

Pan-cancer prevalence of microsatellite instability and Lynch syndrome in India

Harsh Sheth

harsh.sheth@frige.co.in

Foundation for Research in Genetics and Endocrinology

Jyoti Sadhwani

Foundation for Research in Genetics and Endocrinology

Bhawna Sirohi

Vedanta Medical Research Foundation (BALCO Medical Centre)

Senthilkumar Ramasamy

Vedanta Medical Research Foundation (BALCO Medical Centre)

Yashwant Kashyap

Vedanta Medical Research Foundation (BALCO Medical Centre)

Pankaj Shah

Zydus Cancer Hospital

Mithun Shah

Zydus Cancer Hospital

Suresh Advani

Sushrut Hospital

Lidiya Thomas

Sushrut Hospital

Vipul D. Yagnik

Nishtha Surgical Hospital & Research Centre

Avinash Tank

Dwarika Hospital

Chirag Shah

Mission Gastro Hospital

Raja Pramanik

All India Institute of Medical Sciences

Bhavesh Thakkar

Marengo CIMS Hospital

Darshan Bhansali

Marengo CIMS Hospital

Manish Gandhi

Marengo CIMS Hospital

Tarang Patel

Marengo CIMS Hospital

Natoo Patel

Marengo CIMS Hospital

Ashok Patel

Marengo CIMS Hospital

Ruchir Patel

Gastro1 Hospital

Ravindra Gaadhe

Gastroplus Hospital

Chintan Shah

HOC Vedanta

Bhavin Shah

HOC Vedanta

Gaurang Modi

Oncowin Cancer Center

Ankita Jakhar

Gujarat Cancer & Research Institute

Ava Desai
Zydus Cancer Hospital

Chirantan Bose
MOC Cancer Care & Research Centre

Amit Sehrawat
All India Institute of Medical Sciences Rishikesh

Dipak Limbachiya
Eva Women's Hospital

Chandni Patel
Foundation for Research in Genetics and Endocrinology

Prachi Soni
Foundation for Research in Genetics and Endocrinology

Sunil Trivedi
Foundation for Research in Genetics and Endocrinology

Kshipra Chauhan
Gujarat Technological University

Frenny Sheth
Foundation for Research in Genetics and Endocrinology

Jayesh Sheth
Foundation for Research in Genetics and Endocrinology

D. Timothy Bishop
Leeds Institute of Medical Research, University of Leeds

John Burn
Translational and Clinical Research Institute, Newcastle University

Abhinav Jain
Gastro1 Hospital

Research Article

Keywords: Lynch syndrome, Microsatellite instability, Mismatch repair deficiency, Founder variant, Tumour first screening

Posted Date: June 30th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9894863/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Familial Cancer on October 7th, 2026. See the published version at <https://doi.org/10.1007/s10689-026-00607-7>.

Abstract

Lynch syndrome (LS) is underdiagnosed in India, particularly outside colorectal cancer (CRC), because pan-cancer data on microsatellite instability (MSI), mismatch repair deficiency (MMRd), and germline predisposition remain limited. We sequentially recruited 519 patients with CRC and extra-colonic cancers from 17 centres in India between 2019 and 2025, and applied a tumour-first approach using MSI testing by PCR-fragment length analysis (PCR-FLA) across all tumours, MMR immunohistochemistry (MMR-IHC) in extra-colonic tumours, *BRAF* V600E testing in MSI-high CRCs, and germline whole exome sequencing in MSI-high/*BRAF* wildtype CRCs and MSI-high/MMRd extra-colonic tumours. MSI-high/MMRd status was identified in 25.6% (133/519) tumours, and 49.6% (66/133) patients harboured pathogenic or likely pathogenic germline variants consistent with LS. CRC showed the highest LS yield, followed by endometrial and ovarian cancers, whereas, several extra-colonic tumour groups showed lower yields, in part reflecting cohort composition and tumour-specific MSI/MMRd prevalence. *MLH1* accounted for most LS diagnoses (69.7%, 46/66) followed by *MSH2* (18.2%, 12/66). Two recurrent *MLH1* variants, c.306G > T and c.156delA, together accounted for 36.4% of all LS cases. Haplotype analysis supported c.156delA as a novel founder variant originating in Gujarat. Concordance between PCR-FLA and MMR-IHC in extracolonic tumours was 93.8%, supporting combined tumour-based screening in pan-cancer workflows. These findings establish the clinical utility of pan-cancer tumour-first LS detection in India and support broader implementation of reflex MSI/MMRd testing, germline confirmation, and cascade testing to improve surveillance, chemoprevention, and risk-reducing management in India.

Introduction

Lynch syndrome (LS), previously referred to as hereditary nonpolyposis colorectal cancer, is one of the most common hereditary cancer predisposition syndrome, with an estimated population prevalence of approximately 1 in 279 individuals in the general population [1–3]. It is inherited in an autosomal dominant pattern and results from germline pathogenic or likely pathogenic (P/LP) variants in the mismatch repair (MMR) genes- *MLH1*, *MSH2*, *MSH6*, *PMS2* and *EPCAM* - which play critical role in maintaining genomic integrity by correcting errors that occur during DNA replication [4]. Loss of MMR function permits accumulation of replication errors, particularly within short repetitive DNA sequences, producing the characteristic tumour phenotype of microsatellite instability (MSI) [5]. Although LS is classically associated with colorectal and endometrial cancers, the syndrome encompasses a broader tumour spectrum that includes ovarian and other gastrointestinal malignancies, with gene-specific differences in penetrance, age at onset, and tumour distribution [6]. In particular, variants in *MLH1* and *MSH2* are associated with higher cancer risks and earlier age of onset, whereas *MSH6* and *PMS2* exhibit comparatively lower penetrance and later presentation [7].

Despite its clinical importance, LS remains substantially underdiagnosed. Recent population based genomic studies suggest that approximately 63.2% of individuals carrying pathogenic mismatch repair gene variants remain undetected in routine care [8]. Similar concerns have been raised in health systems with established hereditary cancer pathways, where many affected individuals remain undetected and therefore do not benefit from risk-reducing surveillance, chemoprevention, or cascade testing in relatives [9–14]. For this reason, tumour-based screening using MSI testing and/or mismatch repair immunohistochemistry (MMR-IHC) has become central to LS case-finding, particularly in colorectal and endometrial cancer. However, the performance of these pathways depends not only on the availability of testing, but also on appropriate reflex analysis and timely germline follow-up.

An additional challenge is that LS is not confined to the classical colorectal-endometrial presentation. Pan-cancer analyses have shown that MMR-deficient tumours and germline MMR pathogenic variants occur across a wider range of malignancies than previously appreciated, including small bowel, gastric, pancreatic, biliary, urothelial, sebaceous, and other extracolonic tumours [15–17]. These data suggest that approaches focused only on family history or on a narrow set of index tumours may fail to identify a meaningful proportion of LS carriers. At the same time, differences in tumour spectrum observed between cohorts may reflect not only biology, but also case ascertainment, referral pathways, tumour mix, and the tumour-first workflow used to trigger germline testing. A molecular epidemiology framework is therefore needed when interpreting LS detection across tumour sites.

Population-specific factors may further modify the observed distribution of LS. Founder effects, demographic history, and regional clustering of pathogenic variants can influence both the relative contribution of different MMR genes and the tumour types through which carriers are first ascertained [18, 19]. Recent data from the Prospective Lynch Syndrome Database have highlighted important gene-specific differences in cancer risk and clinical outcomes among LS carriers, underscoring the need for population-relevant data to inform detection strategies and downstream management [7, 20, 21]. In settings where recurrent founder variants are present, the apparent tumour distribution of LS in clinic-based cohorts may differ materially from that reported in predominantly European populations.

In the Indian context, robust estimates of LS and MSI/MMRd remains limited due to the majority of the studies focussing on colorectal cancer cohorts. While national guidelines emphasize structured management of colorectal cancer, integration of universal mismatch repair (MMR) testing and hereditary cancer testing is not yet implemented in the clinical settings [22]. Additionally, multicentre data spanning both colorectal and extracolonic index tumours, with systematic tumour-based screening and germline confirmation, are sparse. As a result, the frequency of MSI/MMRd across commonly encountered solid tumours in Indian practice, the germline yield of a tumour-first approach, and the contribution of recurrent or founder MMR variants remain insufficiently defined. This is a clinically relevant gap because access to molecular testing is heterogeneous, extracolonic tumours are less consistently incorporated into LS evaluation pathways, and regional founder variants may influence both detection yield and apparent tumour spectrum.

In this prospective multicentre hospital-based cohort, we therefore assessed MSI and/or MMR deficiency across a range of commonly observed solid tumours recruited in India, evaluated the yield of germline LS detection following tumour triage, and examined evidence for recurrent founder variants in MMR genes. By doing so, we aimed to define the molecular epidemiology of MSI/MMRd associated LS detection in Indian clinical cohort and to clarify how tumour site and population specific variant structure may share case ascertainment.

Materials and methods

Study population

A total of 519 patients with primary diagnosis of cancer in any one of the anatomical sites- colorectum, endometrium, ovaries, oesophagus, stomach, liver, pancreas, appendix, prostate, cervix and small intestine- were sequentially recruited from 17 centres between 2019 and 2025. Clinical parameters such as age at cancer diagnosis, tumour site and histopathology, together with family history details were collected from all patients. Tumour site for colorectal cancer was divided as follows: (a) right sided or proximal colon which includes caecum, ascending colon, and hepatic flexure; (b) transverse colon; (c) left sided or distal colon which included splenic flexure, descending colon and sigmoid colon; and (d) rectum. Written informed consent was collected from all patients in accordance with the Helsinki declaration. Additionally, first degree relatives of the patients diagnosed with LS (with or without prior cancer history) were invited for germline testing of which 42 members participated. The study was approved by the ethics committee at FRIGE's Institute of Human Genetics (FRIGE/IEC/23/2024), CIMS Hospital, BALCO Medical Centre, Chattisgarh (BMCIRB/2024/03), AIIMS, Delhi (AIIMS1322/03.05.2024, RP-20/2024), and AIIMS, Rishikesh (AIIMS/IEC/25/206).

