Received: 28 October 2024 | Revised: 18 February 2025 | Accepted: 4 April 2025

DOI: 10.1002/ijc.35451

RESEARCH ARTICLE

Tumor Markers and Signatures

IJC INTERNATIONAL JOURNAL of CANCER | UICC

Comprehensive genetic and epigenetic characterization of Lynch-like syndrome patients

Francesca Pirini[1] | Luciano Calzari[2] | Gianluca Tedaldi[1] | Michela Tebaldi[1] | Valentina Zampiga[1] | Ilaria Cangini[1] | Rita Danesi[3] | Mila Ravegnani[3] | Valentina Arcangeli[3] | Alessandro Passardi[1] | Elisabetta Petracci[1] | Sara Bravaccini[1,4] | Giorgia Marisi[1] | Alessandra Viel[5] | Daniela Barana[6] | Monica Pedroni[7] | Luca Roncucci[7] | Daniele Calistri[1] | Davide Gentilini[2,8]

[1]IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy

[2]Bioinformatics and Statistical Genomic Unit, IRCCS Istituto Auxologico Italiano, Milan, Italy

[3]Romagna Cancer Registry, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy

[4]Faculty of Medicine and Surgery, University of Enna "Kore", Enna, Italy

[5]Unit of Oncogenetics and Functional Oncogenomics, Centro di Riferimento Oncologico di Aviano (CRO), IRCCS, Aviano, Italy

[6]Oncology Unit, Local Health and Social Care Unit, ULSS8 Berica, Vicenza, Italy

[7]Department of Medical and Surgical Sciences, University of Modena and Reggio Emilia, Modena, Italy

[8]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy

Correspondence
Francesca Pirini, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy.
Email: francesca.pirini@irst.emr.it

Funding information
Italian ministry of health, Ricerca Corrente

Abstract

Lynch-like syndrome (LLS) presents very similar clinicopathological characteristics to Lynch syndrome (LS) but the mechanism for cancer predisposition remains unknown. The present study aims to investigate the causal mechanism of LLS by a comprehensive genetic and epigenetic approach. Thirty-two LLS and 34 LS patients with colorectal cancer (CRC) fitting the Amsterdam and Bethesda criteria were included, along with 29 CRC sporadic patients, and analyzed for the presence of pathogenic variants

Abbreviations: ALK, Anaplastic Lymphoma Kinase; ALPK, Alpha Kinase 1; APC, Adenomatous Polyposis Coli; ASCL2, Achaete-Scute Family BHLH Transcription Factor 2; ATM, ATM Serine/Threonine Kinase; BARD1, BRCA1 Associated RING Domain 1; BRAF, B-Raf proto-oncogene, serine/threonine kinase; BRCA1, BRCA1 DNA Repair Associated; BUB1B, BUB1 Mitotic Checkpoint Serine/Threonine Kinase B; CAVIN3, Caveolae Associated Protein 3; CDC73, Cell Division Cycle 73; CDH1, Cadherin 1; CRC, Colorectal cancer; CTC1, CST Telomere Replication Complex Component 1; DCC, DCC Netrin 1 Receptor; DICER1, Dicer 1, Ribonuclease III; dMMR, Mismatch repair deficient; EGFR, Epidermal Growth Factor Receptor; EML, Epi-Mutation-Load; ENTPD1-AS, Ectonucleoside Triphosphate Diphosphohydrolase 1- Antisense; EPCAM, Epithelial Cell Adhesion Molecule; ERBB2, Erb-B2 Receptor Tyrosine Kinase 2; ERBB4, Erb-B2 Receptor Tyrosine Kinase 4; ERCC2, Exonuclease 1; EZH2, Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit; FA, Fanconi anemia; FAN1, FANCD2 And FANCI Associated Nuclease 1; FANCA, FA Complementation Group A; FANCC, FA Complementation Group C; FANCI, FA Complementation Group I; FANCM, FA Complementation Group M; HBOC, Hereditary Breast and Ovarian Cancer; HOXA11-AS, Homeobox A11 Antisense RNA; KCNK12, Potassium Two Pore Domain Channel Subfamily K Member 12; KIT, KIT Proto-Oncogene, Receptor Tyrosine Kinase; LLS, Lynch Like Syndrome; LS, Lynch Syndrome; MAP, MUTYH-associated polyposis; MCM8/9, minichromosome maintenance 8/9 homologous recombination repair factor; MLH1, MutL Homolog 1; MLH3, MutL Homolog 3; MMR, Mismatch repair; MSH2, MutL Homolog 2; MSH6, MutL Homolog 6; MSI, microsatellite instability; MUTYH, MutY DNA Glycosylase; NBN, Nibrin; NF1, Neurofibromin 1; OPCML, Opioid Binding Protein/Cell Adhesion Molecule Like; PCA, principal component analysis; PMS2, PMS1 Homolog 2, Mismatch Repair System Component; POLD1, DNA Polymerase Delta 1, Catalytic Subunit; POLE, DNA Polymerase Epsilon, Catalytic Subunit; PPARG, Peroxisome Proliferator Activated Receptor Gamma; PRKDC, Protein Kinase, DNA-Activated, Catalytic Subunit; PTPN13, Protein Tyrosine Phosphatase Non-Receptor Type 13; RCF1, Replication Factor C Subunit 1; RECQL4, RecQ Like Helicase 4; RPA1, RP1 Axonemal Microtubule Associated; RUNX1, RUNX Family Transcription Factor 1; SEM, Stochastic Epimutation; SLX4, SLX4 Structure-Specific Endonuclease Subunit; TFCP2, Transcription Factor CP2; WRN, WRN RecQ Like Helicase; XPC, XPC Complex Subunit, DNA Damage Recognition And Repair Factor.

Francesca Pirini, Luciano Calzari, Daniele Calistri and Davide Gentilini have contributed equally to this study.

in 94 genes associated with hereditary tumors. The cohorts were also characterized for the methylation profile and examined through a sample group analysis and a Stochastic Epigenetic Mutations (SEMs) analysis in comparison with 29 age-matched healthy controls. The multigene panel analysis revealed the presence of pathogenic variants in non-mismatch repair (MMR) genes and three variants classified as pathogenic/likely pathogenic possibly predisposing to LLS. The epigenetic analysis showed epivariations targeting genes associated with LS or DNA repair, most of them associated with the Fanconi Anemia pathway, which could explain the susceptibility to cancer. Our results highlight the need for using extended genetic and epigenetic analyses to understand the causal mechanism of LLS.

KEYWORDS
colorectal cancer, Lynch syndrome, Lynch-like syndrome, methylation, stochastic epigenetic mutations

What's New?

Lynch-like syndrome is associated with development of colorectal cancer (CRC) with microsatellite instability and loss of expression of certain mismatch repair (MMR) genes, similar to Lynch syndrome, but unlike Lynch syndrome, the genetic cause of Lynch-like syndrome (LLS) remains unknown. Here, the authors conducted a genetic and epigenetic study of 32 patients with LLS, 34 with Lynch syndrome, 29 with sporadic CRC, and 29 age-matched controls. A multigene panel revealed pathogenic variants in non-MMR genes, and the epigenetic analysis revealed variations affecting genes associated with the Fanconi anemia pathway.

