Study of Mutation Landscape and Stepwise Carcinogenesis in Hereditary Colorectal Cancers

By

NIKHAT NAHID KHAN

[LIFE09201004003]

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

March 2018

Homi Bhabha National Institute

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by Ms. Nikhat Nahid Khan titled **"Study of mutation landscape and stepwise carcinogenesis in Hereditary Colorectal Cancer"** and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chairman – Dr. Neelam Shirsat — Date: 27-03-2018

Guide/Convener – Dr. Rajiv Sarin — Date: 27-03-18

External Examiner – Dr. Murali Dharan Bashyam — Date: 27-3-2018

Member – Dr. Mukta Ramadwar — Date: 27/03/18

Member – Dr. Shailesh Shrikhande — Date: 27/3/18

Technical Advisor – Dr. Rita Mulherkar — Date: 27.03.18

Technical Advisor – Dr. Pradnya Kowtal — Date: 27/3/2018

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I hereby certify that I have read this thesis prepared under my direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: 27-03-18

Place: Navi Mumbai

Dr. Rajiv Sarin
Guide

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.

Nikhat.

Nikhat Nahid Khan

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.

Nikhat.

Nikhat Nahid Khan

List of Publications arising from the thesis

Journal

1. "Novel mutations and phenotypic associations identified through APC, MUTYH, NTHL1, POLD1, POLE gene analysis in Indian Familial Adenomatous Polyposis cohort", **Nikhat Khan,** Anuja Lipsa, Gautham Arunachal, Mukta Ramadwar & Rajiv Sarin, *Scientific Reports*, **2017**.

Chapters in books and lectures notes: **NIL**

Conferences

1. Presented a poster on "**Familial Adenomatous Polyposis (FAP) due to *APC* gene germline mutation: Genotype - Phenotype correlation in an Indian FAP cohort**" at *31st Annual Convention of the Indian Association of Cancer Research (IACR)* held at ACTREC, TMC in January, 2012.

2. Oral presentation on **"Molecular Analysis of the Adenomatous Polyposis Coli (APC) & MutYH Gene and its Phenotypic correlation in an Indian Familial Adenomatous Polyposis (FAP) Cohort"** at *32nd Annual Convention of the Indian Association of Cancer Research (IACR)* held at ACBR Institute, University of Delhi in February 2013.

3. Invited lecture on **"Genetics of colorectal cancer"** at *Indian chapter of IHPBA* organized by TMC and held at Renaissance Convention Center Hotel, Mumbai in January 2014.

4. Poster presentation on **"Comprehensive molecular analysis of Mismatch Repair (MMR) Gene in Indian Hereditary Nonpolyposis Colorectal cancer (HNPCC) families"** at *2nd Indian Cancer Genetics Conference and Workshop (ICGCW)* held at ACTREC, TMC in November 2014.

5. Invited lecture on **"NGS for identifying genetic basis of rare Mendelian disorders"** at *2nd Indian Cancer Genetics Conference and Workshop (ICGCW)* held at ACTREC, TMC in November 2014.

6. Best poster award received for poster presentation on **"A novel and recurrent Lynch syndrome associated Indian Founder mutation in Mismatch Repair (MMR) gene MLH1"** at *40th Annual Conference of Indian Society of Human Genetics (ISHG)* organized by NIIH and held at Nehru Science Centre, Mumbai in January 2015.

7. Best poster award for poster presentation on **"Establishing the founder effect of novel and recurrent MMR gene mutations in Indian Lynch Syndrome patients"** at *3rd Indian Cancer Genetics Conference and Workshop (ICGCW)* held at ACTREC, TMC in December 2016.

Nikhat.

Nikhat Nahid Khan

This thesis is dedicated to my Parents,

Jalaluddin Hawaldar and Najma Jalaluddin,

My husband Rashid Ahmad

and

My Beloved Daughter Rihaam Rashid Khan

ACKNOWLEDGEMENTS

It gives me an immense pleasure to express my gratitude to all of them who extended their help towards the successful completion of my PhD project. This tenure has been a long journey of 7 years in which I associated with so many people, each of whom taught me various lessons of life. Apart from learning science, I also learnt important philosophies of life that have had a great impact on my perception of life.

I begin by thanking the Almighty ALLAH for the wonderful life, family, friends and opportunities that He has given me. He has always showered His mercy on me and I shall always strive to have faith in Him and shall abide by His orders.

First and foremost, I begin to pay my sincere gratitude to my Guide, Professor Rajiv Sarin. This work would not have been possible without his guidance, support and encouragement. He gave me the opportunity to work with him, constantly provided with important inputs and arranged for any infrastructure whenever needed. He is an institution in himself from whom you have a lot to learn and I tried my best to absorb many skills from him, though I succeeded in only a few. Right from improving my presentation skills to enhancing my confidence, he has contributed everywhere. He always shared his life experiences to enrich my scientific knowledge as well as general understanding of life and taught me how to withstand pressure in life. He has been very kind and patient during the times of crisis in my personal life, of which there have been so many episodes during this tenure of PhD. He gave me all the help, support, and more importantly freedom and flexibility to work as per my convenience especially towards the end of my PhD tenure. His encouragement, good wishes and blessings will always be cherished by me. Thank you Sir for always understanding my problems and standing by my side. I owe you a lot.

I am also very grateful to Dr. Pradnya Kowtal (Co-Investigator, Sarin Lab). She has been very helpful and has sometimes gone out of the way to help me. She taught me the basics of molecular biology and always helped me with her valuable suggestions. Being trained under her supervision have helped me immensely as she imbibed in me good lab practices, management of time and resources and being economical

wherever you can. Her sincerity and dedication towards work is appreciable and inspiring. Hope I can work like her in the future.

I would like to express my sincerest gratitude to the Director of ACTREC Dr. S.V. Chipulnkar for providing me such a wonderful infrastructure to work with. I am also thankful to Dr. Sudeep Gupta (Deputy Director) and Dr. Surekha Zingde (Ex-Deputy 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.

My heartfelt thanks go to my Doctoral Committee members. I am thankful to Dr. Neelam Shirsat, Dr. Rita Mulherkar, Dr. Mukta Ramadwar, Dr. Shailesh Shrikhande, and Dr. Pradnya Kowtal for their valuable suggestions and encouragement to continue my work. I am also thankful to my ex-Chairperson Dr. Sheshadri and Ex-DC member, Dr Chauhan for their suggestions and best wishes.

I would like to appreciate the hard work done by the Genetic Counselors at the Cancer Genetics Clinic in recruiting the patients on which this thesis work was carried out. Without this integral help, this thesis would not have been possible. I thank the current and ex-members of Cancer Genetics Clinic, from the counselors Nina Bhatnagar, Payal Manek, Gouri Pandit, Areeba Khan, Vandana Kembhavi, Gayathri Jeyaraman and Ravindra Reddy to the trainees Meera Clytone, Mithila Belekar, Sanchi, Dhvani and Nidhi. I would also like to thank our collaborator, Dr. Gautham Arunachal (CMC, Vellore) for sending us samples. I also thank Mr. Santosh Nikam for his help.

I would like to acknowledge all the help I received from past and present members of the ICGC-Lab (ACTREC). Be it any reagent, machines, protocols or advice, the members of this lab are always ready to help. Special thanks to Dr. Poonam Gera for helping me with histopathological evaluation of the tissue samples. On the personal note also, she has been very motivating and always shown me a positive outlook towards life. Also, special mention is deserved by Dr. Amin, Dr. Kshama, Dr. Rajshri,

Sayee, Malcolm, Ranjeeta, Nishanth, Dr. Pooja, Rahul and others for their help and co-operation.

I would like to thank the DMG-GI team (Tata Memorial Hospital) for referring the patients to CGC and special thanks to Dr. Shrikhande and Dr. Saklani for helping me in collecting the tumor tissue samples. I am also thankful to Dr. Amin, Mr. Dudhal, Mr. Madan from Biorepository (ACTREC), and Mr. Anand and Ms Lalita (Tumor Tissue Repository TMH) for their help in co-ordinating and collecting 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. Special mention to Mr. Sakpal from Animal House Facility (ACTREC) for his timely help in making slides and cutting sections from FFPE blocks.

I am out of words to acknowledge the help extended by Mr. Naresh Mahida and Mrs. Sharada Haralkar from Genomics Facility, ACTREC. I thank them for not only helping me with sequencing and fragment analysis applications but also for the encouragement, trust, love, care and affection that they gave me during my PhD tenure. I have cherished a good 7 years long friendship and bonding with them and they will always be an important part of my life forever.

This acknowledgement section would be incomplete if I do not thank the members of Sarin Lab. I acknowledge the rich and supportive environment provided by Sarin Lab members. I feel very lucky to be a part of this lab where you will never be left alone with your own problems and I am glad that people here just don't mind their own business. If an experiment of one member fails, the entire lab will discuss the issue and help troubleshoot the problems. This is how your problem may never seem a problem to you. I thank each and every member of Sarin lab for the big and small helps provided to me in my time of need. The friendly and healthy atmosphere in this lab nourished me to grow and develop my personality. I would like to thank Anjana, Amrit, Mandar, Jyoti, Jaymin, Disha, Priyanka, Suresh, Roshan, Shujath, Moquit, Amit, Deepak, Charnit, Imran, Arpita, Asmita, Kushal, Vasudha, Dr. Pallavi Shukla, Anuja, Prerana Dange, Afrin and Mr. Parag Madankar. I am thankful to my trainees,

Prateik, Zeenat, Kaivalya and Ruchika and other trainees in the lab including Bhanu, Pallavi, Purvi, Haritha, Sachin, Tushar, Priyanka, Rucha, Himani, Satish, Aradhana, Gunjan, Nivedita, Flavia, Bhagyashree and the list goes on.

My heart is filled with joy while expressing my deepest gratitude to Jyoti, Roshan and Moquit for the tremendous, immeasurable and uncountable support that they gave me here in ACTREC. They are the perfect definition of what people call as "Friend". They truly are the assets of my life and I treasure their friendship more than anything in life. They have been my shield, my voice, my support system and I could not have asked God for better friends than them. I just can't thank them enough for whatever they have done for me. Whenever, I was stuck at any experiment or I fell short of time, their helping hands were always there to support me. They have been by my side through all the thicks and thins of my life for which I will always be indebted to them.

I acknowledge the academic environment of ACTREC for supporting my work in any possible way. I thank people from Genomics, Common Instrument Facility, Library, Stenopool, SCOPE cell, Photography, IT Department, Security, Accounts and Administration for their timely support. A special thanks to Ms. Maya Dolas for all her help. Mr. Nikhil Gadewal (BTIS), Rohan Chaubal and Pallavi Pandharipande deserves special credit for helping me in various Bioinformatics based problems.

I appreciate the generosity and cooperation shown by all the patients and healthy donors by giving their blood samples and consenting to be a part of the study. I would also like to thank Mr. Habib Dawood and Mr. Noor Mohammad Kadiwala and the entire Shia Jafferi Mashaikhi Momin Welfare Association for helping me in carrying out community level study for founder mutations.

My joy has no limits when I am acknowledging my family for the life they gave me. I feel very blessed to be a member of this family where everyone loves me unconditionally. I have no words to express my feelings for my father who has always treated me like a Prince and not a Princess. He has been the sole reason and motivation for me to pursue Ph.D. He has always been a torch bearer in my life and I

cannot forget what all he has sacrificed in life so that I can be successful in all my endeavours. My mother has made a supreme contribution for me to accomplish my dreams. The sense of joy and pride that she shows whenever I achieve something in life is the fuel that drives me through all the hurdles of life. My parent in laws also deserves great appreciation for their patience. My husband's patience and the freedom that he gave me to pursue my degree are exemplary. I owe a lot to him and my daughter Rihaam as they had to sacrifice their special moments whenever I could not be around with them. My brother Suhail Khan and my bhabhis Ayesha and Fahmida deserve a very BIG THANK YOU as they are the ones because of whom I could complete my PhD as they took care of my daughter and house whenever I could not. I would also like to thank the junior members of the family, Humaira, Faiza, Tuba, Zubiya, Aliza, Zoya, Uzair, Aquib and Yahya for their love, care and support. My entire family is a boon for me for they have always been there to share, be it my happiness or sorrows.

Nikhat.

Nikhat Nahid Khan

TABLE OF CONTENTS

SYNOPSIS

Homi Bhabha National Institute

SYNOPSIS OF Ph.D. THESIS

1. **Name of the Student:** Nikhat Nahid Khan
2. **Name of the Constituent Institution:** ACTREC-TMC
3. **Enrolment No. and Date of Enrolment:** LIFE09201004003, 27th July 2010
4. **Title of the Thesis:** "Study of mutation landscape and stepwise carcinogenesis in Hereditary Colorectal Cancer"
5. **Board of Studies**: Life Science, HBNI

SYNOPSIS

Introduction:

Colorectal cancer (CRC) is the third most common cancer in men and second most common cancer in women worldwide [1]. There is a wide variation in the incidence of CRC across the globe with the highest incidence in New Zealand, Australia, Europe and North America and lowest in Western Africa and Asia [1]. Nearly 55% of all the cases occur in the developed countries but the mortality rates are relatively higher in the less developed parts of the world. In India, CRC has an incidence of more than 5 per 100000 in both men and women (36917 CRC cases in men and 27415 CRC cases in women in India in 2012) [2].

Colorectal cancer is a complex disease with the etiological and pathogenetic mechanisms underlying CRC being very heterogenous. Contributory agents and mechanisms in CRC include dietary and lifestyle factors and inherited and somatic mutations. The lifetime risk of CRC is 5-6% in the general population [3]. Three quarters of the CRC cases are sporadic and 25% of the cases may have a major hereditary component. In about 15% of the cases, there is aggregation of CRC and/or

adenomas in the families, but with no apparent association with an identifiable hereditary syndrome, collectively known as familial CRC. However, 5-10% of the CRC cases occur in a setting with family history and clinical features that are suggestive of highly penetrant Mendelian syndrome that predisposes to CRC development. 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).

Familial Adenomatous Polyposis (FAP) is a rare, autosomal dominant colorectal cancer predisposition syndrome which is characterized by the presence of hundreds and thousands of colorectal polyps [4]. As the penetrance is nearly 100%, one of these polyps may progress to colorectal cancer in the absence of any surgical intervention. Although CRC stands out as the most prevalent complication, FAP is actually a multisystem disorder. Individuals with FAP therefore, also have an increased risk of developing upper GI cancers, papillary thyroid cancers, hepatoblastoma, brain tumors (particularly medulloblastoma), pancreatic cancers as well as some benign manifestations like desmoid tumors, osteomas, congenital hypertrophy of the retinal pigment epithelium (CHRPE) and dental abnormalities [5]. In the attenuated variant of FAP (AFAP), which is characterized by much fewer polyps (usually 10-30 polyps), cancer may develop a decade later as compared to classical FAP.

FAP is caused due to germline mutations in *APC*, a tumor suppressor gene located on chromosome 5q21. The gene spans 15 exons with exon 15 accounting for 75% of the coding sequence. *APC* gene encodes a multifunctional protein 2843amino acid long having a molecular weight of 310Kda [6]. APC is an integral part of the Wnt-signalling pathway and in the colonic epithelium its most important function is to sequester the growth stimulatory effects of ß-catenin, a protein that transcriptionally activates growth-associated genes in conjunction with tissue-coding factors. Majority

(60-70%) of the FAP associated *APC* mutations are truncating (frameshift or nonsense) mutations and located in the Mutation Cluster Region (codon 1286-1580). Up to 25% of FAP cases seem to be caused by *de novo* germline mutations in *APC* gene and therefore do not show characteristic pattern of inheritance. Various studies have shown that in 20% of the *APC* mutation negative FAP cases, biallelic germline mutations are identified in the MutYH (MYH) gene [7]. Recently, two more adenomatous polyposis syndrome have been described; the autosomal recessive NTHL1 associated polyposis (NAP) and the autosomal dominant Polymerase Proofreading Associated Polyposis (PPAP) due to germline mutations in exonuclease domain of POLD1 and POLE genes [8-9].

Lynch Syndrome is an autosomal dominant, highly penetrant (80-85%) syndrome which accounts for 2-5% of all CRC cases. Lynch Syndrome (LS) patients develop cancer of the colorectum and endometrium (second leading cancer in females of HNPCC family) and, less frequently, cancer of the stomach, urinary tract, ovaries, small bowel, and brain (Glioblastomas). HNPCC is clinically characterized by a family history of colorectal cancer at early age, predominance of tumors in the proximal colon, a high frequency of synchronous and metachronous colorectal cancers, and an association with extracolonic cancers. For the clinical diagnosis of HNPCC, various classification systems like the Amsterdam Criteria I & II and Bethesda & Revised Bethesda guidelines have been developed which uses clinical features like the age of onset and site of cancer in the proband and affected family members [10].

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* [11]. Mutations in MMR genes are found throughout these genes, without any obvious hot spots. Mutation-detection strategies must therefore cover the entire gene. MMR genes encode proteins that survey the newly replicated DNA for errors and repair all mismatched bases. Defects in MMR-genes result in replication errors and genetic instability, which can easily be observed in repetitive sequences such as microsatellites. Microsatellite instability (MSI), also known as replication error (RER), is defined as a change of

length due to either insertion or deletion of repeated units in a microsatellite within a tumor as compared with normal tissue. In 1997, a National Cancer Institute sponsored workshop recommended a panel of five microsatellite markers for detection of MSI consisting of two mononucleotide markers (*BAT-25* and *BAT-26*) and three dinucleotide repeat markers (*D2S123*, *D5S346* and *D17S250*) [12]. This panel of markers is referred to as the Bethesda panel. The selected markers exhibit high sensitivity and specificity. Using this reference panel, tumors having instability in two or more markers are defined as MSI-H (high instability), tumors having instability in one marker are defined as MSI-L (low instability), and tumors where none of the markers exhibit MSI are defined as MSS (microsatellite stable) [12].

Mutation spectrum of MMR genes and APC gene: The mutation spectrum, phenotypic characteristics and genotype-phenotype correlations for the major hereditary colorectal cancer syndromes are derived from the Caucasian studies and are extrapolated in other populations [13-16]. However, many studies on various populations have shown that though most of the features remain common, there exist certain distinct and novel genetic and phenotypic features that are unique to that population [17-18]. This underscores the need to characterize the mutation and clinical spectrum of these syndromes in various populations thereby expanding the current knowledge. Moreover, knowledge about mutation spectrum especially the recurrent and founder mutations prevalent in a particular population is very useful for developing rapid and cost-efficient strategy for genetic screening in specific syndromes. There is scarcity of data on hereditary colorectal cancer syndromes in South Asians countries including India. The magnitude of this paucity can be estimated from the fact that germline mutation in MMR genes has been reported in only 28 Indian families by three studies so far [19-21] and there is no FAP cohort reported from India.

Stepwise carcinogenesis in HCRC: As in all cancers, carcinogenesis due to germline mutations in MMR, MYH or APC gene requires accumulation of key genetic alterations. In 1990, Fearon and Vogelstein proposed the widely accepted adenoma-carcinoma sequence model for colorectal cancers which is considered as the genetic paradigm of CRC tumorigenesis [22]. However, this model was built by extrapolating

observations from individual adenomas and carcinomas collected from different patients. On the basis of the increasing frequency of various mutational events at differing stages, empirical evidence for a preferred order of stepwise genetic change was postulated. However, many studies after this model was proposed have shown that multiple alternative genetic pathways to colorectal cancer exist [23-24]. Currently, three pathways are described to explain colorectal carcinogenesis: Chromosomal instability (CIN), Microsatellite instability (MIN) and CpG island methylator phenotype (CIMP) pathway [25]. Though recent studies based on high-throughput technologies have addressed the issue of CRC molecular complexity, revealing high level of heterogeneity among tumors, stepwise carcinogenesis in the hereditary forms of CRC like FAP have been scarcely defined.

Hypothesis:

The widely accepted Fearon-Vogelstein genetic model of stepwise colorectal carcinogenesis is not representative of the majority of hereditary colorectal cancers, including FAP, and the genetic pathways are likely to be more complex than predicted by this model.