Sample collection and DNA extraction

Formalin-fixed paraffin-embedded (FFPE) tumour tissue blocks or fresh tumour tissue biopsies and matched normal blood samples were obtained from all participating patients. For each selected block, five to ten 5µm sections were cut using fresh blades to prevent cross-contamination. Sections were placed into sterile tubes and deparaffinized using three xylene washes followed by a 100% ethanol wash and air-dried prior to DNA extraction.

Genomic DNA from FFPE and fresh tumour tissue was extracted using the QIAamp DNA FFPE Tissue Kit (Qiagen, Germany) and DNeasy blood and tissue kit (Qiagen, Germany), respectively, following the manufacturer's protocol. Briefly, tissue was incubated overnight with proteinase K in ATL buffer at 56°C, followed by heating at 90°C for 1 hour to reverse formalin cross-linking. DNA was isolated using silica membrane spin columns and eluted in AE buffer. Genomic DNA from the peripheral blood sample was extracted using the desalting method [23]. DNA quantity was measured using the Qubit dsDNA HS Assay Kit (Thermo Fisher, USA), and quality assessment was performed using QiaXpert (Qiagen, Germany).

Microsatellite instability analysis

MSI testing was performed on genomic DNA from tumour and matched normal samples of all 519 patients. MSI testing was carried out using the Promega MSI Analysis System v1.2 (Promega, USA), which includes five mononucleotide repeat markers (BAT-25, BAT-26, NR-21, NR-24, MONO-27) and two pentanucleotide markers (Penta C and Penta) for sample identification. PCR amplification was conducted using 2ng of template DNA according to the manufacturer's protocol. PCR products were denatured with Hi-Di Formamide, and were analysed via capillary electrophoresis on the ABI SeqStudio Genetic Analyzer (Thermo Fisher Scientific, USA) using ILS600 as the size marker. Data from the SeqStudio instrument were interpreted using GeneMapper v5.0 (RRID:SCR_014290). Tumours were classified as microsatellite unstable (MSI-High) if ≥ 2 markers showed instability.

Immunohistochemistry

MMRd testing was performed across all non-colorectal cancer (non-CRC) tumours, based on the prior evidence of inferior performance of PCR based MSI testing in non-CRC tumours [24, 25]. Immunohistochemistry (IHC) for mismatch repair (MMR) proteins was performed on 4µm FFPE tissue sections using the Ventana BenchMark ULTRA automated platform (Roche, USA), as per the manufacturer's protocol. Slides were deparaffinized and subjected to antigen retrieval using Cell Conditioning 1 buffer at 95°C for 64 minutes. Primary antibodies included MLH1 (clone M1) (Roche Cat# 760–5091, RRID:AB_3669002), PMS2 (clone A16-4) (Roche Cat# 790–5094, RRID:AB_3669003), MSH2 (clone G219-1129) (Roche Cat# 790–5093, RRID:AB_2936886), and MSH6 (clone SP93) (Roche Cat# 790–5092, RRID:AB_2936885). Detection was carried out using the OptiView DAB IHC Detection Kit (Ventana), and counterstaining was done with hematoxylin. Complete loss of nuclear staining in tumour cells, with preserved staining in internal controls, was considered MMRd. Each result was evaluated by a pathologist.

BRAF V600E analysis

In all colorectal cancer cases showing MSI-High status, somatic *BRAF* V600E variant analysis was performed to distinguish sporadic from LS tumour, in accordance with the NICE guidelines (DG27). *BRAF* V600E variant was detected using ARMS-PCR [26]. The *BRAF* V600E ARMS-PCR assay contained 4 primers in a single PCR reaction tube with the two outside primers designed to amplify a common fragment of 200 bp flanking the variant site, which served as an internal amplification control. The two internal primers were either specific for the mutant sequence or the wildtype sequence. The mutant and wildtype sequences were distinguishable by the difference of fragment size, as the amplified wildtype fragment was 97 bp, while the mutant fragment was 144 bp. The amplification protocol was carried out in the manner as described by the author [26].

Germline whole exome sequencing

Only those CRC tumours with MSI-high and *BRAF* V600E negative status were analysed using germline whole exome sequencing (WES) for detection LS, as per the NICE DG27 guidelines. Furthermore, any non-CRC tumours with MMRd status were tested using germline WES to identify LS. Genomic DNA was subjected to selective capture and sequencing of the protein coding regions that included exons and exon-intron boundaries of genes using Twist Bioscience's Human Core Exome V2 kit (Twist Bioscience, USA). Sequencing was performed either on the NovaSeq X Plus, NovaSeq 6000, or HiSeq platforms (Illumina, USA) at an average depth of 80-100x. Sequenced reads were aligned against the human reference genome GRCh38/hg38 using BWA-MEM (v0.7.12) (RRID:SCR_022192) [27], followed by deduplication, base quality score recalibration, and single nucleotide variant (SNV) and small indel calling using the GATK HaplotypeCaller (v4.12) in accordance with the GATK's best practice guidelines [28].

Variant annotation, filtration, and prioritisation was performed using Exomiser v13.1.0 (RRID:SCR_002192) [29] Exomiser uses the hiPHIVE prioritisation method that incorporates mode of inheritance (recessive and dominant), protein-protein interaction networks, and multi-species ontologies along with

ranking candidate genes based on the predicted variant pathogenicity associated with the phenotype. The phenotype information was coded in uniform HPO terminologies [30]. Common variants were filtered based on the minor allele frequency in the 1000Genomes Phase 3, TopMed, and gnomAD v4 databases. The minor allele frequency cut-off was set at 0.02 (2%); which was set assuming monogenic recessive as well as dominant inherited or *de novo* forms of hereditary cancer predisposition syndromes, especially commonly occurring founder variants in MMR genes. Only non-synonymous variants with a depth of > 20x were used for analysis and clinical correlation. Various *in silico* prediction tools such as REVEL [31], MVP [32], SIFT (RRID:SCR_012813) [33], MutationTaster2 [34], PolyPhen-2 (PolyPhen: Polymorphism Phenotyping (RRID:SCR_013189)) [35], and CADD [36] were used to predict variant pathogenicity. Post gross filtering, variants were prioritised based on the following: (a) known disease causing variants previously reported in the ClinVar (RRID:SCR_006169) [37] and/ or InSiGHT databases [38]; and (b) novel variants in known genes based on the Z-score for missense and pLOUEF score for loss of function variants available in the gnomAD database. All candidate variants were assessed using the Integrative Genomics Viewer (IGV) [39] to evaluate their quality.

Copy number variant (CNV) calling was performed across the exome data using the CNVRobot (v4.1) tool (<https://github.com/AnetaMikulasova/CNVRobot>), which uses a custom GATK 4-based pipeline. At least five “normal” samples were used as controls for the normalisation step. CNVs were annotated using AnnotSV [40]. In the case of candidate CNVs, variants were primarily screened for population frequency and known disease associations using publicly available databases like the gnomAD database [41], DGV [42], and OMIM [43]. CNVs present in more than 1% of the samples of the aforementioned datasets and present in more than 10% of the patients were excluded from the analysis. CNVs were validated using SYBR green based quantitative PCR using HiMedia's Insta Q48 Real Time PCR system (HiMedia Labs, India). CNVs encompassing the *PMS2* gene were orthogonally validated using long-range PCR, as previously described [44], due to the presence of the pseudogene.

The identified SNVs and CNVs were classified according to the American College of Medical Genetics- American College of Pathologists (ACMG-AMP) guidelines and the ClinGen framework [45, 46]. Carrier screening was carried out using Sanger sequencing or quantitative-PCR for SNV and CNV, respectively, for first degree family members (parents, siblings and children) who provided their consent.

Haplotype analysis for *MLH1* c.156delA variant

Two recurrent variants c.306G > T (p.Glu102Asp) and c.156delA (p.Glu53ArgfsTer4) in the *MLH1* gene were noted in the current study. We have previously demonstrated *MLH1* c.306G > T to be a founder variant in India [19]. To assess *MLH1* c.156delA as a putative founder variant, 24 individuals being carriers for this variant (rsID: rs587778927/rs63750028) underwent genome-wide genotyping using the Illumina Infinium Global Screening Array v3.0 (Illumina, USA). Genotyping was conducted as per the manufacturer's protocol, and data was analyzed using Illumina's DRAGEN array v1.3.0 (Illumina, USA). Genotype data was subjected to standard quality control, including filtering for missing alleles and minor allele frequency thresholds. VCF files were exported and merged into a singular multi-sample VCF using bcftools using GRCh38/hg38 reference genome build.

Pairwise LD was computed for the variant rs587778927/rs63750028 to identify any SNPs in linkage disequilibrium (LD) within ± 1 Mb region of the variant. As all carriers were observed to have originated from a singular geographic location (Gujarat) in India, genetic relatedness between these individuals was estimated using pairwise identity-by-descent (IBD) analysis using PLINK (RRID:SCR_001757) [47]. The -genome function was used to calculate pairwise PI_HAT values, representing the proportion of the genome shared by IBD. To reduce bias due to linked SNPs, LD pruning was performed using PLINK ('--indep-pairwise 50 5 0.2'). This step ensures that only independent SNPs are used for IBD estimation.

Subsequently, these VCFs were phased and analyzed using BEAGLE v5.4 (RRID:SCR_001789) [48] to identify pairwise IBD segments. The analysis focused on segments longer than 1 cM, which were considered reliable for detecting recent shared ancestry [49]. To estimate the time to the most recent common ancestor (TMRCA), IBD segment data were further analyzed using BEAGLE's Refined IBD algorithm [50]. Refined IBD also measures relatedness, but unlike genome-wide values such as PI_HAT, its estimates are segment-specific [50]. Each IBD segment was assigned a LOD score, which did or did not support identity-by-descent. Higher LOD scores indicated stronger evidence for true IBD sharing and usually segments with LOD > 3 are considered reliable and are retained for downstream analysis [49]. TMRCA for each shared segment was estimated as: $g = 1/(2L)$, where g is the number of generations and L is the segment length in Morgans [49, 51]. Visualisation of the shared haplotype structure surrounding the variant rs587778927 in the phased VCF was carried out manually in the region spanning ± 1 Mb around the variant (Variant genomic co-ordinate: chr3:36996655TA > T). For each individual, putative haplotypes were constructed manually by aligning phased alleles for the selected informative markers. The resulting haplotypes were represented as color-coded matrices to visually identify shared segments across all cases.

