Genomic Landscape of Hereditary and Early Onset Sporadic Colorectal Cancer

By
Anuja Lipsa
LIFE09201404015

Tata Memorial Centre, Mumbai

A thesis submitted to the

Board of Studies in Life Sciences
In partial fulfillment of requirements

for the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE

January, 2020

Version approved during the meeting of Standing Committee of Deans held during 29-30 Nov 2013

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Anuja Lipsa

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

Anuja Lipsa

List of Publications arising from the thesis

Journal

1. **Lipsa A**., Kowtal P. and Sarin R. "Novel germline STK11 mutations and breast cancer phenotype identified in an Indian cohort of Peutz-Jeghers syndrome". Human Molecular Genetics, **2019**, Vol. 28, No. 11

Chapters in books and lectures notes: NIL

Conferences

1. Awarded **3rd prize in Poster presentation** for poster entitled, "CMMRD & beyond: Varied mutation spectrum of *PMS2* gene producing diverse phenotype" at **3rd Indian Cancer Genetics Conference & Workshop (ICGCW)** held at ACTREC, Navi Mumbai in Nov, 2016.
2. Poster presentation on "Germline mutations of the STK11 gene in Indian Peutz-Jeghers syndrome Patients" at **Indian Society of Human Genetics (ISHG)** hosted by IISc Bangalore in March, 2017.
3. **Best poster award** for poster entitled, "Comprehensive genotypic and phenotypic characterization of the rare CMMRD syndrome in the Indian Cohort" at **2nd International Conference on Founder Population** held at Kochi in Nov, 2017.
4. Oral presentation on "Novel, Recurrent and Founder Mutations in Mismatch Repair Genes in a large Indian Hereditary Non-polyposis Colorectal Cancer (HNPCC) Cohort" at **Indian Association for Cancer Research** hosted by Bose Institute Kolkata in Feb, 2018.
5. Poster presentation on, "Comprehensive genotypic and phenotypic characterization of heterozygous versus homozygous MMR gene mutation carriers in an Indian Cohort" at

European Human Genetics Conference 2018" held at MiCo Milano Congressi, Milan, Italy in June 2018.

Others

i. Participated as Instructor in 3rd *Indian Cancer Genetics Conference and Workshop (ICGCW)* held at ACTREC, TMC in November 2016.

ii. Participated as an organiser in *12th National Research Student Meet (NRSM)* held at ACTREC, TMC in December 2016

iii. Attended NGS workshop in Indian Women Scientists Association, MUMBAI in Feb 2017

Anuja Lipsa

This thesis is dedicated to my caring parents,

Dr. Ajay Kumar Singh and Mrs. Neelam Singh,

My loving siblings,

Prasoon Kumar Singh and Ruchitra Lipsa

and

My beloved husband, Gaurav Mittal

ACKNOWLEDGEMENTS

Undertaking this PhD has truly been a life changing experience for me and it gives me immense pleasure to express my sincere gratitude to several people associated with it. It has been a roller coaster ride for me and it would not have been possible without the support and guidance from many people during my entire tenure.

I would like to express my sincere gratitude to my guide Professor Rajiv Sarin for all the support, guidance and encouragement he gave me throughout my tenure. I am grateful to him for all his contributions, insightful discussions and for providing the infrastructure to complete my thesis work. I am indebted to him for showing confidence in me when I needed the most, for giving me a chance when I lost all hopes and for always appreciating any small or big achievement of mine. Under his supervision I have learned a lot starting from how to analyse data with a research aptitude, how to present results and how to do multitasking. I have never seen a person with so much positive energy despite of doing work at so many fronts and excelling in all. I have always been inspired by his passion and attitude towards both work and life. He has not only contributed in my research work but has always been a big support during times of crisis in my personal life.

I am also very grateful to Dr. Pradnya Kowtal (Co-Investigator, Sarin Lab) for her constant support and valuable suggestions. I have learned many things from her like good lab practices, management of time and resources, being economical and resourceful whenever needed. I thank her for motivating me during tough times both in career and personal life.

I would like to express my sincerest gratitude to the Director of ACTREC Dr. Sudeep Gupta for providing me such a wonderful infrastructure to work with. I am also thankful to Dr. S.V. Chipulnkar (Former Director) for giving me such a great opportunity to pursue my PhD

degree in this institute. I am very thankful to ACTREC-DAE, India for my PhD fellowship and ICMR-CAR for funding the project.

I would also like to thank my doctoral committee members, Dr. Neelam Shirsat, Dr. Sorab Dalal and Dr. Mukta Ramadwar for their time, advice and suggestions in all my DC meetings. Their valuable insights and constant feedback has helped me tremendously in shaping my thesis work.

I would like to express my sincere gratitude and appreciation for all the hard work done by the Genetic Counsellors at the Cancer Genetics Clinic starting from recruiting the patients, collecting their samples and providing all the consents needed to carry out this thesis work. Without their integral help and involvement, this thesis would not have been possible.

I would like to acknowledge all the help I received from past and present members of the ICGC-Lab (ACTREC). I would like to give a special mention to Dr. Poonam Gera for helping me with tissue sample collection, their histopathological evaluation and interpretation of the IHC results. She has been very helpful and has many times gone out of the way to help me. On the personal level also, she has been very motivating and caring.

I would like to thank the DMG-GI team (Tata Memorial Hospital) for referring the patients to CGC. I would also like to thank Dr. Amin, Mr. Dudhal, Mr. Madan from Biorepository (ACTREC), and Mr. Anand and Mrs Manisha (National Tumor Tissue Repository TMH) for their help in providing the tissue samples. I would also like to thank the Pathology Department of ACTREC and TMH for their co-operation in retrieving the FFPE blocks for my study.

A good infrastructure and facility is instrumental to carry out any work. I thank Mr. Naresh Mahida and Mrs. Sharada Haralkar from Genomics Facility for helping me with sequencing and fragment analysis applications. I would also like to mention people from Common Instrument Facility, Library, Stenopool, SCOPE cell, Photography, IT Department, Security,

Accounts and Administration for their timely support. Mr. Nilesh Gardi and Sunil Pachakar from Clinician Scientists Laboratory, ACTREC deserves special credit for helping me in the Bioinformatics based analysis and problems.

All the patients who gave their blood samples and consented to be a part of the study deserves a special mention as without whom this work would not have been possible.

I would also like to thank all the past and present members of Sarin lab for providing a nice and friendly environment to work and share ideas. A special mention to Mr. Parag Madankar for carrying out all the DNA extraction from blood samples of patients that was used in this study. I am thankful to my trainees, Aisha, Zeeshan, Shivani, Bhagyashree, Shabnam, Ishita and Sakshi.

My heart is overwhelmed and filled with joy while expressing my gratitude to my 2013 batchmates. They have provided me a home away from home and have been my support in tough times. I would not have made this far if it was not for their love, care affection and unconditional support. I will cherish all the beautiful memories and funny moments spend with this bunch of people I am proud to call my friends. I would like to give special thanks to my dearest friend, Dhanashree Mundhe for her unconditional love and affection. She has gone out of her way to help me in any personal or professional problems I faced during my tenure. She has been my shield, my voice and my support system without whom it would have been impossible to survive the difficult times. I would also like to give a special mention to the student community of ACTREC; I have never seen such a friendly and helpful work environment anywhere else.

The acknowledgment would be incomplete without thanking the people who mean the world to me, my beloved family. Words cannot expression how grateful I am to my mother and father for all the sacrifices they’ve made on my behalf. Their prayer and believe in me was what sustained me thus far. It was my father’s dream that I pursue PhD and my mother left no

stone unturned in making sure that I do it. I am forever indebted to both of you for always letting me pursue my dreams, treating me like a princess and bearing all the hardships in life so that I can get a good one. I would also like to thank my in-laws for always showering me with love and affection and for being patient with me during my final year. My joy has no limits while acknowledging my mother- in law for being my second mother, accepting me as her daughter and teaching me so many precious lessons in life. She is an epitome of love, care and strength and without whose guidance and support this PhD was not possible. I cannot thank enough my siblings for their unconditional love and support. My sister has been my best friend all my life and I love her dearly for being my support system and my 3 am go to person. My little brother has been the best brother anyone can ask for, he has gone out of his way to do things that would make me happy, be it my birthday or wedding. I'm short of words to express my gratitude to my beloved husband and my amazing life partner who had been by my side throughout this PhD, living every single minute of it be my pain or my joy, and without whom, I would not have the courage to embark on this journey in the first place. These past years have not been an easy ride, both personally and academically. I truly thank Gaurav for instilling in me the confidence and faith with which I could get through it. Thank you for sacrificing so much, being so patient with me and letting me finish my PhD. These five years of long distance has taught me what commitment and determination towards each other means. You are my God send angel and my magic man. I can't tell you enough how much I love you.

Finally I want to thank God for guiding me through my difficult times. You are the one who made sure I get through all the failures and finish my PhD. I know I am your favourite child; I will keep my faith in you forever. Thank you Lord.

Anuja Lipsa

CONTENTS

SYNOPSIS

Homi Bhabha National Institute

SYNOPSIS OF Ph.D. THESIS

1. Name of the Student: Anuja Lipsa

2. Name of the Constituent Institution: ACTREC-TMC

3. Enrolment No.: LIFE09201404015

4. Title of the Thesis: "Genomic landscape of Hereditary and Early onset sporadic Colorectal Cancer"

5. Board of Studies: Life Sciences

6. DC meeting: 14th June, 2019

SYNOPSIS

Introduction:

Colorectal cancer (CRC) is the third most commonly diagnosed cancer in males and the second in females worldwide, with 1.8 million new cases and almost 861,000 deaths in 2018 according to the World Health Organization GLOBOCAN 2018 update. The incidence rate of colorectal cancer is more in developed countries as compared to developing countries. Globally, the incidence rates vary 10-fold with highest rate in Australia, New Zealand, Northern/Western Europe and North America. The incidence rate is intermediate in Eastern and Southern Europe while it is lowest in Africa and South Central Asia. [1, 2] Despite its low incidence rate, the mortality rate is quite high in India. Late stage presentation due to poor awareness, low socio-economic status and, and geo-ethnic differences may account for the same. While the overall age-adjusted incidence rate of CRC is quite low in India, the incidence of early onset recto-sigmoid cancers is disproportionately higher in India. [3, 4, 5]

CRC results from the progressive accumulation of genetic and epigenetic alterations, which lead to the transformation of normal colonic epithelia into benign adenomas, and finally malignant adenocarcinomas. The central dogma concerning the molecular

pathogenesis of colorectal cancer is: 1) Genetic and epigenetic alterations promotes cancer formation process, 2) Cancer develops via a multistep progression both at molecular and morphologic levels, 3) Loss of genomic instability is the key molecular step in cancer formation and, 4) Hereditary cancer syndromes frequently correspond to germ line forms of key genetic defects whose somatic occurrences drive the emergence of sporadic colon cancers. [6]

Based on increasing hereditary influence and cancer risk, colorectal cancer can be classified into three categories: Sporadic CRC, accounting for 70% of cases, where there is no family history of cancer and no deleterious germline mutation in any gene implicated in hereditary colorectal cancer. Approximately 20% cases are Familial CRC comprising of patients with at least one blood relative with CRC or an adenoma, but with no specific germline mutation or clear pattern of inheritance. The remaining 10% are Hereditary CRC cases which result from inheritance of germline mutation in cancer susceptibility genes or the families showing autosomal dominant pattern of inheritance or other related cancers.

Based on the phenotype and underlying molecular mechanisms, **Hereditary Colorectal Cancer (HCRC)** syndromes can be divided mainly into Nonpolyposis and Polyposis syndromes. Lynch syndrome (also known as Hereditary Nonpolyposis Colorectal Cancer) is the major form of Nonpolyposis syndrome and Polyposis syndromes are further classified into adenomatous (e.g. Familial Adenomatous Polyposis) and hamartomatous polyposis syndromes (e.g. Peutz Jeghers Syndrome) [7, 8, 9].

Peutz-Jeghers syndrome (PJS; OMIM #175200) is an autosomal dominant hereditary cancer predisposition syndrome with characteristic phenotypic features of mucocutaneous pigmentation and multiple hamartomatous polyps in the gastrointestinal (GI) tract [10, 11]. It is a rare disorder with an estimated prevalence of 1/50 000 to 1/200 000 [12, 13]. The most prominent features of PJS patients are mucocutaneous pigmentation and/or small bowel intussusception due to GI polyps that occur in the first decade of life [14].

PJS is caused by germline mutations in the tumour suppressor gene, STK11 (OMIM 602216), a serine/threonine kinase localized on chromosome 19p13.3. Around 400

distinct pathogenic or likely pathogenic mutations in STK11 gene in PJS patients have been recorded in Human Gene Mutation database (HGMD, http://www.hgmd.cf.ac.uk) and majority are in the catalytic kinase domain (amino acids 49–309). The reported STK11 mutation detection rate ranges from as low as 10% to as high as 90% depending upon the criteria used for defining the PJS cases and the genetic screening method employed [15,16]

Most of the genotype and phenotype data in PJS families is from the Caucasian population. The genotype–phenotype correlation in PJS has not been examined in detail in the Asian population, particularly the South Asians. So far only a single PJS family with pathogenic mutation in STK11 has been reported from India [17].

Lynch Syndrome (LS, OMIM #120435), also known as hereditary non-polyposis colon cancer syndrome (HNPCC), is an autosomal dominant, highly penetrant (80-85%) syndrome which accounts for 2-5% of all CRC cases. HNPCC is clinically characterized by a family history of colorectal cancer at early age, predominance of tumours in the proximal colon, a high frequency of synchronous and metachronous colorectal cancers, and an association with extracolonic cancers. The tumour spectrum of LS is thus highly heterogeneous and is continuously evolving as more and more studies are coming up. HNPCC is a genetically heterogeneous disease and is caused due to germline mutations in one of the DNA-mismatch-repair (MMR) genes, predominantly in *MLH1*, *MSH2*, and *MSH6* and rarely in *PMS1* and *PMS2* [18]. Mutations in *MMR genes* are found throughout these genes, without any obvious hot spots. The genetic and phenotypic heterogeneity associated with this syndrome poses a challenge for identification of the individuals at risk. Various clinical criteria like Amsterdam criteria and Bethesda guidelines were developed and subsequently revised to facilitate syndromic diagnosis of LS [19]. Also, various screening methodologies like immunohistochemistry (IHC) and microsatellite instability (MSI) testing have been employed to aid in the process of genetic testing. Most of the molecular and phenotypic data on LS comes from Western and some Asian populations [20]. It is now being increasingly recognized that the frequency and the types of mutations differ in different geographical areas, and therefore, the frequency of extra-colonic cancer associated with LS is also likely to differ in different populations [21]. Therefore an

understanding of LS mutation and tumour spectrum and genotype-phenotype correlation is critical to the development of efficient strategies for LS diagnosis and genetic testing in each population. A mutation of LS caused by biallelic MMR gene mutation has been recognized in the recent past.

Constitutional Mismatch Repair Deficiency (CMMRD; OMIM #276300) syndrome is a rare autosomal recessive distinct childhood cancer predisposition syndrome with less than 200 cases reported worldwide [(22)]. A constitutively defective MMR system in CMMRD syndrome results in a wide variety of malignant tumours in early childhood, mainly haematological malignancies, brain tumours, early-onset CRCs, and HNPCC-associated tumours [(23)]. CMMRD syndrome is caused by biallelic germline mutations in any one of the four MMR genes, *MLH1* (OMIM #120436), *MSH2* (OMIM #609309), *MSH6* (OMIM #600678) and *PMS2* (OMIM #600259). CMMRD is also characterized by several non-neoplastic features like that of NF-1, particularly café-au-lait macules (CALMs), hyper- or hypopigmentation, axillary freckling and Lisch nodules [(24, 25)]. Due to its rarity and an unusual tumour spectrum which overlaps with several other syndromes, CMMRD is an under diagnosed syndrome for which the genotype-phenotype data are still emerging. Only about 200 CMMRD cases with confirmed biallelic MMR gene mutations have been reported globally with very few reports from South-Asia and no report from India.

Sporadic colorectal cancer (SCRC) is thought to be influenced by diet, lifestyle, environmental factors, and acquired somatic mutations in genes involved in hereditary CRC. However, early onset sporadic colorectal cancer (EOSCRC) is quite different from sporadic CRC. The three major genetic mechanisms responsible for sporadic CRC, namely chromosomal instability (CIN); microsatellite instability (MSI) and the CpG island methylator phenotype (CIMP) pathways are not commonly identified in EOSCRC [(26)]. Recent reports suggest that the incidence rate of early onset sporadic CRC has increased in India. The etio-pathogenesis of early onset CRC, in the absence of family history or polyposis, is not known. And the major drawback of the few studies done on the molecular pathogenesis of this subset of CRC is the age cut off which is variable and not properly defined. However, increase in the incidence of

sporadic CRC, especially Sporadic Recto-sigmoid cancer in very young age group (<40 yrs) has been reported from India. The established pathways of CRC carcinogenesis, such as chromosomal instability/aneuploidy, microsatellite instability, APC-mutation, *KRAS*-mutation accounts only for very small subset of early onset sporadic Recto-sigmoid cancer patients, thereby suggesting the existence of an alternative genetic pathway [(27)]. Our understanding of sporadic early-onset Recto-sigmoid cancer is meagre and their survival rate remains poor. It is imperative therefore to identify molecular pathways that drive tumourigenesis in this poorly studied CRC subtype in order to improve patient management and treatment strategies.

Colorectal cancer: Indian Scenario

Molecular studies on CRC from developing nations have been sparse with most of the clinical and genetic data on the hereditary CRCs from Western and East Asian populations. There is a lack of knowledge in the South Asian patients with only 28 MMR families from India from 3 studies and no cohort studies on FAP other than from our group. There is only one report of STK11 mutation carrier PJS family and no report of *PMS2* or any other MMR gene biallelic mutation carrier CMMRD family from India**.**

In the past decade increased incidence of sporadic rectal cancers in early age group has been observed from the Indian context and elsewhere however a suggestive cause has not been identified yet. The established pathways of CRC carcinogenesis- CIN, MIN, CIMP accounts only for very small subset of EOS sporadic Recto-sigmoid cancer patients, thereby suggesting the existence of an entirely distinct genetic pathway.

Hypothesis:

In the multi-ethnic Indian population, the genetic landscape and phenotypic manifestations of genes underlying major colorectal cancer predisposition syndromes may be distinct from the Caucasian population and this may have clinical relevance.

In early onset sporadic recto-sigmoid cancers, there may be completely different and yet unknown genetic factors responsible which needs to be elucidated for establishing novel diagnostic/prognostic/therapeutic regime for young onset patients

Aim of the study: We propose to study the key genetic alterations in various forms of Hereditary Colorectal Cancers in an Indian cohort and to elucidate the somatic landscape of early onset recto-sigmoid tumours from patients below 40 yrs of age with no family history of CRC associated cancers

Objectives:

1) To define the mutation spectrum of key genes- MMR and STK11 in different types of Hereditary Colorectal Cancers in the Indian cohort and study genotype-phenotype correlation

   1.1) To identify novel/recurrent germline mutations in Peutz-Jeghers syndrome-STK11/LKB1 and study the genotype-phenotype correlation

   1.2) To identify novel/recurrent germline mutations in Lynch Syndrome/Hereditary Non-Polyposis Colorectal Cancer (HNPCC) – MMR study known and novel genotype-phenotype correlation

   1.3) To identify novel/recurrent germline mutations in CMMRD syndrome- *PMS2* and other MMR genes

2) To identify the somatic mutational spectrum of early onset Sporadic Rectal cancer in tumour and matched paired normal tissue samples using NGS technology (whole exome sequencing)

Patients and Methods

Patients were registered at Cancer Genetics Clinic in TMH after syndromic diagnosis of PJS, LS and CMMRD based on the evaluation of family and medical history of patients. Detailed clinical characterization was done for each patient to understand the phenotypes and genotype-phenotype correlations. Blood or Mouthwash samples were collected after taking written informed consent. DNA was extracted from the lymphocytes or buccal cells isolated from these samples using QiaAmp Blood Mini kit or by Phenol Chloroform method.

Mutation analysis of STK11 and MMR genes

PCR amplification of the coding region of STK11 gene for PJS cases and MMR genes (MLH1, MSH2 and MSH6) in case of Lynch syndrome was carried out using specific

primers designed for each exon of the genes. *PMS2* analysis is quite difficult largely due to the presence of multiple pseudo-genes, the majority of which share homology with the 5' region of *PMS2*, covering exons 1-5. Another pseudo-gene, *PMS2*CL, has strong homology to exons 9 and 11–15. For amplification of *PMS2* gene, we adopted the established long range PCR technique to avoid *PMS2* pseudogene coamplifiaction [28, 29]. Amplification of long-range PCR products was confirmed by gel electrophoresis prior to exon-specific amplification. Exon-specific amplification of *PMS2* gene was carried out using long range amplified product as template and exon specific primers. 1% agarose gel electrophoresis was performed to check if the amplification was successful. Sanger sequencing of the amplified products was done after purification of PCR products with EXO-SAP IT. Chromatograms were analysed using Chromas Lite software and matched with reference sequences to identify the germline mutations in these genes. MLPA analysis was carried out using SALSA MLPA kits from MRC-Holland; P101 (STK11) and P003 (MLH1 and MSH2) and to detect the large genomic rearrangements (LGRs) in STK11 and MMR genes in the cases where no point mutation and small indels could be identified through Sanger sequencing. Data was analysed using Coffalyser.Net software from MRC-Holland.

To understand the somatic landscape of early onset sporadic rectal cancer (EOS-RC)

To understand the somatic landscape, tumour and adjacent normal mucosa tissue samples were collected from 30 EOS-CRC patients below the age of 40 years, with no family history and fitting our criteria i.e. with poorly differentiated tumour with mucinous type / signet ring cell type, mostly aggressive /advanced stage. We adopted a 2 step approach for genetic analysis of EOS-RC. The first step was to screen 10 such samples to evaluate the status of the three known canonical pathways to rule out their role in Indian EOS-RC cohort: The Mismatch repair pathway (Microsatellite Instability pathway), the Wnt mediated pathway (Chromosomal Instability pathway) and the CpG island methylator pathway (Serrated pathway).

We collected FFPE blocks, blood and frozen tissue samples from all these patients after their informed consent. FFPE blocks were collected after reviewing the slides for the same to choose tumour rich blocks with the help of the collaborating pathologist.

DNA was isolated from FFPE blocks, tumour and normal tissues. The isolated DNA was quantified using spectrophotometric assays. The quantity and quality of the isolated DNA was determined using A260 /280and A260 /230 values and by analysing on a 0.8% agarose gel. In order to evaluate the MSI and CIMP pathway, IHC was performed for MMR proteins using standard protocol. STR markers (BAT25, BAT26, D2S123 and D17S250) were used to study genetic instability. The approach was to PCR amplify these 5 STR markers from the tumour DNA as well as the paired normal DNA followed by fragment analysis using automated sequencer. As BRAF mutation usually occurs in concordance with hypermethylation of MLH1 promoter, we also screened Exon 15 of BRAF, where the most common and only mutation found in CRC occurs i.e. in codon 600. Next, to evaluate the status of CIN pathway, we carried out IHC for β-catenin to look for nuclear stability under background of APC mutation. In addition, we also screened for the commonly occurring somatic mutations in CIN pathway- the mutation cluster region of APC (Exon 15), exon 3 of CTNNB1, exon 2 of *KRAS* and exon 15 of BRAF gene.

For **Whole exome sequencing,** we selected 15 EOS-RC patients for whom treatment-naïve tumour samples were available**.** DNA was extracted from the 15 paired tumour/normal samples using the PAXgene tissue DNA kit (Qiagen). Quality check and quantification was done by Nanodrop Spectrophotometer and by Agarose gel Electrophoresis. Exome capture was carried out using SureSelectXT Human All Exon V5 (Agilent) which covers 50Mb target region and sequencing was done using Illumina Hiseq 150. Both exome capture and sequencing was done at MedGenome, Bangalore. Sequencing was carried out at 100X coverage for both tumour and normal mucosa tissue. The FASTQ files were demultiplexed and analysed using FASTQC software to check for the quality of reads generated. The fastq files were then aligned against the reference human genome hg19 using BWA mem which generates sam files. After conversion of sam to bam files using samtools, sequence PCR duplicates were removed using Picard tools. Mpileup files were generated using Samtools, which is the acceptable format for Varscan2 that was used for mutation calling. The mutations were then annotated using ANNOVAR. Mutation prioritization was carried out by

filtering in all the known and novel exonic mutations which includes frameshift mutations, splice site mutations and nonsynonymous mutations.

Results & Discussion

I. Objective 1:

1.1 To identify novel/recurrent germline mutations in Peutz-Jeghers syndrome-STK11/LKB1

A total of 19 unrelated South Asian PJS families were included in this study. 11/19 patients had family history of PJS associated cancers or polyps. 17/19 probands had PJS-associated cancers characterized by mucocutaneous pigmentation or hamartomatous polyps or both. The median age of onset of the disease in the probands was 39 years (range, 22–56 years). The primary cancer site was breast in eight cases, GI tract in five cases (small bowel in two and colorectal in three cases), ovary in two cases, uterus and nasopharynx in one case each. Pathogenic or likely pathogenic germline heterozygous mutations in the STK11 gene were identified in 7/19 probands. Screening of relatives in these seven families with pathogenic mutations identified three additional individuals harbouring the family-specific STK11 pathogenic mutation. All three carriers had mucocutaneous pigmentation and GI polyps were seen in two carriers who have undergone GI endoscopy till date.

Table 1. Phenotypic features and variant spectrum of STK11 gene found in the Indian PJS cohort

Family (proband/ relative)	Gender/age at last FU or death	Upper GI polyps	OCP	Cancer site (age at diagnosis)	Syndrome (FH+/−)	Known or [a]novel STK11 variants	STK11 polymorphisms (polymorphisms co-occurring in multiple cases)	Other gene variants
PJ1 (Proband)	F/46	Yes	Yes	IDC III bilateral breast (38, 46) ER/PR/Her2 +/−/−	Classical PJS (FH−)	[a]Exon 7 del	c.290+36G→T; (c.375–49G→A; c.464+40_46dup)	NT
PJ1a (Nephew)	M/31	Yes	Yes	Nil	Classical PJS (FH−)	Exon 7 del		NT
PJ1b (Brother)	M/61	Yes	Yes	Nil	Classical PJS (FH−)	Exon 7 del		NT
PJ2 (Proband)	F/45	Yes	Yes	PDA ascending colon (22)	Classical PJS (FH+)	Exon 2–3 del		NT
PJ2a (Mother)	F/62	NK	Yes	IDC II breast (60) ER/PR +/+	Suspected PJS	Obligate carrier		
PJ2b (Sister)	F/38	NK	Yes	Small bowel (23)	Suspected PJS	Obligate carrier		
PJ3 (Proband)	M/41	Yes	Yes	MDA small bowel (24)	Classical PJS (FH−)	c.842_843insC	[b]c.478 C→T (L160L)	NT
PJ3a (Son)	M/14	NK	Yes	Nil	Classical PJS (FH−)	c.842_843insC		NT
PJ4 (Proband)	M/21	Yes	Yes	Nil	Classical PJS (FH−)	Exon 2–3 del		NT
PJ5 (Proband)	F/34	Yes	No	Gonadoblastoma ovary (25)	Suspected PJS (FH−)	c.842_843insC		NT
PJ6 (Proband)	F/47	NK	Yes	IDC III breast (26) ER/PR/Her2 +/+/−	Suspected PJS (FH−) DD HBOC	[a]c.542 A→T	(c.290+36G/G→T/T); c.290+78C→T); c.375–49G/G→A/A; c.465–51T/T→C/C	No BRCA1 hotspots
PJ7 (Proband)	F/54	NK	Yes	IDC III breast (49) ER/PR+ve, Her2−ve	Suspected PJS (FH−) DD HBOC	[a]c.542 A→T		No BRCA1 hotspots

Abbreviations: FU indicates follow-up; NK, not known; OCP, oro-cutaneous pigmentation; IDC, infiltrating ductal carcinoma; PDA, poorly differentiated adenocarcinoma; MDA, moderately differentiated adenocarcinoma; FH, family history of first-degree relative with PJS-associated cancer; DD, differential diagnosis of syndrome; del, deletion; ins, insertion; dup, duplication; NT, not tested; HMZ, homozygous [a]novel variant; [b]novel likely benign STK11 variant. Homozygous variants are represented by both alleles (e.g. GG→TT), and heterozygous variants are represented by single allele (e.g. C→T). The STK11 reference sequence used was NM_000455.4 from NCBI.

Interestingly, the breast cancer histology and IHC in all the three carriers of STK11 pathogenic mutations (PJ1, PJ6, PJ7) and one STK11 obligate carrier (PJ2a) was nearly identical—infiltrating duct carcinoma grade III in three cases and grade II in one case and positive for oestrogen receptor (ER) in all four cases, progesterone receptor (PR)-positive in three of four cases and Her2-negative in all three cases in

whom this information was available. Despite breast being one of the most common PJS-associated cancer sites, the immunohistochemical (IHC) and molecular features of these tumours in carriers of STK11 mutation is not known. In a large cohort of over 10 000 women with triple-negative breast cancer evaluated for germline pathogenic mutations with multigene Next Generation Sequencing (NGS) panels, which included STK11, pathogenic mutations were identified in 22 genes but not in the STK11 gene [(30)]. This inverse association between TNBC and STK11 germline pathogenic mutations derived from a very large TNBC cohort supports the finding of our study where all the four cases were ER-positive.

Status of objective 1.1: Finished and published- **Lipsa A**., Kowtal P. and Sarin R. Novel germline STK11 mutations and breast cancer phenotype identified in an Indian cohort of Peutz-Jeghers syndrome. **Human Molecular Genetics**, 2019, Vol. 28, No. 11

1.2 To identify novel/recurrent germline mutations in Hereditary Colorectal Cancers – MMR study known and novel genotype-phenotype correlation

Germline mutation analysis of the MMR genes was undertaken for 85 Lynch syndrome cases. Deleterious germline mutations in MMR genes identified in 82 HNPCC families. This high mutation detection rate of 96.47% in our cohort can be attributed to the use of comprehensive syndromic diagnostic criteria, genetic screening approach and the use of IHC as pre-screening technique. MLH1 and MSH2 mutations accounted for 90% of all the mutations identified in LS families. 62/82 was distinct deleterious germline mutations, of which 34 are novel. Of the 62 MMR mutations, 34 were in MLH1 gene (48 families), 26 were in MSH2 gene (32 families) and 2 were in MSH6 gene (2 families). 90% MMR mutations were in MLH1 or MSH2 genes, with MLH1 mutations scattered throughout the gene while MSH2 mutation clustered in proximal part of the gene encoding the DNA binding domain and MSH6/MSH3 interaction domain. Different types of mutation were prevalent in the two predominantly mutated genes, MLH1 and MSH2. Truncating mutations were the most common in both MLH1 and MSH2 gene followed by splice site (26% of all mutations) mutations in the MLH1 gene, while LGRs in the MSH2 gene (16% of all mutations).

Phenotyping was carried out in 215 affected members from 82 mutation carrier families. The phenotype and genotype-phenotype correlations in the MLH1 and MSH2 mutation carriers are largely in accordance with the known literature. The mean number of cancers per family was same in both groups. No difference in the mean age at diagnosis between males and females among MLH1 and MSH2 mutation carriers**.**

CRC was the most common cancer in both MLH1 and MSH2 while extracolonic cancers were more common in MSH2 carriers. As breast cancer was second most common extracolonic malignancy in our cohort (identified in 12 cases) we advocate the inclusion of breast cancer in the LS tumour spectrum.

Status of Objective 1.2: Finished and manuscript in preparation

1.3 To identify novel/recurrent germline mutations in *PMS2* associated syndromes (CMMRD syndrome)

In this study of comprehensive genomic study to identify novel and recurrent germline mutations in CMMRD syndrome, a total of 5 unrelated South Asian CMMRD suspected families were included: 4 from North Indian states of UP, Bihar and Rajasthan and 1 from South Indian state of Tamil Nadu. 19 tumours in 13 patients from 5 *PMS2* mutations carrier and 1 MLH3 mutation carrier families : 5 high grade gliomas, 6 haematological malignancies, 6 Lynch-syndrome associated tumours and 2 mediastinal tumour. Family history of CMMRD associated cancers was noted in 4/5 families, parental consanguinity in 2/5 families while 2 families belonged to same community and Café-au-lait spots: 2/3 probands for which data was available. We identified 6 individuals with confirmed biallelic pathogenic *PMS2* mutations and 3 obligate *PMS2* mutation carriers from 5 families screened for *PMS2* mutations. We also report a rare case of a 30y old female with endometrial cancer and a family history of colon cancer suggestive of LS was analysed for the four MMR genes, *MLH1, MSH2, MSH6* and *PMS2* however when no pathogenic or likely pathogenic mutation was identified in these genes. It was then taken up for hereditary panel testing which revealed a novel homozygous likely pathogenic mutation (c.320delT, p.Val107GlyfsTer11) in exon 2 of *MLH3* gene. Our data is supported by another report of a biallelic MLH3 mutation in a distinct polyposis syndrome, where a

germline nonsense mutation was identified in 4 patients with polyposis syndrome and a distinct clinical and molecular phenotype [31].

Status of Objective 1.3: Finished and manuscript in preparation

II. Objective 2:

I] To identify the somatic mutational spectrum of early onset Sporadic Rectal cancer in tumour and matched paired normal tissue samples using NGS technology (whole exome sequencing)

We adopted a 2 step approach: I] Screening of 10 EOS-RC cancers below 40yrs of age for the known canonical pathways to rule out their role in EOS-RC -a) MSI pathway: All 10 tumours exhibited nuclear positivity for the 4 MMR proteins on IHC: MLH1, MSH2, MSH6 and *PMS2* suggestive that all these 10 tumours are MMR-proficient. STR markers (BAT25, BAT26, D2S123 and D17S250) were used to study genetic instability. The approach was to PCR amplify these 5 STR markers from the tumour DNA as well as the paired normal DNA followed by fragment analysis using automated sequencer. In one of the case, instability was seen at two markers (D5S346 and D17250). In another case, genetic instability was seen at D2S123. No instability was observed at any marker in the remaining cases. b) CIN pathway: We screened for the canonical gene mutations- the mutation cluster region of APC (Exon 15), exon 3 of CTNNB1 and exon 2 of *KRAS*. No mutation was found in any of these genes in any of the 10 tumours. However, in order to further validate the results and get a clear picture of what percentage of EOS-CRC is driven by Wnt, we carried out IHC for β-catenin to look for nuclear stability under background of APC mutation. All 10 tumours exhibited membranous positivity for β-catenin protein indicating that these tumours are not driven by the Wnt pathway. Comparative analysis of all these results suggests that almost all (> 80%) of the EOS-RC are not driven by the known canonical pathways and indicate towards the possibility of an alternative pathway driving their pathogenesis.

II] To define the somatic mutational landscape of early onset (<40years) sporadic rectal cancer in the Indian population, we performed exome capture DNA sequencing on 15 tumour/normal pairs. Quality of the raw sequence data was analysed using

FastQC. Mean and median base Phred quality score for all bases was above 30. All exomes were sequenced at >100X average depth of coverage in both tumour and normal tissue except for 1 samples which was sequenced at 53X coverage for tumour and 85X for normal tissue. More than 95% of the targeted exome region was covered at least 10X and more than 80% of the targeted exome was covered at least 30X in each sample except for 2 samples covered at 64.76X and 73.66X. In total we detected 5311 high confidence somatic mutations in the 15 EOS-RC paired samples. After removing intergenic, intronic, unknown and synonymous mutations, there were 2581 exonic mutations (which includes truncating, splicing and nonsynonymous mutations). Mutation burden per sample- 0.28-1.88 per Mb for 14 samples (median no. of somatic mutations= 68, non-hypermutated tumour) and 32.8/Mb for 1 sample (total no. of mutations= 1640, hypermutated tumour). This hypermutator phenotype in 1 sample is accounted to presence of POLE somatic mutation as it has been reported that tumours with POLE or POLD1 mutations are characterized by an extremely high mutation frequency (>1 million per genome). C/G>T/A transitions were the most common type of substitution in 13/15 all cases as opposed to G/C to T/A transversions being most frequent in the TCGA Rectal Adenocarcinoma (RC)-US data. A second more stringent mutation filtering including only truncating mutations, splice sites and nonsynonymous missense mutations predicted to be pathogenic by ≥5/11 in-silico tools revealed 15 recurrently mutated genes identified in more than 1 sample. Interestingly, this somatic mutation distribution differs greatly from the ones reported in TCGA RC. The mutation detection rate in the canonical genes also known as "hill" genes were either significantly lower, except TP53 and SMAD4 or mostly absent in our cohort of early onset RC. The mutation rate of TP53 was similar to the TCGA RC data; however the rate of SMAD4 mutation was significantly higher (40%) when compared to US late onset RC (12%) from TCGA data. There are few reports where loss of SMAD4 was shown to be a feature of early onset CRC [32, 33]. One report also showed that SMAD4 mutation was higher in younger age group than older age group [34].These results suggest that the absence of SMAD4 could be a marker of the worse behaviour of early onset CRC, particularly in the microsatellite stable (MSS) group. Notably one of the recurring genes was a E3 ubiquitin ligase RNF43 which negatively regulates Wnt by ubiquitination and subsequent degradation of the Frizzled receptor.

Somatic mutations of *RNF43* are reported in only 0.7% of US late onset rectal cancers TCGA. These 3 cases do not have APC mutation which is in concordance with literature that RNF43 and other RNF family members may serve as potential alternative to APC mutation as a mechanism for altering the Wnt signaling pathway in EOS-RC. Altogether, the data points towards a possibility that tumourigenesis in sporadic early onset rectal tumours proceeds via an alternate pathway other than the known canonical pathways of CRC. However, no definite conclusions can be made owing to the relatively low discovery set which merits further studies on more such samples and validation on an independent cohort.

Summary and Conclusion

In summary, we report the first comprehensive study on an Indian PJS cohort including 12 carriers/obligate carriers of STK11 mutations. This is the first study to investigate and identify the association between germline STK11 pathogenic mutation and ER and Her2 status of PJS-associated breast cancer. In Lynch syndrome cohort, we have achieved a high mutation detection rate for MMR gene mutations in our Lynch Syndrome cohort which showed that MLH1 and MSH2 gene mutations account for majority of the mutations in LS families. A high frequency of novel mutations MMR genes has been identified in our study. The genotype and phenotype and the genotype-phenotype associations in our LS cohort largely concurs with the known spectra with few notable exceptions. Ours is the first report of first comprehensive study on 5 *PMS2* mutation carrier CMMRD families from India which added to the existing knowledge of CMMRD genotype and phenotype with an emphasis on the importance of syndromic differential diagnosis. In addition, we also report a possible association of MLH3 biallelic mutation with CMMRD which needs to be explored further on additional cohorts to establish the significance of MLH3 as a possible cause of cancer predisposition in CMMRD patients. In the early onset rectal cancers, there was significantly high frequency of SMAD4 as compared to late onset US RC TCGA cohort which indicates towards a possibility of SMAD4 being a biomarker for the worse behavior of these subset of rectal cancers. Identification of RNF43 mutation occurring mutually exclusive with APC in 3 cases points towards a possibility of an alternative mechanism of WNT deregulation in early onset rectal

cancers. Study on early onset rectal cancer points towards a possibility of distinct somatic events in these tumours due to very low frequency of the known canonical gens and identification of several other less known genes. However, this needs to be confirmed and validated on a larger cohort which may help in delineating the alternate pathway that explains CRC tumourigenesis in early onset sporadic rectal cancer.

References:

1. Parkin DM et al. *"Estimates of the worldwide incidence of eighteen major cancers in 1985"*. Int J Cancer 1993; 54:594-606.
2. Finlay IG and Pickford IR. "Colorectal carcinoma, in Lawson DH (ed): Current Medicine 4th edition", Edinburgh, Scotland, Churchill Livingstone, 1994; 51-64.
3. Curado MP et al.: "Cancer Incidence in Five Continents, Volume IX." IARC Scientific Publication, No 160 IARC, Lyon 2007.
4. Nath J. et al. "*Rectal cancer in young adults: a series of 102 patients at a tertiary care centre in India*". Colorectal Disease 2008; 11:475–479.
5. National Cancer Registry Programme. Accessed at: http://www.icmr.nic.in/ncrp/cancer_reg.htm on November 4, 2010.
6. Kheirelseid A.H. et al. *"Molecular biology of Colorectal Cancer".* American Journal of Molecular biology 2013; 3:72-80.
7. Fearnhead N.S. et al. *"Genetics of colorectal cancer: hereditary aspects and overview of colorectal tumourigenesis"*. British Medical Bulletin 2002; 64:27–43.
8. Hisamuddin I.M. and Yang V.W. "*Genetics of Colorectal Cancer*". Medscape General Medicine 2004; 6(3):13.
9. Jasperson K.W. et al. *"Hereditary and Familial Colon Cancer".* Gastroenterology 2010; 138:2044 –2058.
10. Amos C.I. et al. "*Fine mapping of a genetic locus for Peutz–Jeghers syndrome on chromosome 19p".* Cancer Res., 1997; 57:3653–3656.
11. Hemminki A. et al. *"Localization of a susceptibility locus for Peutz–Jeghers syndrome to 19p using comparative genomic hybridization and targeted linkage analysis".* Nat. Genet., 1997; 15:87–90.
12. Jenne D.E. et al. *"Peutz–Jeghers syndrome is caused by mutations in a novel serine threonine kinase".* Nat. Genet., 1998; 18:38–43.

13. Hemminki A. et al. *"A serine/threonine kinase gene defective in Peutz–Jeghers syndrome.Nature"*. 1998; 391:184–187.
14. Kopacova M. et al. "*Peutz–Jeghers syndrome: diagnostic and therapeutic approach"*. World J. Gastroenterol., 2009; 15:5397–5408.
15. Hearle N., et al. *"Frequency and spectrum of cancers in the Peutz–Jeghers Syndrome"*. Clin. Cancer Res., 2006; 12:3209–3215.
16. Papp J. et al. *"High prevalence of germline STK11mutations in Hungarian Peutz–Jeghers Syndrome patients"*. BMC Med. Genet., 2010; 11:169.
17. Thakur N. et al. *"A novel mutation in STK11 gene is associated with Peutz–Jeghers Syndrome in Indian patients".* BMC Med. Genet., 2006; 7:73.
18. Cohen S.A. and Leininger A. *"The genetic basis of Lynch syndrome and its implications for clinical practice and risk management".* The Application of Clinical Genetics, 2014; 7:147-158.
19. Lin-Hurtubise K.M. et al. *"Lynch syndrome: expanded tumour spectrum, universal screening and multimodal treatment strategies for colon cancer".* Transl Gastrointest Cancer 2015; 4(5):367-372.
20. Chew M.H. et al. *"Phenotypic characteristics of hereditary non-polyposis colorectal cancer by the Amsterdam criteria: an Asian perspective".* Anz J Surg. 2008; 78(7):556–60.
21. Wei W. et al. *"Racial differences in MLH1 and MSH2 mutation: an analysis of yellow race and white race based on the InSiGHT database".* J Bioinform Comput Biol. 2010; 8:111–25.
22. Suerink M. et al. "*Constitutional mismatch repair deficiency as a differential diagnosis of neurofibromatosis type 1: consensus guidelines for testing a child without malignancy".* J Med Genet, 2019; 56(2):53-62.
23. Wimmer K. and Kratz C.P. "*Constitutional mismatch repair-deficiency syndrome".* Haematologica, 2010; 95(5): 699-701.
24. Polubothu S. et al: "*Atypical dermal melanocytosis: a diagnostic clue in constitutional mismatch repair deficiency syndrome".* Br J Dermatol, 2017; 177(5):e185-e186.

25. Tabori U. et al: "*Clinical Management and Tumour Surveillance Recommendations of Inherited Mismatch Repair Deficiency in Childhood".* Clin Cancer Res, 2017; 23(11):e32-e37.
26. Stigliano V. et al. "Early-onset colorectal cancer: A sporadic or inherited disease?" World J Gastroenterol 2014; 20(35):12420-12430.
27. You Y.N. et al. *"Young-Onset Rectal Cancer: Presentation, Pattern of Care and Long-term Oncologic Outcomes Compared to a Matched Older-Onset Cohort".* Ann Surg Oncol 2011; 18:2469–2476.
28. Clendenning M. et al. *"Long-range PCR facilitates the identification of PMS2-specific mutations*". Human Mutation, 2006.
29. Vaughn C.P. et al. "*Clinical Analysis of PMS2: Mutation Detection and Avoidance of Pseudogenes*". Human Mutation, 2010.
30. Shimelis H. et al. *"Triple-negative breast cancer risk genes identified by mutligene hereditary cancer panel testing"*. J. Natl. Cancer Inst., 2018; 110:855–862.
31. Olkinuora A. et al. "*Biallelic germline nonsense mutation of MLH3 underlies polyposis predisposition".* Genet Med. 2018, doi: 10.1038/s41436-018-0405-x.
32. Royce S.G. et al. *"The role of SMAD4 in early-onset colorectal cancer".* Colorectal Dis., 2010; 12(3):213–219.
33. Perea J. et al. *"SMAD4 in early onset colorectal cancer".* Colorectal Dis. 2010; 12(9):948.
34. You Y.N. *et al. "Somatic mutations in young onset colorectal cancer".* Journal of Clinical Oncology, 2015; 33:623-623.

Publications in Refereed Journal:

a. Published:

1) **Lipsa A**., Kowtal P. and Sarin R. "Novel germline *STK11* mutations and breast cancer phenotype identified in an Indian cohort of Peutz-Jeghers syndrome". **Human Molecular Genetics**, 2019, Vol. 28, No. 11
2) Khan N., **Lipsa A.**, Arunachal G., Ramadwar M. and Sarin R. "Novel mutations and phenotypic associations identified through *APC*, *MUTYH*, *NTHL1*, *POLD1*, *POLE* gene analysis in Indian

Familial Adenomatous Polyposis cohort". **Scientific Reports,** 22nd May, *2017*

b. Accepted: NA

c. Communicated: NA

Other Publications:

a. Book/ Book Chapters: NA

b. Conference/Symposium:

1. Awarded **3rd prize in Poster presentation** for poster entitled, "CMMRD & beyond: Varied mutation spectrum of *PMS2* gene producing diverse phenotype" at **3rd Indian Cancer Genetics Conference & Workshop (ICGCW)** held at ACTREC, Navi Mumbai in Nov, 2016.
2. Poster presentation on "Germline mutations of the *STK11* gene in Indian Peutz-Jeghers syndrome Patients" at **Indian Society of Human Genetics (ISHG)** hosted by IISc Bangalore in March, 2017.
3. **Best poster award** for poster entitled, "Comprehensive genotypic and phenotypic characterization of the rare CMMRD syndrome in the Indian Cohort" at **2nd International Conference on Founder Population** held at Kochi in Nov, 2017.
4. Oral presentation on "Novel, Recurrent and Founder Mutations in Mismatch Repair Genes in a large Indian Hereditary Non-polyposis Colorectal Cancer (HNPCC) Cohort" at **Indian Association for Cancer Research** hosted by Bose Institute Kolkata in Feb, 2018.
5. Poster presentation on, "Comprehensive genotypic and phenotypic characterization of heterozygous versus homozygous MMR gene mutation carriers in an Indian Cohort" at **European Human Genetics Conference 2018"** held at MiCo Milano Congressi, Milan, Italy in June 2018.

Participation:

1. Participated as Instructor in 3rd *Indian Cancer Genetics Conference and Workshop (ICGCW)* held at ACTREC, TMC in November 2016.
2. Participated as an organiser in *12th National Research Student Meet (NRSM)* held at ACTREC, TMC in December 2016
3. Attended NGS workshop in Indian Women Scientists Association, MUMBAI in Feb 2017

Signature of Student: Date: 21/06/2019

Doctoral Committee:

S. No.	Name	Designation	Signature	Date
1.	Dr. Neelam Shirsat	Chairperson		21-6-19
2.	Dr. Rajiv Sarin	Guide & Convener		21.6.19
3.	Dr. Sorab Dalal	Member		21/6/19
4.	Dr. Mukta Ramadwar	Member		21/6/19

Forwarded Through:

8/7/19

Dr. S.V. Chiplunkar
Chairperson,
Academic &Training Program, ACTREC

(Prof. Dr. S. V. Chiplunkar)
Chairperson, Academic &
Training Programme , ACTREC

Dr. S. D. Banavali
Dean, Academics
T.M.C.

PROF. S. D. BANAVALI, MD
DEAN (ACADEMICS)
TATA MEMORIAL CENTRE
MUMBAI – 400 012.

To,
Dean,
Homi Bhabha National Institute
BARC

Encl: as above

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ACF	Aberrant crypt foci
AC-I/II	Amsterdam Criteria I/II
ALL	Acute lymphocytic leukaemia
AML	Acute myeloid leukaemia
APC	Adenomatous polyposis coli
BAM	Binary Alignment Map
BRRS	Bannayan-Riley Ruvalcaba syndrome
BWA	Burrows Wheelers Aligner
CCS	Cronkhite-Canada syndrome
CDK	Cyclin dependent kinase
CFS	Cancer Family Syndrome
CGC	Cancer Genetics Clinic
CHRPE	Congenital Hypertrophy of the Retinal Pigment Epithelium
CIMP	CpG Island Methylation Pathway
CIN	Chromosomal Instability
CLL	Chronic lymphocytic leukaemia
CML	Chronic myeloid leukaemia
CMMRD	Constitutional Mismatch Repair Deficiency syndrome
CRC	Colorectal cancer
CS	Cowden syndrome
CTD	C-terminal domain
dbSNP	Database of Single Nucleotide Polymorphism
EBV	Epstein-Barr virus
EDM	exonuclease domain mutation
EOSCRC	early onset sporadic colorectal cancer
ERK	Extracellular signal Regulated Kinase
ExAC	Exome Aggregation Consortium
FAP	Familial Adenomatous Polyposis
FFPE	Formalin Fixed Paraffin Embedded

GBD	Global Burden of Diseases
GI	Gastrointestinal
HBOC	Hereditary Breast and Ovarian Cancer
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HGMD	Human Gene Mutation database
HIV	Human Immunodeficiency Virus type-1
HMPS	Hereditary Mixed-Polyposis Syndrome
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
HPR	Histopathological review
HPV	Human Papilloma Virus
HRP	Horse Radish Peroxidase
HUGO	Human Genome Organization
HVP	Human Variome Project
IBD	Inflammatory Bowel Disease
IDC	Infiltrating Ductal Carcinoma
IHC	Immunohistochemistry
ILO	International Labour Organization
InSiGHT	International Society for Gastrointestinal Hereditary Tumors
IRB	Institutional Review Board
IRB	Institutional Review Board
JPS	Juvenile polyposis syndrome
LFS	Li–Fraumeni syndrome
LGR	Large Genomic Rearrangements
LHS	Laugier-Hunziker syndrome
LS	Lynch Syndrome
LOH	Loss of heterozygosity
MAP	MUTYH associated polyposis
MAPKK	Mitogen-activated protein kinase kinase
MDA	Moderately differentiated adenocarcinoma
MMR	Mismatch Repair

MSI	Microsatellite instability
MSS	Microsatellite stable
NHL	Non-Hodgkin's lymphoma
NGS	Next Generation Sequencing
NLS	Nuclear localization signal
NTD	N-terminal domain
NTTR	National Tumour Tissue Repository
PCNA	Proliferating Cell Nuclear Antigen
PDA	Poorly differentiated adenocarcinoma
PIKCA	Phosphatidyl Inositol 4,5-bisphosphate 3-kinase catalytic subunit α
PJS	Peutz-Jeghers Syndrome
PPAP	Polymerase Proofreading-Associated Polyposis
OCP	Oro-cutaneous pigmentation
RPA	Replication protein A
RFC	Replication Factor C
SAM	Sequence Alignment Map
SCTAT	"sex-cord" tumours with annular tubules
SEER	Surveillance Epidemiology and End Results
SNV	Single Nucleotide Variants
SPSS	Statistical Package for the Social Sciences
TBE	Tris/Borate/EDTA
TCGA	The Cancer Genome Atlas
TGFβ	Transforming growth factor β
TIL	Tumour Infiltrating Lymphocytes
MB	Tumour Mutational Burden
TNM	Tumour, Node, Metastasize
UICC	Union for International Cancer Control
UTR	Untranslated Region
UV	Ultraviolet
WES	Whole Exome Sequencing
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

1.1 Introduction to Cancer

Cancer has existed through all of human history. Edwin Smith Papyrus which was written in 3000 BC is the first documented proof of cancer with description of breast cancer. Hippocrates (460-375 BC) described cancer by a Greek word καρκίνος which means crab or crayfish due to the appearance of the cut surface of a solid malignant tumour which resembles a moving crab [1]. Cancer can be described as a large group of disease that results from abnormal and uncontrolled cell division due to failure of mechanisms that control growth and proliferation. The human body is made of 37.2 trillion cells which are strictly regulated by signals for growth, apoptosis, differentiation, cell–cell interactions and cell–extracellular matrix interactions. Cancer cells, on the other hand disregard these signals in order to survive thereby resulting in uncontrolled growth and proliferation [2]. Unlike normal cells, cancer cells lack contact inhibition therefor have the potential to invade, disseminate from their primary site to surrounding tissue or distant organs through blood or lymphatic system [3].

1.1.1 Classification of cancer

Cancer can arise in most of the cell types and organs of the human body and is classified based on the type of cell where the tumour originated:

- Carcinoma: tumour that arise from the epithelial tissue lining of internal organs. This category includes the most common cancer like breast, prostate, lung and colon
- Sarcoma: tumour that arise from connective tissue. These include bone, muscle, fat, cartilage, tendon, ligament etc.

- Leukaemia: type of cancer that arise from the blood and bone marrow. These include Acute myeloid leukaemia (AML), Chronic myeloid leukaemia (CML), Acute lymphocytic leukaemia (ALL) and Chronic lymphocytic leukaemia (CLL)
- Lymphoma: malignancies of the lymphatic system. Lymphomas are broadly classified into Hodgkin's lymphoma and Non-Hodgkin's lymphoma (NHL)
- Myeloma: cancer of the plasma cells of bone marrow
- Blastoma: tumour that arise in embryonic tissue
- Central nervous system cancer: cancers that arise in the tissue of brain and spinal cord

Cancer can also be classified based on increasing hereditary influence and cancer risk into following groups:

- **Hereditary cancer:** Account for 5% of all cancers. It is caused due to germline mutations in cancer susceptibility gene that are inherited from parents. It is characterized by higher cancer risk and early age of onset than the general population, risk of multiple cancers in the same individual and blood relatives with same type or related type of cancer
- **Familial cancer:** Usually account for 20% of all cancers. It is likely caused by a combination of genetic and environmental risk factors. Individuals with familial cancer may have one or more relatives with the similar cancer; however, there is no specific pattern of inheritance. It is characterized by variable age of onset and classical features of hereditary cancer syndromes are usually not present
- **Sporadic cancer:** Almost 75% of all cancers are sporadic. It is caused due to nonhereditary factors which include environmental exposures, lifestyle factors such

as diet, smoking, natural factors like hormones, aging and other influences that may cause DNA damage. Usually there is no family history and late age of disease onset.