Aim of the project:

We propose to study the key genetic alterations in various forms of Hereditary Colorectal Cancers in an Indian cohort and undertake comprehensive genomic study on a unique model of paired normal – polyp – carcinoma samples to understand the stepwise carcinogenesis in hereditary colorectal cancers.

Objectives of the Study:

1 To characterize the genetic basis and pathways underlying different types of Hereditary Colorectal Cancers – MMR, APC and MYH
2 Identify novel, founder and recurrent germline mutations, if any, in Indian HCRC and study known and novel genotype-phenotype correlation.
3 To understand the key genetic alterations in stepwise colorectal carcinogenesis using the unique model system of paired blood, normal mucosa, polyps and cancer samples from clinically characterized FAP patients.

Patients and Methods

Patients were registered at Cancer Genetics Clinic in TMH after syndromic diagnosis of FAP or Lynch Syndrome 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 MMR genes and APC/MutYH genes

PCR amplification of the coding region of MMR genes (MLH1, MSH2 and MSH6) in case of Lynch syndrome and APC or MYH genes in case of FAP cases was carried out using specific primers designed for each exon of the genes. 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; P003 (MLH1 and MSH2) and P043 (APC) to detect the large genomic rearrangements (LGRs) in MMR and APC genes in the cases where no point mutation and small indels could be identified through Sanger sequencing. Data was analyzed using Coffalyser.Net software from MRC-Holland.

Haplotyping to establish founder effect of novel and recurrent mutations

In order to establish the frequency of one of the novel mutations in MLH1 gene; c.156delA identified in two unrelated families from the Shia Momin community, a cancer awareness and blood sample collection camp was organized at Jogeshwari, Mumbai with the help of community members, after authorization was given by the elder members of this community. Blood samples were collected from 400 participants of Shia Momin community after taking written informed consent from each of them. Personal and family history of cancers or other illness and other details

like age, gender and diet preferences were also documented. Part of the blood samples were spotted on the FTA cards for rapid extraction of DNA and lymphocytes were separated from rest of the sample and stored at -20°C for further analysis. Screening of the c.156delA mutation was carried out by PCR-CSGE method followed by Sanger sequencing of the samples showing aberrant band pattern on CSGE so as to confirm the mutation.

Haplotyping was carried out using a panel of 13 tightly linked polymorphic microsatellite markers to prove the founder effect of all five suspected founder mutations in MLH1 gene. These markers (Telomere-D3S3564-D3S1298-D3S3623-D3S1007-D3S1611-D3S1561-D3S2411-D3S3718-D3S3512-D3S4153-D3S3718-D3S3936-D3S2432-Centromere) spans a length of 10MB on Chromosome 3p flanking both sides of the mutations on MLH1 gene. The methodology was to PCR amplify the microsatellite markers using the fluorescently labeled primers (only one primer was labeled) followed by capillary electrophoresis (fragment analysis) of these products to identify the allele sizes of these markers for each sample. Results were analysed using GeneMapper software from Thermo Fisher.

To study stepwise carcinogenesis in hereditary colorectal cancer (FAP)

To understand the stepwise carcinogenesis, blood and tissue samples (normal/polyp/cancerous) were collected from eight unrelated FAP cases. Of these, for the 4 FAP cases, all blood-normal mucosa-polyp-carcinoma was available (complete paired samples) but in the remaining 4 cases, either polyp or tumor was not available (incomplete paired samples). Two complementary strategies were used to study the genetic alterations in stepwise colorectal carcinogenesis –

1. Sanger sequencing: For critical evaluation of the Fearon Vogelstein model, key genetic alterations indicated in this model were studied in these paired blood-normal-adenoma-carcinoma samples using Sanger sequencing. Mutation analysis of entire coding region of *KRAS* and *TP53* and MCR region of *SMAD4* genes and Exon 3 of *CTNNB1* gene (beta-catenin) were carried out in all the samples by PCR-Direct

sequencing approach. KRAS and TP53 analysis was further extended on FFPE blocks of adenoma (7 FAP cases) and paired adenoma-carcinoma (5 FAP cases) samples.

2. Whole exome sequencing: DNA was extracted from the 4 complete paired FAP 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 Nextera Rapid capture Expanded Exome kit (Illumina) which covers 62Mb target region including exons, UTRs and miRNAs and sequencing was done using Illumina Hiseq 2000. Both exome capture and sequencing was done at NIBMG, Kolkata. Sequencing was carried out at 30X coverage for blood and normal mucosa while the polyps and tumor samples were sequenced at 60X coverage. 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 version-0.6.2. Sequence duplicates were removed using Picard tools and local realignment around indels was carried out using GATK. Mpileup files were generated using Samtools, which is the acceptable format for Varscan2 that was used for variant calling. The variants were then annotated using ANNOVAR. Variant prioritization was carried out by filtering all the exonic variants.

RESULTS AND DISCUSSION

Objective 1: To characterize the genetic basis and pathways underlying different types of Hereditary Colorectal Cancers – MMR, APC and MYH

1.1 Genetic and clinical characterization of FAP patients

Mutation analysis of the APC gene in 47 FAP cases (44 classical FAP and 3 AFAP) identified 31 distinct germline mutations in 40 cases leading to mutation detection rate of 85%. Higher mutation detection in our cohort reflects the appropriateness of the clinical criteria used for syndromic diagnosis of FAP and the use of comprehensive mutation screening strategy (including MLPA analysis). Of the 31 distinct germline mutations, 14 are novel. Majority of the mutations, 28 out of 31 (90%) were truncating (16 frameshift and 12 nonsense mutations), 1 splicing mutation and 2 large

genomic rearrangements. No missense APC mutations were detected in this series of FAP patients. The mutations were found to occur between codons 197 to 1538 of the APC gene. Exon 15 harbored 70% of mutations with one third occurring in the Mutation Cluster Region of APC gene. LGRs identified in this study included a large heterozygous deletion of 11kb encompassing exon 9-13 of APC gene in one case and duplication of Promoter1B identified in 2 FAP cases. The identification of LGRs in 3 families reinforces that MLPA analysis must be included in the genetic analysis approach. Germline mutation analysis of MutYH gene was carried out in 7 APC mutation negative FAP cases which identified a homozygous nonsense mutation E466X in one AFAP family and compound heterozygous mutations G286E and R245H in another family with multiple polyposis. Mutation analysis of NTHL1 and POLD1/POLE mutations in the APC and MutYH mutation/LGR negative cases did not identify mutation in any of the cases. This can be explained by the fact that none of these cases manifested the phenotypes described for the syndromes associated with these genes.

1.2 Genetic and clinical characterization of Lynch Syndrome patients

Germline mutation analysis of the MMR genes was undertaken for 81 Lynch syndrome cases. This high mutation detection rate of 93% 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. Fifty-eight distinct deleterious germline mutations, of which 23 are novel, were identified in MMR genes in 75 LS families. Of the 58 MMR mutations, 32 were in MLH1 gene (43 families) and 24 were in MSH2 gene (30 families) and 2 mutations in MSH6 (2 families). Different types of mutation were prevalent in the two predominantly mutated genes, MLH1 and MSH2. Missense (25% of all mutations) and splice site (28% of all mutations) mutations predominate in the MLH1gene, while there is a preponderance of LGRs in the MSH2 gene (38% of all mutations). The mutations in MLH1 gene is evenly scattered throughout the gene with majority of the mutations being unique. The MSH2 mutation spectrum shows the clustering of mutations (including LGRs) in the proximal part of the gene which encodes the DNA binding domain and MSH6/MSH3

interaction domain. All the mutations in MSH2 gene were identified in one family each; except two mutations, c.942+3A>T (Intron 5) and c.340G>T (Exon 2) that were identified in 6 and 2 families respectively.

Objective 2: Identify novel, founder and recurrent germline mutations, if any, in Indian HCRC and study known and novel genotype-phenotype correlation.

Of the 31 mutations identified in APC gene, 14 were novel mutations. The worldwide recurrent mutations at codon 1309 and codon 1061 were also identified in our cohort at high frequency in 7 (18%) and 4 (9%) families respectively. This study also identified a new Indian mutational hotspot at codon 935 identified in 4 (9%) FAP families. No founder mutation was identified in APC gene. The phenotypic characteristics and the genotype-phenotype correlations observed in our cohort largely concur with the Caucasian data with some notable exceptions. Eight very rare FAP phenotype or phenotypes rarely associated with mutations outside specific regions of the APC gene were identified. Interestingly, three of the cases with novel phenotypes or phenotypic associations harbor novel mutations. This suggests that novel mutation expands the knowledge on genotype and phenotypic spectrum. Also such deviations from the established genotype-phenotype correlations reiterates that these correlations, though useful for guiding genetic testing in some cases, are not absolute and must be used in combination with clinical data for taking important decisions about genetic testing, surveillance and treatment.

There were 15 novel mutations in MLH1, 7 in MSH2 gene and 1 novel mutation in MSH6 gene which together accounted for 40% of all the MMR gene mutations. The phenotype and genotype-phenotype correlations in the MLH1 and MSH2 mutation carriers are largely in accordance with the known literature. All the mutations in MSH2 gene were identified in one family each; except two mutations, c.942+3A>T (Intron 5) and c.340G>T (Exon 2) that were identified in 6 and 2 families respectively. Six mutations in MLH1 gene were found in more than one family indicating the recurrent nature of these mutations. Five of these six recurrent mutations in MLH1 gene were identified in unrelated families from specific geo-ethnic backgrounds. A novel mutation c.156delA in Exon 2 was identified in 5 unrelated families from Shia

Momin community hailing from Patan district of Gujarat and in one Hindu family from Gujarat. Three Maratha families from Sindhudurg region (Maharashtra) harbored the c.1558+2insG mutation in Intron 13. A novel frameshift mutation c.46insG (Exon 1) was identified in three unrelated Hindu families from Bankura district of West Bengal. A missense mutation E102D in Exon 3 which has earlier been reported 14 times in the InSiGHT database was identified in two unrelated Sindhi families who reported their ancestry from Sindh region of Pakistan. Another novel mutation c.1389_1390delAC (Exon 12) was identified in 2 Hindu families from Bihar. The founder effect of all these 5 mutations was confirmed by the presence of conserved haplotypes that were present in all the mutation carriers but that was not frequently seen or absent in the control population.

Objective 3: To understand the key genetic alterations in stepwise colorectal carcinogenesis using the unique model system of paired blood, normal mucosa, polyps and cancer samples from clinically characterized FAP patients.

After whole exome sequencing, 78.27%–86.13% of reads were mapped to targets, and the mean target coverage was between 25X and 40X for the blood and mucosa samples whereas it was between 81X to 98X for the adenoma (polyp) and carcinoma samples (tumor). Exome analysis revealed a total of 20610 somatic variants within 12 samples analyzed (4 pairs of mucosa-adenoma-tumors) with the average number of variants per sample being 1716 (range: 1368-2231). Most of the variants were intergenic and intronic variants and the exonic variants accounted for only 3% of the total variants (634 exonic variants), of which 208 were exonic non-synonymous SNVs. The overall mean frequency of exonic variants in all samples was 53 (range: 32-80). The mean number of exonic variant in the 4 polyp samples were 46 while it was 55 for the tumor samples. Remarkably, the vast majority of the identified variants were private events with a small fraction of variants that were shared by more than one sample at any stage. Moreover, most of the shared variants were intergenic, intronic or silent variants or were reported at a very high frequency in the publicly available databases (dbSNP and 1000 genomes). A striking observation was the absence of somatic variants in the KRAS, TP53 and SMAD genes, all of which are reported to be key players that are implicated in the CRC carcinogenesis pathways

(Vogelstein's model) [22]. We also did not identify any variants in the other members of the wnt-signaling pathway and TGF-β signaling pathway.

Sanger sequencing of KRAS and TP53 genes in a total of 18 adenomas and 11 carcinomas from 20 FAP cases revealed the presence of KRAS mutation G12V in one carcinoma sample and TP53 mutations in 3 carcinomas (2 missense and 1 frameshift mutation). This frequency of KRAS and TP53 mutations is not in accordance with the reported frequencies of mutations in these genes in CRC [22]. The absence of KRAS mutation in the 18 adenoma samples studied is noteworthy as KRAS mutation is regarded as an early event in CRC tumorigenesis and is identified in up to 50% of the late stage adenomas. This observation can be supported by other studies which have also shown a low frequency or absence of KRAS mutations in adenoma or carcinoma samples from FAP patients harboring germline APC mutations [26-27]. Altogether, the exome sequencing and the Sanger sequencing data raises a strong possibility that the adenoma-carcinoma sequence model proposed by Fearon and Vogelstein does not explain a large fraction of FAP tumorigenesis. However, no definite conclusions can be made owing to the small sample size studied in this project and merits further studied with a large number of paired samples from well characterized FAP patients.

Summary and Conclusions

The first report on genetic spectrum in Indian FAP patients has confirmed the predominance of APC mutations and rarity of MYH, NTHL1 and POLD1/POLE gene mutations. 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 in APC and MMR genes has been identified in our study. The genotype and phenotype and the genotype-phenotype associations in our FAP and LS cohort largely concurs with the known spectra with few notable exceptions. Ours is the first study from India that has identified 5 founder mutations in MLH1 gene. Establishment of rapid and cost-effective mutation screening strategies for APC and MMR genes seem plausible owing to the identification of large number of novel, recurrent and founder mutations in our cohort. Study on stepwise carcinogenesis using paired samples from

FAP patients point towards a possibility that the widely accepted Fearon Vogelstein model is not the mechanistic pathway for CRC tumorigenesis in FAP cases with germline APC mutations as the driver mutation. However, this needs to be confirmed by extending the study with a large number of paired samples from FAP patients, which may also help in delineating the alternate pathway that explain CRC tumorigenesis in FAP and other hereditary colorectal cancer.

REFERENCES

1. Ferlay J et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136, E359–E386 (2015).
2. Mallath M K et al. The growing burden of cancer in India: epidemiology and social context. Lancet Oncology Series Cancer burden and health systems in India 1 (2014).
3. Jemal A et al. Cancer statistics, 2008. CA: a cancer journal for clinicians 58 (2): 71–96 (2008).
4. Hisamuddin I M and Yang V W. Genetics of Colorectal Cancer. Medscape General Medicine 6(3): 13 (2004).
5. Groen E J et al. Extra-Intestinal Manifestations of Familial Adenomatous Polyposis. Annals of Surgical Oncology 15(9): 2439–2450 (2008).
6. Koji Aoki and Makoto M Taketo**.** Adenomatous polyposis coli (APC): a multi-functional tumor suppressor gene. Journal of Cell Science 120: 3327-3335 (2007).
7. Skrzypczak M et al. MYH gene status in Polish FAP patients without APC gene mutations. Hereditary Cancer in Clinical Practice 4: 43-47 (2006).
8. Weren R D et al. A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. Nature Genetics 47: 668–671 (2015).
9. Palles C et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nature Genetics 45(2): 136–144 (2013).
10. Vasen H F A. Clinical Diagnosis and Management of Hereditary Colorectal Cancer Syndromes**.** *Journal of Clinical Oncology*, Vol 18, No 21S: 81s-92s (2000).
11. 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 7: 147-158 (2014).
12. 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 58: 5248–5257 (1998).

13. Friedl W and Aretz S. Familial Adenomatous Polyposis: Experience from a Study of 1164 Unrelated German Polyposis Patients. Hereditary Cancer in Clinical Practice 3: 95–114 (2005).

14. Nieuwenhuis M H and Vasen H F. Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): A review of the literature. Critical Reviews in Oncology/Hematology 61: 153–161 (2007).

15. Mangold E et al. Spectrum and frequencies of mutations in MSH2 and MLH1identified in 1,721 German families suspected of hereditary nonpolyposis colorectal cancer. International Journal of Cancer 116: 692-702 (2005).

16. Kurzawski G et al. Germline MSH2 and MLH1 mutational spectrum including large rearrangements in HNPCC families from Poland (update study). Clinical Genetics 69: 40-47 (2006).

17. Tao H et al. Identification of 5 novel germline APC mutations and characterization of clinical phenotypes in Japanese patients with classical and attenuated familial adenomatous polyposis. BMC Research Notes 3: 305–313 (2010).

18. Kim D W et al. Mutation Spectrum of the APC Gene in 83 Korean FAP Families. Human Mutation 26: 281–292 (2005).

19. Rajkumar T et al. Mutation analysis of hMSH2 and hMLH1 in colorectal cancer patients in India. Genetic Testing 8(2): 157-62 (2004).

20. Rajender S et al. R659X mutation in the MLH1 gene in hereditary non-polyposis colorectal cancer (HNPCC) in an Indian extended family. IJMR 131: 64-70 (2010).

21. Bhasyam M et al. Evidence for presence of mismatch repair gene expression positive Lynch syndrome cases in India. Molecular Carcinogenesis 54 (12): 1807-1814 (2015).

22. Fearon and Vogelstein. A Genetic Model for Colorectal Tumorigenesis. Cell 61: 759-767 (1990).

23. Frattini M et al. Features Associated with Colon and Rectal Carcinogenesis. Clin Cancer Res 10: 4015-4021 (2004).

24. Smith G et al. Mutations in APC, Kirsten-ras, and p53—alternative genetic pathways to colorectal cancer. PNAS 99 (14): 9433-9438 (2002).

25. Worthley D et al. Colorectal carcinogenesis: Road maps to cancer. World J Gastroenterol 13(28): 3784-3791 (2007).

26. Lamlum et al. APC mutations are sufficient for the growth of early colorectal adenomas. PNAS 97 (5): 2225-2228 (2000).

27. McLellan E A et al. High frequency of K-ras mutations in sporadic colorectal adenomas. Gut 34: 392-396 (1993).

28. Kucherlapati R et al. Mouse models for human familial adenomatous polyposis. Seminars in Cancer Biology 11: 219-225 (2001).

Publications in Refereed Journal:

a. Published:

1. "Novel mutations and phenotypic associations identified through APC, MUTYH, NTHL1, POLD1, POLE gene analysis in Indian Familial Adenomatous Polyposis cohort" **(Nikhat Khan *et al.*, Scientific Reports, 2017)**

b. Accepted: **NA**

c. Communicated: **NA**

Other publications:

a. Book/ Book chapters: NA

b. Conference/ Symposium:

1. Presented a poster on "**Familial Adenomatous Polyposis (FAP) due to *APC* gene germline mutation: Genotype - Phenotype correlation in an Indian FAP cohort**" at *31st Annual Convention of the Indian Association of Cancer Research (IACR)* held at ACTREC, TMC in January, 2012.

2. Oral presentation on **"Molecular Analysis of the Adenomatous Polyposis Coli (APC) & MutYH Gene and its Phenotypic correlation in an Indian Familial Adenomatous Polyposis (FAP) Cohort"** at *32nd Annual Convention of the Indian Association of Cancer Research (IACR)* held at ACBR Institute, University of Delhi in February 2013.

3. Invited lecture on **"Genetics of colorectal cancer"** at *Indian chapter of IHPBA* organized by TMC and held at Renaissance Convention Center Hotel, Mumbai in January 2014.

4. Poster presentation on **"Comprehensive molecular analysis of Mismatch Repair (MMR) Gene in Indian Hereditary Nonpolyposis Colorectal cancer (HNPCC) families"** at *2nd Indian Cancer Genetics Conference and Workshop (ICGCW)* held at ACTREC, TMC in November 2014.

5. Invited lecture on **"NGS for identifying genetic basis of rare Mendelian disorders"** at *2nd Indian Cancer Genetics Conference and Workshop (ICGCW)* held at ACTREC, TMC in November 2014.

6. Best poster award received for poster presentation on **"A novel and recurrent Lynch syndrome associated Indian Founder mutation in Mismatch Repair (MMR) gene MLH1"** at *40th Annual Conference of Indian Society of Human Genetics (ISHG)* organized by NIIH and held at Nehru Science Centre, Mumbai in January 2015.