1 | INTRODUCTION

Lynch-like syndrome (LLS) is a disorder characterized by the development of CRC with microsatellite instability (MSI) and loss of expression of one of the mismatch repair (MMR) genes *MSH2*, *MLH1*, *MSH6*, and *PMS2* detected by immunohistochemistry (IHC). Unlike Lynch syndrome (LS), these two features are not due to the presence of germline alterations in the MMR genes, and the causal mechanism remains unknown.[1] The clinical diagnosis of LS and LLS is reached by applying the Amsterdam criteria[2] and integrating them with the Bethesda criteria.[3] According to the most recent guidelines, all CRC patients are tested for MSI and/or deficiency of MMR genes (dMMR) in IHC and, possibly, for *BRAF* V600E mutation[4] and *MLH1* promoter hypermethylation. If the tests show MSI and/or dMMR (without *BRAF* mutation and *MLH1* hypermethylation), the patients are sent to genetic counseling in order to evaluate the family history and the individual clinical and histopathological features.[2,3] The application of this screening is not sufficient to make a diagnosis, but it allows the identification of high-risk patients.[5] The following step in the diagnostic path is the identification of possible MMR gene alterations. When germline testing confirms the presence of a pathogenic/likely pathogenic variant, the patient is diagnosed with LS; otherwise, they are indicated as LLS. About 55% of the patients suspected of LS, after the genetic test, are classified as LLS cases.[6] LLS patients and their relatives show an increased risk of CRC, even if lower compared to LS patients, suggesting a hereditary component. Moreover, the age at CRC diagnosis is slightly higher compared with LS patients (53.7 years vs. 48.5 years), but still lower compared to sporadic CRC. LLS patients are considered a heterogeneous group with characteristics between LS and sporadic CRC due to the lack of a causal mechanism. Double somatic hits in MMR genes can explain a fraction of LLS cases. However, they are usually found in patients without a family history of cancer; therefore, they should be considered sporadic CRCs. There are at least four potential causes of LLS. First, the presence of alterations on MMR genes still considered as variants of uncertain significance (VUS) or, second, not detected with the conventional diagnostic methods, such as rearrangements or variants in regulatory regions.[7,8] Third, alterations in non-MMR genes like *MUTYH*,[9] *EXO1*, *POLE*, *POLD1*, *MCM8/9*, *WRN*, *BARD1*, *RCF1*, *RPA1*, *MLH3*, and recently also *PPARG*, *CTC1*, *DCC*, *ALPK*, and *PRKDC*[10–12] have been associated with the LLS phenotype. Fourth, additional mechanisms, such as the presence of constitutional epigenetic alterations, can cause a MMR-deficient phenotype and be inherited. Although less frequently, constitutional epimutations in *MLH1* and *MSH2* have been associated with LS through the literature.[13,14] It is not excluded that epimutations in other genes can play a causative role in the LLS.

In the present study, we explored the possible causes of LLS in patients selected after genetic counseling and therefore, with a proven cancer predisposition. Moreover, we used a panel of 94 genes frequently mutated in hereditary tumors and a comprehensive epigenetic characterization to evaluate possible causes and genetic/epigenetic similarities with LS and sporadic cases.

2 | MATERIALS AND METHODS

2.1 | Patients

A total of 95 CRC cases and 29 age-matched controls have been recruited in five different Italian centers. Thirty-two patients were diagnosed with Lynch syndrome, 34 patients were included in the Lynch-like syndrome cohort and 29 patients were diagnosed with sporadic CRC. LS and LLS patients were recruited by the "IRCCS, Istituto Romagnolo per lo Studio dei Tumori (IRST) Dino Amadori" of Meldola, the "Centro di Riferimento Oncologico" (CRO) of Aviano, the "Ospedale Vittorio Emanuele III" of Montecchio Maggiore, and the "University of Modena and Reggio Emilia" (UniMoRe) in Modena. The sporadic CRC patients were recruited by the IRST and the healthy controls by the "Istituto Auxologico Italiano" of Milan. All patients with suspected LS who showed high instability of microsatellites and/or MMR deficiency in IHC were evaluated by genetic counseling and assessed for the risk of hereditary syndromes based on the application of the Amsterdam and Bethesda criteria between 1996 and 2020. The patients that showed high risk were genetically tested to identify possible pathogenic variants in the *MSH2*, *MLH1*, *MSH6*, *PMS2* and *EPCAM* genes. The patients with genetic alterations were diagnosed with LS, and those without alterations were diagnosed with LLS.

2.2 | DNA extraction

For all the cohorts were collected only peripheral blood samples were collected, and all the following experiments were conducted on germline DNA. DNA isolation was performed using QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). DNA quantity and quality were tested by NanoDrop ND 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) with Qubit dsDNA BR Assay Kit.

2.3 | Gene panel sequencing

Sequencing libraries were created starting from 50 ng of germline DNA of LS, LLS, and Sporadic CRC patients, following the TruSight Cancer (Illumina, San Diego, CA) protocol, which allows the analysis of 94 genes associated with hereditary cancers (Table S1). Sequencing was performed on the MiSeq platform (Illumina) with MiSeq Reagent Kit v2 configured for 2 × 150 cycles, according to the manufacturer's instructions.

2.4 | Gene panel bioinformatics analysis

The raw data generated by the sequencer were analyzed with a customized bioinformatic pipeline.[15] FastQ files were aligned to the UCSC-Build37/hg19 reference genome version using BWA software.[16] After alignment, duplicates were removed; sequences were locally realigned around regions with indels using GATK indelRealigner and remapped against Trusight Cancer Panel reference. Unified Genotyper software, GATK version 3.2.2[17] was used to identify single nucleotide variants (SNVs) and indels for variant analysis. The identified variants were subjected to genomic and functional annotation using Annovar.[18]

2.5 | Copy number variation assessment

The presence of large deletions/duplications of MMR genes and *EPCAM* was assessed by Multiplex Ligation-dependent Probe Amplification (MLPA) technique with the P003-MLH1-MSH2, P072-MSH6, and P008-PMS2 kits (MRC Holland, Amsterdam, The Netherlands) or by Hereditary Cancer Solution (HCS)_v1_1 (Sophia Genetics). Coffalyser.net software (MRC Holland) was used for the quantitative analysis of the electropherograms, and the software SOPHIA DDM (Analytics Platform for Genomics) was used for HCS_v1_1 analyses. Sequencing Coverage and Quality Statistics are reported in Table S2.

2.6 | Microsatellite Instability Assessment

MSI analysis has been performed for diagnostic purposes using the Promega MSI Analysis System version 1.2 (Promega, Madison, WI), which amplifies five mononucleotide repeat markers (BAT-25, BAT-26, NR-21, NR-24, and MONO-27) and two pentanucleotide repeat markers (Penta C and Penta D). The fragment analysis of the Multiplex PCR amplification products was carried out using the automated sequencer 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA) The results were visualized with GeneMapper software (Applied Biosystems). As reported in the guidelines,[3] tumors with high microsatellite instability (MSI-H) are defined by at least two unstable microsatellites, cases with one unstable microsatellite are classified as tumors with low microsatellite instability (MSI-L), while all the others are classified as tumors with microsatellite stability (MSS).

2.7 | Variant classification

According to the International Agency for Research on Cancer (IARC), all the genetic variants identified were divided into five classes: Benign (class 1), Likely Benign (class 2), Variants of Uncertain Significance (VUS) (class 3), Likely Pathogenic (class 4) and Pathogenic (class 5).[19] The process of variant classification was performed in accordance with the guidelines of the American College of Medical Genetics (ACMG).[20] The interpretation of MMR gene variants was done using the Leiden Open Variation Database (LOVD) [https://www.lovd.nl/], the International Agency of Research on Cancer–IARC database [https://www.iarc.who.int/], the InSIGHT database [https://www.insight-group.org/variants/databases/], and the ClinVar database [https://www.ncbi.nlm.nih.gov/clinvar/]. The variants of other genes were classified using comprehensive variant databases dbSNP [https://www.ncbi.nlm.nih.gov/snp/], ClinVar, and VarSome [https://varsome.com/].

2.8 | Genome-wide methylation profiling

Sodium bisulfite conversion of 900 ng of high-quality germline DNA was performed by the EZ DNA Methylation Kit (Ref: D5001, Zymo Research Corporation) and Illumina incubation conditions were used. In order to assess conversion yield, a single-strand quantification of bisulfate-converted DNA (bsDNA) was performed using an N60 Implen Nanophotometer. Approximately 200 ng/μl of bisulfite-converted DNA was used for hybridization on Illumina Infinium Methylation EPIC BeadChips. Fluorescent signals were scanned using the Illumina iScan scanner and saved as intensity data files (*.idat). The level of methylation for each CpG site is represented as β-values based on the fluorescent intensity ratio between methylated and unmethylated probes. β-values may range between 0 (non-methylated) and 1 (completely methylated).