1.1.2 Causes of cancer

Cancer is caused by accumulation of defects in the DNA of a transformed cell. These defects may be caused either due to known environmental and lifestyle risk factors or some unknown causes. It may also be influenced by a person's genetic makeup. All these causes either independently or in combination may result in cancer development. Cancer risk factors can be broadly classified into four groups:

- Biological agents such as bacteria, viruses and parasites which are responsible for up to 20% of human cancers worldwide. These include Human Immunodeficiency Virus type-1 (HIV) causing Kaposi's sarcoma and Non-Hodgkin's lymphoma. Hepatitis B Virus (HBV) and Hepatitis C Virus (HCV) cause hepatocellular carcinoma (liver cancer). Human Papilloma Virus (HPV) may cause cervical, vaginal, vulvar, oropharyngeal and penile cancers. Epstein-Barr virus (EBV) causes Burkitt lymphoma, non-Hodgkin lymphoma, Hodgkin lymphoma and nasopharyngeal carcinoma. *Helicobacter pylori* which causes gastric carcinoma. Schistosomes (*Schistosoma hematobium*), a parasite fluke is associated with bladder cancer and liver flukes (*Opisthorchis viverrin*) parasite cause liver cancer [4].
- Occupational risk factors include chemicals, radioactive materials and asbestos. As per the International Labour Organization (ILO) around 2 million deaths per year worldwide, among 2.7 billion workers, are attributed to workplace exposure.
- Environmental factors such as chemicals like asbestos and benzene are known to cause cancer. Prolonged exposure to asbestos fibres is associated with

mesothelioma. Various sources of radiation such gamma and X-rays may also cause cancer such as bone cancer, leukaemia, and central nervous system cancers. Prolonged exposure to UV radiation can cause melanoma and other skin cancers. Lifestyle related factors such as tobacco smoking and chewing which is associated with many forms of cancers such as mouth, larynx, oesophagus, head, neck, lung, stomach, bladder, kidney and pancreas; particularly 90% of lung cancer is caused due to tobacco smoking. Excessive consumption of alcoholic beverages may also cause cancer. Chemicals such as nitrites and poly aromatic hydrocarbons generated by barbecuing food may also cause cancer. Among other lifestyle factors such as red meat, processed meats, salted fish, low fibre diets and obesity are also associated with susceptibility to cancer [5, 6].

- Genetic risk factors that act through accumulation of genetic alterations during cell division may result in cancer development. These genetic alterations can either be inherited from parents, known as germline mutations or can be acquired during one's lifetime as a result of error during cell division or from exposure to DNA damaging carcinogens, known as somatic or acquired changes.

1.1.3 Cancer: Global and Indian Scenario

Cancer is the major cause of mortality and morbidity worldwide. It is the second leading cause of death globally after cardiovascular diseases. Cancer is one of the major public health burdens in both developed and developing countries. The burden is expected to continue rising due to ageing and population growth [7]. Globally, around 1 in 6 deaths occur due to cancer. An estimated 18.1 million new cases and 9.6 million cancer associated deaths occurred in 2018 as per GLOBOCAN update [8]. Lung cancer is found to be the most commonly diagnosed cancer and leading cause of

cancer related death in both sexes combined, followed by breast, prostate and colorectal [8]. As per World Health Organization (WHO), there will be 27.5 million new cases of cancer each year by 2040 and an estimated 12 million yearly deaths by 2030 worldwide (WHO 2011). In India over 1.1 million new cancer patients registration and 0.78 million people cancer death were recorded in 2018. The most common cancers in both sexes combined were stomach cancer (9·0%), breast cancer (8·2%), lung cancer (7·5%), lip and oral cavity cancer (7·2%), pharynx cancer other than nasopharynx (6·8%), colon and rectum cancer (5·8%), leukaemia (5·2%), and cervical cancer (5·2%) as per Global Burden of Diseases (GBD), injuries and risk factors study 2016 [7]. In India more than 80% of cancers present in advanced stages which makes their management difficult [9]. Approximately 70% of cancer associated deaths occur in low- and middle-income countries like India due to lack of cancer awareness, late diagnosis and lesser access to affordable curative services [7].

1.1.4 Hallmarks of cancer

Cancer is an immensely complex and diverse disease which is generally classified based on the cell or organ type from which they originate resulting in 100 different types of cancers. However, a set of characteristics are shared among almost all these different types of cancers. These characteristics, known as hallmarks of cancer, was first presented by Hanahan and Weinberg in the year 2000 as general rules that govern the transformation of normal cells into malignant cancers. The originally proposed hallmarks of cancer include self-sufficient growth promoting signals, insensitivity to growth-inhibitory signals, resisting programmed cell death, unlimited replicative potential, inducing angiogenesis, and activating tissue invasion and metastasis [10]. The list was further extended after decades of research thereby including emerging

hallmarks such as deregulating cellular energetics and avoiding immune response. Additionally, enabling characteristics were proposed, which include tumour promoting inflammation, and genome instability and mutation [11]. The hallmarks and enabling characteristics are schematically represented in Fig 1.1

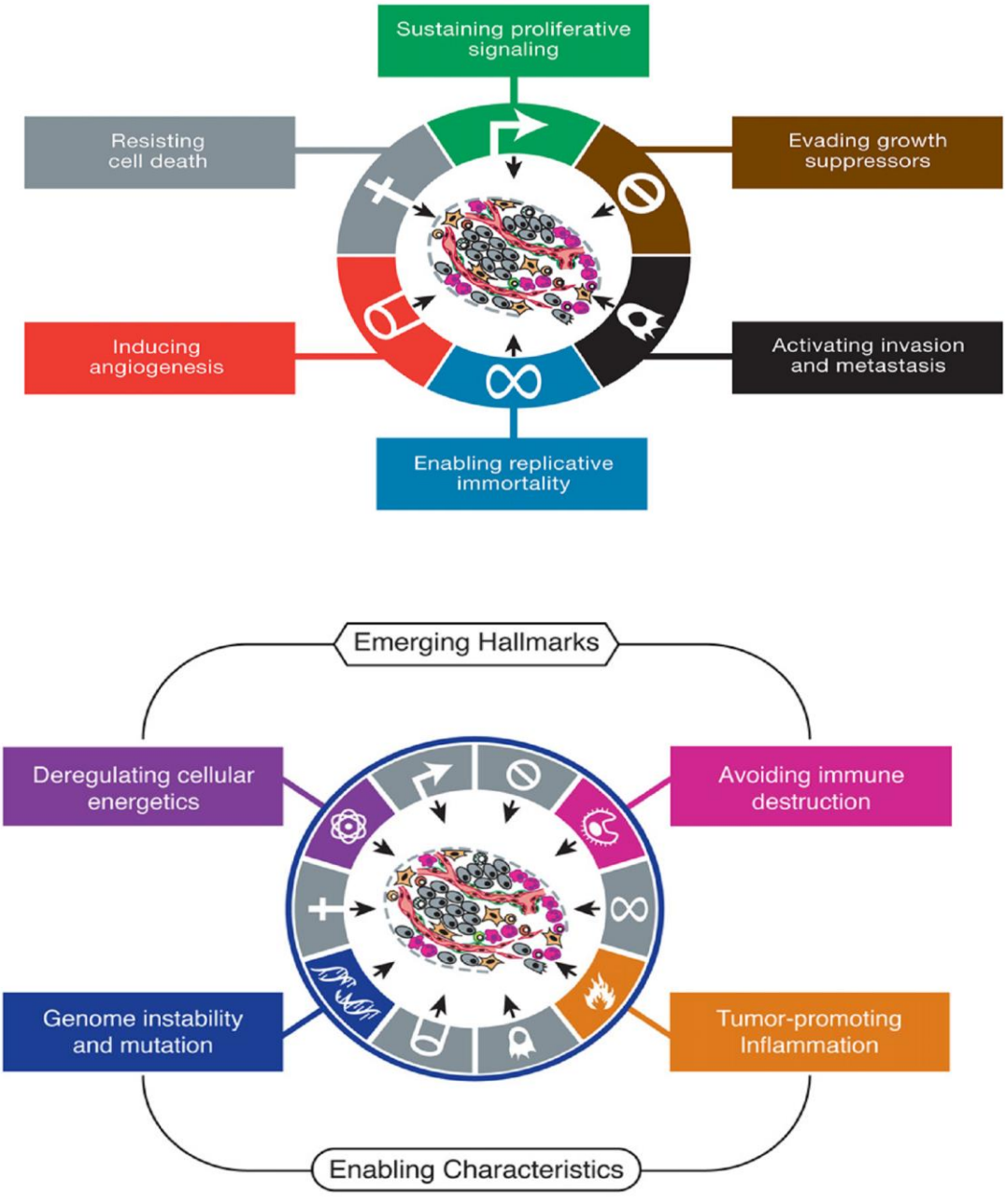


Figure 1.1 Hallmarks of Cancer: Upper Panel-First set of Hallmarks [10]

Bottom Panel: Emerging Hallmarks and Enabling Characteristics [11]

1.1.5 Cancer: a genetic disease

Cancers arise as a result of numerous alterations that have occurred in the DNA of a cancerous cell. These alterations commonly known as mutations can either be transmitted through the germline of an individual and result in cancer susceptibility or can be somatically acquired. It has been suggested that the great majority of cancers arise when two to eight sequential alterations have occurred, during several decades, in genes with functions relevant to cancer [12]. There are about 22,000 genes in each human cell however random mutation within any gene cannot cause cancer. Only when mutations occur in certain key genes that play important role in tumour growth and suppression will lead to cancer development. These key genes can be grouped into two classes depending on their mutation patterns and the effect of the mutations on gene function and cellular processes:

- **Proto-Oncogenes:** Also known as growth promoting genes, these are dominant genes that play a role in regulation of gene expression, signal transduction and have a role in cell proliferation or inhibition of programmed cell death. Activation of a proto-oncogene into an oncogene usually occurs via gain-of-function mutation. The patterns of mutations tend to be highly non-random, with most of the mutations enriched in certain regions of the protein. Oncogenic activation occurs via point mutation, gene amplification and chromosomal translocation. Oncogenes are altered in such a way in cancer that they become constitutively active thereby resulting in excessive stimulation of cell proliferation or inhibition of apoptosis and both contribute to the initiation and progression of cancer [13]. On cellular level these alterations act in a dominant manner which means that one allele is usually sufficient to confer a selective growth advantage to the cell. Since the identification

of the first human oncogene *HRAS*, with a glycine to valine substitutions at codon 12, several human oncogenes have been discovered like *MYC*, *EGFR*, *SRC* etc. [14].

- **Tumour Suppressor genes:** In normal cells, tumour-suppressor genes inhibit cell proliferation and stimulate cell death. In cancer, these genes are frequently altered leading to loss of function or reduction in protein activity. Tumour suppressor genes are inactivated in three ways, from missense mutations that alter the amino acid residues essential for its activity, frameshift or nonsense mutations that result in a truncated protein, from large deletions or insertions and from epigenetic silencing [13]. Tumour-suppressor genes are recessive in nature therefore mutations in both alleles are generally required to confer a selective growth advantage to the cell. This principle is known as the "two-hit" hypothesis and was first proposed by Alfred Knudson [15]. According to this hypothesis, hereditary cancers may arise by two inactivating alterations of which the first is inherited germline mutation and the other is acquired somatically. Conversely, sporadic cancers require two somatically acquired hits and thus such cancers usually develop at a later age [15]. The inherited inactivated allele tends to show small intragenic mutations, whereas the remaining allele is usually inactivated by similar mutations or by loss of heterozygosity (LOH), caused by for instance mitotic recombination [16]. Several classical tumour suppressor genes include *RB1*, *TP53* and *APC*. There are exceptions to the classical two-hit hypothesis like in some occasions; a single-copy event may be preferentially selected for in tumour evolution, instead of biallelic inactivation that might lead to cell death or senescence. The term haploinsufficiency refers to the scenario when inactivation of

a single allele is enough for aberrant protein function and promotion of cancer [17]. Another exception to the classical two-hit hypothesis is when a single-copy mutation functions in a dominant negative manner, interfering with the normal protein produced by the remaining wild-type allele [18].

Several tumour suppressor genes have been identified till date however it is difficult to clearly classify the role of these genes in tumour development. Broadly tumour suppressor genes can be classified into three types: 'gatekeeper', 'caretaker' and 'landscaper' tumour suppressor genes [19]. The concept of gatekeeper genes was first introduced to explain the role of a tumour suppressor gene, adenomatous polyposis coli (*APC*) gene in colorectal carcinogenesis. Gatekeepers are the subset of tumour suppressor genes that act as guards by preventing cancer cell growth and by inducing apoptosis. Alterations in the genome that inactivate these genes will result in abnormal growth and subsequent restoration of the gene function shows marked suppression of cancer cell growth [19, 20]. Caretaker genes are the stability genes that ensure effective DNA repair or prevent genomic stability thereby suppressing tumourigenesis [21]. Normally these genes function to keep the number of genetic alterations low but upon their inactivation the mutation burden is significantly increased. The resultant increased mutation rate affects all genes however only mutations targeting oncogenes or tumour suppressor genes will be preferentially selected for and have a tumour promoting effect. Similar to classical tumour suppressor genes, both alleles are generally inactivated in the tumour [20]. A large number of mismatch repair genes along with nucleotide-excision repair and base-excision repair genes are caretaker genes. Also genes involved in mitotic recombination and chromosomal segregation belong to this class, for example

BRCA1 and *ATM* [13, 22]. The 'landscaper' hypothesis was first postulated following the study of inherited mutations in Juvenile polyposis syndrome (JPS) which seems to be a result of an altered terrain of stromal cells and appears to be a landscaper defect. The landscaper gene is predicted to control the microenvironment in which the tumour growth takes place. It occurs by direct/indirect regulation of extra cellular matrix proteins, adhesion proteins, secreted growth factors and cell surface markers. Loss of function of the 'landscaper' gene will result in aberrant change in the microenvironment and subsequently promote abnormal growth of adjacent epithelial cells thereby increasing the risk of neoplastic transformation [19, 20].

1.1.6 Inherited Predisposition to Cancer

Majority of cancers arise sporadically and are highly influenced by environment, diet and lifestyle. Some individuals are born with a genetic defect that results in an increased lifetime risk of developing specific types of cancer in them, often at younger ages than people in the general population. These individuals are said to have a hereditary cancer syndrome, and it is believed that about 5-10% of all cancers are inherited via this mechanism [23]. There are more than 100 genes reported to cause Mendelian inherited cancer syndromes and at least 45 identified syndromes with clear genetic causes that confer an increased lifetime risk of developing cancer [24]. These syndromes can differ on many dimensions including the types of cancer, the magnitude of the risks, the inheritance pattern, and whether individuals have other physical symptoms. Many types of cancer are associated with hereditary syndromes including breast, colon, ovarian, pancreatic, and kidney cancer. Some hereditary

cancer syndromes have been well described and studied for many years, while others are poorly understood.

1.2 Colorectal Cancer

Colorectal cancer (CRC) is one of the most common malignancies worldwide. It is the third leading cause of cancer related death in males and second in females worldwide, with 1.8 million new cases and almost 861,000 deaths in 2018 according to the World Health Organization (WHO) GLOBOCAN 2018 updates [8]. The incidence of colorectal cancer is more in developed countries as compared to developing countries. The global distribution or the incidence rates of CRC vary 10-fold with highest rate in high-income countries like Australia, New Zealand, Northern/Western Europe and North America. The incidence rate is intermediate in Eastern and Southern Europe while it is lowest in low income countries like Africa and South Central Asia [25, 26]. These geographic differences may be attributed to the differences in dietary and environmental exposures coupled with genetic susceptibility [27]. Despite its low incidence rate, the mortality rate is quite high in India. Late stage presentation due to poor awareness, low socio-economic status and, and geo-ethnic differences may account for the same.

The common symptoms of CRC include rectal bleeding, significant changes in the colour of stool (especially dark or black-coloured stools), irregular bowel habits, diarrhoea or constipation, pain or cramp in the lower abdomen, decreased appetite, unintended weight loss, weakness or fatigue, and certain types of anaemia. CRC is a complex disease that is influenced by genetic, lifestyle, environmental and/or other factors. The lifetime risk of CRC is around 5% in the general population [28].

The genetic factors include:

- **Family history:** A personal or family history of colorectal cancer or GI polyps may increase risk of CRC. There is not always a clear reason for increased risk in a family however inherited genes, shared environmental factors or a combination of these may be responsible for cancers in such family.
- **Inherited syndromes:** Around 5% of all CRC patients have an inherited gene mutation that is responsible for their cancer. The two most common inherited CRC syndromes are familial adenomatous polyposis (FAP) and Lynch Syndrome also known as hereditary non-polyposis colorectal cancer (HNPCC). Other syndromes that are linked with colorectal cancer include Turcot Syndrome, Peutz-Jeghers Syndrome (PJS) and Juvenile Polyposis Syndrome (JPS).
- **Racial and ethnic background:** Clear difference in the incidence of colorectal cancer, stage at diagnosis, and mortality has been observed worldwide. Age-adjusted incidence and mortality are highest for African Americans and lowest for Hispanics and Asians/Pacific Islanders.

The lifestyle risk factors include [29]:

- **Diet:** Increased intake of red and processed meats (e.g., beef, lamb, hot dogs) is associated with high colorectal cancer risks. Meat when cooked by frying, grilling, boiling at very high temperatures release carcinogenic chemicals such as polycyclic aromatic hydrocarbons, heterocyclic amines and dietary N-nitroso compounds may also contribute to CRC. A high-fibre diet, increased intake of fruits and vegetables may help reduce risk of developing colorectal cancer
- **Inactive lifestyle:** Individuals who do not indulge in physical activity and live a sedentary life have an increased chance of developing colorectal cancer.

- **Smoking:** Smoking has been associated with increased risk of colorectal cancer with a prolonged latency period.
- **Alcohol use:** Heavy alcohol consumption has been associated with an increased risk of colorectal cancer.

Other inherent factors associated with an increased risk of CRC are [30]:

- **Age:** Age is one of the major risk factor in CRC and the chances of developing the disease increases with advancing age. Nearly 95% of all CRCs occur in patients 50 years or older. The median age of CRC onset is 68, according to the National Cancer Institute.
- **History of inflammatory bowel disease (IBD):** Patient with personal history of IBD, including ulcerative colitis and Crohn's disease, are at an increased risk of developing colorectal cancer than the general population [31, 32].
- **Obesity:** Abdominal obesity may increase the risk of developing colorectal cancer in both sexes, with stronger association seen between colon cancer and obesity.
- **Type II diabetes:** Individuals with type II diabetes is associated with a higher risk of colorectal cancer and the shared risk factors between these two conditions include obesity, sedentary lifestyle and high caloric diet. It may also affect prognosis.

Staging for CRC: Accurate cancer staging is important not only for appropriate evaluation of therapies, prediction of survival and prognosis, but also for cancer research in general. There are two widely used staging systems when diagnosing CRC; the Union for International Cancer Control (UICC) TNM (tumour, node, metastasize) staging system and the Dukes Classification [33] (Table 1.1). The TNM staging is based on the following three information:

- The extent (size) of the **tumour (T):** How far has the cancer spread into the wall of the colon or rectum? These layers, from the inner to the outer, include:
  - The inner lining (mucosa), which is the layer in which nearly all colorectal cancers start. This includes a thin muscle layer (muscularis mucosa)
  - The fibrous tissue beneath this muscle layer (submucosa)
  - A thick muscle layer (muscularis propria)
  - The thin, outermost layers of connective tissue (subserosa and serosa) that cover most of the colon but not the rectum
- The spread of cancer to nearby lymph **n**odes **(N)**
- The spread (**m**etastasis) to distant lymph nodes or distant organs **(M)**

Table 1.1 TNM staging and Dukes Classification for Colorectal cancer

Stage	**Definition**	
T0	No evidence of primary tumour	
Tis	Carcinoma in situ: intraepithelial or intermucosal	
T1	Tumour invasion into submucosa	
T2	Tumour invasion into muscularis propria	
T3	Tumour invasion through muscularis propria	
T4	Tumour invasion into other organs or through visceral peritoneum	
N0	No evidence of regional lymph node metastasis	
N1	Metastasis into 1-3 regional lymph nodes	
N2	Metastasis into ≥4 regional lymph nodes	
M0	No evidence of distant metastasis	
M1	Distant metastasis	
Stage Grouping	**Dukes stage**	**5-year survival (%)**
Stage I: T1-2, N0, M0	Dukes A	80-95
Stage II: T3-4, N0, M0	Dukes B	65-75
Stage III: Any T, N1-2, M0	Dukes C	25-60
Stage IV: Any T, Any N, M1	Dukes D	0-7

1.3 Colorectal tumourigenesis

CRC develops from rapidly dividing epithelial cells lining the colon or rectum of the gastrointestinal tract. It is now widely accepted that CRC results from the accumulation of genetic and epigenetic alterations, which results in the transformation of normal colonic epithelium to colorectal adenocarcinoma. The development of colorectal adenocarcinoma is characterized by a series of genetic and epigenetic events in tumour suppressor genes and oncogenes which when acquired by a normal epithelium results in stepwise transformation to hyperproliferative mucosa which later gives rise to a benign adenoma that eventually becomes a carcinoma and metastasizes over an average of 10 years [34-36]. A normal epithelial cell progresses in a clonal fashion to carcinoma by acquiring at least five to seven major deleterious molecular alterations. CRC cells can acquire increased mutability of their genomes through several different molecular pathways that result in loss of genomic stability. Genomic instability in CRC can occur via chromosomal instability (CIN), microsatellite instability (MSI), CpG Island Methylation (CIMP) [37, 38]. More recently, tumours have been subcategorized based on their mutation rate. The TCGA study recently described CRCs to be either nonhypermutated or hypermutated based on the number of mutations on nucleotide level. Hypermutated tumours have mutation rates of 10-100 per 10^6 bases, whereas nonhypermutated tumours show mutation rates of less than 10 per 10^6 bases. Non-hypermutated cancers, which represent the large majority of CRCs (84%), are usually microsatellite stable (MSS) and show CIN [39].

- **Chromosomal Instability Pathway (CIN):** CIN, defined as the presence of structural aberrations or abnormal chromosome numbers, arises in about 70-80% of CRCs. It reflects the classical adenoma-carcinoma model proposed by Fearon and

Vogelstein [40] which suggests that CRC develops through step-wise carcinogenesis. Step-wise carcinogenesis occurs by progressive accumulation of point mutations in key genes such as *APC*, *KRAS* and *TP53*, frequent chromosomal losses and gains, especially losses on chromosome arms 5q, 17p and 18q [36, 40](Fig 1.2). The three key principles of this model are: 1) multiple genetic hits are required, 2) discreet intermediates in the progression to cancer, and 3) aberrant crypt foci (ACFs) develop before colorectal polyps and are the earliest detectable change of the adenoma-carcinoma sequence [41]. It is estimated that the entire process from ACFs to invasive carcinomas takes between 20-40 years during which there is a constant increase in CIN [42]. CIN results in an imbalance in chromosomal number (aneuploidy), subchromosomal genomic amplifications, and a high frequency of loss of heterozygosity (LOH) [41]. CIN tumours are non-hypermutated tumours. Majority of sporadic CRCs and Familial Adenomatous Polyposis (FAP), a hereditary form of CRC follows this pathway.

- **Microsatellite Instability pathway (MSI):** A subset of CRC cancers have hypermutated genomes and exhibit a so called “mutator phenotype”, which occurs due to defects in DNA mismatch repair (MMR) genes that function in the maintenance of genomic stability. Approximately 15% of CRCs develop through the microsatellite instability (MSI) pathway, which is driven by defects in the mismatch repair system. These defects when inherited causes a type of hereditary CRC, Lynch syndrome and when acquired result in sporadic MSI tumours. The mechanism of tumourigenesis in MSI tumours involves either germline mutations in DNA mismatch repair (MMR) or through somatic mutation or aberrant promoter methylation as in sporadic CRC. (Fig 1.2)

- **CpG Island Methylator phenotype (CIMP):** CIMP arises due to hypermethylation in the promoter region of genes that have tumour suppressive roles or are involved in the cell cycle resulting in their transcriptional inactivation [43]. Hypermethylated promoters are associated with *BRAF* V600E mutation. Hypermethylation of MLH1, one of the MMR genes is the major cause of MSI in sporadic CRC [44] (Fig 1.2). Many other genes have also been identified to be affected in CIMP that may have important functions in the cell. Several studies have classified CIMP tumours into three group based on the degree of methylation- CIMP-High, CIMP-Low and CIMP-No. Majority of studies have commonly contained the classic panel: *hMLH1*, *p16*, *MINT1*, *MINT2*, and *MINT31*, however many more methylation markers are emerging with increasing research [45].

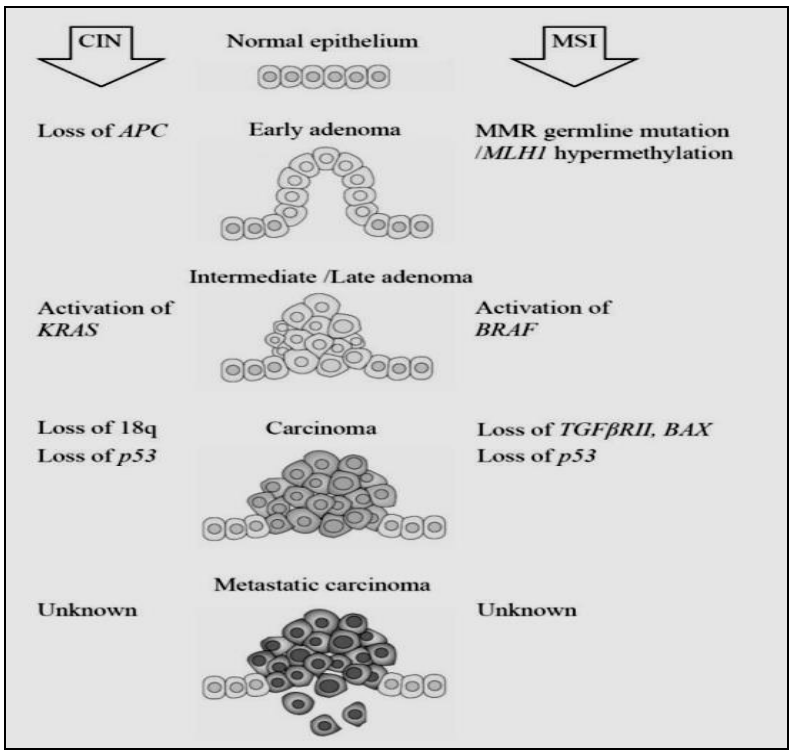


Figure 1.2 Stepwise progression of CRC (Taken from Ref. [46])

Recent studies have identified a small novel class of hypermutated CRCs that result from exonuclease domain mutation (EDM) in *POLE* and *POLD1* [39]. Germline mutation in these genes result in a rare condition termed polymerase proofreading-associated polyposis (PPAP). Somatic mutations in *POLE* have been reported in CRCs as well as endometrial cancer, however no somatic pathogenic *POLD1* mutation have been reported. Both germline and somatic mutations in these genes result in an "ultramutated" phenotype, with mutation rates of over 50 per 10^6 bases.

Alteration in several molecular signalling pathways such as the Wnt, TGF-β, PI3K, RAS/MAPK, and NF-κB pathways or a defective DNA MMR pathway result in CRC (Fig 1.3) [47, 48]. These alterations may result in resistance to antitumor agents, and they can confer individual susceptibility to cancers upon inheritance.

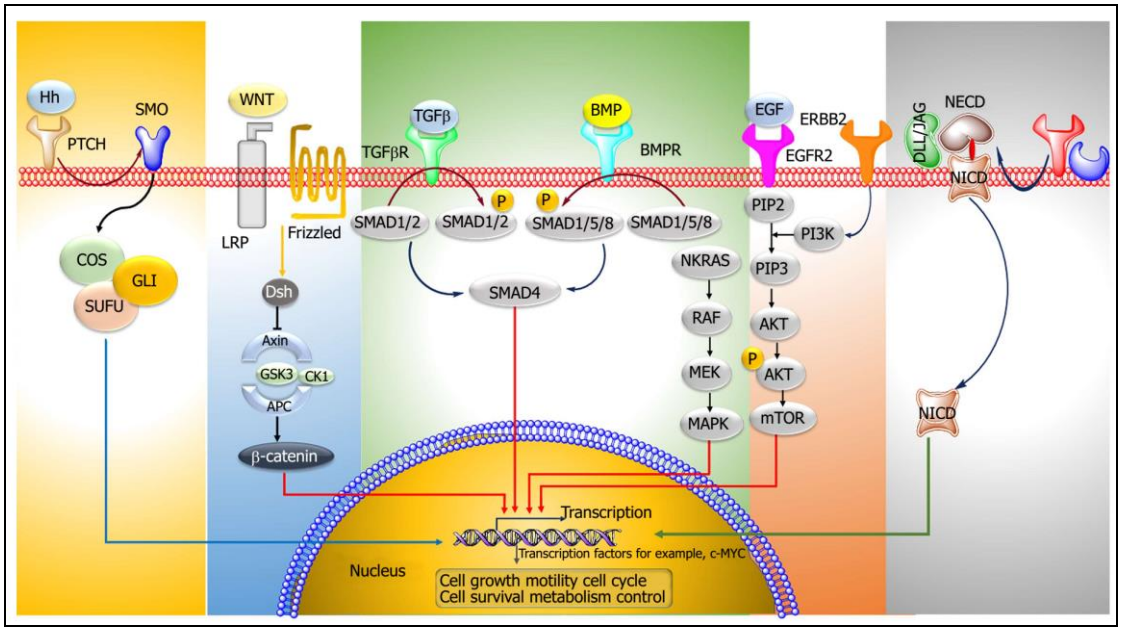


Figure 1.3 Different signalling pathways involved in colorectal cancer (Taken from Ref. [49])

1.4 Classification of colorectal cancer

Based on different risk factors, CRC can be presented in three forms: Sporadic, Familial and Hereditary. Sporadic CRC, in which there is no family history, occurs due to gene mutations in colorectal cells but not as germline mutations. It usually occurs in the later stages of life and dietary and environmental factors have been associated with increased risk [50]. Less than 5% of patients have an inherited predisposition to CRC due to germline mutations in highly penetrant cancer genes with specific pattern of inheritance.

1.4.1 Hereditary Colorectal cancer: It occurs in a well-defined hereditary setting with well characterized clinical features and germline mutations in highly penetrant genes conferring high lifetime risk of CRC. Hereditary CRC syndromes have classically been divided into two groups based on the presence or absence of gastrointestinal polyps: Polyposis and Nonpolyposis syndrome. Polyposis syndromes, usually identified clinically, are defined by the presence of multiple polyps in the colon and are further classified into adenomatous (e.g. Familial Adenomatous Polyposis) and hamartomatous polyposis syndromes (e.g. Peutz-Jeghers Syndrome). The most common type of nonpolyposis syndrome is Lynch syndrome (also known as Hereditary Nonpolyposis Colorectal Cancer, HNPCC) inherited in an autosomal dominant fashion and a rare mutation of Lynch syndrome is Constitutional Mismatch Repair Deficiency syndrome (CMMRD) which is an autosomal recessive childhood cancer predisposition syndrome. The broad classification of hereditary colorectal cancers is represented in Figure 1.4. The clinical and genetic features of major hereditary CRC are described in Table 1.2

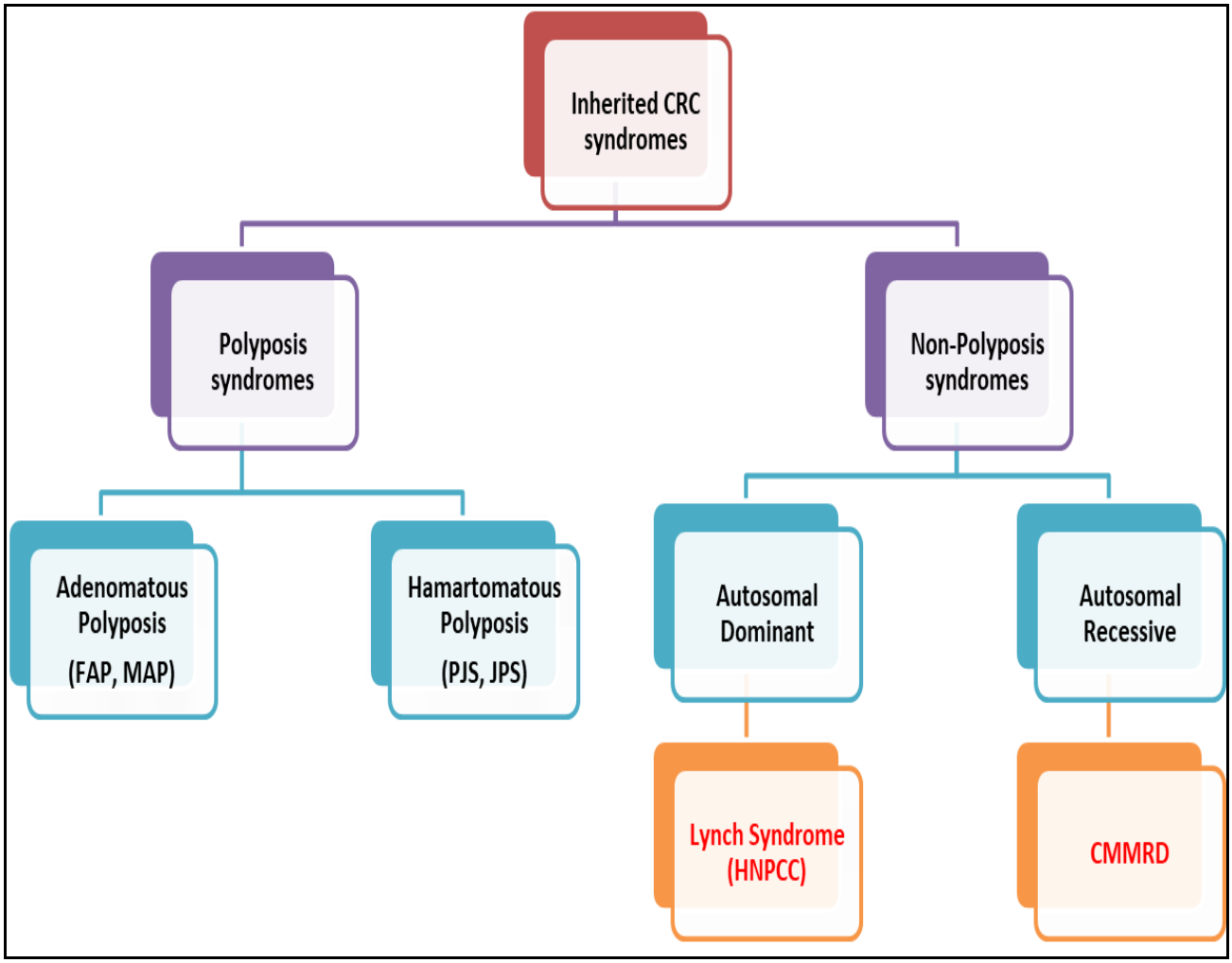


Figure 1.4 Classification of Hereditary Colorectal cancers

Abbreviations: **FAP**- Familial Adenomatous Polyposis; **MAP**- MUTYH associated Polyposis; **PJS**- Peutz-Jeghers Syndrome; **JPS**- Juvenile Polyposis Syndrome; **HNPCC**- Hereditary Nonpolyposis Colorectal Cancer; **CMMRD**- Constitutional Mismatch Repair Deficiency

Table 1.2 Clinical and genetic features of hereditary colorectal cancers

Polyposis syndrome	Clinical features	Genes involved
Adenomatous Polyposis Syndrome		
Familial Adenomatous Polyposis (FAP)	• Autosomal dominant inheritance • 0.5-1% of all CRC • 100-1000 colorectal polyps • Early age of onset • General disorders of FAP includes: Gardner's syndrome (soft tissue tumours like fibromas, desmoid tumours) and Turcot syndrome (Brain tumours) • Extracolonic manifestations include congenital hypertrophy of the retinal pigment epithelium (CHRPE), supernumerary teeth, osteomas, cutaneous lipomas and cysts, thyroid tumours, desmoid tumours, adrenal cortical adenomas, and hepatoblastomas	APC
Attenuated FAP	• Lower polyp number (1-50) • Later age at diagnosis • Less or no extracolonic manifestations	APC
MUTYH associated polyposis (MAP)	• Autosomal recessive inheritance • Fewer than 100 polyps	MYH
Hamartomatous Polyposis		
Peutz-Jeghers syndrome (PJS)	• Autosomal dominant condition • Characterized by mucocutaneous pigmentation and hamartomatous polyps • Usually presented with small bowel intussusception and gastrointestinal haemorrhage • High risk of GI, breast, ovary, uterus, pancreas, testis, oesophagus cancers	STK11

Juvenile Polyposis syndrome (JPS)	• Autosomal dominant inheritance • Multiple hamartomatous (juvenile) polyps in the digestive tract • Increased risk of cancer of stomach, colorectum, small intestine and breast	PTEN, SMAD4 (DPC4) or BMPR1A gene
Cowden syndrome	• Hamartomatous lesions of skin and mucous membrane of mouth and nose • Increased risk of early onset breast cancer, thyroid cancer, ovarian cancer and endometrial cancer	PTEN
Non-Polyposis syndrome	**Clinical features**	**Genes involved**
Lynch syndrome (Hereditary Nonpolyposis Colorectal Cancer, HNPCC)	• Autosomal dominant inheritance • Very few or no polyps • Adult age of onset • Right sided and microsatellite unstable tumour • High lifetime risk of colorectal cancer and significant risk of endometrial, ovarian, urological, central nervous system cancers	MLH1, MSH2, MSH6, *PMS2* (90% mutation in MLH1/MSH2)
Constitutional Mismatch Repair Deficiency (CMMRD) syndrome	• Autosomal recessive inheritance • Early age of onset (0-30 yrs) • Characterized by skin manifestations like café-au-lait spots, freckles, hypopigmentation • Broad tumour spectrum with brain tumours and digestive tract tumours being most common followed by haematological malignancies	MLH1, MSH2, MSH6, *PMS2* (60% mutation in *PMS2* and 40% in rest of the MMR genes)

There is very limited information on molecular characterization of CRC from the developing nations with most of the data from Caucasian population. Among the polyposis syndrome, PJS is a rare condition with no clear genotype-phenotype

correlation and has not been examined in detail in the Asian population, particularly the South Asians. Lynch syndrome, hereditary nonpolyposis colorectal cancer is the most common form of hereditary CRC and has been studied extensively in the Western population with lack of knowledge in the South Asian population. Only 28 MMR mutation carrier Lynch syndrome families reported from India. There is lack of knowledge about the rare hereditary forms of CRC like FAP, PJS and CMMRD also. The only South-Asian report on FAP is from our group [51]. Single report of *STK11* mutation carrier PJS family and biallelic PMS2 CMMRD family is published from India. Therefore, we have carried out a comprehensive genotype and phenotype characterization of these syndromes in this study. PJS, Lynch syndrome and CMMRD syndrome are described in detail in Chapter 3, 4 and 5 along with the results from this study.

1.4.2 Sporadic colorectal cancer: It comprises the largest group of CRC accounting for 75%. It occurs at the median age of 70-75 years with lack of relevant family history and inherited gene mutation that may accelerate cancer development. It is caused by a series of genetic events involving loss of tumour suppressor genes and activation of oncogenes which provides an evolutionary growth advantage to cells resulting in tumour development. Sporadic CRC is influenced by diet, lifestyle and environmental factors [52]. Though the age adjusted incidence rate of CRC is low in India, there has been a disproportionate increase in the incidence of early onset CRC (EOS-CRC), especially rectal cancers without any evident family history [53]. EOS-CRC is an aggressive disease with poor differentiation, mucinous signet ring cell histology and is classically located in the left side of the colon. This subset of early onset sporadic rectal cancer exhibit distinct clinical and genetic features as compared

to late onset CRC and majority of them do not show involvement of the three canonical pathways of CRC tumourigenesis- CIN, MSI and CIMP. Our understanding of EOSCRC is meagre with no clear understanding of its clinical and genetic features. It is therefore imperative to carry out a comprehensive characterization of EOSCRC to identify the genomic mutational landscape involved in its tumourigenesis.

In the light of the discussed literature, the present study was undertaken with the aim to study the key genetic alterations in various forms of Hereditary Colorectal Cancers in an Indian cohort, to carry out genotype-phenotype correlation and to elucidate the somatic landscape of early onset recto-sigmoid tumours from patients below 40 years of age with no family history of CRC associated cancers.

With this aim, the objectives were framed as follows:

1) To define the mutation spectrum of key genes- MMR and *STK11* in different types of Hereditary Colorectal Cancers in the Indian cohort and study genotype-phenotype correlation

   1.1) To identify novel/recurrent germline mutations in Peutz-Jeghers syndrome-STK11/LKB1 and study the genotype-phenotype correlation

   1.2) To identify novel/recurrent germline mutations in Lynch Syndrome/Hereditary Non-Polyposis Colorectal Cancer (HNPCC) – MMR study known and novel genotype-phenotype correlation

   1.3) To identify novel/recurrent germline mutations in CMMRD syndrome- *PMS2* and other MMR genes

2) To identify the somatic mutational spectrum of early onset Sporadic Rectal cancer in tumour and matched paired normal tissue samples using NGS technology (whole exome sequencing)

CHAPTER 2

MATERIALS AND METHODS

2.1 Patients and Samples

Patients of all the hereditary colorectal cancer syndromes included in this study were enrolled at the Cancer Genetics Clinic of Tata Memorial Hospital, Parel after approval from Institutional Review Board (IRB) of Tata Memorial Centre from 2008-2019. Patients were diagnosed on the basis of family history and clinico-pathological characteristics and were referred by their treating oncologists to the Cancer Genetics Clinic of Tata Memorial Hospital. After genetic counselling and written informed consent, blood sample was collected for genetic testing. All experiments were carried out in accordance with the approved guidelines and regulations. This study included three forms of hereditary colorectal cancer syndromes:

1) PJS- 20 families (registered from 2009-2019)
2) Lynch syndrome- 91 families (registered from 2014-2019)
3) CMMRD- 5 families (registered from 2014-2019)

Details of patients with sporadic CRC are discussed in section 2.3

2.1.1 Overview of the patient enrolment process

Pre-test counselling: Genetic counselling is a communication process that deals with the occurrence or risk of occurrence of a possible genetic disorder in the family. Patients referred to the Cancer Genetics Clinic (CGC) undergo a pre-test counselling session which was conducted between the patients, their accompanying family members and the genetic counsellor at CGC. During this session, the patients and their family members were educated on basic cancer genetics and what cancer risk means. They were explained about the risk, benefits, limitations and possible outcome, medical, reproductive, and psychosocial implications of genetic testing including potential role of testing for other family members. Clinical features, personal details,

medical history of patients and other family members were obtained from the patients based on which a pedigree chart was drawn. Based on these features, cases with a suspected syndromic diagnosis of PJS, Lynch syndrome or CMMRD were enrolled and 3 ml of blood was collected in EDTA tubes after taking written informed consents from patients or their parents if they were minors.

Post-test counselling: In this session, the results of the genetic test were disclosed to the patient by providing a printed genetic test report and explained about the significance of the test. Patients found to have a mutation associated with an increased risk of malignancy were counselled regarding their age-related and lifetime risks of particular types of cancer associated with the syndrome and the surveillance and management strategies available for risk reduction and health maintenance. Screening of the first and second degree relatives of mutation carrier patients were also offered during this session.

2.2 Methodology to study mutation spectrum of STK11 and MMR genes in different hereditary cancer syndromes

- DNA extraction from blood samples using Qiagen DNA extraction kit and quantification by Nanodrop
- Exon wise PCR amplification of individual genes followed by agarose gel electrophoresis of amplified PCR products
- Enzymatic clean-up of PCR products and Sanger sequencing to detect point mutations and small indels
- MLPA analysis to detect large genomic rearrangements (LGRs) and data analysis by Coffalyser.Net software from MRC-Holland

2.2.1 DNA extraction

Genomic DNA was extracted from 200µl of blood sample using QIAamp Blood DNA Mini kit as per manufacturer's protocol using the following steps**:**

1. 20µl QIAGEN protease or Proteinase K was added into to a 1.5ml microcentrifuge tube to which 200µl whole blood was added.

2. 200µl Buffer AL was added to the sample. It was mixed by pulse-vortexing for 15 seconds and then incubated at 56°C for 10 minutes.

3. The sample was briefly centrifuged and 200µl of ethanol (96-100%) was added to the sample and mixed by pulse-vortexing for 15 seconds. After mixing, the 1.5ml microcentrifuge tube was briefly centrifuged to remove drops from inside the lid.

4. The mixture from step 3 was carefully applied to the QIAamp Mini spin column. The cap was closed and centrifuged at 6000 x g (8000 rpm) for 1 minute.

5. The QIAamp Mini spin column was placed in a 2 ml collection tube and the tube containing the filtrate was discarded.

6. 500µl Buffer AW1 was added and centrifuged at 6000 x g (8000 rpm) for 1 minute. The QIAamp Mini spin column was placed in a 2ml collection tube and the collection tube containing the filtrate was discarded.

7. 500µl Buffer AW2 was added and centrifuged at full speed (20,000 x g; 14000 rpm) for 3 minutes.

8. Recommended: Place the QIAamp Mini spin column in a new 2ml collection tube and discard the old collection tube with the filtrate. Centrifuge at full speed for 1 min.

9. The QIAamp Mini spin column was placed in a 1.5ml microcentrifuge tube and the collection tube containing the filtrate was discarded. 200µl Buffer AE or distilled water was added to the column.

10. The elution buffer (AE) was incubated at room temperature (15-20°C) for 1 min, and then centrifuged at 6000 x g (8000 rpm) for 1 minute to collect the DNA.

11. DNA was quantified using Nanodrop spectrophotometer or by loading the DNA samples on 0.5% agarose gels.

12. After estimation of the purity and concentration of the DNA samples, the DNA samples were diluted to a working concentration of 20ng/µl.

2.2.2 Polymerase Chain Reaction (PCR)

Exon wise PCR amplification of the entire coding region and flanking intronic region was carried out using specific primers designed for respective genes. PCR amplification was carried out in a 25µl reaction with 10pmol of primers (Sigma), 100ng of DNA and using 0.5 units of Taq polymerase (Axygen). The components of the mastermix are given in Table 2.1 and the cycling conditions are summarized in Table 2.2. The primer sequences and their respective annealing temperatures for all the genes studied in this thesis are given in Table 2.3-2.7

Table 2.1 PCR mastermix composition

Components	Volume (µl)/reaction
10X PCR buffer	**2.5**
2.5 mM dNTP	**1**
5 U/µl Taq Polymerase	**0.1**
20 ng/ml Template DNA	**5**
10pmole/µl Forward primer (FP)	**1**
10pmole/µl Reverse primer (RP)	**1**
MilliQ	**14.4**
Total reaction volume	**25**

Table 2.2 PCR cycling condition

Steps	Temperature	Time
1. Initial denaturation	95˚C	5 minutes
2. Denaturation	95˚C	45 seconds
3. Primer annealing, Ta*	X	45 seconds
4. Extension	72˚C	45 seconds
5. Go to steps 2-4: 34 cycles	-	-
6. Final Extension	72˚C	5 minutes

***Ta: Annealing temperatures variable for different primer pairs**

Table 2.3 Primer sequences and annealing temperature for STK11

Exon	Forward Primer	Reverse Primer	Ta
1	CACAAGGAAGGACCGCTCAC	CCGCTGCGACAACTGGCCTT	62˚C
2	GGCCCTTTCCCACAGCACT	AGGCCCCGCGGTCCCAACA	58˚C
3	GAGGAGGGGCAAGGTGGGT	GTGTGGCCTCACGGAAAGGAG	61.5˚C
4+5	GCTGGGCCTGTGGTGTTTGG	GACGGGCCAGGCTGCACTTC	60˚C
6	TCAACCACCTTGACTGACCA	ACACCCCCAACCCTACATTT	60˚C
7	CAGCTGACAGGCTCCTCGC	CTCAACCAGCTGCCCACAT	61˚C
8	CCCTTGCACGGCCTGGTCC	TGGGACATCCTGGCCGAGT	60˚C
9	TGGATACACCTGGGCCTGAC	GGGCTATGCTCACGGCTGGC	66.3˚C

Table 2.4 Primer sequences and annealing temperature for MLH1

Exon	Forward Primer	Reverse Primer	Ta
1	CTGTCCAATCAATAGCTGCC	TGCGGAAAAGGAGAAGGCCTG	62˚C
2	GTATGAGCCTGTAAGACAAAGG	GCCCAGCAAATAATAGGTAC	60˚C
3	CAAGAAATGGAATTCAAAGAGATTTG	TTTGAAAGTTAAGTTCATTAAGTTTG	60˚C
4	AAAGTGCTCATCGTTGCC	CACCTAATAATCATCCTTGAG	60˚C
5	GATATGATTTTCTCTTTTCCC	GCCAATAGTCATTTATCTTGC	55.5˚C
6	AATGCTGTCTTATCCCTGGCC	ACCTTGACCAGAAACTATCTG	60˚C

7+8	TAAAAGTAGAGAGGAGTCTGTG	CCTAGAAAGTGTTGATTACGTG	57.2°C
9	CTGAGTAGGGTAGGTGGGTG	CAACCAGCAATGAGCACATGTG	63-56°C TD
10	CGATAGTAAGATAGTGGGCTGG	AGGCTCTTAGTGAGGTTCTGC	62°C
11	CAAATGAAGAGACTGAGGC	CTAAGCCTAGGAACAACAGC	58.5°C
12	CGGGCAGAATTGCTTCTAT	GGTCAAAGGCAGACAGTGG	63-56°C TD
13	GGGTTGTCAGATAAGCAGTC	GCTGATGCTATTGTGGGTTA	62°C
14	GTTCGTTTTCACCAGGAGG	CTGACTCCAAAGCCTGTGCC	63-56°C TD
15	CCCTGGTTGAAGACGTTG	GATACCTCCATATGCAAATC	54.3°C
16	TGACAAGAGGAGGAAAGGG	TTTCATCATGTTGGCCAGC	63-56°C TD
17	GTTCTGCCGTGCTGTTTGTC	TGGGACTGCTGAATATTGCTGG	62°C
18	CCAGCAATATTCAGCAGTCC	CAACCTCCCATTTCTCACTGTG	62°C
19	CTTGTGTTCAGGCCTGTGGGATC	GGTCAGTGCCATCAGAGCC	62°C

TD: Touch Down temperature

Table 2.5 Primer sequences and annealing temperature for MSH2

Exon	Forward Primer	Reverse Primer	Ta
1	GCTGAT TGGGTGTGGTCGC	CGCACAAGCACCAACGTTC	60.3°C
2	GAAGTCCAGCTAATACAGTGC	GTGTCTCAAACCATTCTAC	56.2°C
3	GGTTCATAGAGTTTGGATTTTTCC	GGGGAGAAAAGATCTGAGGT	62.4°C
4	CAGTACATCATATCAGTGTC	CATTGATACACAGTTTAGG	53.3°C
5	GGATTGGGAAGGAACACC	GGGAGAGAAAAATACAGCCA	55.5°C
6	GGAAGAGGAACTTTTTGTGG	CATGTTCGATACCATACCATC	62.4°C
7	GCCCAGCAGATTCAAGCT	TGAGTCACCACCACCAAC	60.3°C
8	ACTTTGGAGACCTGCTGTAC	CCACTGTCCACAAAGGTGCT	62.4°C
9	CTCTAAAGTCCTAATGAACAG	TCATCTTGGGGACAGGGAAC	57.9°C
10	CATTCATAAGGGAGTTAAGG	GTTGCGACAGCTGACTGCTCTATG	57.9°C
11	TGTCCCTAAGGAGTTGTTCG	TCAGAATGTAATGGCTTGCG	55.5°C

12	CTATGTTGAGTTTTAGGTGG	CTTCCCTCTAAACCAAATGTG	57.9°C
13	GCTGTGGTTCTGCCTTTATATGC	AGTCCACAGGAAAACAACT	62.4°C
14	GTTTGTGGCATATCCTTCCC	GTGGTCCTACTATGAGATACAG	57°C
15	CTAATGACAAGGTGAGAAGG	GCACTAGAAACACAGAGG	57.9°C
16	GGGTGGGCTAATGTGGGAGGAG	GATAGCCCATGGGCACTGA	67°C

Table 2.6 Primer sequences and annealing temperature for MSH6

Exon	Forward Primer	Reverse Primer	Ta
1	TGTTGATTGGCCACTGGG	CAACCCCCTGTGCGAGCCTC	58-51°C TD
2	TGCCAGAAGACTTGGAATTC	TTACTGGGGTAAAATACACTTAATTTC	58-51°C TD
3	CTGGTCTTGAACTGCTGGGAT	CCCCTTTCTTCCCCCATC	58-51°C TD
4.1	TGCACGGGTACCATTATAAAGTCA	GTATTCTTGGTTTCTGATGAAATGCTAG	58-51°C TD
4.2	GAAGGAAACGCCCTCAGC	CAGTTGCCTTTCATGAATACCAG	58-51°C TD
4.3	CCACATGGATGCTCTTATTGGA	TCATCTGAAAACTGACCTATGAAAAACT	58-51°C TD
4.4	TTTGTTGATACTTCAGTGGGAAAGTT	CTCCTGATCAATAAGGCATTTTTTG	55-48°C TD
4.5	CTCTAGGTGGTTGTGTCTTCTACCTC	TGAGTAGCCTCTCAACATCTGGAA	58-51°C TD
4.6	CGAAGTTGTAGAGCTTCTAAAGAAGCT	GTCCTACAGCCAATTCTGTTGC	58-51°C TD
4.7	AGCCTCCTGGAATACCTAGAGAAAC	ACTTATTTTTAGGGATAATATACAGCTGGC	58-51°C TD
5	CACTTAGGCTGATAAAACCCCC	GTATGTTATTCCTAATGTCACAAATGACTTT	58-51°C TD
6	AAGACAAAAGTTTATGAAACTGTTACTACCA	AGAAGCAAATATCTTTTATCACATCTAAATG	58-51°C TD
7	TAACCTAGAAGATGAATTTATGTAATATGATTT	TTCAGATAATCTTCTATAAAAATAGTTATTTGT	55-48°C TD
8	TGAGTTACTTCCTTATGCATATTTTACT	AATATTAGCGATACATGTGCTAGCA	55-48°C TD

9	TGCTAGCACATGTATCGCTAATATT	GCATCATCCCTTCCCCTTTTA	58-51°C TD
10	GAAGGGATGATGCACTATGAAAAA	GTAGAAGGTAGATAAGAATTAAAAGGGTTTAA	58-51°C TD

TD: Touch Down temperature

Table 2.7 Primer sequences and annealing temperature for MLH3

Exon	Forward Primer	Reverse Primer	Annealing temperature
2A	TGAGCTGTGCCTAGAGATC	GAAATGGAAGGGTGCATGAG	60.4°C
2B	GCCCTGAAAGCTTGTGAA	GACCATTCTTTGGCTTGCAT	60.4°C
2C	GGTTCCATGGTTCTTCAGCTCC	GGTAGCTTCTGAATCCCTAG	60.4°C
2D	GTGACTTCCGATGAGAGGAGC	CTGTCTGAGCACTATGTACTCC	60.4°C
2E	ATGCTACTGAAGTGGGATGC	AAAGATCCTAGCTGTGAACTC	60.4°C
2F	GGTATAGACACGTTTCCAATG	CGGAACCCTTCAGTCTGG	60.4°C
2G	CCTTTGGACCTTGAGAAGTC	ACCTGTGGCATCTTCTACCGG	60.4°C
2H	TCCCAACATCAGATTCTGCC	TCAAAAGCATCTCATGCACA	60.4°C

2.2.3 Long Range Polymerase Chain Reaction (PCR)

Mutations within *PMS2* are considerably more difficult to identify because of the presence of a large family of pseudogenes, which are located on the same chromosome as the true *PMS2* gene. We have used previously described and established long range PCR method which selectively amplifies the *PMS2* genomic region, while avoiding the amplification of sequences from the pseudogene loci. Long range primers are designed in such a way that either the forward or the reverse primers are positioned at regions of divergence between the real gene and the pseudogenes. Amplicons spanning exons 1–5, 7-9 and 11-15 are amplified by long range primer 1 (LR1), LR2 and LR3 respectively. Using each of the long-range primer sets, 100ng of

DNA was amplified in 25µl reaction using 0.2mM each primer, 1.25U Takara LA Taq (Cat#), 10X LA PCR Buffer II (Mg2+free), 25mM $MgCl_2$ (final 2.5mM) and 400mM each dNTP. The components of PCR mastermix are given in Table 2.8 and cycling conditions are summarized in Table 2.9. The appropriate long-range PCR product was diluted 1:10, and 2mL of this dilution was used as the template for exon-specific amplification using the previously described PCR method in section 2.2.2. The primer sequences and their respective annealing temperatures for *PMS2* gene studied are given in Table 2.10.