Statistical analysis

For all proportion estimations, 95% confidence intervals (CI) were calculated using the Clopper-Pearson exact method. For comparison of proportions between two dichotomous variables, Pearson's chi-square test (χ^2) statistic was used. For assessment of concordance between two assays, Cohen's κ statistic was used. A two-sided p-value of < 0.05 was considered statistically significant. All data were analysed using Stata v18 (RRID:SCR_012763) (Stata Corp, USA). Graphs were created in R v4.5.2 (R Project for Statistical Computing (RRID:SCR_001905)) using the “ggplot” function. Anonymised clinical, MSI, MMRd, and germline Lynch syndrome results are available in Supplementary File 1.

Results

Clinical details of the patient cohort

A total of 519 patients representing a broad spectrum of solid tumours in any of the anatomical sites: colorectum, endometrium, ovaries, oesophagus, stomach, liver, pancreas, appendix, prostate, cervix and small intestine; were sequentially recruited on the study (Fig. 1). The mean age at diagnosis across

the cohort was 52.9 ± 13.9 years. Colorectal cancers constituted the largest subgroup, accounting for 56.8% of cases (295/519). Within this group, tumours of the descending colon were most frequent (116/295, 39.3%), followed by the ascending colon (91/295, 30.8%), rectum (71/295, 24.1%), and transverse colon (17/295, 5.8%). Amongst extracolonic tumours, gastric cancer comprised 14.6% (76/519) of the cohort, followed by ovarian (46/519, 8.9%), endometrial (40/519, 7.7%), and oesophageal cancers (32/519, 6.2%). Less frequent tumour types included pancreatic (13/519, 2.5%), small intestinal (4/519, 0.8%), hepatobiliary (3/519, 0.6%), appendiceal (3/519, 0.6%), prostate (2/519, 0.4%), and cervical cancers (1/519, 0.2%), as well as a small number of stomach-oesophagus and endometrium-ovary synchronous tumours (Fig. 2, Table 1).

Table 1
Demographic characteristic of the 519 cancer patient cohort

Clinical details (N = 519)	n (%)
Age at diagnosis (years)	
≤ 25	16 (3.08)
> 25 and ≤ 50	207 (39.88)
> 50 and ≤ 75	280 (53.95)
> 75	16 (3.08)
Sex	
Female	230 (44.32)
Male	289 (55.68)
Cancer site	
Descending colon	116 (22.35)
Ascending colon	91 (17.53)
Transverse colon	17 (3.28)
Rectum	71 (13.68)
Endometrium	40 (7.71)
Ovarian	46 (8.86)
Stomach	76 (14.64)
Oesophagus	32 (6.17)
Small intestine	4 (0.77)
Liver	3 (0.58)
Pancreas	13 (2.50)
Appendix	3 (0.58)
Prostate	2 (0.39)
Endometrial and ovarian (synchronous)	3 (0.58)
Stomach and oesophageal (synchronous)	1 (0.19)
Cervical	1 (0.19)
Synchronous or metachronous cancer	
Yes	19 (3.7)
No	500 (96.3)
Family history of cancer	
Yes	88 (16.9)
No	431 (83.04)

Analysis of recruitment by geographical site showed higher recruitment of CRCs from Gujarat and Maharashtra versus stomach and ovarian cancers from Chattisgarh (Supplementary File 1). 3.7% (19/519) of the patients had synchronous or metachronous cancer (Table 1). 16.9% (88/519) of the patients had family history of cancer whereby either their first degree relative had a history of cancer at less than 50 years of age or their first- and second-degree relatives were diagnosed with one of the Lynch syndrome associated cancers regardless of age at diagnosis (Table 1).

Distribution and MSI and MMRd across tumour types

Of 519 tumours assessed in the study, MSI result by PCR-FLA was available for 96.3% (500/519) of the cases. Poor DNA quality/ poor amplification, unavailability of matched normal samples, or unavailability of tumour blocks were observed in 73.7% (14/19), 5.3% (1/19), and 15.8% (3/19) of the cases, respectively. In all 19 cases, MMR-IHC was conducted either during the study or was conducted previously by the referring clinician (Supplementary File 1).

In the remaining 500 tumours, 24.8% (124/500; 95% CI: 21–29%) were MSI-high and 75.2% (376/500; 95% CI: 71–79%) were microsatellite stable (MSS) (Fig. 3a). The highest proportion of MSI-high was noted in CRC, endometrium, and gastric tumours. Whilst both small intestine and prostate tumours showed 50% proportion for MSI-high signature, they had a small sample size thereby limiting robustness in estimation of MSI-high proportion. All tumours were MSS for pancreas, cervical, and appendix. In accordance with the NICE diagnostic pathway (DG27) for LS detection, somatic *BRAF* V600E testing was carried out in 91.8% (89/97) of MSI-high CRC tumours, with the remaining 8 samples not being tested due to paucity of tumour DNA sample. Of 89 MSI-high CRC samples, *BRAF* V600E somatic variant was detected in 4.5% (4/89) and absent in 95.5% (85/89) cases. Germline testing for LS detection was therefore conducted in 85 MSI-high *BRAF* V600E negative cases.

In contrast to MSI by PCR-FLA, MMR-IHC was conducted only in non-CRC tumours. This is because MSI by PCR-FLA is considered superior to MMR-IHC for CRC, but neither is considered “gold standard” for non-CRC tumours [52]. MMR-IHC was conducted in 46.8% (243/519) of tumour samples, with MMRd observed in 15.2% (37/243; 95% CI: 11–20%) and MMR proficient (MMRp) in 84.8% (206/243; 95% CI: 80–89%) of cases (Fig. 3b).

When restricted to samples where MSI/MMR data from PCR-FLA and MMR-IHC was available, assessment of concordance between both techniques in non-CRC tumours showed a significant concordance of 93.8% (195/208) (Cohen’s κ =0.71, Z = 9.1, p < 0.001; Table 2 and Supplementary File 1). Of 13 discordant cases, germline testing showed P/LP variants in the *MSH2* and *PMS2* gene in cases of prostate and endometrial cancer, respectively. In both cases, PCR-FLA showed MSS and MMR-IHC showed MMRd status (Supplementary File 1 and Supplementary Table 1), thus suggesting higher sensitivity of IHC-MMR based MMRd testing in non-CRCs for LS detection.

Table 2
Concordance between MSI by PCR-FLA and MMRd by MMR-IHC in extracolonic tumours.

MSI Status (N = 224)	MMR deficient (n = 25)	MMR proficient (n = 197)	MMR status not available (n = 2)*
MSI-High (n = 27)	19	8	0
MSS (n = 183)	5	176	2
MSI status not available (n = 14)#	1	13	0

* IHC could not be performed for 2 cases due to insufficient tumour tissue.
MSI status was not available for 14 cases due to either poor DNA quality/ poor amplification, unavailability of matched normal samples, or unavailability of tumour blocks.

Lynch syndrome prevalence

Overall, prevalence of LS was 49.6% (66/133; 95% CI: 0.41–0.58) in MSI-high/MMRd tumours. Of patients with LS and MSI-high/MMRd tumour, 11% (7/66) had a primary tumour other than CRC (Fig. 4a and Supplementary Table 1). Highest LS prevalence was observed in CRC tumours 20% (59/295; 95% CI: 16–25%) followed by endometrial tumour 9.3% (4/43; 95% CI: 2.6–22%) (Fig. 4a). Stratification by anatomical site in colorectum showed significant difference in the proportion of LS observed between ascending colon, descending colon, transverse colon, and rectum (29.7%, 18.1%, 52.9%, 2.8%, respectively; χ^2 = 30.2, p < 0.001).

Amongst LS, highest proportion of P/LP variants was observed in the *MLH1* gene (69.7%, 46/66) followed by *MSH2* (18.2%, 12/66), *MSH6* (4.6%, 3/66), *PMS2* (4.6%, 3/66), and *EPCAM* (3.0%, 2/66) gene (Fig. 4b). Critically, 2 variants- c.306G > T and c.156delA- in the *MLH1* gene accounted for 36.4% (24/66) of all LS patients in the current cohort (Fig. 4c). Previous study by our group has shown *MLH1* c.306G > T to be a founder variant in India, with an estimated 6.4% of all LS carriers in the country to have this variant [19]. A high proportion of *MLH1* c.156delA variants suggested a putative second founder variant in the country. Furthermore, almost all patients identified with either of the 2 variants presented with colorectal cancer as their primary site (Supplementary Table 1).

Of 66 cases where LS was detected, 42 first degree relatives from 15 families consented to undergo cascade testing for LS carrier detection. Of these, 40.5% (17/42) and 59.5% (25/42) were non-carriers and carriers for LS, respectively (Supplementary File 2, Supplementary Tables 2 and 3).

MLH1 c.156delA haplotype analysis

Ancestors of families carrying the *MLH1* c.156delA variant came from Gujarat region in India. Common geographic origin of carrier families and their relatedness (Supplementary Table 4) suggested that the variant could have emerged as a unique event in a single founder individual. This hypothesis was confirmed by TMRCA, IBD, and haplotype analysis using phased VCFs (Supplementary Fig. 1 and Supplementary Table 5). The existence of a clear shared haplotype in carrier individuals permitted estimation of variant age. In this estimation for the *MLH1* c.156delA variant, we considered disease haplotypes in the region extending from rs9877525 to rs267526 (Table 3), which has a length of 0.7032 cM (approximately 703.2kb). Of 24 LS patients, 7 were first degree relatives corresponding to 6 families. Assessment of haplotype structure in the first degree relatives showed a haplotype recombination map length 0.9207 cM (approximately 920.7kb) extending from rs9817732 to rs267526. From the haplotype segment length, TMRCA was estimated to be 70 and 54 generations for all 24 LS carriers and first degree relatives, respectively. Assuming 25 years per generation, TMRCA is estimated to be 1750 years.