7. Best poster award for poster presentation on "Establishing the founder effect of novel and recurrent MMR gene mutations in Indian Lynch Syndrome patients" at *3rd Indian Cancer Genetics Conference and Workshop (ICGCW)* held at ACTREC, TMC in December 2016.

Signature of Student: Nilaat

Date: 21/8/17

Doctoral Committee

S. No.	Doctoral Committee	Name	Signature	Date
1.	Dr. Neelam Shirsat	Chairman		21-8-17
2.	Dr. Rajiv Sarin	Guide & Convener		18.8.17
3.	Dr. Mukta Ramadwar	Member		21.08.17
4.	Dr. Shailesh Shrikhande	Member		21/8/17
5.	Dr. Rita Mulherkar	Technical Advisor		21-8-17
6.	Dr. Pradnya Kowtal	Technical Advisor		21/8/17

Forwarded Through:

S.Chiplunkar 28/8/17

Dr. S. V. Chiplunkar
Director
Advanced Centre for Treatment, Research & Education in Cancer (ACTREC)
Tata Memorial Centre
Kharghar, Navi Mumbai-410210.

Dr. S.V. Chiplunkar
Director, ACTREC
Chairperson,
Academic & training Program, ACTREC

Prof. K. Sharma
Director, Academics
T.M.C.

PROF. K. S. SHARMA
DIRECTOR (ACADEMICS)
TATA MEMORIAL CENTRE,
PAREL, MUMBAI

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

aa	amino acids
AA	Australo-Asiatic
ACF	Aberrant Crypt Foci
ADP	Adenosine diphosphate
AFAP	Attenuated FAP
AJ	Ashkenazi Jews
ANI	Ancestral North Indian
APC	Adenomatous Polyposis Coli
ASI	Ancestral South Indian
BRCA1	Breast Cancer 1
BRCA2	Breast Cancer 2
CFS	Cancer Family Syndrome
CGC	Cancer Genetics Clinic
CHRPE	Congenital Hypertrophy of Retinal Pigment Epithelium
CIMP	CpG Island Methylator Phenotype
CIN	Chromosomal Instability
CMMRD	Constitutional Mismatch Repair Deficiency
CNS	Central Nervous System
CRC	Colo-Rectal Cancer
CSGE	Conformation Sensitive Gel Electrophoresis
CTC	Computed Tomographic Colonography
DNA	Deoxyribonucleic acid
DQ	Dosage Quotient
DR	Dravidian
EDTA	Ethylene Diamine Tetraacetic Acid
FAP	Familial Adenomatous Polyposis
FFPE	Formalin fixed Paraffin Embedded
FIT	Fecal Immunochemical Test
FV	Fearon Vogelstein
GATK	Genome Analysis Tool Kit
gFOBT	guaiac based Fecal Occult Blood Test
GI	Gastro-Intestinal
HBOC	Hereditary Breast and Ovarian Cancer
HCRC	Hereditary Colo-Rectal Cancer
HGVS	Human Genome Variation Society
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
ICMR	Indian Council for Medical Research
IE	Indo-European
iFOBT	immunochemical Fecal Occult Blood Test

IHC	Immunohistochemistry
InSiGHT	International Society for Gastrointestinal Hereditary Tumors
JPS	Juvenile Polyposis Syndrome
LFS	Li-Fraumeni Syndrome
LGR	Large Genomic Rearrangements
LOH	Loss of Heterozygosity
LS	Lynch Syndrome
MAP	MUTYH Associated Polyposis
MCR	Mutation Cluster Region
MEN1	Multiple Endocrine Neoplasia 1
MEN2	Multiple Endocrine Neoplasia 2
MIN	Microsatellite Instability
MLH1	MutL Homologue 1
MLPA	Multiplex Ligation dependent Probe Amplification
MMR	Mismatch Repair
MSH2	MutS Homologue 2
MSH6	MutS Homologue 3
MSI-H	Microsatellite Instability High
MSI-L	Microsatellite Instability Low
MSS	Microsatellite Stable
MUTYH	MutY Homologue
NAP	NTHL1 Associated Polyposis
NCBI	National Center for Biotechnology Information
NTHL1	Nth Like Endonuclease 1
OMIM	Online Mendelian Inheritance in Man
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reaction
PJS	Peutz Jeghers Syndrome
POLD1	Polymerase delta 1
POLE	Polymerase epsilon
PPAP	Polymerase Proofreading Associated Polyposis
PTC	Papillary Thyroid Cancer
RBC	Red Blood Cell
RET	Rearranged during Transfection
RFC	Replication Factor C
RNA	Ribonucleic Acid
SM	Shia Momin
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
STK11	Serine Threonine Kinase 11

Taq	Thermus aquaticus
TBE	Tris Borate EDTA
TCF	T-cell factor
TCGA	The Cancer Genome Atlas
TD	Touchdown PCR
TNM	Tumor Node Metastasis
TTE	Tris Taurine EDTA
UAP	Unexplained Adenomatous Polyposis
UCSC	University of California Santa Cruz
US	United States
UV	Ultraviolet
VUS	Variant of Unknown Significance

CHAPTER 1

Introduction

1.1 INTRODUCTION TO CANCER

Human body is made up of about one hundred trillion (10^{14}) cells, all of which are supposed to abide by certain rules and regulations that govern the growth and survival of these cells. This strictly controlled governance keeps each cell in right size, shape and architecture and helps in proper functioning of the cells [1]. Normal cells in the body divide in a controlled fashion by carefully instructing the production and release of proliferation and survival signals at respective stages of cell cycle and cell division [1]. These signals are generally the growth factors that bind to the cell surface receptors and carry on the further downstream signaling via a network of several molecules [2]. These cells die when they are worn out or are damaged, after which they no longer receive signal for cell proliferation and survival rather they receive signals for growth inhibition and cell death. The process of maintenance of cell homeostasis is outlined in **Fig. 1.1.**

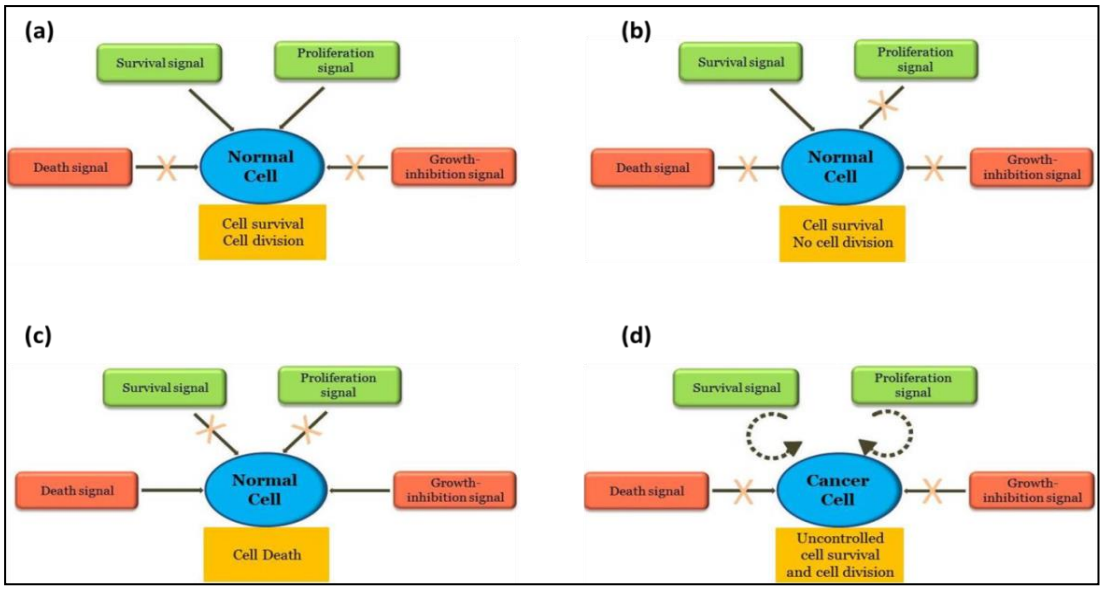


Fig.1.1. Maintenance of cell homeostasis by regulation of cell survival, cell division and cell death.

Under normal circumstances, the rate and time of new cell growth and old cell death are kept in balance in accordance with the needs of the body. However, when the normal cells become a 'rebel' breaking all the rules, then they begin to divide recklessly; stops undergoing apoptosis; develop new characteristics, starts invading other tissues and seizes body's resources thus leading to the emergence of *"Emperor of all Maladies"*, ***Cancer***.

Cancer is a major public health concern and a leading cause of death in both, the developed and the developing parts of world [3]. The burden is expected to grow worldwide owing to the growth and aging of population [4]. Other factors that contribute to the rising cancer incidences is the increasing prevalence of established risk factors which includes smoking, alcohol consumption, obesity, chronic bacterial and viral infections, change of reproductive patterns (low parity and late age of first birth), inappropriate dietary habits, sedentary lifestyle and exposure to environmental and occupational hazards including chemical carcinogens, ionizing and non-ionizing radiations [5].

In India, more than 1 million new cases of cancers are diagnosed every year and this number is expected to double by 2035 [6]. The most common cancers in Indian men include oral, lung, stomach, colorectal, pharyngeal and esophageal cancers while in women, cancers of the breast, cervix and colorectum are more common [6]. Other cancers are rare in both Indian men and women with an incidence of five or less than five per 100000 individuals respectively [6]. It is important to recognize that cancer related mortality is decreasing across all age groups in the developed countries [3]; however, no evidence exists for such decrease in developing countries like India [3] which can be attributed to the contrasting etiology like recurrent infections and unique local pattern of tobacco use, limited access to early diagnosis, treatment and lack of knowledge on preventive measures and prophylactic intervention [6].

1.1.1. Hallmarks of Cancer

Cancer represents an umbrella of term which comprises of a plethora of atleast 100 diseases affecting nearly every tissue type and organ. Although, all of these cancers are different in terms of prevalence, prognosis, histological origin and pathology, certain features are fundamental to all cancers that are elucidated as hallmarks of cancer [7].

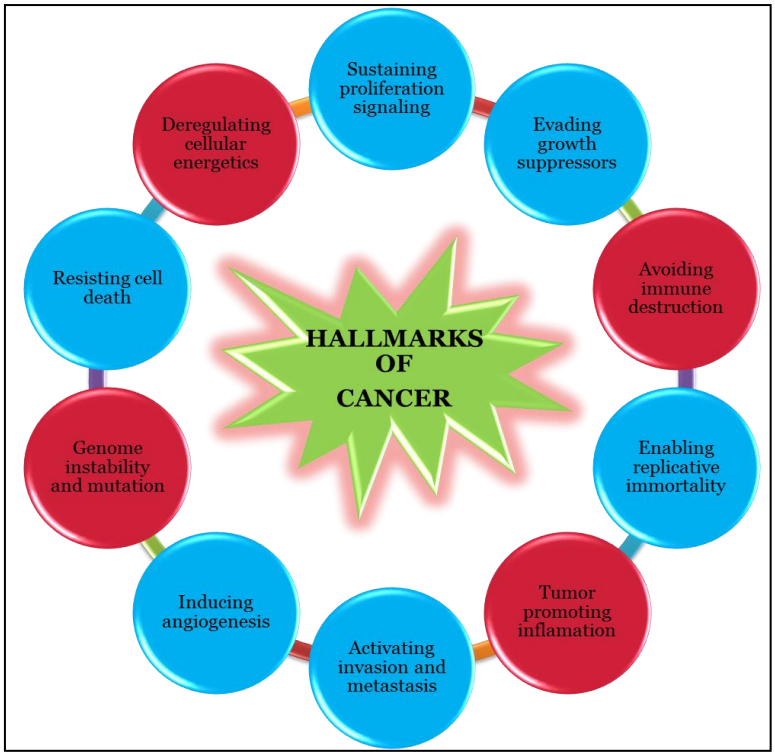


Fig 1.2 Hallmarks of Cancer (adapted from Ref. 7)

The hallmarks in the blue circles in Fig 1.2 are the ones that were originally proposed in 2000 by Hanahan and Weinberg. However, after a decade of research, four more hallmarks were added in 2011 version that is represented in the red circles in Fig 1.2.

1.1.2. Cancer Classification

Cancers are classified by the type of cell that resembles the tumor and, therefore, the tissue presumed to be the origin of the tumor. These are the histology and the location, respectively. Examples of general categories include:

- **Carcinoma:** Malignant tumors derived from epithelial cells. This group represents the most common cancers. Carcinomas can be further subdivided into adenocarcinoma (arising from glandular epithelium) or squamous cell carcinoma (arising from squamous epithelium)
- **Sarcoma:** Malignant tumors derived from connective tissue (i.e. bone, cartilage, fat, nerve). Examples include osteosarcoma, chondrosarcoma, fibrosarcoma, synovial sarcoma, liposarcoma, rhabdomyosarcoma
- **Lymphoma** and **leukemia:** Malignancies derived from hematopoietic (blood-forming) cells.

- **Germ cell tumor:** Tumors derived from pluripotent cells most often presenting in the testicle (seminoma) or ovary (dysgerminoma).
- **Blastic tumor or blastoma:** A tumor (usually malignant) which resembles an immature or embryonic tissue.

Cancer can also be classified as:

- **Hereditary Cancer**
  - Apparently autosomal dominant or recessive transmission of specific cancer(s) – family history in atleast 2-3 generations.
  - Earlier age of onset of cancers than is typical for that cancer
  - Multiple primary cancers in an individual
  - Clustering of rare cancers
  - Bilateral or multifocal cancers

- **Familial Cancer**
  - More cases of a specific type(s) of cancer within a family than statistically expected, but no specific pattern of inheritance
  - Age of onset variable
  - May result from common genetic background, similar environment and/or lifestyle factors

- **Sporadic Cancer**
  - Cancers due to non-hereditary causes occurring in a setting of no family history of cancers
  - Later age of onset (due to aging process)
  - Mostly unilateral
  - Even if there is more than one case in the family, there is no particular pattern of inheritance.

1.1.3. Cancer: A genetic disease

Cancer is the most common genetic disease which means that cancer development is driven by mutations in several genes. It is thought that cells would acquire atleast 2-8 alteration over several decades in relevant genes to initiate the process of tumorigenesis [8]. Somatic mutations in DNA can occur in the genomes of the cell upon mitotic cell division [9] or because of other reasons including chemical

carcinogens and radiations. Most of the mutations are quickly repaired by the cell's repair machinery but this machinery is not always foolproof and hence mutations may accumulate in the cell. There are about 25,000 genes in each human cell, but, in most cases, a mutation within a gene will not lead to the development of cancer. It is only when mutations occur in certain key genes that cancer develops. These key genes can be grouped broadly into two classes [10]:

- **Proto-oncogenes:**

  Proto-oncogenes code for proteins that help to regulate cell growth and differentiation. In their native state proto-oncogenes do not possess the ability to transform cells but upon "activation" due to point mutations at specific positions [8], chromosomal translocation or gene amplification, a proto-oncogene becomes a tumor inducing agent, an oncogene [11]. Oncogenes are altered in cancers in ways that render the gene constitutively active or active under situations when the wild-type is not. *RAS* genes are the most commonly mutated oncogenes in human cancers and upon acquiring a point mutation they lead to constitutive mitogenic signaling which is one of the most fundamental trait of cancer cells [12]. On cellular level these alterations act in a dominant manner, meaning that one allele is usually sufficient to confer a selective growth advantage to the cell. Other examples of proto-oncogenes include WNT, MYC, ERK, and TRK.

- **Tumor Suppressor Genes:**

  Tumor-suppressor genes encode proteins that either have a inhibiting or suppressive effect on the regulation of the cell cycle or promote apoptosis, and sometimes do both. When this gene is mutated to cause a loss or reduction in its function, the cell can progress to cancer, usually in combination with other genetic alterations. Tumor suppressor genes are generally recessive in nature which means that both the alleles need to be mutated to confer the growth advantage to cells. This principle was first put forth by Alfred Knudson in 1971 and is known as “two-hit hypothesis” [13]. According to this hypothesis, the hereditary forms of cancer may arise by two inactivating alterations, one of which is inherited germline and the other is acquired somatically later in life. However, in case of sporadic cancers, both the alterations have to be acquired somatically and therefore such cancers develop at a late age. Certain exceptions to this hypothesis

include the cases of *haploinsufficiency* wherein loss of only one allele is enough for the aberrant protein function and development of tumor [14]. Another scenario where this hypothesis is not followed is when mutation in the single copy hinders the function of the protein expressed from the wild type copy. This is known as *dominant-negative effect* of mutation [15].

In 1997, Kinzler and Bert Vogelstein grouped the tumor suppressor genes into two classes: "caretakers" and "gatekeepers" [16]. In 2004, a third classification of tumor suppressor genes was proposed; "landscaper" genes [17].

- ***Gatekeeper genes:*** Encode gene products that act to prevent growth of potential cancer cells and prevent accumulation of mutations that directly lead to increased cellular proliferation. In many cases, gatekeeper genes encode a system of checks and balances that monitor cell division and death.
- ***Caretaker genes:*** They encode products that stabilize the genome. Fundamentally, mutations in caretaker genes lead to genomic instability. These genes are critical for maintaining low mutation burden, failure of which leads to a mutator phenotype in tumor cells.
- ***Landscaper genes:*** Encode products that, when mutated, contribute to the neoplastic growth of cells by fostering a stromal environment conducive to unregulated cell proliferation. Landscaper genes encode gene products that control the microenvironment in which cells grow.

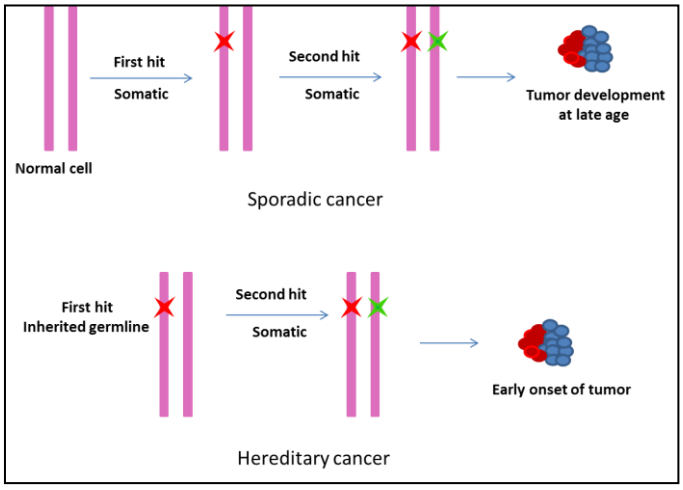


Fig 1.3 Knudson two hit hypothesis

1.1.4 Hereditary Cancer Syndromes

An estimated 5-10% of all cancers are inherited, due to highly penetrant germline mutations that cause rare inherited cancer syndromes. To date, more than 100 genes have been reported to cause inherited predisposition to cancers. While most of the syndromes are transmitted in an autosomal dominant fashion, certain syndromes follow the autosomal recessive inheritance pattern. Table 1.1 summarizes the features of some common hereditary cancer syndromes.