Table 2.8 Long Range PCR mastermix composition

Components	Volume (µl)/reaction
10X LA PCR buffer II	2.5
25mM MgCl2	2.5
2.5 mM dNTP	4
1.25 U/µl Taq Polymerase	0.3
20 ng/ml Template DNA	5
10pmole/µl Forward primer (FP)	1
10pmole/µl Reverse primer (RP)	1
MilliQ	8.7
Total reaction volume	25

Table 2.9 Long Range PCR cycling condition

Steps	Temperature	Time
1. Initial denaturation	94˚C	1 minute
2. Denaturation	94˚C	15 seconds
3. Primer annealing	65˚C (0.5↓/cycle)	30 seconds
4. Extension	68˚C	15 minutes

5. Go to steps 2-4: 13 cycles	-	-
6. Denaturation	94°C	15 seconds
7. Primer annealing	58°C	30 seconds
8. Extension	68°C	15 minutes
9. Go to steps 6-8: 25 cycles	-	-
10. Final Extension	72°C	10 minutes

Table 2.10 Primer sequences and annealing temperature for *PMS2*

Exon	Forward Primer	Reverse Primer	Annealing temperature
LR1	ACGTCGAAAGCAGCCAATGGGAGTT	CTTCCACCTGTGCATACCACAGGCT	65-58°C TD
LR2	GGTCCAGGTCTTACATGCATACTGT	CTGACTGACATTTAGCTTGTTGACA	65-58°C TD
LR3	GCGTTGATATCAATGTTACTCCAGA	CCTTCCATCTCCAAAACCAGCAAGA	65-58°C TD
1	ACGTCGAAAGCAGCCAATGGGAGTT	CAGGTAGAAAGGAAATGCATTCAGT	65-58°C TD
2	GTAAGGATCTGTTGAATTTGAAG	TTCTTAGCATAACACCTGCCTGGCA	58-51°C TD
3+4	CTGGGCTAGTAAATAGCCAGAAAG	TATGACTTAGATTGGCAGCGAGACA	58-51°C TD
5	CTTGATTATCTCAGAGGGATCGTCA	TCTCACTGTGTTGCCCAGTCCTAAT	58-51°C TD
6	TGCTTCCCTTGATTTGTGCGATGAT	CTACTGGAAGGGACAATGGAAACC	60°C
7	ATTGTACTCCAGCCTGGGCAATAG	ATTGTAGTTCTCTTGCCAGCAATC	63-56°C TD
8	AGATTTGGAGCACAGATACCCGTGA	TGCGGTAGACTTCTGTAAATGCACA	63-56°C TD
9	CCTTCTAAGAACATGCTGGTTGGTT	ATCTCATTCCAGTCATAGCAGAGCT	63-56°C TD
10	AATTAGCCAGTGTGGTGGCACTTG	AGCTTTAGAAGCTGTTTGTACAC	60°C
11a	TCACATAAGCACGTCCTCTCACCAT	CTGGTTTGAATGGCAGTCCACATC	63-56°C TD
11b	TCGCAGGAACATGTGGACTCTCAG	GCGCAACAGA GCAAGACTCT	63-56°C TD
12	TTACAGTGTTCTATAACATAATCAG	AGTAGATACAAGGTCTTGCTGTGTT	63-56°C TD
13	GTGACACTTAGCTGAGTAGTGTTGT	ATGTTAGCCAGGCTGGTCTCAAACT	63-56°C TD
14	GGTCTGTATCTCCTGACCTCATGAT	GCACGTAGCTCTCTGTGTAAAATGA	63-56°C TD
15	GCTGAGATCTAGAACCTAGGCTTCT	ACACACGAGCGCATGCAAACATAGA	63-56°C TD

TD: Touch Down temperature

2.2.4 Agarose Gel Electrophoresis

1. The required amount of agarose was weighed to make a 1% solution (1% is 0.5 gm in 50 ml TBE or 2gm in 200 ml TBE) and dissolved in 1X TBE.

2. The gel casting tray was prepared and the combs were adjusted to the required depth.

3. The agarose solution was boiled in a microwave to digest the agarose powder. It was cooled to less than 40 °C.

4. Ethidium bromide was added from 10mg/ml stock to a final concentration of 0.5µg/µl.

5. The solution was poured on the tray and allowed to solidify.

6. The stipulated volume of samples to be loaded was mixed with 6X loading dye to obtain a final concentration of 1X.

7. The samples were loaded and electrophoresis was started at the required voltage

8. The electrophoresis was allowed to run until the dye traverses $3/4^{th}$ of the gel.

9. The DNA/PCR products bands were visualized using a UV transilluminator and pictures were documented.

2.2.5 Purification of PCR products

1. The PCR products were diluted before sequencing such that 30-50ng of template was available for sequencing.

2. 5µl of a diluted PCR product was mixed with 2µl of ExoSAP-IT reagent for a combined 7µl reaction volume. It was mixed thoroughly and tap spun.

3. This mix was incubated at 37°C for 15 minutes for the enzymes to degrade unincorporated primers and dNTPs.

4. It was further incubated at 80°C for 15 minutes to inactivate ExoSAP-IT enzymes.

2.2.6 Sanger sequencing of cleaned up PCR products

Step1: Cycle sequencing

1. 2µl of template and 1µl of primer (both in the desired concentration) was added in the 96-well sequencing plate (or 0.2ml flat cap PCR tube).
2. The cycle sequencing reaction mastermix was prepared as:
    - MilliQ water - 4.75 µl
    - 5X Sequencing Buffer - 1.75 µl
    - Ready Reaction mix - 0.50 µl
3. 7µl of mastermix was added to each sample
4. The cycle sequencing reaction was setup in the thermal cycler as:

    STEP 1: 96ºC for 2 minutes

    STEP 2: 96ºC for 10 seconds

    STEP 3: 1ºC/second to 55ºC

    STEP 4: 55 ºC for 0.05 minutes (5 seconds)

    STEP 5: 1ºC/second to 60ºC

    STEP 6: 60 ºC for 4 minutes

    STEP 7: Go to steps 2-5 for 24 times

    STEP 8: 4 ºC forever

Step 2: Post cycle sequencing clean up

In the BigDye Xterminator clean up, following steps were carried out:

1. A mixture of the two BigDye Xterminator reagents (Premix) was prepared as:
    - Xterminator Solution- 10µl
    - SAM solution - 45µl
2. 55µl of the premix was added to each sample and the reaction plate was tap spun.

3. The reaction plate was vortexed for 30 minutes.
4. The reaction plate was centrifuged at 2500 rpm for 2 minutes.
5. The reaction plate was tap spun and loaded on the sequencer.

2.2.7 Mutation analysis

The chromatograms were analysed by using Chromas Lite software [Technelysium Pty Ltd]. The data was compared with the reference sequence of the respective genes taken from the NCBI (genomic DNA sequence), UCSC and HGVS (cDNA sequences) databases to identify mutations. All identified mutations were searched in PubMed and different databases including the InSiGHT database (https://www.insight-group.org/mutations/databases/), HGMD, NCBI dbSNP, ClinVar and ExAC. All pathogenic germline mutations identified were confirmed by bi-directional sequencing. Also, the mutations were confirmed on a second independent sample, whenever available. The pathogenicity of the mutations was inferred based on the nature of mutations. In case of splice site mutation outside the canonical splice donor and acceptor sites, transcripts analysis was carried out to infer the pathogenicity of mutations whenever RNA samples were available for analysis.

2.2.8 MLPA analysis to detect LGRs

In the absence of point mutations or small indels upon Sanger sequencing, MLPA analysis was carried out to detect large genomic rearrangements (LGRs) using SALSA MLPA kit following manufacturer's instructions. SALSA MLPA P003 MLH1/MSH2 Probemix and SALSA MLPA P101 STK11 Probemix were used to detect LGRs in MLH1/MSH2 and STK11 respectively.

Protocol:

Step 1: Denaturation of DNA sample

1. Add 5 µl of DNA sample to each tube.

2. Place the tubes in a thermocycler and start the MLPA thermocycler program (given at the end of protocol).

3. Denature sample DNA for 5 minutes at 98°C and cool to 25°C.

Step 2: Hybridization of Probes to sample DNA

1. Prepare a hybridization master mix containing for each reaction: 1.5µl MLPA buffer + 1.5µl probemix. Mix the hybridization master mix well by pipetting or vortexing.

2. After DNA denaturation, add 3µl of the hybridization master mix to each sample tube. Mix well by pipetting up and down.

3. Continue the thermocycler program: incubate for 1 minute at 95°C, then for 16 – 20 hours at 60°C.

Step 3: Ligation of hybridized probes

1. Prepare a Ligase-65 master mix by adding for each reaction: 25µl H2O + 3µl Ligase buffer A + 3µl Ligase buffer B. Then add 1µl Ligase-65 enzyme. Mix well by pipetting gently up and down. (Do not vortex enzyme solutions).

2. Continue the thermocycler program: pause at 54°C.

3. When the samples are at 54°C, add 32 µl of the ligase master mix to each reaction tube. Mix by gently pipetting up and down.

4. Continue the thermocycler program: 15 minutes incubation at 54° (for ligation), followed by 5 minutes at 98°C for heat inactivation of the Ligase-65 enzyme and then pause at 20°C.

Step 4: PCR amplification of Ligated Probes

1. Prepare a polymerase master mix by adding for each reaction: 7.5µl dH2O + 2 µl SALSA PCR primer mix + 0.5 µl SALSA Polymerase. Mix well by pipetting up and down; do not vortex. Keep on ice until use.

2. At room temperature add 10µl polymerase to each tube. Mix by pipetting up and down. Continue the thermocycler program: 35 cycles: 30 seconds 95°C; 30 seconds 60°C; 60 seconds 72°C. End with 20 minutes incubation at 72°C and then pause at 15°C.

Step 5: Fragment Separation by Capillary Electrophoresis

1. Add 1.0µl PCR products with 0.15µl LIZ size standard and 10µl of formamide

2. Denature the samples at 95°C for 5minutes and then place the plate on ice rapidly to avoid renaturation.

3. Load the plate on the sequencer.

2.2.9 MLPA Data Analysis

Data analysed using the Coffalyser.Net Software provided by MRC-Holland. This software calculates the Dosage quotient (representative of the copy number) for each probe in the kit by:

- Intra-sample normalization: Peaks of probes specific for the gene of interest are compared with the peaks of reference probes in a sample which are expected to have a normal copy number.
- Inter-sample normalization: Peaks pattern of the sample of interest is compared to that of the control DNA samples (derived from healthy volunteers) that were included in the same experiment.

- In this way, it is possible to detect abnormal probe signals that indicate deletions or duplications of sequences detected by MLPA probes.

Table 2.11 Dosage Quotients values for copy number status in MLPA

Copy Number Status	Dosage Quotient
Normal	0.85 < DQ < 1.15
Heterozygous duplication	1.35 < DQ < 1.55
Homozygous duplication	1.70 < DQ < 2.20
Heterozygous deletion	0.35 < DQ < 0.65
Homozygous deletion	0

2.3 Screening methodology for early onset sporadic rectal cancer patients using IHC, MSI and PCR

2.3.1 Patients and samples

A total of 31 histologically confirmed rectal or recto-sigmoid cancer samples and their adjacent normal tissues were collected retrospectively along with archived formalin fixed paraffin embedded (FFPE) blocks from National Tumour Tissue Repository (NTTR), TMH Parel under the Hereditary Cancer Consortium Project and Founder mutation project approved by the IRB of Tata Memorial Centre (TMC). These patients were enrolled from 2007 to 2015 in TMH, Parel. Clinico-pathological data like age, gender, family history of CRC, location (low/high rectal tumours), stage, differentiation, and metastasis were collected prospectively from electronic medical record of patients from TMC website. Primary rectal cancer or recto-sigmoid cancer samples from patients below the age of 40 years, with no known family history and

with aggressive /advanced stage poorly differentiated tumour with mucinous type / signet ring cell type carcinoma were included in this study.

2.3.2 DNA extraction

Protocol for DNA extraction from tissue samples: Genomic DNA was extracted using the PAXgene Tissue DNA Kit (Qiagen) as per the manufacturer's protocol.

1. 20mg of tissue were minced in 180µl of TD1 buffer and were transferred to a 2mL microcentrifuge tube.
2. 20µl of proteinase K (25mg/mL) (Invitrogen Inc., Carlsbad, CA, USA) was added to the mix, vortexed and incubated at 56°C overnight.
3. The microcentrifuge tube was centrifuged to remove drops from inside the lid.
4. 200µl of TD2 buffer was added to the mix, pulse-vortexed for 15s and were incubated at 70ºC for 10 minutes.
5. 200µl of absolute ethanol was added to the mix immediately followed by mixing thoroughly on by vortexing.
6. The lysate was allowed to pass through PAXgene DNA spin column placed in a 2 ml processing tube by centrifugation at 6000 X g for 1 minute.
7. After discarding the flow-through, the mix was washed with 500µl of TD3 and TD4 buffers at 6000X for 1 minute in 2 steps.
8. After washing, the column membrane is dried by centrifuging 20,000X for 3 minutes.
9. The DNA bound to the PAXgene DNA spin column was retrieved by eluting with 50-200µl of TD5 buffer under centrifugation at 20,000X for 1 minute.
10. Repeated elutions were performed to improve the total yield of DNA.

Protocol for DNA extraction from FFPE blocks: Genomic DNA was extracted from FFPE blocks using Qiagen FFPE DNA extraction kit as per the manufacturer's protocol.

1. 10-15 sections of 5–10µm thickness were cut and immediately placed in a 1.5ml microcentrifuge tube and 1ml xylene was added to the sample. The lid was closed and vortexed vigorously for 10 seconds.
2. The tube was centrifuged at full speed for 2 mins at room temperature (15–25°C).
3. The supernatant was removed and 1ml ethanol (96–100%) was added to the pellet, and mixed by vortexing.
4. The pellet is centrifuged at full speed for 2 minutes at room temperature and the supernatant was removed by pipetting.
5. Any residual ethanol was removed using a fine pipet tip and incubated at room temperature till all residual ethanol was evaporated.
6. The pellet was resuspended in 180µl of Buffer ATL to which 20µl proteinase K was added, and mixed by vortexing.
7. The sample is incubated at 56°C overnight followed by 90°C incubation for 1 hour.
8. 200µl Buffer AL was added to the sample, and mixed thoroughly by vortexing. Then 200µl of ethanol (96–100%) was added and mixed again thoroughly.
9. The lysate was allowed to pass through QIAamp MinElute column placed in a 2 ml processing tube by centrifugation at 6000 x g for 1 minute.
10. After discarding the flow-through, the mix was washed with 500µl of AW1 and AW2 buffers at 6000 x g for 1 minute in 2 steps.
11. After washing, the column membrane is dried by centrifuging 20,000 x g for 3 minutes.

12. The DNA bound to the QIAamp MinElute column was retrieved by eluting with 50-200µl of ATE buffer under centrifugation at 20,000 x g for 1 minute.

DNA quantification and quality analysis were performed using Nanodrop ND-1000 and electrophoresis on 0.5% agarose gels respectively.

2.3.3 Immunohistochemistry (IHC)

IHC staining for MLH1, MSH2, MSH6, *PMS2* and β-catenin was carried out to determine the status of MMR and Wnt pathway respectively using a two-step EnVision polymer based immunohistochemistry developed by DAKO (DAKO EnvisionPlus HRP kit, Dako Denmark A/S, Glostrup, Denmark)

1. A 4µ thin section of FFPE tissue was cut using a microtome and spread on to a coated slide using 60°C tissue floatation bath.
2. The sections were allowed to attach to the glass slide by incubating at 37°C incubator for 2 hours.
3. The slides were heated with sections for 37°C overnight and then at one hour at 60°C in a hot air oven.
4. Deparaffinization was carried out in three changes of Xylene for 10 minutes each.
5. The tissue sections were then rehydrated in graded alcohol of 100%, 70%, and 50% one change of 5 minutes in each.
6. The slides were then rinsed in distilled water for 5 minutes.
7. Antigen retrieval was done with 10mM Tris-1mM EDTA buffer (pH 9.0) using pressure cooker (2 whistles).
8. After antigen retrieval the sections were brought to room temperature.
9. The sections were washed twice in working wash buffer made from EnVision™ FLEX Wash Buffer (20x) for 5 minutes each.

10. 50µl-100µl of EnVision FLEX Peroxidase-Blocking Reagent was added to the sections & incubation was carried out at room temperature for 30 minutes in dark.
11. The sections were washed twice in working wash for 5 minutes each.
12. Excess of buffer was wiped and the slides were arranged in a humid chamber.
13. The slides were then incubated overnight with appropriately diluted primary antibodies (50-100µl) at 4^{o}C. (Antibody Diluent - Da Vinci Green pH 7.3; MLH1 →1: 25, MSH2 →1:250, MSH6 → 1: 25, *PMS2* →1: 40, β-catenin→1:150)

    Note: Be sure that sections are completely covered.
14. The sections were washed twice in working wash for 5 minutes each.
15. Excess buffer was tapped, wiped and the slides were arranged in the humid chamber.
16. The slides were incubated with EnVision FLEX/HRP detection reagent for 30 minutes at room temperature.
17. The sections were washed twice in freshly prepared working Tris buffer saline (TBS) (pH 7.4) for 5 minutes each.
18. Excess of buffer was wiped and the slides were arranged in a humid chamber.
19. Slides were incubated in EnVision FLEX Substrate Working Solution (prepared by mixing thoroughly 1drop EnVision FLEX DAB+ Chromogen per 1 mL EnVision FLEX Substrate Buffer) for 3-10 minutes at room temperature (Incubation time is monitored under the microscope).
20. The slides were washed in TBS for one minute to stop the reaction.
21. The tissue sections were counterstained lightly with Haematoxylin
22. The slides were washed under running tap water.

23. The tissue sections were again dehydrated using two changes of 95% ethanol, 2 changes of absolute ethanol for 5 minutes each.
24. The slides were cleared in 4 changes of Xylene of 5 minutes in each.
25. The slides were mounted in DPX and allowed to dry, following which they were examined using a microscope.

2.3.4 Screening for Microsatellite Instability (MSI)

Screening of MSI was conducted as per guidelines set by the National Cancer Institute (NCI, USA) reference panel. Genotyping was performed using the NCI panel of 5 microsatellites (BAT25, BAT26, D5S346, D17S250, D2S123). The strategy was to PCR amplify the microsatellite markers using a pair of primers in which Forward primer is labelled with a fluorescent dye. The primer sequences and annealing temperatures for all these markers is given in Table 2.12. The amplicons were then loaded on sequencers for fragment analysis in which PCR products of different sizes are separated by capillary electrophoresis. A fluorescently labelled size standard is added along with the sample to allow molecular size comparison of fragments. The data can be analysed using GeneMapper software from Thermo Fisher to determine the allele sizes of the PCR product.

Table 2.12 Primer sequences and annealing temperature of MSI markers

MSI marker	Forward primer	Reverse primer	Annealing temperature
BAT26	TCGCCTCCAAGAATGTAAGT	TCTGGATTTTAACTATGGCTC	55-48°C TD
BAT25	TGACTACTTTTGACTTCAGCC	AACCATTCAACATTTTTAACC	55-48°C TD
D2S123	AAACAGGATGCCTGCCTTTA	GGACTTTCCACCTATGGGAC	55-48°C TD
D5S346	ACTCACTCTAGTGATAAATCGGG	AGCAGATAAGACAAGTATTACTAG	55-48°C TD
D17S250	GGAAGAATCAAATAGACAAT	GCTGGCCATATATATATTTAAACC	55-48°C TD

TD: Touch Down temperature

2.3.5 Mutation screening of key genes involved in CRC carcinogenesis

Screening for mutations were performed by PCR-DNA sequencing on DNA isolated either from frozen tissue samples or archived FFPE blocks following protocol described in section 2.2. Primer sequences of all the genes screened and their respective annealing temperature are listed in Table 2.13.

Table 2.13 Primer sequences and annealing temperatures of key genes

Gene	Forward primer	Reverse primer	Ta
APC (Mutation Cluster Region, MCR)			
APC MCR 1	GCTCAAGCTTGCCATCTCTT	TATGGGCAGCAGAGCTTCTT	62◦C
APC MCR 2	CCAGGAACTTCTTCAAAGCG	GTGAAGGACTTTGCCTTCCA	62◦C
APC MCR 3	GTCAATACCCAGCCGACCTA	AGGCTGATCCACATGACGTT	59◦C
APC MCR 4	AACGTCATGTGGATCAGCCT	TGCTGGATTTGGTTCTAGGG	62◦C
APC MCR 5	CAGACGACACAGGAAGCAGA	GCAGCTTGCTTAGGTCCACT	62◦C
APC MCR 6	GTGAACCATGCAGTGGAATG	TGTTGGCATGGCAGAAATAA	59◦C
APC MCR 7	TTTGCCACGGAAAGTACTCC	TATCATCCCCCGGTGTAAAA	59◦C
KRAS Exon 2			
KRAS	AAGGTGAGTTTGTATTAAAAGGTACTGG	TGGTCCTGCACCAGTAATATGC	63.2◦C
BRAF Exon 15			
BRAF	TCATAATGCTTGCTCTGATAGGA	GGCCAAAAATTTAATCAGTGGA	63◦C
CTNBB1 Exon 3			
CTNBB1	GCTGATTTGATGGAGTTGGACATGGC	CCAGCTACTTGTTCTTGAGTGAAGG	63.2◦C

Ta: Annealing temperature

2.4 Statistical analysis

Data for calculation of cumulative risk were submitted for statistical testing using the Statistical Package for the Social Sciences (SPSS Inc, Chicago, IL). The chi square test and log rank test were used to compare differences between MLH1 and MSH2 mutation carriers. Penetrance for age was calculated using the Kaplan Meier survival analysis method and included the 106 mutation carriers. In case of multiple or recurrent colorectal carcinoma or endometrial adenocarcinoma, only the age at first cancer diagnosis was included in the analysis. The observation time for the different cancers was from birth until the date of first cancer diagnosis. A p value below .05 was considered statistically significant. The results from SPSS were also compared with Kaplan Meier analysis using R statistical package.

2.5 Exome sequencing pipeline

Whole-Exome sequencing libraries of genomic DNA from the 28 EOS-RC samples (tumour and adjacent normal) were prepared using Agilent SureSelectXT Human All Exon V5 (which targets 50 Mb of genomic content). Both the library preparation and sequencing of captured library was carried out in MedGenome Labs Ltd, Bangalore, India using the Illumina HiSeq platform to generate 2X150bp sequence reads at 100X sequencing depth.

Data analysis was carried out in ACTREC using the following pipeline:

a. **Quality Check of Fastq files**: FastQC software was used to assess the quality of each base based on Phred score which predicts the probability of an error in base calling.

b. **Alignment with reference genome: Raw paired end reads in** FASTQ format were aligned to the hg19 reference genome using BWA-mem (Burrows Wheelers Aligner) (v.0.7.16a) software. Post-alignment Sequence Alignment Map (SAM) files were generated which were then compressed into Binary Alignment Map (BAM) files.

c. **Post-alignment processing**: Unusual flag information from sample files were removed using SAMtools (v. 1.6-1) fixmate program [54]. Picard tools (v.2.10.0) (https://broadinstitute.github.io/picard/) were used for sorting and duplicate removal steps. Qualimap (http://qualimap.bioinfo.cipf.es/), a tool to evaluate BAM file metrics was used to carry out further Q.C.

d. **Variant calling**: SAMtools mpileup (v. 1.6-1) was used to locate non-reference positions in tumour and germline samples using processed bam files as input. Resulting mpileup files were inputted directly to [55] somatic (version 2.4.3) program to identify somatic variants in tumour. Parameters used were *--min-var-frequency 0.02,--min-coverage-normal 8, --min-coverage-tumour 6, --min-coverage 8, --somatic-p-value 0.05*. VarScan2 [55] processSomatic was used to extract the high confidence somatic variants using parameters as maximum variant allele frequency in germline sample was 0.05 and minimum variant allele frequency for tumour was 0.02. Further, false positive SNVs were removed using fpfilter.pl from VarScan2, which takes input metrics of readcounts (generated using bam-readcount) of variants identified in previous step.

e. **Variant Annotation and filtering**: These high confidence somatic variants were further processed with latest version of ANNOVAR software [56] to functionally annotate these. The variants which were present in non-coding

region (intronic, intergenic, ncRNA, UTR etc.) were filtered out so as to enrich the exonic variants or flanking splice site variants. Amongst the exonic variants, the synonymous variants that were predicted to be silent mutations and the ones which were present at a very high frequency in the dbSNP, ESP6500 or 1000Genome databases were also excluded. Further, only those non-synonymous (missense) variants were included that was predicted to be pathogenic by two or more in silico analysis tools out of 11 tools used to assess the pathogenicity of missense variants (SIFT, PolyPhen2_HDIV, Polyphen2_HVAR, LRT, Mutation Assessor, Mutation Taster, FATHMM, PROVEAN, VEST3, CADD, DANN). Truncating mutations like frameshift and non-sense variants were also included in the study.

CHAPTER 3

PEUTZ-JEGHERS SYNDROME

3.1: Peutz–Jeghers syndrome

Peutz–Jeghers syndrome (PJS; OMIM 175200) is an autosomal dominant hereditary cancer predisposition syndrome with characteristic phenotypic features of mucocutaneous pigmentation and multiple hamartomatous polyps in the gastrointestinal (GI) tract [57].

3.1.1 History of PJS

PJS was first studied in identical twin sisters with spots on the lips and the buccal mucosa by Connor in 1895 and illustrated by Hutchinson in 1896. They described the phenotype of the sisters in their report however did not recognise the connection between polyps and these spots. J.L.A. Peutz made this link in 1921 and published the clinical description of PJS in one Dutch family (the Harrisburg family) by describing it as a gastrointestinal familial polyposis syndrome with pigmentations. The pedigree of this original Dutch family is shown in Fig 3.1. Later H. Jeghers provided more detailed description of the syndrome by studying 10 cases from different families in 1949 [57] in whom approximately half of the affected members suffered from gastrointestinal malignancy. He defined the correlation between pigmented lesions, gastrointestinal polyposis and increased risk of carcinoma. The eponym Peutz-Jeghers Syndrome (PJS) was first used in 1954 by Andre Bruwer [58].

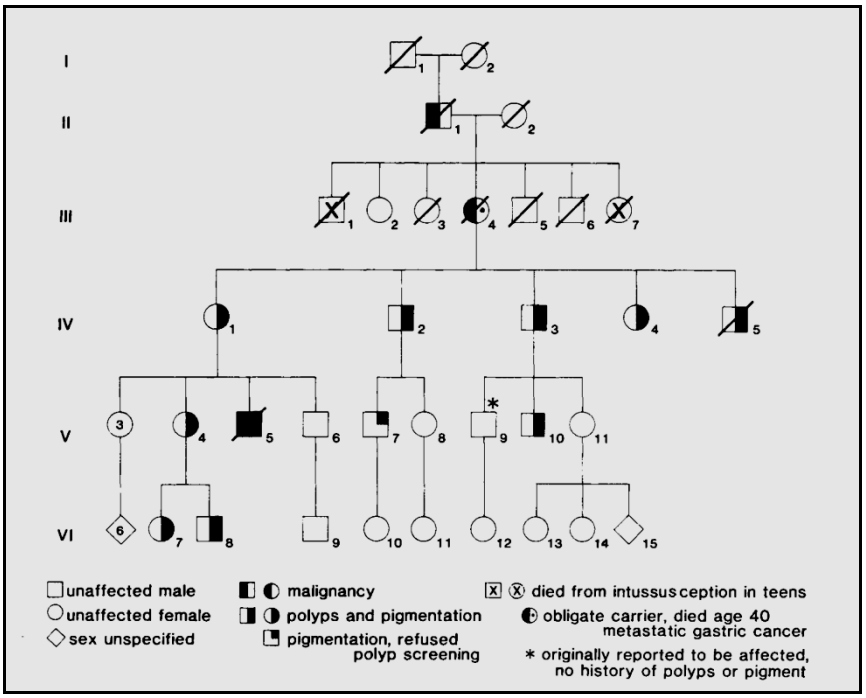


Fig 3.1 Detailed pedigree of the Harrisburg family with a 49 year follow-up (Taken from Ref. [59])

Horrilleno and colleagues gave the first histological description of hamartomatous polyps in 1957 [60]. Since then, descriptions of several different syndromes with hamartomatous polyps in the gastrointestinal tracts have been reported.

In 1997 PJS was linked to loci 19p13.3 with the combination of comparative genome hybridization, loss of heterozygosity (LOH) studies, and targeted linkage analysis [61]. A year later *LKB1* gene germline mutations at that locus were identified in PJS patients by two independent groups [61, 62]. A 78 year follow-up study of the Peutz's original Dutch family led to the identification of an *LKB1* mutation in affected family members [63]. The original gene designation is *LKB1* which is still used. *SKT11* is the official designation for *LKB1* by the Human Genome Organization (HUGO) (http://www.genenames.org/data/hgnc_data.php?hgnc_id=11389).

3.1.2 Clinical manifestations of PJS

Peutz-Jeghers syndrome is characterised by clinical manifestations like mucosal melanin pigmented macules, multiple hamartomatous polyps of the gastrointestinal tract and an increased risk of cancer.

- **Hyperpigmentation**: Mucocutaneous pigmented macules (melanin spots) are present in more than 95% of individuals with PJS and are caused by pigment-laden macrophages in the dermis. These macules appear as small, flat, brown or dark-blue spots present primarily around the mouth crossing the vermilion border, eyes, and nostrils; and sparsely on the perianal area intestinal mucosa, digits, and the dorsal and volar aspects of hands and feet [64]. A photographic representation of these spots is given in Fig 3.2. Mucocutaneous pigmentation usually occurs during the first one to two years of life, increases in size and number over the ensuing years, and finally fades after puberty with the exception of those on the buccal mucosa. Hyper-pigmentation has not yet been shown to undergo malignant transformation.
- **Hamartomatous Polyposis:** Gastrointestinal hamartomatous polyps are present in most patients with PJS. They occur mainly in the small bowel with the jejunum being the most common site followed by the ileum and duodenum. These polyps can become very large in size often causing intussusception [65]. Extraintestinal polyps are also reported and nasal polyposis is supposedly a rare complication. On histology, PJ polyps are hamartomas that characteristically contain a proliferation of smooth muscle extending into the lamina propria giving the characteristic frond-like structure; the overlying epithelium is normal. (Fig 3.2)

❖ **Cancer Risk:** PJS is associated with an increased risk of gastrointestinal as well as extra-intestinal malignancies. In a systematic review of 1644 PJS patients, 349 patients developed 384 malignancies at an average age of 42 years; the most common cancer was colorectal followed by breast, stomach, small bowel and pancreas with the risk for any cancer at the age of 60–70 years varying from 37 to 93% [66]. The risk of extra-intestinal cancers is also increased in individuals with PJS. In a meta-analysis of six studies that included 210 individuals with PJS, extracolonic cancers accounted for 55 of the 66 malignancies (83%) between the ages of 15 and 64 years [67].

Women with PJS have an increased lifetime risk for cancers of the breast (32%-54%), ovary (21%), and cervix (10%). In addition, small, asymptomatic, benign ovarian tumours known as "sex-cord" tumours with annular tubules (SCTAT tumours) occur commonly in women with PJS. Men with PJS have an increased lifetime risk of Sertoli cell testicular tumours (9%) [68].

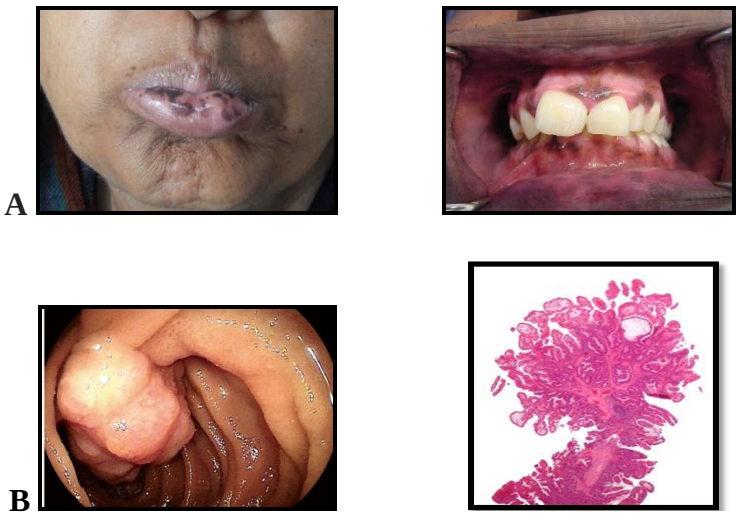


Fig 3.2 A) Melanin pigmentation on the lips & buccal mucosa
B) Hamartomatous polyps [64]

3.1.3 Diagnosis and Management of PJS

❖ **PJS Clinical and differential diagnosis:** In an individual, a clinical diagnosis of PJS may be made when histologically verified hamartomatous polyps are present with atleast two of the following features:

a) characteristic mucocutaneous pigmentation

b) small bowel polyposis

c) PJS associated family history

Careful physical examination should be done to identify melanin spots. Polyps can be detected by endoscopy, x-ray examination, or wireless capsule endoscopy and should be histologically verified as hamartomatous. Genetic testing to identify germline mutation in the *STK11* gene in an individual who meets clinical criteria for PJS is required to confirm the diagnosis of PJS and counsel at-risk family members. A study by Aretz et.al. showed a correlation between PJS diagnostic criteria and *STK11* mutation detection rates. 94% of the patients who met the criteria for PJS were found to harbour *STK11* mutation (64% point mutation, 30% deletions) thereby highlighting the importance of syndromic diagnosis and genetic testing [69]. PJS can be differentiated from other disorders that may present with pigmented mucocutaneous macules or multiple hamartomatous gastrointestinal polyps or both based on the clinical presentation and/or genetic testing.

- Hamartomatous polyps of the small intestine can be associated with hereditary mixed-polyposis syndrome (HMPS) and Cronkhite-Canada syndrome (CCS) Cowden syndrome (CS), Bannayan-Riley Ruvalcaba syndrome (BRRS), and juvenile polyposis syndrome (JPS) [70]. Although the polyps may appear histologically similar in these syndromes they also exhibit distinct morphological

features, which should be taken into consideration to avoid any misclassification. Pigmented spots in BRRS and CS characteristically occur on the glans penis in males and not on the lips as seen in PJS, however CCS closely mimic PJS due to the presence of both hamartomatous polyps and mucocutaneous pigmentation [71]. Unlike PJS, which is associated with germline mutations in the *STK11* gene, CS and BRRS are associated with germline mutations in *PTEN1*. It has been reported that both mutation of *BMPR1A* or a duplication of 15q15.3q22.1 that leads to increased expression of *GREM1* may cause HMPS. Some families with mixed hereditary polyposis syndrome that includes JPS are associated with germline mutations in *SMAD4*, *BMPR1A*, and *ENG* genes. The genetic basis of CCS has not been examined and as of now it is considered to be a non-hereditary condition

- Mucocutaneous pigmentation may be associated with Laugier-Hunziker syndrome (LHS). LHS is an acquired, sporadic, benign disorder that is characterized by intraoral hyperpigmentation on the lips, hard and soft palate, and buccal mucosa [72]. Unlike pigmentation in PJS, which occurs in the first few years of life, LHS lesions are progressively acquired in young or middle-age adults. In addition, LHS is not associated with hamartomatous gastrointestinal polyps or a pathogenic mutation in the *STK11* gene.
- The tumour spectrum of PJS overlaps with several other syndromes like familial adenomatous polyposis (FAP), Lynch syndrome (LS), hereditary breast and ovarian cancer (HBOC) and Li–Fraumeni syndrome (LFS) however they can be distinguished based on the phenotypic features seen in PJS.

❖ **Management**: Guidelines for cancer screening in individuals with Peutz-Jeghers syndrome (PJS) have been proposed by several groups and are largely based on expert opinion and limited observational data.

- Patients should under upper colonoscopy and endoscopy along with small bowel examination (MR enterography or wireless capsule endoscopy) starting at the age of 8 years or when symptoms occur
- Polyps that are causing obstruction or bleeding needs to be removed surgically
- Small bowel resection should be as conservative as possible to avoid the risk of short bowel syndrome
- In women, gynaecologic and breast examinations should be done annually starting from the age of 25 years
- In men, testicular examination and testicular ultrasound examination should be done annually, if clinically indicated
- Consultation with a clinical geneticist and/or genetic counsellor is recommended

3.2 Genetic basis of PJS

PJS is caused by germline mutations in the tumour suppressor gene, *STK11*(OMIM 602216), a serine/threonine kinase. The reported STK11 mutation detection rate ranges from as low as 10% to as high as 90% depending upon the criteria used for defining the PJS cases and the genetic screening method employed. Germline pathogenic *STK11* mutations are identified in 70–90% of patients fulfilling the diagnostic criteria of PJS—hamartomatous polyps, mucocutaneous pigmentation and PJS-associated cancers.

3.2.1 STK11 gene

The *STK11* gene, localized on chromosome 19p13.3 extends over 23 kb of genomic DNA and consists of 9 coding exons and 1 non-coding exon, which occurs within the 3' Untranslated Region (UTR) region of the gene. The 3'-UTR extends over 1449 bp and the 5'-UTR is approximately 338 bp long. Gene transcription takes place in telomere to centromere direction. This gene is spliced in an unusual U12 snRNA dependant manner because intron 2 of this gene begins with ATATCCTT and ends with CCCAC thus deviating from the usual GT/AG splice junctions [61]. A graphical representation of the *STK11* geen is given in Fig 3.3

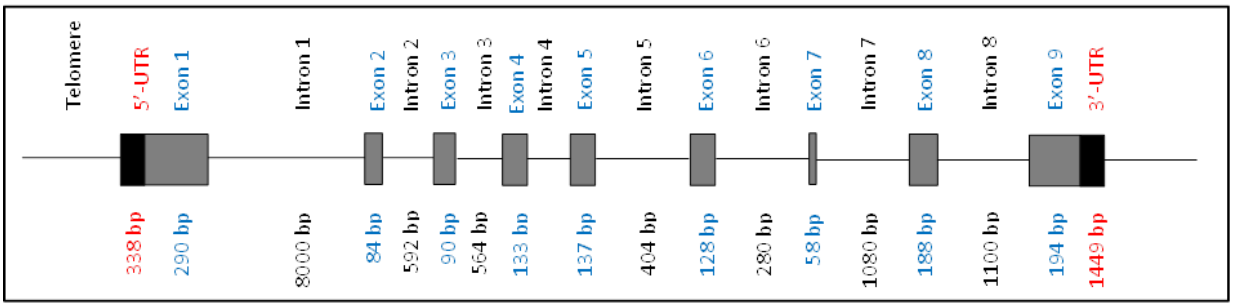


Fig 3.3 Schematic illustration of *STK11* genomic structure

Black boxes indicate the 5'-UTR and 3'-UTR and grey boxes represent exons

3.2.2 STK11 Protein

STK11 encodes a 48.6-kDa serine threonine kinase protein consisting of 434 amino acid residues [73]. In adult humans, STK11 is ubiquitously expressed with higher expression observed in testis and foetal liver [74]. Due to this widespread expression there is an increased risk of several cancer types associated with PJS. The STK11 protein is comprised of three domains: an N-terminal domain (NTD) with the nuclear localization (NL) signal and cytoplasmic retention signal, a catalytic kinase domain which is essential for ATP binding, and a C-terminal non-catalytic regulatory domain

that contains a prenylation motif (CAAX-box) [75]. There are four autophosphorylation sites; threonine 185 (T185), T189, T336, T402, and additionally four phosphorylation sites; serine 31 (S31), S325, T363 and S428, in human STK11/LKB1 [76]. In the CAAX motif, C is the amino acid cysteine (C430), the two 'A' residues are any aliphatic amino acid except alanine and X which represents the C-terminal amino acid can be any amino acid. The CAAX motif mediates the prenylation of many proteins in three sequential steps [77], where prenylation is the addition of the 15 carbon farnesyl group or the 20 carbon geranylgeranyl group to acceptor proteins. Both farnesyl as well as geranylgeranyl are isopropenoid groups derived from the cholesterol biosynthetic pathway. Prenylation results in covalent attachment of either of these isoprenoid groups to the cysteine in the CAAX motif. Prenylation is followed by proteolytic cleavage of the last three amino acids (AAX). The cysteine (C430) in STK11 is no exception to this rule and is prenylated by the addition of a farnesyl group. This prenyalation however was not found to be essential for STK11's ability to suppress cell growth, but it was suggested that it could play a role in cellular location, interaction with a regulatory substrate or perhaps the stability of the protein [77]. Fig 3.4 shows a schematic representation of STK11 protein.

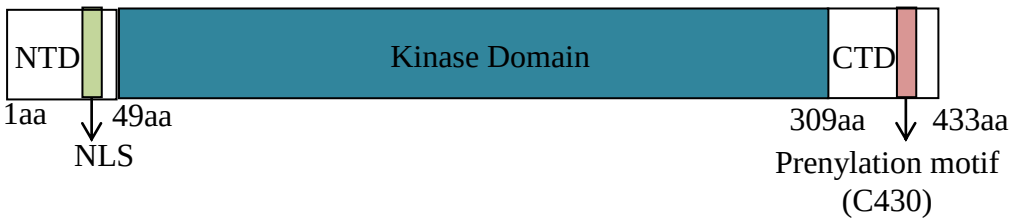


Fig 3.4 Domains of STK11 protein

(NTD: N-terminal domain, NL: nuclear localization signal, CTD: C-terminal domain Catalytic kinase domain is encoded by amino acids 49-309)

STK11 is normally localized in the nucleus which is probably due to its nuclear localization signal [74, 78]. STK11 is translocated to the cytoplasm, on forming a heterotrimeric complex with STE20-related adaptor (STRAD) and scaffolding mouse protein 25 (MO25). The kinase domain and amino acids within the CTD of STK11 are important for binding STRAD. MO25 stabilizes the STK11-STRAD interaction [79, 80]. In the cytoplasm, STK11 acts as a "master kinase" and phosphorylates the "T"/activation loop of up to 13 downstream proteins which are part of an AMPK (5' adenosine monophosphate-activated protein kinase) subfamily termed ARK (AMPK related kinases) [81, 82]. AMPK is a protein kinase cascade that plays an important role in regulating energy homeostasis. Fig 3.5 shows translocation and activation of STK11 by STK11-STRAD-MO25 complex.

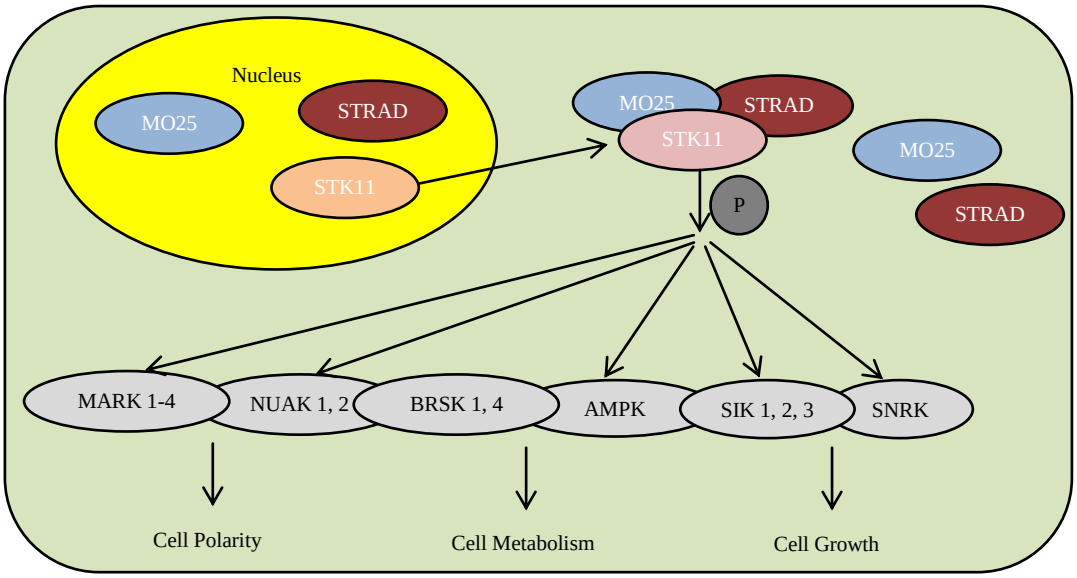


Fig 3.5 Translocation, activation and downstream signalling of STK11

3.2.3 STK11 downstream signalling

STK11 has been reported to play a role in many processes including cell polarity, cell metabolism and cell growth through activation of ARKs. In addition, STK11 has been shown to be involved in cell cycle, cell death, anoikis and gene expression.

- **Cell Metabolism and Growth:** STK11 acts upstream of AMPK, an important sensor and modulator of cellular energy homeostasis. AMPK is a heterotrimeric complex composed of α catalytic subunit and β and γ regulatory subunits. When ATP levels are low, AMP or ADP bind the γ subunit causing a conformational change. This allows STK11 or calcium/calmodulin-dependent protein kinase kinase β (CaMKKβ) [83-85]to phosphorylate AMPK at T172 within the "T" loop of the α subunit [86-88]. AMPK restores ATP levels by stimulating catabolism (glycolysis) and inhibiting anabolism (protein synthesis, lipogenesis) [89, 90].
- **Cell cycle:** STK11 is shown to play a role in cell cycle arrest by affecting expression of tumour protein 53 (TP53) and the cyclin dependent kinase (CDK) inhibitors; CDKN1A (p21), CDKN1B (p27) and CDKN2A (p16) [91]. A study showed that in the G361 cell line, ectopic LKB1 expression increased p21 promoter activity in a TP53 dependent manner, to induce G1 cell cycle arrest [91]. In another study the Peutz-Jeghers gene product STK11 was shown to mediate p53-dependent apoptosis [92]. Other studies have shown that ectopic expression of STK11 stimulated p27 thereby inducing G1 cell cycle arrest and knock down of STK11 led to decrease in TP53 and p16 stimulated G1/S transition [93, 94].
- **Cell Death and Survival:** STK11 has been reported to play a role in apoptosis and anoikis. Studies have shown that kinase activity of STK11 is important in stimulating TP53-dependent apoptosis [92]. LKB1 also plays a role in anoikis,

which is apoptosis stimulated by anchorage independent growth. LKB1 activates salt inducible kinase 1 (SIK1) by phosphorylating T182 which stimulates SIK1 autophosphorylation at S186. SIK1 is required for TP53 dependent anoikis [95].

- **Cell cytoskeleton, polarity and adhesion:** STK11 phosphorylates MARK, BRSK and NUAK which play roles in cell cytoskeleton, cell polarity and cell adhesion. STK11 mediated phosphorylation and activation of Brain Selective kinases (BRSK), BRSK1 and BRSK2 is involved in cortical neuron polarization leading to axon initiation and specification [96]. The phosphorylation of other kinases like MARK (microtubule- affinity-regulating-kinase), NUAK family SNF1-like kinase 1 (NUAK) by STK11 regulates cellular polarity by remodelling the actin cytoskeleton [97].

3.3 Mutation Spectrum of STK11

Germline mutations in STK11 gene have been identified in 30%-80% of PJS patients depending upon the diagnostic criteria used and screening method employed. Around 400 distinct pathogenic or likely pathogenic mutations in *STK11* gene in PJS patients have been recorded in Human Gene Mutation database (HGMD, http://www.hgmd.cf.ac.uk) and most of these mutations are in the catalytic kinase domain (amino acids 49–309). These mutations can affect STK11 kinase activity or its ability to interact with STRAD or MO25 [98]. Most mutations are frameshift or nonsense, which result in an abnormal truncated protein and the consequent loss of kinase activity. Missense and splice site mutations are also reported in PJS however in lower frequency (http://www.hgmd.cf.ac.uk). A significant proportion of patients do not appear to have a family history of disease and possibly represent de novo

mutations. Large Genomic Rearrangements (LGRs) were reported to account for up to a third of all *STK11* pathogenic mutations in a series of PJS cohorts [69, 99, 100]. Studies of Alu elements in the *STK11* gene have highlighted the high rate of large genomic deletions. A case series by De Rosa et al. (2000) and later by Resta et al. shows that non-homologous recombination is a putative mechanism for deletions within *STK11* [101, 102]. Out of 30 breakpoints 16 were located in Alu-elements in large deletions of STK11. High Alu density in *STK11* (26% c.f. average density in human genome of 10%) confers instability to the region. The density and spread of Alu elements across the gene may explain distribution of mutations throughout [103, 104]. Alu-elements have been shown to be associated with non-homologous recombination which is known to predispose to copy number variations [101]. Alu elements are present in all deletions of exons 2–3 and these deletions are the most recurrent large genomic rearrangements in PJS.

3.4 Genotype-phenotype correlation in PJS

Several studies have come to a consensus regarding the risk of developing PJS associated cancer, type of cancer developed and the age of onset, however no significant correlation could be established between the mutation type and site or effect of these on clinical and phenotypic manifestations [68]. A study suggested that cancer risk may be different in patients with and without *STK11* mutation [105]. Several small studies have suggested association between the type and site of mutation with the onset and risk of malignancy [106]. Truncating mutations may result in early onset as compared to missense mutation or absence of mutation [105]. Truncating mutations were also shown to have a positive association with the polyp burden [107].

Studies also assessed mutations *affecting* the 3′end of the kinase domain which negatively influence the action of STK11 by impairing its ability to bind the STRAD-MO25 complex [108]. However, in a survey conducted on 419 PJS patients, cancer risks was found to be similar in patients with PJS and *STK11* mutations and those without any mutation, suggesting that the type or site of *STK11* mutation may not have a significant influence on the risk of malignancy [109].

3.5 Hypothesis and Objective of the study

Due to the rareness of PJS, correlation between the clinic-pathological features and the mutation spectrum is not well established. There is also wide variability in *STK11* mutation frequencies from as low as 10% to as high as 90% depending upon the criteria used for defining the PJS cases and the genetic screening method employed [109]. The current knowledge regarding the spectrum of *STK11* gene mutation and the genotype phenotype correlations is derived mainly from studies in Caucasian cohort [66, 68, 100]. The genotype–phenotype correlation in PJS has not been examined in detail in the Asian population, particularly the South Asians. So far only a single PJS family with pathogenic mutation in *STK11* has been reported from India [110]. Moreover, the symptoms of PJS show great variation among patients, even intra-familiarly, highlighting the syndrome's extended phenotypic spectrum and the importance to study individual families in every population. *STK11* genotype-phenotype association studies in different geo-ethnic groups can enrich the existing knowledge about phenotypic consequences of distinct *STK11* mutations and guide counselling and risk management in different populations.

We hypothesized that the Indian PJS patients may have some novel mutations and novel phenotypes or genotype-phenotype correlations that are distinct from the Caucasian population. One of the objectives of this study was therefore to characterize the mutation spectrum and delineate the phenotypic features in Indian cohort of 20 PJS families. The methodology was to PCR amplify the entire coding and flanking intronic region of *STK11* gene followed by Sanger sequencing to identify mutation. MLPA analysis was carried out to study LGRs in *STK11* in cases where no point mutation or small insertion/deletion was identified in the *STK11* gene. Detailed phenotypic characterization was carried out based on the pedigree and medical records.

3.6 Results

3.6.1 Clinical Characteristics of Patients

The 20 unrelated South Asian PJS families reported here represent the diverse regions of the Indian subcontinent with 8 hailing from Eastern Indian states of Assam, Orissa, West Bengal or from Bangladesh; 5 from North Indian states of UP and Punjab; 3 from Western Indian state of Maharashtra; and 4 from Central Indian states of MP and Chhattisgarh. Of the 20 probands, 14 were females and 6 were males. Mucocutaneous pigmentation was noted in 18/20 probands, while polyps were seen in 8/10 probands in whom upper GI endoscopy was done. Of the 8 probands with upper GI polyps, 6 have histologically confirmed hamartomatous polyps, 1 proband with classical features of PJS was reported to have adenomatous polyps and for another case with classical PJS features the polyp histology was not available. Only 6/20 probands had documentation of both the characteristic features of PJS—hamartomatous polyps and mucocutaneous pigmentation (representative pedigree in Fig 3.6). In the remaining 14

probands, PJS was suspected due to the presence of only one feature—mucocutaneous pigmentation (n = 12) or hamartomatous polyps (n = 2) along with personal or family history of PJS-associated cancer (representative pedigree in Fig 3.7). All except two probands had PJS-associated cancer diagnosed at a median age of 39 years (range, 22–56 years). The primary cancer site was breast in nine cases, GI tract in five cases (small bowel in two and colorectal in three cases), ovary in two cases, uterus and nasopharynx in one case each. In 8/20 families, PJS-associated cancers or polyps were noted in 2 or more first-degree relatives of the proband.