Table 3
Haplotypes associated with the *MLH1* c.156delA variant in carrier patients and first degree relatives

Genomic coordinates (hg38)		36553952	36554544	36560861	36771445 36788489 36993455	36996655	37311751 37372948 37432765 37455234	37459167 37471156	37474042	37474645
RSID		rs9817732 GSA- rs2361278	rs1879545		rs9877525 rs987190 rs1800734	rs587778927 c.156delA	rs6550471 rs6773378 rs1029236 rs197730	rs199279 rs11709385	rs169111	rs267526
REF	C	T	G	CGG	TA	TATA	GG	C	A	
ALT	T	C	T	AAA	T	CCCG	AC	T	G	
Case 1	D-7065 a1	T	C	T	AAA	TA/T	CCCG	AC	T	G
		C	T	T	CAG	TA/TA	TATA	GG	T	G
	D-7065 a2	T	C	T	AAA	TA/T	CCCG	AC	T	G
Case 2	D-7066 a1	C	T	G	AAG	TA/TA	CACG	AC	T	A
	D-7066 a2									
Case 3	MSI-AA-145 a1	T	C	T	AAA	TA/T	CCCG	AC	T	G
		T	C	T	AAG	TA/TA	CACG	AC	T	A
	MSI-AA-145 a2	T	C	T	AAA	TA/T	CCCG	AC	T	G
Case 4	MSI-HFA-144 a1	T	C	T	AAG	TA/TA	CACG	AC	T	A
	MSI-HFA-144 a2									
Case 5	MSI-ZSA-029 a1	T	C	T	AAA	TA/T	CCCG	AC	T	G
	MSI-ZSA-029 a2	C	T	T	AAG	TA/TA	TATG	AG	C	G
Case 6	MSI-HM-034 a1	T	C	T	AAA	TA/TA	TCTA	AC	T	G
	MSI-HM-034 a2	T	C	T	AAA	TA/T	CCCG	AC	T	G
Case 7	MSI-SA-035 a1	T	C	T	AAA	TA/T	CCCG	AC	T	G
	MSI-SA-035 a2	T	C	T	AAA	TA/TA	TCTA	AC	T	G
Case 8	MSI-KMV-134 a1	T	T	T	AAA	TA/TA	TATA	GG	T	G
	MSI-KMV-134 a2	T	C	T	AAA	TA/T	CCCG	AC	T	G
Case 9	MSI-ZID-207 a1	T	C	T	AAA	TA/T	CCCG	AC	T	G
	MSI-ZID-207 a2	C	T	G	CGA	TA/TA	TCTA	AC	T	G
Case 10	MSI_AD_070 a1	C	T	G	AAG	TA/TA	CACG	AG	T	A
	MSI_AD_070 a2	T	C	T	AAA	TA/T	CCCG	AC	T	G
Case 11	MSI-MF-165 a1	T	C	T	AAA	TA/T	CCCG	AC	T	G
	MSI-MF-165 a2	C	T	G	AAA	TA/TA	TCTA	AC	C	G
Case 12	MSI-RP-012 a1	T	C	T	AAA	TA/TA	TCTA	AC	T	G
	MSI-RP-012 a2	T	C	T	AAA	TA/T	CCCG	AC	T	G
Case 13	MSI_HP_176 a1	T	C	T	AAA	TA/T	CCCG	AC	T	G

	Genomic coordinates (hg38)	36553952	36554544	36560861	36771445 36788489 36993455	36996655	37311751 37372948 37432765 37455234	37459167 37471156	37474042	37474645
	RSID	rs9817732 GSA-rs1879545 rs2361278			rs9877525 rs987190 rs1800734	rs587778927 c.156delA	rs6550471 rs6773378 rs1029236 rs197730	rs199279 rs11709385	rs169111	rs267526
REF	C	T	G	CGG	TA	TATA	GG	C	A	
ALT	T	C	T	AAA	T	CCCG	AC	T	G	
	MSL_HP_176 a2	T	C	T	AAA	TA/TA	TCTA	AC	T	G
Case 14	D-7063 a1	T	C	T	AAA	TA/T	CCCG	AC	T	G
	D-7063 a2	C	T	G	CAG	TA/TA	CATG	AG	C	G
Case 15	MSI-ARP-271 a1	T	T	T	AAG	TA/TA	TATG	AG	C	G
	MSI-ARP-271 a2	TTT			AAA	TA/T	CCCG	AC	T	G
Case 16	MSI-SMP-164 a1	C	T	T	AAA	TA/T	CCCG	AC	T	G
	MSI-SMP-164 a2	T	C	T	CGA	TA/TA	TCTA	AC	T	G
Case 17	MSI_EAAM_058 a1	T	C	T	AAA	TA/T	CCCG	AC	T	G
	MSI_EAAM_058 a2	C	T	T	AAG	TA/TA	CACG	AC	T	A
Case 18	MSI_GAF_037 a1	T	C	T	AAA	TA/T	CCCG	AC	T	G
	MSI_GAF_037 a2	C	T	T	AAA	TA/TA	TATA	GC	T	G
Case 19	MSI_HMD_145 a1	CTG			AAA	TA/T	CCCG	AC	T	G
	MSI_HMD_145 a2	T	C	T	AAG	TA/TA	CACG	AC	T	A
Case 20	MSI_KRP_048 a1	T	C	T	AAA	TA/T	CCCG	AC	T	G
	MSI_KRP_048 a2	T	T	T	AAA	TA/TA	TATA	GG	T	G
Case 21	MSI_MHRP_102 a1	CTG			AAA	TA/T	CCCG	AC	T	G
	MSI_MHRP_102 a2	T	C	T	AAG	TA/TA	CACG	AC	T	A
Case 22	MSI_PGP_001 a1	C	T	T	AAG	TA/TA	TCTA	AC	T	G
	MSI_PGP_001 a2	T	C	T	AAA	TA/T	CCCG	AC	T	G
Case 23	MSI_PNJ_142 a1	T	C	T	AAA	TA/T	CCCG	AC	T	G
	MSI_PNJ_142 a2	T	C	T	CAA	TA/TA	TCCG	AG	T	G
Case 24	MSI_SVP_071 a1	TTT			AAA	TA/T	CCCG	AC	T	G
	MSI_SVP_071 a2	T	C	T	AAG	TA/TA	CATG	AG	C	G

Discussion

Our comprehensive assessment of MSI/MMR status in India, spanning 10 cancer types and 519 tumours, demonstrates that the presence of MSI-high/MMRd is predictive of LS, with 49.6% of patients with MSI-high/MMRd tumours harbouring germline P/LP variants in the MMR genes. In our study, 43.2% (224/519) of the tumours were of non-CRC origin. With the FDA's tissue site-agnostic approval of immunotherapies for advanced MSI-high/MMRd solid tumours [53], and the increasing availability of PCR-FLA, MMR-IHC, and tumour NGS platforms testing [54, 55], increase in routine MSI testing across a broad spectrum of cancers is anticipated in India. Our results support that all MSI-high/MMRd tumours should undergo germline assessment for LS, with cascade testing of at-risk relatives, given clinical implications for increased cancer surveillance, cancer chemoprevention interventions, and potential risk-reducing surgeries that may be warranted [21]. Although we used PCR-FLA across all tumour types and MMR-IHC only in non-CRC tumours, we observed a concordance of 93.8% between the two techniques in non-CRC tumours, suggesting that either technology can sufficiently capture LS molecular signature in non-CRC tumours. Additionally, identification of two recurrent founder variants c.306G > T and c.156delA in the *MLH1* gene has diagnostic and management implications in India.

In the present cohort, MSI-high/ MMRd status was observed in 25.6% (133/519) of tumours, which is significantly higher than that observed by Latham *et al.* (25.6% v/s 2.2%) [15]. However, Latham *et al.*'s cohort consisted of majorly non-canonical LS tumours (n = 10706, 71.2%). When focussing on the tumour types assessed in the present cohort, MSI-high/MMRd status was observed in 10% (436/4339) in the tumours by Latham *et al.*, which is significantly lower than that observed in the present study ($\chi^2 = 108.8$, $p < 0.01$). Furthermore, assessment of LS prevalence amongst MSI-high/MMRd tumours showed significant enrichment for LS in the current cohort compared to Latham *et al.* (49.6% versus 10.3%, $\chi^2 = 100.3$, $p < 0.01$). A significant proportion of the enrichment for MSI-high/MMRd signature and LS can be attributed to the two founder variants c.156delA and c.306G > T in the *MLH1* gene, which was observed in 36.4% (24/66) of all LS patients observed in the current study. Enrichment of these variants was likely due to the majority of the patients being recruited from Gujarat (52.6%, 273/519) where *MLH1* c.156delA variant was observed exclusively. Critically, the average age at diagnosis in the present cohort was 52.9 ± 13.7 years, which is significantly lower than that reported in patients of European/ non-Hispanic white ethnicity. However, when compared with studies from India, the average age at diagnosis is in congruence with data reported from tertiary centres [56–59], suggesting minimal effect of recruitment bias by age on MSI-high/MMRd and LS prevalence in the current study. A noteworthy observation is that of 88 patients with a family history of cancer in accordance with the Bethesda/ Amsterdam II criteria, 65 (73.9%) patients were detected with LS. This suggests that while recruitment bias due to family history and founder variants may have conflated LS prevalence estimates in the current study, family history could be used as a proxy to screen for LS carriers. A comparative assessment of LS detection by molecular signatures (MSI-high/MMRd) versus LS prediction models incorporating family history [60–62], is beyond the scope of this study.