Table 1.1: Mendelian cancer predisposition syndromes

Syndrome	Mode of inheritance	Tumor spectrum	Genes involved
Hereditary Breast and Ovarian Cancer (HBOC)	Autosomal Dominant	Breast, Ovarian*, Prostate#, Pancreas	*BRCA1, BRCA2*
Lynch Syndrome (LS) aka Hereditary Nonpolyposis Colorectal Cancer (HNPCC)	Autosomal Dominant	Colorectum, Endometrium*, Stomach, small intestine, ovary*, liver, pancreas, ureter, brain (glioblastoma), breast*, prostate#	*MLH1, MSH2, MSH6, PMS2*
Li-Fraumeni Syndrome (LFS)	Autosomal Dominant	Sarcoma (mainly osteosarcoma), Breast*, Brain, Blood, Adrenal gland	*TP53*
Familial Adenomatous Polyposis (FAP)/Attenuated FAP (AFAP)	Autosomal Dominant	Colorectum, Upper GI, Brain (medulloblastoma***), Papillary Thyroid**, Hepatoblastoma***	*APC/ MUTYH*
MUTYH-associated polyposis (MAP)	Autosomal Recessive	Colorectum and rarely upper GI cancers	*MUTYH*
Multiple Endocrine Neoplasia (MEN1 and MEN2)	Autosomal Dominant	Thyroid (Medullary, Follicular), Pituitary, Parathyroid, Adrenal, Pancreas	*RET*
Retinoblastoma	Autosomal Dominant	Retina, pituitary	*RB1*
Xeroderma Pigmentosa	Autosomal Recessive	Skin cancers (particularly of face, eyelids, lips and sometimes tongue tip, eyes, scalp)	*XPA-XPG, XPV*

Legend:

*:only in females, #; only in males, **: preponderance in females, ***: mainly during childhood

1.2 COLORECTAL CANCER

Colorectal cancer is the disease that arises from the epithelial lining of the colon and rectum. Colon and rectum are part of the large intestine which is the terminal portion of the gastro-intestinal (GI) tract. Colon is involved in the absorption of water and nutrients from the partially digested food that is received from the small intestine, after which the waste is passed onto the rectum from where it is expelled out of the body from anus. The colon is a large muscular tube of about 1.5 meters in length and 5 centimeters in diameter and it forms the major part of large intestine while the rectum which is 6 centimeters long forms the last part. As shown in *Fig. 1.4*, the colon has four sections:

1. The ***ascending colon*** which begins with the caecum and extend upward on the right side of the abdomen.
2. The ***transverse colon*** which is so called because it traverses the body from the right to the left side. The bend between the ascending colon and the transverse colon is the *hepatic flexure*. The ascending and the transverse colon are collectively referred to as the *proximal colon*.
3. The ***descending colon*** descends on the left side of the abdomen. The bend between the transverse colon and descending colon is the *splenic flexure*.
4. The ***sigmoid colon*** named for its 'S' shape is the final part of the colon which joins the rectum. The descending and the sigmoid colon are together referred to as *distal colon*.

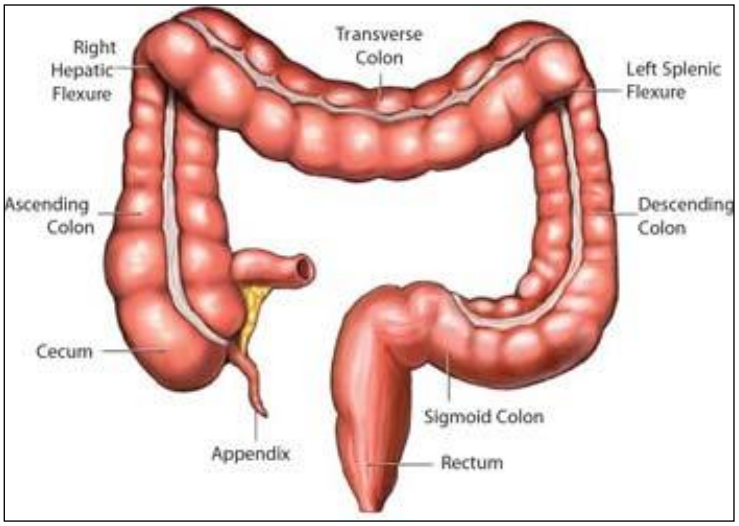


Fig 1.4 **Anatomy of the colon**

1.2.1 Epidemiology of colorectal cancer

Colorectal cancer (CRC) is the third most common cancer in men and second most common cancer in women worldwide [3]. There is a wide variation in the incidence of CRC across the globe with the highest incidence in New Zealand, Australia, Europe and North America and lowest in Western Africa and Asia [3]. Nearly 55% of all the cases occur in the developed countries but the mortality rates are relatively higher in the less developed parts of the world. In India, CRC has an incidence of more than 5 per 100000 in both men and women (36917 CRC cases in men and 27415 CRC cases in women in India in 2012) [6]

1.2.2 Symptoms of CRC

Early CRC has no specific symptoms which makes it important to undergo regular screening for early diagnosis. However, as the tumor grows, it starts to bleed and obstruct the intestine and may produce certain symptoms like:

- Blood in stools
- Dark or black stools
- Change in bowel habits
- Decreased appetite
- Diarrhea or constipation
- Unexplained weight loss
- Cramping or abdominal discomfort
- Anemia (weakness and excessive fatigue)

1.2.3 Screening for CRC

Optimal screening helps in reducing CRC related mortality both by decreasing the incidence and by increasing the overall survival likelihood. For the general population, the screening is recommended to start at the age of 50 years. However, for an individual with inherited predisposition to CRC, screening should initiate at an earlier age. Various diagnostic tests that are used for screening of colorectal cancers are outlined in Table 1.2.

Table 1.2 Diagnostic tests for CRC screening

Category	Test	Advantages	Limitations
Stool tests	gFBOT	No bowel preparation Can be performed at home Low cost Non-invasive	Requires multiple stool samples Prone to false positive results Will miss most polyps Needs confirmation by colonoscopy
	iFBOT (FIT)	No bowel preparation Can be performed at home Low cost Non-invasive	Requires multiple stool samples Prone to false positive results Will miss most polyps Needs confirmation by colonoscopy
	FIT-DNA	No bowel preparation Can be performed at home Requires only a single sample Non-invasive	More false positive results than other tests Costlier than gFBOT and iFBOT Needs confirmation by colonoscopy
Endoscopy	Flexible sigmoidoscopy	Minimal bowel preparation Quick with minimal complications Does not require sedation	Examines only one-third of colon Cannot remove large polyps
	Colonoscopy	Examines entire colon Can biopsy and remove polyps Can diagnose other diseases	Sedation usually required Requires full bowel cleansing Expensive
Radiological imaging	Barium enema	Examines entire colon Does not require sedation	Requires cleansing of entire bowel False positive results Needs confirmation by colonoscopy Exposure to low dose radiation
	Computed Tomographic Colonography (CTC)	Examines entire colon Quick No sedation required Non-invasive	Requires cleansing of entire bowel Needs confirmation by colonoscopy Exposure to low dose radiation

1.2.4 Risk factors for CRC

Colorectal cancer is a complex disease with the etiological and pathogenetic mechanisms underlying CRC being very heterogenous. The lifetime risk of CRC is 5-6% in the general population [18]. The most important risk factor for CRC like other cancers is the age as it is a disease of elderly. Apart from age, the contributing agent includes modifiable factors like diet and lifestyle or non-modifiable factors related to family and medical history [19-26]. *Table 1.3* describes the established risk factors for colorectal cancer.

Table 1.3 Established risk factors for CRC

Heredity and medical history
One or more family member diagnosed with CRC (especially at age less than 45y)
Inflammatory bowel disease
Presence of polyps
Diabetes
Behavioral factors
Alcohol consumption and smoking
Diet rich in unsaturated fatty acids (red meat) and low in dietary fibers
Obesity
Lack of physical activity

1.2.5 Staging of colorectal cancer

Staging is defined as the extent to which the cancer has spread at the time of diagnosis. It is important to determine the stage of cancer to make appropriate decisions about the treatment regime to be followed and to assess the prognosis. There are two widely used staging systems that are used for CRC, ***TNM staging*** and ***Dukes Classification*** that are described in *Table 1.4*

Table 1.4 TNM staging and Dukes Classification for CRC

Stage	Definition
T0	No evidence of primary tumor
Tis	Carcinoma in situ: intraepithelial or intramucosal
T1	Tumor invasion into submucosa
T2	Tumor invasion into muscularis propria
T3	Tumor invasion through muscularis propria
T4	Tumor 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 Group	**Dukes Classification**
Stage I: T1-2, N0, M0	Dukes A
Stage II: T3-4, N0, M0	Dukes B
Stage III: Any T, N1-2, M0	Dukes C
Stage IV: Any T, Any N, M1	Dukes D

1.3 HEREDITARY CRC SYNDROMES

While the majority (85%) of CRC cases are sporadic in nature caused due to environmental and lifestyle factors, current estimate is that 15-30% of all CRCs have a familial component given the occurrence of CRC in first and second degree relatives [28, 29]. Nearly 5% of cases occur in the background of family history and certain clinical or histopathological features that are indicative of highly penetrant, inherited CRC predisposition syndromes that are transmitted in Mendelian fashion [30, 31]. Most of the CRCs arise from polyps which are mushroom like tissue growth that protrudes from the lining of the colorectum [32]. These are usually benign but have the tendency to become malignant. There are mainly two types of colorectal polyps;

adenomatous polyps that arise from the glandular epithelium and *hamartomatous polyps* that are composed of a mixture of tissues. The adenomatous polyps may have tubular, villous and tubulovillous histology. Another class of polyps are hyperplastic polyps that are generally benign and do not have a potential for malignancy. Hereditary CRC (HCRC) syndromes are classified broadly on the basis of the presence or absence of polyps and the type of polyps present [31, 33]. *Fig 1.4* represents the broad classification of hereditary colorectal cancer predisposition syndromes and their clinical features and underlying genetic basis are explained in *Table 1.5*.

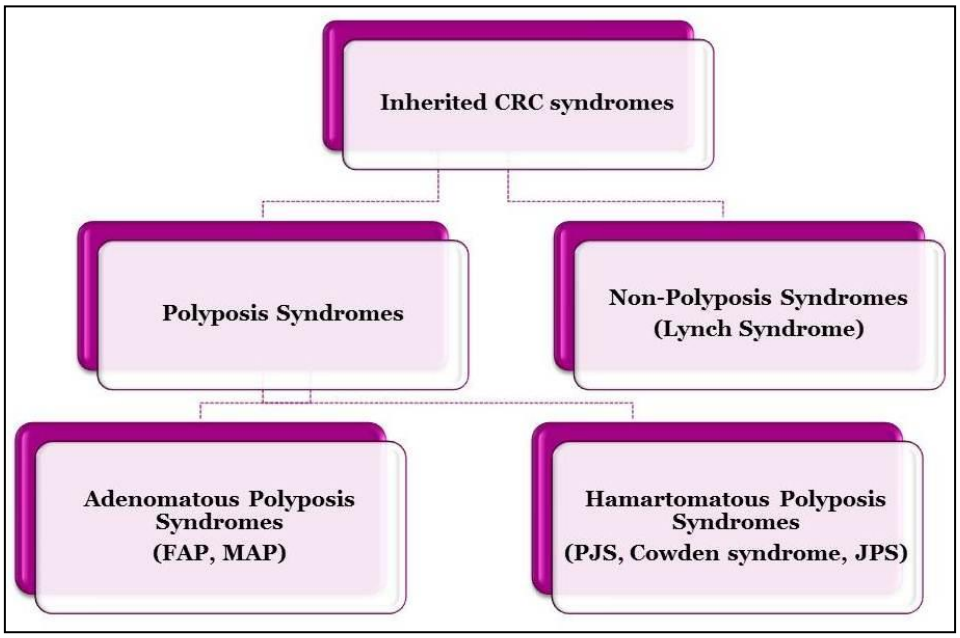


Figure 1.5 Classification of Inherited CRC syndromes

Abbreviations: FAP – Familial Adenomatous Polyposis
MAP – *MUTYH* associated polyposis
PJS – Peutz Jeghers Syndrome
JPS – Juvenile Polyposis Syndrome

Table 1.5 Features of common hereditary colorectal cancer syndromes

Type of HCRC	Syndrome	Clinical Features	Genes involved
Non-polyposis syndrome	Lynch Syndrome	High lifetime risk for CRC (without extensive polyposis) and extracolonic cancers like endometrium, stomach, ovary and rarely hepatobiliary, ureteric, small intestine and brain tumors (glioblastoma – Turcot syndrome); also manifests skin neoplasms (Muir-Torre syndrome)	*MLH1, MSH2, MSH6, PMS2*
Adenomatous Polyposis Syndromes	Familial Adenomatous Polyposis	CRC in the background of multiple polyps (>100); extracolonic cancers like upper GI cancers, papillary thyroid cancer and brain tumors (medulloblastoma – Turcot syndrome); also risk for benign extracolonic manifestations (upper GI polyps, desmoid tumor, CHRPE, osteomas, dental anomalies)	*APC*
	Attenuated FAP	CRC with less than 100 polyps; later age of onset than FAP; less or no incidence of extracolonic manifestations	*APC* (mutations in 5' part of gene)
	MUTYH Associated Polyposis	Multiple colorectal polyps (≤100) and colorectal cancer; autosomal recessive inheritance	*MUTYH*
Hamartomatous Polyposis Syndrome	Peutz-Jeghers syndrome	Multiple hamartomatous polyps throughout GI (predominantly in jejunum and stomach); mucocutaneous pigmentation mostly on lips, face, hands and genitals; increased risk for cancers of breast, colon, stomach and pancreas	*STK11*
	Juvenile Polyposis syndrome	Multiple hamartomatous (juvenile) polyps of colon and stomach; increased risk for cancers of colon, stomach, small intestine and breast	*SMAD4, BMPR1A, PTEN*
	Cowden syndrome	Hamartomas of skin & mucous membranes of mouth and nose; increased cancer risk particularly of breast, thyroid and endometrium; lower incidence of melanoma, colorectal and renal cancer	*PTEN*
	Banyan-Riley-Ruvalcaba syndrome	Hamartomatous polyps in intestine; dark freckles on penis in males; development of lipomas, angiolipomas and hemangiomas; macrocephaly, macrosomnia with intellectual disability and skeletal abnormalities	*PTEN*

Lynch syndrome which is one of the most common cancer predisposition syndromes is also the most common HCRC syndrome that accounts for nearly 3-4% of all CRC cases [34]; followed by FAP, AFAP and MAP which together accounts for nearly 1% of all CRCs [30, 31]. As these syndromes are relatively more common, the underlying mutation spectrum of the associated genes is very well characterized for Caucasian population and for some Asian countries [35-38]. According to a report, although the

incidence of CRC in India is one eighth as compared to the western population, the proportion of patients with younger onset is more in India pointing towards the larger burden of HCRC [39]. Although this inference was drawn from single-institution study, yet it cannot be ignored that frequency of hereditary CRCs (comparable to western data) amount to a significant number of families in India; for which genetic testing and counseling becomes mandatory. However, the real scenario of genetic testing for these syndromes shows that there exists a large gap in knowledge regarding the mutation spectrum and clinical features of FAP and Lynch syndrome patients in India. FAP & Lynch syndrome and the study on these syndromes in Indian patients (including this study) have been discussed in detail in *Chapter 3* and *Chapter 4* of this thesis.

1.4 PATHWAYS OF COLORECTAL TUMORIGENESIS

The development of colorectal adenocarcinoma is characterized by the transformation of normal mucosa into carcinoma through various phases of aberrant crypt foci (ACF) and increasing stages of adenoma [40]. In 1990, Fearon and Vogelstein put forth a multistep genetic model of colorectal carcinogenesis which stated that the histological progression of the normal epithelium to the adenoma and carcinoma sequence is paralleled by the accumulation of specific genetic alterations at each step [41]. The Fearon-Vogelstein (FV) model proposed the *APC* mutations as the initiating event for colorectal carcinogenesis followed by mutations in the *KRAS* and *TP53* gene that were proposed to be critical for the progression to adenoma and carcinoma stages respectively. FV model also reported a high frequency of allelic losses at various loci particularly of chromosome 5q, 17p and 18q. FV model is widely accepted and still forms the genetic paradigm of colorectal tumorigenesis.

Several studies few years after this model was proposed have shown that though this model holds true for a subset of colorectal cancer, there exist multiple alternative genetic pathways of CRC development [42, 43]. It is now universally appreciated that genomic instability is the key feature of CRC development that leads to the acquisition of multiple *genetic* and *epigenetic* alterations that then drive malignant transformation [44]. Based on the evidence from these studies, three major pathways of CRC are now described, *chromosomal instability pathway (CIN)*, *microsatellite instability pathway (MIN)* and *CpG island methylation pathway (CIMP)*.

- ***Chromosomal Instability pathway (CIN)***

CIN is the most common molecular pathway of CRC development [45]. This pathway reflects the classical adenoma carcinoma sequence (FV model) and is defined by the accumulation of numerical (aneuploidy) and structural chromosomal abnormalities [46]. This pathway is characterized by frequent loss of heterozygosity (LOH) at various tumor suppressor loci, accumulates mutation in tumor suppressor genes like *APC, TP53* and *SMAD4* and oncogenes like KRAS. CIN tumors are also defined as non-hypermutated and are thought to arise from defects in pathways involved in chromosome segregation [47]. Most of the sporadic tumors and FAP tumors are assumed to follow this pathway of colorectal carcinogenesis.

- ***Microsatellite Instability pathway (MIN)***

Microsatellite instability is caused by dysfunction of the MMR genes leading to mismatches in the DNA that are not repaired, which leads to an accumulation of mutations. The subset of CRCs with MSI is fairly stable at chromosomal level and present with near-diploid genome but they show high mutation rates (10-100 per 10^6 bases) at nucleotide level leading to the mutator phenotype in cell. These hypermutated tumors generally show a low frequency of *APC* and *KRAS* mutations, rather they accumulate mutations in *BRAF* and *TGFβ* signaling pathway genes [48]. MSI can arise either due to germline defects in mismatch repair (MMR) genes or due to *MLH1* promoter hypermethylation (epigenetic silencing) as in the case of sporadic MSI tumors [49].

- ***CpG Island Methylation pathway (CIMP)***

CRC tumors can also be classified based on methylation of CpG islands [50]. CpG-island methylation can occur at many sites. Two distinct types of methylation are seen in colorectal cancer, depending on the tissues affected. Type A (age-related) CpG-island methylation is seen mostly in normal colonic mucosa while Type C methylation occurs exclusively in colorectal cancer and this type is perhaps involved in tumorigenesis [51]. Most of sporadic CRCs have a pervasive hypermethylated phenotype and can be classified as CIMP positive while tumors having fewer methylated CpG islands which also show lower level of methylation

at individual loci, these are classified as CIMP– or CIMP-low [52]. Most CIMP+ tumors have a BRAF mutation; the one without BRAF mutation will have mutation in KRAS [53]. However BRAF and KRAS mutations are mutually exclusive. CIMP+ tumors also correlate with microsatellite instability (MSI). CIMP+ tumors with MLH1 promoter hypermethylation will always have MSI-H status [51].

Over the past few years, with the advent of new high throughput technologies, many studies are coming up leading to the classification of CRC tumors based on the massive amount of genomic data. Thus many new alternate pathways of CRC carcinogenesis are emerging and this necessitates the study of carcinogenesis in various model systems of CRC with different driver mutations and different underlying pathologies.

In the light of the discussed literature, the present study was undertaken with the aim of studying the mutational landscape and stepwise carcinogenesis in hereditary colorectal cancer. The objectives of this study were:

1. To characterize the genetic basis and pathways underlying different types of Hereditary Colorectal Cancers – MMR, APC, MYH.
2. Identify novel, founder and recurrent germline mutations, if any, in Indian HCRC and study known and novel genotype-phenotype correlation.
3. To understand the key genetic alterations in stepwise colorectal carcinogenesis using the unique model system of paired blood, normal mucosa, polyps and cancer samples from clinically characterized FAP patients.

CHAPTER 2

Materials and Methods

2.1 Patients and Samples

The hereditary colorectal cancer patients studied in this thesis were recruited at the Cancer Genetics Clinic situated at Homi Bhabha Block, Tata Memorial Centre, Parel, Mumbai between 2008 to 2016. The patients were recruited under the Hereditary Cancer Consortium Project and Founder Mutation Project, both of which were approved by the Institutional review Board (IRB) of the Tata Memorial Centre. This thesis dealt with the study of two major forms of hereditary colorectal cancers; Lynch Syndrome and FAP. The study sample size was:

- FAP cohort : 47 families
- Lynch syndrome cohort: 81 families

2.1.1 Overview of patient recruitment process

Pre-test counselling session: is a communication session between the patients and their family members who accompanied them and the genetic counselor at CGC. During this session, the counselors explained about the genetic aspects of the disease and educated the patient about what and why are the tests performed, risks of various cancers, advantages of genetic testing, the expected outcomes and its possible implications. A pedigree chart was drawn based on the detailed family history as narrated by the patients and their family members. Detailed medical records were also checked to document all the clinico-pathological phenotypes. Based on these details, syndromic diagnosis of FAP, AFAP, MAP or LS was made, following which 6 ml of blood in EDTA Vacutainers or 50 ml of mouthwash samples in Falcon tube with Normal Saline were collected from the patient after taking written informed consent.