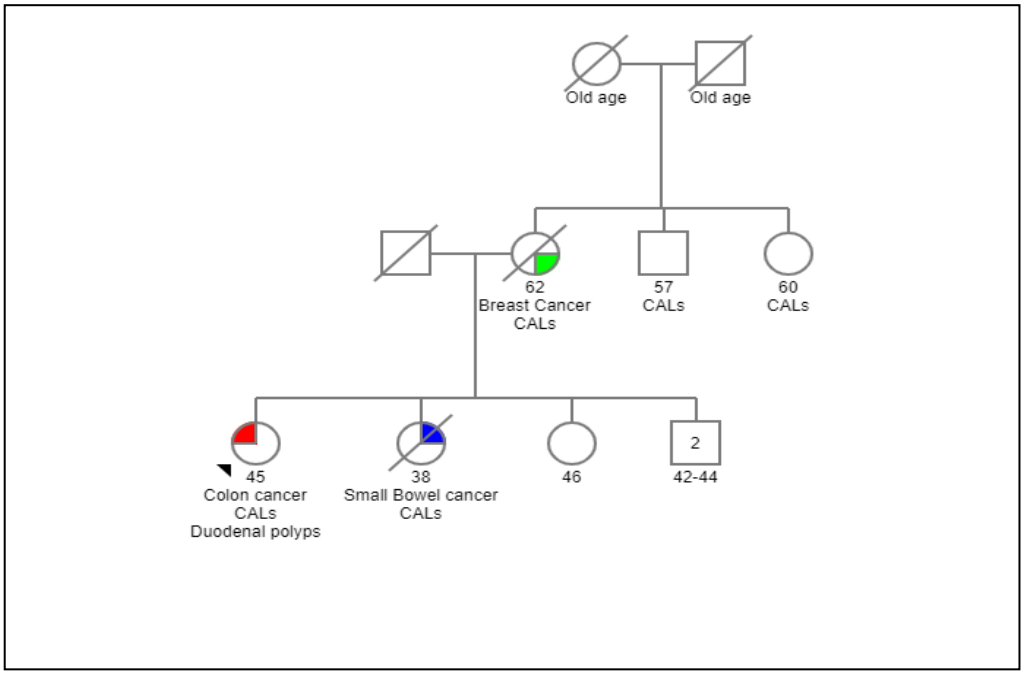


Fig 3.6 Representative Pedigree of a classical PJS family (Proband had histologically confirmed hamartomatous polyps and mucocutaneous pigmentation)

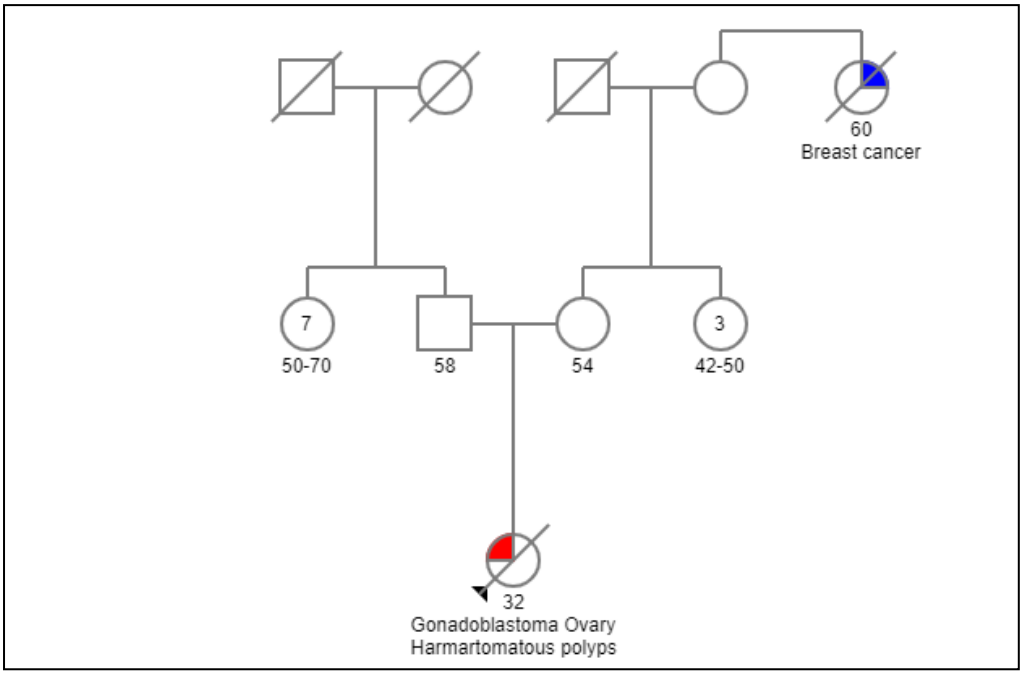


Fig 3.7 Representative Pedigree of a suspected PJS case (Proband had histologically confirmed multiple hamartomatous polyps)

3.6.2 Mutation spectrum of STK11

Pathogenic or likely pathogenic germline heterozygous mutations in the *STK11* gene were identified in 8/20 probands. A known exon 6 frameshift mutation (c.842c.843insC) was identified in two unrelated families. For one proband, both parents were negative for this mutation, implying it to be a *de novo* pathogenic mutation. This mutation occurred at the mononucleotide repeat (C6) mutational hotspot, resulting in a frameshift that leads to premature termination of the STK11 protein 3 amino acids downstream to codon 282 (p.L282Afs∗3). Another known exon 4 frameshift mutation (c.574dup) was identified in one patient. This mutation inserts a single-base (A) insertion between nucleotides 574 and 575 of exon 4 thereby introducing a frameshift at codon K191 and premature termination of the STK11 protein. In two suspected PJS patients we identified an exon 4 missense mutation

(c.542A>T), which has not been reported in literature or mutation databases such as InSiGHT (LOVD), HGMD, NCBI-dbSNP, ClinVar and Exome Aggregation Consortium (ExAC). The substitution that replaces asparagine with isoleucine at codon 181 (p.N181I) has not been functionally characterized but was predicted to be disease associated by all three in-silico tools (Polyphen, Align GVGD and Mutation Taster). The asparagine residue is highly conserved and lies within the protein kinase domain of STK11 protein. Another substitution at the same position, c.542A>G is reported as a likely pathogenic mutation in 3 PJS patients in the ClinVar database. Hence, we considered this *STK11* missense mutation as likely pathogenic. MLPA analysis was performed in all but the two cases with the frameshift mutation. Large genomic deletions were identified in three probands—deletion of exons 2 and 3 in two families and deletion of exon 7 in one family.

3.6.3 Genotype-phenotype data of PJS families

Of the eight probands in whom a pathogenic or likely pathogenic *STK11* mutation was identified, six fulfilled all the established clinico-pathological criteria for PJS. Two probands (PJ6 and PJ7) with germline *STK11* pathogenic mutation did not have any GI symptoms and there was no documentation of presence or absence of hamartomatous polyps as GI endoscopy had not been performed as yet. One proband (PJ5) carrying a *STK11* pathogenic mutation with hamartomatous small bowel polyps and ovarian gonadoblastoma did not have any mucocutaneous pigmentation on examination. Screening of relatives in these eight families with pathogenic mutations identified three additional individuals harbouring the family-specific *STK11* pathogenic mutation. All three carriers had mucocutaneous pigmentation and GI polyps were seen in two carriers who have undergone GI endoscopy till date. The

genotype–phenotype data on the 11 carriers of germline *STK11* pathogenic mutation is given in Table 3.1. In these eight families with an identified *STK11* pathogenic mutation, a total of ten PJS-associated cancers were noted with seven cancers in six probands (bilateral breast cancer in PJ1) and two cancers in two obligate first-degree relatives of PJ2 (Table 3.1). The ten PJS-associated cancers were breast (6), small bowel (2), ascending colon (1) and ovarian gonadoblastoma (1). The family and phenotype details of these carriers of *STK11* pathogenic mutations are shown in Table 3.1.

Table 3.1 Phenotypic features and mutation spectrum of STK11 gene found in the Indian PJS cohort

Family (Proband/ Relative)	Gender / Age at last FU or death	Upper GI Polyps	OCP	Cancer Site (age at diagnosis)	Syndrome (FH +/-)	Known or *Novel STK11 mutations	STK11 polymorphisms [Polymorphisms co-occurring in multiple cases]	Other gene mutations
PJ1 (Proband)	F/46	Yes	Yes	IDC III Bilateral Breast (38, 46) ER+ve, PR/Her2–ve	Classical PJS (FH -)	*Exon 7 deletion	c.290+36 G>T; [c.375-49 G>A; c.464+40_464+46dup]	NT
PJ1a (Nephew)	M/31	Yes	Yes	Nil	Classical PJS (FH -)	Exon 7 deletion		NT
PJ1b (Brother)	M/61	Yes	Yes	Nil	Classical PJS (FH -)	Exon 7 deletion		NT
PJ2 (Proband)	F/45	Yes	Yes	PDA Ascending Colon (39)	Classical PJS (FH+)	Exon 2-3 deletion		NT
PJ2a (Mother)	F/62	NK	Yes	IDC II Breast (60) ER/PR+ve, Her2 status-NK	Suspected PJS	Obligate carrier		
PJ2b (Sister)	F/38	NK	Yes	Small Bowel (38)	Suspected PJS	Obligate carrier		
PJ3 (Proband)	M/41	Yes	Yes	MDA Small Bowel (40)	Classical PJS (FH -)	c.842_843insC	#c.478 C>T (L160L)	NT
PJ3a (Son)	M/14	NK	Yes	Nil	Classical PJS (FH -)	c.842_843insC		NT
PJ4 (Proband)	M/21	Yes	Yes	Nil	Classical PJS (FH -)	Exon 2-3 deletion		NT
PJ5 (Proband)	F/34	Yes	No	Gonadoblastoma Ovary (32)	Suspected PJS (FH -)	c.842_843insC		NT
PJ6 (Proband)	F/47	NK	Yes	IDC III Breast (42) ER/PR+ve, Her2-ve	Suspected PJS (FH -) DD HBOC	*c.542 A>T	[c.290+36 G>T(HMZ); c.290+78 C>T]; c.375-49 G>A (HMZ); c.465-51 T>C (HMZ)	No BRCA1 hotspots

PJ7 (Proband)	F/54	NK	Yes	IDC III Breast (49) ER/PR+ve, Her2-ve	Suspected PJS (FH -) DD HBOC	*c.542 A>T		No BRCA1 hotspots
PJ20 (Proband)	F/43	Yes	Yes	Breast (42)	Suspected PJS (FH -) DD HBOC	c.574dup	c.290+36 G>T; c.464+40_464+46dup	NT

Abbreviations: FU: Follow Up; NK: Not Known; OCP: Oro-cutaneous pigmentation; IDC: Infiltrating Ductal Carcinoma; PDA: Poorly Differentiated Adenocarcinoma; MDA: Moderately Differentiated Adenocarcinoma; FH: Family history of first degree relative with PJS associated cancer; DD: Differential Diagnosis of syndrome; del- deletion; ins- insertion; dup- duplication; NT: Not Tested; MMR- Mismatch Repair genes; HMZ- homozygous [*]Novel mutation; [§]Novel likely benign *STK11* mutation. Homozygous mutations are represented by both alleles (eg. GG→TT) and heterozygous mutations are represented by single allele (eg. C→T). The *STK11* reference sequence used was NM_000455.4 from (NCBI)

In the remaining 12 families, (1 with classical PJS and 11 with suspected PJS), no pathogenic or likely pathogenic *STK11* mutation was identified. In the three breast cancer patients with suspected PJS in whom no *STK11* pathogenic mutation was identified, screening for *BRCA1* gene identified a pathogenic *BRCA1* mutation in a young lady with breast cancer and mucocutaneous pigmentation along with family history of breast cancer. Genetic screening of hotspot pathogenic mutations in mismatch repair (MMR) genes *MLH1* and *MSH2* in four Colorectal Cancer (CRC) patients with suspected PJS did not identify pathogenic mutation in any MMR gene. The genotype–phenotype data on these 12 PJS families with no *STK11* pathogenic mutation is shown in Table 3.2.

Table 3.2: Phenotypic and genotypic features of the Indian PJS cohort with no STK11 mutation

Family (Proband/Relative)	Gender / Age at last FU or death	Upper GI Polyps	OCP	Cancer Site (age at diagnosis)	Syndrome (FH +/-)	#Novel likely benign STK11 variants	STK11 polymorphisms [Polymorphisms co-occurring in multiple cases]	Other gene mutations
PJ8 (Proband)	M/39	Yes	Yes	MDA Caecum (35)	Classical PJS (FH +)	#c.478 C>T (L160L)	[c.290+36 G>T; c.290+78 C>T] [c.375-49 G>A; 464+40_464+46dup];	No MMR hotspots
PJ9 (Proband)	F/41	NK	Yes	IDC III Bilateral Breast (22) ER/PR -ve	Suspected PJS (FH +) DD HBOC	#c.290+74 T>C	c.375-49 G>A; c.920+7 G>C	No BRCA1 mutation
PJ10 (Proband)	F/34	NK	Yes	IDC III Breast (25) ER/PR/Her2 -ve	Suspected PJS (FH -) DD HBOC	#c.920+45 C>G		No BRCA1 hotspots
PJ11 (Proband)	F/59	NK	Yes	Poorly differentiated carcinoma Nasopharynx (46)	Suspected PJS (FH -)	#c.478 C>T (L160L)		NT
PJ12 (Proband)	M/35	NK	Yes	Signet Ring cell Adenocarcinoma Rectal (27)	Suspected PJS (FH +) DD Lynch syndrome	#c.374+20 G>A	c.375-49 G>A; c.920+7 G>C	No MMR hotspots
PJ13 (Proband)	F/62	NK	Yes	Carcinoma Ovary (56)	Suspected PJS (FH +)	#c.478 C>T (L160L)	c.375-49 G>A; c.920+7 G>C	NT
PJ14 (Proband)	M/32	No	Yes	Nil	Suspected PJS (FH +)	No	[c.290+36 G>T; c.290+78 C>T]; [c.375-49 G>A;	NT

							c.464+40_464+46dup]; c.465-51 T>C	
PJ15 (Proband)	M/40	Yes	No	Small bowel (39)	Suspected PJS (FH -) DD Lynch Syndrome	No	c.290+36 G>T; [c.375-49 G>A (HMZ); c.464+40_464+46dup]; c.465-51 T>C; c.920+7 G>C	No MMR hotspots
PJ16 (Proband)	F/69	NK	Yes	IDC III Breast (54) ER/PR +ve, Her2 status-NK	Suspected PJS (FH -) DD HBOC	No	c.290+36 G>T	No BRCA1 hotspots
PJ17 (Proband)	F/54	NK	Yes	Uterine (45)	Suspected PJS (FH -)	No	c.920+7 G>C	NT
PJ18 (Proband)	F/56	NK	Yes	IDC II Breast (47) ER/PR +ve, Her2 -ve	Suspected PJS (FH -) DD HBOC	No	c.375-49 G>A; c.920+7 G>C	No BRCA1 hotspots
PJ19 (Proband)	F/27	NK	Yes	IDC III Breast (23) ER/PR/Her2 -ve	Suspected PJS (FH -) DD HBOC	No	[c.375-49 G>A; c.464+40_464+46dup]; c.465-51 T>C	BRCA1 Exon 2: c.185delAG

Abbreviations: NK: Not Known; OCP: Oro-cutaneous pigmentation; IDC: Infiltrating Ductal Carcinoma; PDA: Poorly Differentiated Adenocarcinoma; MDA: Moderately Differentiated Adenocarcinoma; PDC: Poorly Differentiated Carcinoma; SRCA: Signet Ring Cell Carcinoma; FH: Family history of first degree relative with PJS associated cancer; DD: Differential Diagnosis of syndrome; del- deletion; ins- insertion; dup- duplication; NT: Not Tested; MMR- Mismatch Repair genes; HMZ- homozygous [ǁ]Age at last follow up or death; [¶]Age at diagnosis; [§]Novel likely benign *STK11* mutation. Homozygous mutations are represented by both alleles (eg. GG→TT) and heterozygous mutations are represented by single allele (eg. C→T). The *STK11* reference sequence used was NM_000455.4 from (NCBI)

3.6.4 STK11 polymorphisms and likely benign mutations

Five known STK11 intronic polymorphisms were identified in multiple cases. The minor allele frequency of these polymorphisms was in the range of 0.16 to 0.46 in the ExAC and 1000 Genome database. In addition, we identified three intronic mutations and one exonic mutation (c.478C>T) that have not reported in any database or literature. The c.478C>T mutation identified in four cases was categorized as likely benign since it was a synonymous mutation (L160L) and co-occurred with a deleterious frameshift *STK11* pathogenic mutation in one case (PJ3). The intronic mutations were also considered to be likely benign as they were deep intronic and were not predicted to affect splicing by the Human Splice Finder tool. The details of cases in which these polymorphisms or likely benign novel mutations were identified are provided in Tables 3.1 and 3.2.

3.7 Discussion

PJS is a rare autosomal dominant hereditary cancer syndrome that requires detailed phenotypic characterization and comprehensive genetic analysis. Several small cohort studies and meta-analysis have been carried out to study the genotype–phenotype correlation; however, a clear consensus could not be reached. Moreover, most of this information is largely derived from the Caucasian population [66, 68, 109]. So far only a single case of PJS with *STK11* pathogenic mutation has been identified in the two studies reported from the populous South Asia, which includes India [110, 111]. PJS patients typically show heterogeneous phenotypes with regards to oral and GI lesions, including different onset times, disease severity, distribution of GI symptoms and level of malignancy which enhances the difficulty for proper diagnosis and

management. Being quite a rare syndrome, small cohort studies and overlapping features with other syndromes have added to this difficulty. Therefore, we carried out detailed genotyping and phenotyping characterization of 20 Indian PJS families for a better understanding of the syndrome and expand our existing knowledge about the phenotypic associations of *STK11* mutation carriers.

Using a combination of Sanger sequencing of *STK11* gene with MLPA analysis to study large genomic rearrangements, we were able to identify deleterious germline *STK11* mutations in 8 families. Extended testing of family members of the mutation carrier patients identified additional 3 confirmed and 2 obligate carriers of family specific *STK11* mutation. Of the eight families with pathogenic or likely pathogenic *STK11* mutations, the identified mutations in three families were novel and not reported in any databases or literature. In addition, likely benign but novel *STK11* mutations never reported earlier were identified in 8/20 PJS cases we genotyped. This underscores the need for detailed genetic analysis of different populations across the world and collation of detailed genotype phenotype data in large international databases such as the InSiGHT database.

The seemingly low *STK11* pathogenic mutation detection rate of 8/20 (40%) in our cohort is possibly a reflection of the PJS syndromic clinical characterization in our study. The pathogenic mutation detection rate was 5/6 (83%) in the classical PJS families with mucocutaneous pigmentation plus hamartomatous upper GI polyps and personal or family history of PJS associated cancer. This is in accordance with 73–94% *STK11* pathogenic mutation detection rate in series of studies which performed comprehensive *STK11* sequencing and LGR analysis in cases fulfilling the established PJS diagnostic criteria [69, 99, 100, 112]. As expected, the detection rate of

pathogenic mutation in STK11 was only 3/14 (21.5%) in cases with suspected PJS in whom all the established diagnostic criteria of PJS were not fulfilled. These three carriers of *STK11* pathogenic mutation were categorized as suspected PJS due to lack of documentation of hamartomatous polyps in two breast cancer patients with characteristic mucocutaneous pigmentation (PJ6 and PJ7) and absence of mucocutaneous pigmentation in one case with ovarian gonadoblastoma and hamartomatous polyps (PJ5). Hence, in cases with a PJS associated cancer, even if one of the two syndromic features is evident, *STK11* testing is warranted.

Interestingly, the breast cancer histology and IHC in all the four carriers of *STK11* pathogenic mutations (PJ1, PJ6, PJ7, PJ20) and one *STK11* obligate carrier (PJ2a) was nearly identical—infiltrating duct carcinoma grade III in four cases and grade II in one case and positive for oestrogen receptor (ER) in all five cases, progesterone receptor (PR)-positive in three of five cases and Her2-negative in all four cases in whom this information was available. Breast cancer is the second most common cancer in PJS with a risk of 45% at the age of 70 years [109]. Surprisingly, the ER/PR/Her2 status despite being the most important prognostic factor having a direct correlation with the underlying molecular subtype has not been described for *STK11* germline pathogenic mutation carriers. Germline pathogenic mutations in specific genes produce distinct breast cancer phenotype. While germline BRCA1 pathogenic mutations are associated with triple-negative breast cancer, the BRCA2-associated breast cancers are often ER-positive and Her2-negative [113]. In a large cohort of over

10 000 women with triple-negative breast cancer evaluated for germline pathogenic mutations with multigene Next Generation Sequencing (NGS) panels, which included *STK11*, pathogenic mutations were identified in 22 genes including *BRCA1*, *BRCA2*,

TP53 and several other genes but not in the *STK11* gene [114]. This inverse association between TNBC and *STK11* germline pathogenic mutations derived from a very large TNBC cohort supports the finding of our study where all the four cases were ER-positive. Strong association between breast cancer ER positivity and HER2 negativity if confirmed in a larger pooled cohort of *STK11* pathogenic mutation carriers would firmly establish the *STK11* breast cancer phenotype and facilitate their genetic counselling and risk management advice. The NCCN guidelines (https://www.nccn.org/) recommend breast screening for women with *STK11* pathogenic mutation but do not discuss options of chemoprevention or risk reducing salpingo-oopherectomy (RRSO). If the strong association between *STK11* pathogenic mutation and ER positive status of breast cancer is confirmed, similar to the carriers of pathogenic mutation in *BRCA2* gene, women with *STK11* pathogenic mutations may be offered the option of tamoxifen chemoprevention or RRSO. RRSO would not only reduce their risk of gynaecological malignancies but also lower the breast cancer risk.

Several reasons have been put forth regarding the absence of detectable *STK11* germline pathogenic mutations even after comprehensive sequencing and LGR analysis. The most important cause may be syndromic misclassification due to incomplete information or overlapping features with other syndromes as discussed in section 3.1.3 on differential diagnosis of PJS. Presence of only one and sometimes both characteristic features of PJS may therefore be considered as suggestive of a PJS syndromic diagnosis and needs confirmatory genetic analysis to identify *STK11* pathogenic mutation. Negative genetic testing for *STK11* in cases with a distinct possibility of a benign syndrome like LHS would help these individuals as lifelong screening for PJS-associated cancers can be avoided for them. The identification of a

germline *BRCA1* pathogenic mutation in one of our cases (PJ19) with triple-negative breast cancer and PJS-like mucocutaneous pigmentation the importance of making syndromic differential diagnosis and genetic screening of *BRCA1*/2 in suspected PJS patients with breast cancer and MMR genes in suspected PJS patients with GI cancer if no pathogenic mutation in *STK11* is identified.

Classic PJS cases without a detectable *STK11* pathogenic mutation may also be explained by genetic heterogeneity with other genetic loci that may produce the PJS phenotype. Linkage between PJS and a second locus at 19q13.4 has been reported [111, 115]. To investigate the possibility of chromosome 19q13.4 as the second locus, a study [116] screened eight *STK11* mutation negative PJS families for genetic alterations in 4MARK genes that are present at this locus and are also a part of the AMPK-related kinase family of which *LKB1*/*STK11is* an upstream activator [81]. No pathogenic mutation was found in any of these genes in eight cases. Germline exome sequencing in four *STK11* mutation negative PJS cases in two studies, identified several mutations with possible role but they have to be characterized to determine their causative role [117, 118].

Somatic mosaicism was reported in four of the 300 PJS families in whom germline *STK11* pathogenic mutation was not identified earlier [119]. A recent study reported mosaicism in a female negative for the pathogenic *STK11* mutation that was found in her son who developed PJS-associated cancer. The mother was later found to have mosaicism with the mutant allele being detected in the DNA extracted from her gametes and GI tract but not from the blood [120]. The possibility of *STK11* inactivation through germline pathogenic mutations in the promoter region of *STK11* has also been ruled out in a study cohort of 33 *STK11*-negative PJS cases [121].

The pathogenic mutations identified in our cohort including frameshift, missense and LGRs are scattered from exon 2 to 7, which encodes the *STK11* functional kinase domain and is in concordance with the published literature [68, 122]. Of the 18/20 PJS probands with muco-cutaneous pigmentation over lips and adjoining buccal mucosa, 5 cases also had pigmentation over palms, finger tips, breast and upper back. However, we could not identify any difference in the pattern or extent of pigmentation between carriers and non-carriers. This is concordance with one of the largest systematic review of 1644 PJS cases where no difference in cancer risk with the type and site of pathogenic *STK11* mutations was identified [109].

In summary, we report the first comprehensive study on an Indian PJS cohort including 13 carriers/obligate carriers of *STK11* mutations. With the identification of several novel *STK11* mutations, this study has expanded the spectrum of pathogenic mutation in *STK11* gene and highlights the need for studying different populations across the world and pooling of genotype phenotype data. This is the first study to investigate and identify the association between germline *STK11* pathogenic mutation and ER and Her2 status of PJS-associated breast cancer. This needs to be confirmed in a larger pooled cohort of breast cancers in women with germline *STK11* pathogenic mutation as this knowledge can help devise guidelines for better risk management of female carriers of *STK11* pathogenic mutation with RRSO and/or chemoprevention.

CHAPTER 4

LYNCH SYNDROME

4.1 Introduction

Lynch syndrome (LS), also known as hereditary non-polyposis colon cancer (HNPCC), is an autosomal dominantly inherited cancer predisposition syndrome that accounts for approximately 2-5 % of all CRC cases [123, 124]. It occurs at an early age (around 45 years) with multiple generations of the family usually affected with CRC. Lynch syndrome patients develop tumours predominantly in the proximal colon and show a propensity for multiplicity (synchronous and metachronous tumours). In addition, they have an increased risk for extracolonic cancers, such as endometrial, ovarian, gastric, and pancreatic cancers [125, 126]. The penetrance of this syndrome is high; the lifetime risk of developing CRC is up to 80% in men and 50% in women. In addition, women have approximately a 40-60% risk of developing endometrial cancer [127, 128].

4.1.1 History of Lynch syndrome

Reportedly inspired by his fatalistic seamstress, Aldred Warthin, from the University of Michigan, first identified this multi-cancer syndrome in mid-western USA in 1895. In 1913 he published the first known case report of her family's pedigree including multiple cases of gynaecological and gastrointestinal cancers in the absence of polyposis under the name of Family 'G' [129]. Pedigree of Family 'G' remains one of the longest family cancer histories ever recorded (Pedigree shown in Fig 4.1). It was rediscovered by Henry Lynch in 1966, who reported two families from mid-west USA (Families N and M) whose members had very similar spectrum of tumours to Family G's so he argued for the recognition of a new syndrome referred to as autosomal dominant cancer family syndrome (CFS) [130]. The disease was initially called "family cancer syndrome" by Lynch who later coined the term HNPCC, since

associated polyposis is rare [131]. Lynch with the help of James French; then Chairman of the Department of Pathology at the University of Michigan and Anne Krush, a medical social worker conducted a detailed investigation of Family 'G' with phenotypic characterization of over 650 family members. This syndrome's salient features includes: (1) increased incidence of adenocarcinomas, mainly of the colon and endometrium; (2) increased risk for multiple tumours; (3) autosomal dominant inheritance; and (4) early onset of cancer [132].

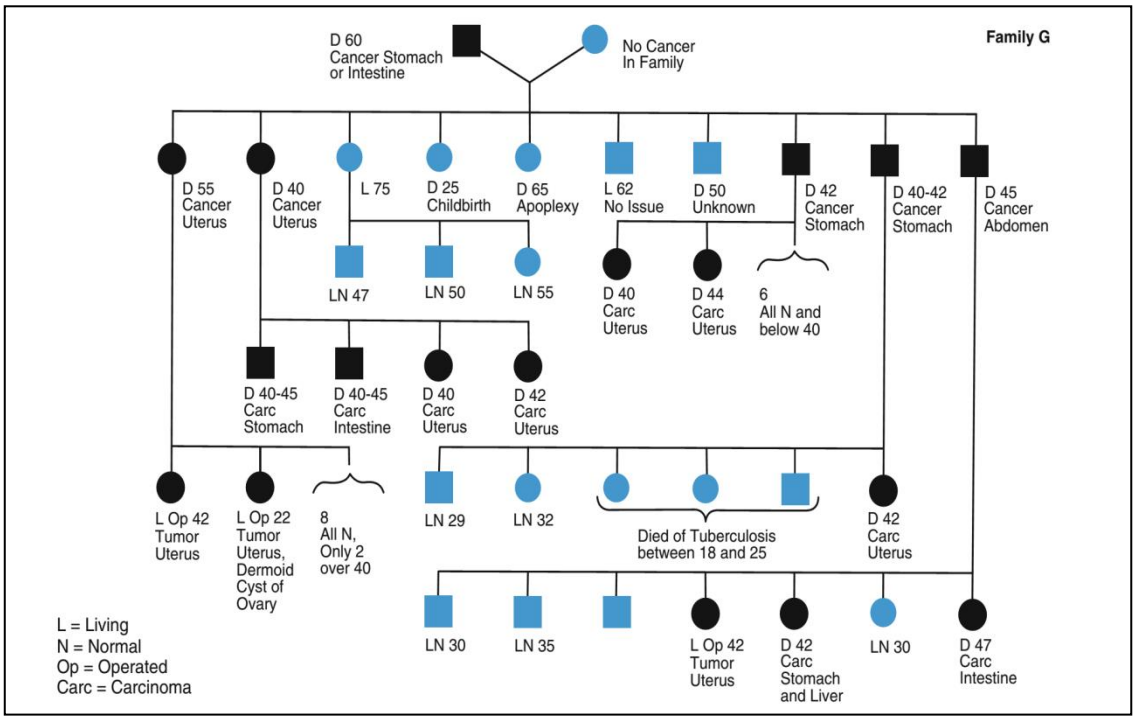


Fig 4.1 Pedigree of Family 'G' (Taken from Ref. [129])

It was not until 1984 that this syndrome was coined as Lynch syndrome [133]. In 1991, the term hereditary non-polyposis colorectal cancer (HNPCC) was forged by an international collaborative group of researchers to distinguish them from FAP [134]. Due to prevalence of extra-colonic cancers with endometrial cancer being the sentinel cancer in affected women [135], Lynch Syndrome (LS) is now favoured over HNPCC.

4.2 Clinical Features of Lynch Syndrome

Despite its heterogeneous nature, LS has a clinical signature that eases its identification. Affected individuals generally develop at young age with the median age of first tumour diagnosis being 45years, 24 years earlier than the general population. LS colorectal tumours are predominantly right sided. LS patients have an elevated risk of synchronous and metachronous cancers.

At the histological level, colorectal cancers (CRCs) generally are poorly differentiated, have a mucinous component, with signet ring cells, tumour infiltrating lymphocytes (TILs) and intense lymphocytic reaction (Crohn's-like) [136]. CRC in LS patients are associated with a better prognosis than that of sporadic cases despite of aggressive histological features of most LS carcinomas [137] [138].

Furthermore, although LS individuals form less adenoma and develop at an older age than FAP cases, precursor lesions evolve to malignancy in a highly accelerated manner, with the adenoma-carcinoma sequences lasting less than 3 years, in contrast with CRCs from sporadic origin where usually carcinoma arises after 7-10 years [131, 139, 140].

4.2.1 Lynch syndrome spectrum of cancer

Individuals diagnosed with Lynch syndrome have a 70 to 85% risk of developing CRC by the age of 65 years and an elevated risk of developing a second primary. The most common extra-colonic cancers in the Lynch syndrome spectrum is endometrial cancers in females. The lifetime risk varies in the range of 30-40% risk depending upon the MMR gene mutated [141]. Other LS-associated cancers such as small bowel cancer are considered rare with lower lifetime risk (1-4%), however it is significant compared to only about 0.01% lifetime risk in the general population [142]. There is

an increased risk of extracolonic cancers like urothelial, renal, biliary, pancreatic, breast, prostate, and rare adrenocortical tumours in LS-associated MMR-mutation positive individuals have also been reported and therefore they be considered as part of LS spectrum [142].

4.2.2 Clinical criteria for diagnosis of Lynch syndrome

Identification of Lynch syndrome patients and families has significant effect on their clinical management and may impact the surgical approach, recurrence of cancer surveillance and screening for extracolonic malignancies. In 1990 the first diagnostic guidelines called as the Amsterdam Criteria (AC-I) was established by a panel of experts (the International Collaborative Group on Hereditary Non-polyposis Colorectal Cancer – ICG-HNPCC) for recruiting HNPCC patients for collaborative studies [134]. The initial criteria was restricted to common characteristics of LS-associated CRCs (i.e. young age of onset and family history), therefore was revised (AC-II) to incorporate extra-colonic cancers [143]. Both these criteria imply special familial aggregation and identify around 60% to 80% of LS patients; however they are too stringent for clinical use [134, 144, 145]. Therefore in 1997, the National Cancer Institute established the Bethesda Guidelines to aid in the identification of LS in cases not identified by the Amsterdam Criteria. The criteria are based on the clinicopathological aspects of LS and were established to direct MSI testing in suspected Lynch-associated colorectal tumours for early diagnosis [146]. In 2004 the Bethesda guidelines were revised in order to broaden the criteria and improve LS detection [147] [148]. These clinico-pathological guidelines are detailed in Table 4.1.

Table 4.1 Clinical criteria for diagnosis of Lynch syndrome

Amsterdam Criteria I (est. 1990) [134]
Individuals must meet ALL the following criteria: 1. At least three relatives with colorectal cancer 2. At least two successive generations should be affected 3. At least one colorectal cancer should be diagnosed <50 years of age 4. One of the individuals should be a first-degree relative to the other two 5. Familial adenomatous polyposis should be excluded 6. The tumours should be verified by pathological examination
Amsterdam Criteria II (est. 1998) [143]
Individuals must meet ALL the following criteria: 1. At least three relatives with a Lynch-associated cancer (colorectal, endometrial, small bowel, ureter/renal pelvis) 2. At least two successive generations should be affected 3. At least one tumour should be diagnosed <50 years of age 4. One of the individuals should be a first-degree relative to the other two 5. Familial adenomatous polyposis cases should be excluded 6. The tumours should be verified by pathological examination
Bethesda Guidelines for testing colorectal tumours for MSI (est. 1997) [146]
Tumours should be tested for MSI in any of the following situations: 1. Cancer in a family that meets Amsterdam Criteria I 2. Two Lynch-associated cancers, including synchronous and metachronous colorectal cancers or extracolonic cancers (endometrial, ovarian, gastric, eratoacantho, small bowel cancer, transitional cell carcinoma of the renal pelvis/ureter) 3. Colorectal cancer and a first-degree relative with colorectal cancer and/or Lynch-associated extracolonic cancer and/or a colorectal adenoma; one of the cancers diagnosed <45 years of age, and the adenoma diagnosed <40 years of age 4. Colorectal cancer or endometrial cancer diagnosed <45 years of age 5. Right-sided colorectal cancer with an undifferentiated (solid/cribriform) tumour histology, diagnosed <45 years of age 6. Signet-ring-cell-type colorectal cancer diagnosed <45 years of age 7. Adenomas diagnosed <40 years of age

Revised Bethesda Guidelines for testing colorectal tumours for MSI (est. 2004) [147]
Tumours should be tested for MSI in any of the following situations: 1. Colorectal cancer diagnosed <50 years of age 2. Presence of synchronous/metachronous colorectal, or other Lynch-associated tumours (colorectal, endometrial, gastric, ovarian, pancreas, small bowel, ureter and renal pelvis, hepatobiliary tract, glioblastoma in Turcot syndrome, sebaceous adenomas and eratoacanthomas in Muir-Torre syndrome) regardless of age 3. Colorectal cancer with the MSI-H histology diagnosed <60 years of age 4. Colorectal cancer diagnosed in one or more first-degree relatives with a Lynch-associated tumour, with one of the cancers diagnosed <50 years of age 5. Colorectal cancer diagnosed in two or more first- or second-degree relatives with Lynch associated tumours, regardless of age

4.2.3 Microsatellite Instability: Hallmark of Lynch Syndrome tumours

In 1992, Manuel Perucho and colleagues carried out PCR amplification using DNA extracted from colon cancer and matched normal tissue samples using random primers [149]. In 1993 the data was published where they reported that 12 % of the tumours had bands that were shorter in length compared to their normal counterparts. Analysis of the sequences from these bands revealed the presence of simple repetitive elements (i.e., microsatellites), primarily in polyadenine (A_n) tracts associated with Alu sequences.

Concurrently, work from Stephen Thibodeau's laboratory identified deletion mutations in $[CA]_n$ sequences in chromosomes 5q, 15q, 17p, and 18q in colorectal tumours and coined the term microsatellite instability [150]. They detected MSI in 28% of colorectal tumour similar to Perucho's findings and showed that these alterations were heterogeneous in different tumours. Both the above work recognized

that microsatellite instability could be a unique pathway to CRC development in absence of chromosomal instability.

Microsatellites are repeated-sequence motifs, consisting of simple mono-, di-, tri and tetra-nucleotide DNA repeats, found all across the genome in large numbers [151]. Mismatch repair deficiency causes microsatellite instability, which is characterized by clonal global deletion or duplication of microsatellite repeat sequences. Polymerase errors are relatively common in microsatellites resulting in insertion or deletion of nucleotides [152]. Failure of DNA mismatch repair protein activity in LS, results in the accumulation of these errors especially in these repetitive sequences, therefore producing microsatellite instability. It is seen that MSI results in increased mutation rate in the order of 100 to 1,000 fold [153-155]. This high mutation rate increases the probability of other tumour suppressor genes or oncogenes to be also affected [156]. LS tumours have a hypermutated genome due to overall increase in mutation burden and exhibit a so called, “mutator phenotype”. The process of microsatellite instability is represented in Fig 4.2

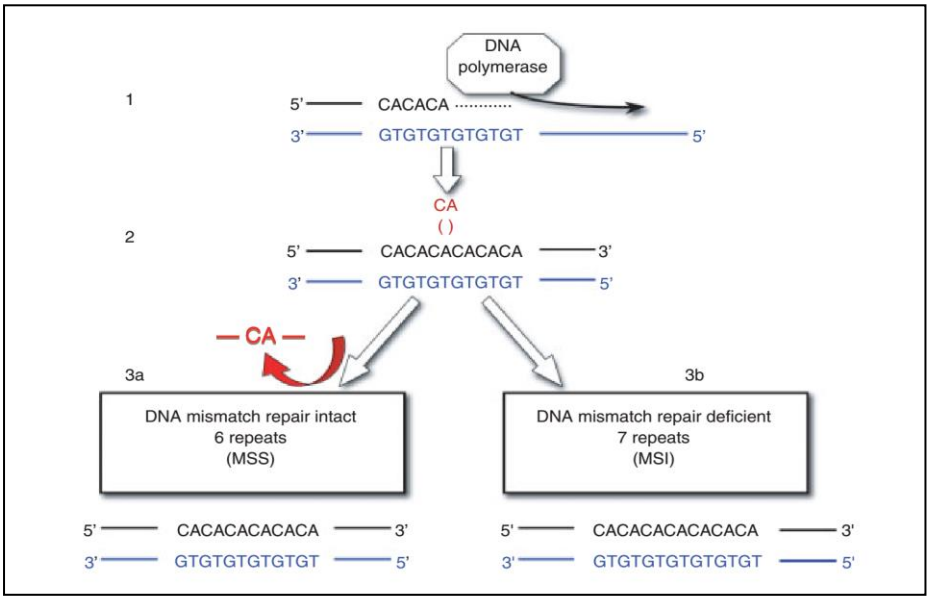


Fig 4.2 Mechanism of Microsatellite Instability (Taken from Ref [157]

More than 80% of Lynch syndrome associated tumours display MSI and is therefore regarded as the hallmark of LS tumours. MSI testing is therefore used as one of the diagnostic criteria for identifying LS cases and a pre-screening technique for genetic testing in these cases. A panel consisting of three dinucleotide repeats (D2S123, D5S346, D17S250) and two mononucleotide repeats (BAT26, BAT25) known as the Bethesda panel is used as a standard test for MSI as recommended by National Cancer Institute in 1997 [158]. If two or more of the five marker show instability, the tumour is MSI-H; if only one marker is unstable, the tumour is MSI-L; and if all five markers are stable, the tumour is MSS. Numerous published reports showed that mononucleotide markers were superior to the dinucleotide markers for assessment of MSI as the changes in repeat length are longer in mononucleotides and therefore easier to interpret. It soon became apparent that mononucleotide markers are quasimonomorphic markers that in principle obviate the need to study unaffected or normal tissue for MSI testing [159, 160]. In 2002, a comprehensive study which was undertaken to improve the sensitivity of MSI testing proposed a pentaplex marker panel of 5 quasi-monomorphic mononucleotide marker (BAT25, BAT26, NR-21, NR-22, and NR-24) [159], By consensus, microsatellite status has been divided into three groups: MSI-H would be called if 40% or more of the markers tested were unstable, low-level instability (MSI-L) with less than 40% instability and microsatellite stable (MSS), with no instability seen. A description of Bethesda Panel and pentaplex panel is given in Table 4.2.

Table 4.2 Microsatellite markers used for MSI testing

Panel	Marker	MS repeat	Location
Bethesda Panel	BAT25	A (25)	Chromosome 4, Intron 16 of c-kit
	BAT26	A (26)	Chromosome 2, Intron 5 of MSH2
	D2S123	CA (n)	Chromosome 2, Linked to MSH2
	D5S346	CA (n)	Chromosome 5, Linked to APC
	D17S250	CA (n)	Chromosome 17, Linked to P53
Pentaplex Panel	BAT25	A (25)	Chromosome 4, Intron 16 of c-kit
	BAT26	A (26)	Chromosome 2, Intron 5 of MSH2
	NR21	T (21)	Chromosome 14, 5'UTR of SLC7A8
	NR24	T (24)	Chromosome 2, 5'UTR of ZNF2
	NR27	A (27)	Chromosome 11, 5'UTR of IAP-2

4.2.4 Immunohistochemical analysis as a prescreening tool

Immunohistochemistry (IHC) has been reported to have a good sensitivity (>90%) as loss of tumour MMR protein expression is reported to be evident in significant number of cases with pathogenic MMR gene mutations and shows greater than 90-95% correlation with the MSI-H phenotype [161]. IHC has higher sensitivity and specificity over STR based MSI analysis for the molecular screening of LS [162]. IHC is carried out by using a four-antibody panel including MLH1, MSH2, MSH6, and PMS2 which is used to test for loss of protein expression. MMR protein stability is based on its in-vivo heterodimerization and can explain the staining pattern observed in MMR IHC. When MLH1 protein is defective there is loss of both MLH1 and PMS2, because PMS2 stability depends on its ability to form a complex with MLH1.

Similarly, there is MSH2 & MSH6 loss when MSH2 is defective. The opposite, however, does not apply, because tumours with defective PMS2 or MSH6 may maintain expression of MLH1 or MSH2, respectively. A tumour with loss of expression of MLH1/PMS2 may be either due to promoter methylation or germline mutation in MLH1. IHC of each MMR protein has a sensitivity of 74%, 91%, 55% and 77% for the detection of mutation in *MLH1*, *MSH2*, *MSH6* and *PMS2* respectively [163, 164]. The specificity varies from 80-100% depending upon the antibody and the panel used [164-166]. Several reports have proven that MMR proteins IHC provides a faster, cost-effective, sensitive and highly specific screening technique for MSI analysis and subsequent LS diagnosis. However, the debate regarding the methods used to start screening for Lynch syndrome is still ongoing (IHC and MSI) as some studies strongly favours its use demonstrating the diagnosis of a significant fraction of LS cases that would otherwise have been missed [167]. On the contrary, other studies discourage its application by outweighing its advantages over the limitations and challenges associated with the implementation of this concept [163].

4.3 Genetic features of Lynch Syndrome

Lynch Syndrome is caused due to inherited germline mutations in one of four mismatch repair (MMR) genes *MLH1*, *MSH2*, *MSH6*, and *PMS2* [168] (Table 4.3), as well as epimutations in MLH1 and MSH2, the later associated to EPCAM deletions resulting in a defective MMR machinery.

Table 4.3 Characteristics of MMR genes involved in LS

Gene Symbol	Name	Reference number	Chromosomal position	Coding exons	DNA length (bp)	RNA length (bp)	Protein length (aa)	Protein domains
MLH1	MutL homolog 1	NM_00249.3	3p21.3	19	75,557	2,752	756	ATPase domain Interaction domain for MSH2, MSH3, MSH6 Interaction domain for *PMS2*, MLH3, PMS1
MSH2	MutS homolog 2	NM_00251.2	2p21	16	80,259	3,307	934	DNA binding domain Interaction domain for MSH3 and MSH6 Interaction domain for MLH1 and *PMS2*
MSH6	MutS homolog 6	NM_00179.2	2p16	10	23,871	7,476	1360	Helix-turn-helix domain associated with a Walker- A motif (adenone with Mg binding motif) with ATPase activity PCNA binding motif PWWP domain that bind to dsDNA
PMS2	Postmeotic segregation increased 2	NM_00535.5	7q22.1	15	35,886	2,885	862	ATP interaction domain MLH1 binding domain Exonuclease domain

Abbreviations: bp- base pair; aa- amino acid; ATP- adenosine triphosphate; PCNA- proliferating Cell Nuclear Antigen; PWWP- proline-tryptophan-tryptophan-proline; ds- double stranded

4.3.1 DNA Mismatch Repair Pathway

MMR genes are involved in different cellular processes. They modulate DNA recombination, DNA damage signalling, and have a role in apoptosis regulation. However, their most important function is to restore replication fidelity when the polymerase fails. Polymerase errors occur during DNA replication and can happen either by simple mismatches or by strand slippage, that convey small insertions or deletions in the newly synthesized DNA strand [152]. MMR is a well conserved pathway, fundamental to maintain genomic integrity by correcting replication or recombination errors. There are four basic steps to repair a mismatch: a) recognition of a DNA base mismatch which causes distortion in the DNA double helix structure or insertion/deletion loop (IDL – caused by polymerase slippage during replication) by MutSα, which is made up of a heterodimer of *MSH2* and *MSH6*, or MutSβ consisting of a *MSH2* and *MSH3* heterodimer. MutSα recognizes mismatches and single base loops while MutSβ recognizes indels. The function of the two MutS heterodimers is overlapping, and therefore partly redundant. b) Upon recognition of the mismatch, MutSα (or MutSβ) recruits MutLα (*MLH1* and *PMS2* heterodimer) in an ATP dependent reaction [169]. The tetrameric complex slides on the DNA clamp. MutLα acts as a mediator as it interacts with and modulates the activity of several MMR proteins. c) *PMS2* introduces a nick in the daughter strand at the 5' end of the mismatch using its endonuclease activity thereby facilitating the excision of the mismatch-containing strand of DNA in a 5' to 3' direction by exonuclease EXO1. The excision intermediate is protected from nuclease degradation by replication protein A (RPA). d) The excised DNA strand is then resynthesized by DNA polymerase (Pol δ or ε) and DNA ligase in the presence of RPA, proliferating cell

nuclear antigen (PCNA) and the clamp loader replication factor C (RFC) [170-172]. A schematic representation of the mismatch repair pathway is shown in Fig 4.3.

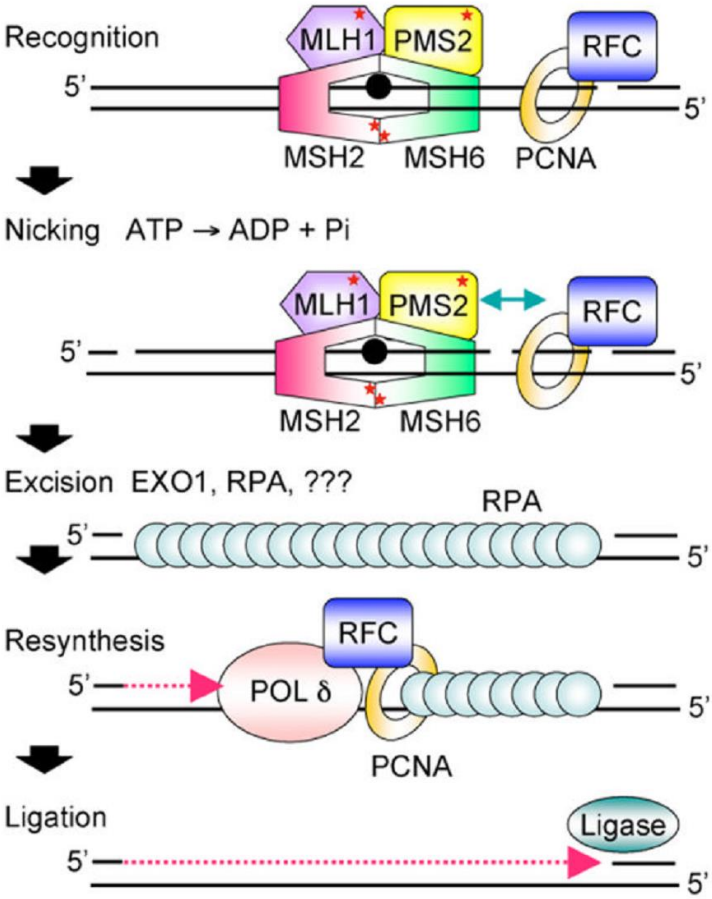


Fig 4.3 Schematic representation of MMR pathway (Taken from Ref [173])

4.3.2 Mutation spectrum of MMR genes

MLH1 and *MSH2* mutations account for more than 70% of LS cases. *MSH6* represents ~18%, *PMS2* represent ~6% and deletion of the 3' end of *EPCAM* (formerly TACSTD1), the gene adjacent to *MSH2*, are responsible for about 1-2% of LS families [174]. The deletions at the end of the *EPCAM* gene lead to epigenetic inactivation of the cis MSH2 allele [175, 176]. *MLH1* and *MSH2* are the major genes owing to their indispensable role in MMR pathway followed by *MSH6* and *PMS2*, the minor genes which are the chief

interacting partners of *MLH1* and *MSH2*. The mutations in *MLH3* are very rare and no LS-associated *MSH3*, *MSH4* and *MSH5* mutations have been reported so far in the literature, as *MSH4* and *MSH5* proteins have no role in mismatch repair, rather they are involved in meiotic recombination [177]. There is currently no evidence regarding the involvement of any other mismatch repair genes in Lynch syndrome tumour development. MMR gene mutation affects all races, although mutation frequencies vary among genders and geographic regions.

LS pathogenic genetic mutations usually are nonsense, frameshift or splicing mutations. Large genomic rearrangements are also causative and have been reported to make nearly a third of MMR gene mutations, especially frequent in *MSH2* due to high content of Alu elements [178-180]. Fig 4.4 shows type of mutation found in MMR genes as per InSiGHT database [181]. Additionally, a rare constitutional *MLH1* epimutation has been reported that leads to promoter hypermethylation and is usually associated with sporadic CRC cases [182, 183].

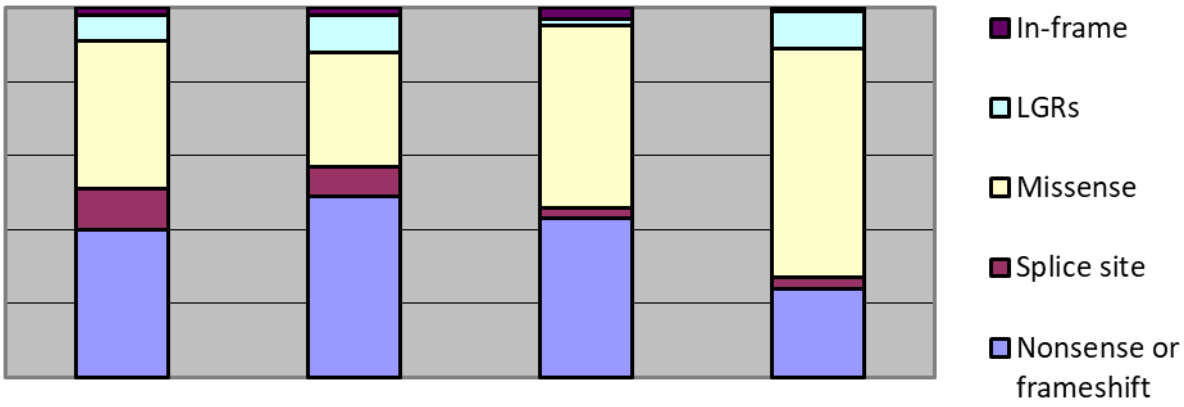


Fig 4.4 Types of mutation in MMR genes (InSiGHT database)

MMR deficiency in tumour tissue occurs due to inactivation of both alleles. The second allele can be inactivated by a variety of mechanisms including, whole gene deletion, gene conversion, and point mutation (the least likely mechanism) [184]. The colon

microenvironment may have mutagens that greatly increase the chance of losing the wild-type allele of the mutated MMR gene, either by somatic mutation or loss of heterozygosity (LOH) [185-187]. MMR gene mutation increases the risk of malignant transformation especially in the gastrointestinal epithelium and endometrium, due to the high proliferation rate of these cells [188].

Usually mutations in MMR genes are scattered throughout the gene with no obvious hotspots and most mutations are unique. Interestingly, there are recurrent mutations that reappear due to genetic circumstances or other factors, like the A>T transversions in a splice site of Intron 5 of *MSH2* gene (c.942+3A>T), explained by the fact that this adenine is the first of 26 adenines in a stretch, creating a hotspot for this particular change, possible by *de novo* mutation produced by polymerase slippage during replication [189]. The same A_{26} repeat is a part of an important MSI marker, BAT26. Moreover, there are also mutations shared by unrelated cases inherited from a common ancestor many generations before, recognized as founder mutations. The likelihood for them to become common is greater in isolated or rapidly grown populations. Based on the extent of haplotype conversion, the age of MMR gene founder mutations may vary from few hundreds to few thousand years [190]. To date around 50 proven founder mutations have been detected in MMR genes, in specific populations and may account for 50% of all LS families in some populations [190].

The International Society for Gastrointestinal Hereditary Tumours (InSiGHT, www.insight-group.org) formed the InSiGHT Colon Cancer Gene Mutation Databases as part of a pilot program with the Human Variome Project (HVP) to collate all inherited alterations affecting CRC susceptibility genes and undertook a massive scale study of MMR mutations to infer the pathogenicity of all the mutations using a 5-tiered

classification system. The Class 5 and Class 4 mutations are the “pathogenic” and “likely pathogenic” mutations. Class 3 represents “mutations of unknown significance (VUS)” whereas the Class 2 and Class 1 signify the “likely benign” and “benign” mutations [174].

4.3.3 Cancer risk of MMR gene mutation carriers

Carriers of a MMR gene mutation have been reported to have an estimated lifetime 40-75% risk of developing cancer depending on the gender and mutated MMR gene [191]. Among the different cancers associated with Lynch, the highest lifetime risk in *MLH1* and *MSH2* mutation carriers is for colorectal cancer followed by endometrial cancer and other extracolonic cancer. The cumulative risk associated with *MLH1*, *MSH2* and *MSH6* mutation carriers to age 70 for CRC ranges between 30-75% for men and 25-50% for women, and between 30-45% for women with endometrial cancer [192-196]. The associated lifetime risk of CRC and endometrial cancer varies depending upon the MMR gene mutated. The estimated colorectal cancer risks in *MLH1* and *MSH2* mutation carriers by age 70 years vary, ranging from 28-74% for males and 23-61% in females, and 18-54% for endometrial cancer [193, 197]. *MSH6* and *PMS2* display a reduced age specific penetrance compared to *MLH1* and *MSH2* mutations which reflects functional redundancy of *MSH6* with *MSH3* and *PMS2* with *MLH3* and PMS1 [181]. The corresponding CRC risks reported for a mutation in *MSH6* are 22-36% and 10-18% for males and females respectively [198, 199] and for *PMS2* mutation carriers the risks are 20% and 15% [200]. *MSH6* mutation carriers show a higher association with endometrial cancer than *MLH1*/*MSH2* mutation carriers and a later age of onset [198, 201-204] . The estimated risk of endometrial cancer by age 70 reported for different studies is between 16 and 49% [128, 192, 198, 205].

Several other cancers are included in the Lynch spectrum such as gastric, small bowel, ovarian, pancreatic, hepatobiliary tract, urothelial (including renal pelvis, ureter and bladder), and brain tumours; and are associated with a more modest penetrance. Table 4.4 summarizes the comparative risks of various cancers in Lynch syndrome and general population with the mean age of onset of each cancer [206]. Breast cancer and prostate cancer are the comparatively newly recognized cancers in Lynch syndrome spectrum with moderate increase associated with MMR gene mutation carriers, however inclusion of these cancers in the syndrome is still a debatable issue [207, 208].