Assessing distribution of MSI-high/MMRd molecular signature by tumour site showed highest prevalence in colorectum (33.9%) and endometrium (44.2%) which is congruence with the data observed in European/ non-Hispanic white as well as Indian cohorts [15, 59]. Additionally, proportion of MSI-high/MMRd signatures in extracolonic tumours such as gastric and ovaries showed congruence with prior published data. However, unlike European/ non-Hispanic white population data, the highest prevalence of LS was observed in MSI-high/MMRd CRC, endometrium, and ovarian cancers in the present study. Approximately a third of all LS patients (36.3%, 24/66) in the study were detected with founder variants in the *MLH1* gene with predominance in CRCs. This sharp contrast to previously reported distribution of LS in extra colonic tumours reflect differences in the underlying makeup of the cohort whereby prior study in the European/ non-Hispanic white population consisted 94.5% of extra-colonic tumours whereas the current study only consisted of 43.2% [15]. In our tumour-first cohort of patients with index colorectal or extracolonic cancers, LS was identified by MSI/MMR screening followed by germline testing of MSI-high/MMRd cases. This difference would lead to a “denominator effect” whereby low MSI-high/MMRd prevalence in extra-colonic tumours would lead to lower proportion of LS positive cases for these tumours. Additionally, each MMR gene confers different penetrance risk by site and sex [7, 20], with the highest conferred by *MLH1* and *MSH2* genes for CRC in males. In the present cohort we observed a significant proportion of males in CRC versus extra-colonic tumours (69.5% versus 37.5%, $\chi^2 = 52.8$, $p < 0.01$), which in part explains highest proportion of LS in CRC tumours.

In 208 extra-colonic tumours, both MSI by PCR-FLA and MMR-IHC were carried out to ascertain concordance between both assays for MSI-high/MMRd molecular phenotype. Indeed, 93.8% concordance observed in the current study is in congruence with that observed for gynaecological and gastro-esophageal cancers [63, 64], despite there being reports of discordance due to pre-analytical factors [25]. In 13 discordant cases, LS was detected in prostate and endometrial cancer patients where MMR-IHC had predicted MMRd status. This further supports ASCO endorsement, CAP guidelines, and prior literature evidence supporting MMR-IHC as first line test for endometrial cancer, and MMR-IHC and PCR-FLA for gastroesophageal and small bowel cancer [52, 65].

Lastly, to the best of our knowledge, we are the first group to report *MLH1* c.156delA as a founder variant in India. The variant has not been observed in any other populations on the gnomAD v4.1.1 database except for the SAS population with an allele frequency of 0.00001098. This suggests that like the first founder variant *MLH1* c.306G > T [19], this variant arose in a common ancestor based in Gujarat approximately 70 generations ago. Furthermore, with the availability of samples from affected relatives, IBD analysis could be conducted in the present study. Together with c.306G > T, the founder variants were present in approximately one-third of all LS cases in the present cohort. This has direct implications for molecular diagnostic labs providing genetic testing, genetic counselling, and clinical care [7, 18, 66].

This study has several limitations. First, it was a prospective hospital-based multicentre cohort rather than a population-based study [8, 17, 67], therefore, the reported frequencies of MSI-high/MMRd and LS should not be interpreted as population prevalence estimates for India. Second, recruitment was geographically skewed, with a substantial proportion of patients originating from Gujarat where recurrent *MLH1* founder variants were concentrated. This may have inflated the overall LS yield, *MLH1* predominance, and founder variant contribution. Third, the cohort was enriched for CRCs while several extra-colonic tumour groups were represented by small numbers, limiting the precision of site-specific prevalence estimates and complicating direct comparison with pan-cancer cohorts from other populations. Fourth, germline testing was restricted to MSI-high and *BRAF* V600E wildtype CRCs and MSI-high/MMRd extra-colonic tumours, so LS cases in tumours classified as MSS/MMRp may have been missed. Finally, not all germline-negative MSI-high/MMRd tumours

underwent comprehensive somatic resolution, including systematic assessment for double-somatic MMR inactivation or *MLH1* promoter hypermethylation across tumour types, which limits complete discrimination between hereditary and sporadic causes of MMR deficiency.

In conclusion, our study demonstrates that a pan-cancer tumour-first approach using MSI/MMRd screening followed by germline testing can substantially improve detection of LS in Indian cancer patients, while also revealing population-specific features of LS architecture. The marked predominance of *MLH1* and *MSH2* gene carriers, and the identification of *MLH1* c.156delA as a novel founder variant, highlight the need for regionally informed diagnostic pathways in India that combine reflex tumour testing with rapid germline confirmation and cascade testing. Such an approach has direct translational relevance, as timely diagnosis of LS can guide tailored treatment with immunotherapies, tailored surveillance for CRC and extra-colonic tumours, support evidence based cancer chemoprevention, and influence surgical management and risk reducing procedures. These findings therefore provide a strong rationale for integrating pan-cancer LS detection into routine oncology practice in India.

Tables, Figures, and Supplementary Data

Figure 1 Schematic diagram of the somatic and germline testing pathway in 519 cancer patients. *Immunohistochemistry was performed for 224 extracolonic tumours and microsatellite instability testing could not be carried out for 19 patients due to poor DNA quality/ poor amplification, unavailability of matched normal samples, or unavailability of tumour blocks

Figure 2 Distribution of patients with index cancer by cancer site

Figure 3 Proportion of (a) microsatellite unstable (MSI-high) and (b) mismatch repair deficient (MMRd) tumours by cancer site. Whiskers represent 95% confidence intervals

Figure 4 (a) Distribution of Lynch syndrome detection by cancer site; (b) Proportion of cancer site within each mismatch repair (MMR) gene; and (c) Number of Lynch syndrome patients with a pathogenic/ likely pathogenic variant within the *MLH1* gene

Table 1: Demographic characteristic of the 519 cancer patient cohort.

Table 2: Concordance between MSI by PCR-FLA and MMRd by MMR-IHC in extracolonic tumours.

Table 3: Haplotypes associated with the *MLH1* c.156delA variant in carrier patients and first degree relatives.

Supplementary File 1

Tabular dataset containing clinical and molecular genetic details of 519 cancer patients and 42 first degree relatives of Lynch syndrome patients

Supplementary File 2

Pedigree charts of 25 family members diagnosed with Lynch syndrome

Supplementary Fig. 1

TMRCa estimation between 24 individuals detected with *MLH1* c.156delA variant

Supplementary Table 1

Germline variants detected in 66 cancer patients diagnosed with Lynch syndrome

Supplementary Table 2

Germline variant details of 25 Lynch syndrome carriers that are first degree relatives of Lynch syndrome patients with cancer

Supplementary Table 3

List of primer sequences for variant validation by Sanger sequencing/ Q-PCR.

Supplementary Table 4

PI_HAT values for 24 individuals detected with *MLH1* c.156delA variant

Supplementary Table 5

Phased VCF file for 17 Lynch syndrome patients with cancer and 7 first degree relatives detected with *MLH1* c.156delA variant

Declarations

Competing interests

All authors certify that they have no affiliations with or involvement in any organisation or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Ethics approval

Written informed consent was collected from all patients. All methods were carried out in accordance with the Helsinki declaration. This study was approved by the Ethics Committee at FRIGE's Institute of Human Genetics, Ahmedabad (FRIGE/IEC/23/2024), CIMS Hospital, Ahmedabad, BALCO Medical Centre, Chattisgarh (BMCIRB/2024/03), AIIMS, Delhi (AIIMSA1322/03.05.2024, RP-20/2024) and AIIMS, Rishikesh (AIIMS/IEC/25/206). The study was registered with the Clinical Trials Registry - India (CTRI) with registration number CTRI/2024/03/063560.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent to publish

Patients signed informed consent regarding publishing their data.

Funding

This study was funded by the Gujarat State Biotechnology Mission, Department of Science and Technology (GSBTM/JDR&D/604/2019/299 and GSBTM/JDR&D/663/2023-24/02003575) to Harsh Sheth. Jyoti Sadhwani is supported by the Bakeri Foundation fellowship.

Author Contribution

HS, JS, FS, AJ, DTB, and JB conceived the idea for this study. HS, AJ, BS, SK, YK, PS, MS, SA, LT, VDY, AT, CS, RP, BT, DB, MG, TP, NP, AP, RuP, RG, ChS, BhS, GM, AJ, AD, CB, AS, and DL carried out the patient recruitment. HS, JyS, CP, and PrS carried out the laboratory work. HS, JyS, DTB, JB, and AJ carried out the data analysis and interpretation. HS, JyS, and DTB wrote the manuscript. All authors approved the final version of the manuscript.

Acknowledgement

We thank all participating patients and their family members. Their participation would greatly aid the progress of Indian medical science towards cancer prevention.

Data Availability

All data generated or analysed during this study are included in this published article as Supplementary File 1. FASTQ files from germline sequencing and genome-wide genotyping data are available from the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS50000001449 and EGAS00001008405, respectively. Further information about EGA can be found at <https://ega-archive.org> and The European Genome-phenome Archive in 2021: 10.1093/nar/gkab1059

References

1. Lynch HT, de la Chapelle A (2003) Hereditary colorectal cancer. *N Engl J Med* 348:919–932. <https://doi.org/10.1056/NEJMra012242>
2. Nagtegaal ID, Odze RD, Klimstra D et al (2020) The 2019 WHO classification of tumours of the digestive system. *Histopathology* 76:182–188. <https://doi.org/10.1111/his.13975>
3. Win AK, Jenkins MA, Dowty JG et al (2017) Prevalence and Penetrance of Major Genes and Polygenes for Colorectal Cancer. *Cancer Epidemiol Biomarkers Prev* 26:404–412. <https://doi.org/10.1158/1055-9965.EPI-16-0693>
4. Abildgaard AB, Nielsen SV, Bernstein I et al (2023) Lynch syndrome, molecular mechanisms and variant classification. *Br J Cancer* 128:726–734. <https://doi.org/10.1038/s41416-022-02059-z>
5. Peltomäki P (2003) Role of DNA Mismatch Repair Defects in the Pathogenesis of Human Cancer. *J Clin Oncol* 21:1174–1179. <https://doi.org/10.1200/JCO.2003.04.060>
6. Boland CR, Goel A (2010) Microsatellite Instability in Colorectal Cancer. *Gastroenterology* 138:2073. <https://doi.org/10.1053/j.gastro.2009.12.064>
7. Dominguez-Valentin M, Haupt S, Seppälä TT et al (2023) Mortality by age, gene and gender in carriers of pathogenic mismatch repair gene variants receiving surveillance for early cancer diagnosis and treatment: a report from the prospective Lynch syndrome database. *eClinicalMedicine* 58:101909. <https://doi.org/10.1016/j.eclinm.2023.101909>
8. Park J, Karnati H, Rustgi SD et al (2025) Impact of population screening for Lynch syndrome insights from the All of Us data. *Nat Commun* 16:523. <https://doi.org/10.1038/s41467-024-52562-5>
9. Lowry C, Laurent C, Bindra AK et al (2026) Utilization of care and alignment of screening with NCCN guidelines for patients with Lynch syndrome: a retrospective cohort study. *Hered Cancer Clin Pract.* <https://doi.org/10.1186/s13053-026-00334-1>
10. Crean C, McConnell L, Quinn C et al (2026) Population-based screening for Lynch syndrome in patients with colorectal cancer: implementing the NICE diagnostic pathway. *BMJ Connect Clin Genet Genomics* 3:e000054. <https://doi.org/10.1136/bmjccg-2025-000054>