2.9 | Differential methylation analyses

The β-value dataset of all samples was generated using the ChAMP package[21]: the quality control step discarded 125,078 probes: 7875 probes with a detection *p*-value above 0.01, 2082 probes with a bead-count <3 in at least 5% of samples, 2961 NoCG probes, 95,575 probes potentially associated with SNPs as identified in Zhou's,[22] 11 MultiHit probes and 16,574 located on XY chromosomes. After the filtering procedure, 740,840 CpG sites were retained. After BMIQ normalization, data were adjusted for batch effect by using ComBat library.[23] An estimate of the proportions of blood cells including $CD8^+$ T cells, $CD4^+$ T cells, natural killer (NK) cells, B cells, monocytes, granulocytes, and others, was assessed by using the DNA Methylation Age Calculator analysis software (https://dnamage.genetics.ucla.edu/).[24] At the group level, differential methylation analysis was conducted by computing *p*-values using the limma method integrated in RnBeads package[25] for both CpG sites and regions (Genes, promoters, CpG island, tiling). To adjust for potential confounding factors, principal component analysis (PCA) was performed to evaluate the association of age, sex, and blood cell estimations with both dependent (disease groups) and independent (methylation values) variables: significant associations were used as covariates in the differential methylation analysis. Individual sample analyses were carried out by identifying Stochastic Epigenetic Mutations (SEMs) as described in[26,27] SEMs represent extreme aberrant methylation data points and were identified, for each CpG site, by comparing the methylation profile of each case to a reference methylation range, calculated from a control population as follows: upper value $= Q3 + (k * \text{IQR})$, lower value $= Q1 - (k * \text{IQR})$, where Q1 represents the first quartile, Q3 corresponds to the third quartile, IQR (Interquartile Range) equals $Q3 - Q1$, and k is set at 3. Outlier values were then classified as hyper-methylated or hypo-methylated with respect to the median values of the controls' corresponding probes. The identification of SEMs enriched regions (Epivariations) was carried out at a genome-wide level through the application of hypergeometric distribution statistical tests on (11-sites) sliding windows. Gene annotation of SEMs/Epivariations was obtained using the web tool wANNOVAR.[28] Organization/investigation of results was conducted according to the disease phenotypes/keywords by using VarElect (The Next Generation Sequencing Phenotyper).[29] Data/results were visualized using specific packages in the R environment. The "ggplot2" package produced PCA charts, boxplots; "Pheatmap" package produced the heatmaps, and "CMplot" the manhattan plots. Linear regressions or the Wilcoxon–Mann–Whitney function were used to evaluate differences in age, cell-type composition, and SEMs number between cases and controls. Unless otherwise stated, the statistical significance threshold was set to False Discovery Rate (FDR) <0.05.

2.10 | Statistical analyses

Patients characteristics were summarized by the median, interquartile range, reporting the first (IQ) and third (IIIQ) quartiles, and minimum and maximum values for continuous variables and by means of absolute frequencies and percentages for categorical ones. Comparisons among patients with LS, LLS, and Sporadic CRC were performed using Pearson's χ^2 test of the Fisher exact test, as appropriate, for categorical variables and through the Analysis of Variance (ANOVA) or the Kruskal–Wallis test for continuous ones. All analyses were performed with STATA 15.0 (College Station, TX).

3 | RESULTS

3.1 | Patient's characteristics

The study recruited LS, LLS, CRC patients, and healthy subjects. The LS cohort was composed of 32 patients diagnosed with LS according to Amsterdam and Bethesda guidelines and for the presence of pathogenic variants in MMR genes identified by sequencing and copy number variant (CNV) analyses. The LLS cohort was composed of 34 patients who fulfilled the Amsterdam and Bethesda criteria but did not present pathogenic variants in MMR genes. The Sporadic CRC cohort was composed of 29 patients diagnosed with CRC at an age >60 years and no relatives with LS-associated cancers, so defined as "Sporadic" (Table 1). Twenty-nine age-matched healthy controls (median age: 55 years, min–max: 35–67 years) were also recruited for the epigenetic analysis. The LS cohort showed a higher percentage of female patients (73.5%) compared with the LLS cohort (43.8%) and sporadic CRC cohort (34.5%). The groups showed significant differences in terms of age and frequency of second tumors. The median age of tumor onset was 43 years for the LS, 56.5 for the LLS, and 68 for Sporadic. The presence of secondary tumors was significantly higher in the LS cohort (68.7%), followed by LLS (29.4%) and sporadic (6.9%). Among LS patients, 31.3% presented a third tumor and 6.2% a fourth tumor. The second tumors were frequent in colon (31.8%) and endometrium (22.7%), in line with the literature and the guidelines[30] while in LLS patients, the tumors were colon (40%) and breast (20%). Overall, LLS patients showed intermediate characteristics between LS and Sporadic.

TABLE 1 Patients characteristics.

	Lynch		Lynch-like		Sporadic		
	32		34		29		
	n	%	*n*	%	*n*	%	*p*
Sex							
F	14	43.8	25	73.5	10	34.5	0.005
M	18	56.3	9	26.5	19	65.5	
Age at diagnosis							
Median [IQ–IIIQ]	43 [36.5–53.5]		56.5 [48–62]		68 [60–77]		<0.001
Min–max	24–74		31–79		50–85		
Tumor site							
Colon (right)	12	41.4	17	51.5	13	48.2	0.469
Colon (left)	3	10.3	6	18.2	6	22.2	
Colon (transverse)	4	13.8	2	6.1	1	3.7	
Rectum	3	10.3	2	6.1	2	7.4	
Sigmoid	1	3.5	2	6.1	4	14.8	
Cecum	3	10.3	4	12.1	1	3.7	
Duodenum	3	10.3	0	–	0	–	
Missing	3		1		2		
T							
T1	3	13.0	2	6.9	2	6.9	0.048
T2	2	8.7	9	31.0	6	20.7	
T3	12	52.2	18	62.1	16	55.2	
T4	6	26.1	0	–	5	17.2	
Missing	9		5		0		
N							
N0	19	73.1	26	81.3	20	69.0	0.733
N1	6	23.1	5	15.6	6	20.7	
N2	1	3.9	1	3.1	3	10.3	
Missing	6		2		0		
M							
M0	22	95.7	29	90.6	22	95.7	0.727
M1	1	4.4	3	9.4	1	4.4	
Missing	2		2		6		
MMR IHC							
MLH1	1	5.0	10	37.0	0	–	<0.001
MSH2	6	30.0	2	7.4	0	–	
MSH6	1	5.0	0	–	0	–	
PMS2	0	–	2	7.4	0	–	
MSH2/MSH6	11	55.0	2	7.4	0	–	
MLH1/PMS2	1	5.0	12	44.4	5	100.0	
Missing	12		6		24		
Microsatelite instability							
MSI-H	22		32		3		
Missing	10		2		26		
Second tumor							
No	10	31.3	24	70.6	27	93.1	<0.001
Yes	22	68.8	10	29.4	2	6.9	

TABLE 1 (Continued)

	Lynch		Lynch-like		Sporadic		
	32		34		29		
	n	%	*n*	%	*n*	%	*p*
Site of the second tumor							
Breast	1	4.6	2	20	1	33.3	
Colon	7	31.8	4	40	0	–	
Endometrium	5	22.7	1	10	0	–	
Uterus	1	4.6	1	10	0	–	
Renal pelvis	2	9.1	0	–	0	–	
Liver	2	9.1	0	–	0	–	
Ovary	1	4.6	0	–	0	–	
Bladder	1	4.6	0	–	0	–	
Prostate	1	4.6	0	–	0	–	
Skin	1	4.6	0	–	0	–	
Tongue	0	–	1	10	0	–	
Melanoma	0	–	1	10	0	–	
Leukemia	0	–	0	–	1	33.3	
Third tumor							
No	22	68.8	32	94.1	29	100.0	<0.001
Yes	10	31.3	2	5.9	0	–	
Site of the third tumor							
Breast	1	10	0	–	0	–	
Colon	2	20	0	–	0	–	
Endometrium	1	10	0	–	0	–	
Uterus	1	10	2	100	0	–	
Hodgkin's lymphoma	1	10	0	–	0	–	
Bladder	3	30	0	–	0	–	
Skin	1	10	0	–	0	–	
Forth tumor							
No	30	93.8	0	–	0	–	–
Yes	2	6.3	0	–	0	–	
Site of the fourth tumor							
Colon	1	50	0	–	0	–	
Kidney	1	50	0	–	0	–	