Table 4.4 The comparative risks of various cancers in Lynch syndrome and general population with the mean age of onset of each cancer

Cancer risk by age 70 years					**Mean age of onset (years)**			
Cancer	Population risk	MLH1/ MSH2	MSH6	*PMS2*	Population risk	MLH1/ MSH2	MSH 6	*PMS2*
Colorectal	2%	40-70%	10-22%	15-20%	70	40-61	54	61-66
Endometrial	<2%	35-40%	17-44%	15%	70	47-62	55	50
Ovarian	<1%	4-11%	1-11%	#	63	41-51	46	42
Stomach	0.9%	5-13%	<3%	#	69	55	63	70-78
Pancreatic	1.5%	3.7%	<1%	Not clear	71	-	-	-
Hepatobiliary	0.9%	Upto 4%	NA	#	63	50-57	-	-
Urinary Tract	<1%	Upto 6%	0.7%	#	64	54-60	30-75	-
Small Bowel	0.2%	Upto 6%	NA	#	66	<50	40-73	-
Brain/CNS	0.6%	Upto 3%	<1%	#	57	50	-	-
Sebaceous	<<1%	1-9%	Not clear	Not clear	-	55-62	-	-

The combined risk for ovary, stomach, hepatobiliary, urinary tract, small bowel and brain is 6%

Muir-Torre syndrome and Turcot's syndrome are rare variants of LS. Muir-Torre is characterized by the presence of multiple benign/malignant skin lesions such as sebaceous carcinomas and keratoacanthomas along with with other LS-associated tumours [209, 210]. Turcot's syndrome, a rare syndrome is associated with FAP or LS and may also be included in constitutional mismatch repair deficiency syndrome (CMMR-D). It is characterized by association of colonic polyps and brain tumours, such as glioblastoma/astrocytoma, and CRC [211, 212]. CMMRD is caused by bi-allelic MMR gene germline mutations and is detailed in Chapter 5. Brain tumours are less common in heterozygote mutation carriers and predominantly occur in MSH2 mutation

carriers [197, 213]. It has even been suggested that the original Turcot's syndrome cases (two siblings with multiple adenomatous polyps, CRC and brain tumours) could have been CMMRD cases [211] [214]. It is worth mentioning that Turcot syndrome could also be caused by *APC* mutations, when so, affected cases develop different tumour features from the MMR gene mutated tumours, with a special predisposition to manifest polyps [212]. The LS tumour spectrum is quite wide and some rare tumours like leiomyosarcoma, adrenocortical carcinoma and malignant fibrous histiocytoma are seen more frequently in LS than the general population [215-217].

4.4 Genotype-phenotype correlation in Lynch syndrome

Several studies have reported a difference in the risk of developing colorectal cancer and other extracolonic cancers among different MMR gene mutation carriers and also among carriers of different types of MMR gene mutation [218]. Difference in age of onset and severity of disease has also been noted depending upon which MMR gene is mutated [218]. *MLH1* or *MSH2* mutation carriers typically develop the LS characteristics which 'fulfils the Amsterdam I Criteria with a mean age of CRC onset of 43–46 years and with tumours exhibiting MSI, although *MSH2* mutation carriers have a higher risk of developing extracolonic tumours [219, 220]. In contrast, *MSH6* or *PMS2* mutation carriers develop atypical LS phenotype, with *MSH6* mutation carrier females at a higher risk of endometrial cancer than CRC with mean age of onset above 50 years [204, 221]. Moreover, these tumours do not consistently exhibit MSI expect at the mononucleotide repeats [222]. *PMS2* mutation carrier usually develop CRC at a later age of onset and sometimes in the absence of family history, however these tumours do exhibit MSI [200, 223, 224]. The reduced penetrance and later age of onset of various cancers associated

with *MSH6* and *PMS2* mutation as compared to *MLH1* and *MSH2* mutations may partly be explained by the functional redundancy of *MSH6* and *PMS2* with *MSH3* and *MLH3* as opposed to the obligate partners, *MLH1* and *MSH2* whose loss results in the destabilization of their respective partners [225]. LS patients with *EPCAM* deletions exhibit similar risk of developing CRC like *MSH2* mutations carriers however show a reduced risk of extracolonic cancers [226]. If *EPCAM* deletions extend as far as *MSH2* promoter, it confers a higher risk of endometrial cancers in female carriers [227]. MLH1 epimutations show similar phenotypic features like *MLH1* gene mutation carriers [228]. No clear correlation has been observed between the type or location of mutation and clinical phenotype of LS [181]. While monoallelic MMR gene mutation causes LS, homozygous or biallelic mutations in MMR genes are associated with a more severe cancer phenotype, known as constitutional mismatch repair deficiency syndrome (CMMRD). Individuals with biallelic MMR gene mutation are associated with childhood haematological malignancies, brain tumours and LS-associated tumours with features of neurofibromatosis from infancy or young adulthood. CMMRD syndrome is discussed in detail in Chapter 5.

LS phenotype exhibits heterogeneity among identical germline MMR gene mutation carriers also which suggests that it could be influenced by additional factors. For example, Peltomäki et al. reported a variable age of cancer onset among Finnish families carrying the same founder mutations of *MLH1* [220]. Another study showed that tumour morphology and IHC expression of *β-catenin* (the nuclear localization of which is a hallmark of cancers driven by chronic activation of the WNT signaling pathway) varied extensively within families and even between synchronous or metachronous CRCs from the same individual [229]. Proximal CRCs more frequently exhibit poor tumour

differentiation, expanding growth pattern and increased tumour infiltrating lymphocytes, whereas distal CRCs often lack these distinct LS-associated morphological features [225]. Heterogeneity in clinical phenotype and gene mutation frequencies is evident among different populations For example; endometrial cancer is the most common extracolonic LS cancer in western countries whereas stomach cancer is more frequent in Southeast Asia [230]. It is therefore important to take into consideration these specificities as they may affect clinical diagnosis criteria, warranted genetic testing and the follow-up for affected patients and healthy mutation carriers. Therefore, it is imperative to study genotype-phenotype in each population as it could provide more specific surveillance program focused on the individualized risk.

4.5 Indian Scenario of LS and objective of this study

Lynch syndrome being the most commonly inherited colorectal cancer syndrome worldwide has been extensively studied with huge amount of data on its clinical and molecular aspects. However, most of these data comes from the Caucasian population [231, 232], with lack of knowledge from the Asian population, particularly South Asian. So far only 3 studies have been reported, describing a very small number of 28 Indian LS families with MMR gene mutations [233-235]. Of these 3 studies, the major study on 48 cases used a comprehensive genetic screening approach (Sanger sequencing combined with MLPA) of the LS cases which identified MMR germline mutations in 24 families [233]. The other two studies includes a case study on an extended Indian family with R659X mutation in *MLH1* gene [234] and a report of *MLH1* and *MSH2* mutations identified in 3 families after preliminary screening for MSI in 31 individuals [235]. Therefor one of the objectives of this study was to characterize the mutation spectrum of

MMR genes in a large Indian cohort and to study their genotype-phenotype correlation in detail.

The families were registered in the Cancer Genetics Clinic after making syndromic diagnosis of Lynch syndrome based on the family history and other clinical features as per the medical records available. Majority of the cases in this study was preselected based on their IHC results. The cases showing MMR deficiency were taken on priority; though some cases with classical family history were also taken up irrespective of their IHC status. The methodology used for genetic screening of MMR genes was to amplify the entire coding and flanking intronic regions of the MMR gene by PCR as guided by IHC. In case of unavailability of IHC results, *MLH1* gene was first screened followed by *MSH2* and *MSH6*. PCR products were subjected to Sanger sequencing in order to identify the germline mutation.

4.6 Results

4.6.1 Clinical Characteristics of MMR mutation carrier families

This study was conducted on 232 individuals from 91 unrelated suspected Lynch syndrome families. This cohort included patients who fulfilled criteria defined in accordance with the Amsterdam or Bethesda guidelines. Informed consent, family history up to 3 generations at least and detailed clinical features were obtained from these 91 families of which 13 meet the Amsterdam criteria and 78 families meet the Bethesda guidelines. This cohort represents patients from all parts of the Indian subcontinent and belonging to Hindu, Muslim, Christian, Sikh, Jain and Sindhi religions. Of the 91 probands, 45 were males and 46 were females. Of the 91 probands, 79 had CRC of which 9 had metachronous CRC, 11 had CRC along with extracolonic cancer, 6

had synchronous CRC and remaining 12 had only extracolonic cancers. Classical Lynch syndrome cancers were noted in 64 families and the remaining 27 probands with no or unrelated family history were included in the study because they met the Revised Bethesda guidelines and tumour of 27/28 patients showed MMR deficiency on IHC. A representative pedigree of a classical LS family is shown in Fig 4.5. The suspected cases could be due to a *de-novo* mutation in a MMR gene (representative pedigree shown in Fig 4.6). Comprehensive analysis was carried out in these families using a combination of Sanger sequencing and MLPA approach for genetic analysis, as outlined in Fig. 4.7.

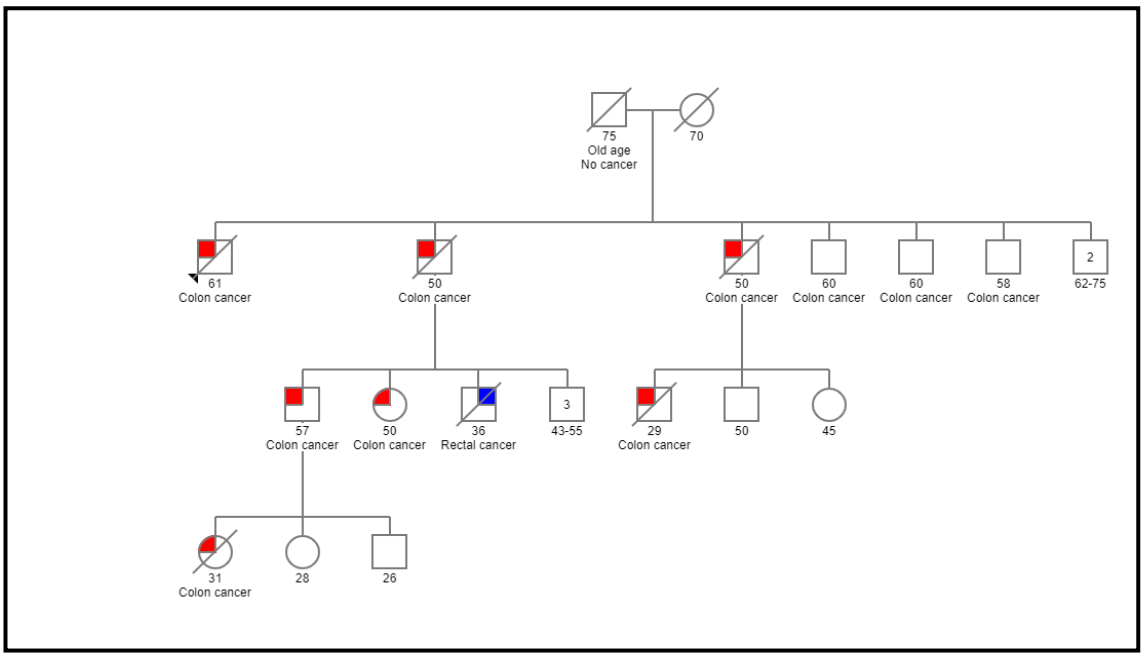


Fig 4.5 Classical Lynch syndrome family

This is a Hindu Brahmin family from Madhya Pradesh. A total of 11 members across 3 generations were affected with colon cancer between the ages of 29-61 years. Deleterious germline mutation in MLH1 gene was identified in the proband followed by testing of 4 at-risk individuals for the family specific mutation (FSM). Out of these 4 family members, two were tested negative for the mutation while the other two were positive for FSM and were unaffected so far and have been kept under surveillance

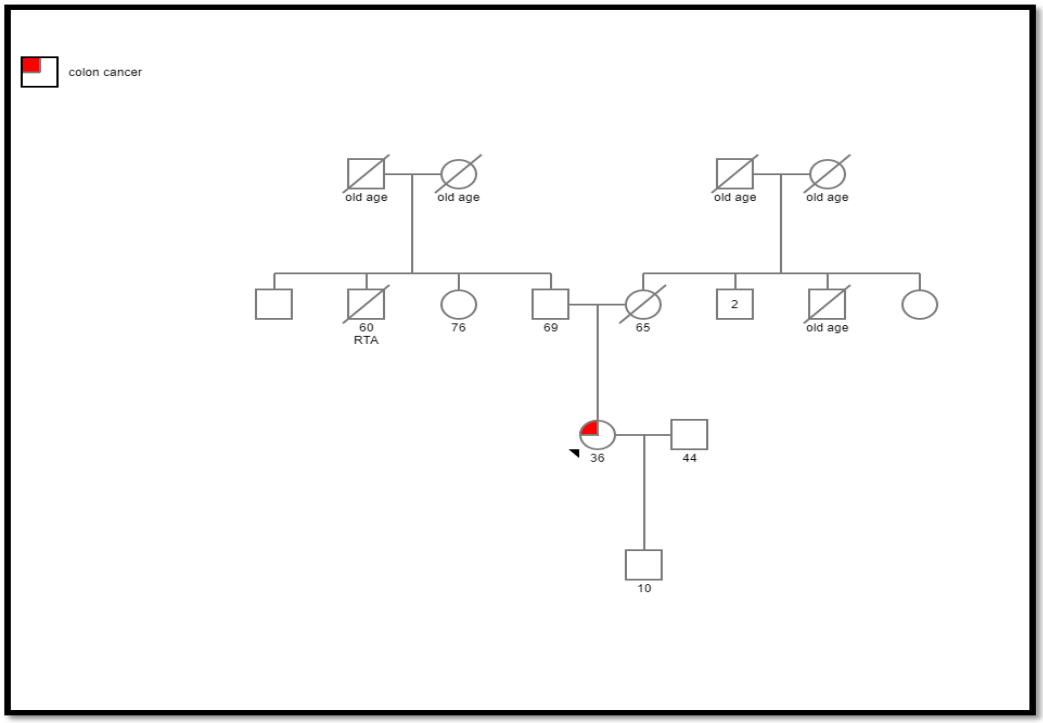


Fig 4.6 Suspected de novo Lynch Syndrome case

The proband is a 30 year old Hindu female from Maharashtra who was diagnosed with moderately differentiated adenocarcinoma of transverse colon. The tumour also showed loss of expression of MLH1/*PMS2* proteins on IHC. She reported no history of cancers in her family but the early age of onset and the clinic-pathological features of tumours were suggestive of Lynch syndrome and therefore genetic testing was undertaken in this case which identified a deleterious frameshift mutation in *MLH1* gene (*de novo* mutation).

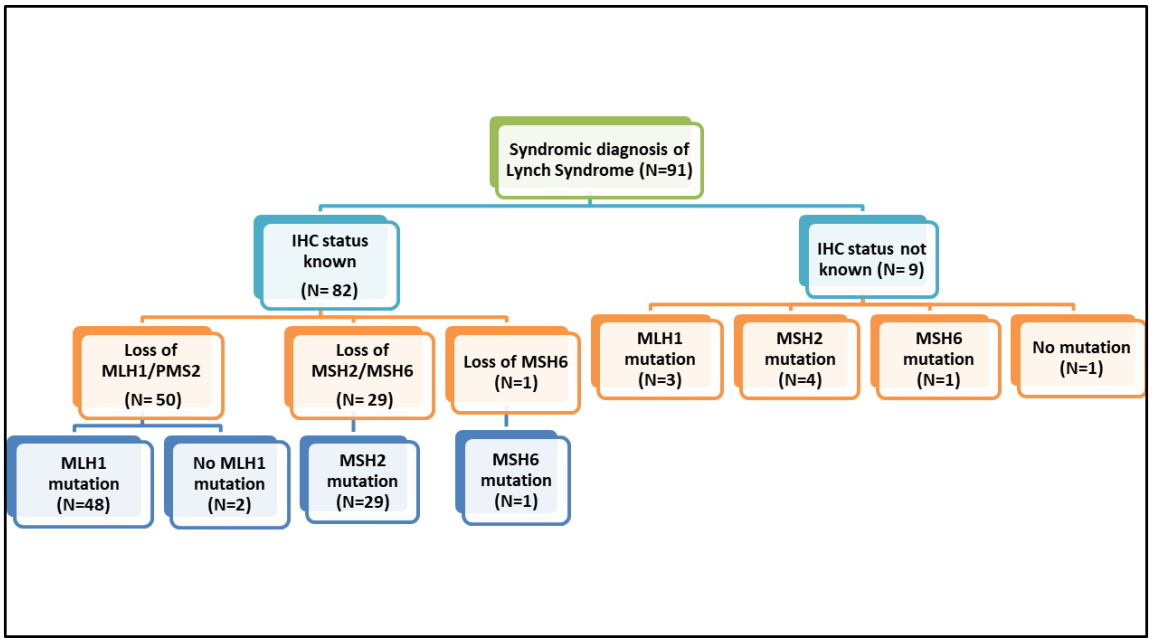


Fig 4.7 Summary of the genetic analysis approach and results in Indian LS patients

4.6.2 Mutation spectrum of MMR genes

Genetic analysis revealed identification of MMR gene mutation in 88/91 families resulting in a mutation detection rate of 96.7% in our cohort. 68/88 was distinct deleterious germline mutations, of which 38 were in *MLH1* gene (53 families), 28 were in *MSH2* gene (33 families) and 2 were in *MSH6* gene (2 families). *MLH1* and *MSH2* mutations accounted for 90% of all the mutations identified in LS families which in concordance with the available literature. Most *MLH1* and *MSH2* mutations are truncating (including both frameshift and nonsense mutations) with splice site mutations being second most common type of mutation in *MLH1* while there is a preponderance of Large Genomic Rearrangement (LGRs) in MSH2. The missense changes, which lead to single amino acid substitutions, included in this study which accounted for 12% of all mutations were Class 5 pathogenic mutation as per InSiGHT database (Fig 4.8)

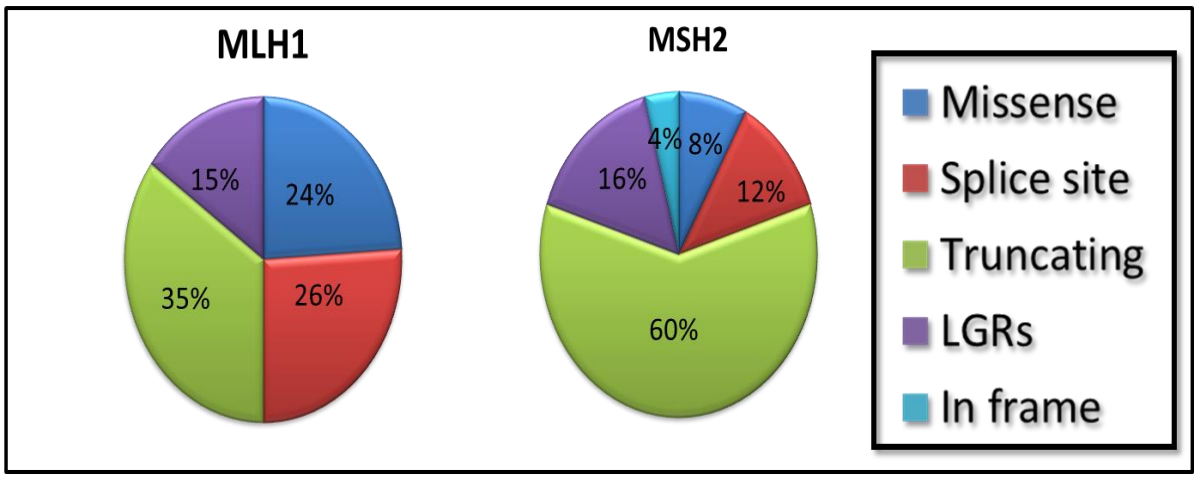


Fig 4.8 Distributions of the types of germline mutations across MLH1 and MSH2

A total of 24 novel mutations were identified, of which 15 were in *MLH1*, 8 in *MSH2* and 1 in *MSH6* gene. *MLH1* mutations scattered throughout the gene while *MSH2* mutation clustered in proximal part of the gene encoding the DNA binding domain and *MSH6*/*MSH3* interaction domain. 7 mutations were recurrent in *MLH1* i.e. were found in more than 1 family, IVS9+1 G>A (Intron 9), c.306 G>T (Exon 3), c.350C>T (Exon 4), c.793C>T (Exon 10), c.2041G>A (Exon 18), c.1459C>T (Exon 13) and c.1416dupA (Exon 13) (Fig 4.9). All the mutations in *MSH2* gene were identified in one family each; except three mutations, c.942+3A>T (Intron 5) identified in 4 families, c.181C>T (Exon 1) and c.1165C>T (Exon 7) were identified in 2 families each (Fig 4.10). *MLH1* and *MSH2* mutation details are given in Table 4.5 and 4.6.

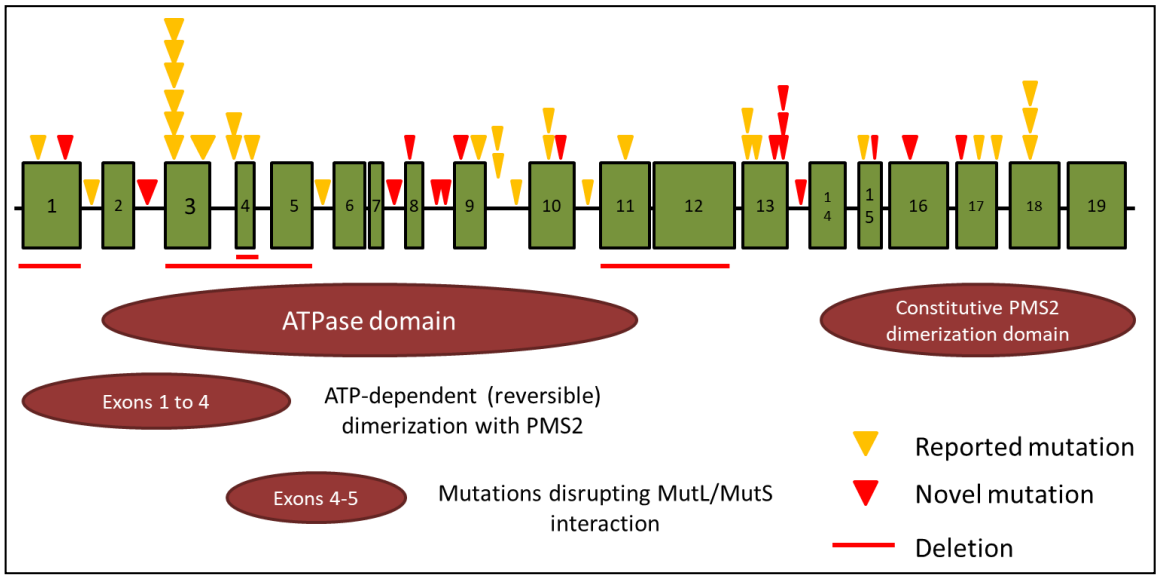


Fig 4.9 Exon wise distribution of mutations in MLH1 gene

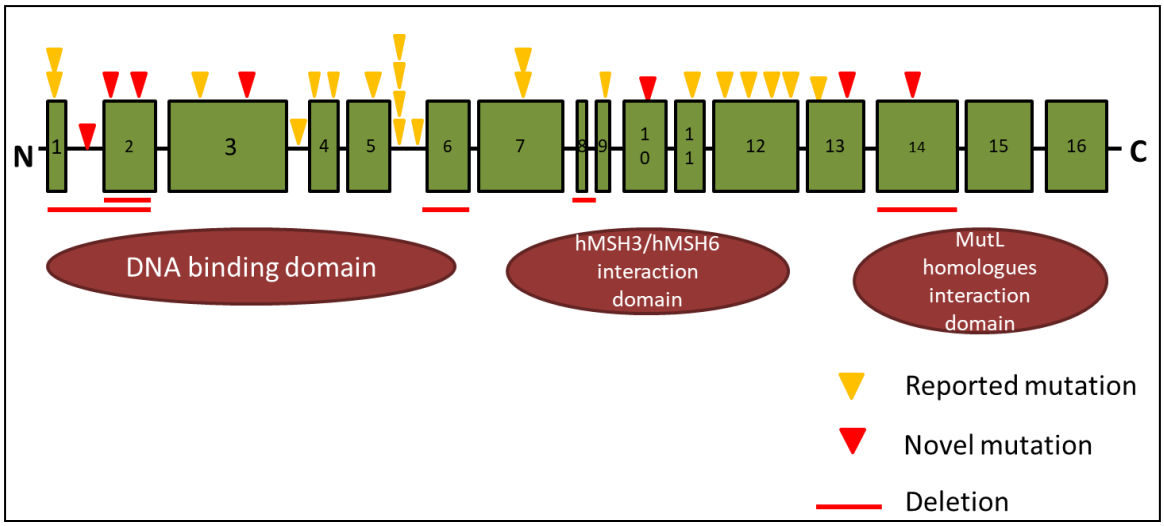


Fig 4.10 Exon wise distribution of mutations in MSH2 gene

Table 4.5 MLH1 germline mutations in Indian LS cohort

Sr. No	Exon	DNA change	Protein change	Type of mutation	No. of families	Reported in InSiGHT/ ClinVar database
1	Exon 1	c.46insG	V16GfsX13	Frameshift	1	Novel
2	Exon 1	c.83C>T	P28L	Missense	1	Reported 37 times
3	Intron 1	c.117-3C>G	-	Splice site	1	Reported once
4	Intron 2	c.208-2 A>C	-	Splice site	1	Novel
5	Exon 3	c.298 C>T	R100X	Nonsense	1	Reported 30 times
6	Exon 3	c.306 G>T	E102D	Missense	6	Reported 18 times
7	Exon 4	c.350C>T	T117M	Missense	2	Reported 132 times
8	Exon 4	c.346delA	T116Qfs*20	Frameshift	1	Reported 2 times
9	Intron 5	c.454-13A>G	-	Splice site	1	Reported 8 times
10	Intron 7	c.589-2A>T	-	Splice site	1	Novel
11	Exon 8	c.676delC	R226Efs*3	Frameshift	1	Novel
12	Intron 8	c.678-3 T>G	-	Splice site	1	Novel
13	Intron 8	c.677+2 T>G	-	Splice site	1	Novel
14	Exon 9	c. 731G>A	G244D	Missense	1	Reported 35 times
15	Exon 9	c.766_776del	C256Tfs*46	Frameshift	1	Novel
16	Intron 9	c.790+1 G>A	-	Splice site	3	Reported 54 times
17	Intron 9	c.790+2dupT	-	Splice site	1	Reported 19 times
18	Exon 10	c.793C>T	R265C	Missense	2	Reported 75 times
19	Exon 10	c.879C>G	Y293X	Nonsense	1	Novel
20	Intron 10	c.884+4 A>G	-	Splice site	1	Reported 7 times
21	Exon 11	c.955G>T	K319X	Nonsense	1	Reported 3 times
22	Exon 13	c.1416dupA	H473Tfs*6	Frameshift	3	Novel
23	Exon 13	c.1459 C>T	R487X	Nonsense	3	Reported 64 times
24	Exon 13	c.1480delT	C494Vfs*14	Frameshift	1	Novel
25	Exon 13	c.1491delG	R498Efs*10	Frameshift	1	Reported 4 times
26	Intron 13	c.1558+2insG	-	Splice site	1	Novel
27	Exon 15	c.1719delT	L574Sfs*15	Frameshift	1	Novel
28	Exon 15	c.1731G>A	S566S	Substitution	1	Reported 62 times
29	Exon 16	c.1742delC	P581Rfs*10	Frameshift	1	Novel
30	Exon 16	c.1852_1854delAAG	K618del	Frameshift	1	Reported 54 times
31	Exon 17	c.1916dup	L639Ffs*6	Frameshift	1	Reported 6 times
32	Exon 17	c.1949dupTT	L650Ffs*12	Frameshift	1	Novel
33	Exon 17	c.1976 G>C	R659P	Missense	1	Reported 53 times
34	Exon 18	c.2041G>A	A681T	Missense	3	Reported 90 times
35	Deletion of Exon 1	-	-	LGR	1	Reported 5 times
36	Deletion of Exon 3-5	-	-	LGR	1	Reported 12 times
37	Deletion of Exon 4	-	-	LGR	1	Reported 2 times
38	Deletion of Exon 11-12	-	-	LGR	1	Novel

Table 4.6 MSH2 germline mutations in Indian LS cohort

Sr. No.	Exon	DNA change	Protein change	Type of mutation	No. of families	Reported in InSiGHT database
1	Exon 1	c.181C>T	Q61X	Nonsense	2	Reported 6 times
2	Intron `1	c.211+2T>G	-	Splice site	1	Novel
3	Exon 2	c.268A>T	K90X	Nonsense	1	Novel
4	Exon 2	c.270dupA	D91Rfs*9	Frameshift	1	Novel
5	Exon 3	c.443_444delTG	V148Gfs*3	Frameshift	1	Novel
6	Exon 3	c.595T>C	C199R	Missense	1	Reported 30 times
7	Intron 3_Exon 4	c.646-4_648delinsT	-	Splice site	1	Novel
8	Exon 4	c.687dup	A230Sfs*2	Frameshift	1	Reported 5 times
9	Exon 4	c.754C>T	Q252X	Nonsense	1	Reported 8 times
10	Exon 5	c.811delTCTG	S271Rfs*2	Frameshift	1	Reported 10 times
11	Intron 5	c.942+1G>T	-	Splice site	1	Reported 4 times
12	Intron 5	c.942+3 A>T	-	Splice site	4	Reported 220 times
13	Exon 7	c.1165C>T	R389X	Nonsense	2	Reported 49 times
14	Exon 9	c.1477 C>T	Q493X	Nonsense	1	Reported 11 times
15	Exon 10	c.1616_1617delTT	F539fsX	Frameshift	1	Novel
16	Exon 11	c.1705_1706del	E569Ifs*2	Frameshift	1	Reported 21 times
17	Exon 12	c.1786_1788delAAT	N596del	Frameshift	1	Reported 65 times
18	Exon 12	c.1801C>T	Q601X	Nonsense	1	Reported 6 times
19	Exon 12	c.1807G>A	D603N	Missense	1	Reported 29 times
20	Exon 12	c.1983dupA	Q662Tfs*14	Frameshift	1	Reported once
21	Exon 13	c.2038C>T	R680X	Nonsense	1	Reported 55 times
22	Exon 13	c.2162delG	G721Efs*24	Frameshift	1	Novel
23	Exon 14	c.2434_2453del	T812Efs*5	Frameshift	1	Novel
24	Deletion of Ex 6	-	-	LGR	1	Reported 16 times
25	Deletion of Exon 1-2	-	-	LGR	1	Reported 32 times
26	Deletion of Ex 14	-	-	LGR	1	Reported 3 times
27	Deletion of Ex 2	-	-	LGR	1	Reported 39 times
28	Deletion of Ex 8	-	-	LGR	1	Reported 2 times

4.6.3 Phenotypic characterization of MMR genes mutation carrier LS families

This study included a total of 118 individuals with confirmed pathogenic mutation from 88 families, of which 22% are healthy carriers. This included 75 *MLH1*, 39 *MSH2* and 4 *MSH6* gene mutation carriers. Phenotypic characterization of these 88 mutation carrier families revealed an additional 135 obligate carriers or presumed to be mutation carriers in view of their relationship to confirmed mutation carriers and personal history of LS associated caner but not tested due to unavailability of samples. A total of 273 cancers were recorded in these 229 individuals from 88 mutation carrier families.

Tumour frequencies and spectrum in Indian LS families

Among the 273 tumours, CRC was the most common cancer in both *MLH1* and *MSH2* mutation carriers, accounting for 74% versus 48% of tumours in *MLH1* and *MSH2* mutation carrier families respectively (Table 4.7). Proportion of CRC was higher in males than females in both *MLH1* (69/127, 54%) and *MSH2* (29/46, 63%) mutation carriers (Fig 4.11). The mean number of cancers recorded per family was similar for *MLH1* and *MSH2* mutation carrier families. In *MLH1* mutation carrier families, 1-11 tumours were recorded per family (mean: 3) whereas in the *MSH2* mutation carrier families there were 1-7 tumours per family (mean: 2.8). Multiple primary cancers were observed in 18/140 (13%) of MLH1 family members and in 6/85 (7%) of MSH2 families (Table 4.7).

Table 4.7 Spectrum of LS-associated cancers in families with MLH1 and MSH2 mutation

Tumour location	MLH1 (N=171)	MSH2 (N=98)
Colorectal	74%	48%
Endometrium	5.3%	23%
Stomach	1.3%	3%
Breast	5.3%	4%
Ovary	0.6%	5%
Intestine	0	1%
*Hepatobiliary	3%	2%
Adrenal gland cancer	0	1%
Pancreas	0.6%	0
#Urinary tract	3%	5%
Brain & CNS	1.2%	2%
Haematological malignancies	0.5%	2%

*Include cancers of liver, bile duct, ampulla and gall bladder
#Include cancers of ureter, bladder, kidney and prostrate

Colorectal cancer is the most common cancer in males and females mutation carriers. Extracolonic cancers were more common in *MSH2* (49%) as compared to the *MLH1* (25%) mutation carrier families. Endometrial cancer was the most common extracolonic cancer in females in *MSH2* carriers and second most common extracolonic cancer after breast in *MLH1* mutation carriers (Fig 4.11, Fig 4.12). In addition, breast, ovarian, stomach, hepato-biliary tract, urinary tract, brain and haematological malignancies were also observed in both *MLH1* and *MSH2* carrier mutation families. In addition, intestinal

cancers were seen only in *MSH2*. Interestingly, breast cancer was also found to be common in our cohort in both *MLH1* (n=9) and *MSH2* (n=4) mutation carrier families.

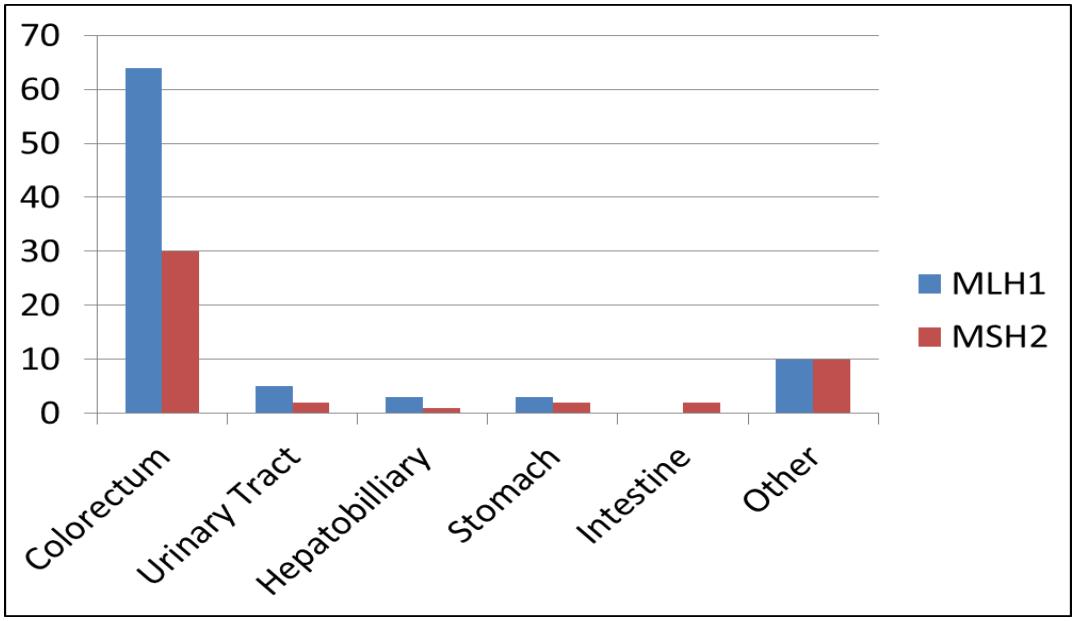


Fig 4.11 Extra-colonic cancers in males with mutations in MLH1 and MSH2

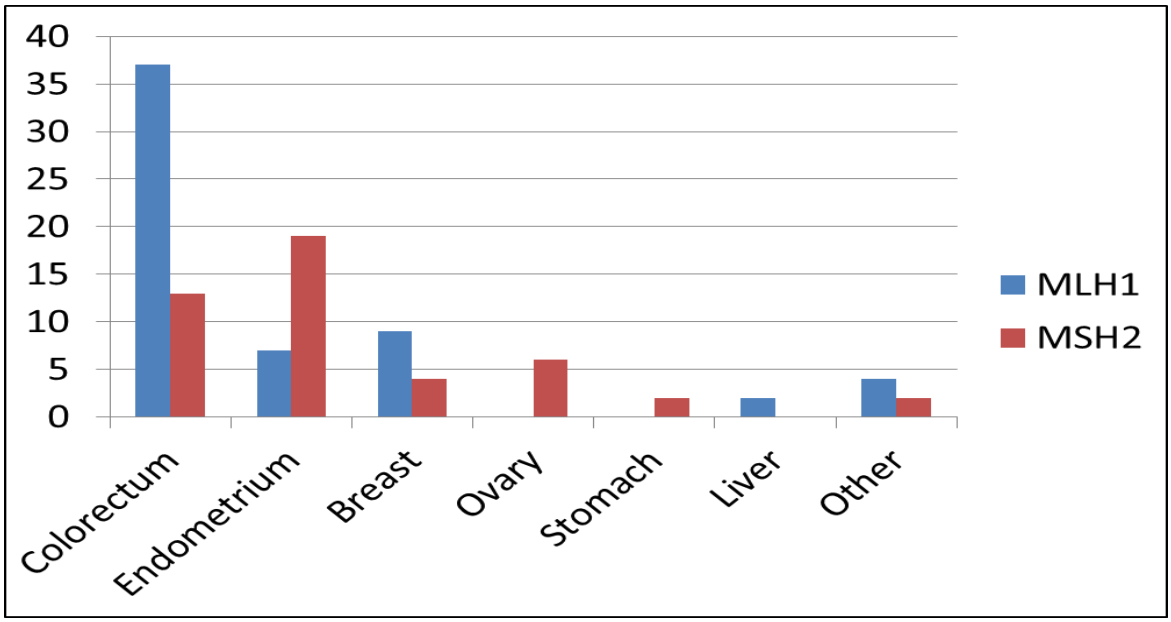


Figure 4.12 Extra-colonic cancers in females with mutations in MLH1 and MSH2

Age at diagnosis

The age at diagnosis of first cancer was similar in families with mutations in *MLH1* (48.3±13.6 years, range: 13y-84y) and *MSH2* (46.1±12.8 years, range: 3y-82y)**.** There was no significant difference in the mean age at diagnosis at first cancer between males and females among *MLH1* and *MSH2* mutation carriers. (Fig 4.13)

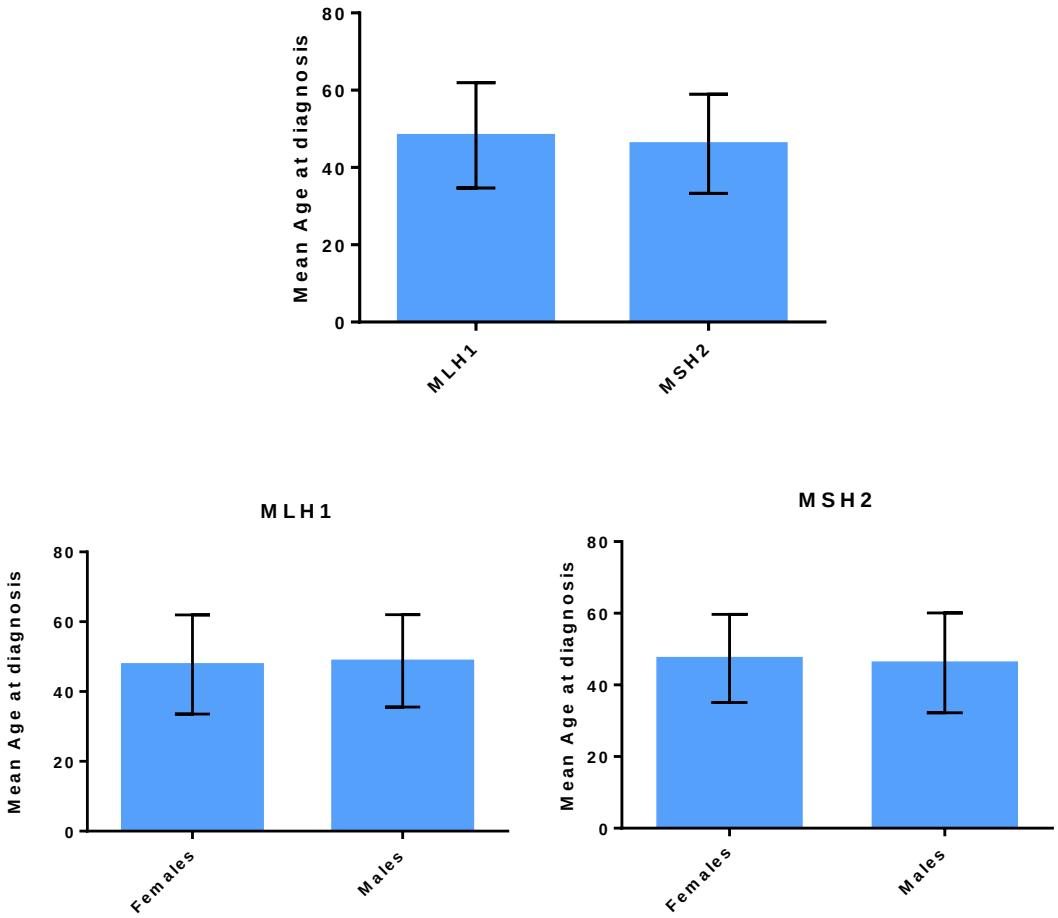


Fig 4.13 Mean age at diagnosis in MLH1 and MSH2 mutation carriers

The genotype-phenotype correlation of *MLH1* and *MSH2* mutation carriers are represented in Table 4.8. CRC was diagnosed at a mean age below 50 years in *MLH1* and *MSH2* mutation carriers in both genders. The extracolonic cancers showed

significantly higher age of diagnosis (57 years) as opposed to CRC (46.76 years) in MLH1 mutation carrier families (*p=0.0007).

Table 4.8 Genotype-phenotype correlation among MLH1 and MSH2 mutation carriers

	MLH1		**MSH2**		***p* value**
Families	53		33		
No. of patients with cancer	140		85		
Males	82 (58.5%)		42(49.4%)		
Females	58 (41.4%)		43 (50.6%)		
No. of tumours	n	(%)	N	(%)	
1 cancer	117	(83.6)	76	(89.4)	
>1 cancer	23	(18.6)	9	(10.6)	
Type of tumour					
CRC	103 (69 males & 34 females)	(73.6)	41 (29 males &12 females)	(48)	
EC	37 (13 males & 24 females)	(26.4)	44 (14 males & 30 females)	(51)	
Onset					
≤50 years	84	(60)	57	(67)	
> 50 years	56	(40)	28	(33)	
Age at primary tumour	Average (SD)		Average (SD)		
Total	48.44±13.62 years		46.50±13.12 years		0.3066
CRC (Both genders)	46.63±12.56[#] years		43.76±11.89 years[$]		0.2090
CRC (Males)	47.9±11.83 years		45.096±11.70 years		0.2840
CRC (Females)	44.72±13.47 years		43.85±14.05 years		0.3718
EC (Both genders)	56.08±14.50[#] years		48.86±13.62 years[$]		0.0236
EC (Males)	54.09±18.90 years		47.77±17.03 years		0.3974
EC (Females)	54.29±12.73 years		48.85±10.69 years		0.0902

EC: Extracolonic cancer

[#]comparison of age at onset of CRC and EC in *MLH1* mutation carriers; p-value: 0.0007

[$]comparison of age at onset of CRC and EC in *MSH2* mutation carriers; p-value: 0.0621

Penetrance Estimates

The lifetime risk for any LS associated caner and for CRC was estimated for the entire cohort and both genders individually for *MLH1* and *MSH2* mutation carrier families using the Kaplan Meier analysis (Fig 4.14 and Fig 4.15). The combined cumulative risks in both males and females of Lynch syndrome–associated cancer in *MLH1* and *MSH2* mutation carriers by age 70 years was estimated to be 93% and 95% respectively with no significant difference between the two groups (p=0.89).

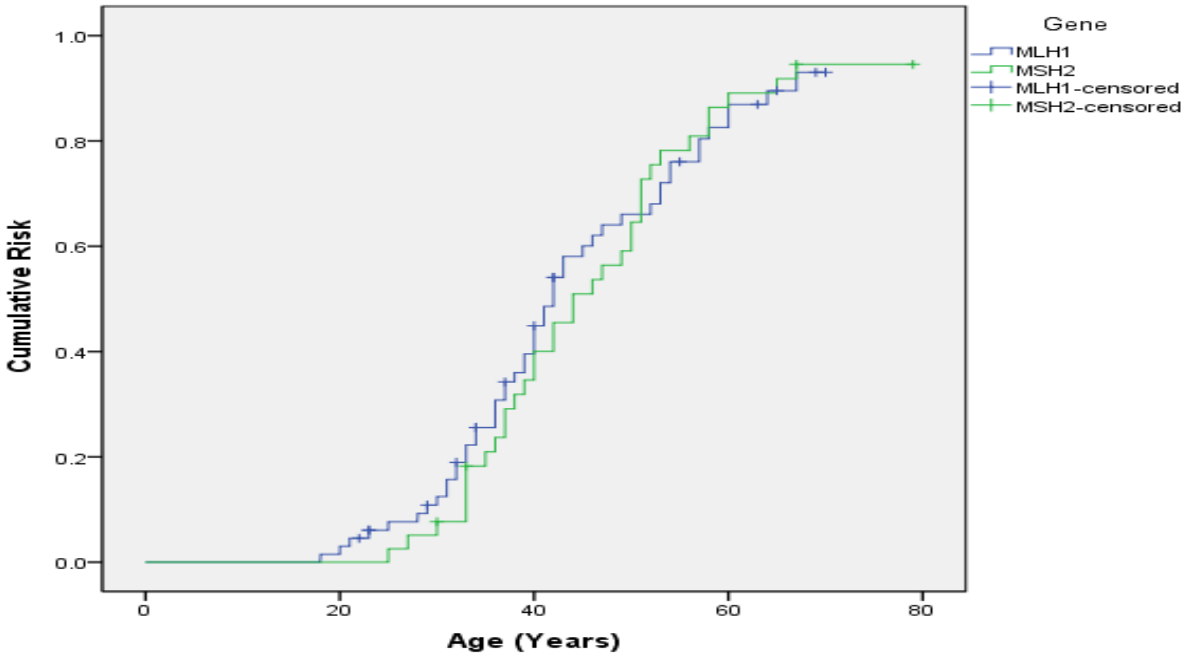


Fig 4.14 Any Lynch Syndrome associated cancers in MLH1 and MSH2 carriers.

Gender wise age related cumulative risk for CRC by age 70 years is shown for *MLH1* and *MSH2* mutation carriers in figure 4.15. For males, the cumulative risk for *MLH1* mutation carriers was 91% and for *MSH2* mutation carriers was 95%. The cumulative risks for CRC in females were 87% and 94% for *MLH1* and *MSH2* mutation carriers respectively.

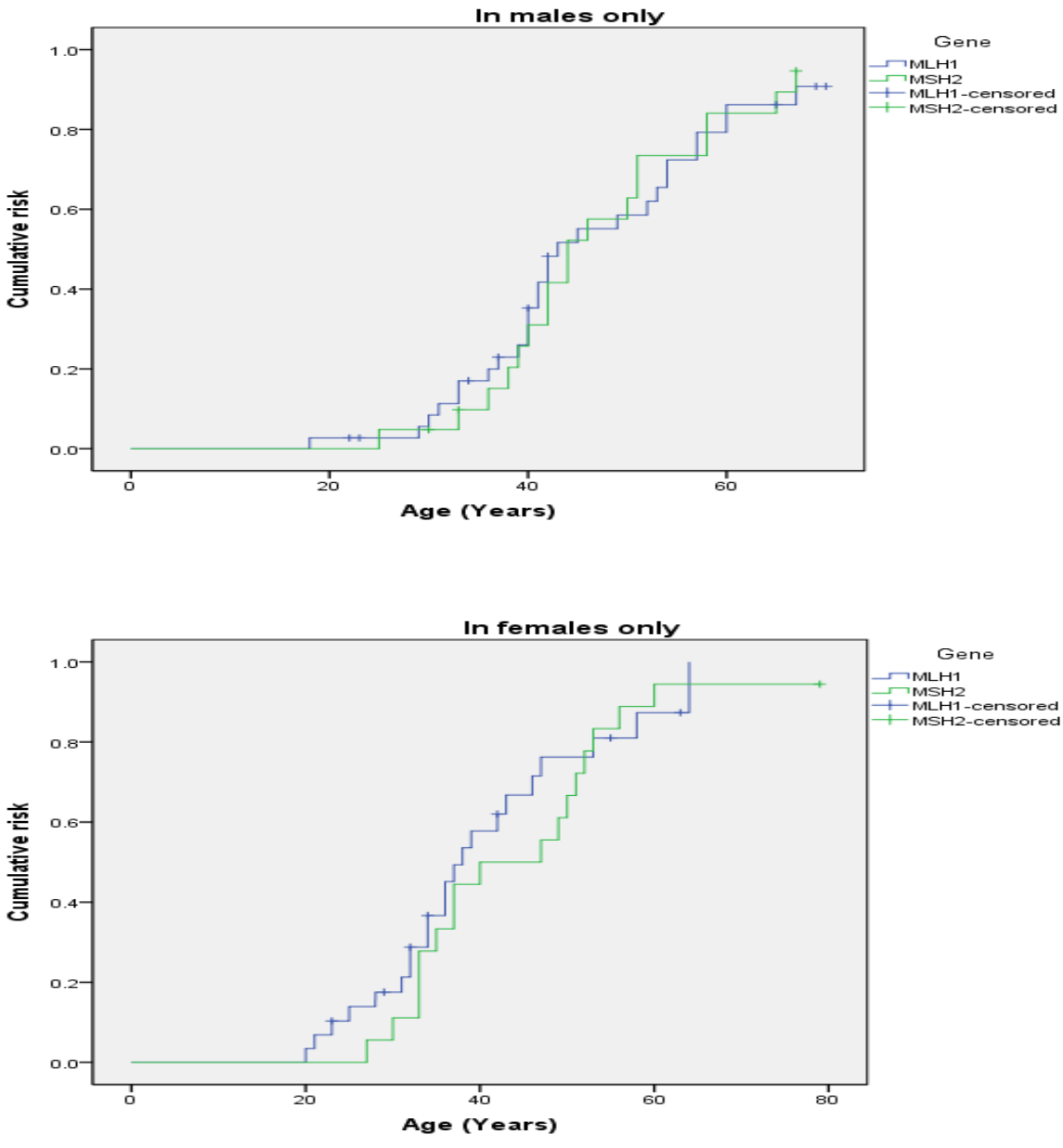


Fig 4.15 Cumulative CRC risks for MLH1 and MSH2 mutation carriers:

Top panel- males; Bottom panel-females

4.7 Discussion

Lynch syndrome is the second most common hereditary cancer after hereditary breast ovarian cancer. It carries a high lifetime risk of colorectal cancer as well as a number of extracolonic cancers, most prominently endometrial cancer and is caused by germline mutations in one of the mismatch repair genes. LS exhibit genetic heterogeneity in clinical phenotypes and gene mutation frequencies among various populations and even between different families of same population. Most of the genotype-phenotype data is from Caucasian populations and East Asian population [231, 232, 236, 237] with only 28 MMR gene mutation carrier Indian families reported so far [233-235].

The understanding about molecular basis and phenotypic manifestations of Lynch syndrome has evolved over time and continues to do so. The spectrum of LS associated extracolonic malignancies appears to be more diverse than previously described, and differs in various studies reported so far. This variation in phenotypic expression of LS can be explained by various factors like geography, ethnicity, gender, and underlying genetic mutation. Given this genetic and phenotypic heterogeneity in LS, we undertook a comprehensive analysis of the correlation between these clinic-pathological features and genetic parameters in an Indian cohort of 91 LS families, in order to establish a possible phenotype–genotype correlation.

Comprehensive genetic analysis using Sanger sequencing for point mutations and small indels, and MLPA for LGRs resulted in the identification of MMR gene mutations in 88 of the 91 suspected Lynch syndrome families. The mutation detection rate of >90% in our study is higher than other reports which vary from as low as 7% [238] to as high as 88% [239]. In the earlier study from India, the mutation detection rate was only 50% [233]. The high mutation detection rate in our cohort can be attributed to the use of

multimodal approach for selection of patients using different clinical criteria for syndromic diagnosis and pre-screening techniques like IHC and comprehensive genetic testing involving Sanger sequencing and MLPA.

The predominance of *MLH1* and *MSH2* gene mutations which account for 90% of all MMR gene mutation in our cohort is in accordance with their frequency reported in other populations [181]. In our study, majority of the families had distinct deleterious germline mutations highlighting the genetic heterogeneity of Lynch syndrome. The high frequency of novel mutation (20/83) in our study confirms that the mutation spectrum in Indian LS families is very distinct from the Caucasian population. These results underscore the need to study mutational spectrum in each population which can facilitate appropriate diagnostic algorithms and targeted surveillance programmes.

Mutations in the *MLH1* were scattered throughout the gene while *MSH2* mutations are limited up to Exon 12 covering the region which encodes the DNA binding and *MSH6*/*MSH3* interaction domains. This observation reiterates that genetic analysis approaches in Lynch syndrome must cover the entire coding region of the MMR genes and inclusion of MLPA analysis is mandatory. Truncating mutations predominate in both *MLH1* and *MSH2* gene with splice site mutations and LGRs being second most common type of mutation in *MLH1* and *MSH2* genes respectively. The high frequency of LGRs in *MSH2* is due to the presence of higher number of *Alu* repeat sequences in the *MSH2* gene [240].

Despite the high degree of heterogeneity observed, 7 recurrent mutations in *MLH1* and 3 recurrent mutations in *MSH2* gene were identified in 21 and 8 unrelated families respectively, together accounting for 34.9% (28/83) of all mutation carrier families. It is interesting to note that the mutation spectrum including the recurrent mutations in our

cohort is different than that reported in other LS cohort reported from South India [233]. For example, the most recurrent mutation worldwide c.942+3A>T in Intron 5 of *MSH2* gene reported 220 times in the InSiGHT database [189] was also identified in 4 families in our cohort (12% of all *MSH2* carriers families); however this mutation was absent in the South Indian LS cohort. Nonetheless, our findings may lead to the establishment of cost-effective LS screening protocol for Indians. For the *MLH1* gene, the 7 recurrent mutations in exons 3, 4 , 9, 10, 13 and 18 can be screened first followed by full gene analysis and MLPA. For the *MSH2* gene, the recurrent mutation in Intron 5 must be screened first followed by analysis of LGRs and then sequencing of the entire coding region if required. As the mutations in *MLH1*/*MSH2* are scattered throughout the gene, it is important to sequence the entire coding region of these genes, in cases where hotspot mutations are not identified after initial screening. IHC showed a high concordance with germline mutations and high sensitivity of 97.4% in our cohort therefore emphasizing the use of IHC as a pre-screening technique to guide genetic testing in Lynch syndrome which otherwise is time consuming owing to the genetic heterogeneity associated with it. Such high sensitivity and specificity has been reported for IHC in previous studies also [163]. Use of IHC also facilitated identification of 30% (25/83) of MMR *de novo* mutation positive LS families in our cohort with no family history therefore highlighting the importance of this technique in LS diagnosis and screening.

Phenotypic characterization in 229 individuals from 88 MMR mutation positive families in our cohort showed clinical characteristics that largely concur with the known data; though many noteworthy exceptions were identified. In accordance with the previous studies, colorectal cancer accounted for majority of the tumours and had a lower age at diagnosis as compared to extracolonic tumours [241]. We observed a predominance of

CRCs in both *MLH1* and *MSH2* mutation carriers in our cohort that is in line with previous studies showing similar frequencies. There was no significant difference in cumulative risk of LS-associated tumours between *MLH1* and *MSH2* mutation carriers with more than 90% risk in both groups. This is higher than the cumulative risk reported in most studies, and most likely due to the relatively smaller number of extended family members tested which would have identified more healthy mutation carriers in our cohort.