Post-test counselling session: is a session wherein the genetic test results were communicated to the patients by giving a printed report. During this session, the patients were explained about the test results based on which appropriate surveillance and prophylactic measures were advised in case an individual has been detected with a deleterious mutation. Screening of the family specific mutation in first and second degree relatives of mutation carriers were also offered during this session.

2.2. Methodology to study mutation spectrum of MMR and APC genes in hereditary colorectal cancer patients

The overall strategy for this studying the mutation spectrum of these genes included the following steps:

1. DNA extraction from blood/mouthwash samples & Quantification of DNA
2. PCR amplification of the coding region of MMR and APC/MUTYH genes.
3. Agarose gel electrophoresis of PCR products
4. Clean-up of PCR products
5. Sanger sequencing to detect point mutations and small indels.
6. Analysis of chromatograms using Chromas Lite software
7. MLPA analysis to detect LGRs
8. Data analysis using Coffalyser.Net software from MRC-Holland.

2.2.1 DNA extraction

DNA was extracted from the 200µl of blood or 50 ml of Moutwash samples using the QiaAmp Blood DNA Minikit as per the manufacturer's protocol.

Procedure:

1. Pipet 20 µl QIAGEN Protease (or proteinase K) into the bottom of a 1.5 ml microcentrifuge tube.

2. Add 200 µl sample to the microcentrifuge tube.

3. Add 200 µl Buffer AL to the sample. Mix by pulse-vortexing for 15 s.

4. Incubate at 56°C for 10 min.

5. Briefly centrifuge the 1.5 ml microcentrifuge tube to remove drops from the inside of the lid.

6. Add 200 µl ethanol (96–100%) to the sample, and mix again by pulse-vortexing for 15 s. After mixing, briefly centrifuge the 1.5 ml microcentrifuge tube to remove drops from the inside of the lid.

7. Carefully apply the mixture from step 6 to the QIAamp Mini spin column (in a 2 ml collection tube) without wetting the rim. Close the cap, and centrifuge at 6000 x g

(8000 rpm) for 1 min. Place the QIAamp Mini spin column in a clean 2 ml collection tube (provided), and discard the tube containing the filtrate.

8. Carefully open the QIAamp Mini spin column and add 500 µl Buffer AW1 without wetting the rim. Close the cap and centrifuge at 6000 x g (8000 rpm) for 1 min. Place the QIAamp Mini spin column in a clean 2 ml collection tube (provided), and discard the collection tube containing the filtrate.

9. Carefully open the QIAamp Mini spin column and add 500 µl Buffer AW2 without wetting the rim. Close the cap and centrifuge at full speed (20,000 x g; 14,000 rpm) for 3 min.

10. Recommended: Place the QIAamp Mini spin column in a new 2 ml collection tube (not provided) and discard the old collection tube with the filtrate. Centrifuge at full speed for 1 min.

11. Place the QIAamp Mini spin column in a clean 1.5 ml microcentrifuge tube (not provided), and discard the collection tube containing the filtrate. Carefully open the QIAamp Mini spin column and add 200 µl Buffer AE or distilled water. Incubate at room temperature (15–25°C) for 1 min, and then centrifuge at 6000 x g (8000 rpm) for 1 min.

12. DNA was quantified using Nanodrop spectrophotometer or by loading the DNA samples on 1% agarose gels.

14. 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 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 composition of mastermix is given in Table 2.1 and the cycling conditions are summarized in Table 2.2. The primer sequences and annealing temperatures for various genes studied in this thesis are given in Table 2.3 – 2.10

Table 2.1 Mastermix composition for PCR

Components	Volume /reaction
10 X PCR buffer	2.5µl
2.5 mM dNTP	1.0µl
10 pmoles/µl P (F)	1.0µl
10 pmoles/µl P (R)	1.0µl
20 ng/µl DNA template	5.0µl
5U/µl Taq polymerase	0.1 µl
MilliQ	14.4 µl
Total volume	25 µl

Table 2.2 Cycling conditions for PCR

Step	Temperatures	Time
1. Initial Denaturation	95ºC	5 mins
2. Denaturation	95ºC	45 secs
3. Primer annealing*	XºC	45 secs
4. Extension	72ºC	45 secs
5. Final Extension	72ºC	5 mins
6. End	4ºC	2 mins
Go to Step 2-4: **34 cycles**		

* annealing temperature variable for various primer pairs

Table 2.3 Primer sequences and annealing temperatures for MLH1 gene

Exon	Sequence	Annealing temp (°C)
MLH1_1F	CTGTCCAATCAATAGCTGCC	62ºC
MLH1_1R	TGCGGAAAAGGAGAAGGCCTG	
MLH1_2F	GTATGAGCCTGTAAGACAAAGG	60ºC
MLH1_2R	GCCCAGCAAATAATAGGTAC	
MLH1_3F	TGGGTGACAGAGCAAGAC	55-48ºC TD
MLH1_3R	AGTTTGCTCAGATTTGC	
MLH1_4F	AAAGTGCTCATCGTTGCC	60ºC
MLH1_4R	CACCTAATAATCATCCTTGAG	
MLH1_5F	TGATTATGGAAGTAGTGGAG	55-48ºC TD
MLH1_5R	GATATCTTGGGACCTCC	
MLH1_6F	AATGCTGTCTTATCCCTGGCC	60ºC
MLH1_6R	ACCTTGACCAGAAACTATCTG	
MLH1_7+8F	TAAAAGTAGAGAGGAGTCTGTG	57.2ºC
MLH1_7+8R	CCTAGAAAGTGTTGATTACGTG	
MLH1_9F	CTGAGTAGGGTAGGTGGGTG	63-56ºC TD
MLH1_9R	CAACCAGCAATGAGCACATGTG	
MLH1_10F	CGATAGTAAGATAGTGGGCTGG	62ºC
MLH1_10R	AGGCTCTTAGTGAGGTTCTGC	
MLH1_11F	CAAATGAAGAGACTGAGGC	58.5ºC
MLH1_11R	CTAAGCCTAGGAACAACAGC	
MLH1_12F	CGGGCAGAATTGCTTCTAT	63-56ºC TD
MLH1_12R	GGTCAAAGGCAGACAGTGG	
MLH1_13F	GGGTTGTCAGATAAGCAGTC	62ºC
MLH1_13R	GCTGATGCTATTGTGGGTTA	
MLH1_14F	GTTCGTTTTCACCAGGAGG	63-56ºC TD
MLH1_14R	CTGACTCCAAAGCCTGTGCC	
MLH1_15F	CCCTGGTTGAAGACGTTG	54.3ºC
MLH1_15R	GATACCTCCATATGCAAATC	
MLH1_16F	TGACAAGAGGAGGAAAGGG	63-56ºC TD
MLH1_16R	TTTCATCATGTTGGCCAGC	
MLH1_17F	GTTCTGCCGTGCTGTTTGTC	62ºC
MLH1_17R	TGGGACTGCTGAATATTGCTGG	
MLH1_18F	CCAGCAATATTCAGCAGTCC	62ºC
MLH1_18R	CAACCTCCCATTTCTCACTGTG	
MLH1_19F	CTTGTGTTCAGGCCTGTGGGATC	62ºC
MLH1_19R	GGTCAGTGCCATCAGAGCC	

Table 2.4 Primer sequences and annealing temperatures for MSH2 gene

Exon	Sequence	Annealing temp (°C)
MSH2_1F	AACAGCTTAGTGGGTGTGGG	55-48ºC TD
MSH2_1R	ACTCTCTGAGGCGGGAAAG	
MSH2_2F	GAAGTCCAGCTAATACAGTGCTTGA	55-48ºC TD
MSH2_2R	AAACACAATTAAATTCTTCACATTTTTATTTT	
MSH2_3F	AGAGTTTGGATTTTTCCTTTTTGC	58-51ºC TD
MSH2_3R	TCATGTCAATTAAAGAGCCTTTCC	
MSH2_4F	TTCATTTTTGCTTTTCTTATTCCTTTT	53ºC
MSH2_4R	ATATGACAGAAATATCCTTCTAAAAAGTCACTA	
MSH2_5F	TCTTGGTTTGGATTGGGAAG	58-51ºC TD
MSH2_5R	GCCATTTAAAGCTAGTTATCTAATCC	
MSH2_6F	GCGTAGTAAGGTTTTCACTAATGAGC	55-48ºC TD
MSH2_6R	CATGTGGGTAACTGCAGGTTACA	
MSH2_7F	TGAGACTTACGTGCTTAGTTGATAAATTT	55-48ºC TD
MSH2_7R	GCACATTTGCCAAGTATATATTGTATGAG	
MSH2_8F	TGATGCTTGTTTATCTCAGTCAAAATT	58-51ºC TD
MSH2_8R	AATCTACAAACTTTCTTAAAGTGGCCTT	
MSH2_9F	GAAAACAGTAAAATTTAAGTGGGAGG	58-51ºC TD
MSH2_9R	CATCTTGGGGACAGGGAAC	
MSH2_10F	ATTGAAAAATGGTAGTAGGTATTTATGGAA	53ºC
MSH2_10R	CACATCATGTTAGAGCATTTAGGGA	
MSH2_11F	ATATGTTTCACGTAGTACACATTGCTTCTA	58-51ºC TD
MSH2_11R	TCAAATATCATGATTTTTCTTCTGTTACC	
MSH2_12F	GGGTTTTGAATTCCCAAATG	58-51ºC TD
MSH2_12R	AAGAACTGGGAATTTTCTCCATC	
MSH2_13F	CTTCTCTTGATGAAAGGCCC	58-51ºC TD
MSH2_13R	TCACAGGACAGAGACATACATTTCTATCT	
MSH2_14F	GTGGCATATCCTTCCCAATG	53ºC
MSH2_14R	GCAATTACTGATGATTTCAAGGG	
MSH2_15F	TTTTCTAATGACAAGGTGAGAAGG	58-51ºC TD
MSH2_15R	AAATAACACAGAGATAGATTCTTTGCC	
MSH2_16F	TGTGATATGTTTAGATGGAAATGAAAC	58-51ºC TD
MSH2_16R	GGCACTGACAGTTAACACTATGG	

TD: Touchdown PCR

Table 2.5 Primer sequences and annealing temperatures for MSH6 gene

Exon	Sequence	Annealing temp (°C)
MSH6_1F	TGTTGATTGGCCACTGGG	58-51°C TD
MSH6_1R	CAACCCCCTGTGCGAGCCTC	
MSH6_2F	TGCCAGAAGACTTGGAATTC	58-51°C TD
MSH6_2R	TTACTGGGGTAAAATACACTTAATTTC	
MSH6_3F	CTGGTCTTGAACTGCTGGGAT	58-51°C TD
MSH6_3R	CCCCTTTCTTCCCCCATC	
MSH6_4.1F	TGCACGGGTACCATTATAAAGTCA	58-51°C TD
MSH6_4.1R	GTATTCTTGGTTTCTGATGAAATGCTAG	
MSH6_4.2F	GAAGGAAACGCCCTCAGC	58-51°C TD
MSH6_4.2R	CAGTTGCCTTTCATGAATACCAG	
MSH6_4.3F	CCACATGGATGCTCTTATTGGA	58-51°C TD
MSH6_4.3R	TCATCTGAAAACTGACCTATGAAAAACT	
MSH6_4.4F	TTTGTTGATACTTCAGTGGGAAAGTT	55-48°C TD
MSH6_4.4R	CTCCTGATCAATAAGGCATTTTTTG	
MSH6_4.5F	CTCTAGGTGGTTGTGTCTTCTACCTC	58-51°C TD
MSH6_4.5R	TGAGTAGCCTCTCAACATCTGGAA	
MSH6_4.6F	CGAAGTTGTAGAGCTTCTAAAGAAGCT	58-51°C TD
MSH6_4.6R	GTCCTACAGCCAATTCTGTTGC	
MSH6_4.7F	AGCCTCCTGGAATACCTAGAGAAAC	58-51°C TD
MSH6_4.7R	ACTTATTTTTAGGGATAATATACAGCTGGC	
MSH6_5F	CACTTAGGCTGATAAAACCCCC	58-51°C TD
MSH6_5R	GTATGTTATTCCTAATGTCACAAATGACTTT	
MSH6_6F	AAGACAAAAGTTTATGAAACTGTTACTACCA	55-48°C TD
MSH6_6R	AGAAGCAAATATCTTTTATCACATCTAAATG	
MSH6_7F	TAACCTAGAAGATGAATTTATGTAATATGATTT	55-48°C TD
MSH6_7R	TTCAGATAATCTTCTATAAAAATAGTTATTTGT	
MSH6_8F	TGAGTTACTTCCTTATGCATATTTTACT	55-48°C TD
MSH6_8R	AATATTAGCGATACATGTGCTAGCA	
MSH6_9F	TGCTAGCACATGTATCGCTAATATT	58-51°C TD
MSH6_9R	GCATCATCCCTTCCCCTTTTA	
MSH6_10F	GAAGGGATGATGCACTATGAAAAA	55-48°C TD
MSH6_10R	GTAGAAGGTAGATAAGAATTAAAAGGGTTTAA	

TD: Touchdown PCR

Table 2.6 Primer sequences and annealing temperatures for APC gene

Exon	Primer sequences	Annealing temp (°C)
APC 1F	TTTCTTTAAAAACAAGCAGCCA	59°C
APC 1R	CACAGAAAACCTTGCCTCAG	
APC 2F	AAGGTGCGTGCTTTGAGAGT	59°C
APC 2R	ACCAACACCCAAATCGAGAG	
APC 3F	CCAAGTGGACTTTTCAGGGA	59°C
APC 3R	CTGGAGTACACAAGGCAATGTT	
APC 4F	GCTCTTCTGCAGTCTTTATTAGCA	59°C
APC 4R	CCTAGTTGAACCCTGAGGTCC	
APC 5F	AAGCCACTTGTGACTTTGGC	62°C
APC 5R	GTTGCTCAGCAGCCATGATA	
APC 6F	TGCGGTGAGCTGAGATTATG	62°C
APC 6R	ACCCACAAACAAGAAAGGCA	
APC 7F	GCAGCTCTAATGCTCAAGGG	59°C
APC 7R	TGGTACTGAATGCTTCTGGAAA	
APC 8F	CCATTCTGCAGTTTAATGCTCA	59°C
APC 8R	TAGAGATGGGGTTTTGCCAC	
APC 9F	CTGGAAAGGTTTTCCGGTTT	59°C
APC 9R	TGCTTTGAAACATGCACTACG	
APC 10F	GTCAAGGGCAGATGAGTGGT	65.6°C
APC 10R	TTCTATGCTGGAAACCAGGG	
APC 11F	TTGTCTTTTTCCTCTTGCCC	59°C
APC 11R	AGCGAATGTGAAGCACAGGT	
APC 12F	CCTGTTGCTTATCATTTCTCACC	64.2°C
APC 12R	AGAGTGAGACCCTGCCTCAA	
APC 13F	CAGCCTCCCAAAGTGATAGG	50.2°C
APC 13R	ATGGCTAAAAGAAGGCAGCA	
APC 14F	AGGGACGGGCAATAGGATAG	59°C
APC 14R	CATTGCTTACAATTAGGTCTTTTTGA	
APC 15AF	AGAGTGGCACCCAACCATAG	59°C
APC 15AR	TCCCATAATGCTTCCTGGTC	
APC 15BF	GTTACTGCATACACATTGTGAC	55°C
APC 15BR	GCTTTTTGTTTCCTAACATGAAG	
APC 15CF	GCTCAAGCTTGCCATCTCTT	62°C
APC 15CR	TATGGGCAGCAGAGCTTCTT	
APC 15DF	CCAGGAACTTCTTCAAAGCG	62°C
APC 15DR	GTGAAGGACTTTGCCTTCCA	
APC 15EF	GTCAATACCCAGCCGACCTA	59°C

APC 15ER	AGGCTGATCCACATGACGTT	
APC 15FF	AACGTCATGTGGATCAGCCT	62°C
APC 15FR	TGCTGGATTTGGTTCTAGGG	
APC 15GF	CAGACGACACAGGAAGCAGA	62°C
APC 15GR	GCAGCTTGCTTAGGTCCACT	
APC 15HF	GTGAACCATGCAGTGGAATG	59°C
APC 15HR	TGTTGGCATGGCAGAAATAA	
APC 15IF	TTTGCCACGGAAAGTACTCC	59°C
APC 15IR	TATCATCCCCCGGTGTAAAA	
APC 15JF	CCCAGACTGCTTCAAAATTACC	59°C
APC 15JR	GAGCCTCATCTGTACTTCTGC	
APC 15KF	CCCTCCAAATGAGTTAGCTGC	59°C
APC 15KR	TTGTGGTATAGGTTTTACTGGTG	
APC 15LF	ACCCAACAAAAATCAGTTAGATG	59°C
APC 15LR	GTGGCTGGTAACTTTAGCCTC	
APC 15MF	ATGATGTTGACCTTTCCAGGG	59°C
APC 15MR	ATTGTGTAACTTTTCATCAGTTGC	
APC 15NF	AAAGACATACCAGACAGAGGG	59°C
APC 15NR	CTTTTTTGGCATTGCGGAGCT	
APC 15OF	AAGATGACCTGTTGCAGGAATG	59°C
APC 15OR	GAATCAGACGAAGCTTGTCTAGAT	
APC 15PF	CCATAGTAAGTAGTTTACATCAAG	55°C
APC 15PR	AAACAGGACTTGTACTGTAGGA	
APC 15QF	CAGCCCCTTCAAGCAAACATG	59°C
APC 15QR	GAGGACTTATTCCATTTCTACC	
APC 15RF	CAGTCTCCTGGCCGAAACTC	62°C
APC 15RR	GTTGACTGGCGTACTAATACAG	
APC 15SF	TGGTAATGGAGCCAATAAAAAGG	59°C
APC 15SR	TGGGAGTTTTCGCCATCCAC	
APC 15TF	TGTCTCTATCCACACATTCGTC	59°C
APC 15TR	ATGTTTTTCATCCTCACTTTTTGC	
APC 15UF	GGAGAAGAACTGGAAGTTCATC	59°C
APC 15UR	TTGAATCTTTAATGTTTGGATTTGC	
APC 15VF	TCTCCCACAGGTAATACTCCC	59°C
APC 15VR	GCTAGAACTGAATGGGGTACG	
APC 15WF	CAGGACAAAATAATCCTGTCCC	53°C
APC 15WR	ATTTTCTTAGTTCTATTCTTCCTC	

Table 2.7 Primer sequences and annealing temperatures for MUTYH gene

Exons	Primer sequences	Annealing temp (ºC)
MYH_2F	AAAATTTGGCCTCATTGTGAC	60-53ºC
MYH_2R	TATCACAATCCCTTCCCAGC	
MYH_3+4+5F	AAGGGGGTTAGTTGGGGGAAGC	60-53ºC
MYH_3+4+5R	CAAGGGTGAAGGTGGTAGAGGAAGC	
MYH_6+7+8F	TTTGGGGTGGGTGTAGAGAAGG	60-53ºC
MYH_6+7+8R	GCACAGAGGGGCCAAAGAGTTAG	
MYH_9+10+11F	CAGCCCACCCCCACTTTGT	60-53ºC
MYH_9+10+11R	GCTTTGGCCGGGTTCTGC	
MYH_12+13F	TCTAGGTTGGCCCCTAAAGC	60-53ºC
MYH_12+13R	GTCAAGGGGTTCAAATAGGC	
MYH_14F	TTGGCTTTTGAGGCTATATCC	60-53ºC
MYH_14R	ACATGTAGGAAACACAAGGAAGTA	
MYH_15F	TGGGACATGAAGTTAAGGGC	60-53ºC
MYH_15R	GAGTGGAGAATGTTCACCCAG	
MYH_16F	GAGAGGATTCTCTGCTCCCC	60-53ºC
MYH_16R	TCGAAACCAGTCTGAGCAAC	