3.2 | Genomic characterization of LS, LLS and Sporadic CRC

All LS patients presented a high microsatellite instability and/or an aberrant MMR IHC staining, predominantly the loss of expression of MSH2 or MSH2/MSH6. The analysis of the LS cohort with the 94-gene panel (Table S1) and CNV analysis revealed that 78% of the pathogenic variants in MMR genes targeted *MSH2*, followed by *MLH1* (16%), *EPCAM* (6%) and *MSH6* (3%). Four LS patients showed a double pathogenic variant, *EPCAM/MSH2*, *ATM/MSH2*, *BUB1B/MSH2*, *PALB2/MSH2*. One patient showed a 187 Kb novel duplication including *EPCAM* and the first seven exons of *MSH2*, described by our group in 2019.[31] The genetic results were consistent with the MMR IHC results, confirming the prevalence of alterations in the *MSH2* gene. In total, the 32 cases presented 94 VUS (Figure S1). The highest frequency of VUS was detected in the *FANCA* gene (25%), followed by *CDH1* and *ATM* (19%) and *APC* and *DICER1* (12%). All patients in the LLS cohort showed a high microsatellite instability and/or the loss of expression of MLH1 or MLH1/PMS2. Only three pathogenic variants were found in non-MMR genes: *MUTYH* (NM_001128425.2:c.734G > A), *FANCM* (NM_020937.4:c.5101C > T) and *XPA* (NM_000380.3:c.820T > G). In the LLS cohort, 108 VUS were reported, and the genes most frequently mutated were *FANCA* (21%), *ATM* (26%), *SLX4* (18%), *ALK* and *MSH2* (15%) and *WRN* (12%) (Figure S2). The sporadic cohort showed the presence of two pathogenic/likely pathogenic variants in *FANCC*

TABLE 2 List of the VUS classified as likely pathogenic or pathogenic by VarSome Premium.

Position	Ref	Alt	Gene	Classification	Transcript	Codon change	Samples
chr17:29686035	T	G	NF1	Likely Pathogenic	NM_000267	c.8097 + 2 T > G	S15, S3, S20
chr17:41219639	ACA	–	BRCA1	Pathogenic	NM_007294	c.5058_5060del	1396
chr6:35434020	G	C	FANCE	Likely Pathogenic	NM_021922	c.1510-1G > C	M196/01

(NM_001243743.2:c.346-1G > A) and *FANCI* (NM_001113378.2: c.1973delT). Ninety-seven VUS were found in the sporadic cases, and the genes most frequently mutated were *SLX4*, *FANCA*, *MSH2*, *NF1*, *RECQL4* (17%), *ATM*, ERCC2, NBN, and XPC (14%, Figure S3). For research purposes only, all the VUS have been classified by VarSome Premium, and three of them, all located in splicing regions in three sporadic (S15, S3, S20), one LLS (1396) and one Lynch patient (M196/01) were classified as likely pathogenic/pathogenic (Table 2).

3.3 | Methylation profile of LS, LLS, or Sporadic cases compared to healthy subjects

To evaluate the epigenetic contribution to the LLS phenotype, we performed a two-step analysis, which included a sample group analysis (RnBeads) and SEMs analysis comparing the results of the three cohorts with the healthy controls (Figure S4).

3.3.1 | Sample-group characterization: differential methylation analysis between the cohorts

First, an explorative principal component analysis (PCA) was performed. At the CpG site level, no significant separations between the cases (LS, LLS or Sporadic) and the reference population were noted (Figure 1A).

A differential methylation analysis was conducted between groups at the site level by adjusting for potential confounding factors (sex, age and cellular components). The analysis identified 2121 differentially methylated sites (1279 hypermethylated and 842 hypomethylated) in the LS cohort, 266 (214 hypermethylated and 52 hypomethylated) in LLS cohort, and 650 (64 hypermethylated and 587 hypomethylated) in the Sporadic cohort (Figure 1B). Annotation of differentially methylated CpG sites was performed for each comparison (Table S3). Initially, we focused on gene loci, and we found isolated deregulations without any specific gene loci enrichment. To refine our prioritization strategy, we cross-referenced the three gene lists with the genes included in the NGS panel (Table S1). For the LLS patients, a correspondence was found in the *MET* and *SLX4* genes; for the LS cohort in the *SLX4* gene, along with *ALK*, *CDC73*, *EZH2*, *FANCC*, *KIT*, *RECQL4*, *RUNX1*, and *WRN*, while for Sporadic in *EGFR*. For all of these genes, the epigenetic deregulations were observed as isolated positions, scattered heterogeneously across the gene body or within the gene promoter region. At the regional level (genes and promoters), no significant epigenetic deregulation was observed.

3.3.2 | Single case characterization: epigenetic drift

The analysis of epigenetic drift was carried out by examining the burden of SEMs defined Global Epi-Mutation-Load (EML), which occur, for each patient, at specific sites when the methylation status exceeds a reference methylation range. Methodology details are provided in the methods section. The average SEM number in the healthy control group (Reference) was 2270 (median: 1096; IQR: 906–2446), while in the LLS cohort was 6215 (median: 3028; IQR: 1705–8160). For the LS and Sporadic cohorts, the mean values were 9753 (median: 4034; IQR: 1879–12,803) and 4699 (median: 3408; IQR: 1833–6389), respectively. The multiple regression models, adjusted for covariates including sex, age, and cellular components, confirmed a statistically higher burden of SEMs across all three groups (LS: $p = 3.1 \times 10^{-5}$, LLS: $p = 3.8 \times 10^{-3}$, Sporadic: $p = 1.9 \times 10^{-3}$) (Figure 2).

We then focused on the burden of SEMs at the gene level by identifying SEMs that enrich gene loci, called epivariations. For this purpose, we used the validated methodology developed by Gentilini et al. described in materials and methods. To enhance the robustness of the deregulated genes identified across the three cohorts, we removed all gene loci that showed enrichment in SEMs within the control cohort, resulting in univocal gene lists (available as Tables S4--S6). The frequency of hypermethylated and hypomethylated genes across the three cohorts is shown in Table S7.

Restricting prioritization to hypermethylation status, the Venn analysis revealed 1071 unique genes for LS, 593 genes for LLS, and 311 genes for the Sporadic group (Figure 3A). A prioritization of the LLS 593 unique genes for the LS phenotype with VarElect identified 46 genes directly related to the syndrome (Table S8). The gene with the highest association is *MLH1*, followed by *FAN1*, *EPM2AIP1*, and *ERBB2*. The VarElect prioritization of the 1071 unique genes in LS identified 70 genes directly related to the syndrome (Table S9), and the top four genes with the highest association are *MSH2/KCNK12*, *FAN1*, *ERBB2*, and *PTPN13*, all hypermethylated in one patient each. *MSH2/ KCNK12* hypermethylation belongs to a patient with a CNV mutation in *MSH2*. VarElect prioritization for the sporadic cohort indicated the presence of hypermethylation of *FAN1*, *TFCP2*, *PRKDC*, and *HOXA11-AS* in, respectively, 2, 1, 1, and 2 patients (Table S10). The comparison of epivariations among the three patient cohorts did not highlight significant differences. However, we noticed 55 epivariations in common (Figure 3B). The prioritization with VarElect of the 55 genes highlighted the presence of five genes correlated with LS (*FAN1*, *HOXA11-AS*, *OPCML*, *ENTPD1-AS*, *ERBB4*) but also other genes such as *ASCL2 and CAVIN3* are in the list (Table S11).