We also observed a comparatively higher frequency of extracolonic cancers in *MSH2* mutation carriers with endometrial cancer being the first and second common extracolonic in *MSH2* and *MLH1* mutation carrier families respectively, similar to previously reported studies [128, 241]. The role of breast cancer in Lynch syndrome has been controversial [242-244]. Our finding of breast cancer being the second most common extracolonic malignancy in our cohort, identified in 14% cases supports the inclusion of breast cancer in the LS tumour spectrum. There was no significant difference in the frequencies of various extracolonic cancers between the *MLH1* and *MSH2* mutation positive families except ovarian and intestinal cancers, which were only seen in *MSH2* mutation carrier.

While the high frequency of urothelial cancers in our cohort is in agreement with previous studies [241], the high frequency of bladder cancers (5/88 cases) in our study provides further evidence for including them in LS spectrum [245] We also report the identification of adrenal gland cancers in one male with *MSH2* mutation at the age of 55 years. This adds to the increasing reports of adrenal gland cancers in Lynch syndrome families [246] providing an additional evidence for this association.

In summary, we report genetic and clinical characterization of 88 Indian Lynch syndrome families with significantly high mutation detection rate and high frequency of novel and recurrent mutations prevalent in Indian population. We also report the clinical characteristics of 229 individuals from these 88 MMR mutation positive Indian LS families that are slightly different from other studies thereby expanding the knowledge of phenotypic spectrum associated with LS and shed a light on the emerging genotype-phenotype correlations in Lynch Syndrome.

In addition to the work done for this thesis, earlier work from our group has identified additional MMR gene mutation carrier families (unpublished), independent from the present study [335]. A pooled analysis of these two independent Lynch Syndrome cohorts sequentially enrolled at TMC is being currently performed to obtain more robust genotype-phenotype correlations.

CHAPTER 5

CONSTITUTIONAL MISMATCH REPAIR DEFICIENCY SYNDROME

5.1 Introduction

Constitutional Mismatch Repair Deficiency (CMMRD; OMIM #276300) syndrome is a rare autosomal recessive distinct childhood cancer predisposition syndrome with less than 200 cases reported worldwide [247]. Unlike Lynch syndrome, this disorder usually occurs in infancy or adolescence at an incidence of 1 per million [248]. In contrast to Lynch syndrome (LS) patients with monoallelic mismatch repair (MMR) gene mutations, CMMRD syndrome is a result of constitutional biallelic homozygous or compound heterozygous mutations in the genes encoding MMR proteins.

5.1.1 History of CMMRD

CMMR-D syndrome was recognized relatively recently as a 'variant' of Lynch syndrome – it is likely that earlier cases were misclassified or missed, since CMMR-D patients have overlapping features with other paediatric cancer syndromes. Nearly 20 years ago the original report of CMMRD was published by Ricciardone and Tayfun (1999) who described three children of consanguineous Turkish parents [249] with haematological malignancies by the age of 3 years and clinical characteristics of neurofibromatosis type 1 (NF1: OMIM; 162200) e.g. café-au-lait (CAL) spots. Mutation analysis revealed an inherited homozygous mutation in *MLH1* in each of the children. Both the parents who; although unaffected carried the same cancer-predisposing *MLH1* mutation in heterozygous state. The pedigree analysis revealed that both parents had a family history of Lynch syndrome. Further analysis suggested that the presence of *NF1* symptoms was due to somatic and not germline mutations in the *NF1* or Neurofibromin gene on chromosome 17. The presence of the haematological malignancies also suggested that these cases were not classical NF1 patients.

Soon after, another report came up which also described paediatric patients with similar phenotypes, where the parents had a consanguineous marriage [250]. The description was of a family from North Africa, with a confirmed history of LS spectrum tumours. The first CMMR-D case in this family was that of a child, who developed malignant Non-Hodgkin's lymphoma (NHL; OMIM: 605027) at the age of 2 years. Subsequently, her sister was diagnosed with acute myeloid leukaemia (AML; OMIM: 61626) at the age of 6 years, followed by medulloblastoma (MDM; OMIM: 155255) at the age of 7 years. Both sisters had CAL spots with no previous family history of NF1. Genetic testing in one of the children confirmed the presence of the homozygous mutation in *MLH1* previously described as pathogenic and disease causing in LS [250]. Genetic testing for the other sibling could not be done due to unavailability of DNA but the observation that the children had similar characteristics and disease presentations led the authors to conclude that the other child in all likelihood must be a presumed carrier of the same biallelic *MLH1* mutation. Parents had the same mutation in *MLH1* in heterozygous state.. In both families it was speculated that a defective MMR system caused a mutator phenotype, which resulted in somatic mutations in the relatively large *NF1* gene on chromosome 17 (374,244 bp).

The first two reported cases of CMMR-D had homozygous MLH1 mutations. The first report of homozygous mutation in another MMR gene, MSH2 came in 2002 [251]. The patient was diagnosed with acute lymphoblastic leukaemia (ALL) and showed signs of NF1, though did not meet the diagnostic criteria for an NF1. The parents were not related and there was no family history of cancer, which added to the complexity of making a syndromic diagnosis. Till date, this syndrome has been shown to exhibit a variable

phenotype and it keeps on expanding with identification of more cases therefore it is possible that some features and phenotypes are yet to be described.

Although unproven at the molecular level, it is suspected that Jacques Turcot could have described the first cases of CMMRD in 2 Canadian siblings with multiple adenomatous polyps, colorectal cancer and malignant brain tumours in 1959 [211]. In 1995 another group of researchers classified Turcot syndrome into dominantly and recessively inherited types [212]. The eponym "Turcot syndrome" denoting a combination of colorectal polyposis and Central Nervous System cancers was used for several years which included both dominant cases with heterozygous APC mutations and recessive cases with homozygous MMR mutations. One interesting observation was that Turcot syndrome cases with APC gene mutations often develop medulloblastomas compared to MMR deficient cases, most of whom develop glioblastomas [212, 252, 253]. Following the description of CMMRD syndrome, some of the previously described cases of Turcot syndrome have been re-classified as CMMR-D. The term Turcot syndrome has now largely been replaced by CMMRD because the definition is too restrictive as CMMRD also includes childhood hematologic malignancies and cafe-au-lait spots suggestive of NF1.

5.1.2 Epidemiology of CMMRD syndrome

The first comprehensive report of 78 cases of CMMR-D syndrome from 46 different families was published in 2008 [254]. Subsequently, a European Consortium Care for CMMRD (C4CMMRD) reported a worldwide total number of 146 patients from 91 families published in 63 papers in 2014 and described the clinical and diagnostic criteria of CMMRD syndrome. Most of the cases reported to date are from the North

American and European population, however data from countries in the Middle East and Africa have also emerged recently [255-260].

In countries of Western Asia like Saudi Arabia, Israel etc. the numbers of CMMR-D cases are higher than expected, probably due to prevalence of consanguinity in those regions [261]. A retrospective pooled analysis of reports published between 2003 and 2013 identified 42 cases that met the CMMR-D syndrome criteria of which 51% cases had childhood gliomas [262]. It is presumed that CMMRD could be highly prevalent in South Asia as consanguinity is common in these regions, however very few case reports have been published.

5.2 Clinical features associated with CMMRD syndrome

A constitutively defective MMR system in CMMRD syndrome results in a wide variety of malignant tumours in early childhood [263]. The tumour spectrum associated with CMMR-D can be grouped into four categories; (i) haematological malignancies, (ii) brain tumours, (iii) gastric cancers (including the Lynch syndrome spectrum of cancers) and (iv) others [264]. Brain and haematological malignancies are mostly diagnosed in the first decade of life and LS-associated tumours in second and third decades of life [265]. The average age at diagnosis for gliomas, haematological malignancies and colorectal cancers (CRCs) is 9.5 years, 5 years and 16 years respectively [266].

One of the most common phenotypic features of CMMRD syndrome is the presence of NF-1 like café-au-lait spots (CAL). However, the phenotypic features of NF1 are quite distinct, comprising largely of e.g. CAL macules on the skin while in CMMR-D there is hyperpigmentation of the CAL macules with hypo-pigmentation in the inner areas; the skin lesions are more diffuse and irregular when compared to the classic CALs in

inherited NF1 [267]. The number of skin lesions in CMMR-D varies, ranging from mostly one to two focal areas, and sometimes with more diffuse areas of skin pigmentation [267]. In a report by Bakry et al. (2014), 97% of CMMR-D patients who were followed up as part of the BMMRD consortium, had developed CAL macules [258]. 'Ash leaf spots' or 'vitiligo' are another pigmentary abnormality seen in CMMRD which are areas of skin hypopigmentation.

Additionally developmental venous anomalies, pilomatricomas agenesis of the corpus callosum, and mild immunodeficiency with decreased levels of immunoglobulins IgG2/4 and IgA, are also observed among other clinical features [268].

5.2.1 Family history in CMMRD

Parental consanguinity is a relatively common feature of CMMR-D patients however the rate of consanguinity varies with high rate observed among homozygous cases and no consanguinity seen in compound heterozygous cases [268]. Review of literature suggests that parents of offspring with CMMR-D are often unaffected especially among families with *PMS2* or *MSH6* mutation. This observation may be because CMMR-D-related cancer diagnosis is made in very young children (e.g. <5 years of age) when their parents may well be in their third (20s) or fourth (30s) decade of life while the average age of LS diagnosis is around 40-60 years. However, some CMMR-D families may have a history of the Lynch syndrome associated cancers in extended family members. Therefore, family history may not always be sufficient to make CMMRD diagnosis. The absence of family history could also be due decreased penetrance associated with certain genes and mutations. The penetrance of *PMS2* mutation which is the most commonly mutated gene in CMMRD is relatively lower compared to *MLH1* and *MSH2*. This may be the likely

explanation for lack of a clear dominant family history among individuals who harbour biallelic *PMS2* mutations and develop CMMR-D associated cancers.

5.2.2 Diagnosis of CMMRD and related challenges

It should be noted that due to the overlapping features of CMMRD with many other inherited syndrome like NF-1, LFS, FAP and LS and with no definitive clinical features to distinguish sporadic brain tumour or leukaemia patients from CMMRD patients, they are usually misdiagnosed or misclassified. Though the specific features of the CAL macules differ from the classic NF1 as described above, it can only be correctly evaluated by experienced clinician by direct physical examination of the patient. Other NF-1 features like axillary freckling, Lisch nodules and plexiform neurofibromas have also been reported in BMMR-D, however only a small subset of these meet established NF-1 diagnostic criteria [269]. CMMRD tumour spectrum overlaps with LFL with prevalence of childhood brain tumours and haematological malignancies which may result in misdiagnosis therefor LFL cases with no TP53 mutation or any other unexplained childhood cancer should be taken up for MMR testing. Similarly, CMMRD patients with GI cancers often develop polyps and therefore can be misdiagnosed as FAP [270]. However, the presence of high-grade dysplasia, type of cancer and early onset distinguishes this syndrome from the classic familial adenomatous polyposis (FAP) [271]. Therefore, FAP-like cases with no pathogenic *APC* or *MUTYH* mutation should be suspected to be CMMRD case. Due to the complex nature of CMMRD, it is often misdiagnosed therefore appropriate surveillance measures could not be given timely and individuals end up dying of potentially preventable malignancies. Another challenge is that CMMRD patients who survive their first malignancy usually have a risk of developing a second primary tumour therefore CMMRD diagnosis is important for

providing appropriate anticancer treatment, surveillance and genetic counselling to parents regarding the risk of recurrence of second primary in their children and their own risk of developing LS associated cancers. This issue was addressed in a newly established European consortium named "Care for CMMRD" where a 3 point scoring system for clinical criteria was proposed to confirm the diagnosis of CMMRD (Table 5.1). According to this scoring system, CMMRD should be suspected in all individuals who scores ≥3 points [264].

Table 5.1 Three-point scoring system for diagnsois of CMMRD

Indication for CMMRD testing	**≥3 points**
Malignancies/pre-malignancies: at least one is mandatory, if more than 1 is present add points	
LS spectrum tumours at age <25years	3 points
Multiple polyps with absence of APC/MUTYH mutations at age <25 years	3 points
WHO grade III or IV glioma at age <25 years	2 points
NHL of T-cell lineage or sPNET at age <18 years	2 points
Any malignancy at age <18 years	point
Additional features, if more than one of the following is present then add points	
Clinical features of NF1 and/or hyperpigmented or hypopigmented spots	2 points
Diagnosis of LS in first or second degree relative	2 points
Diagnosis of LS spectrum tumours in 1st, 2nd or 3rd degree relative <60 years of age	1 point
A sibling with LS-spectrum tumours, high grade gliomas, NHL or sPNET	2 points
A sibling with any childhood malignancy	1 point
Multiple pilomatricomas in patient	2 points
One pilomatricoma in patient	1 point
Agenesis of the corpus callosum or non-therapy induced cavernoma in the patient	1 point
Consanguineous parents	1 point
Deficiency or reduced levels of IgG2/4 and/or IgA	1 point

LS-spectrum tumours: Colorectal, small bowel, endometrial, uterine, renal pelvis, biliary tract, stomach, bladder carcinoma

Abbreviations- CMMRD: Constitutional Mismatch Repair Deficiency, LS: Lynch syndrome, NHL: Non-Hodgkin's Lymphoma, sPNET: supratentorial primitive neuro-ectodermal tumours, WHO: World Health Organization

However, genetic testing is recommended to confirm the diagnosis of CMMRD. When a biallelic germline mutation is identified, the next challenge is to determine the pathogenicity of the mutation for which we have to rely on previous literature. Though InSiGHT database curates MMR gene mutations but they are mostly associated with Lynch syndrome. Therefore, a combination of pre-screening tools like IHC and MSI along with well-defined diagnostic algorithms and confirmation of germline mutations in CMMRD patients is warranted to identify such families so that they can benefit from surveillance measures.

5.3 Spectrum of tumours in CMMRD

In CMMRD syndrome brain and digestive tract tumours account for 50% and 40% of all cancers, followed by haematological malignancies in over 30% of patients [265]. The tumour type is influenced by the MMR gene involved, with brain tumours being more prevalent in case of *PMS2/MSH6* gene mutations while haematological malignancies occur more often in case of *MLH1/MSH2* gene mutations [264, 267]. There is no significant difference for LS associated tumours with respect to the MMR gene [264].

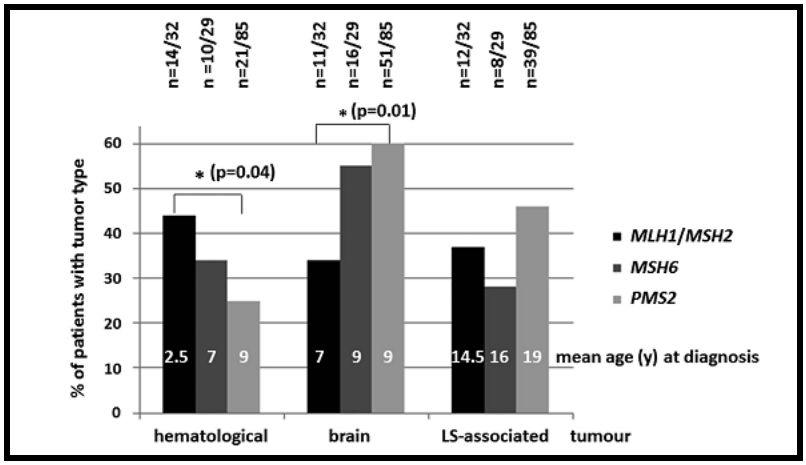


Fig 5.1 Differences in the prevalence of tumour type depending upon the MMR gene involved (Taken from Ref [264])

5.3.1 Haematological malignancies

In general, haematological malignancies arise from bone marrow and plasma cells [272, 273] and are divided into leukaemias, lymphomas and plasma cell neoplasms [273]. The most common haematological malignancies in CMMR-D are Non-Hodgkin's lymphoma (NHL) and acute leukaemia [254]. T cell NHL more frequently observed than B cell NHL [268]. Based on histological subtypes the age at diagnosis of sporadic NHL varies from 5 to 12 years [274]. The median age at diagnosis of NHL in CMMR-D is 5 years (range 0.4–17 years) [264].

The most common type of acute leukaemia seen in CMMRD is acute lymphoblastic leukaemia (ALL). It is a heterogeneous cancer, which originates from both B and T lymphoid progenitors. The mean age at diagnosis of ALL in CMMR-D is 6 years (range 2–21 years) [264]. It has been proposed that development of other types of acute leukaemia, such as acute myeloid leukaemia (AML) might be due to the administration of chemotherapy [254].

5.3.2 Brain tumours

Brain tumours are the most common malignancy in CMMR-D cases. The most common type of brain tumours are the glioblastomas [254]. Glioblastomas originate from glia cells and are classified as grade IV according to the WHO grading system (2007). The mean age at diagnosis of glioblastoma in CMMR-D patients is 9 years [264]. Other types of brain cancers reported as part of the CMMR-D spectrum include medulloblastoma and sPNET (supratentorial primitive neuroectodermal tumours) [265]. Medulloblastomas are believed to arise in the cerebellum and sPNET in other parts of the brain. The mean age at diagnosis in of these tumours in CMMR-D patients is 7 years.

5.3.3 Lynch syndrome associated tumours

The most prevalent LS cancers seen in CMMR-D patients are colorectal carcinoma although small bowel, endometrial, ovarian and urinary tract cancers are also observed. Majority of CMMR-D patients with CRC have >10 adenomas and typically diagnosed between 5-10 years of age [268, 275]. The average age of diagnosis of CRC in CMMR-D patients is around 16 years (range of 23-35 years) which is relatively late as compared to other CMMRD cancers [254, 264] but much lower as compared to LS. Additionally, the progression of adenoma to carcinoma is quite rapid in CMMRD patients and they mostly present with advanced metastatic GI cancers. Renal and bladder cancers also occur but at a lower frequency in CMMR-D [264].

Few other cancers within this syndrome have also been classified under ‘emerging phenotypes’, including hepatic carcinomas, neuroblastomas, Wilms tumour and sarcomas such as osteosarcoma and rhabdomyosarcoma [268]. Tumours associated with CMMRD are listed in Table 5.2

Table 5.2 Spectrum of cancers reported in CMMRD

Type of cancer	Median age at onset in years (Age range)
Haematological Malignancies	
Non-Hodgkin's Lymphoma (NHL)	5 (0.4-17)
Lymphoid leukaemia	6 (2-21)
Acute myeloid leukaemia	9.5 (6-17)
Malignant lymphoma	15
Atypical chronic myeloid leukaemia	1
Acute leukaemia	2
Overall Average age	**6 (0.4-21)**
Brain tumours	
High grade gliomas	9.5 (2-40)
sPNET	8 (4-17)
Medulloblastoma	7 (4-12)
Unspecified	8
Overall Average age	**8 (4-24)**
LS-associated cancers	
CRC	16 (8-48)
Small bowel	28 (11-42)
Endometrium	28 (23-44)
Urinary bladder	19(15-22)
Ovary	15
Overall Average age	**17 (8-48)**
Other cancers	
Neuroblastoma	13
Wilms tumour	4
Ovarian neuroectodermal tumour	21
Infantile myofibromatosis	1
Rhabdomyosarcoma	4
Osteosarcoma	24

Adapted from (Wimmer and Etzler, 2008; Wimmer et. al., 2014

5.4 Ultrahypermutator phenotype- Hallmark of CMMRD tumours

Cancer is caused by mutations that result in unregulated cell division, invasion and metastasis. The low mutation rate in normal cells is insufficient to cause cancer which otherwise requires tens to hundreds of thousands of genetic alterations therefore cancers must exhibit an increased mutation rate early in their evolution [276]. The total mutation count, known as the mutation burden is dependent on the tumour type and may also be affected by the mutagenic processes shared across different tumour types [277]. The concept of progressive accumulation of mutations in cancers was initially proposed as the "The Mutator Phenotype Hypothesis" based on theoretical models [278]. Proof of concept of this hypothesis came initially from the findings of Lynch syndrome and in a subset of sporadic colorectal cancers [38, 149].

Until recently, the concept of mutator phenotype was not widely considered as a major contributor to tumour initiation or progression however there is emerging evidence that high mutation burden exists for many cancers. The strongest evidence in support of the mutator phenotype in human cancers is the data presented by The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC) which has confirmed hypermutation in several tumour types. It has been shown by the international bMMRD (biallelic mismatch repair deficiency) consortium that all malignant CMMRD cancers are ultrahypermutant [279]. As opposed to most childhood cancers with mutation rates of <10/MB, CMMRD exhibits quite a high mutational burden in the tumour with mutation rates of 100/MB which may play a role in future diagnostic algorithms [268].

5.5 Genetic features of CMMRD

CMMRD is caused by biallelic mutations in one of the four main MMR genes, namely *MLH1*, *PMS2*, *MSH6* and MSH2. These MMR proteins play an important role in MMR activity (Chapter 4, section 4.3). In contrast to LS, *PMS2* and *MSH6* are the most commonly mutated genes reported in patients diagnosed with CMMR-D while mutations in *MLH1* and *MSH2* are rare (Fig 5.2). It could be due to lethality of homozygous null mutations in *MLH1* and *MSH2* as they are obligate partners of the MMR system while lower penetrance and clinical severity of *PMS2* and *MSH6* due to their redundant role [268]. The InSiGHT website maintains a database of all MMR mutations however they are mostly associated with LS and not CMMRD. Due to the rarity of the syndrome, single case reports and small cohort studies, there is no dedicated database for CMMRD associated MMR gene mutations.

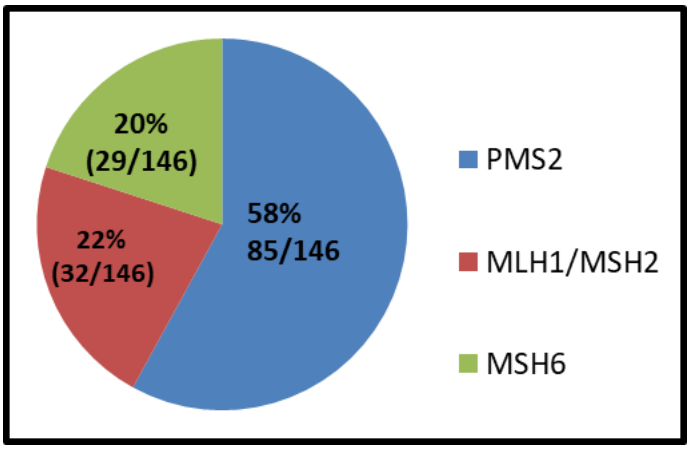


Fig 5.2 Frequency of mutations in MMR genes in CMMRD

(Adopted from Ref [264])

5.5 Genotype-phenotype correlation in CMMRD

Due to the rarity of the syndrome and mostly single case reports, genotype-phenotype correlations in CMMRD are difficult to assess. As per the most recent and largest review of CMMRD cases by the CMMRD consortium in 2014, individuals with either *MLH1* or *MSH2* mutations are more likely to develop haematological primary malignancies, and at younger age of onset compared to cases with mutations in either *MSH6* or *PMS2* [264]. About 60% of the *PMS2* and *MSH6* mutation carriers developed brain tumours as their primary malignancy as opposed to only about 30% of the *MLH1* or *MSH2* mutation carriers [248, 264]. Although patients with either *MSH6* or *PMS2* mutations were likely to survive their first malignancy, 40% of the patients with homozygous *PMS2* mutations tend to develop second primary malignancy as opposed to patients with homozygous *MLH1*/*MSH2* mutations. This may be due to more aggressive and lethal haematological malignancies in homozygous *MLH1*/*MSH2* mutation carriers [248]. The literature on CMMRD is still emerging and more research is needed to establish a clear genotype and phenotype correlations for this devastating group of diseases.

5.6 Hypothesis and objective of the study

CMMRD is a very rare childhood cancer predisposition syndrome with mostly small case series and their pooled analysis. It displays a wide spectrum of childhood cancers which continues to emerge with new case reports. Due to its rarity and an unusual tumour spectrum which overlaps with several other syndromes, CMMRD is an under diagnosed syndrome for which the genotype-phenotype data are still emerging [263, 267, 280]. The lower number of cases reported so far can also be attributed to the fact that genetic analysis of *PMS2*, the most commonly involved gene in CMMRD poses specific

technical challenges due to the presence of multiple pseudogenes which shares homology with the *PMS2* gene. Only about 200 CMMRD cases with confirmed biallelic MMR gene mutations have been reported globally with very few reports from South-Asia with only one report from India [281]. As consanguinity is commonly practiced in several regions and religions in India, the prevalence of CMMRD is expected to be higher in India. Therefore, one of the objectives of this study was to identify CMMRD families and to characterize the mutation spectrum of MMR genes, especially *PMS2* gene as it accounts for majority of mutations in CMMRD and to study the phenotypic features of the mutation carrier CMMRD families.

The hereditary cancer syndrome families registered in the Cancer Genetics Clinic especially with unexplained childhood cancers were screened using 3 point scoring system based on comprehensive pedigree analysis and other clinical features. Depending upon the availability, IHC status was also taken into account for genetic testing of a CMMRD suspected family. The methodology used for genetic screening of MMR genes was to PCR amplify the coding and flanking intronic regions of the MMR gene as guided by IHC, in case of unavailability of IHC results, *PMS2* gene was first studied followed by MLH1, MSH2 and MSH6. *PMS2* analysis was carried out using long range PCR to avoid pseudogene amplification followed by nested PCR to amplify individual exons. PCR products were subjected to Sanger sequencing in order to identify the germline mutation.

5.7 Results

5.7.1 Clinical characteristics of the patients

A total of 12 unrelated South Asian CMMRD suspected families based on their clinical features and three point scoring system were included in this study. The clinical description, family history and the CMMRD score for probands in these 12 families is shown in Table 5.3.

Table 5.3 Clinical features, family history and three point score of probands with suspected CMMRD syndrome

Probands (Gender)	Cancer site	Age at diagnosis	CALs	Parental consanguinity	CMMRD associated FH	Three point Score	Mutation Status (Y/N)
CM1 (M)	GBM; Colon	11y; 15y	NA	Yes	Yes	8	Y
CM2 (M)	T-ALL; ASPS	6y; 8y	No	No	Yes	5	Y
CM3 (F)	GBM	9y	Yes	No	No	4	Y
CM4 (F)	Rectum	10y	No	Yes	No	5	N
CM5 (F)	TLBL	4y	Yes	Yes	Yes	6	Y
CM6 (M)	GBM	4y	Yes	Yes	No	5	Y
CM7 (F)	GBM	5y	Yes	Yes	Yes	8	Y
CM8 (M)	T-ALL	4y	NA	No	Yes	2	N
CM9 (M)	Pilocytic astrocytoma; Rectum	11y; 13y	No	Yes	No	5	N
CM10 (F)	GBM	11y	No	Yes	No	3	N
CM11 (M)	MBM	2y	NA	No	Yes	3	N
CM12 (M)	Nasopharynx	13y	NA	Yes	Yes	3	N

Abbreviations- CALs: Café-au-lait spots; GBM: Glioblastoma; T-ALL: T-cell Acute Lymphoblastic Leukaemia; ASPS- alveolar soft part sarcoma; TLBL: T-cell lymphoblastic lymphoma

After comprehensive MMR gene analysis in these families, biallelic MMR gene mutation was identified in 7 families - 1 from Eastern Indian state of West Bengal, 1 from Western Indian state of Maharashtra, 1 from South Indian state of Karnataka and 4 from North Indian states of UP, Bihar and Rajasthan. A total of 17 malignancies were

diagnosed in 12 patients from 6 *PMS2* mutations carrier families and 1 from MSH2 mutation carrier family. The tumour spectrum included 6 high grade gliomas, 6 haematological malignancies, 2 mediastinal tumours, 1 colon cancer, medulloblastoma and alveolar soft part sarcoma each. A family history of CMMRD associated cancers was noted in 4/7 families. Parental consanguinity was present in 4 families and in 2 families the parents were unrelated but belonged to the same inbred community. Café-au-lait spots were present in 4/6 probands for whom data was available.

5.7.2 Genotypic and phenotypic characteristic of MMR gene mutation carrier CMMRD families

We identified 8 individuals with CMMRD associated malignancies with confirmed biallelic pathogenic MMR gene mutations from 7 families. Four siblings of biallelic MMR gene mutation carriers could not be tested as they had died earlier of characteristic CMMRD cancers and were presumed to be biallelic MMR gene mutation carriers. A known *PMS2* biallelic frameshift mutation (c.1500delC; p.V501Wfs*94) in exon 11 was identified in the proband of CM1 family and his brother. This frameshift mutation leads to a frameshift in codon 501 resulting in a premature stop codon 94 codons downstream and has previously been described in the ClinVar database as a pathogenic mutation with respect to Lynch syndrome and in literature as a Pakistani founder mutation [266]. The proband was diagnosed with glioblastoma (GBM) at 11yrs and died within months of developing a second malignancy (adenocarcinoma of colon) at 15 years. The brother (CM1a) of proband developed medulloblastoma (MDM) at the age of 9 years and later died of Non-Hodgkin's lymphoma (NHL) at the age of 11yrs. The parents who had a consanguineous marriage were found to be heterozygous carriers of the mutation and were healthy in their thirties. The eldest sibling (CM1d) was found to be a heterozygous

carrier and is healthy at the age of 17 years. This pedigree of this classical CMMRD family is shown in Fig 5.3

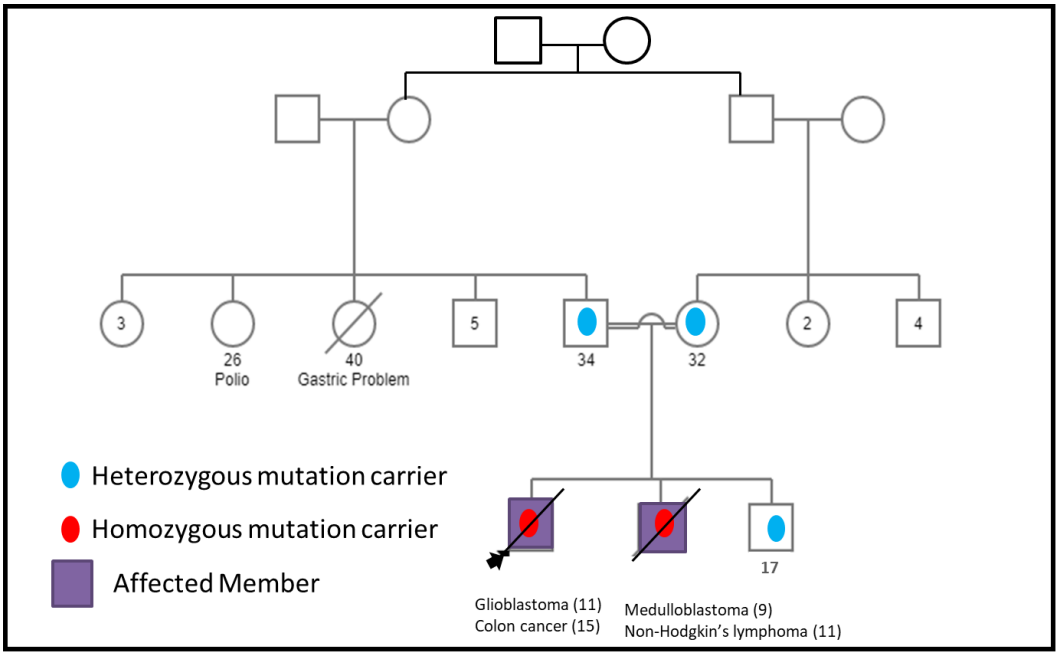


Fig 5.3 Pedigree of a classical CMMRD family (CM1)

In the proband of second family (CM2), a novel biallelic frameshift mutation (c.1487delG; p.G495Gfs*99) in exon 11 of *PMS2* gene was identified which has not been reported in literature or mutation databases such as InSiGHT (LOVD), HGMD, ClinVar, NCBI-dbSNP and Exome Aggregation Consortium (ExAC). The proband developed T-cell acute lymphoblastic leukaemia (T-ALL) and alveolar soft part sarcoma (ASPS) at 6 years and 8 years respectively and died within few months of developing the second malignancy. The sister (CM2c) of proband who was not tested was first diagnosed with T-ALL at the age of 8 years and Glioblastoma at 12 years of age and died 2 years later. Both the parents (CM2a & CM2b) were found to be heterozygous carriers of the mutation. Considering that both siblings had characteristic CMMRD associated

double primary cancers, the sister though not tested in all likelihood was also a biallelic carrier of the mutation identified in her brother. There is a strong history of haematological malignancies and LS-associated malignancy in the family with 5 additional affected members who could be putative carriers of the mutation. For the above two families (CM1 & CM2) a clinical diagnosis of Li-Fraumeni Syndrome (LFS) was initially made based on the tumour spectrum in the family, but TP53 gene analysis revealed no pathogenic or likely pathogenic mutation. Later, they were suspected to have CMMRD syndrome and *PMS2* gene analysis was carried out thereby highlighting the importance of syndromic differential diagnosis.

A known homozygous nonsense pathogenic mutation, c.2404C>T (p.R802X) in exon 14 of *PMS2* was identified in 2 unrelated families. In the third family (CM3), the proband (homozygous PMS2 mutation carrier) is an 11 year old female affected with glioblastoma (GBM) and multiple CALs. Her parents (CM3A & CM3B), who were healthy at the time of testing, were found to be heterozygous carriers of this mutation. In the fourth family (CM4), the proband (homozygous PMS2 mutation carrier) is a 10 year old female affected with rectal cancer. This pathogenic nonsense mutation inserts a premature stop codon at codon 802 and has previously been described in the ClinVar database as a pathogenic mutation with respect to Lynch syndrome and in literature as a Pakistani founder mutation [280]. In the fifth family (CM5) there are three siblings affected with TLL, ALL and mediastinal tumour at the age of 15yrs, 4yrs and 7yrs respectively. All siblings had Café-au-lait spots and are a result of consanguineous marriage. All these features were strongly suggestive of CMMRD and a *PMS2* biallelic pathogenic mutation (c.325dupG; p.E109Gfs*28) in *PMS2* Exon 4 was identified. The others 2 siblings (CM5A & CM5B) with characteristic childhood CMMRD cancers are

presumed to be biallelic carriers of the same mutation. In the sixth family (CM6), the proband is a 4 year old male affected with GBM and has multiple CALs. IHC analysis showed loss of MLH1 and *PMS2* protein expression on both tumour and normal sample thereby raising a suspicion for a germline silencing alteration of these genes. *PMS2* genetic analysis revealed a novel homozygous nonsense pathogenic mutation, (c.478C>T; p.Q160X) in Exon 5. This pathogenic nonsense mutation inserts a premature stop codon at codon 160 and has not been reported in literature or any mutation databases.

The proband of sixth family (CM7), is a 5 year old female with GBM along with corpus callosal agenesis and multiple CALs. IHC analysis showed loss of MSH2 protein expression on both tumour and normal sample thereby suggesting germline silencing alteration of MSH2 gene. MSH2 genetic analysis revealed a novel biallelic frameshift mutation (c.221_231del) in exon 2 of MSH2 gene which has not been reported in literature or mutation databases. The brother of the proband had CMMRD associated high grade glioma (HGG) at the age of 15yrs and CALs and therefore presumed to be biallelic carriers of the same mutation (CM7A).

In addition , we also identified a 30y old female with MSI-H endometrial cancer and a family history of colon cancer suggestive of LS was analysed for the four MMR genes, *MLH1*, *MSH2*, *MSH6* and *PMS2* however when no pathogenic or likely pathogenic mutation was identified in these genes. Multigene NGS panel testing (with Sanger sequencing for confirmation) revealed a novel homozygous likely pathogenic mutation (c.320delT, p.Val107GlyfsTer11) in exon 2 of *MLH3* gene.

The median age of diagnosis for patients with brain tumours was 7.5 years (4-11 years) and haematological malignancies is 7 years (4-12 years). Among the 9 patients who survived their first malignancy, 6 developed a second malignancy and 4 out of 6 patients

(67%) died at a median age of 13 years (8-16 years). Median survival after the diagnosis of the first malignancy was 48 months. The genotype-phenotype data on these 7 families are given in Table 5.4.

Table 5.4 Clinical features and mutation spectrum of patients suspected of CMMRD in an Indian cohort

Family (Proband/ Relative)	Known or *Novel Mutation	Zygosity	Malignancy (age at diagnosis)	CALs (Y/N)	Other variants identified in MMR genes	Other genes tested
CM1 (Proband)	PMS2 Exon 11 c.1500delC; p.V501Wfs*94	Homozygous	GBM (11y); Colon cancer (15y)	NA	PMS2 [Exon 14 c.2324A>G, p.N775S (Benign)]	No mutation in TP53
CM1A (Brother 1)	c.1500delC; p.V501Wfs*94	Homozygous	MBM (9y); NHL (11y)			
CM1B (Father)	c.1500delC; p.V501Wfs*94	Heterozygous	Healthy			
CM1C (Mother)	c.1500delC; p.V501Wfs*94	Heterozygous	Healthy			
CM1D (Brother 2)	c.1500delC; p.V501Wfs*94	Heterozygous	Healthy			
CM2 (Proband)	PMS2 Exon 11 *c.1487delG; p.G495Gfs*99	Homozygous	ALL (6y); ASPS (8y)	N		No mutation in TP53
CM2A (Father)	c.1487delG; p.G495Gfs*99	Heterozygous	Healthy			
CM2B (Mother)	c.1487delG; p.G495Gfs*99	Heterozygous	Healthy			
CM2C (Sister)	Obligate carrier		ALL (8y); Astrocytoma (12y)			
CM3 (proband)	PMS2 Exon 14 c.2404C>T, p.R802X	Homozygous	GBM (9y; 10y)	Y; >2	PMS2 [Exon 14 c.2324A>G, p.N775S (Benign)]	NT
CM3A (Father)	PMS2 Exon 14 c.2404C>T, p.R802X	Heterozygous				
CM3B (Mother)	PMS2 Exon 14 c.2404C>T, p.R802X	Heterozygous				
CM4 (Proband)	PMS2 Exon 14 c.2404C>T, p.R802X	Homozygous	Rectum (10y)	N	PMS2 [Exon 7 c.780C>G HMZ, Exon 14: c.2324A>G & c.2445+9A>C]	NT

CM5 (Proband)	PMS2 Exon 4 c.325dupG; p.E109Gfs*28	Homozygous	T-LBL (4y, 8y)	Y; >15	PMS2 [Exon 4 HMZ c.288 C>T, p.A96A; Intron 11 HMZ c.2006+6 G>A; Exon 14 c.2324A>G, p.N775S (Benign)]	NT
CM5A (Brother)	Obligate carrier		ALL (4y); Mediastinal tumour (5y)	Y		
CM5B (Brother)	Obligate carrier		Mediastinal tumour (7y)	Y		
CM6 (Proband)	PMS2 Exon 4 *c.478C>T; p.Q160X	Homozygous	GBM (4y)	Y; multiple		No mutation in TP53
CM6A (Mother)						
CM7 (Proband)	MSH2 Exon 2 *c.221_231del	Homozygous	GBM (5y)	Y; multiple	MSH2 [c.216+9 C>G, c.1661+12 G>A]	
CM7A (brother)	Obligate carrier		HGG (6y)			
CM8 (Proband)	MLH3 Exon 2 *c.320delT (p.Val107GlyfsTer11)	Homozygous	Endometrium (30y)	NA		No mutation in MLH1, MSH2, MSH6, PMS2, EPCAM, PMS1, MSH3

Abbreviations- GBM: Glioblastoma, MBM: Medulloblastoma, NHL: Non-Hodgekin's Lymphoma, ALL: Acute Lymphobastic leukemia, ASPS: alveolar soft part sarcoma, T-LBL: T-cell Lymphoblastic Lymphoma, HGG: High Grade Glioma, CALs: Café-au-lait spots

5.8 Discussion

In the last decade CMMRD has emerged as a very rare childhood cancer predisposition syndrome due to constitutional defects in the mismatch repair machinery due to biallelic homozygous or compound heterozygous mutations in one of the four MMR genes- *MLH1, MSH2, MSH6* or *PMS2* [248]. Most of the genotype-phenotype data on CMMRD is from small case series [266, 282-284] and their pooled analysis [264, 267, 270, 275]. There are no large studies from single centre and there is very scant information from South Asia with only few families reported from Pakistan [266, 280] and India [281].

With its phenotypic features being quite broad [285], CMMRD is still an under recognized clinical syndrome with no clear understanding of the clinical manifestations, diagnostic criteria and cancer prevention strategies. All the CMMRD associated features like CALs, parental consanguinity, early onset of tumours are mostly indicative and not confirmatory to diagnose the disease. It is therefore important to identify CMMRD patients especially in developing countries like India due to their high prevalence of consanguinity, to provide proper patient management and care. The most widely used criteria for clinical diagnosis of CMMRD is a three point scoring system proposed by Care for CMMRD (C4CMMRD) consortium [264]. In our cohort, the CMMRD score showed good correlation with germline biallelic MMR gene mutation being identified in 6/12 probands with CMMRD score of ≥3 and only 1 proband with the score of <3.

We report detailed genotype and phenotype of 8 confirmed and 4 presumed carriers of biallelic pathogenic MMR gene mutations carriers in 7 Indian families. Four of the 7

pathogenic mutations identified were novel and never reported in any database or literature.

As per the review of 146 cases, there is a preponderance of *PMS2* gene mutations (60%) in CMMRD syndrome [264]; however *PMS2* gene analysis is often neglected due to technical difficulty in detecting *PMS2* mutations as it shares homology with several pseudogenes resulting in under reporting of *PMS2* mutation carrier patients. In our cohort also 5/7 (71%) families had biallelic mutation in *PMS2* gene. In addition, due to the phenotypic overlap with Neurofibromatosis-1 (NF-1) and Leigus syndrome [285] and overlap of tumour spectrum with several other syndromes including FAP, MAP, LFS many patients with CMMRD are misdiagnosed and therefore are not offered proper genetic testing, they might undergo inappropriate surveillance protocols and early detection of CMMRD associated tumours might be missed [267]. In our cohort, diagnosis of LFS was made initially in 2 families however no pathogenic mutations in TP53 were identified. Identification of biallelic pathogenic *PMS2* mutations in these 2 families highlights the importance of making a syndromic differential diagnosis and genetic screening of *PMS2* gene in unexplained childhood cancer patients.

The tumour spectrum is similar to what is reported in literature with high grade gliomas 9(6/18) and haematological malignancies (6/18) being the most common type of cancers. The median age of diagnosis for brain tumours and haematological malignancies was 7.5 years and 7 years respectively which in concordance with the reported literature. The median survival after the diagnosis of the first malignancy was 48 months thereby highlighting the aggressiveness of the disease and importance of timely diagnosis.

We also report a novel biallelic *MLH3* frameshift mutation in a 30 year lady with endometrial cancer and family history of colon cancer. The syndromic diagnosis of LS was made. However no pathogenic mutation was identified in any of the four MMR genes usually tested but a biallelic mutation was identified in the *MLH3* gene. MLH3, as a DNA mismatch repair gene was identified as mammalian homologue of yeast which maps to chromosome band 14q24.3 and encodes a protein that interacts with MLH1 and repair short insertion-deletion loops single-stranded DNA [286, 287]. There are few likely pathogenic missense and truncating mutations of MLH3 with unclear clinical significance reported in InSiGHT database. *MLH3* mutations comprise less than 2% of LS cases [288], however no biallelic mutations in *MLH3* has been reported in CMMRD [254, 258, 263, 264]. Our results suggest a possible link between biallelic germline likely pathogenic mutation of *MLH3* and CMMRD syndrome. Due to unavailability of parents' sample, the inheritance of this mutation could not be established. Absence of any other pathogenic or likely pathogenic mutation in other LS-associated genes on a large NGS hereditary cancer panel analysis is suggestive of the causative role of this *MLH3* mutation in this patient with a characteristic early onset LS associated cancer. Our data is supported by another report of a biallelic *MLH3* mutation in a distinct polyposis syndrome, where a germline nonsense mutation was identified in 4 patients with polyposis syndrome and a distinct clinical and molecular phenotype [289]. In another report of 57 endometrial patients, 6 likely pathogenic mutations of *MLH3* were identified which suggests a possible role of *MLH3* in endometrial cancers [290]. CMMRD is characterized by variable penetrance, variable age of onset and diverse clinical manifestations which include many LS-associated tumours like colorectal cancer, endometrial cancer and uterine cancer.

Therefore we suggest *MLH3* gene analysis in CMMRD suspected patients in whom no pathogenic mutation is identified in the 4 major MMR genes.

In summary, we report the first comprehensive study on 7 CMMRD families which include 8 confirmed and 10 presumed carriers of biallelic germline pathogenic or likely pathogenic mutation in MMR genes which include 4 novel mutations. Our study has added to the existing knowledge of CMMRD genotype and phenotype with an emphasis on the importance of syndromic differential diagnosis to avoid any misclassification and proper management of CMMRD patients and their family. The CMMRD scoring system is also validated in our cohort. In addition, we also report a novel *MLH3* associated phenotype in a CMMRD patient which needs to be explored further on additional cohorts to establish the significance of *MLH3* as a possible cause of cancer predisposition in CMMRD patients.

CHAPTER 6

EARLY ONSET SPORADIC COLORECTAL CANCER

6.1 Sporadic colorectal cancer

Sporadic colorectal cancers are the subset of cancers that arise in the colon and rectum without any notable family history [291]. It comprises around two-thirds of all CRC cases. In the Caucasian population, the median age of onset of sporadic colon cancer is 68 years and 72 years in male and female and median age of patients with rectal cancer is 63 years in both genders. It occurs more frequently in the distal colon comprising approximately 70% cases [52].

Sporadic CRC is often a somatic genetic disease which develops either due to abnormality in the local colonic environment or background genetic makeup or a combination of both [291]. Genetically, sporadic CRC develops by a progressive accumulation of genetic alterations in the tumour suppressor genes and oncogenes. CRC usually develops over few decades due to this progressive accumulation of alterations and undergo a stepwise progression from normal colonic epithelium to an adenomatous intermediate and finally adenocarcinoma, the so-called "adenoma-carcinoma sequence theory".

6.1.1 Colorectal adenoma to carcinoma progression

In the normal colorectum, homeostasis is maintained by a continuous replenishment of undifferentiated epithelial or transit cells from the intestinal crypts (crypts of Liberkuhn) [292]. This replenishment process involves rapid and continuous proliferation of the epithelial cells of the crypt and their subsequent migration along the crypt-villus axis. This huge replenishment of epithelial cells requires rapidly dividing multipotent intestinal stem or progenitor cells which are present at the base of the intestinal crypt [293]. To ensure this lifelong tissue homeostasis, stem cells must have an unlimited capacity to self-renew and differentiate. Therefore at least one of

the progeny must possess the property of stem cells after each cell division. The terminally differentiated cells are pushed towards the surface of gut lumen by young daughter cells and shed regularly. This continuous process of shedding and self-renewal to maintain tissue homeostasis is under strict regulation of a small number of evolutionary highly conserved pathways, among which Wnt signalling pathway is a key player [294]. The disruption of this tightly regulated process occurs due to accumulation of mutation and it has been hypothesized that the first mutational hit occurs in the stem cells of the crypt followed by several mutagenic events over years. The mutated stem cells undergo rapid proliferation giving rise to daughter cells which ultimately give rise to cancer cells that have lost their self-renewal capacity [294]. These daughter cells are pushed towards the lumen at a rapid rate disrupting the sac like architecture of crypt resulting in small outgrowth in the luminal wall, termed as adenomas. These adenomas have malignant potential and upon accumulation of further aberrations transforms into colorectal adenocarcinoma.

6.1.2 Colorectal tumourigenesis

Colorectal cancer develops by deregulation of multiple pathways through independent genomic events leading to the loss of key regulatory mechanisms resulting in proliferation, invasion and metastasis. Decades of research has identified three major distinct genetic pathways implicated in the development of CRC:

1. Chromosomal instability (CIN)
2. Microsatellite instability (MSI)
3. CpG island methylator phenotype (CIMP)

Chromosomal Instability (CIN): Approximately 70% of sporadic CRC follows the CIN pathway, which refers to rapid gain or loss of a part or whole of chromosome

resulting in an imbalance in chromosome number i.e. aneuploidy. This chromosomal imbalance along with progressive mutational activation of tumour suppressor genes and oncogene is referred to as chromosomal instability [295]. Approximately 70% of sporadic CRCs develop along the CIN pathway. CIN pathway is based on the adenoma-carcinoma sequence theory proposed by Fearon and Vogelstein in 1990. The key genetic changes in the CIN pathway include activation of proto-oncogenes like *KRAS*, c-*Myc*, c-src, inactivation of tumour suppressor gene such as *APC* gene, *TP53* gene and allelic loss of 18q [40].

Fearon and Vogelstein model

This model is derived from studies based on 172 colorectal specimens including 80 adenomas representing various stages (early, intermediate and late adenomas) and 92 carcinomas from both FAP and non-FAP patients [36]. The model proposed that there are four major genetic alterations that occur in stepwise manner resulting in clinical progression of CRC tumours. The genetic model by Fearon and Vogelstein is represented in Fig 6.1

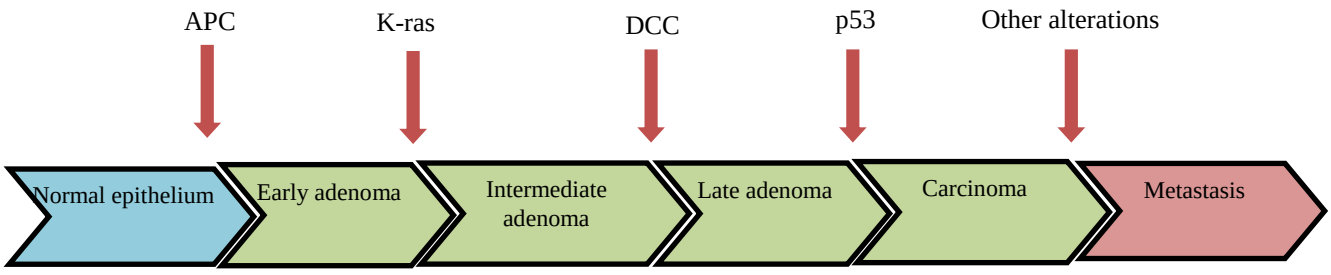


Fig 6.1 Genetic model of colorectal carcinogenesis (Fearon & Vogelstein)

Key events in the CIN pathway

Loss of function of the APC gene is one of the earliest events in sporadic CRC progression. APC function is lost either by genetic disruption or hypermethylation of its promoter which result in activation of the Wnt/β signalling, the key event for adenoma initiation [296, 297]. Somatic mutations in APC is reported in 5% of dysplastic ACF, 30%–70% of sporadic adenomas, and approximately 70% of sporadic colorectal tumour [298-300]. In as many as 50% of CRCs where APC gene is intact, gain of function mutations in *β-catenin* (CTNNB1) activate the Wnt signalling thereby reflecting its importance [295].

Another key event in the CIN pathway is oncogenic gain of function mutation in *KRAS* which is reported in 30-50% of CRC [301]. Approximately 90% of *KRAS* mutations occur in codon 12 and 13; additionally mutations in codon 61 and 146 are also reported though in lower frequencies [302]. Activated RAS regulates multiple cellular processes through different signalling pathways of which the most characterized is Raf–MEK–ERK pathway, which is involved in the control of cell cycle progression [303]. Mutations in *KRAS* impairs its GTPase activity keeping it in active state thereby resulting in constitutive activation of its downstream signaling such as *PIK3CA* which translocate to the plasma membrane to generate phosphoinositol lipids. *PIK3CA* has also been shown to play a role in CRC development.

Loss of function of *TP53* gene and 18q LOH has also been shown to be major contributors to the CIN phenotype [40]. In non-hypermutated CRC, TP53 mutations rate is 55-60% as opposed to ~20% in hypermutated tumours which are mostly driven by MSI pathway. *TP53* alteration frequency is 15-30% in adenomatous polyps as

opposed to 80% in advanced colorectal carcinoma thereby suggesting the fact that loss of TP53 mutation is a late event in CRC progression [304]. Allelic loss at chromosome 18 (18q) is an important genetic alteration in the CIN pathway. One of the candidate genes at this locus is *SMAD4* which is reported to be mutated in 10% of sporadic CRCs [305]. Somatic *SMAD4* mutations are more frequent in later stages of CRC development [306], and LOH at the *SMAD4* locus has been proposed to be a worse prognostic marker for patient survival [295].

Microsatellite Instability Pathway (MSI): MSI refers to genetic hypermutability caused due to a defective mismatch repair system. It is seen in 10-15% of sporadic CRCs [295]. Microsatellites, that are repeat sequences present throughout the genome, are prone to DNA replication errors due to polymerase slippage. These regions accumulate mutations due to abnormally functioning MMR resulting in microsatellite instability and subsequently lead to development of sporadic CRC [30]. It is the most common tumourigenesis pathway in Lynch syndrome (LS), a hereditary form of CRC. Loss of MMR function can occur in two ways, either by mutation in one of the MMR genes or by hypermethylation of the promoter region of hMLH1 gene. The latter being a more common mechanism in sporadic CRCs accounting for 90% of all MSI-H sporadic CRCs [307]. Most MSI-H sporadic CRCs are diploid and lack LOH.

Though a small proportion of hereditary CRC, LS and sporadic CRC are driven by the MSI pathway, there exist various genetic differences among them. APC mutation is one of the earliest events in MSI-H sporadic CRCs as opposed to CTNNB1 mutation in LS [308]. Additionally, *BRAF* (V600E) mutations have been found to be more frequent in MSI-H sporadic CRC than LS [309]. *KRAS* mutations are rare in MSI-H

sporadic CRCs [310]. These differences suggest that these two types of MSI-H CRC, sporadic CRC and Lynch syndrome CRC which evolve via alternate pathways.

CpG Island Methylator Phenotype (CIMP) pathway: CIMP pathway is characterized by widespread CpG island methylation. CpG sequences usually cluster in a region known as island present in the 5' promoter region of several genes. Methylation of cytosine in these CpG islands result in silencing of respective protein expression. *hMLH1* promoter hypermethylation is most common in sporadic CRCs therefore CIMP-positive CRCs are often MSI-H [52]. CIMP is associated with BRAF mutations in both MSI and MSS CRCs. Hypermethylation result in loss of expression of other genes as well which are involved in CRC carcinogenesis like p16 (cell cycle regulator), ER (growth suppressor), THBS1 (angiogenesis inhibitor), and MGMT (DNA repair gene) [311]. There exist few panels for defining CIMP like Issa panel consisting of p16, hMLH1, MINT1, MINT2, and MINT31 markers and the new Laird panel comprising of CACNA1G, IGF2, NEUROG1, RUNX3, and SOCS1markers however a unified panel remains to be established.