TD: Touchdown PCR

Table 2.8 Primer sequences and annealing temperatures for POLE gene

Exon	Primer sequences	Annealing temp (ºC)
POLE_Exon 9F	AACCAGAGGGAGGTAGAGCA	63.2 ºC
POLE_Exon 9R	CTCCCTGTTGGTGATGAGGT	
POLE_Exon 10+11F	GCACTTTCACATTGCTGTGG	63.2 ºC
POLE_Exon 10+11R	CCTAAGTCGACATGGGAAGC	
POLE_Exon 12F	ACACGTCCAGGAGACCAAAC	63.2 ºC
POLE_Exon 12R	TTGCAGCTGCCATACTCTTG	
POLE_Exon 13F	GGTGCCTGTTAGGAACTTGC	61.4 ºC
POLE_Exon13R	GAGCGGGCTGGCATACAT	
POLE_Exon 14F	GGCTTTGCTTTCTGTGCTTC	63.2 ºC
POLE_Exon 14R	AGCACTCCTGGGACATCCAC	

Table 2.9 Primer sequences and annealing temperatures for POLD1 gene

Exon	Primer sequences	Annealing temp (ºC)
POLD1_Exon 6+7+8F	CTCCCGAGAGAGTGAGTGCT	63.2 ºC
POLD1_Exon 6+7+8R	ATCTCCACACCCTCTGTGCT	
POLD1_Exon 9F	AGGTGAGAGCAGAGCAGGAG	63.2 ºC
POLD1_Exon 9R	AGGAGCTGATGGCTCAGGAC	
POLD1_Exon 10F	GGTTCTGCAGGATTTTCAGG	63.2 ºC
POLD1_Exon 10R	GTGGAGAGGGAGTGGGAAG	
POLD1_Exon 11+12F	TGTCCCCAAATCTCTTCCTG	61.4 ºC
POLD1_Exon 11+12R	CCAGGAAAGCAGAGACAAGG	
POLD1_Exon 13F	TCCCTGACCCCATCCGTG	63.2 ºC
POLD1_Exon 13R	GGACAAGTCTCGGCTACTGA	

Table 2.10 Primer sequences and annealing temperatures for NTHL1 gene

Exons	Primer sequences	Annealing temp (ºC)
NTHL1_1F	GGCCGCATGGGCCGCCGGG	63ºC
NTHL1_1R	AGCCTGGAGTGGAGAGTCC	
NTHL1_2F	GTCGCTGGCATAAGGAGG	64.2ºC
NTHL1_2R	CGAGCACGAGGCCCTAAACC	
NTHL1_3F	AAGTGCTGGGATGACGGGTG	64.2ºC
NTHL1_3R	GCCTGAGATGCTTGACCCT	
NTHL1_4F	GAGCCTACACGTGCATCATTG	64.2ºC
NTHL1_4R	GGTGCTCAGCCCATGTGAC	
NTHL1_5+6F	GATGGTCACATGTAGGCTTGC	57ºC
NTHL1_5+6R	CCATCTGCAAACACACCAAAG	

2.2.3 Agarose Gel Electrophoresis

1. The required amount of agarose was weighed to make a 1 % solution and was dissolved in 1X TBE (Tris, Boric acid, EDTA buffer).
2. The agarose solution was boiled in a microwave to enable complete digestion of the powder. It was cooled to less than 50ºC.
3. Ethidium bromide (from 10mg/ml stock) was added to a final concentration of 0.5 µg/µl.
4. The solution was poured on the tray and allowed to solidify.
5. The samples to be loaded were mixed with 1X loading dye.
6. The samples were loaded and the apparatus was run at the required voltage ($5V/cm^2$).
7. It was allowed to run until the dye traversed $3/4^{th}$ of the gel.
8. The DNA bands were visualized using a UV Transilluminator.
9. Pictures were taken and stored using Gel Documenter.

2.2.4 Purification of PCR products

The PCR products were diluted according to the intensity of the band such that 30-50ng of template is available for sequencing. Two microliter of EXO-SAP IT [USB] was added to 5µl of diluted PCR product. This mix was incubated at 37ºC for 15 mins followed by incubation at 80ºC for 15 mins.

2.2.5 Sanger sequencing of PCR products

Step 1: Cycle sequencing

- Add 2µl of template and 1µl of primer (both in the desired concentration) in the 96-well sequencing plate.

- Prepare the cycle sequencing reaction master mix as:
  - MilliQ water - 4.75 µl
  - 5X Sequencing Buffer - 1.75 µl
  - Ready Reaction mix - 0.50 µl

- Add 7µl of mastermix to each sample

- Setup the cycle sequencing reaction in the thermal cycler as:
    1. 96ºC for 2 minutes
    2. 96ºC for 10 seconds
    3. 1ºC/second to 55ºC
    4. 55ºC for 0.05 minutes (5 seconds)
    5. 1ºC/second to 60ºC
    6. 60ºC for 4 minutes
    7. Go to steps 2-5 for 24 times
    8. 4ºCforever

Step 2: Post cycle sequencing cleanup

1. A mixture of the two BigDye XTerminator reagents (Premix) is prepared as:
    - XTerminator Solution - 10µl
    - SAM solution - 45µl
2. Add 55µl of the premix to each sample and tap spin the reaction plate
3. Vortex the reaction plate for 30 minutes
4. Centrifuge the reaction plate at 2500 rpm for 2 minutes
5. Tap spin and load the reaction plate on the sequencer.

2.2.6 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. Locus specific database for both APC and MMR genes – InSiGHT database (https://www.insight-group.org/) was used to check if the mutations were reported or not. Whenever, a deleterious germline mutation was identified it was 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.

2.2.7 MLPA analysis to detect LGRs

In case a point mutation or small indel was not identified by Sanger sequencing, MLPA analysis was carried out to study the large genomic rearrangements in the specific genes as per manufacturer's protocol described below. SALSA MLPA Kits from MRC-Holland are used for this analysis. For APC, SALSA MLPA P043 APC Probemix and for MLH1/MSH2, SALSA MLPA P003 MLH1/MSH2 Probemix kit was used.

Procedure

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 H_2O + 3µl Ligase buffer A + 3µl Ligase buffer B. Then add 1µl Ligase-65 enzyme. Mix well by pipetting gently up and down. (Never 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 5minutes 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 dH_2O + 2 µl SALSA PCR primer mix + 0.5 µl SALSA Polymerase. Mix well by pipetting up and down; do not vortex. Store on ice until use.
2. **At room temperature**, add 10 µl polymerase mix 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.

2.2.8 MLPA Data Analysis

Data analyzed using the Coffalyser.Net Software created 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
Equivocal copy number	All other values

2.3 Haplotyping analysis to study founder mutations

In order to establish the frequency of one of the novel mutations in MLH1 gene; c.156delA identified in two unrelated families from the Shia Momin community, a cancer awareness and blood sample collection camp was organized at Jogeshwari, Mumbai with the help of community members, after taking permission from the elder members of this community. Blood samples were collected from 400 participants of Shia Momin community after taking written informed consent from each of them. Personal and family history of cancers or other illness and other details like age, gender and diet preferences were also documented. Part of the blood samples were spotted on the Whatman FTA cards for rapid extraction of DNA and lymphocytes were separated from rest of the sample and stored at -20°C for further analysis.

2.3.1 Separation of lymphocytes from blood

Procedure

1. The blood was transferred to 15 ml Falcon tubes.
2. 3 volumes of RBC lysis buffer (containing ammonium chloride, ammonium bicarbonate, EDTA) was added and mixed by rotating for 10 minutes.
3. The tubes were centrifuged at 1500 rpm for 15 minutes.
4. The supernatant was discarded
5. To the pellet three volumes of RBC lysis buffer was added and steps 3 to 5 were repeated until a clear supernatant and a clean white pellet was obtained.
6. After final centrifugation, the supernatant was discarded completely and the pellet was resuspended in 1ml of PBS (NaCl, KCl, Na_2HPO_4 and KH_2PO_4).
7. The solution was transferred to two 1.5ml fresh Eppendorf tubes, distributing half the volume in each.
8. The tubes were then centrifuged at 12,000rpm for 10 minutes.
9. The supernatant was discarded and the pellet was resuspended in 500µl of fresh PBS.
10. The tubes were again centrifuged at 12,000rpm for 10 minutes and the supernatant was completely discarded.
11. The pellets were stored at -80ºC for banking and other was taken for DNA extraction.

2.3.2 DNA extraction from FTA cards

FTA cards utilize Whatman FTA technology that simplifies the handling and processing of nucleic acids. FTA Cards contain chemicals that lyse cells, denature proteins and protect nucleic acids from nucleases, oxidation and UV damage.

Procedure

1. Apply sample to FTA elute matrix; dry thoroughly.
2. Punch 3mm discs (3-5) and place in 0.5ml centrifuge tube.
3. Rinse punches in 500µl MilliQ, pulse vortex 3 times – 5secs each time.
4. Remove water and tap spin for 5secs; pipette off excess water.
5. Add 30µl of MilliQ; heat at 95ºC for 30 mins.
6. Pulse vortex 60 times and tap spin.
7. Use 5-10µl eluted DNA in a PCR reaction.

2.3.3 PCR amplification of MLH1 Exon 2

PCR amplification of MLH1 Exon 2 was carried out in all the 400 samples of Shia Momin individuals using specific primers described in Table 2.3. A total of 7µl of DNA extracted from FTA cards were used for amplification in a 25µl reaction using 10pmol of forward and reverse primers and 2.5mM of dNTPs.

2.3.4 CSGE analysis

CSGE is used as a prescreening technique for specific and sensitive detection of mutations. CSGE was used to screen for c.156delA mutations in Exon 2 of MLH1 gene in the Shia Momin samples.

Protocol:

1. **Sample preparation**
   - Dispense the PCR products into 0.2ml tubes and seal the tubes with parafilm.
   - Denature the PCR products in boiling water for 10-12 minutes.
   - Incubate the PCR products at 67ºC for partial reannealing of the denatured products.

2. Casting of Gel

- Wash the glass plates thoroughly with detergent and tap water
- Wipe with 70% ethanol and assemble the apparatus using side clamps and insert the comb in to the assembly.
- Pour the gel mixture very slowly through the sides (avoid bubbles).
- Allow the gel to polymerize for 1 – 2 hours.
- Wash the wells to remove any unpolymerized acrylamide.
- Transfer the entire assembly in to the gel running chamber and fill the upper and lower tanks with running buffer (0.5X TTE).
- Load 5-10µl of samples with 6X loading dye into the wells.
- Run the gel at 200V for 4 hours.
- Disassemble the gel apparatus carefully (the gel sticks to one of the glass plate).
- Transfer the gel to a staining tray gently and stain with ethidium bromide for 10 mins.
- Destain with MilliQ water for 20-30 mins
- Visualize the gel and capture the gel image using Gel Documentation system.
- Interpret results based on the banding pattern of samples in comparison with the positive control band pattern.

The samples showing aberrant band pattern on CSGE was sequenced in order to confirm the mutation.

2.3.5 Fragment analysis of microsatellite markers

Panel of 13 microsatellite markers were used to study the founder effect of 5 recurrent mutations identified in MLH1 gene. The strategy was to PCR amplify the microsatellite markers using a pair of primers in which Forward primer is labeled 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 labeled size standard is added along with the sample to allow molecular size comparison of fragments. The data can be analyzed using GeneMapper software from Thermo Fisher to determine the allele sizes of the PCR product.

Table 2.12 Primers and Annealing temperatures of haplotyping markers

Sr. No	Marker	Primer sequences	Label	Size range (bp)	Ta (°C)
1	D3S3564_F	AGCTAAACACAGTCTAACTGCAT	FAM	190-220	55°C
	D3S3564_R	CCCACAGAGTGATAGGGA	-		
2	D3S1298_F	AGCTCTCAGTGCCACCCC	PET	200-220	59°C
	D3S1298_R	GAAAAATCCCCTGTGAAGCG	-		
3	D3S3623_F	CCCCCATGTTGGTTAAAGGC	VIC	210-220	62°C
	D3S3623_R	TCTCTGAACTGAAGTGACCTCCTG	-		
4	D3S1611_F	CCCCAAGGCTGCACTT	NED	250-270	59°C
	D3S1611_R	AGCTGAGACTACAGGCATTTG	-		
5	D3S1007_F	GAAGGGTCACTTGAGTCTAGGAG	FAM	60-80	62°C
	D3S1007_R	ATTTGCCACCATGCCTGGCTAG	-		
6	D3S1561_F	TAAGTCCCAGAGGCAAAGG	FAM	220-240	56°C
	D3S1561_R	CGCTAAACTATCCACAGGACAC	-		
7	D3S2411_F	GTTGGGTTTCTTTCCTGGTT	FAM	220-240	59°C
	D3S2411_R	TGAGTGCCTACTATGTGCCA	-		
8	D3S3512_F	AGATCCCACACCTGCCTCC	PET	130-150	59°C
	D3S3512_R	ACCAAGTTTTTAAGGCCAATGC	-		
9	D3S3518_F	CCTTTCGGGATATGAATG	FAM	150-170	53.6°C
	D3S3518_R	TGTGCAGGGTCCTAGA	-		
10	D3S3718_F	CATTTTTGGCAGATTCTTT	NED	150-170	53-46°C TD
	D3S3718_R	ACTTCCAGAGTGCTTAGACA	-		
11	D3S2432_F	GGCAGGCAGGTAGATAGACA	FAM	120-160	56°C
	D3S2432_R	ACACTAAACAAGCATAGTCAGGC	-		
12	D3S3936_F	TTTCACCTGAGGTTTCCCTG	FAM	110-130	56°C
	D3S3936_R	ATTCCAGCCCTGCAGATATG	-		
13	D3S4153_F	TGGACTTAGGTGCCTACTGGA	FAM	260-280	56°C
	D3S4153_R	AGACCCACACACCCACTACAT	-		

TD: Touchdown PCR

CHAPTER 3

Study of Mutation spectrum of APC and MUTYH genes in Indian Familial Adenomatous Polyposis Patients

3.1 Familial Adenomatous Polyposis – An Introduction

Familial Adenomatous Polyposis (FAP) is the second most common form of hereditary colorectal cancer predisposition syndrome, the first being Lynch Syndrome [31]. It is an autosomal dominant syndrome characterized by the presence of hundreds and thousands of adenomatous polyps in the colorectum with an inevitable progression to carcinoma in absence of any surgical intervention as it has a nearly complete penetrance.

3.1.1 History of FAP

The first histologically verified case of adenomatous polyposis was published in 1881 by Sklifasowski in Russia. In 1882, the inherited predisposition for what Cripps termed "disseminated polypus of the rectum" garnered much attention. He gave the name on the basis of 20-30 adenomas discovered in two affected siblings. The extreme rarity of multiple polypi, unlike the singular polyp was emphasized. In 1890, Handford confirmed the association of this syndrome with cancer, by histologically documenting the progression of adenoma to adenocarcinoma, setting the stage for more accurate histological investigation [54].

In 1925, Lockhart-Mummery stated that adenomas should be distinguished from inflammatory polyps, and that the hereditary factor in FAP is not cancer, but multiple adenomas having a marked tendency to undergo malignant change. On the basis of his early polyposis series, the polyposis registry was established at St. Mark's Hospital as the first in the world [54]. In 1927, Cockayne stated that FAP is inherited as a dominant condition [54]. In 1951, Gardner described what was later termed ''Gardner syndrome'' characterized by the presence of colorectal adenomas, desmoid tumors, bone tumors, and soft cyst-like surface tumors [55]. The possible association of FAP to a gene on chromosomal band 5q21-22 was suggested in 1987 following linkage analysis of families with FAP [56, 57]. This gene was then cloned, identified and characterized in 1991 [58, 59].

3.1.2 CLINICAL MANIFESTATIONS OF FAP

The cardinal feature of FAP is the development of multiple adenomatous polyps throughout the colon and rectum, usually beginning in the late childhood or adolescence. As FAP has a near complete penetrance, one of these adenomas will eventually progress into an adenocarcinoma by the age of 50y. Depending upon the number of polyps present, age of onset of polyposis and age of colorectal cancer development, FAP phenotypes are classified as [60]:

- ***Profuse polyposis:*** presence of more than 5000 polyps by the first or second decade of life. The average age of onset of CRC is 34 years.
- ***Classical polyposis:*** development of 100-1000 polyps in the second/third decade with the mean age of CRC being 40years in untreated individuals.
- ***Attenuated polyposis:*** presence of fewer polyps, generally 10-100 and late age of onset of both polyps and colorectal cancer.

Although colorectal cancer stands out as the most prevalent complication, FAP is actually a multisystem disorder. Patients with FAP have an increased predisposition to certain extracolonic cancers and benign manifestations that are observed at a relatively higher frequency in FAP patients as compared to the general population.

❖ **Extracolonic cancers in FAP**

- ***Thyroid cancer:*** Thyroid cancer in FAP patient was first reported by Crail in 1949 [61]. FAP patients develop papillary thyroid cancers (PTC) at an average age of 27years. Young women are at an increased risk of developing PTC at a rate of 160 times that of the general population. There is a striking female preponderance with the female to male ratio being 17:1 [62].

- ***Hepatoblastoma:*** The first association between FAP and hepatoblastoma was shown by Kingston et al., in 1983 [63] and more than 50 cases of hepatoblastoma has been reported so far in literature. Risk of hepatoblastoma is several folds higher in children from FAP families (predominantly in boys of 1-4 years) as compared to the general population [62].

- ***Brain tumors:*** The association between brain tumors and FAP is described as "Turcot syndrome" as it was first reported by Turcot et al., in 1959 [64].

Medulloblastoma is the most common form of brain tumor seen in FAP patients, however, high grade astrocytomas and ependymomas have also been described [62]. Like sporadic medulloblastoma, FAP associated medulloblastoma are also diagnosed most commonly in the first decade of life [62].

- ***Upper GI cancers:*** Upper GI tumors are commonly seen in FAP patients with periampullary cancers being one of the most common reasons for death in patients who underwent prophylactic colectomy [65]. The risk of duodenal and periampullary cancers are several hundred fold increased in FAP patients, though these are rare cancers in general population [66]. Gastric cancers are less commonly seen in Caucasian FAP patients; however, there exist a 3-4 fold excess risk of gastric carcinomas in Japanese and Korean FAP patients [67].

- ***Other rare cancers:*** Apart from the extracolonic cancers discussed above, FAP patients may also develop adrenal tumors [68] or pancreatic cancers [69], though at a lower frequency than other cancers.

❖ **Extracolonic manifestations in FAP**

More than 70% of the FAP patients develop extracolonic manifestations. These manifestations are generally benign with little significance but some of them may have serious morbidity and sometimes even mortality. Most common extracolonic manifestations include:

- ***Upper GI polyps:*** Along with colorectal polyps, upper GI polyps also manifest as a characteristic feature of FAP. Fundic gland polyps are very common and seen in nearly 90% of FAP patients. Even more common are the polyps of the duodenum and periampullary region with a lifetime risk reported to reach 100% in FAP [70]. Small bowel polyps also develop, though at a relatively lower frequency than duodenal and fundic gland polyps. The prevalence of most common small bowel polyps (jejunal and ileal polyps) has been estimated to be between 30-75% [71].

- ***Congenital hypertrophy of the retinal pigment epithelium (CHRPE):*** It is patch/patches of discoloration in the ocular fundus that has no malignant potential and does not affect sight. It is the most common extracolonic manifestation that is

identified in nearly 60% of all FAP patients [72]. While CHRPE can also occur in absence of FAP, multiple or bilateral patches are highly specific for FAP and can be used as very sensitive phenotypic marker for FAP [72, 73]. CHRPE can be detected easily by direct, non-invasive ophthalmologic examination at any age.