11. McDonald FE, Pethick J, Santaniello F et al (2024) Identification of people with Lynch syndrome from those presenting with colorectal cancer in England: baseline analysis of the diagnostic pathway. *Eur J Hum Genet* 32:529–538. <https://doi.org/10.1038/s41431-024-01550-w>
12. Burn J, Sheth H, Elliott F et al (2020) 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. [https://doi.org/10.1016/S0140-6736\(20\)30366-4](https://doi.org/10.1016/S0140-6736(20)30366-4)
13. Seppälä TT, Latchford A, Negoi I et al (2021) European guidelines from the EHTG and ESCP for Lynch syndrome: an updated third edition of the Mallorca guidelines based on gene and gender. *Br J Surg* 108:484–498. <https://doi.org/10.1002/bjs.11902>
14. Mathers JC, Elliott F, Macrae F et al (2022) Cancer Prevention with Resistant Starch in Lynch Syndrome Patients in the CAPP2-Randomized Placebo Controlled Trial: Planned 10-Year Follow-up. *Cancer Prev Res (Phila Pa)* 15:623–634. <https://doi.org/10.1158/1940-6207.CAPR-22-0044>
15. Latham A, Srinivasan P, Kemel Y et al (2019) Microsatellite Instability Is Associated With the Presence of Lynch Syndrome Pan-Cancer. *J Clin Oncol Off J Am Soc Clin Oncol* 37:286–295. <https://doi.org/10.1200/JCO.18.00283>
16. Li S, Luo N, Jin G et al (2025) Clinical and genomic features of Lynch syndrome differ by tumor site and disease spectrum. *Nat Commun* 16:10179. <https://doi.org/10.1038/s41467-025-65164-6>
17. Mizukami K, Usui Y, Iwasaki Y et al (2025) Pan-cancer prevalence, risk, and clinical and demographic characteristics of Lynch Syndrome-associated variants in BioBank Japan. *Commun Med* 5:522. <https://doi.org/10.1038/s43856-025-01231-9>
18. Ponti G, Castellsagué E, Ruini C et al (2015) Mismatch repair genes founder mutations and cancer susceptibility in Lynch syndrome. *Clin Genet* 87:507–516. <https://doi.org/10.1111/cge.12529>
19. Sheth H, Sadhwani J, Jain A et al (2024) Haplotype analysis detects MLH1 founder variant in Indian Lynch syndrome patient cohort. *Fam Cancer* 24:13. <https://doi.org/10.1007/s10689-024-00436-6>
20. Møller P, Seppälä T, Dowty JG et al (2022) Colorectal cancer incidences in Lynch syndrome: a comparison of results from the prospective lynch syndrome database and the international mismatch repair consortium. *Hered Cancer Clin Pract* 20:36. <https://doi.org/10.1186/s13053-022-00241-1>
21. Prospective Lynch Syndrome Database (2025) Metachronous colorectal cancer risks after extended or segmental resection in MLH1, MSH2, and MSH6 Lynch syndrome: multicentre study from the Prospective Lynch Syndrome Database. *BJS* 112:znaf061. <https://doi.org/10.1093/bjs/znaf061>
22. Saklani A, Kazi M, Desouza A et al (2024) Tata Memorial Centre Evidence Based Management of Colorectal cancer. *Indian J Cancer* 61:S29–S51. https://doi.org/10.4103/ijc.ijc_66_24
23. Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215–1215. <https://doi.org/10.1093/nar/16.3.1215>
24. Dedeurwaerdere F, Claes KB, Van Dorpe J et al (2021) Comparison of microsatellite instability detection by immunohistochemistry and molecular techniques in colorectal and endometrial cancer. *Sci Rep* 11:12880. <https://doi.org/10.1038/s41598-021-91974-x>
25. Nádorvári ML, Kenessey I, Kiss A et al (2024) Comparison of standard mismatch repair deficiency and microsatellite instability tests in a large cancer series. *J Transl Med* 22:150. <https://doi.org/10.1186/s12967-024-04960-y>
26. Huang T, Zhuge J, Zhang WW (2013) Sensitive detection of BRAF V600E mutation by Amplification Refractory Mutation System (ARMS)-PCR. *Biomark Res* 1:3. <https://doi.org/10.1186/2050-7771-1-3>
27. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
28. Van Der Auwera GA, Carneiro MO, Hartl C et al (2013) *Curr Protoc Bioinforma* 43. <https://doi.org/10.1002/0471250953.bi1110s43>. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline
29. Smedley D, Jacobsen JOB, Jäger M et al (2015) Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc* 10:2004–2015. <https://doi.org/10.1038/nprot.2015.124>
30. Robinson PN, Köhler S, Bauer S et al (2008) The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease. *Am J Hum Genet* 83:610–615. <https://doi.org/10.1016/j.ajhg.2008.09.017>
31. Ioannidis NM, Rothstein JH, Pejaver V et al (2016) REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* 99:877–885. <https://doi.org/10.1016/j.ajhg.2016.08.016>
32. Qi H, Zhang H, Zhao Y et al (2021) MVP predicts the pathogenicity of missense variants by deep learning. *Nat Commun* 12:510. <https://doi.org/10.1038/s41467-020-20847-0>
33. Sim N-L, Kumar P, Hu J et al (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 40:W452–W457. <https://doi.org/10.1093/nar/gks539>
34. Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014) MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 11:361–362. <https://doi.org/10.1038/nmeth.2890>
35. Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr Protoc Hum Genet* 76. <https://doi.org/10.1002/0471142905.hg0720s76>
36. Rentzsch P, Witten D, Cooper GM et al (2019) CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 47:D886–D894. <https://doi.org/10.1093/nar/gky1016>
37. Landrum MJ, Lee JM, Benson M et al (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 46:D1062–D1067. <https://doi.org/10.1093/nar/gkx1153>

38. Plazzer JP, Sijmons RH, Woods MO et al (2013) The InSiGHT database: utilizing 100 years of insights into Lynch Syndrome. *Fam Cancer* 12:175–180. <https://doi.org/10.1007/s10689-013-9616-0>
39. Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14:178–192. <https://doi.org/10.1093/bib/bbs017>
40. Geoffroy V, Herenger Y, Kress A et al (2018) AnnotSV: an integrated tool for structural variations annotation. *Bioinformatics* 34:3572–3574. <https://doi.org/10.1093/bioinformatics/bty304>
41. Karczewski KJ, Francioli LC, Tiao G et al (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581:434–443. <https://doi.org/10.1038/s41586-020-2308-7>
42. MacDonald JR, Ziman R, Yuen RKC et al (2014) The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* 42:D986–D992. <https://doi.org/10.1093/nar/gkt958>
43. Hamosh A (2004) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 33:D514–D517. <https://doi.org/10.1093/nar/gki033>
44. Clendenning M, Hampel H, LaJeunesse J et al (2006) Long-range PCR facilitates the identification of PMS2-specific mutations. *Hum Mutat* 27:490–495. <https://doi.org/10.1002/humu.20318>
45. Richards S, Aziz N, Bale S et al (2015) 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* 17:405–424. <https://doi.org/10.1038/gim.2015.30>
46. Thaxton C, Good ME, DiStefano MT et al (2022) Utilizing ClinGen gene-disease validity and dosage sensitivity curations to inform variant classification. *Hum Mutat* 43:1031–1040. <https://doi.org/10.1002/humu.24291>
47. Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* 81:559–575. <https://doi.org/10.1086/519795>
48. Browning BL, Tian X, Zhou Y, Browning SR (2021) Fast two-stage phasing of large-scale sequence data. *Am J Hum Genet* 108:1880–1890. <https://doi.org/10.1016/j.ajhg.2021.08.005>
49. Thompson EA (2013) Identity by Descent: Variation in Meiosis, Across Genomes, and in Populations. *Genetics* 194:301–326. <https://doi.org/10.1534/genetics.112.148825>
50. Browning BL, Browning SR (2013) Improving the Accuracy and Efficiency of Identity-by-Descent Detection in Population Data. *Genetics* 194:459–471. <https://doi.org/10.1534/genetics.113.150029>
51. Ralph P, Coop G (2013) The Geography of Recent Genetic Ancestry across Europe. *PLoS Biol* 11:e1001555. <https://doi.org/10.1371/journal.pbio.1001555>
52. Vikas P, Messersmith H, Compton C et al (2023) Mismatch Repair and Microsatellite Instability Testing for Immune Checkpoint Inhibitor Therapy: ASCO Endorsement of College of American Pathologists Guideline. *J Clin Oncol* 41:1943–1948. <https://doi.org/10.1200/JCO.22.02462>
53. (2017) First Tissue-Agnostic Drug Approval Issued. *Cancer Discov* 7:656–656. <https://doi.org/10.1158/2159-8290.CD-NB2017-078>
54. Gallon R, Sheth H, Hayes C et al (2020) Sequencing-based microsatellite instability testing using as few as six markers for high-throughput clinical diagnostics. *Hum Mutat* 41:332–341. <https://doi.org/10.1002/humu.23906>
55. Gallon R, Holt G, Alfaiakawi W et al (2025) Novel microsatellite instability test of sebaceous tumours to facilitate low-cost universal screening for Lynch syndrome. *Clin Exp Dermatol* 50:1155–1162. <https://doi.org/10.1093/ced/llaf005>
56. Daga P, Rawal T, Gupta PK, Khatri NK (2021) Incidence of Colorectal Cancer in North-Western India over 5 Years. *J Radiat Cancer Res* 12:165–167. https://doi.org/10.4103/jrcr.jrcr_29_21
57. Deo SVS, Kumar S, Bhoirwal S et al (2021) Colorectal Cancers in Low- and Middle-Income Countries—Demographic Pattern and Clinical Profile of 970 Patients Treated at a Tertiary Care Cancer Center in India. *JCO Glob Oncol* 1110–1115. <https://doi.org/10.1200/GO.21.00111>
58. Mathew Thomas V, Baby B, Wang K et al (2020) Trends in colorectal cancer incidence in India. *J Clin Oncol* 38:e16084–e16084. https://doi.org/10.1200/JCO.2020.38.15_suppl.e16084
59. Kaler AK, Fatima S, Dhurat O et al (2024) Role of Microsatellite Instability as Pan-Cancer Agnostic Marker of Solid Tumours in Personalized Medicine. *Curr Tissue Microenviron Rep* 5:69–77. <https://doi.org/10.1007/s43152-024-00054-5>
60. Barnetson RA, Tenesa A, Farrington SM et al (2006) Identification and Survival of Carriers of Mutations in DNA Mismatch-Repair Genes in Colon Cancer. *N Engl J Med* 354:2751–2763. <https://doi.org/10.1056/NEJMoa053493>
61. Chen S, Wang W, Lee S et al (2006) Prediction of Germline Mutations and Cancer Risk in the Lynch Syndrome. *JAMA* 296:1479. <https://doi.org/10.1001/jama.296.12.1479>
62. Kastrinos F, Uno H, Ukaegbu C et al (2017) Development and Validation of the PREMM5 Model for Comprehensive Risk Assessment of Lynch Syndrome. *J Clin Oncol Off J Am Soc Clin Oncol* 35:2165–2172. <https://doi.org/10.1200/JCO.2016.69.6120>
63. Park Y, Nam SK, Seo SH et al (2023) Comprehensive Study of Microsatellite Instability Testing and Its Comparison With Immunohistochemistry in Gastric Cancers. *J Gastric Cancer* 23:264. <https://doi.org/10.5230/jgc.2023.23.e5>
64. Smithgall MC, Remotti H, Hsiao SJ et al (2022) Investigation of discrepant mismatch repair immunohistochemistry and microsatellite instability polymerase chain reaction test results for gynecologic cancers using next-generation sequencing. *Hum Pathol* 119:41–50. <https://doi.org/10.1016/j.humpath.2021.10.004>