These mechanisms are not mutually exclusive and may coexist therefore sporadic CRC may develop through three distinct clinicopatholoigical evolution pathway [52]. The first is the traditional pathway which starts with APC mutation resulting in formation of tubular adenomas from normal mucosa and subsequently develops in CRC with TP53 mutations and CIN. The second is the serrated pathway which is characterized by CIMP and BRAF mutations resulting in the formation of serrated polyps from normal mucosa and ultimately CRC due to loss of MLH1 function and MSI. The third one known as the alternative pathway, starts from normal mucosa via villous, partly serrated adenomas harbouring mutations in *KRAS*, BRAF, *APC* and

CIMP-L and subsequently develops colon cancer (with CIMP) [52]. The three pathways are represented in Fig 6.2

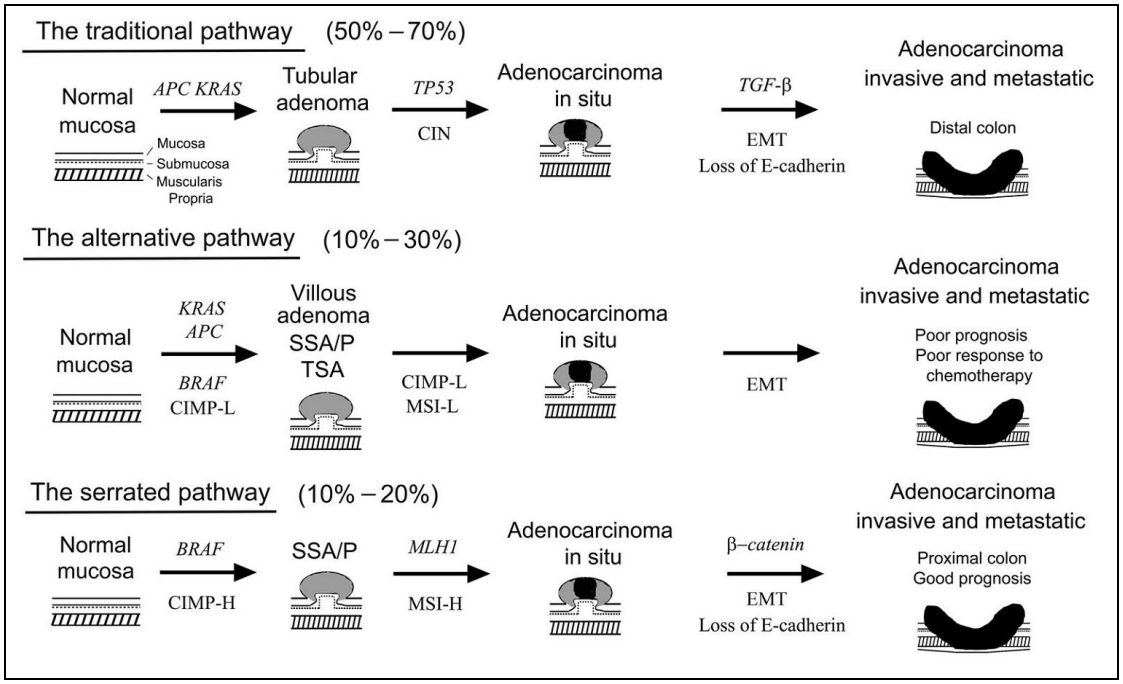


Fig 6.2 Three evolutionary pathways of sporadic colorectal tumourigenesis

(Taken from Ref [52])

6.2 Sporadic early onset colorectal cancer

There is a huge geographical variation in the incidence rate of CRC with developed nations in the Western world carrying most of the CRC burden. CRC being an age-related disease mostly develop in the 6th decade of life with early onset of CRC seen mostly in the hereditary setting of FAP and LS. With increased awareness, lifestyle modification and screening programmes, the incidence of CRC in individuals above 50 years of age has declined significantly over the last decade [312, 313]. Despite an overall decrease in CRC incidence and mortality, a significant increase in the incidence of CRC in young individuals with no notable family history (early-onset sporadic, EOS-CRC) has been observed worldwide [314]. Analysis of data from both

USA (SEER) and UK (Office for National statistics) revealed an increased incidence of rectal and left sided tumour in the age group below 70 years [315]. This trend was well established by the early 2000's and several studies have been undertaken to better understand this subset of CRC [316]. In contrast to the Western population, data from developing countries like India is sparse. The few etiological studies that have been done, report a disproportionate increase in the incidence of young onset CRC in India with the incidence rates even higher than the Western population [53, 317]. The incidence of rectal cancer was more than colon cancer in young individuals in the developing countries, including India [53, 317-319].

6.2.1 Clinicopatholoigical features of EOS-CRC

In an attempt to understand the biology of EOS-CRC, different population based studies have evaluated the clinicopathological and molecular origin of this subset of CRC. These studies revealed a distinctive tumour location, stage at presentation and histologic features in young onset CRC. Young onset CRC tumours were found to be more prevalent in the sigmoid colon or rectum (50-80%) as compared to late onset CRC [320]. CRC in young individuals exhibit histologically unfavourable disease characteristics such as poor differentiation, advanced stage, mucinous/or signet ring cell carcinoma and therefore are typically associated with worst outcome.

Studies from India who have evaluated the clinicopathological features of EOS-CRC reported similar features such as poorly differentiated, advanced stage, mucin secreting aggressive tumours with poor survival compared to other subgroups [321-323]. Rectal tumours in young population was found to be less resectable when compared to the older population either due to advanced stage or physical adherence

to the pelvic bone, therefore leaving radiotherapy and chemotherapy as the only treatment options [323].

The reasons for these histologic differences between young and old onset CRC is not known yet, however a better understanding of their molecular biology may give some explanation.

6.2.2 Age-cut off of EOS-CRC

While young age at onset of CRC is a hallmark of hereditary CRC syndromes, in recent years there is a distinct trend of sporadic CRC among young individuals. There is a wide variation in the age cut off adopted by different centres for defining early onset CRC but majority of studies have used an age cut off between 40-50 years [324]. Few studies have reported difference in the histopathological characteristics between early and late onset CRC and also among different subset of early onset group like <30 years and 30-50 years age group. One such observation is that survival is worst in patients younger than 30 years old, whereas it is comparable or even better among patients between 40 and 50 when compared with those older than 50 years [325]. Therefore, it is important to define the age groups to have a better understanding of non-hereditary CRC in young patients. The differences observed in the pathological features and clinical outcome at different ages may be due to difference in the underlying molecular mechanisms.

6.2.3 Molecular profiling of EOS-CRC

Molecular events in colorectal cancer occurring in young individuals have not been well elucidated and very few studies have compared the molecular profile of early and late onset CRC cases. Despite of a higher incidence of young onset CRC in the developing countries, very few studies have been conducted on these patients, limiting

our understanding of this poor prognosis subset of CRC. The first combined genetic, epigenetic and clinical profiling of rectal cancer in India was carried out in 2014 which revealed that these tumours are microsatellite stable (MSS), CIMP-low and exhibit differential *KRAS* mutations. A significantly lower frequency of *KRAS* codon 12 and 13 mutations and higher frequency of codon 15 and 18 mutations were observed in early onset rectal cancers versus late onset cases [321]. Another study of molecular genetic profiling of CRC revealed that in a significant proportion of early onset CRC, the tumour is not driven by the known canonical pathways i.e. CIN or MSI. In addition, early onset CRCs does not harbour hotspot *KRAS* mutations as opposed to their late onset counterpart [326]. Similar differential molecular profile is reported recently from the largest study from USA where they carried out a comprehensive analysis of genomic landscape of younger (<40 years) and older onset CRC (≥50 years) [327]. The commonly mutated CRC genes like APC, *KRAS*, BRAF, PIK3CA and FAM123B were found to be more frequently altered in older patients with CRC while TP53 and CTNNB1 alterations were more frequent in younger patients.

However, apart from the lower frequency of the most commonly mutated genes and lesser involvement of the canonical pathways in CRC, the route to tumourigenesis meaning the somatic mutational landscape, other genes that are mutated and the path to malignancy in early onset CRC is poorly understood.

6.3 Hypothesis and objective of the study

The genetic and epigenetic events driving the tumourigenesis of early onset CRC, in the absence of a hereditary influence, has not been fully elucidated. The limited

number of comprehensive molecular genetic analysis studies reported so far for early onset CRC suggests involvement of a non-canonical pathway. This may explain the distinct clinical, pathological and molecular features seen in EOS-CRC. One of the major drawbacks of these studies is the wide variability in the age cut off used to define early onset CRC. Another shortcoming is the lack of clear distinction between early onset hereditary or LS associated CRC and early onset sporadic CRC. Despite several reports showing different molecular profiles for colon and rectal cancer, most studies have analysed them as single entity of colorectal cancer. It is noteworthy that the rise in young onset CRC patients in developing countries like India shows a higher incidence rate of rectal cancer. Additionally, the studies from India are limited to few tertiary hospitals and are centre specific therefore the results could be biased as they are not population based. Studies on sporadic-RC occurring in the young have been limited and our understanding of this subset of CRC is meagre and their survival rate remains poor.

In the light of the aforementioned points, we hypothesize that early onset sporadic rectal cancer follow an entirely distinct genetic pathway that are more heterogeneous and complex to be explained by the three existing canonical pathways. It therefore becomes imperative to identify molecular pathways that drive tumourigenesis in this poorly studied CRC subtype. One of the objectives of this project was therefore to understand the somatic mutational landscape of early onset rectal cancer using a two-step approach. After ruling out the known canonical pathways as a major component of early onset rectal cancer through limited genetic analysis of 10 early onset sporadic rectal cancers, we undertook comprehensive genetic analysis via whole exome

sequencing on a set of 28 early-onset RC to define the somatic mutational landscape and identify any other pathway.

6.4 Results

6.4.1 Clinical characteristics of patients

As mentioned in section 2.3.1 of Chapter 2, this study included advanced stage poorly to moderately differentiated rectal/recto-sigmoid tumours from patients below 40 years of age with no notable family history. Patient samples were segregated into categories based on treatment status:

a. Set I: Rectal or Recto-sigmoid paired tumour/adjacent normal samples from 10 cases obtained during surgery done after pre-operative radio-therapy with or without chemotherapy (median age: 30 years). Two of these samples which were MMR proficient on IHC showed Micro-Satellite Instability on STR profiling and was not subjected to further molecular characterization
b. Set II included 21 treatment naïve paired samples from patients with rectal cancer (median age: 27 years) and 7 paired samples from set I (RS_31, 32, 33, 34, 36, 37)

The clinicopathological details of patients from both sets are summarized in Table 6.1

Table 6.1 Clinicopathological features of EOS-Rectal Cancer patients

	Sr. No.	Case ID	Age at diagnosis	Gender	Cancer site	HPR of tumor
Set 1	1	RS_31	39y	M	Rectum	PDA
	2	RS_32	18y	M	Rectum	PDA
	3	RS_33	30y	M	Rectum	PDA with extracellular mucin
	4	RS_34	27y	M	Recto-sigmoid	PDA with signet ring cell features
	5	RS_35	27y	F	Rectum	PDA with mucinous and signet ring cell features
	6	RS_36	37y	M	Rectum	MDA
	7	RS_37	30y	F	Rectum	MDA
	8	RS_38	37y	M	Rectum	MDA-PDA
	9	RS_39	30y	M	Recto-sigmoid	PDA
	10	RS_40	38y	F	Rectum	PDA
Set 2	1	RS_02	22y	F	Rectum	MDA
	2	RS_03	30y	M	Rectum	Mucin secreting with signet ring cell feature
	3	RS_04	30y	F	Rectum	PDA
	4	RS_05	21y	F	Rectum	MDA
	5	RS_06	36y	M	Rectum	PDA with signet ring cell features
	6	RS_08	29y	F	Rectum	MDA
	7	RS_09	35y	M	Rectum	MDA
	8	RS_10	27y	M	Rectum	MDA
	9	RS_11	25y	M	Rectum	PDA with mucinous and signet ring cell features
	10	RS_12	23y	M	Rectum	MDA with mucinous and signet ring cell features
	11	RS_13	33y	M	Rectum	PDA with extracellular mucin
	12	RS_14	32y	M	Rectum	PDA
	13	RS_16	27y	F	Rectum	PDA with mucinous and signet ring cell features

	14	RS_17	35y	M	Rectum	MDA
	15	RS_18	35y	F	Recto-sigmoid	MDA
	16	RS_19	26y	M	Rectum	Mucinous signet ring cell
	17	RS_20	35y	M	Rectum	MDA-PDA
	18	RS_21	21y	F	Rectum	MDA
	19	RS_22	24y	F	Rectum	MDA
	20	RS_23	35y	M	Rectum	PDA with mucinous and signet ring cell features
	21	RS_28	38y	F	Rectum	PDA with mucinous and signet ring cell features

Abbreviation- HPR: Histopathology report, PDA: poorly differentiated adenocarcinoma, MDA: moderately differentiated adenocarcinoma

Two-step approach was used to understand the genetic and molecular profile of early onset sporadic rectal cancer (EOS-RC) in the Indian population as:

1) Screening for the known canonical pathways in Set I tumours to investigate their role in EOS-RC
2) Whole exome sequencing of 15 paired tumour and adjacent normal samples to identify the somatic mutational landscape of EOS-RC

6.4.2 Screening for the known canonical pathways of CRC

6.4.2.1 Screening for CIN in the Indian EOS-RC patients

CIN in CRC patients results from inactivation of *APC* or activating mutations of *β-catenin* gene *(CTNNB1*) causing aberrant Wnt signaling. The first step was to verify whether aberrant Wnt activation is involved in the tumourigenesis of EOS-RC in Indian patients. All 10 tumours exhibited membranous positivity for *β-catenin* protein indicating that these tumours are not driven by the Wnt pathway (Fig 6.3).

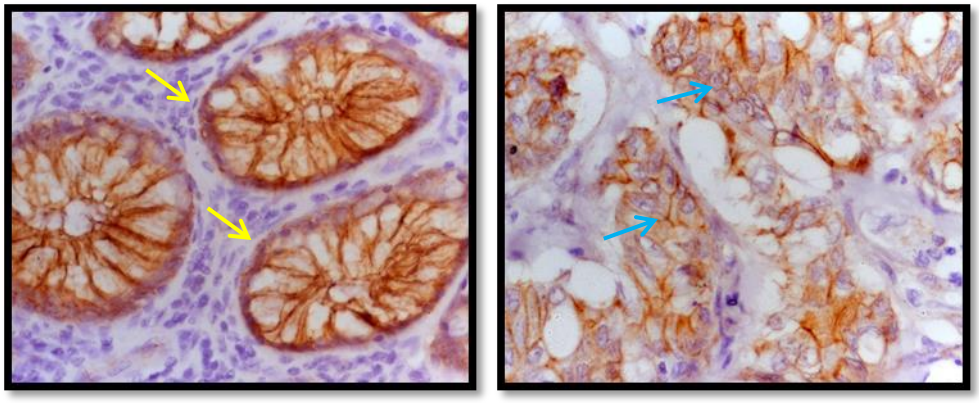


Fig 6.3 Representative photomicrographs of *β-catenin* IHC. Left Panel: Normal colonic mucosa showing membrane staining of *β-catenin* indicated by yellow arrows**.** Right Panel: Rectal tumour showing membranous staining for β-catenin indicated by blue arrows

APC mutations are known to be responsible for more than 80% of Wnt activated CRC. 70% of the inactivating mutations of *APC* has been identified in the Mutation Cluster Region (MCR) located in the central region of the *APC* gene (including codon 1250-1450) spanning approximately 1000bp which overlaps with the region encoding the 20 amino acid repeats (20R) that are beta-catenin-binding sites [328]. Mutations in MCR disrupt *β-catenin* binding by *APC* which subsequently results in constitutively activated WNT signaling in tumours. We screened APC-MCR region in our 10 samples from Set I and identified no mutation in any case.

In some CRC without *APC* mutations activating *β-catenin* mutations that involve missense mutation or, occasionally, deletion of exon 3 serine/threonine residues have been shown to be responsible for aberrant Wnt signaling therefore we screened our Set I tumours for the same. No mutation was identified in any of the 10 cases.

Absence of mutations in *APC* MCR and *β-catenin* in all the cases thereby further confirm our IHC based Wnt evaluation.

Mutation in *KRAS* is considered as an early event in the CIN pathway and has been reported to be mutated in 35%-45% of colon carcinomas [329]. Point mutations in codon 12 and 13 of exon 2 account for nearly 90% of all the *KRAS* mutations identified in CRC. No mutation was detected in any case upon *KRAS* exon 2 screening.

6.4.2.2 Screening of MSI in EOS-RC

We next investigated the role of second major CRC driver pathway, the MSI pathway, in Wnt- samples by carrying out IHC of MMR proteins. All the 10 exhibited nuclear staining for all four MMR proteins (Fig 6.4A-D)

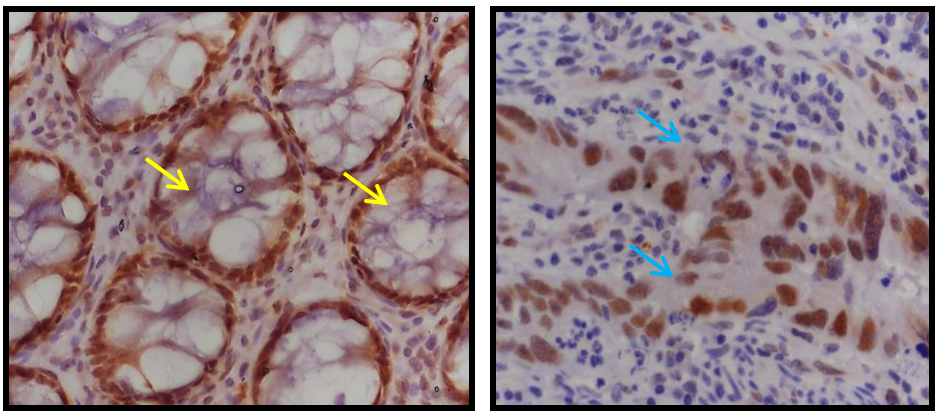


Fig 6.4A Representative photomicrographs of *MLH1* IHC. Left Panel: Normal colonic mucosa showing nuclear staining of MLH1 indicated by yellow arrows**.** Right Panel: Rectal tumour showing nuclear staining for MLH1 indicated by blue arrows.

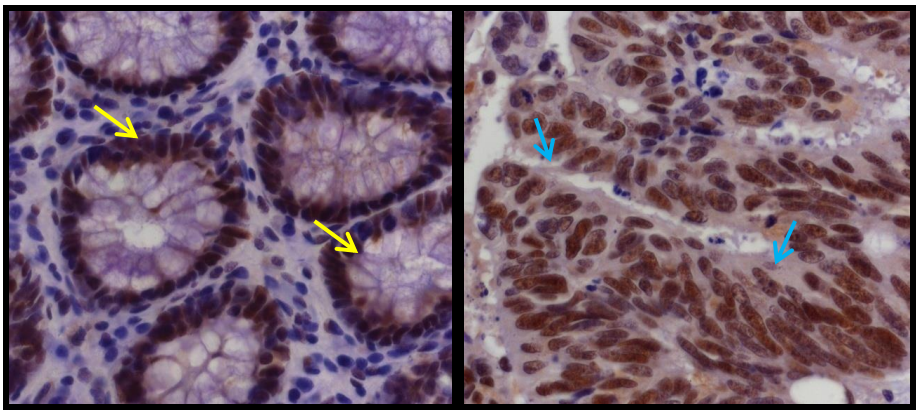


Fig 6.4B Representative photomicrographs of MSH2 IHC. Left Panel: Normal colonic mucosa showing nuclear staining of MSH2 indicated by yellow arrows**.** Right Panel: Rectal tumour showing nuclear staining for MSH2 indicated by blue arrows.

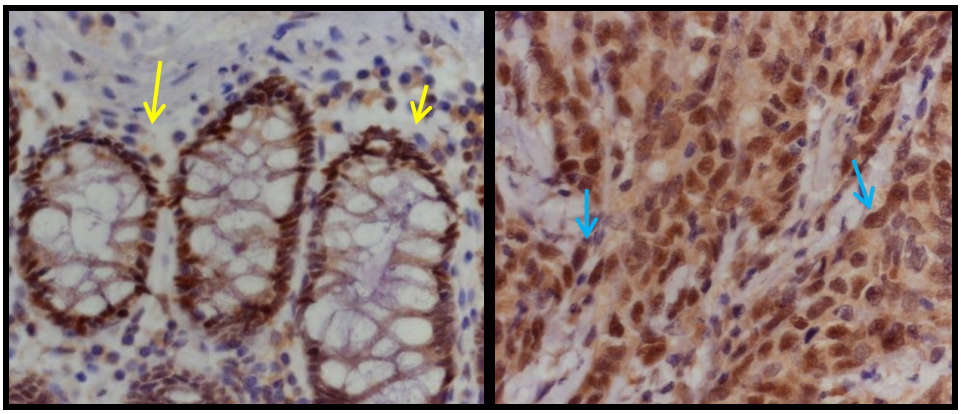


Fig 6.4C Representative photomicrographs of MSH6 IHC. Left Panel: Normal colonic mucosa showing nuclear staining of MSH6 indicated by yellow arrows**.** Right Panel: Rectal tumour showing nuclear staining for MSH6 indicated by blue arrows.

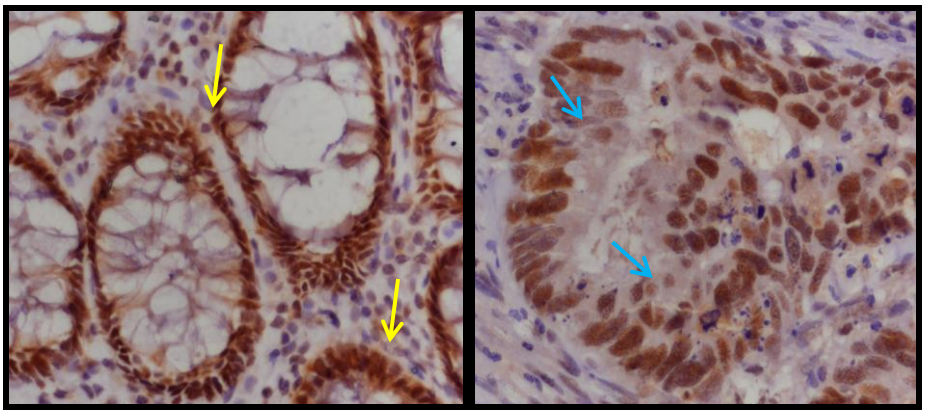


Fig 6.4D Representative photomicrographs of PMS2 IHC. Left Panel: Normal colonic mucosa showing nuclear staining of PMS2 indicated by yellow arrows**.** Right Panel: Rectal tumour showing nuclear staining for PMS2 indicated by blue arrows.

We next investigated the MSI status by PCR based amplification of NCI panel of 5 microsatellite markers. All the 10 tumours were found to be MSS (Fig 6.5).

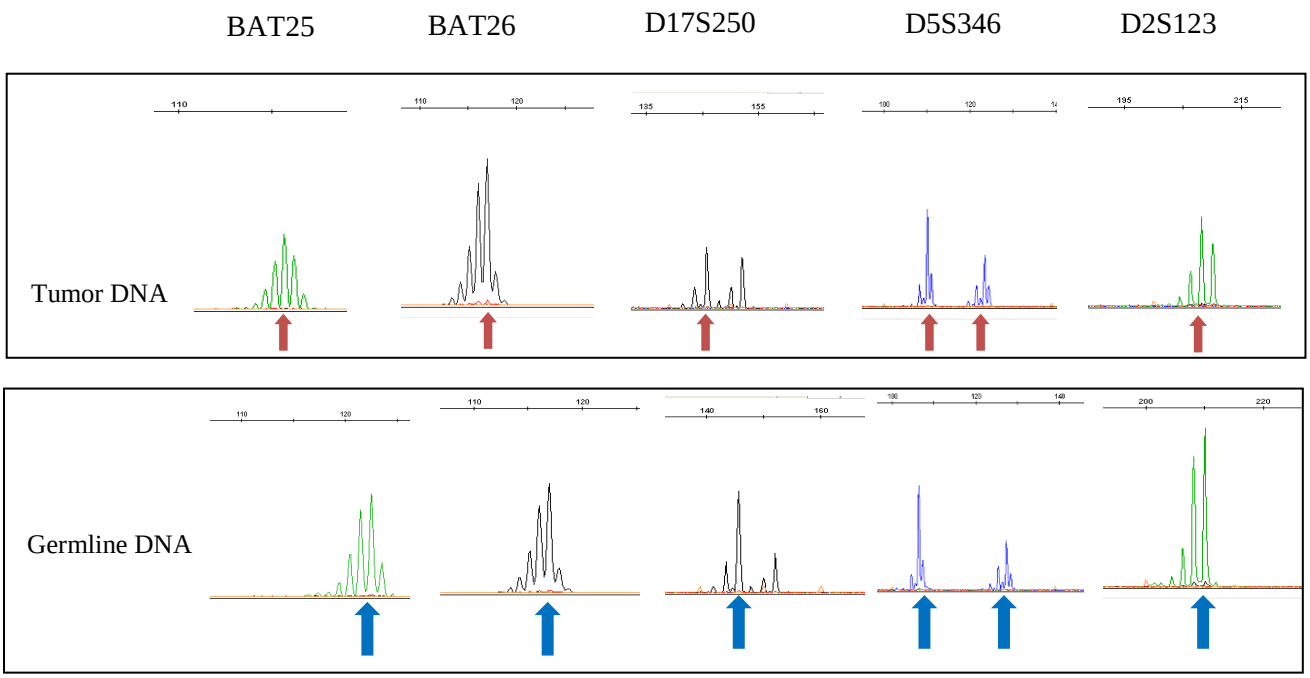


Fig 6.5 Representative Electropherograms of MSI screening

Combined analysis of MMR IHC and STR markers revealed that all these 10 cases are MMR proficient.

hMLH1 promoter hypermethylation is the most common tumourigenesis pathway in sporadic CRCs, however intact nuclear staining of MLH1 protein in these 10 cases further confirm that EOS-RC are not driven by MLH1-associated CIMP pathway. CIMP has been reported to be associated with BRAF mutations in both MSI (30%-50%) and MSS (10%) CRCs with Exon 15, V600E being the most common mutation. No mutation was identified in Exon 15 of BRAF in any of the 10 cases.

The screening results are summarized in Table 6.2

Table 6.2 Summary status of CIN and MSI pathway and key genes in EOS-RC

Sample ID	MMR proficient	MSI analysis	β-catenin nuclear staining	APC MCR	CTNNB1 Exon 3	KRAS Exon 2	BRAF Exon 15
RS_31	Yes	MSS	Yes	WT	WT	WT	WT
RS_32	Yes	MSS	Yes	WT	WT	WT	WT
RS_33	Yes	MSS	Yes	WT	WT	WT	WT
RS_34	Yes	MSS	Yes	WT	WT	WT	WT
RS_35	Yes	MSS	Yes	WT	WT	WT	WT
RS_36	Yes	MSS	Yes	WT	WT	WT	WT
RS_37	Yes	MSS	Yes	WT	WT	WT	WT
RS_38	Yes	MSS	Yes	WT	WT	WT	WT
RS_39	Yes	MSS	Yes	WT	WT	WT	WT
RS_40	Yes	MSS	Yes	WT	WT	WT	WT

Abbreviations:- MSI: Microsatellite Instability, MSS: Microsatellite stable, WT: Wild type (No mutation)

6.4.3 Whole exome sequencing analysis of 28 EOS-RC samples

We performed WES of 28 sporadic rectal adenocarcinoma samples from patients with median age of onset of 27 years. Demutliplexed FASTQ files were used to assess the quality of the sequencing data using the FASTQC software. FASTQC analysis revealed that the reads were of good quality with mean and median base Phred quality score for almost bases in all 28 samples was above 30 (Q>30). Representative FASTQC report image for one of the samples is given in Fig 6.3

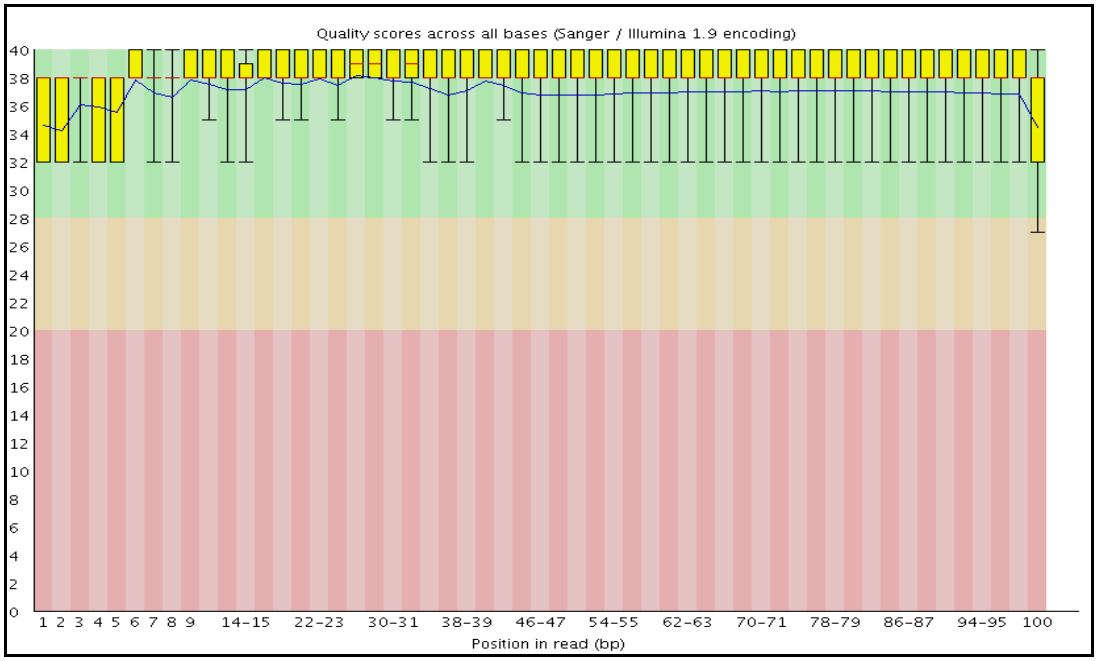


Fig 6.6 FASTQC analysis of raw data shows quality scores above 30 across all bases

After mapping the reads to the reference human genome hg19, nearly 99% of reads were mapped to targets, and the mean target coverage was 90-171 in 24 samples and 40-69 in remaining 4 samples. More than 95% of the bases was covered at least 10X in each sample. More than 80% of the bases were covered at least 30X in each sample except 2 samples covered at 64.76X and 73.66X each. Detailed description of the exome coverage statistics are given in Table 6.3.

Table 6.3 Coverage statistics for WES samples

Sample ID	Total no. of reads	No. of mapped reads	Mapping Rate	Mean Target Coverage	Percent bases covered at 10X	Percent bases covered at 30X
RS_38_T	82187747	82096524	99.89%	171.49	99.47%	96.41%
RS_38_N	74703273	74592083	99.85%	155.31	99.38%	95.1%
RS_37_T	69654314	69537232	99.83%	143.46	99.26%	95.33%
RS_37_N	68450353	68354900	99.86%	146.60	98.77%	91.67%
RS_36_T	60609202	60521134	99.85%	125.91	99.16%	93.04%
RS_36_N	75726928	75633493	99.88%	163.49	99.19%	94.23%
RS_34_T	61800012	61710314	99.85%	126.91	99.15%	93.55%
RS_34_N	69676712	69583899	99.87%	146.89	99.28%	94.18%
RS_33_T	67212958	67099193	99.83%	138.02	99.3%	94.43%
RS_33_N	74054533	73932456	99.84%	151.23	99.4%	95.9%
RS_32_T	65230284	65104835	99.81%	136.77	99.11%	93.61%
RS_32_N	66443334	66342745	99.85%	137.84	99.16%	93.58%
RS_31_T	73829940	73723190	99.86%	154.15	99.39%	95.58%
RS_31_N	65994375	65893957	99.85%	136.97	99.15%	93.7%
RS_28_T	67848477	67754783	99.86%	133.34	95.49%	89.51%
RS_28_N	68287222	68177726	99.84%	133.25	95.52%	89.69%
RS_23_T	73245898	73116182	99.82%	140.61	95.89%	90.51%
RS_23_N	73281078	73166327	99.84%	141.82	95.87%	90.6%
RS_22_T	65756399	65681202	99.89%	129	95.18%	88.53%

RS_22_N	64228759	64157306	99.89%	125.8	95.22%	88.51%
RS_21_T	67950875	67850347	99.85%	131.77	95.47%	89.71%
RS_21_N	68303914	68215292	99.87%	132.78	95.54%	89.83%
RS_20_T	65855732	65748738	99.84%	127.6	95.57%	89.07%
RS_20_N	67103930	67002669	99.85%	129.94	95.63%	89.31%
RS_19_T	64424292	64320054	99.84%	124.2	95.49%	88.90%
RS_19_N	63505006	63389089	99.82%	121.81	95.55%	89.03%
RS_18_T	64259909	64160514	99.85%	131.53	98.95%	93.6%
RS_18_N	65999807	65911990	99.87%	143.56	98.7%	91.4%
RS_17_T	67699010	67608067	99.87%	145.98	98.81%	91.39%
RS_17_N	79139219	79043791	99.88%	171.98	99.29%	94.78%
RS_16_T	67946638	67852520	99.86%	133.43	95.51%	89.69%
RS_16_N	65233306	65142269	99.86%	122.86	95.41%	89.29%
RS_14_N	69159022	69047698	99.84%	143.05	99.33%	95.26%
RS_14_T	64553398	64465372	99.86%	134.14	99.15%	94.12%
RS_13_T	64833898	64742676	99.86%	125.79	95.41%	89.27%
RS_13_N	68998004	68907118	99.87%	134.44	95.61%	90.05%
RS_12_T	67493312	67375984	99.83%	130.65	95.59%	89.25%
RS_12_N	66722806	66622974	99.85%	128.88	95.70%	89.69%
RS_11_T	66543532	66439722	99.84%	129.05	95.59%	89.22%
RS_11_N	62667354	62579788	99.86%	121.60	95.44%	88.46%
RS_10_T	73061721	72962755	99.86%	151.14	99.22%	95.07%
RS_10_N	77968907	77855636	99.85%	164.15	99.24%	94.39%
RS_09_T	75881040	75766521	99.85%	161.78	98.96%	92.68%
RS_09_N	72977344	72905769	99.9%	160.89	98.99%	92.9%

Rs_08_T	65848569	65747497	99.85%	136.24	99.18%	94.6%
RS_08_N	70985079	70901384	99.88%	151.16	99.13%	94.33%
RS_06_T	150747388	150381010	99.76%	166.06	99.28%	97.47%
RS_06_N	37293621	37170760	99.67%	47.30	99.16%	73.66%
RS_05_T	33600027	33469155	99.61%	42.47	94.5%	64.76%
RS_05_N	45805230	45533745	99.41%	60.31	96.80%	82.39%
RS_04_T	58743371	58218466	99.11%	69.40	97.85%	89.68%
RS_04_N	79452399	78982941	99.41%	96.70	98.19%	93.27%
RS_03_T	72442799	72357825	99.88%	141.74	95.73%	90.15%
RS_03_N	90640136	90523515	99.87%	176.94	96.25%	92.23%
RS_02_T	45717384	45451922	99.42%	55.29	96.84%	81.26%
RS_02_N	74232481	73722224	99.31%	92.93	97.99%	92.46%

After processing the variants identified by exome sequencing to include only somatic SNVs and Indels (explained in Section 2.5 in Chapter 2) a total of 29575 high confidence somatic variants were identified within 28 samples analysed. The high confidence somatic exonic variants were further processed with latest version of ANNOVAR software to functionally annotate. After excluding all the variants which were present in non-coding region (intronic, intergenic, ncRNA, UTR etc.), the silent synonymous variants, unknown significance variants and the ones which were present at a very high frequency in the dbSNP, ESP6500 or 1000Genome databases from further study, we restricted our study to only 6727 non-synonymous known and novel variants.

Two tumour sample (RS_21 & RS_33) had a very high number of 1640 and 3262 nonsynonymous exonic variants with a tumour mutational burden (TMB) of 32.8/Mb and 65.24/Mb respectively. In the remaining 26 samples, the median frequency of

nonsynonymous exonic variants were 67 (range: 13-160) resulting in median TMB of 1.34 per Mb (0.26-3.2). As per TCGA, a tumour is designated as hypermutated if the TMB is >12 mutations/Mb and non-hypermutated, if TMB is <8.24 mutations/Mb. Therefore, samples RS_21 with 1640 and RS_31 with 3262 variants are hypermutated tumours.

A detailed summary of variant statistics of all the samples is given in Table 6.4.

Table 6.4 Variant statistics of WES for all samples

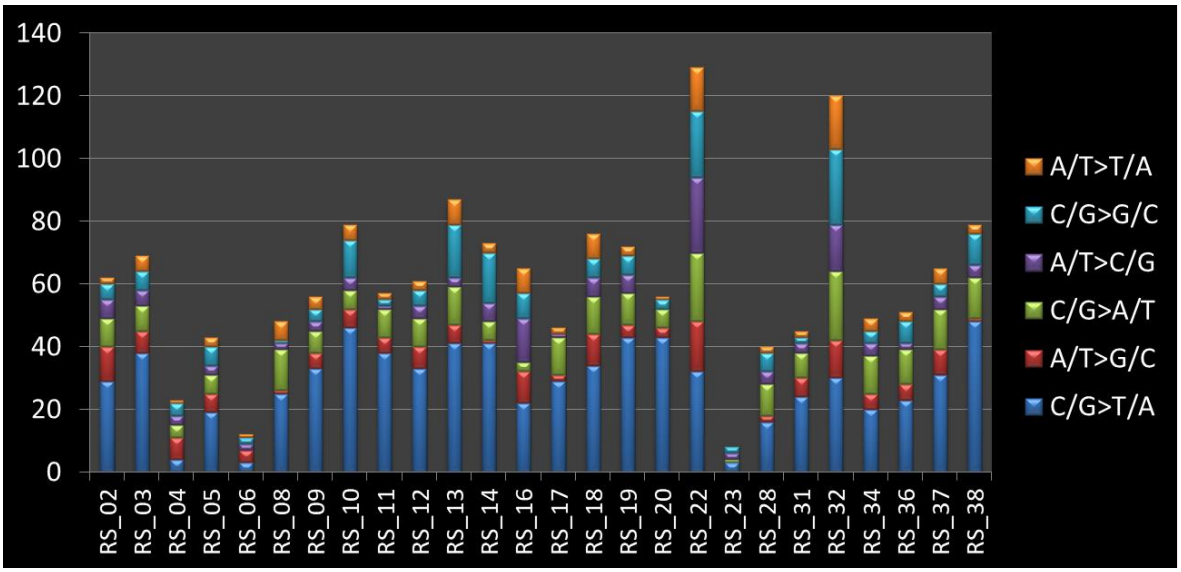
S. No.	Sample	Total no. of somatic-hc variants	All exonic variants (synonymous variants)	Filtered known/novel variants
1	RS_02	157	109 (17)	90
2	RS_03	209	111 (29)	81
3	RS_04	54	50 (10)	40
4	RS_05	109	73 (20)	52
5	RS_06	50	18 (5)	13
6	RS_08	133	74 (18)	56
7	RS_09	156	86 (21)	64
8	RS_10	230	121 (35)	86
9	RS_11	150	94 (23)	68
10	RS_12	178	90 (20)	69
11	RS_13	205	113 (19)	94
12	RS_14	123	91 (35)	55
13	RS_16	150	90 (21)	68
14	RS_17	144	80 (22)	57
15	RS_18	259	135 (48)	86
16	RS_19	187	102 (25)	74
17	RS_20	159	91 (25)	66
18	*RS_21	3164	2115 (458)	1640

19	RS_22	395	188 (27)	160
20	RS_23	37	24 (8)	14
21	RS_28	107	74 (24)	50
22	RS_31	113	73 (16)	55
23	RS_32	382	174 (44)	140
*24	RS_33	22072	6637 (3286)	3262
25	RS_34	144	85 (24)	60
26	RS_36	117	75 (16)	59
27	RS_37	183	105 (26)	78
28	RS_38	208	116 (26)	86

Somatic-hc: somatic-high confidence; *Hypermutated tumour

While defining the somatic mutation spectrum of tumour samples, C/G>T/A transitions were found to be the most common in 23/26 cases. The somatic mutation signatures by the type of nucleotide changes is represented in Fig 6.7

Fig 6.7 Somatic mutation signatures by type of nucleotide change



In order to identify to genes harbouring potentially pathogenic variants that were observed in more one sample we included only splice sites and nonsynonymous missense variants predicted to be pathogenic by ≥5/11 in-silico tools. The analysis revealed 15 genes that were identified in more than two samples.

We compared the most frequently mutated genes in the Indian Early Onset Rectal Cancer with the USA-TCGA Late onset Rectal Cancer. This comparison reveals a distinct mutational landscape between the two forms of rectal cancers in these two populations (Fig 6.8 and Fig. 6.9)

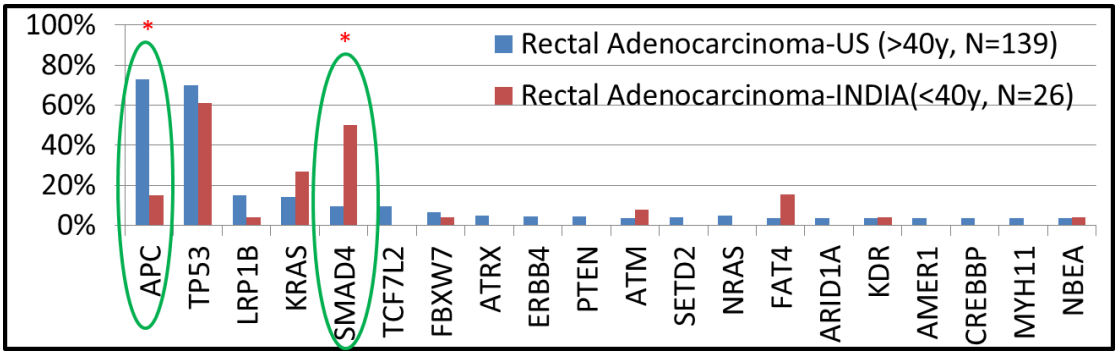


Fig 6.9 Top 15 frequently mutated genes in Indian EOS Rectal Cancer

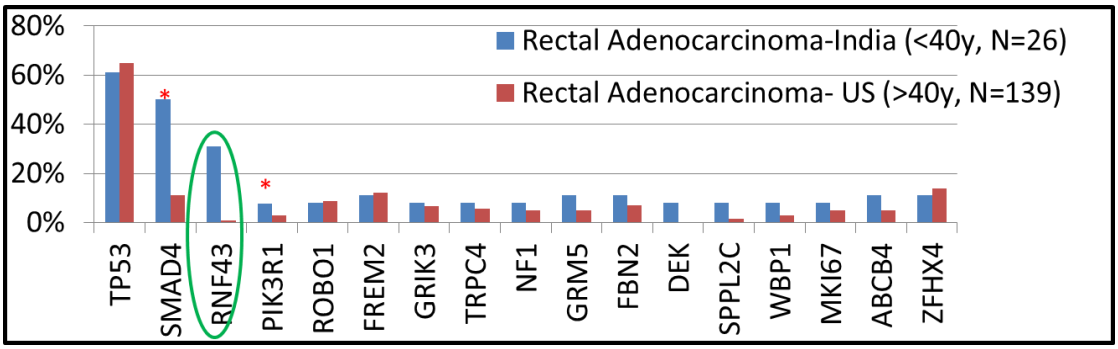


Fig 6.10 Top 20 frequently mutated genes in USA TCGA Late onset Rectal Cancer

The combined comparative analysis of the above 2 set of genes reveals that the somatic mutation distribution of EOS-RC differs greatly from USA TCGA late onset RC. The most significant observation of this study is that the mutation detection rates in the canonical genes also known as “hill” genes are either significantly lower, except TP53 and SMAD4 or mostly absent in our cohort of early onset RC. The mutation rate of TP53 in our cohort was similar to the TCGA rectal cancer data; however the rate of SMAD4 mutation was significantly higher (50%) when compared to TCGA-US late onset rectal cancer cohort (12%).

6.5 Discussion

Despite the low overall age-adjusted CRC incidence in India, the incidence of early onset CRC cancers is disproportionately higher in India along with predominance of rectal cancer [53, 317, 318]. These early onset rectal tumours are associated with unfavourable disease phenotype however there is lack of comprehensive genetic studies on this subgroup of CRC. Only one such study has been reported from India which revealed that known canonical pathways of CRC carcinogenesis, such as CIN and MSI accounts only for a small proportion of early onset sporadic rectal cancer patients [326]. In addition, another study from Indian rectal cancer patients showed that the frequency of alterations in critical CRC genes such as *KRAS*, *TP53* and *BRAF* was significantly lower in this subgroup as compared to late-onset CRC patients [321]. In light of these previous data, we hypothesize that early onset sporadic rectal/recto-sigmoid cancers may have completely different and as yet unknown genetic factors which needs to be elucidated. This may help in establishing novel diagnostic/prognostic/therapeutic regime for young onset patients.

Whole exome sequencing revealed the presence of an average of 67 exonic somatic variants per sample (range 13-160) in 26/28 with a mean TMB of 1.32 per Mb (0.28-1.88). This conforms that EOS-RC tumours are non-hypermutated. Two of these samples were hypermutated with 1640 and 3262 variants and mean TMB of 32.8/Mb and 65.24/Mb respectively. It is known that hypermutated tumour may arise in individuals with germline mutations in the MMR, POLE and MUTYH genes [39]. 75% of hypermutated tumours exhibits MSI while 25% have somatic mutations in MMR genes and/or *POLE or POLD1* genes Somatic *POLE* mutations which affect the exonuclease domain also result in mismatch repair error and subsequent accumulation

of these errors in several genes leading to hypermutation [330]. Tumours with POLE or POLD1 mutations have been shown to be characterized by an extremely high mutation frequency (>1 million per genome) [331]. The hypermutator phenotype observed in two of our study sample in fact harboured a known non-synonymous variant in POLE along with somatic mutation in MSH6 gene in one sample (RS-21) and mutations in *MLH3, MSH3 and PMS2* genes in second sample (RS_33).

Analysis and filtering of variants revealed 15 genes to be recurrently mutated in the 26 non-hypermutated tumours. Interestingly, this somatic mutation distribution differs greatly from the ones reported in US-TCGA RC. The mutation detection rate in the canonical genes also known as "hill" genes except TP53 and SMAD4 were significantly lower or absent in our cohort of early onset RC. The mutation rate of TP53 was similar to the TCGA RC data; however the rate of SMAD4 mutation was significantly higher (50%) when compared to US late onset RC (12%) from TCGA data. There are few reports where loss of SMAD4 was shown to be a feature of early onset CRC [332, 333]. Another report also showed that SMAD4 mutation was higher in younger age group than older age group. These results suggest that the absence of SMAD4 could be a marker of the worse behaviour of early onset CRC, particularly in the microsatellite stable (MSS) group.

The genes found to be recurrently mutated in our cohort are not commonly altered in rectal cancer. Notably one of the recurring genes was an E3 ubiquitin ligase RNF43 which negatively regulates Wnt by ubiquitination and subsequent degradation of the Frizzled receptor. Somatic mutations of *RNF43* are reported in only 0.7% of US late onset rectal cancers TCGA while it was 31% (8/26) in our early onset rectal tumours. These 8 cases do not have APC mutation which is in concordance with literature that

RNF43 and other RNF family members may serve as potential alternative to APC mutation as a mechanism for altering the Wnt signalling pathway in EOS-RC. However, the FFPE blocks were not available for us to examine the IHC status of β-catenin as a marker of Wnt signalling. Therefore the mechanism by which RNF43 is altering Wnt signaling cannot be deduced as RNF43 is shown to be involved in both canonical and non-canonical Wnt signaling pathway [334]. Altogether, the data suggests a possibility that tumourigenesis in sporadic early onset rectal tumours proceeds via an alternate pathway than the known canonical pathways of CRC. However, no definite conclusions can be made owing to the relatively low discovery set which merits further studies on more such samples and validation on an independent cohort.

CHAPTER 7

SUMMARY AND CONCLUSION

Hereditary colorectal cancer syndromes account for 5% of all CRC and include the common hereditary syndromes like LS and rare syndromes like FAP, PJS, JPS and CMMRD. These syndromes are widely studied in the Western countries with most of the clinical and genetic profiling and genotype-phenotype correlations data derived from the Caucasian population. There are few reports from East Asian population as well. However, there is lack of knowledge in the South Asian population. Our group has previously described detailed genotype-phenotype correlation in 54 FAP families [51] Although LS is one of the most common hereditary CRC, only 28 MMR families from India from 3 studies have been reported [233-235]. There is only one report each of STK11 mutation carrier PJS family [110] and PMS2 biallelic mutation carrier CMMRD family from India [281]. The present study therefore examined common as well as rare hereditary CRC syndromes through comprehensive genetic analysis of MMR and STK11 genes in Indian LS, PJS and CMMRD patients.

In the first comprehensive analysis of Indian PJS patients, the STK11 mutation and variant spectrum was quite distinct with several novel mutations**.** This study expands the mutation spectrum of *STK11* gene and highlights the need for studying different populations across the world and collation of data**.** A major observation from this study was the identification of STK11 associated breast cancer phenotype, not previously reported. The characteristic feature of STK11 associated breast cancers was infiltrating duct carcinoma histology, ER positive and Her2Neu negative tumours**.** If this association between STK11 mutation and ER and Her2 status is confirmed in a larger pooled cohort, it could help developing counselling and risk management guidelines and establish a role for chemoprevention and or prophylactic oophorectomy in STK11 mutation carriers similar to their role in BRCA2 mutation carriers.

The very high (96%) mutation detection rate in our Lynch syndrome cohort can be attributed to the use of multimodal approach for selection of patients using different clinical criteria for syndromic diagnosis and pre-screening techniques like IHC and comprehensive genetic testing involving Sanger sequencing and MLPA. The MMR gene mutation spectrum in our cohort included a high frequency of novel mutations highlighting the need for genetic characterization of Lynch syndrome cohorts from different population. While the phenotypes in the MMR mutation carriers in our cohort is similar to that reported in the Caucasian population, the high frequency of breast cancer in our cohort supports the inclusion of breast cancer in the LS tumour spectrum.

Ours is the first comprehensive CMMRD study from India with 8 confirmed and 4 presumed biallelic MMR gene mutation carriers. Due to technical difficulty in screening PMS2, overlapping clinical features and lack of well-established clinical guideline suggests that many CMMRD cases might be missed. Identification of PMS2 biallelic mutations in 2 cases suspected to have LFS further emphasizes the importance of keeping CMMRD as a major differential diagnosis in childhood cancers. In addition, we also report a possible association of MLH3 biallelic mutation with CMMRD which needs to be explored further on additional cohorts to establish the significance of MLH3 as a possible cause of cancer predisposition in CMMRD patients.

This study also examined the somatic landscape of early onset sporadic rectal cancer which is an aggressive and common form of cancer in many developing countries including India. A comprehensive genetic analysis in early onset sporadic rectal cancer has revealed that these cancers are not arising from alteration in any of the

three known somatic canonical pathways of CRC carcinogenesis. Our findings and those from two previous studies from India clearly show that early onset rectal cancers are a distinct subgroup of CRC. Whole exome sequencing reveals this subgroup of CRC has very distinct somatic mutational landscape with very low frequency of mutations in the known canonical genes and identification of recurring mutation in several other less known genes. One notable observation is the significantly high frequency of SMAD4 in the early onset aggressive rectal cancer as compared to late onset rectal cancer data which indicates SMAD4 being a biomarker for the worse behaviour of these subset of rectal cancers. Identification of RNF43 mutation occurring mutually exclusive with APC in eight cases suggests a possibility of existence of an alternative mechanism of Wnt deregulation in early onset rectal caners. However, further studies are required to delineate the cancer development pathway followed by RNF43 and other genes identified in early onset rectal cancer group. These findings need to be confirmed and validated on a larger cohort which may help in delineating the alternate pathway that explains CRC tumourigenesis in early onset sporadic rectal cancer.

CHAPTER 8

BIBLIOGRAPHY

1. Hajdu, S.I., *A note from history: landmarks in history of cancer, part 1.* Cancer, 2011. **117**(5): p. 1097-102.
2. Cooper, J.P. and R.J. Youle, *Balancing cell growth and death.* Curr Opin Cell Biol, 2012. **24**(6): p. 802-3.
3. Rusciano, D. and M.M. Burger, *Why do cancer cells metastasize into particular organs?* Bioessays, 1992. **14**(3): p. 185-94.
4. Bouvard, V., et al., *A review of human carcinogens--Part B: biological agents.* Lancet Oncol, 2009. **10**(4): p. 321-2.
5. Cogliano, V.J., et al., *Preventable exposures associated with human cancers.* J Natl Cancer Inst, 2011. **103**(24): p. 1827-39.
6. Danaei, G., et al., *Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors.* Lancet, 2005. **366**(9499): p. 1784-93.
7. India State-Level Disease Burden Initiative Cancer, C., *The burden of cancers and their variations across the states of India: the Global Burden of Disease Study 1990-2016.* Lancet Oncol, 2018. **19**(10): p. 1289-1306.
8. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA Cancer J Clin, 2018. **68**(6): p. 394-424.
9. Takiar, R., D. Nadayil, and A. Nandakumar, *Projections of number of cancer cases in India (2010-2020) by cancer groups.* Asian Pac J Cancer Prev, 2010. **11**(4): p. 1045-9.
10. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
11. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
12. Vogelstein, B., et al., *Cancer genome landscapes.* Science, 2013. **339**(6127): p. 1546-58.
13. Vogelstein, B. and K.W. Kinzler, *Cancer genes and the pathways they control.* Nat Med, 2004. **10**(8): p. 789-99.
14. Croce, C.M., *Oncogenes and cancer.* N Engl J Med, 2008. **358**(5): p. 502-11.
15. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma.* Proc Natl Acad Sci U S A, 1971. **68**(4): p. 820-3.
16. Knudson, A.G., *Cancer genetics.* Am J Med Genet, 2002. **111**(1): p. 96-102.
17. Santarosa, M. and A. Ashworth, *Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way.* Biochim Biophys Acta, 2004. **1654**(2): p. 105-22.
18. Berger, A.H., A.G. Knudson, and P.P. Pandolfi, *A continuum model for tumour suppression.* Nature, 2011. **476**(7359): p. 163-9.
19. Kinzler, K.W. and B. Vogelstein, *Landscaping the cancer terrain.* Science, 1998. **280**(5366): p. 1036-7.
20. Macleod, K., *Tumor suppressor genes.* Curr Opin Genet Dev, 2000. **10**(1): p. 81-93.
21. Levitt, N.C. and I.D. Hickson, *Caretaker tumour suppressor genes that defend genome integrity.* Trends Mol Med, 2002. **8**(4): p. 179-86.
22. Kinzler, K.W. and B. Vogelstein, *Cancer-susceptibility genes. Gatekeepers and caretakers.* Nature, 1997. **386**(6627): p. 761, 763.

23. Garber, J.E. and K. Offit, *Hereditary cancer predisposition syndromes.* J Clin Oncol, 2005. **23**(2): p. 276-92.
24. Riley, B.D., et al., *Essential elements of genetic cancer risk assessment, counseling, and testing: updated recommendations of the National Society of Genetic Counselors.* J Genet Couns, 2012. **21**(2): p. 151-61.
25. Parkin, D.M., P. Pisani, and J. Ferlay, *Estimates of the worldwide incidence of eighteen major cancers in 1985.* Int J Cancer, 1993. **54**(4): p. 594-606.
26. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer, 2015. **136**(5): p. E359-86.
27. Virk, R., et al., *Racial differences in the incidence of colorectal cancer.* Can J Gastroenterol, 2010. **24**(1): p. 47-51.
28. Jemal, A., et al., *Cancer statistics, 2008.* CA Cancer J Clin, 2008. **58**(2): p. 71-96.
29. Marley, A.R. and H. Nan, *Epidemiology of colorectal cancer.* Int J Mol Epidemiol Genet, 2016. **7**(3): p. 105-114.
30. Marmol, I., et al., *Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer.* Int J Mol Sci, 2017. **18**(1).
31. Eaden, J.A., K.R. Abrams, and J.F. Mayberry, *The risk of colorectal cancer in ulcerative colitis: a meta-analysis.* Gut, 2001. **48**(4): p. 526-35.
32. Canavan, C., K.R. Abrams, and J. Mayberry, *Meta-analysis: colorectal and small bowel cancer risk in patients with Crohn's disease.* Aliment Pharmacol Ther, 2006. **23**(8): p. 1097-104.
33. Compton, C.C. and F.L. Greene, *The staging of colorectal cancer: 2004 and beyond.* CA Cancer J Clin, 2004. **54**(6): p. 295-308.
34. Pancione, M., A. Remo, and V. Colantuoni, *Genetic and epigenetic events generate multiple pathways in colorectal cancer progression.* Patholog Res Int, 2012. **2012**: p. 509348.
35. Ewing, I., et al., *The molecular genetics of colorectal cancer.* Frontline Gastroenterol, 2014. **5**(1): p. 26-30.
36. Vogelstein, B., et al., *Genetic alterations during colorectal-tumor development.* N Engl J Med, 1988. **319**(9): p. 525-32.
37. Kinzler, K.W. and B. Vogelstein, *Lessons from hereditary colorectal cancer.* Cell, 1996. **87**(2): p. 159-70.
38. Aaltonen, L.A., et al., *Clues to the pathogenesis of familial colorectal cancer.* Science, 1993. **260**(5109): p. 812-6.
39. Cancer Genome Atlas, N., *Comprehensive molecular characterization of human colon and rectal cancer.* Nature, 2012. **487**(7407): p. 330-7.
40. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis.* Cell, 1990. **61**(5): p. 759-67.
41. Pino, M.S. and D.C. Chung, *The chromosomal instability pathway in colon cancer.* Gastroenterology, 2010. **138**(6): p. 2059-72.
42. Rajagopalan, H., et al., *The significance of unstable chromosomes in colorectal cancer.* Nat Rev Cancer, 2003. **3**(9): p. 695-701.
43. Toyota, M., et al., *CpG island methylator phenotype in colorectal cancer.* Proc Natl Acad Sci U S A, 1999. **96**(15): p. 8681-6.