- ***Desmoid tumors:*** Desmoid tumors (benign fibromatosis) are soft tissue tumors which are an important cause of FAP related mortality and morbidity. They usually arise in the abdominal wall or bowel mesentery and may continue to grow with extensive intra-abdominal involvement. Extensive fibromatosis may compress the urinary or GI tract, nerves and vasculature system which sometimes results in the death of the individual. It can develop at any age from infancy to 81 years with the median age of diagnosis being 30 years [74]. The lifetime risk for men and women of FAP families are 8% and 13% respectively [75]. Most of the desmoid tumors are triggered after the colorectal surgery and they also have a high recurrence rate. Desmoid tumor can also occur as a hereditary desmoid disease characterized by multiple desmoid tumors in the absence of colonic polyposis [76].

- ***Osteomas and dental abnormalities:*** Osteomas occur in about 20% of the FAP patients as compared to the lower prevalence of 1-2% in the general population, while the dental anomalies have 17% prevalence [77]. Osteomas are benign bone growths most commonly found in skull and mandible. Osteomas can be palpated or can be identified as occult radio-opaque jaw lesions. Dental abnormalities include the presence of supernumerary teeth, dentigerous cysts both of which can interfere with eruption of normal teeth. Also, FAP patients present with a phenomenon called as secondary retention of teeth which erupted teeth are retained at a certain position in the jaw, which is usually caused by ankylosis. This results in a submerged position of the involved tooth. Like other extracolonic manifestations, osteomas and abnormal dentition can also present as the first symptom of FAP before the development of colonic polyps [78].

- ***Other rare manifestations:*** Subcutaneous lesions like epidermoid cysts, lipomas and fibromas represent the other benign manifestations that are identified in FAP patients [79]. They usually do not lead to any complications except for some cosmetic concern. Sometimes the cyst becomes infected making it necessary to

surgically remove it. In few FAP cases, nasopharyngeal angiofibromas have been reported [80].

3.1.3 VARIANTS OF FAP SYNDROME

Depending upon the age of onset of polyposis and CRC, burden of polyps and the types of extracolonic manifestations present, there are three variants of the FAP, described in *Table 3.1*.

Table 3.1 Features of FAP syndrome variants.

Syndrome	Colonic manifestations	Extracolonic manifestations
Attenuated FAP	Polyps ranging from 2-100 in numbers; develop by 4th or 5th decade of life; usually progressing to cancer.	Rare occurrence of extracolonic manifestations in AFAP [81].
Gardener syndrome	CRC in the background of classical polyposis	Presence of epidermoid cysts, osteomas and dental abnormalities.
Turcot syndrome	Classical polyposis and colorectal cancer by 2nd or 3rd decade of life	Increased occurrence of brain tumors (medulloblastoma).

3.2 GENETIC BASIS OF FAP

FAP is caused due to germline mutations in the tumor suppressor gene, *APC* [82]. Deleterious mutations in APC gene are identified in up to 90% of the cases with a classical FAP phenotype.

3.2.1 The APC gene

APC gene located on chromosome 5q21-22 is a large multi-exonic gene with an open reading frame of 8535bp. There are 15 exons in the most abundant transcript that translates into a protein of 2843 amino acids. This transcript lacks the smallest exon, Exon 10A (alternatively spliced) which when transcribed leads to a protein of 2861 amino acids. Exon 15 is the largest exon that encodes for three quarters of the translated protein and is the most common target for both germline and somatic mutations [83].

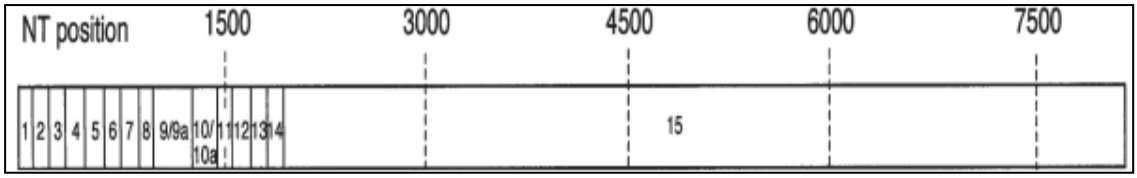


Fig 3.1 Organization of the APC gene

3.2.2 The APC protein – structural organization

APC gene encodes a roughly 310KDa multifunctional protein which is expressed in a variety of fetal and adult tissues, including mammary and colorectal epithelium. The cellular localization of APC is predominantly cytoplasmic [84] though APC proteins can also be localized to nucleus. The APC protein is organized into several domains each of which has a specific function. Full length protein has an oligomerization domain and an armadillo region in the N-terminus, the middle region has several 15 amino acid repeats and 20 amino acid repeats and the C-terminus of protein contains the basic domain, EB1 binding and HDLG binding domain.

- ***Oligomerization Domain:*** contains a heptad repeat (amino acids 6-57) that is necessary for the formation of APC homodimers. Wild type APC can dimerize with either the wild type APC or even with the mutant copy of APC. It is because of this ability, APC is said to exert a dominant negative effect [85].

- ***Armadillo region*****:** named as this region shows homology to the central repeat region of the Drosophila segment polarity protein armadillo. This is a highly conserved region and contains seven arm repeats between amino acids 543 to 767 in APC that is known to bind the regulatory unit of protein phosphatase PP2A, an enzyme that also binds Axin. This interaction suggests that PP2A may act as an antagonist to GSK3β mediated phosphorylation of β-catenin which is necessary for the degradation of β-catenin [86]. This domain also binds to APC-stimulated guanine nucleotide exchange factor (Asef), thereby enhancing the interaction of Asef with Rac (a small GTPase) that controls cell adhesion and motility via modulation of the actin cytoskeleton [87]. Though the armadillo region is important, its role in the tumor suppressor function of APC remains unlikely.

- ***The 15 amino acids repeats*****:** There are three 15 amino acid repeats (imperfect repeats) in the APC protein between the amino acids 1020 and 1169. These repeats are involved in the binding of APC to β-catenin, however unlike the 20

amino acid repeats; this binding does not mark β-catenin for degradation [88].

- ***The 20 amino acids repeats*:** In the region between amino acids 1262-2033, seven 20 amino acids are present which are also involved in the binding to β-catenin. This motif contains the signature residues TPXXFSXXXSL of which SXXXS consensus sequence acts a substrate for GSK3β phosphorylation which is a pre-requisite for β-catenin binding. Binding of β-catenin to 20 amino acids repeats marks it for degradation by proteasomes. Atleast three of these seven repeats must be present for mediating the binding and degradation of β-catenin. Moreover, the third 20 amino acid repeat has the highest affinity for β-catenin binding. The APC mutation cluster region extending from codon 1286 – 1483 coincides with the 3' end of the third 20 amino acid repeat and this possibly explains the most common observation of retention of atleast one 20 amino acid repeat in the mutant APC [89]. There are three SAMP repeats in between the third and fourth, fourth and fifth; and after the seventh 20 amino acid repeat in the central part of APC gene which are the sites for binding of Axin protein. It binds the APC protein via its RGS (regulator of G protein signaling) domain. Axin acts as a scaffold protein to form the functional complex with APC and β-catenin which bring about phosphorylation of both of these proteins by GSK3β. APC phosphorylation enhances the binding affinity for β-catenin thereby leading to efficient degradation of β-catenin [90].

- ***The Basic Domain:*** named because of the presence of a large proportion of basic residues like arginine and lysine is the domain present in the C-terminus of the APC protein between amino acids 2200 and 2400. This domain also contains an unusually high amount of proline residues. This domain binds to the microtubules and has been shown to bring about the polymerization of the tubulin in vitro [91].

- ***The EB1 and HDLG binding Domain:*** The C-terminus of APC (2560-2843 residues) contains binding sites for end binding protein EB1 which has been shown to direct the binding of APC to microtubules end [92]. The last 72 amino acids of APC also interacts with the HDLG (human homolog of the Drosophila discs large tumor suppressor) protein; this interaction has shown to suppresses cell cycle progression from G_0/G_1 to S phase [93].

❖ ***Nuclear Export and nuclear localization signals:*** It has been shown that APC functions as nuclear-cytoplasmic shuttling protein and acts a β-catenin chaperone [94]. There are at least five APC nuclear export signals (NESs) that contain the repeat motif LXXLXL/I/M/V. Among them, two are located at the N-terminal region and the other three are located in the 20-amino acid repeat region of β-catenin binding motif [95]. The nuclear export of APC is mediated by CRM1/Exportin receptor pathway [96]. The APC- β-catenin complex needs to be localized inside the nucleus for the signaling to take place. β-catenin enters the nucleus independently by binding to the nuclear pore machinery and does not require an import signal [97]. However, owing to its large size, APC needs nuclear localization signals (NLSs) which is recognized by the importins that mediate the nuclear translocation of APC in an energy dependent manner. APC contains two classis monopartite basic NLSs in the C-terminus and also an additional domain located between four of the seven armadillo repeats to facilitate its nuclear import [98].

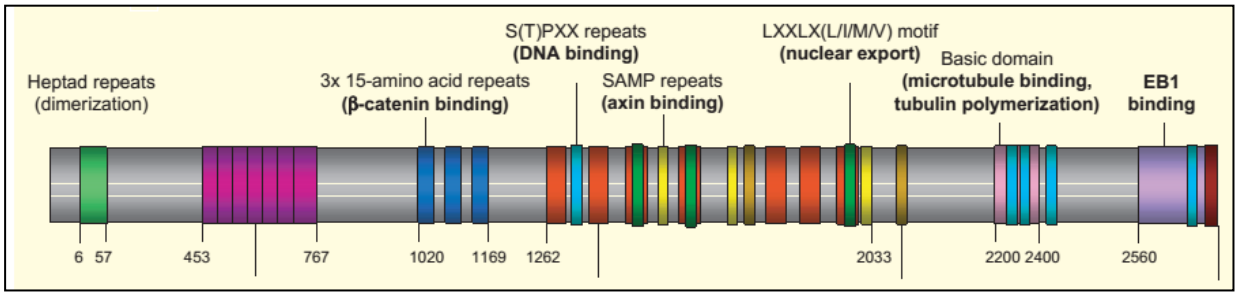


Fig 3.2 APC protein structure (Taken from Ref. 101)

3.2.3 Functions of APC protein

The major role of APC protein is in *wnt* signaling which is the master regulator of cell homeostasis in the colonic epithelium. Apart from this, APC is also involved in other cellular functions including cell-cell adhesion, cell migration, chromosomal segregation and apoptosis in colonic crypts.

❖ ***Regulation of colonic epithelium homoeostasis by wnt pathway***

The flat surface of the colon is covered by an epithelium composed of four cell types; *enterocytes* (absorptive cells), *enteroendocrine cells* (secretes hormones), *goblet cells*

(secretes mucus) and *Paneth cells* (secretes antimicrobial toxins). The colonic epithelium invaginates at regular intervals to form crypts (*Fig 3.3*). Stem cells those are present at the bottom of these crypts gives rise to the actively dividing precursor cells which occupy the bottom two-thirds of the crypt. The precursors migrate upward in an ordered fashion and they stop proliferation when they reach the top third of the crypt. Meanwhile, they continue their migration movement and colonize the surface of the colon. After about a week, epithelial cells undergo apoptosis and are shed in the lumen of the gut. Paneth cells are an exception in that they move downwards and populate at the base of the crypt. Thus, the epithelium of the colon is under perpetual renewal; being one of the most rapidly proliferating tissues of the body [99].

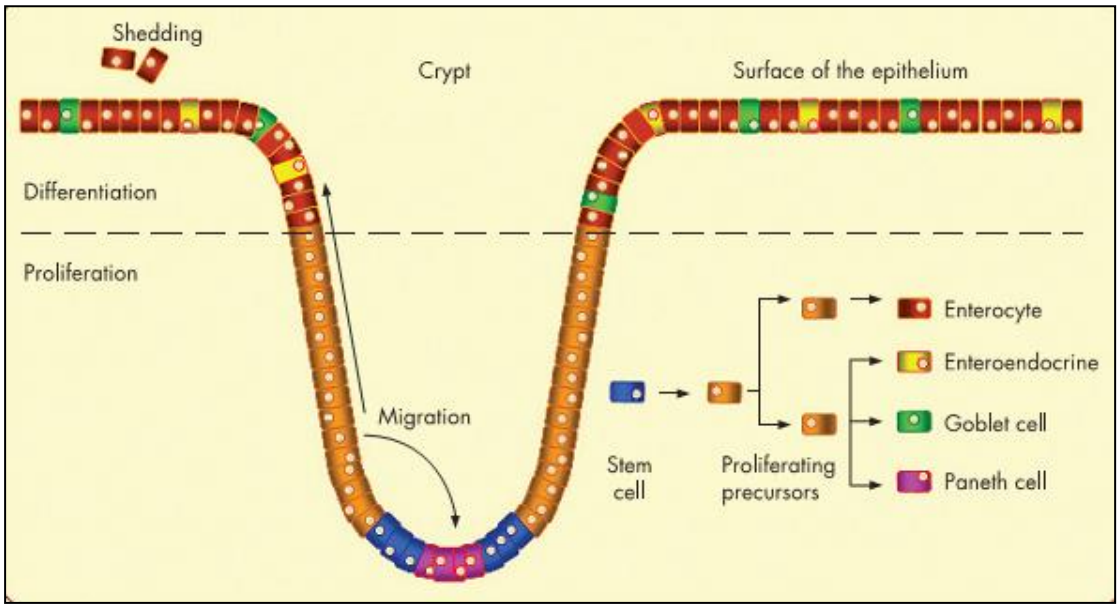


Fig 3.3 Schematic representation of the colonic epithelium (Taken from Ref. 99)

This homeostasis of colonic epithelium is regulated by the canonical Wnt pathway that controls the proliferation, differentiation, migration and sorting of the epithelial cell population. Deregulation of this pathway because of acquired mutations in the major components of this signaling system (predominantly APC/β-catenin mutations) is the initialing event in colorectal carcinogenesis.

❖ *Molecular mechanism of Wnt signal transduction*

The most critical player of canonical *wnt* signaling pathway is the β-catenin which is negatively regulated by the APC protein. The canonical Wnt signaling controls cell proliferation in a β-catenin dependent manner by activating the transcription of target

genes through the T cell factor/lymphoid enhancer factor-1 (TCF) family of activators. Wnt factors can also trigger the activation of β-catenin independent pathways which are referred to as non-canonical *wnt* signaling; however their role in colorectal tumorigenesis is not yet known.

In the absence of Wnt: the destruction complex

Under normal circumstances, the cytoplasmic level of β-catenin is regulated by a multiprotein destruction complex that targets β-catenin for degradation in proteasomes. This complex is assembled over the scaffold component Axin or its homologue Conductin, which contain binding domains for β-catenin, APC and the kinases GSK3β and CK1α/ε. The kinases phosphorylate APC at the 15 amino acid and 20 amino acid repeats which leads to increased binding of APC with β-catenin and simultaneously with Axin thereby acting as a bridge that brings Axin and β-catenin proteins together. Once complexed, dephosphorylation of APC by PP2A phosphatase reduces its affinity for β-catenin which now interacts strongly with Axin. This promotes the GSK3β mediated phosphorylation of β-catenin (at serine and threonine residues in the N-terminal) that creates a recognition motif for the F-box protein β-transducin repeat–containing protein (βTrCP) as part of ubiquitin ligase complex. The ubiquitin ligase complex then promotes polyubiquitination of β-catenin, leading to proteasomal degradation of β-catenin [99, 100, 101].

In the presence of Wnt or in case of APC mutation

Wnts are glycoproteins whose secretion is controlled specifically by the transmembrane protein Wntless/evenness interrupted. Wnt ligands bind to the receptor co-complex composed of transmembrane receptors frizzled and low-density lipoprotein receptor-related protein (LRP)-5 and LRP-6. This leads to the Dsh (disheveled) mediated phosphorylation of LRP by GSK3β which recruits Axin to the plasma membrane after which Axin is degraded and activity of GSK3β is repressed. Because of these events, no functional destruction complex is formed leading to the increasing levels of β-catenin in the cytoplasm. β-catenin now enters the cell nucleus and associates with TCF/LEF transcription factors to activate the transcription of Wnt target genes. TCFs lack transactivation function and repress gene transcription in the absence of β-catenin by interacting with Groucho and other corepressors. The transactivation function of TCF is complemented by β-catenin, which itself lacks any

DNA-binding activity [100-102]. Thus, the transcription of target genes can occur only by functional association of TCFs with β-catenin.

In case of inactivating mutations of APC and Axin and activating mutation of β-catenin in the N-terminal domain, the course of coordinated phosphorylation and destruction of β-catenin are disrupted; this process essentially mimics the constitutive activation of Wnt ligand–mediated signaling. Wnt target genes include proto-oncogenes like c-myc which regulate cell cycle progression, gap junctional protein connexin 43 and metalloproteases like matrilysin. The target genes also include Axin1, TCF1, Naked-1/2, Dickkopf-1 and Wnt inhibitory factor; all of which acts as negative feedback mechanism for this signaling [100-102].

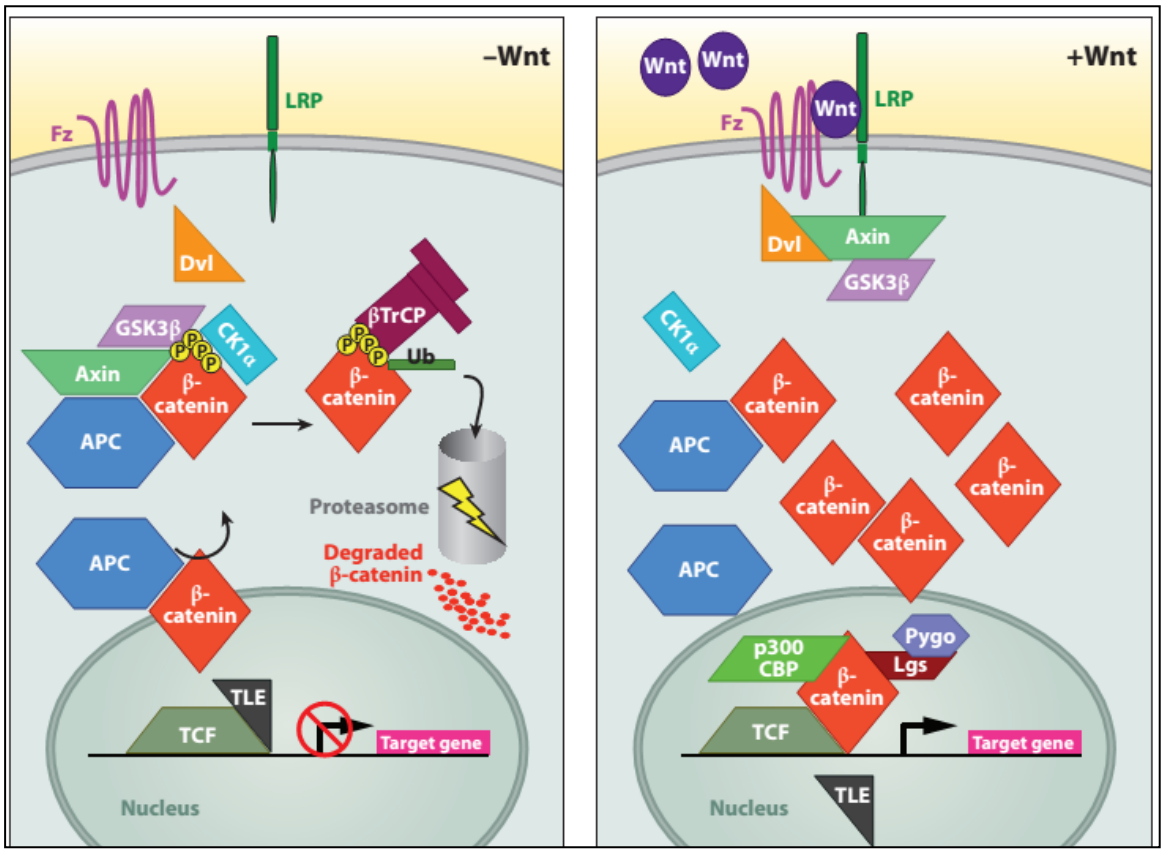


Fig 3.4 Canonical wnt pathway signaling (Taken from Ref. 84)

APC is an important regulator of β-catenin thereby keeping the levels of cytoplasmic levels of β-catenin under check in order to ensure proper proliferation and differentiation of colonic epithelium. APC regulates the β-catenin levels by one of the four mechanisms. Firstly, it targets β-catenin for degradation by participating in the destruction complex. Secondly, APC associates with the transcription co-repressors

like CtBP; guiding them to the protein complex. Thirdly, phosphorylated APC binds to the TCF/LEF factors and competes with β-catenin for binding; thus interfering with the active transcriptional activation complex formation. Lastly, APC assist in the nuclear export of β-catenin thus preventing its build up in the nucleus where it brings about transcription [98, 100].