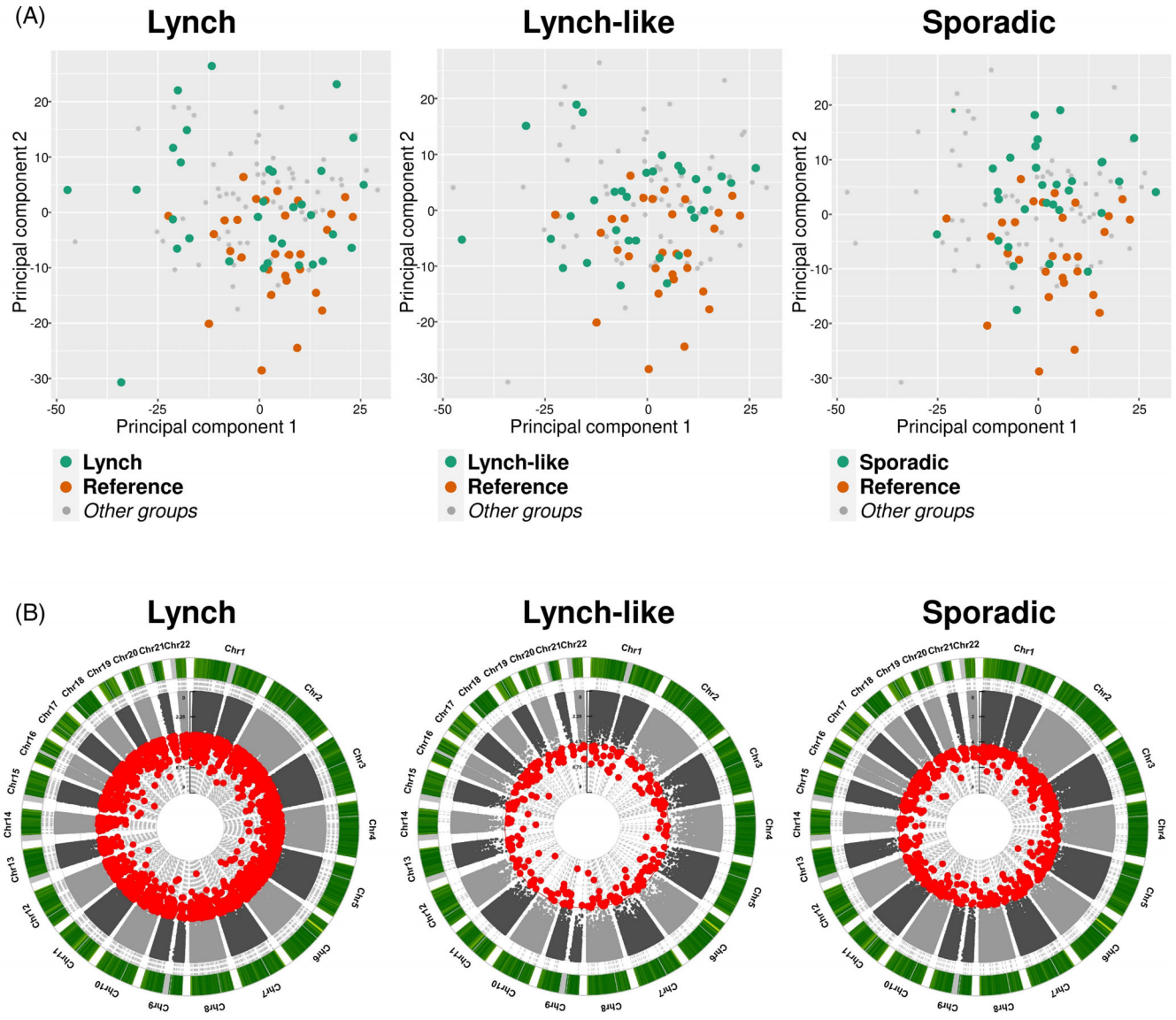


FIGURE 1 (A) Scatter plots from principal component analysis (PCA) depicting the distribution of the sample cohorts across the first two principal components at the CpG site level. (B) Circos plot illustrating the genomic distribution of differentially methylated sites (red dots) across the human genome. These dots are positioned based on their −log10 (unadjusted *p*-value). Notably, the X and Y-chromosomes are excluded from the analysis.

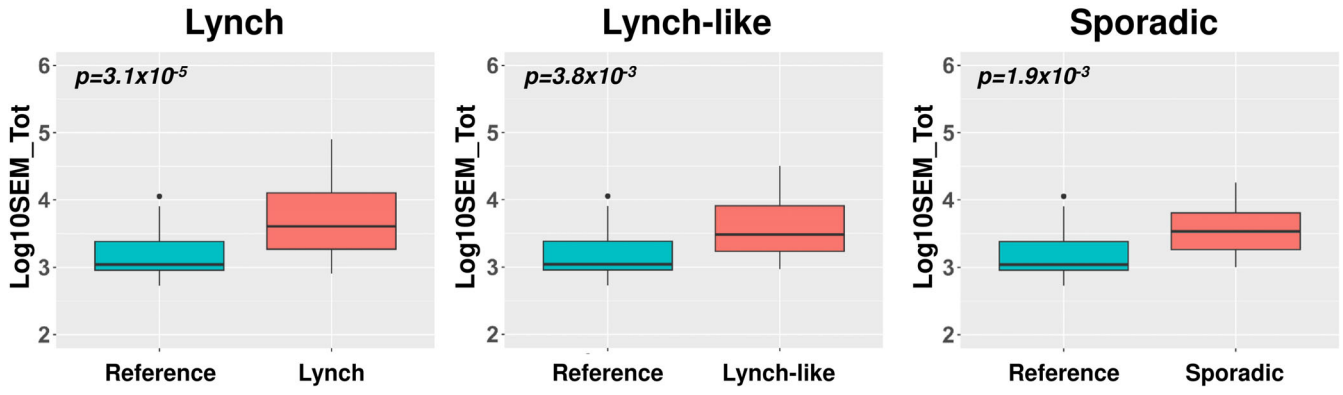
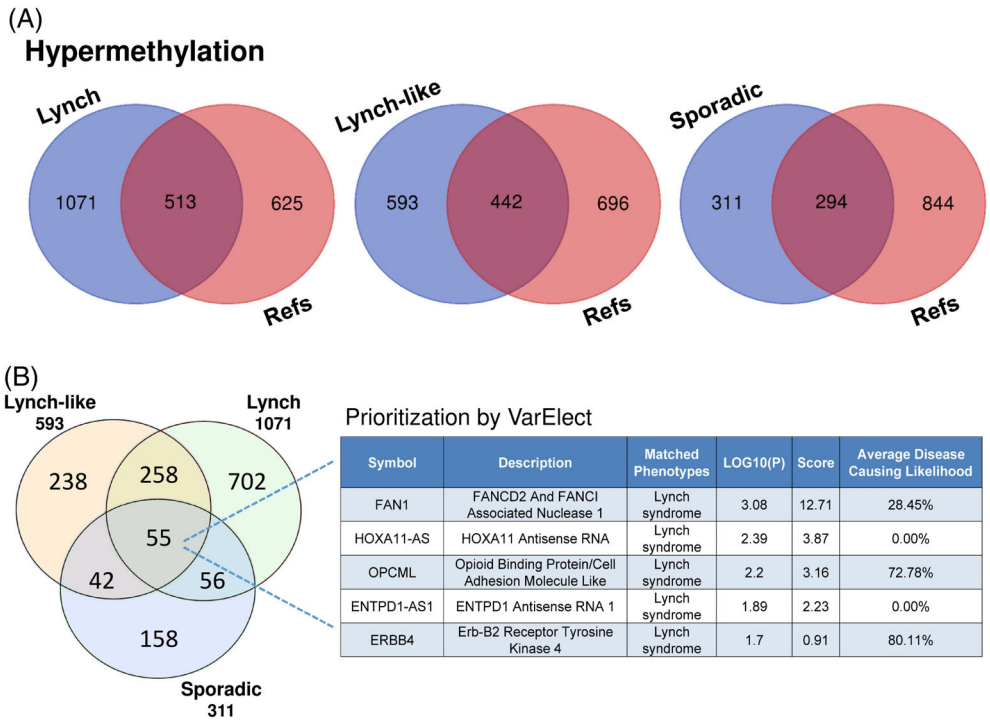


FIGURE 2 SEMs distribution in LS, LLS and Sporadic groups. In each boxplot showing log10 transformed SEMs, the solid horizontal line within the box signifies the median of the dataset, and the box itself illustrates the interquartile range. By default, in the "ggplot" boxplot function, the whiskers stretch to data points that fall within 1.5 times the interquartile range (IQR) from the box. Individual data points beyond this range are depicted as dots, representing outliers.