44. Kane, M.F., et al., *Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines.* Cancer Res, 1997. **57**(5): p. 808-11.
45. Barault, L., et al., *Hypermethylator phenotype in sporadic colon cancer: study on a population-based series of 582 cases.* Cancer Res, 2008. **68**(20): p. 8541-6.
46. Knudson, A.G., *Two genetic hits (more or less) to cancer.* Nat Rev Cancer, 2001. **1**(2): p. 157-62.
47. Nguyen, H.T. and H.Q. Duong, *The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy.* Oncol Lett, 2018. **16**(1): p. 9-18.
48. De Rosa, M., et al., *The biological complexity of colorectal cancer: insights into biomarkers for early detection and personalized care.* Therap Adv Gastroenterol, 2016. **9**(6): p. 861-886.
49. Tiwari, A., et al., *Novel targeting approaches and signaling pathways of colorectal cancer: An insight.* World J Gastroenterol, 2018. **24**(39): p. 4428-4435.
50. Souglakos, J., *Genetic alterations in sporadic and hereditary colorectal cancer: implementations for screening and follow-up.* Dig Dis, 2007. **25**(1): p. 9-19.
51. Khan, N., et al., *Novel mutations and phenotypic associations identified through APC, MUTYH, NTHL1, POLD1, POLE gene analysis in Indian Familial Adenomatous Polyposis cohort.* Sci Rep, 2017. **7**(1): p. 2214.
52. Yamagishi, H., et al., *Molecular pathogenesis of sporadic colorectal cancers.* Chin J Cancer, 2016. **35**: p. 4.
53. Gupta, S., et al., *Colorectal carcinoma in young adults: a retrospective study on Indian patients: 2000-2008.* Colorectal Dis, 2010. **12**(10 Online): p. e182-9.
54. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
55. Koboldt, D.C., et al., *VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing.* Genome Res, 2012. **22**(3): p. 568-76.
56. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.
57. Jeghers, H., K.V. Mc, and K.H. Katz, *Generalized intestinal polyposis and melanin spots of the oral mucosa, lips and digits; a syndrome of diagnostic significance.* N Engl J Med, 1949. **241**(25): p. 993, illust; passim.
58. Bruwer, A., J.A. Bargen, and R.R. Kierland, *Surface pigmentation and generalized intestinal polyposis; (Peutz-Jeghers syndrome).* Proc Staff Meet Mayo Clin, 1954. **29**(6): p. 168-71.
59. Foley, T.R., T.J. McGarrity, and A.B. Abt, *Peutz-Jeghers syndrome: a clinicopathologic survey of the "Harrisburg family" with a 49-year follow-up.* Gastroenterology, 1988. **95**(6): p. 1535-40.
60. Horrilleno, E.G., C. Eckert, and L.V. Ackerman, *Polyps of the rectum and colon in children.* Cancer, 1957. **10**(6): p. 1210-20.
61. Jenne, D.E., et al., *Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase.* Nat Genet, 1998. **18**(1): p. 38-43.

62. Hemminki, A., et al., *A serine/threonine kinase gene defective in Peutz-Jeghers syndrome.* Nature, 1998. **391**(6663): p. 184-7.
63. Westerman, A.M., et al., *Peutz-Jeghers syndrome: 78-year follow-up of the original family.* Lancet, 1999. **353**(9160): p. 1211-5.
64. Kopacova, M., et al., *Peutz-Jeghers syndrome: diagnostic and therapeutic approach.* World J Gastroenterol, 2009. **15**(43): p. 5397-408.
65. McGarrity, T.J., H.E. Kulin, and R.J. Zaino, *Peutz-Jeghers syndrome.* Am J Gastroenterol, 2000. **95**(3): p. 596-604.
66. van Lier, M.G., et al., *High cancer risk in Peutz-Jeghers syndrome: a systematic review and surveillance recommendations.* Am J Gastroenterol, 2010. **105**(6): p. 1258-64; author reply 1265.
67. Giardiello, F.M., et al., *Very high risk of cancer in familial Peutz-Jeghers syndrome.* Gastroenterology, 2000. **119**(6): p. 1447-53.
68. Daniell, J., et al., *An exploration of genotype-phenotype link between Peutz-Jeghers syndrome and STK11: a review.* Fam Cancer, 2018. **17**(3): p. 421-427.
69. Aretz, S., et al., *High proportion of large genomic STK11 deletions in Peutz-Jeghers syndrome.* Hum Mutat, 2005. **26**(6): p. 513-9.
70. Calva, D. and J.R. Howe, *Hamartomatous polyposis syndromes.* Surg Clin North Am, 2008. **88**(4): p. 779-817, vii.
71. Wen, X.H., et al., *Cronkhite-Canada syndrome: report of six cases and review of literature.* World J Gastroenterol, 2014. **20**(23): p. 7518-22.
72. Duong, B.T. and I. Winship, *The role of STK 11 gene testing in individuals with oral pigmentation.* Australas J Dermatol, 2017. **58**(2): p. 135-138.
73. Smith, D.P., et al., *The mouse Peutz-Jeghers syndrome gene Lkb1 encodes a nuclear protein kinase.* Hum Mol Genet, 1999. **8**(8): p. 1479-85.
74. Sanchez-Cespedes, M., *A role for LKB1 gene in human cancer beyond the Peutz-Jeghers syndrome.* Oncogene, 2007. **26**(57): p. 7825-32.
75. Alessi, D.R., K. Sakamoto, and J.R. Bayascas, *LKB1-dependent signaling pathways.* Annu Rev Biochem, 2006. **75**: p. 137-63.
76. Sapkota, G.P., et al., *Identification and characterization of four novel phosphorylation sites (Ser31, Ser325, Thr336 and Thr366) on LKB1/STK11, the protein kinase mutated in Peutz-Jeghers cancer syndrome.* Biochem J, 2002. **362**(Pt 2): p. 481-90.
77. Sapkota, G.P., et al., *Phosphorylation of the protein kinase mutated in Peutz-Jeghers cancer syndrome, LKB1/STK11, at Ser431 by p90(RSK) and cAMP-dependent protein kinase, but not its farnesylation at Cys(433), is essential for LKB1 to suppress cell vrowth.* J Biol Chem, 2001. **276**(22): p. 19469-82.
78. Nezu, J., A. Oku, and M. Shimane, *Loss of cytoplasmic retention ability of mutant LKB1 found in Peutz-Jeghers syndrome patients.* Biochem Biophys Res Commun, 1999. **261**(3): p. 750-5.
79. Boudeau, J., et al., *MO25alpha/beta interact with STRADalpha/beta enhancing their ability to bind, activate and localize LKB1 in the cytoplasm.* EMBO J, 2003. **22**(19): p. 5102-14.
80. Baas, A.F., et al., *Activation of the tumour suppressor kinase LKB1 by the STE20-like pseudokinase STRAD.* EMBO J, 2003. **22**(12): p. 3062-72.
81. Lizcano, J.M., et al., *LKB1 is a master kinase that activates 13 kinases of the AMPK subfamily, including MARK/PAR-1.* EMBO J, 2004. **23**(4): p. 833-43.

82. Jaleel, M., et al., *Identification of the sucrose non-fermenting related kinase SNRK, as a novel LKB1 substrate.* FEBS Lett, 2005. **579**(6): p. 1417-23.
83. Hawley, S.A., et al., *Calmodulin-dependent protein kinase kinase-beta is an alternative upstream kinase for AMP-activated protein kinase.* Cell Metab, 2005. **2**(1): p. 9-19.
84. Hurley, R.L., et al., *The Ca2+/calmodulin-dependent protein kinase kinases are AMP-activated protein kinase kinases.* J Biol Chem, 2005. **280**(32): p. 29060-6.
85. Woods, A., et al., *Ca2+/calmodulin-dependent protein kinase kinase-beta acts upstream of AMP-activated protein kinase in mammalian cells.* Cell Metab, 2005. **2**(1): p. 21-33.
86. Hardie, D.G., D. Carling, and S.J. Gamblin, *AMP-activated protein kinase: also regulated by ADP?* Trends Biochem Sci, 2011. **36**(9): p. 470-7.
87. Oakhill, J.S., et al., *AMPK is a direct adenylate charge-regulated protein kinase.* Science, 2011. **332**(6036): p. 1433-5.
88. Xiao, B., et al., *Structure of mammalian AMPK and its regulation by ADP.* Nature, 2011. **472**(7342): p. 230-3.
89. Mihaylova, M.M. and R.J. Shaw, *The AMPK signalling pathway coordinates cell growth, autophagy and metabolism.* Nat Cell Biol, 2011. **13**(9): p. 1016-23.
90. Hardie, D.G., F.A. Ross, and S.A. Hawley, *AMPK: a nutrient and energy sensor that maintains energy homeostasis.* Nat Rev Mol Cell Biol, 2012. **13**(4): p. 251-62.
91. Tiainen, M., et al., *Growth arrest by the LKB1 tumor suppressor: induction of p21(WAF1/CIP1).* Hum Mol Genet, 2002. **11**(13): p. 1497-504.
92. Karuman, P., et al., *The Peutz-Jegher gene product LKB1 is a mediator of p53-dependent cell death.* Mol Cell, 2001. **7**(6): p. 1307-19.
93. Xie, X., Z. Wang, and Y. Chen, *Association of LKB1 with a WD-repeat protein WDR6 is implicated in cell growth arrest and p27(Kip1) induction.* Mol Cell Biochem, 2007. **301**(1-2): p. 115-22.
94. Liang, X., et al., *Exogenous activation of LKB1/AMPK signaling induces G(1) arrest in cells with endogenous LKB1 expression.* Mol Med Rep, 2014. **9**(3): p. 1019-24.
95. Cheng, H., et al., *SIK1 couples LKB1 to p53-dependent anoikis and suppresses metastasis.* Sci Signal, 2009. **2**(80): p. ra35.
96. Barnes, A.P., et al., *LKB1 and SAD kinases define a pathway required for the polarization of cortical neurons.* Cell, 2007. **129**(3): p. 549-63.
97. Zagorska, A., et al., *New roles for the LKB1-NUAK pathway in controlling myosin phosphatase complexes and cell adhesion.* Sci Signal, 2010. **3**(115): p. ra25.
98. Zeqiraj, E., et al., *Structure of the LKB1-STRAD-MO25 complex reveals an allosteric mechanism of kinase activation.* Science, 2009. **326**(5960): p. 1707-11.
99. Volikos, E., et al., *LKB1 exonic and whole gene deletions are a common cause of Peutz-Jeghers syndrome.* J Med Genet, 2006. **43**(5): p. e18.
100. Hearle, N.C., et al., *Exonic STK11 deletions are not a rare cause of Peutz-Jeghers syndrome.* J Med Genet, 2006. **43**(4): p. e15.

101. De Rosa, M., et al., *Alu-mediated genomic deletion of the serine/threonine protein kinase 11 (STK11) gene in Peutz-Jeghers syndrome.* Gastroenterology, 2010. **138**(7): p. 2558-60.
102. Resta, N., et al., *Breakpoint determination of 15 large deletions in Peutz-Jeghers subjects.* Hum Genet, 2010. **128**(4): p. 373-82.
103. Lim, W., et al., *Further observations on LKB1/STK11 status and cancer risk in Peutz-Jeghers syndrome.* Br J Cancer, 2003. **89**(2): p. 308-13.
104. Mehenni, H., et al., *Cancer risks in LKB1 germline mutation carriers.* Gut, 2006. **55**(7): p. 984-90.
105. Olschwang, S., C. Boisson, and G. Thomas, *Peutz-Jeghers families unlinked to STK11/LKB1 gene mutations are highly predisposed to primitive biliary adenocarcinoma.* J Med Genet, 2001. **38**(6): p. 356-60.
106. Zhang, Y., et al., *Correlation between genotype and phenotype in three families with Peutz-Jeghers Syndrome.* Exp Ther Med, 2017. **13**(2): p. 507-514.
107. Salloch, H., et al., *Truncating mutations in Peutz-Jeghers syndrome are associated with more polyps, surgical interventions and cancers.* Int J Colorectal Dis, 2010. **25**(1): p. 97-107.
108. Boudeau, J., et al., *Analysis of the LKB1-STRAD-MO25 complex.* J Cell Sci, 2004. **117**(Pt 26): p. 6365-75.
109. Hearle, N., et al., *Frequency and spectrum of cancers in the Peutz-Jeghers syndrome.* Clin Cancer Res, 2006. **12**(10): p. 3209-15.
110. Thakur, N., et al., *A novel mutation in STK11 gene is associated with Peutz-Jeghers Syndrome in Indian patients.* BMC Med Genet, 2006. **7**: p. 73.
111. Mehenni, H., et al., *Peutz-Jeghers syndrome: confirmation of linkage to chromosome 19p13.3 and identification of a potential second locus, on 19q13.4.* Am J Hum Genet, 1997. **61**(6): p. 1327-34.
112. Chow, E., et al., *An updated mutation spectrum in an Australian series of PJS patients provides further evidence for only one gene locus.* Clin Genet, 2006. **70**(5): p. 409-14.
113. Chen, H., et al., *Association Between BRCA Status and Triple-Negative Breast Cancer: A Meta-Analysis.* Front Pharmacol, 2018. **9**: p. 909.
114. Shimelis, H., et al., *Triple-Negative Breast Cancer Risk Genes Identified by Multigene Hereditary Cancer Panel Testing.* J Natl Cancer Inst, 2018.
115. Hearle, N., et al., *Mapping of a translocation breakpoint in a Peutz-Jeghers hamartoma to the putative PJS locus at 19q13.4 and mutation analysis of candidate genes in polyp and STK11-negative PJS cases.* Genes Chromosomes Cancer, 2004. **41**(2): p. 163-9.
116. de Leng, W.W., et al., *Genetic defects underlying Peutz-Jeghers syndrome (PJS) and exclusion of the polarity-associated MARK/Par1 gene family as potential PJS candidates.* Clin Genet, 2007. **72**(6): p. 568-73.
117. Wang, H.H., et al., *Exome sequencing revealed novel germline mutations in Chinese Peutz-Jeghers syndrome patients.* Dig Dis Sci, 2014. **59**(1): p. 64-71.
118. Duan, F.X., et al., *Must Peutz-Jeghers syndrome patients have the LKB1/STK11 gene mutation? A case report and review of the literature.* World J Clin Cases, 2018. **6**(8): p. 224-232.
119. McKay, V., et al., *First report of somatic mosaicism for mutations in STK11 in four patients with Peutz-Jeghers syndrome.* Fam Cancer, 2016. **15**(1): p. 57-61.

120. Butel-Simoes, G.I., et al., *Low-level parental mosaicism in an apparent de novo case of Peutz-Jeghers syndrome.* Fam Cancer, 2019. **18**(1): p. 109-112.
121. Hearle, N.C., et al., *Sequence changes in predicted promoter elements of STK11/LKB1 are unlikely to contribute to Peutz-Jeghers syndrome.* BMC Genomics, 2005. **6**: p. 38.
122. Fostira, F., et al., *Genetic analysis and clinical description of Greek patients with Peutz-Jeghers syndrome: Creation of a National Registry.* Cancer Genet, 2018. **220**: p. 19-23.
123. Aaltonen, L.A., et al., *Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease.* N Engl J Med, 1998. **338**(21): p. 1481-7.
124. Lynch, H.T., et al., *Phenotypic and genotypic heterogeneity in the Lynch syndrome: diagnostic, surveillance and management implications.* Eur J Hum Genet, 2006. **14**(4): p. 390-402.
125. Lynch, P.M., *Clinical challenges in management of familial adenomatous polyposis and hereditary nonpolyposis colorectal cancer.* Cancer, 1999. **86**(11 Suppl): p. 2533-9.
126. Aarnio, M., et al., *Cancer risk in mutation carriers of DNA-mismatch-repair genes.* Int J Cancer, 1999. **81**(2): p. 214-8.
127. Aarnio, M., et al., *Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (HNPCC) syndrome.* Int J Cancer, 1995. **64**(6): p. 430-3.
128. Stoffel, E., et al., *Calculation of risk of colorectal and endometrial cancer among patients with Lynch syndrome.* Gastroenterology, 2009. **137**(5): p. 1621-7.
129. *Classics in oncology. Heredity with reference to carcinoma as shown by the study of the cases examined in the pathological laboratory of the University of Michigan, 1895-1913. By Aldred Scott Warthin. 1913.* CA Cancer J Clin, 1985. **35**(6): p. 348-59.
130. Lynch, H.T., et al., *Hereditary factors in cancer. Study of two large midwestern kindreds.* Arch Intern Med, 1966. **117**(2): p. 206-12.
131. Lynch, H.T., et al., *Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications.* Clin Genet, 2009. **76**(1): p. 1-18.
132. Lynch, H.T. and A.J. Krush, *Cancer family "G" revisited: 1895-1970.* Cancer, 1971. **27**(6): p. 1505-11.
133. Lynch, H.T., et al., *Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review.* Gastroenterology, 1993. **104**(5): p. 1535-49.
134. Vasen, H.F., et al., *The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC).* Dis Colon Rectum, 1991. **34**(5): p. 424-5.
135. Ryan, P., et al., *Comparison of clinical schemas and morphologic features in predicting Lynch syndrome in mutation-positive patients with endometrial cancer encountered in the context of familial gastrointestinal cancer registries.* Cancer, 2012. **118**(3): p. 681-8.

136. Risio, M., et al., *Microsatellite instability is associated with the histological features of the tumor in nonfamilial colorectal cancer.* Cancer Res, 1996. **56**(23): p. 5470-4.
137. Watson, P., et al., *Colorectal carcinoma survival among hereditary nonpolyposis colorectal carcinoma family members.* Cancer, 1998. **83**(2): p. 259-66.
138. Stigliano, V., et al., *Survival of hereditary non-polyposis colorectal cancer patients compared with sporadic colorectal cancer patients.* J Exp Clin Cancer Res, 2008. **27**: p. 39.
139. Johnson, R.L. and J.C. Fleet, *Animal models of colorectal cancer.* Cancer Metastasis Rev, 2013. **32**(1-2): p. 39-61.
140. Moreira, L., et al., *Identification of Lynch syndrome among patients with colorectal cancer.* JAMA, 2012. **308**(15): p. 1555-65.
141. Win, A.K., et al., *Risks of primary extracolonic cancers following colorectal cancer in lynch syndrome.* J Natl Cancer Inst, 2012. **104**(18): p. 1363-72.
142. Win, A.K., et al., *Colorectal and other cancer risks for carriers and noncarriers from families with a DNA mismatch repair gene mutation: a prospective cohort study.* J Clin Oncol, 2012. **30**(9): p. 958-64.
143. Vasen, H.F., et al., *New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC.* Gastroenterology, 1999. **116**(6): p. 1453-6.
144. Llor, X., et al., *Differential features of colorectal cancers fulfilling Amsterdam criteria without involvement of the mutator pathway.* Clin Cancer Res, 2005. **11**(20): p. 7304-10.
145. Vasen, H.F., et al., *Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts.* Gut, 2013. **62**(6): p. 812-23.
146. Rodriguez-Bigas, M.A., et al., *A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines.* J Natl Cancer Inst, 1997. **89**(23): p. 1758-62.
147. Umar, A., et al., *Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability.* J Natl Cancer Inst, 2004. **96**(4): p. 261-8.
148. de Jong, A.E., et al., *Microsatellite instability, immunohistochemistry, and additional PMS2 staining in suspected hereditary nonpolyposis colorectal cancer.* Clin Cancer Res, 2004. **10**(3): p. 972-80.
149. Ionov, Y., et al., *Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis.* Nature, 1993. **363**(6429): p. 558-61.
150. Thibodeau, S.N., G. Bren, and D. Schaid, *Microsatellite instability in cancer of the proximal colon.* Science, 1993. **260**(5109): p. 816-9.
151. Ellegren, H., *Microsatellites: simple sequences with complex evolution.* Nat Rev Genet, 2004. **5**(6): p. 435-45.
152. Chung, D.C. and A.K. Rustgi, *The hereditary nonpolyposis colorectal cancer syndrome: genetics and clinical implications.* Ann Intern Med, 2003. **138**(7): p. 560-70.
153. Le, D.T., et al., *PD-1 Blockade in Tumors with Mismatch-Repair Deficiency.* N Engl J Med, 2015. **372**(26): p. 2509-20.

154. Pawlik, T.M., C.P. Raut, and M.A. Rodriguez-Bigas, *Colorectal carcinogenesis: MSI-H versus MSI-L.* Dis Markers, 2004. **20**(4-5): p. 199-206.
155. Shibata, D., et al., *Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation.* Nat Genet, 1994. **6**(3): p. 273-81.
156. Lynch, H.T., et al., *Making sense of missense in Lynch syndrome: the clinical perspective.* Cancer Prev Res (Phila), 2010. **3**(11): p. 1371-4.
157. Soreide, K., et al., *Microsatellite instability in colorectal cancer.* Br J Surg, 2006. **93**(4): p. 395-406.
158. Boland, C.R., et al., *A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer.* Cancer Res, 1998. **58**(22): p. 5248-57.
159. Suraweera, N., et al., *Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR.* Gastroenterology, 2002. **123**(6): p. 1804-11.
160. Zhou, X.P., et al., *Allelic profiles of mononucleotide repeat microsatellites in control individuals and in colorectal tumors with and without replication errors.* Oncogene, 1997. **15**(14): p. 1713-8.
161. Cicek, M.S., et al., *Quality assessment and correlation of microsatellite instability and immunohistochemical markers among population- and clinic-based colorectal tumors results from the Colon Cancer Family Registry.* J Mol Diagn, 2011. **13**(3): p. 271-81.
162. Shia, J., et al., *Lynch syndrome-associated neoplasms: a discussion on histopathology and immunohistochemistry.* Fam Cancer, 2013. **12**(2): p. 241-60.
163. Shia, J., *Immunohistochemistry versus microsatellite instability testing for screening colorectal cancer patients at risk for hereditary nonpolyposis colorectal cancer syndrome. Part I. The utility of immunohistochemistry.* J Mol Diagn, 2008. **10**(4): p. 293-300.
164. Zhang, L., *Immunohistochemistry versus microsatellite instability testing for screening colorectal cancer patients at risk for hereditary nonpolyposis colorectal cancer syndrome. Part II. The utility of microsatellite instability testing.* J Mol Diagn, 2008. **10**(4): p. 301-7.
165. Barrow, E., et al., *Semiquantitative assessment of immunohistochemistry for mismatch repair proteins in Lynch syndrome.* Histopathology, 2010. **56**(3): p. 331-44.
166. Hall, G., et al., *Immunohistochemistry for PMS2 and MSH6 alone can replace a four antibody panel for mismatch repair deficiency screening in colorectal adenocarcinoma.* Pathology, 2010. **42**(5): p. 409-13.
167. Hampel, H., et al., *Feasibility of screening for Lynch syndrome among patients with colorectal cancer.* J Clin Oncol, 2008. **26**(35): p. 5783-8.
168. Wang, Q., et al., *Prevalence of germline mutations of hMLH1, hMSH2, hPMS1, hPMS2, and hMSH6 genes in 75 French kindreds with nonpolyposis colorectal cancer.* Hum Genet, 1999. **105**(1-2): p. 79-85.
169. Pluciennik, A. and P. Modrich, *Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair.* Proc Natl Acad Sci U S A, 2007. **104**(31): p. 12709-13.

170. Kunkel, T.A. and D.A. Erie, *DNA mismatch repair.* Annu Rev Biochem, 2005. **74**: p. 681-710.
171. Fishel, R., *The selection for mismatch repair defects in hereditary nonpolyposis colorectal cancer: revising the mutator hypothesis.* Cancer Res, 2001. **61**(20): p. 7369-74.
172. Harfe, B.D. and S. Jinks-Robertson, *DNA mismatch repair and genetic instability.* Annu Rev Genet, 2000. **34**: p. 359-399.
173. Hsieh, P. and K. Yamane, *DNA mismatch repair: molecular mechanism, cancer, and ageing.* Mech Ageing Dev, 2008. **129**(7-8): p. 391-407.
174. Thompson, B.A., et al., *Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database.* Nat Genet, 2014. **46**(2): p. 107-115.
175. Ligtenberg, M.J., et al., *Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1.* Nat Genet, 2009. **41**(1): p. 112-7.
176. Kovacs, M.E., et al., *Deletions removing the last exon of TACSTD1 constitute a distinct class of mutations predisposing to Lynch syndrome.* Hum Mutat, 2009. **30**(2): p. 197-203.
177. Clark, N., X. Wu, and C. Her, *MutS Homologues hMSH4 and hMSH5: Genetic Variations, Functions, and Implications in Human Diseases.* Curr Genomics, 2013. **14**(2): p. 81-90.
178. Eng, C., *High frequency of large gene deletions and rearrangements in Lynch syndrome--back to the future?* Gastroenterology, 2005. **129**(3): p. 1124-6.
179. Perez-Cabornero, L., et al., *Characterization of new founder Alu-mediated rearrangements in MSH2 gene associated with a Lynch syndrome phenotype.* Cancer Prev Res (Phila), 2011. **4**(10): p. 1546-55.
180. Mauillon, J.L., et al., *Identification of novel germline hMLH1 mutations including a 22 kb Alu-mediated deletion in patients with familial colorectal cancer.* Cancer Res, 1996. **56**(24): p. 5728-33.
181. Peltomaki, P., *Update on Lynch syndrome genomics.* Fam Cancer, 2016. **15**(3): p. 385-93.
182. Hitchins, M.P., et al., *Inheritance of a cancer-associated MLH1 germ-line epimutation.* N Engl J Med, 2007. **356**(7): p. 697-705.
183. Hitchins, M.P., et al., *Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR.* Cancer Cell, 2011. **20**(2): p. 200-13.
184. Boland, C.R. and A. Goel, *Microsatellite instability in colorectal cancer.* Gastroenterology, 2010. **138**(6): p. 2073-2087 e3.
185. Borgdorff, V., et al., *DNA mismatch repair mediates protection from mutagenesis induced by short-wave ultraviolet light.* DNA Repair (Amst), 2006. **5**(11): p. 1364-72.
186. Borgdorff, V., et al., *Spontaneous and mutagen-induced loss of DNA mismatch repair in Msh2-heterozygous mammalian cells.* Mutat Res, 2005. **574**(1-2): p. 50-7.
187. Smith-Roe, S.L., et al., *Mlh1-dependent suppression of specific mutations induced in vivo by the food-borne carcinogen 2-amino-1-methyl-6-phenylimidazo [4,5-b] pyridine (PhIP).* Mutat Res, 2006. **594**(1-2): p. 101-12.

188. Hofstra, R.M., et al., *Tumor characteristics as an analytic tool for classifying genetic variants of uncertain clinical significance.* Hum Mutat, 2008. **29**(11): p. 1292-303.
189. Desai, D.C., et al., *Recurrent germline mutation in MSH2 arises frequently de novo.* J Med Genet, 2000. **37**(9): p. 646-52.
190. Ponti, G., et al., *Mismatch repair genes founder mutations and cancer susceptibility in Lynch syndrome.* Clin Genet, 2015. **87**(6): p. 507-16.
191. Giardiello, F.M., et al., *Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-Society Task Force on colorectal cancer.* Gastroenterology, 2014. **147**(2): p. 502-26.
192. Bonadona, V., et al., *Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome.* JAMA, 2011. **305**(22): p. 2304-10.
193. Barrow, E., J. Hill, and D.G. Evans, *Cancer risk in Lynch Syndrome.* Fam Cancer, 2013. **12**(2): p. 229-40.
194. Alarcon, F., et al., *Estimating cancer risk in HNPCC by the GRL method.* Eur J Hum Genet, 2007. **15**(8): p. 831-6.
195. Dunlop, M.G., et al., *Cancer risk associated with germline DNA mismatch repair gene mutations.* Hum Mol Genet, 1997. **6**(1): p. 105-10.
196. Quehenberger, F., H.F. Vasen, and H.C. van Houwelingen, *Risk of colorectal and endometrial cancer for carriers of mutations of the hMLH1 and hMSH2 gene: correction for ascertainment.* J Med Genet, 2005. **42**(6): p. 491-6.
197. Dowty, J.G., et al., *Cancer risks for MLH1 and MSH2 mutation carriers.* Hum Mutat, 2013. **34**(3): p. 490-7.
198. Baglietto, L., et al., *Risks of Lynch syndrome cancers for MSH6 mutation carriers.* J Natl Cancer Inst, 2010. **102**(3): p. 193-201.
199. Barrow, E., et al., *Colorectal cancer in HNPCC: cumulative lifetime incidence, survival and tumour distribution. A report of 121 families with proven mutations.* Clin Genet, 2008. **74**(3): p. 233-42.
200. Senter, L., et al., *The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations.* Gastroenterology, 2008. **135**(2): p. 419-28.
201. Plaschke, J., et al., *Lower incidence of colorectal cancer and later age of disease onset in 27 families with pathogenic MSH6 germline mutations compared with families with MLH1 or MSH2 mutations: the German Hereditary Nonpolyposis Colorectal Cancer Consortium.* J Clin Oncol, 2004. **22**(22): p. 4486-94.
202. Wijnen, J., et al., *Familial endometrial cancer in female carriers of MSH6 germline mutations.* Nat Genet, 1999. **23**(2): p. 142-4.
203. Berends, M.J., et al., *Molecular and clinical characteristics of MSH6 variants: an analysis of 25 index carriers of a germline variant.* Am J Hum Genet, 2002. **70**(1): p. 26-37.
204. Hendriks, Y.M., et al., *Cancer risk in hereditary nonpolyposis colorectal cancer due to MSH6 mutations: impact on counseling and surveillance.* Gastroenterology, 2004. **127**(1): p. 17-25.
205. Barrow, E., et al., *Cumulative lifetime incidence of extracolonic cancers in Lynch syndrome: a report of 121 families with proven mutations.* Clin Genet, 2009. **75**(2): p. 141-9.

206. Cohen, S.A. and A. Leininger, *The genetic basis of Lynch syndrome and its implications for clinical practice and risk management.* Appl Clin Genet, 2014. **7**: p. 147-58.
207. Win, A.K., N.M. Lindor, and M.A. Jenkins, *Risk of breast cancer in Lynch syndrome: a systematic review.* Breast Cancer Res, 2013. **15**(2): p. R27.
208. Ryan, S., M.A. Jenkins, and A.K. Win, *Risk of prostate cancer in Lynch syndrome: a systematic review and meta-analysis.* Cancer Epidemiol Biomarkers Prev, 2014. **23**(3): p. 437-49.
209. Lynch, H.T. and A. de la Chapelle, *Genetic susceptibility to non-polyposis colorectal cancer.* J Med Genet, 1999. **36**(11): p. 801-18.
210. Suspiro, A., et al., *The Muir-Torre syndrome: a rare variant of hereditary nonpolyposis colorectal cancer associated with hMSH2 mutation.* Am J Gastroenterol, 1998. **93**(9): p. 1572-4.
211. Turcot, J., J.P. Despres, and F. St Pierre, *Malignant tumors of the central nervous system associated with familial polyposis of the colon: report of two cases.* Dis Colon Rectum, 1959. **2**: p. 465-8.
212. Hamilton, S.R., et al., *The molecular basis of Turcot's syndrome.* N Engl J Med, 1995. **332**(13): p. 839-47.
213. Lebrun, C., et al., *Turcot syndrome confirmed with molecular analysis.* Eur J Neurol, 2007. **14**(4): p. 470-2.
214. Holter, S., et al., *Hepatic adenomas caused by somatic HNF1A mutations in children with biallelic mismatch repair gene mutations.* Gastroenterology, 2011. **140**(2): p. 735-6.
215. Medina-Arana, V., et al., *Adrenocortical carcinoma, an unusual extracolonic tumor associated with Lynch II syndrome.* Fam Cancer, 2011. **10**(2): p. 265-71.
216. Medina Arana, V., et al., *Highly aggressive leiomyosarcoma associated with Lynch II syndrome: increasing the range of extracolonic cancers related with hereditary non-polyposis colonic cancer.* Ann Oncol, 2002. **13**(5): p. 807-8.
217. Sijmons, R., et al., *Inclusion of malignant fibrous histiocytoma in the tumour spectrum associated with hereditary non-polyposis colorectal cancer.* Genes Chromosomes Cancer, 2000. **29**(4): p. 353-5.
218. Bandipalliam, P., et al., *Clinical presentation correlates with the type of mismatch repair gene involved in hereditary nonpolyposis colon cancer.* Gastroenterology, 2004. **126**(3): p. 936-7.
219. Vasen, H.F., et al., *Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis.* Gastroenterology, 1996. **110**(4): p. 1020-7.
220. Peltomaki, P., X. Gao, and J.P. Mecklin, *Genotype and phenotype in hereditary nonpolyposis colon cancer: a study of families with different vs. shared predisposing mutations.* Fam Cancer, 2001. **1**(1): p. 9-15.
221. Wagner, A., et al., *Atypical HNPCC owing to MSH6 germline mutations: analysis of a large Dutch pedigree.* J Med Genet, 2001. **38**(5): p. 318-22.
222. Verma, L., et al., *Mononucleotide microsatellite instability and germline MSH6 mutation analysis in early onset colorectal cancer.* J Med Genet, 1999. **36**(9): p. 678-82.
223. Nakagawa, H., et al., *Mismatch repair gene PMS2: disease-causing germline mutations are frequent in patients whose tumors stain negative for PMS2*

protein, but paralogous genes obscure mutation detection and interpretation. Cancer Res, 2004. **64**(14): p. 4721-7.
224. Hendriks, Y.M., et al., *Heterozygous mutations in PMS2 cause hereditary nonpolyposis colorectal carcinoma (Lynch syndrome).* Gastroenterology, 2006. **130**(2): p. 312-22.
225. Lynch, H.T., et al., *Milestones of Lynch syndrome: 1895-2015.* Nat Rev Cancer, 2015. **15**(3): p. 181-94.
226. Kempers, M.J., et al., *Risk of colorectal and endometrial cancers in EPCAM deletion-positive Lynch syndrome: a cohort study.* Lancet Oncol, 2011. **12**(1): p. 49-55.
227. Lynch, H.T., et al., *Lynch syndrome-associated extracolonic tumors are rare in two extended families with the same EPCAM deletion.* Am J Gastroenterol, 2011. **106**(10): p. 1829-36.
228. Ward, R.L., et al., *Identification of constitutional MLH1 epimutations and promoter variants in colorectal cancer patients from the Colon Cancer Family Registry.* Genet Med, 2013. **15**(1): p. 25-35.
229. Halvarsson, B., et al., *Phenotypic heterogeneity in hereditary non-polyposis colorectal cancer: identical germline mutations associated with variable tumour morphology and immunohistochemical expression.* J Clin Pathol, 2007. **60**(7): p. 781-6.
230. Perez-Cabornero, L., et al., *Genotype-phenotype correlation in MMR mutation-positive families with Lynch syndrome.* Int J Colorectal Dis, 2013. **28**(9): p. 1195-201.
231. Mangold, E., et al., *Spectrum and frequencies of mutations in MSH2 and MLH1 identified in 1,721 German families suspected of hereditary nonpolyposis colorectal cancer.* Int J Cancer, 2005. **116**(5): p. 692-702.
232. Kurzawski, G., et al., *Germline MSH2 and MLH1 mutational spectrum including large rearrangements in HNPCC families from Poland (update study).* Clin Genet, 2006. **69**(1): p. 40-7.
233. Bashyam, M.D., et al., *Evidence for presence of mismatch repair gene expression positive Lynch syndrome cases in India.* Mol Carcinog, 2015. **54**(12): p. 1807-14.
234. Rajender, S., et al., *R659X mutation in the MLH1 gene in hereditary non-polyposis colorectal cancer(HNPCC) in an Indian extended family.* Indian J Med Res, 2010. **131**: p. 64-70.
235. Rajkumar, T., et al., *Mutation analysis of hMSH2 and hMLH1 in colorectal cancer patients in India.* Genet Test, 2004. **8**(2): p. 157-62.
236. Chew, M.H., et al., *Phenotypic characteristics of hereditary non-polyposis colorectal cancer by the Amsterdam criteria: an Asian perspective.* ANZ J Surg, 2008. **78**(7): p. 556-60.
237. Wei, W., et al., *Racial differences in MLH1 and MSH2 mutation: an analysis of yellow race and white race based on the InSiGHT database.* J Bioinform Comput Biol, 2010. **8 Suppl 1**: p. 111-25.
238. Loizidou, M.A., et al., *The mutational spectrum of Lynch syndrome in cyprus.* PLoS One, 2014. **9**(8): p. e105501.
239. Wagner, A., et al., *Molecular analysis of hereditary nonpolyposis colorectal cancer in the United States: high mutation detection rate among clinically*

selected families and characterization of an American founder genomic deletion of the MSH2 gene. Am J Hum Genet, 2003. **72**(5): p. 1088-100.
240. Li, L., et al., *Distinct patterns of germ-line deletions in MLH1 and MSH2: the implication of Alu repetitive element in the genetic etiology of Lynch syndrome (HNPCC).* Hum Mutat, 2006. **27**(4): p. 388.
241. Goecke, T., et al., *Genotype-phenotype comparison of German MLH1 and MSH2 mutation carriers clinically affected with Lynch syndrome: a report by the German HNPCC Consortium.* J Clin Oncol, 2006. **24**(26): p. 4285-92.
242. Watson, P., et al., *The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome.* Int J Cancer, 2008. **123**(2): p. 444-449.
243. Harkness, E.F., et al., *Lynch syndrome caused by MLH1 mutations is associated with an increased risk of breast cancer: a cohort study.* J Med Genet, 2015. **52**(8): p. 553-6.
244. da Silva, F.C., et al., *Frequency of extracolonic tumors in Brazilian families with Lynch syndrome: analysis of a hereditary colorectal cancer institutional registry.* Fam Cancer, 2010. **9**(4): p. 563-70.
245. Huang, D., et al., *Systematic Review: An Update on the Spectrum of Urological Malignancies in Lynch Syndrome.* Bladder Cancer, 2018. **4**(3): p. 261-268.
246. Raymond, V.M., et al., *Adrenocortical carcinoma is a lynch syndrome-associated cancer.* J Clin Oncol, 2013. **31**(24): p. 3012-8.
247. Suerink, M., et al., *Constitutional mismatch repair deficiency as a differential diagnosis of neurofibromatosis type 1: consensus guidelines for testing a child without malignancy.* J Med Genet, 2019. **56**(2): p. 53-62.
248. Abedalthagafi, M., *Constitutional mismatch repair-deficiency: current problems and emerging therapeutic strategies.* Oncotarget, 2018. **9**(83): p. 35458-35469.
249. Ricciardone, M.D., et al., *Human MLH1 deficiency predisposes to hematological malignancy and neurofibromatosis type 1.* Cancer Res, 1999. **59**(2): p. 290-3.
250. Wang, Q., et al., *Neurofibromatosis and early onset of cancers in hMLH1-deficient children.* Cancer Res, 1999. **59**(2): p. 294-7.
251. Whiteside, D., et al., *A homozygous germ-line mutation in the human MSH2 gene predisposes to hematological malignancy and multiple cafe-au-lait spots.* Cancer Res, 2002. **62**(2): p. 359-62.
252. Mori, T., et al., *Germ-line and somatic mutations of the APC gene in patients with Turcot syndrome and analysis of APC mutations in brain tumors.* Genes Chromosomes Cancer, 1994. **9**(3): p. 168-72.
253. Huang, H., et al., *APC mutations in sporadic medulloblastomas.* Am J Pathol, 2000. **156**(2): p. 433-7.
254. Wimmer, K. and J. Etzler, *Constitutional mismatch repair-deficiency syndrome: have we so far seen only the tip of an iceberg?* Hum Genet, 2008. **124**(2): p. 105-22.
255. Leung, S.Y., et al., *Microsatellite instability and mutation of DNA mismatch repair genes in gliomas.* Am J Pathol, 1998. **153**(4): p. 1181-8.
256. Kratz, C.P., et al., *Rhabdomyosarcoma in patients with constitutional mismatch-repair-deficiency syndrome.* J Med Genet, 2009. **46**(6): p. 418-20.

257. Bruwer, Z., et al., *Predictive genetic testing in children: constitutional mismatch repair deficiency cancer predisposing syndrome.* J Genet Couns, 2014. **23**(2): p. 147-55.
258. Bakry, D., et al., *Genetic and clinical determinants of constitutional mismatch repair deficiency syndrome: report from the constitutional mismatch repair deficiency consortium.* Eur J Cancer, 2014. **50**(5): p. 987-96.
259. Antelo, M., et al., *Pitfalls in the diagnosis of biallelic PMS2 mutations.* Fam Cancer, 2015. **14**(3): p. 411-4.
260. Shlien, A., et al., *Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers.* Nat Genet, 2015. **47**(3): p. 257-62.
261. Baris, H.N., et al., *Constitutional Mismatch Repair Deficiency in Israel: High Proportion of Founder Mutations in MMR Genes and Consanguinity.* Pediatr Blood Cancer, 2016. **63**(3): p. 418-27.
262. Amayiri, N., et al., *High frequency of mismatch repair deficiency among pediatric high grade gliomas in Jordan.* Int J Cancer, 2016. **138**(2): p. 380-5.
263. Wimmer, K. and C.P. Kratz, *Constitutional mismatch repair-deficiency syndrome.* Haematologica, 2010. **95**(5): p. 699-701.
264. Wimmer, K., et al., *Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium 'care for CMMRD' (C4CMMRD).* J Med Genet, 2014. **51**(6): p. 355-65.
265. Vasen, H.F., et al., *Guidelines for surveillance of individuals with constitutional mismatch repair-deficiency proposed by the European Consortium "Care for CMMR-D" (C4CMMR-D).* J Med Genet, 2014. **51**(5): p. 283-93.
266. Ramchander, N.C., et al., *Homozygous germ-line mutation of the PMS2 mismatch repair gene: a unique case report of constitutional mismatch repair deficiency (CMMRD).* BMC Med Genet, 2017. **18**(1): p. 40.
267. Durno, C.A., et al., *Phenotypic and genotypic characterisation of biallelic mismatch repair deficiency (BMMR-D) syndrome.* Eur J Cancer, 2015. **51**(8): p. 977-83.
268. Tabori, U., et al., *Clinical Management and Tumor Surveillance Recommendations of Inherited Mismatch Repair Deficiency in Childhood.* Clin Cancer Res, 2017. **23**(11): p. e32-e37.
269. Gallinger, S., et al., *Gastrointestinal cancers and neurofibromatosis type 1 features in children with a germline homozygous MLH1 mutation.* Gastroenterology, 2004. **126**(2): p. 576-85.
270. Herkert, J.C., et al., *Paediatric intestinal cancer and polyposis due to bi-allelic PMS2 mutations: case series, review and follow-up guidelines.* Eur J Cancer, 2011. **47**(7): p. 965-82.
271. Levi, Z., et al., *The gastrointestinal manifestation of constitutional mismatch repair deficiency syndrome: from a single adenoma to polyposis-like phenotype and early onset cancer.* Clin Genet, 2015. **88**(5): p. 474-8.
272. Rodriguez-Abreu, D., A. Bordoni, and E. Zucca, *Epidemiology of hematological malignancies.* Ann Oncol, 2007. **18 Suppl 1**: p. i3-i8.
273. Hall, A., et al., *Supportive care needs of hematological cancer survivors: a critical review of the literature.* Crit Rev Oncol Hematol, 2013. **88**(1): p. 102-16.

274. Burkhardt, B., et al., *The impact of age and gender on biology, clinical features and treatment outcome of non-Hodgkin lymphoma in childhood and adolescence.* Br J Haematol, 2005. **131**(1): p. 39-49.
275. Durno, C., et al., *Recommendations on Surveillance and Management of Biallelic Mismatch Repair Deficiency (BMMRD) Syndrome: A Consensus Statement by the US Multi-Society Task Force on Colorectal Cancer.* Gastroenterology, 2017. **152**(6): p. 1605-1614.
276. Prindle, M.J., E.J. Fox, and L.A. Loeb, *The mutator phenotype in cancer: molecular mechanisms and targeting strategies.* Curr Drug Targets, 2010. **11**(10): p. 1296-303.
277. Campbell, B.B., et al., *Comprehensive Analysis of Hypermutation in Human Cancer.* Cell, 2017. **171**(5): p. 1042-1056 e10.
278. Loeb, L.A., C.F. Springgate, and N. Battula, *Errors in DNA replication as a basis of malignant changes.* Cancer Res, 1974. **34**(9): p. 2311-21.
279. Bouffet, E., et al., *Immune Checkpoint Inhibition for Hypermutant Glioblastoma Multiforme Resulting From Germline Biallelic Mismatch Repair Deficiency.* J Clin Oncol, 2016. **34**(19): p. 2206-11.
280. De Vos, M., et al., *PMS2 mutations in childhood cancer.* J Natl Cancer Inst, 2006. **98**(5): p. 358-61.
281. Modi, M.B., et al., *First reported case of alveolar soft part sarcoma in constitutional mismatch repair deficiency syndrome tumor spectrum - diagnosed in one of the siblings with constitutional mismatch repair deficiency.* South Asian J Cancer, 2017. **6**(1): p. 41-43.
282. Rengifo-Cam, W., et al., *A 30-Year-Old Man with Three Primary Malignancies: A Case of Constitutional Mismatch Repair Deficiency.* ACG Case Rep J, 2017. **4**: p. e34.
283. Kaimakliotis, P., et al., *A rare case of Crohn's ileitis in a patient with constitutional mismatch repair deficiency.* Ann Gastroenterol, 2017. **30**(3): p. 370-372.
284. Mork, M.E., et al., *Identification of a novel PMS2 alteration c.505C>G (R169G) in trans with a PMS2 pathogenic mutation in a patient with constitutional mismatch repair deficiency.* Fam Cancer, 2016. **15**(4): p. 587-91.
285. Wimmer, K., T. Rosenbaum, and L. Messiaen, *Connections between constitutional mismatch repair deficiency syndrome and neurofibromatosis type 1.* Clin Genet, 2017. **91**(4): p. 507-519.
286. Flores-Rozas, H. and R.D. Kolodner, *The Saccharomyces cerevisiae MLH3 gene functions in MSH3-dependent suppression of frameshift mutations.* Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12404-9.
287. Lipkin, S.M., et al., *MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability.* Nat Genet, 2000. **24**(1): p. 27-35.
288. Peltomaki, P., *Epigenetic mechanisms in the pathogenesis of Lynch syndrome.* Clin Genet, 2014. **85**(5): p. 403-12.
289. Olkinuora, A., et al., *Biallelic germline nonsense variant of MLH3 underlies polyposis predisposition.* Genet Med, 2018.
290. Taylor, N.P., et al., *MLH3 mutation in endometrial cancer.* Cancer Res, 2006. **66**(15): p. 7502-8.
291. Carethers, J.M. and B.H. Jung, *Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer.* Gastroenterology, 2015. **149**(5): p. 1177-1190 e3.

292. Umar, S., *Intestinal stem cells.* Curr Gastroenterol Rep, 2010. **12**(5): p. 340-8.
293. Blanpain, C., V. Horsley, and E. Fuchs, *Epithelial stem cells: turning over new leaves.* Cell, 2007. **128**(3): p. 445-58.
294. Ricci-Vitiani, L., et al., *Colon cancer stem cells.* J Mol Med (Berl), 2009. **87**(11): p. 1097-104.
295. Mundade, R., et al., *Genetic pathways, prevention, and treatment of sporadic colorectal cancer.* Oncoscience, 2014. **1**(6): p. 400-6.
296. Powell, S.M., et al., *APC mutations occur early during colorectal tumorigenesis.* Nature, 1992. **359**(6392): p. 235-7.
297. Esteller, M., et al., *Analysis of adenomatous polyposis coli promoter hypermethylation in human cancer.* Cancer Res, 2000. **60**(16): p. 4366-71.
298. Otori, K., et al., *Infrequent somatic mutation of the adenomatous polyposis coli gene in aberrant crypt foci of human colon tissue.* Cancer, 1998. **83**(5): p. 896-900.
299. Miyaki, M., et al., *Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors.* Cancer Res, 1994. **54**(11): p. 3011-20.
300. Cottrell, S., et al., *Molecular analysis of APC mutations in familial adenomatous polyposis and sporadic colon carcinomas.* Lancet, 1992. **340**(8820): p. 626-30.
301. Vaughn, C.P., et al., *Frequency of KRAS, BRAF, and NRAS mutations in colorectal cancer.* Genes Chromosomes Cancer, 2011. **50**(5): p. 307-12.
302. Palmirotta, R., et al., *A novel K-ras mutation in colorectal cancer. A case report and literature review.* Anticancer Res, 2009. **29**(8): p. 3369-74.
303. Pruitt, K. and C.J. Der, *Ras and Rho regulation of the cell cycle and oncogenesis.* Cancer Lett, 2001. **171**(1): p. 1-10.
304. Nakayama, M. and M. Oshima, *Mutant p53 in colon cancer.* J Mol Cell Biol, 2019. **11**(4): p. 267-276.
305. Fleming, N.I., et al., *SMAD2, SMAD3 and SMAD4 mutations in colorectal cancer.* Cancer Res, 2013. **73**(2): p. 725-35.
306. Miyaki, M., et al., *Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis.* Oncogene, 1999. **18**(20): p. 3098-103.
307. Cunningham, J.M., et al., *Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability.* Cancer Res, 1998. **58**(15): p. 3455-60.
308. Miyaki, M., et al., *Frequent mutation of beta-catenin and APC genes in primary colorectal tumors from patients with hereditary nonpolyposis colorectal cancer.* Cancer Res, 1999. **59**(18): p. 4506-9.
309. Capper, D., et al., *BRAF V600E-specific immunohistochemistry for the exclusion of Lynch syndrome in MSI-H colorectal cancer.* Int J Cancer, 2013. **133**(7): p. 1624-30.
310. Oliveira, C., et al., *Distinct patterns of KRAS mutations in colorectal carcinomas according to germline mismatch repair defects and hMLH1 methylation status.* Hum Mol Genet, 2004. **13**(19): p. 2303-11.
311. Toyota, M., et al., *Distinct genetic profiles in colorectal tumors with or without the CpG island methylator phenotype.* Proc Natl Acad Sci U S A, 2000. **97**(2): p. 710-5.
312. Ansa, B.E., et al., *Evaluation of Colorectal Cancer Incidence Trends in the United States (2000-2014).* J Clin Med, 2018. **7**(2).

313. Vuik, F.E., et al., *Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years.* Gut, 2019. **68**(10): p. 1820-1826.
314. Siegel, R.L., A. Jemal, and E.M. Ward, *Increase in incidence of colorectal cancer among young men and women in the United States.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(6): p. 1695-8.
315. Meza, R., et al., *Colorectal cancer incidence trends in the United States and United kingdom: evidence of right- to left-sided biological gradients with implications for screening.* Cancer Res, 2010. **70**(13): p. 5419-29.
316. Ahnen, D.J., et al., *The increasing incidence of young-onset colorectal cancer: a call to action.* Mayo Clin Proc, 2014. **89**(2): p. 216-24.
317. Nath, J., et al., *Rectal cancer in young adults: a series of 102 patients at a tertiary care centre in India.* Colorectal Dis, 2009. **11**(5): p. 475-9.
318. Pal, M., *Proportionate increase in incidence of colorectal cancer at an age below 40 years: an observation.* J Cancer Res Ther, 2006. **2**(3): p. 97-9.
319. Peedikayil, M.C., et al., *Colorectal cancer distribution in 220 Indian patients undergoing colonoscopy.* Indian J Gastroenterol, 2009. **28**(6): p. 212-5.
320. Stigliano, V., et al., *Early-onset colorectal cancer: a sporadic or inherited disease?* World J Gastroenterol, 2014. **20**(35): p. 12420-30.
321. Laskar, R.S., S.K. Ghosh, and F.R. Talukdar, *Rectal cancer profiling identifies distinct subtypes in India based on age at onset, genetic, epigenetic and clinicopathological characteristics.* Mol Carcinog, 2015. **54**(12): p. 1786-95.
322. Sudarshan, V., et al., *Colorectal cancer in young adults in a tertiary care hospital in Chhattisgarh, Raipur.* Indian J Cancer, 2013. **50**(4): p. 337-40.
323. Saluja, S.S., et al., *Young onset colorectal cancer: How does it differ from its older counterpart?* Indian J Cancer, 2014. **51**(4): p. 565-9.
324. Kirzin, S., et al., *Sporadic early-onset colorectal cancer is a specific sub-type of cancer: a morphological, molecular and genetics study.* PLoS One, 2014. **9**(8): p. e103159.
325. Mauri, G., et al., *Early-onset colorectal cancer in young individuals.* Mol Oncol, 2019. **13**(2): p. 109-131.
326. Raman, R., et al., *Evidence for possible non-canonical pathway(s) driven early-onset colorectal cancer in India.* Mol Carcinog, 2014. **53 Suppl 1**: p. E181-6.
327. Lieu, C.H., et al., *Comprehensive Genomic Landscapes in Early and Later Onset Colorectal Cancer.* Clin Cancer Res, 2019.
328. Kohler, E.M., et al., *Functional definition of the mutation cluster region of adenomatous polyposis coli in colorectal tumours.* Hum Mol Genet, 2008. **17**(13): p. 1978-87.
329. Tan, C. and X. Du, *KRAS mutation testing in metastatic colorectal cancer.* World J Gastroenterol, 2012. **18**(37): p. 5171-80.
330. Palles, C., et al., *Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas.* Nat Genet, 2013. **45**(2): p. 136-44.
331. Inamura, K., *Colorectal Cancers: An Update on Their Molecular Pathology.* Cancers (Basel), 2018. **10**(1).
332. Royce, S.G., et al., *The role of SMAD4 in early-onset colorectal cancer.* Colorectal Dis, 2010. **12**(3): p. 213-9.

333. Perea, J., et al., *SMAD4 in early onset colorectal cancer.* Colorectal Dis, 2010. **12**(9): p. 948.
334. Tsukiyama, T., et al., *Molecular Role of RNF43 in Canonical and Noncanonical Wnt Signaling.* Mol Cell Biol, 2015. **35**(11): p. 2007-23.
335. Khan N. 2017. Study of Mutation Landscape and Stepwise Carcinogenesis in Hereditary Colorectal Cancers (Unpublished PhD thesis). Homi Bhabha National Institute, Mumbai. India