65. Mendiola M, Heredia-Soto V, Ruz-Caracuel I et al (2023) Comparison of Methods for Testing Mismatch Repair Status in Endometrial Cancer. *Int J Mol Sci* 24:14468. <https://doi.org/10.3390/ijms241914468>

66. Dominguez-Valentin M, Sampson JR, Seppälä TT et al (2020) Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the Prospective Lynch Syndrome Database. *Genet Med* 22:15–25. <https://doi.org/10.1038/s41436-019-0596-9>

67. Fummey E, Navarro P, Plazzer J-P et al (2024) Estimating cancer risk in carriers of Lynch syndrome variants in UK Biobank. *J Med Genet* 61:861–869. <https://doi.org/10.1136/jmg-2023-109791>

Figures

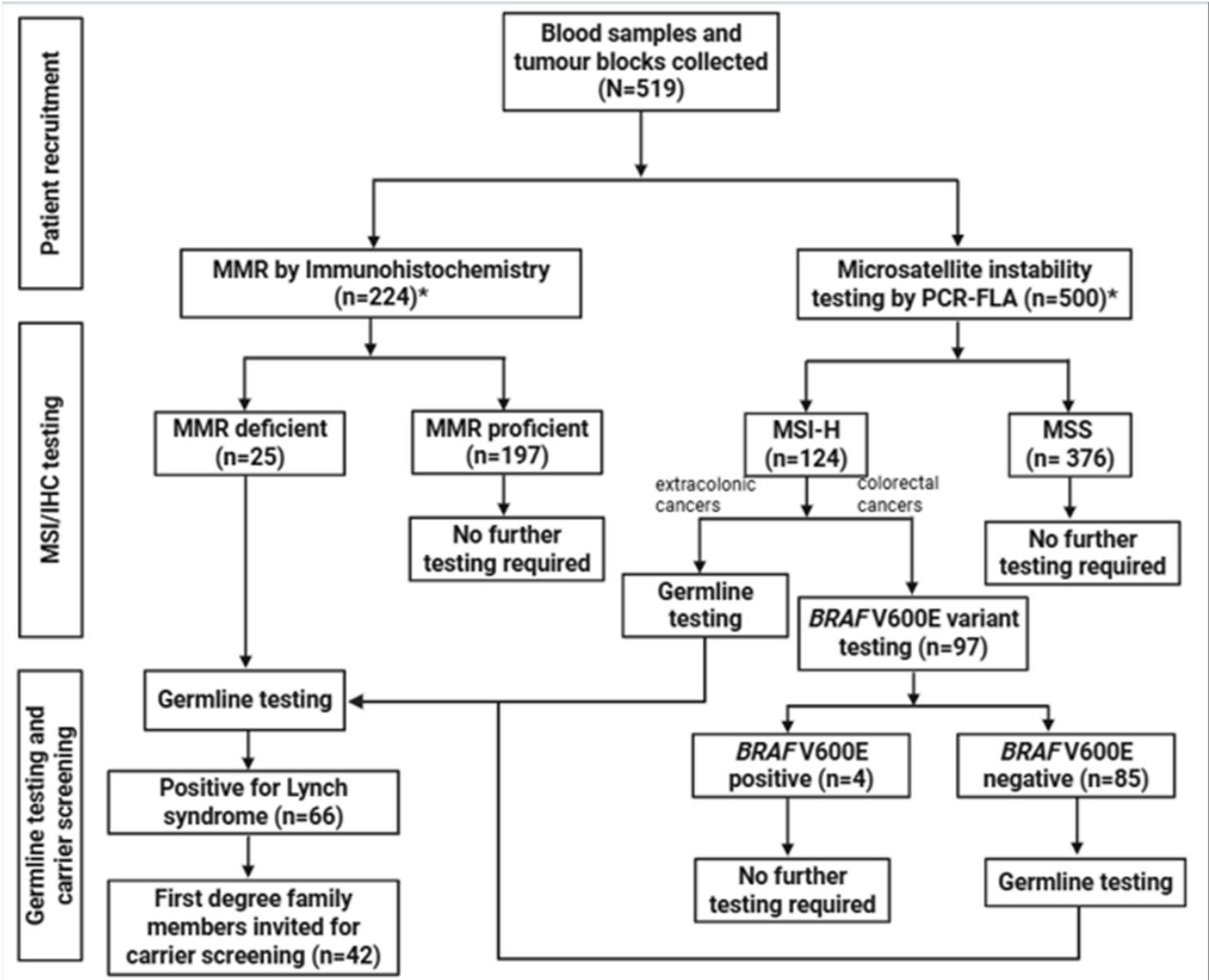


Figure 1

Schematic diagram of the somatic and germline testing pathway in 519 cancer patients. *Immunohistochemistry was performed for 224 extracolonic tumours and microsatellite instability testing could not be carried out for 19 patients due to poor DNA quality/ poor amplification, unavailability of matched normal samples, or unavailability of tumour blocks