Symbol	Description	Matched Phenotypes	LOG10(P)	Score	Average Disease Causing Likelihood
FAN1	FANCD2 And FANCI Associated Nuclease 1	Lynch syndrome	3.08	12.71	28.45%
HOXA11-AS	HOXA11 Antisense RNA	Lynch syndrome	2.39	3.87	0.00%
OPCML	Opioid Binding Protein/Cell Adhesion Molecule Like	Lynch syndrome	2.2	3.16	72.78%
ENTPD1-AS1	ENTPD1 Antisense RNA 1	Lynch syndrome	1.89	2.23	0.00%
ERBB4	Erb-B2 Receptor Tyrosine Kinase 4	Lynch syndrome	1.7	0.91	80.11%

FIGURE 3 (A) Venn diagrams illustrating the distribution of hypermethylated epivariations among the LS, LLS, and Sporadic groups, compared to healthy control (Refs) epivariations. (B) Venn diagrams illustrating the distribution of epivariations among the three cohorts and the prioritization of the 55 epivariations in common for LS phenotype.

4 | DISCUSSION

The causes for LLS are still unclear despite the research effort, but possible reasons are the presence of variants of uncertain significance not yet classified as pathogenic, the presence of pathogenic mutations on non-MMR genes (*MUTYH*, *FANCM* and *XPA*) and the contribution of epigenetics. Here we combined the sequencing of 94 genes involved in hereditary pathologies as well as CNV analysis of MMR genes on three cohorts of patients diagnosed through genetic counseling and genetic testing as LS or LLS patients, or patients with sporadic CRC. We then evaluated the contribution of epigenetics by applying a method of analysis, which allows the identification of methylated genes in every single patient.

Our LLS patients present intermediate clinic-pathological characteristics between LS and Sporadic in terms of age of onset and presence of second and third tumors. Furthermore, the tumor sites of secondary tumors include colon but with a high percentage of breast.

At the genetic level, we identified three mutations on non-MMR genes which could be considered the cause of susceptibility and could explain the moderate penetrance of the disease in LLS patients. One is a monoallelic pathogenic variant in *MUTYH*. Biallelic mutations of the gene predispose to MUTYH-associated polyposis (MAP)[32] but in our case, the patient can not be classified as MAP. However, the association of MUTYH monoallelic mutations with an increased susceptibility to CRC cancer risk has been already reported but is still controversial.[33] The *FANCM* gene has been already reported in association with familial breast cancer.[34] *FANCM* encodes a protein that interacts with several DNA repair proteins and that is fundamental in the repair at stalled replication forks.[35] The variant in *XPA* was a stop-loss never reported before, and VarSome confirmed it as pathogenic. This variant was located in exon 6 inside the binding site of the Transcription Factor of Polymerase II H (TFIIH). XPA is a protein involved in nucleotide excision repair (NER) by verifying DNA damage and stabilizing the DNA as it is repaired. Alterations in this gene are associated with the hereditary disease Xeroderma pigmentosum, but it cannot be excluded that dysfunctions in the NER and therefore in XPA may predispose to the onset of CRC.[36] Furthermore, two Sporadic patients presented pathogenic/likely pathogenic variants in *FANCC* and *FANCI*, all genes involved in DNA repair mechanisms. The alteration in *FANCC* is a loss-of-function variant in the splicing region that has been already reported but not in correlation with LS,[37] while the *FANCI* alteration is a frameshift deletion never reported before. Through VarSome analysis of VUSs, we identified 3 other three variants, albeit to be confirmed, in NF1 and BRCA-Fanconi Anemia (FA) pathway (*FANCE*, *BRCA1*) but not reported before in association with LS. The variant in *NF1* (Neurofibrin 1) (c.8160 + 2 T > G) is reported as VUS in ClinVar; however, a possible correlation between *NF1* dysfunctions and an increased risk of developing GIST, gastrointestinal, and colorectal cancer has been reported.[38,39] *BRCA1* is a well-known gene involved in the susceptibility to Hereditary Breast and Ovarian Cancer (HBOC) syndrome,[40] but the non-frameshift deletion c.5058_5060del has never been reported. In addition, the variant in *FANCE* has never been reported before in association with LS. However, recent studies highlight the role of monoallelic mutations in FA genes in increasing the susceptibility to breast/ovarian cancer and, albeit to a lesser extent, in HNPCC.[41] Due to the limitations of the study, it was not possible to evaluate the presence of the variants identified in LLS cases in family members nor to carry out functional studies. In order to elucidate the role of epigenetics in LLS cases, we performed a methylation characterization and analyzed the

data by a two-step analysis. The principal component analysis (PCA) revealed no discernible patterns of epigenetic variation between the cohorts and control group, at CpG sites level and regions (genes/promoters/CpG islands). At the site level, the analysis highlighted group-specific signatures, enabling the identification of some genes shared with the NGS panel. However, the deregulation appears to be driven by individual single alterations rather than by a pattern of robust enrichment of adjacent single deregulations. By the SEM analysis, we observed a significantly higher epigenetic drift in all patient cohorts. Regarding the LLS cohort, the increased SEM burden was confirmed on the methylation profile of public cohorts (GSE128064 and GSE107353: Illumina 450 K BeadChip, $n = 112$ LLS and $n = 41$ controls, respectively) (data not shown).[14] These findings support the hypothesis that LS, LLS, and Sporadic patients improperly accumulated SEMs. An increased epigenetic drift could potentially exert a notable influence on individual health, leading to enhanced genomic instability/abnormal gene expression. At the gene level, we identified univocal Epivariations on genes associated with LS or belonging to DNA repair systems. The prioritization of the hypermethylated epivariations lists provides a number of possible candidates to explain susceptibility to cancer. In the LLS cohort, the gene with the highest association is *MLH1* that we found hyper-methylated in the promoter region in one case, the same LLS case that presented a pathogenic mutation in *MUTYH*, and that presents the lack of expression of *MLH1* and *PMS2*. A similar case has been reported by Zyla R. and colleagues where the patient was ultimately diagnosed with Lynch Syndrome.[42] Another gene with the highest association with LS was *FAN1* hyper-methylated in two cases and involved in DNA repair. Moreover, previous data showed that the interaction between FAN1 and MLH1 prevents MLH1 binding to MSH3, thus inhibiting the assembly of a functional MMR complex.[43,44] The third gene in the list is *EPM2AIP1* that shares the promoter with *MLH1*; it is located head-to-head with *MLH1* and is transcribed in the opposite direction.[45] In fact, *EPM2AIP1* and *MLH1* are both hypermethylated in the same patient. Another gene in the top four is *ERBB2*, which showed a high mutation frequency in CRC patients fulfilling Bethesda or Amsterdam II criteria considered as LS or LLS, but the impact of epimutations is not known.[46] In the LS cohort, the gene with the highest association is *MSH2/KCNK12*, followed by *FAN1*, *ERBB2*, and *PTPN13*. *KCNK12* is a gene coding for a potassium two-pore domain channel subfamily K member 12 located in between *MSH2* and *MSH6*. No direct associations with the syndrome are currently reported, but the gene is often hypermethylated in CRC patients.[47] *PTPN13* somatic mutations have been reported in MMR-deficient CRC, but their role is not clear.[48] In the Sporadic cohort, the genes with the highest association are *FAN1*, *TFCP2*, *PRKDC*, *HOXA11-AS*. *PRKDC* produces a protein involved in cell cycle control that is able to interact directly with *MSH2*. It is also involved in DNA non-homologous end joining (NHEJ), required for double-strand break (DSB) repair and V(D)J recombination.[49] Recently, founder mutations of the *PRKDC* gene have been associated with an increased mutational load in CRC, and thus the gene has been indicated as a new driver of tumor heterogeneity.[50] Over expression of *HOXA11-AS* is correlated with CRC progression and poor prognosis and may promote metastasis, but no data are reported about its hyper-methylation.[51] Analyzing epivariations based on their association with the disease and in common between the cohorts, we found that all three cohorts have patients with epivariations in *FAN1*, *CAVIN3*, and *ASCL2*. *CAVIN3* is a putative tumor suppressor gene, which has been found inactivated in many cancers. Its role is pivotal for correct DNA repair because its protein interacts with *BRCA1*, stabilizing it and ensuring its correct operation. When *CAVIN3* is not expressed, the level of *BRCA1* in the cell decreases, such as the ability to repair the DNA.[52] *ASCL2* is a gene recently reported by a study on epigenetic characteristics of LS. The author described a high level of H3K27me3 in the promoter of stem cell marker genes including *ASCL2*, suggesting its hyper-methylation.[53] Further studies are needed to understand the extent of the contribution of epivariations in terms of expression of proteins, the heritability of these modifications, and their penetrance. Our study identified pathogenic variants in MMR genes and FA pathways that could reassign some but not all the LLS patients to LS cases. The genetic analysis of sporadic patients is instrumental to the identification of pathogenic variants that could increase cancer susceptibility. Our results reinforce the importance of using extensive genetic panels in the diagnostic work-up in order to avoid misclassifications preventing patients from being enrolled in the family prevention and monitoring system. Furthermore, they suggest that genetic analysis should also be extended to patients with sporadic cancer in order to avoid missed diagnoses. Despite the lack of functional validation and familiar penetrance, our results provide novel insights deserving validation in larger cohorts aiding to verify the role and risk of deleterious variants and hypermethylation on non-MMR specific genes on the onset of the disease and their potential inclusion in diagnostic tests.

