(3):e508–e28. doi: 10.1016/j.cgh.2021.04.013
58. Perkins AT, Haslem D, Goldsberry J, Shortt K, Sittig L, Raghunath S, et al. Universal germline testing of unselected cancer patients detects pathogenic variants missed by standard guidelines without increasing healthcare costs. *Cancers (Basel)* (2021) 13(22):1–10. doi: 10.3390/cancers13225612
59. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* (2004) 96(4):261–8. doi: 10.1093/jnci/djh034
60. Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology* (1999) 116(6):1453–6. doi: 10.1016/s0016-5085(99)70510-x
61. Alatisse OI, Knapp GC, Bebington B, Ayodeji P, Dare A, Constable J, et al. Racial differences in the phenotype of colorectal cancer: A prospective comparison between Nigeria and South Africa. *World J Surg* (2022) 46(1):47–53. doi: 10.1007/s00268-021-06304-z
62. Motsuku L, Chen WC, Muchengeti MM, Naidoo M, Quene TM, Kellett P, et al. Colorectal cancer incidence and mortality trends by sex and population group in South Africa: 2002–2014. *BMC Cancer* (2021) 21(1):129. doi: 10.1186/s12885-021-07853-1
63. Hampel H, Yurgelun MB. Point/counterpoint: is it time for universal germline genetic testing for all GI cancers? *J Clin Oncol* (2022) 40(24):2681–92. doi: 10.1200/JCO.21.02764
64. Neben CL, Zimmer AD, Stedden W, van den Akker J, O'Connor R, Chan RC, et al. Multi-gene panel testing of 23,179 individuals for hereditary cancer risk identifies pathogenic variant carriers missed by current genetic testing guidelines. *J Mol Diagn* (2019) 21(4):646–57. doi: 10.1016/j.jmoldx.2019.03.001
65. Pearlman R, Frankel WL, Swanson BJ, Jones D, Zhao W, Yilmaz A, et al. Prospective Statewide study of universal screening for hereditary colorectal cancer: the Ohio colorectal cancer prevention initiative. *JCO Precis Oncol* (2021) 5:779–91. doi: 10.1200/PO.20.00525
66. Boland CR, Goel A, Patel SG. The genetic and epigenetic landscape of early-onset colorectal cancer. *Colorectal Cancer* (2020) 9(3):CRC23. doi: 10.2217/crc-2020-0005
67. Holowatyj AN, Maude AS, Musa HS, Adamu A, Ibrahim S, Abdullahi A, et al. Patterns of early-onset colorectal cancer among Nigerians and African Americans. *JCO Glob Oncol* (2020) 6:1647–55. doi: 10.1200/GO.20.00272
68. McCabe M, Penny C, Magangane P, Mirza S, Perner Y. Left-sided colorectal cancer distinct in indigenous African patients compared to other ethnic groups in South Africa. *BMC Cancer* (2022) 22(1):1089. doi: 10.1186/s12885-022-10185-3
69. Desmet FO, Hamroun D, Lalonde M, Colod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* (2009) 37(9):e67. doi: 10.1093/nar/gkp215
70. Ruiz JL, Alvarez-Cubero MJ, Rosado FF, Espin EM, Bernal CE. Confirmed pathogenic effect of a splice site variation in the MLH1 gene causing Lynch syndrome. *Int J Colorectal Dis* (2014) 29(8):1019–20. doi: 10.1007/s00384-014-1890-2

71. Okkels H, Lagerstedt-Robinson K, Wikman FP, Hansen TVO, Lolas I, Lindberg LJ, et al. Detection of PMS2 mutations by screening hereditary nonpolyposis colon cancer families from Denmark and Sweden. *Genet Test Mol Biomarkers* (2019) 23(9):688–95. doi: 10.1089/gtmb.2018.0316
72. van der Klift HM, Mensenkamp AR, Drost M, Bik EC, Vos YJ, Gille HJ, et al. Comprehensive mutation analysis of PMS2 in a large cohort of probands suspected of lynch syndrome or constitutional mismatch repair deficiency syndrome. *Hum Mutat* (2016) 37(11):1162–79. doi: 10.1002/humu.23052
73. Kurzawski G, Suchy J, Lener M, Kluszo-Grabowska E, Kladny J, Safranow K, et al. Germline MSH2 and MLH1 mutational spectrum including large rearrangements in HNPCC families from Poland (update study). *Clin Genet* (2006) 69(1):40–7. doi: 10.1111/j.1399-0004.2006.00550.x
74. Pagenstecher C, Wehner M, Friedl W, Rahner N, Aretz S, Friedrichs N, et al. Aberrant splicing in MLH1 and MSH2 due to exonic and intronic variants. *Hum Genet* (2006) 119(1-2):9–22. doi: 10.1007/s00439-005-0107-8
75. Simbolo M, Mafficini A, Agostini M, Pedrazzani C, Bedin C, Urso ED, et al. Next-generation sequencing for genetic testing of familial colorectal cancer syndromes. *Hered Cancer Clin Pract* (2015) 13(1):18. doi: 10.1186/s13053-015-0039-9
76. Tournier I, Vezain M, Martins A, Charbonnier F, Baert-Desurmont S, Olschwang S, et al. A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Hum Mutat* (2008) 29(12):1412–24. doi: 10.1002/humu.20796
77. Chong AS, Chong G, Foulkes WD, Saskin A. Reclassification of a frequent African-origin variant from PMS2 to the pseudogene PMS2CL. *Hum Mutat* (2020) 41(4):749–52. doi: 10.1002/humu.23978
78. Ndugga-Kabuye MK, Issaka RB. Inequities in multi-gene hereditary cancer testing: lower diagnostic yield and higher VUS rate in individuals who identify as Hispanic, African or Asian and Pacific Islander as compared to European. *Fam Cancer* (2019) 18(4):465–9. doi: 10.1007/s10689-019-00144-6
79. Farrington SM, Tenesa A, Barnetson R, Wiltshire A, Prendergast J, Porteous M, et al. Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *Am J Hum Genet* (2005) 77(1):112–9. doi: 10.1086/431213