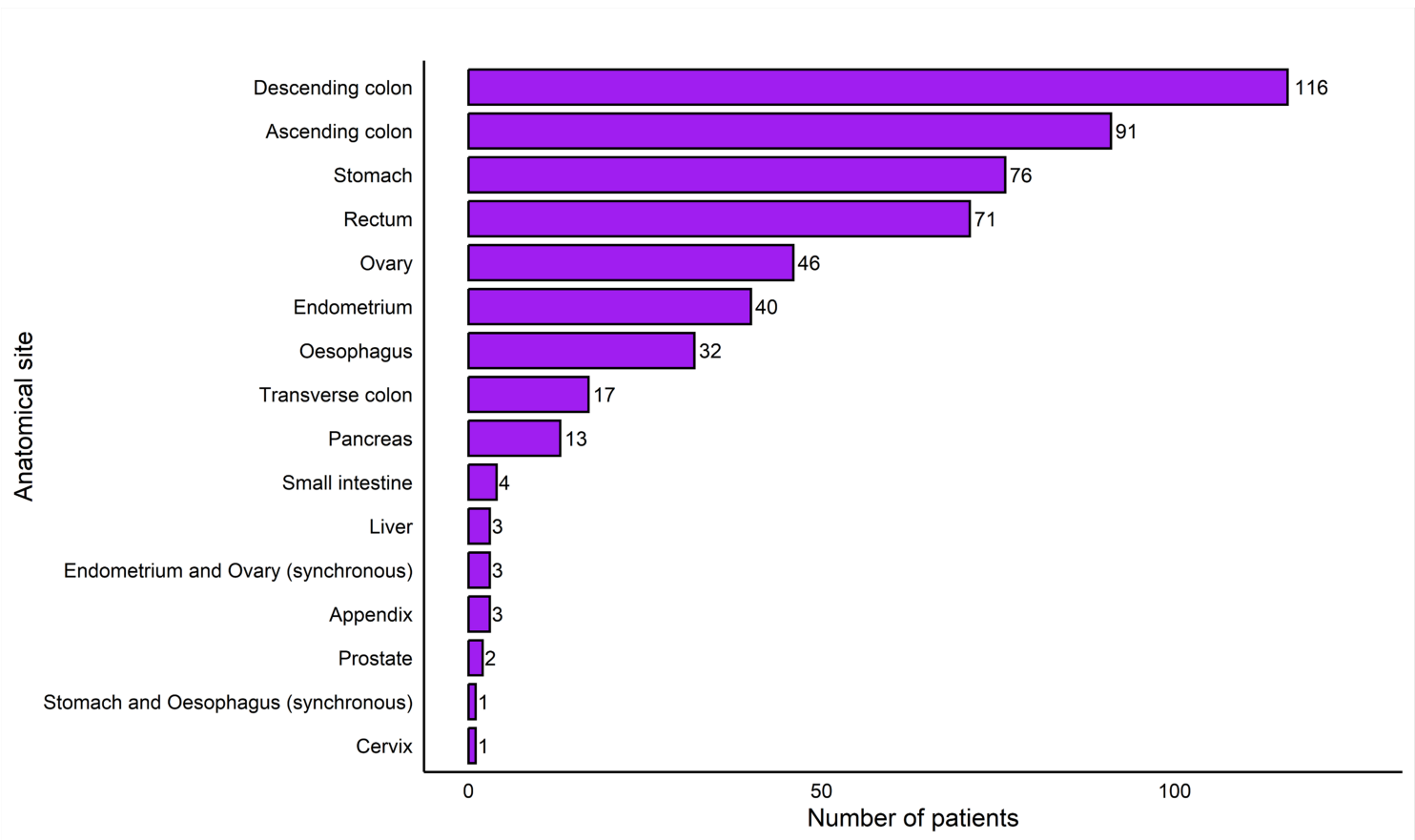


Figure 2
Distribution of patients with index cancer by cancer site

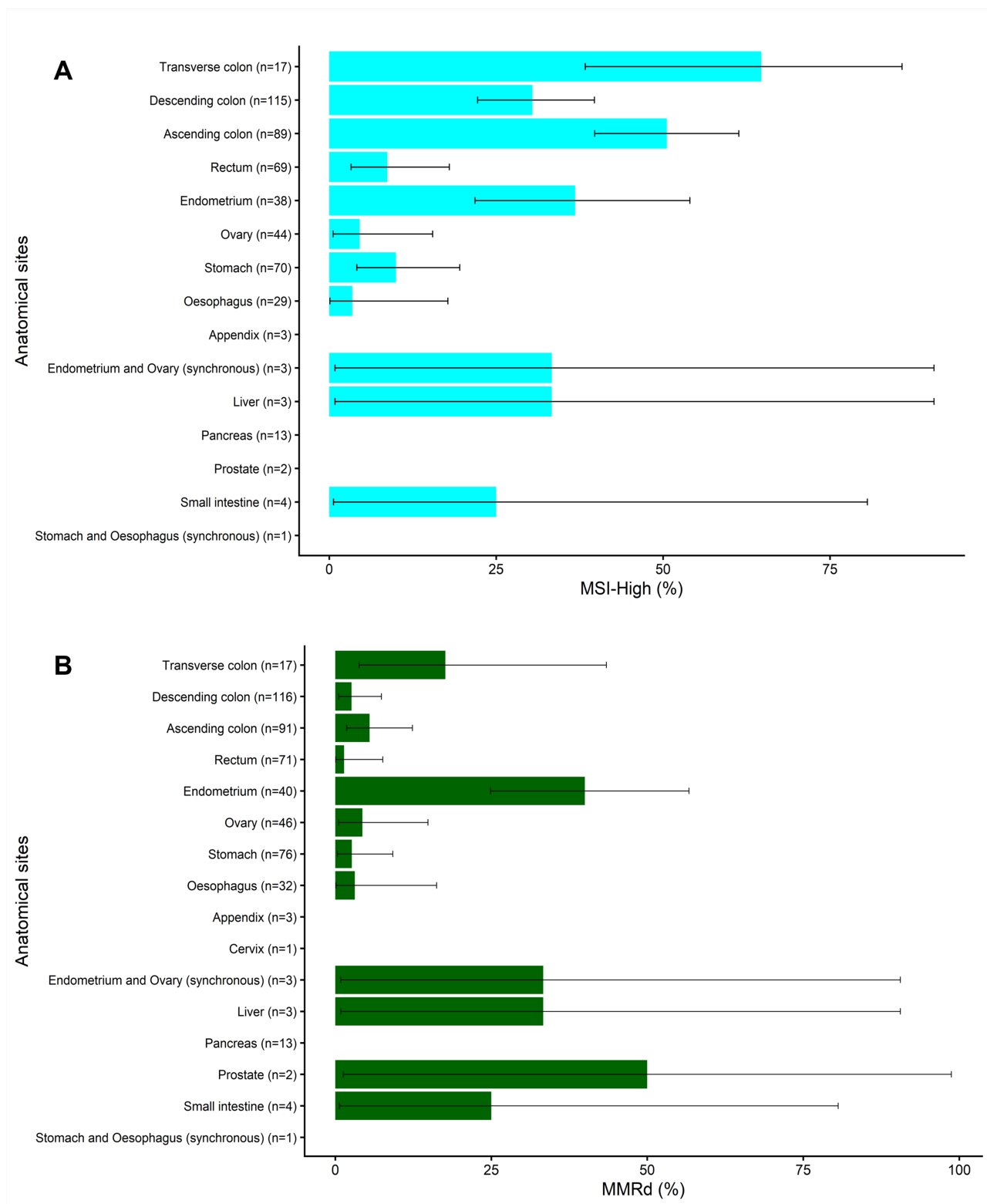


Figure 3

Proportion of (a) microsatellite unstable (MSI-high) and (b) mismatch repair deficient (MMRd) tumours by cancer site. Whiskers represent 95% confidence intervals

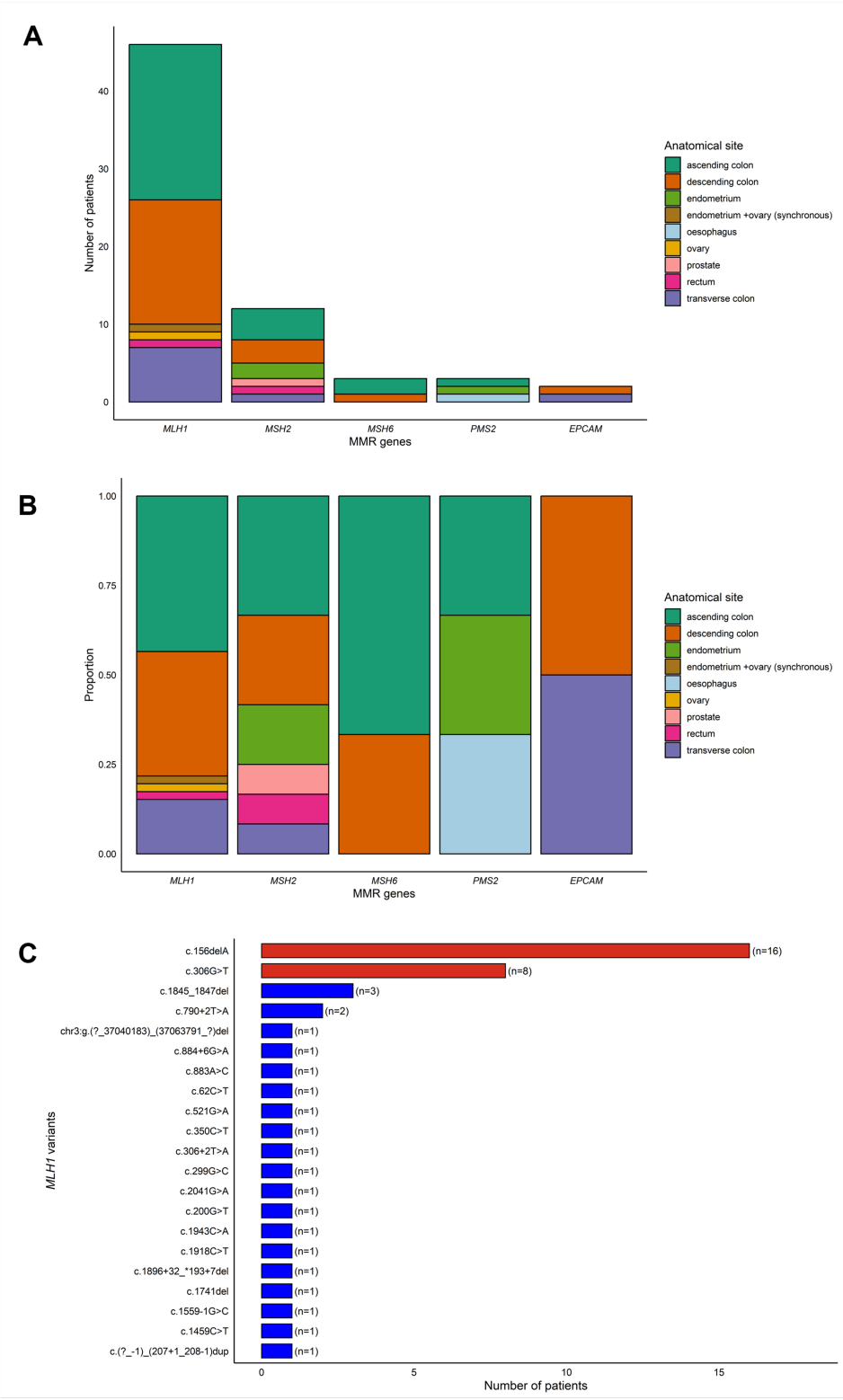


Figure 4

(a) Distribution of Lynch syndrome detection by cancer site; (b) Proportion of cancer site within each mismatch repair (MMR) gene; and (c) Number of Lynch syndrome patients with a pathogenic/ likely pathogenic variant within the *MLH1* gene

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFile1.xlsx](#)
- [SupplementaryFile2Pedigreechartsof25familymembersdiagnosedwithLynchsyndrome.docx](#)

- [SupplementaryFigure1TMRCaestimationbetween24individualsdetectedwithMLH1c.156delAvariant.png](#)
- [SupplementaryTable1Germlinevariantsdetectedin66cancerpatientsdiagnosedwithLynchsyndrome.xlsx](#)
- [SupplementaryTable2Germlinevariantdetailsof25LynchsyndromecarriersatarefirstdegreerelativesofLynchsyndromepatientswithcancer.xlsx](#)
- [SupplementaryTable3ListofprimersequencesforvariantvalidationbySangersequencingQPCR.xlsx](#)
- [SupplementaryTable4PIHATvaluesfor24individualsdetectedwithMLH1c.156delAvariant.xlsx](#)
- [SupplementaryTable5PhasedVCFfilefor17Lynchsyndromepatientswithcancerand7firstdegreerelativesdetectedwithMLH1c.156delAvariant.xlsx](#)