AUTHOR CONTRIBUTIONS

Francesca Pirini: Conceptualization; investigation; methodology; project administration; writing – original draft; writing – review and editing. **Luciano Calzari:** Data curation; methodology; writing – original draft; writing – review and editing. **Gianluca Tedaldi:** Methodology; validation; writing – review and editing. **Michela Tebaldi:** Methodology; formal analysis; data curation. **Valentina Zampiga:** Writing – review and editing. **Ilaria Cangini:** Writing – review and editing. **Rita Danesi:** Writing – review and editing; investigation. **Mila Ravegnani:** Writing – review and editing; investigation. **Valentina Arcangeli:** Writing – review and editing; investigation. **Alessandro Passardi:** Investigation; writing – review and editing. **Elisabetta Petracci:** Formal analysis. **Sara Bravaccini:** Funding acquisition; writing – review and editing. **Giorgia Marisi:** Writing – review and editing. **Alessandra Viel:** Investigation; writing – review and editing. **Daniela Barana:** Investigation; writing – review and editing. **Monica Pedroni:** Investigation; writing – review and editing. **Luca Roncucci:** Investigation; writing – review and editing. **Daniele Calistri:** Investigation; supervision; writing – review and editing. **Davide Gentilini:** Formal analysis; funding acquisition; supervision.

ACKNOWLEDGMENT

Open access funding provided by BIBLIOSAN.

FUNDING INFORMATION

This study was partly supported by a contribution from Ricerca Corrente by the Italian Ministry of Health within the research line "Genetics and environment in the development and progression of tumors and inhibitory mechanisms. Exposomics, primary and secondary prevention".

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The raw EPIC BeadChip methylation data have been deposited in Zenodo (https://doi.org/10.5281/zenodo.11543344) and will be made available only upon reasonable request. Other data that support the findings of this study are available from the corresponding author upon request.

ETHICS STATEMENT

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines. All participants enrolled in the study have signed an informed consent for the genetic analyses and for the use of the results for research purposes. The study was approved by the local Ethics Committee (CE IRST IRCCS-AVR, protocol 3030/2018; Istituto Auxologico Italiano Ethics Committee, approval number: 2020_07_21_01; CE per le sperimentazione cliniche della provincia di Vicenza, approval n. 24/19; CEUR protocol CRO-2018-63) and was conducted in accordance with ethical standards, the Declaration of Helsinki and national and international guidelines.

ORCID

Francesca Pirini https://orcid.org/0000-0002-7629-2439
Sara Bravaccini https://orcid.org/0000-0002-0075-8538

REFERENCES

1. Rebuzzi F, Ulivi P, Tedaldi G. Genetic predisposition to colorectal cancer: how many and which genes to test? *Int J Mol Sci*. 2023;24:2137. doi:10.3390/ijms24032137
2. 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;116:1453-1456.
3. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst*. 2004;96:261-268.
4. Hampel H, Pearlman R, Beightol M, et al. Assessment of tumor sequencing as a replacement for Lynch syndrome screening and current molecular tests for patients with colorectal cancer. *JAMA Oncol*. 2018;4:806-813. doi:10.1001/jamaoncol.2018.0104
5. Hampel H, Frankel WL, Martin E, et al. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med*. 2005;352:1851-1860. doi:10.1056/NEJMoa043146
6. Buchanan DD, Rosty C, Clendenning M, Spurdle AB, Win AK. Clinical problems of colorectal cancer and endometrial cancer cases with unknown cause of tumor mismatch repair deficiency (suspected Lynch syndrome). *Appl Clin Genet*. 2014;7:183-193. doi:10.2147/TACG.S48625
7. Meyer C, Brieger A, Plotz G, et al. An interstitial deletion at 3p21.3 results in the genetic fusion of MLH1 and ITGA9 in a lynch syndrome family. *Clin Cancer Res*. 2009;15:762-769. doi:10.1158/1078-0432.CCR-08-1908
8. Pérez-Cabornero L, Infante Sanz M, Velasco Sampedro E, et al. Frequency of rearrangements in Lynch syndrome cases associated with MSH2: characterization of a new deletion involving both EPCAM and the 5' part of MSH2. *Cancer Prev Res (Phila)*. 2011;4(10):1556-1562. doi:10.1158/1940-6207.CAPR-11-0080
9. Castillejo A, Vargas G, Castillejo MI, et al. Prevalence of germline MUTYH mutations among Lynch-like syndrome patients. *Eur J Cancer*. 2014;50:2241-2250. doi:10.1016/j.ejca.2014.05.022
10. Xavier A, Olsen MF, Lavik LA, et al. Comprehensive mismatch repair gene panel identifies variants in patients with Lynch-like syndrome. *Mol Genet Genomic Med*. 2019;7:e850. doi:10.1002/mgg3.850
11. Xicola RM, Clark JR, Carroll T, et al. Implication of DNA repair genes in Lynch-like syndrome. *Fam Cancer*. 2019;18:331-342. doi:10.1007/s10689-019-00128-6
12. dos Santos W, de Andra ES, Garcia FA d O, et al. Whole-exome sequencing identifies pathogenic germline variants in patients with Lynch-like syndrome. *Cancers Basel*. 2022;14(17):4233. doi:10.3390/cancers14174233
13. Hitchins MP. Constitutional epimutation as a mechanism for cancer causality and heritability? *Nat Rev Cancer*. 2015;15:625-634.
14. Dámaso E, González-Acosta M, Vargas-Parra G, et al. Comprehensive constitutional genetic and epigenetic characterization of Lynch-like individuals. *Cancers (Basel)*. 2020;12:1799. doi:10.3390/cancers12071799
15. Tedaldi G, Pirini F, Tebaldi M, et al. Multigene panel testing increases the number of loci associated with gastric cancer predisposition. *Cancers (Basel)*. 2019;11:1340. doi:10.3390/cancers11091340
16. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25:1754-1760. doi:10.1093/bioinformatics/btp324
17. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297-1303. doi:10.1101/gr.107524.110
18. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164. doi:10.1093/nar/gkq603
19. Plon SE, Eccles DM, Easton D, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat*. 2008;29:1282-1291. doi:10.1002/humu.20880
20. Richards S, Aziz N, Bale S, 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):405-424. doi:10.1038/gim.2015.30
21. Morris TJ, Butcher LM, Feber A, et al. ChAMP: 450k Chip analysis methylation pipeline. *Bioinformatics*. 2014;30:428-430. doi:10.1093/bioinformatics/btt684
22. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res*. 2017;45:e22. doi:10.1093/nar/gkw967
23. Wang M, Huang J, Liu Y, Ma L, Potash JB, Han S. COMBAT: A combined association test for genes using summary statistics. *Genetics*. 2017;207:883-891. doi:10.1534/genetics.117.300257
24. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14:R115. doi:10.1186/gb-2013-14-10-r115
25. Müller F, Scherer M, Assenov Y, et al. RnBeads 2.0: comprehensive analysis of DNA methylation data. *Genome Biol*. 2019;20:55. doi:10.1186/s13059-019-1664-9
26. Gentilini D, Somigliana E, Pagliardini L, et al. Multifactorial analysis of the stochastic epigenetic variability in cord blood confirmed an impact of common behavioral and environmental factors but not of

in vitro conception. *Clin Epigenetics*. 2018;10:77. doi:10.1186/s13148-018-0510-3
27. Gentilini D, Scala S, Gaudenzi G, et al. Epigenome-wide association study in hepatocellular carcinoma: identification of stochastic epigenetic mutations through an innovative statistical approach. *Oncotarget*. 2017;8:41890-41902. doi:10.18632/oncotarget.17462
28. Chang X, Wang K. wANNOVAR: annotating genetic variants for personal genomes via the web. *J Med Genet*. 2012;49:433-436. doi:10.1136/jmedgenet-2012-100918
29. Stelzer G, Plaschkes I, Oz-Levi D, et al. VarElect: the phenotype-based variation prioritizer of the GeneCards suite. *BMC Genomics*. 2016;17-(Suppl 2):444. doi:10.1186/s12864-016-2722-2
30. Vasen HFA. Clinical description of the Lynch syndrome [hereditary nonpolyposis colorectal cancer (HNPCC)]. *Fam Cancer*. 2005;4:219-225. doi:10.1007/s10689-004-3906-5
31. Pirini F, Tedaldi G, Danesi R, et al. Identification of a novel large EPCAM-MSH2 duplication, concurrently with LOHs in chromosome 20 and X, in a family with Lynch syndrome. *Int J Colorectal Dis*. 2019;34:1999-2002. doi:10.1007/s00384-019-03414-y
32. Nielsen M, Infante E, Brand R. MUTYH Polyposis. In: Adam MP, Feldman J, Mirzaa GM, et al., eds. *GeneReviews® [Internet]*. Seattle (WA): University of Washington; 1993.
33. Barreiro RAS, Sabbaga J, Rossi BM, et al. Monoallelic deleterious MUTYH germline variants as a driver for tumorigenesis. *J Pathol*. 2022;256:214-222. doi:10.1002/path.5829
34. Morales-Pison S, Morales-González S, Fernandez-Ramires R, et al. Association of FANCM mutations with familial and early-onset breast cancer risk in a south American population. *Int J Mol Sci*. 2023;24:4041. doi:10.3390/ijms24044041
35. Panday A, Willis NA, Elango R, et al. FANCM regulates repair pathway choice at stalled replication forks. *Mol Cell*. 2021;81:2428-2444.e6. doi:10.1016/j.molcel.2021.03.044
36. Takebayashi Y, Nakayama K, Kanzaki A. Loss of heterozygosity of nucleotide excision repair factors in sporadic ovarian, colon and lung carcinomas: implication for their roles of carcinogenesis in human solid tumors. *Cancer Lett*. 2001;174(2):115-125. doi:10.1016/s0304-3835(01)00690-5
37. Fostira F, Kostantopoulou I, Apostolou P, et al. One in three highly selected Greek patients with breast cancer carries a loss-of-function variant in a cancer susceptibility gene. *J Med Genet*. 2020;57:53-61. doi:10.1136/jmedgenet-2019-106189
38. Lee JM, Lee JM, Hyun JJ, et al. Intraductal papillary bile duct adenocarcinoma and gastrointestinal stromal tumor in a case of neurofibromatosis type 1. *World J Gastroenterol*. 2018;24:537-542. doi:10.3748/wjg.v24.i4.537
39. Xie R, Fu K-I, Chen S-M, Tuo BG, Wu HC. Neurofibromatosis type 1-associated multiple rectal neuroendocrine tumors: A case report and review of the literature. *World J Gastroenterol*. 2018;24:3806-3812. doi:10.3748/wjg.v24.i33.3806
40. Angeli D, Salvi S, Tedaldi G. Genetic predisposition to breast and ovarian cancers: how many and which genes to test? *Int J Mol Sci*. 2020;21:1128. doi:10.3390/ijms21031128
41. Del Valle J, Rofes P, Moreno-Cabrera JM, et al. Exploring the role of mutations in Fanconi anemia genes in hereditary cancer patients. *Cancers (Basel)*. 2020;12:829. doi:10.3390/cancers12040829
42. Zyla R, Graham T, Aronson M, Velsher L, Mrkonjic M, Turashvili G. MLH1 epimutation is a rare mechanism for Lynch syndrome: A case report and review of the literature. *Genes Chromosomes Cancer*. 2021;60:635-639. doi:10.1002/gcc.22957
43. Porro A, Mohiuddin M, Zurfluh C, et al. FAN1-MLH1 interaction affects repair of DNA interstrand cross-links and slipped-CAG/CTG repeats. *Sci Adv*. 2021;7:eabf7906. doi:10.1126/sciadv.abf7906
44. Goold R, Hamilton J, Menneteau T, et al. FAN1 controls mismatch repair complex assembly via MLH1 retention to stabilize CAG repeat expansion in Huntington's disease. *Cell Rep*. 2021;36:109649. doi:10.1016/j.celrep.2021.109649
45. Lin JC, Jeong S, Liang G, et al. Role of nucleosomal occupancy in the epigenetic silencing of the MLH1 CpG Island. *Cancer Cell*. 2007;12:432-444. doi:10.1016/j.ccr.2007.10.014
46. Kloth M, Ruesseler V, Engel C, et al. Activating ERBB2/HER2 mutations indicate susceptibility to pan-HER inhibitors in Lynch and Lynch-like colorectal cancer. *Gut*. 2016;65:1296-1305. doi:10.1136/gutjnl-2014-309026
47. Kober P, Bujko M, Olędzki J, Tysarowski A, Siedlecki JA. Methyl-CpG binding column-based identification of nine genes hypermethylated in colorectal cancer. *Mol Carcinog*. 2011;50:846-856. doi:10.1002/mc.20763
48. Korff S, Woerner SM, Yuan YP, Bork P, von Knebel Doeberitz M, Gebert J. Frameshift mutations in coding repeats of protein tyrosine phosphatase genes in colorectal tumors with microsatellite instability. *BMC Cancer*. 2008;8:329. doi:10.1186/1471-2407-8-329
49. Oksenych V, Kumar V, Liu X, et al. Functional redundancy between the XLF and DNA-PKcs DNA repair factors in V(D)J recombination and nonhomologous DNA end joining. *Proc Natl Acad Sci U S A*. 2013;110:2234-2239. doi:10.1073/pnas.1222573110
50. Pálinkás HL, Pongor L, Balajti M, et al. Primary founder mutations in the PRKDC gene increase tumor mutation load in colorectal cancer. *Int J Mol Sci*. 2022;23:633. doi:10.3390/ijms23020633
51. Chen D, Zhang M, Ruan J, et al. The long non-coding RNA HOXA11-AS promotes epithelial mesenchymal transition by sponging miR-149-3p in colorectal cancer. *J Cancer*. 2020;11:6050-6058. doi:10.7150/jca.49809
52. McMahon K-A, Stroud DA, Gambin Y, et al. Cavin3 released from caveolae interacts with BRCA1 to regulate the cellular stress response. *Elife*. 2021;10:e61407. doi:10.7554/eLife.61407
53. Reyes Uribe L, Bowen CM, Evans K, Wu H, Duzagac F, Sinha K. EVAbstract 721: epigenetic intervention in the prevention of Lynch syndrome colorectal cancer. *AACR Cancer Res*. 2022;82(12_Suppl):721. doi:10.1158/1538-7445.AM2022-721

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pirini F, Calzari L, Tedaldi G, et al. Comprehensive genetic and epigenetic characterization of Lynch-like syndrome patients. *Int J Cancer*. 2025;157(4):788-799. doi:10.1002/ijc.35451