❖ ***Other functions of APC***

APC contributes to the regulation of cytoskeletal proteins as it binds to the microtubules through its basic domain. It also interacts with Asef that enhances the binding between Asef and Rac, a molecule involved in the regulation of actin cytoskeleton. Therefore, loss of APC leads to changes in cell migration, cell orientation, polarity and division [103]. APC has been show to play a role in regulation of cell cycle. This function is mediated by its interaction with HDLG protein which when complexed with APC suppresses the progression of cell cycle from G0/G1 to S phase [104]. Involvement of APC in cell cycle can also be attributed to the β-catenin mediated transcription of S-phase regulators such as cyclin D and c-myc [105]. Apart from this, APC has also been shown to be involved in apoptosis, though the exact mechanism behind this remains elusive [106].

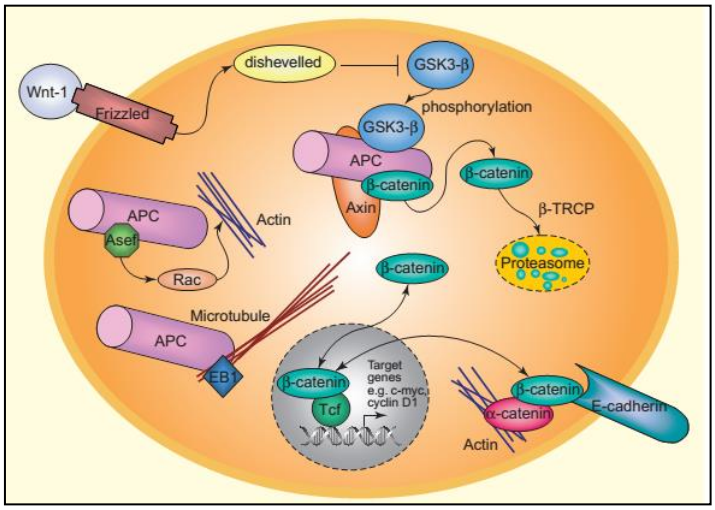


Fig 3.5 APC: the multifaceted protein (Taken from Ref. 101)

3.3 MUTATION SPECTRUM OF APC GENE

Germline mutations have been reported throughout the *APC* gene, which may differ in their penetrance and may have distinct phenotypic effects. More than 1500 mutations have been reported in the *APC* gene, with almost 75% being reported in

Exon 15. Nearly 95% of all the mutations lead to truncated protein formation; 67% being frameshift and the remaining 28% being nonsense mutations [107]. Such high frequency of frameshift mutations due to small insertions or deletions is due to the large number of repetitive elements within this gene. In APC gene, pathogenic missense mutations are rarely identified but large genomic rearrangements may contribute to up to 20% of all mutations. There exists a mutation cluster region (MCR) in the proximal end of Exon 15 between codon 1286-1583 [108]. FAP associated mutations are seldom reported beyond codon 1600 [108]. Two codon, 1061 and 1309 are mutational hotspots in different populations and have been reported in varying frequencies [109]. Irrespective of their frequencies, these two mutations together account for nearly 30% of all *APC* mutations reported [110]. Interestingly, the type of germline mutation in *APC* appears to determine the nature of second hit in the tumor. When the germline mutation lies in the MCR, second hit is usually the allelic loss of *APC* locus. However, when the germline mutation fall outside the MCR, second hit takes place in the MCR region [111]. In about a quarter of cases, germline APC mutations arise in the absence of family history suggesting a high frequency of *de novo* mutation [112].

3.4 GENOTYPE-PHENOTYPE CORRELATIONS IN FAP

It has been reported that there exist a good correlation between the site of APC mutation and the phenotype, in terms of the extent of polyposis, age of onset and occurrence of extracolonic manifestations [60]. Profuse polyposis characterized by the presence of thousands of polyps, is associated with mutation between codons 1250-1464 of APC [113]. Moreover, germline mutation in the codon 1309 is associated with most severe form of the disease and the younger age of onset [114]. While intermediate (classical) polyposis with thousand to hundreds of polyps is associated with mutation between codon 157 to codon 1595 of APC (excluding the MCR); the attenuated polyposis with less than 100 polyps is associated with mutation in the 5' or 3' end of the gene or in the alternatively spliced region of Exon 9 of APC gene [115]. Extracolonic manifestations are also reported to correlate with the location of mutation [60, 62]. Although good correlations have been established which is generally followed, it should be kept in mind that contradictions and inconsistencies have been reported and that these correlations are not absolute [60, 62].

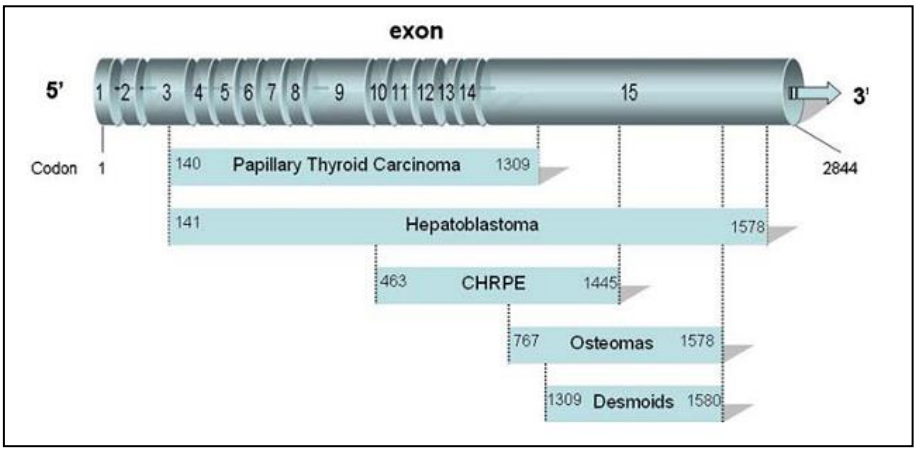


Fig 3.6 Genotype-phenotype correlations in FAP syndrome (Taken from Ref. 62)

3.5 MUTYH ASSOCIATED POLYPOSIS

MUTYH-associated polyposis (MAP) (OMIM #608456) is an autosomal recessive disease associated with adenomas and cancers of the colorectum. It is caused due to biallelic germline mutations in the base excision repair gene *MUTYH* that lead to an increase in 8-oxoG–induced somatic G:C>T:A transversions in other genes, including tumor suppressors such as the APC gene. This condition was first reported on three members from a single Welsh family [116]. *MUTYH* mutations are also identified in up to 23% of the APC negative FAP cases [117].

MAP is characterized by the appearance of multiple adenomas throughout the colorectum, usually numbering between dozens and sometimes exceeding hundred. Colorectal adenomas or colorectal cancer (CRC) usually become symptomatic between the 4th and 7th decade of life and the cumulative lifetime risk for CRC has been estimated to be up to 100% [118, 119]. Because of the overlap of the associated features, differential diagnosis of MAP with attenuated FAP must be done usually on the basis of the inheritance pattern. In contrast to the FAP, there is scarcity of information on the spectrum of extracolonic manifestations in MAP. One study reported increased occurrence of extracolonic malignancies in MAP patients with highest incidence of duodenal cancer [120]. They also concluded that the tumor spectrum in MAP resembles that of the Lynch syndrome; especially the presence of sebaceous gland tumors. Moreover, the FAP specific manifestations like desmoids and osteomas were not reported in MAP patients [120].

3.6 POLYMERASE PROOFREADING ASSOCIATED POLYPOSIS (PPAP)

PPAP is an autosomal dominantly inherited CRC predisposition syndrome characterized by the presence of adenomatous polyps (ranging from 5-100) and early onset of CRC [121]. It is caused due to germline mutations in the exonuclease domain of *POLD1* and *POLE* genes. Specifically two mutations; p.L424V in *POLE* and p.S478N in *POLD1* genes have been identified in most cases with this syndrome [122, 123]. Both these genes encode the catalytic (exonuclease domain) unit of the Polδ and Polε enzymes that are part of the mismatch repair pathway. Therefore, mutation in this protein domain affects the fidelity of the DNA replication that leads to hypermutated phenotype, a characteristic feature of tumors from POLD1 and POLE mutation carriers. The extracolonic tumor spectrum also includes cancers of brain, pancreas, ovary and small intestine in POLE mutation carriers while POLD1 mutation carriers are at an increased risk of developing breast and brain tumors along with CRC and endometrial cancers [123, 124].

3.7 NTHL1 ASSOCIATED POLYPOSIS (NAP)

The causal association of adenomatous polyps and CRC with bi-allelic germline mutations in *NTHL1* gene was first described in three Dutch families [125]. The name NAP was then proposed to define this autosomal recessive CRC predisposition syndrome [126]. Only ten NTHL1 mutation carriers have been reported in literature; c.268C>T (p.Q90X) identified in homozygous state in 9 patients and compound heterozygous c.268C>T and c.709+1G>A identified in case [125, 127, 128]. NAP is characterized by the presence of up to 50 colorectal polyps, most of which progresses to CRC. Another noteworthy feature of NAP is the development of multiple primary tumors of either different or same type. Female carriers are at an increased risk of developing breast and endometrial cancers. The other common manifestations include the occurrence of basal cell carcinoma, meningiomas and bladder tumors [128].

3.8 HYPOTHESIS AND OBJECTIVE OF THE STUDY

Even though FAP is a well characterized syndrome and most cases undergo comprehensive sequencing of APC and MYH gene followed by MLPA analysis, the underlying causative mutations is not identified in 6- 20% of all FAP cases [129, 130] Only a fraction of unexplained adenomatous polyposis (UAP) cases can now be explained by the germline mutations in *POLD1*, *POLE* and *NTHL1* genes identified through recent exome sequencing studies. The current knowledge regarding the spectrum of APC gene mutation, mutational hotspots and the genotype phenotype correlations is derived mainly from studies in Caucasian cohorts [131-133]. There have also been reports on East Asian FAP cohorts [134-136]. However, the mutation spectrum and phenotypic characteristics of South Asian FAP patients remains unknown as there has been only one case report so far on a FAP family from South India [137]. Moreover, in recent years, studies from different geo-ethnic groups have identified several novel APC genotypes, phenotypes and genotype-phenotype associations which reinforces that the mutation spectrum must be defined for every population. The underlying reason for differences in phenotypic associations has not been investigated but may be due to difference in the underlying genetic background or dietary habits. APC genotype-phenotype association studies in different geo-ethnic groups can enrich the existing knowledge about phenotypic consequences of distinct APC mutations and guide counseling and risk management in different populations.

We hypothesized that the Indian FAP 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 47 FAP families. The methodology was to PCR amplify the entire coding and flanking intronic region of APC gene followed by Sanger sequencing to identify the mutation. MLPA analysis was carried out to study LGRs in APC/MUTYH in cases where no point mutation or small insertion/deletion was identified in APC gene. If still the causative genetic defect was unidentified, MUTYH genetic analysis followed by NTHL1, POLD1 and POLE screening was carried out. Detailed phenotypic characterization was carried out based on the pedigree and medical records.

3.9 RESULTS

The 47 unrelated Indian FAP families reported here represent the diverse regions and religions of the Indian subcontinent with 15 hailing from northern, 12 from eastern, 13 from western and 7 from southern states of India and belonging to Hindu (41), Muslim (2), Christian (2) and Jain (2) religions. Of the 47 probands, 23 had no family history of polyposis or cancer suggesting a de novo mutation (representative pedigree in Fig 3.7). The remaining 24 probands reported a family history of polyposis with or without CRC or other extracolonic manifestations (representative pedigree in Fig 3.8). All the probands had classical polyposis except three attenuated FAP cases with less than 100 adenomatous polyps.

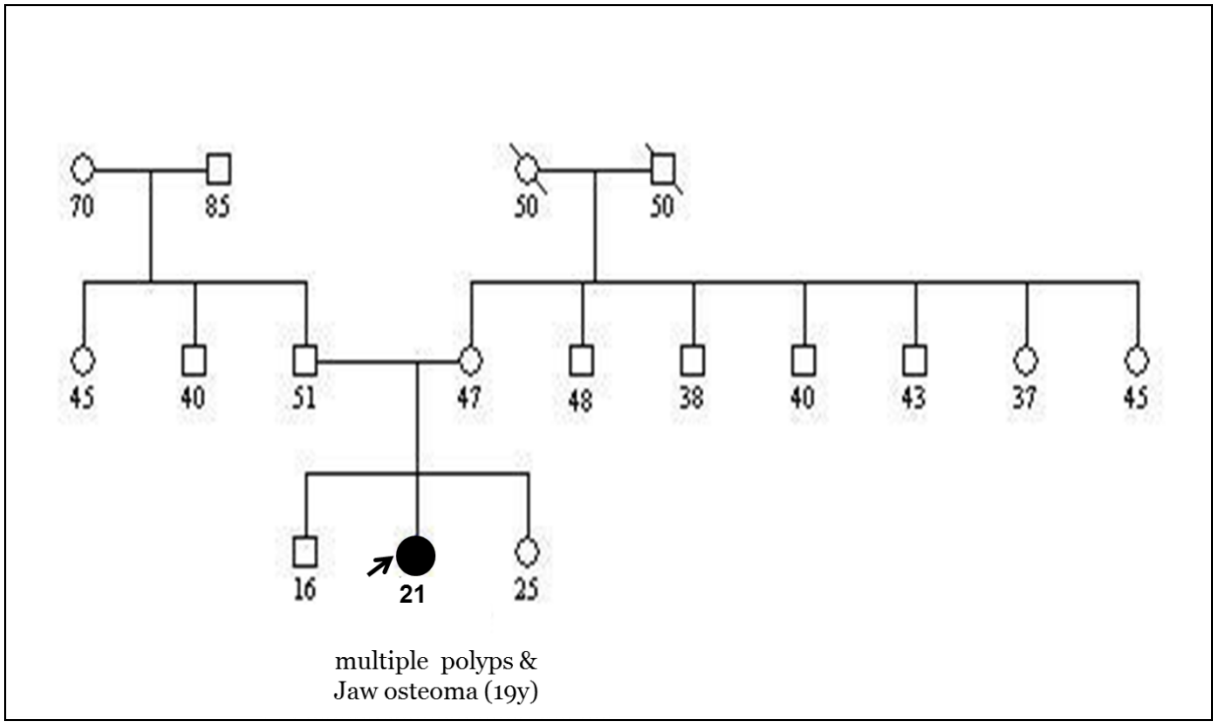


Fig 3.7 Representative pedigree of de novo FAP case (parents were tested negative for mutation)

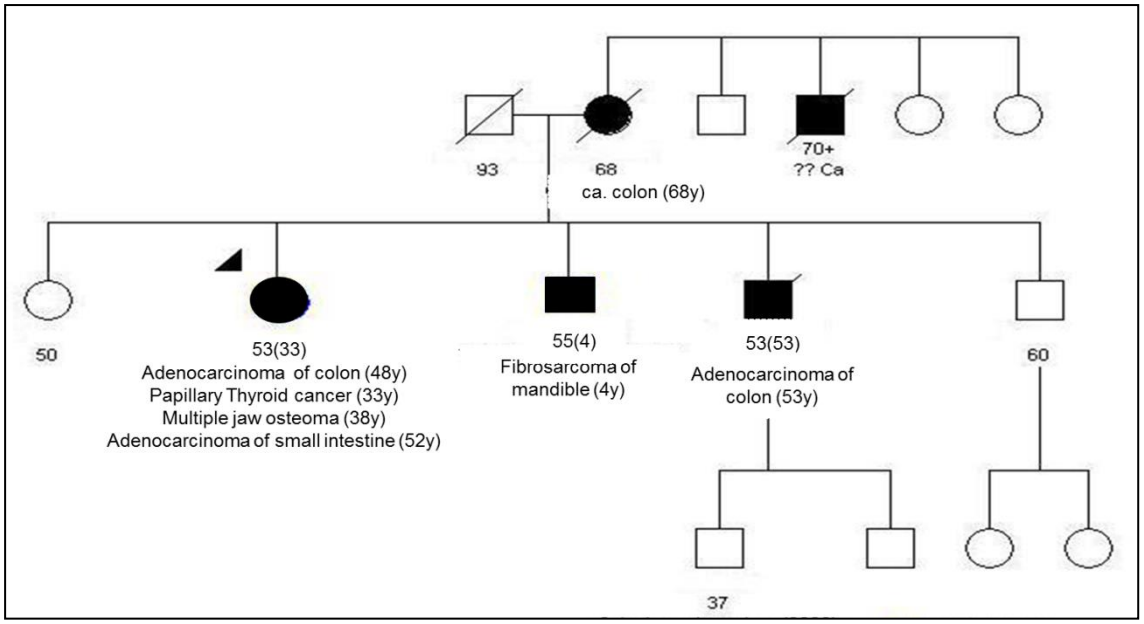


Fig 3.8 Representative pedigree of classical FAP family (all affected members had multiple polyps)

Through Sanger sequencing and MLPA of *APC* and *MUTYH* genes, 31 distinct deleterious germline mutations were identified in 40 families and 4 distinct germline bi-allelic *MUTYH* gene mutations identified in 2 families. With extended testing of 106 family members from 42 families, a total of 54 carriers of *APC* mutation and 4 carriers of bi-allelic *MUTYH* mutations were identified. In a combined analysis in 54 APC mutation carriers and their 50 untested relatives with FAP associated cancer or benign manifestation, the phenotypic features observed were 68 CRC, 5 upper GI cancers, 3 thyroid cancer, 2 brain tumors, 10 desmoid tumors/fibromatosis. CHRPE was noted in 9/29 APC mutation carriers for whom fundus examination details were available.

3.9.1 Mutation spectrum of APC and MUTYH gene.

Of the 31 distinct APC mutations described in Table 3.2 and Fig. 3.9, fourteen (45%) are novel mutations not previously described in the literature or the InSiGHT database. Vast majority of the mutations were truncating (16 frameshift & 12 nonsense), 1 splice site and 2 large genomic rearrangements (LGR). All the mutations were between codons 197 to 1538. The proximal exon 15 harbored 22 (71%) of all the mutations. A five base pair deletion at codon 1309 (c.3927_3931delAAAGA) was the most frequent mutation, identified in 7 unrelated families. Codon 1061 mutation (c.3183_3187delACAAA) was identified in 3 families. Interestingly, 4 distinct

truncating mutations at codon 935 occurred due to 4 different nucleotide alterations (c.2804dupA, c.2805_2815del11, c.2805 C >A and c.2802_2805delTTAC) in 4 families. The remaining 24 mutations were rare and identified in one family each. The APC LGRs identified were a duplication of the Promoter1B identified in two families and deletion of exons 9–13 in one family.

Table 3.2: Spectrum of APC mutations in Indian FAP cohort

Sr. No	Nucleotide change	Exon	Consequence	Type of mutation	No. of families with this mutation	Reported in InSiGHT* database or novel
1	c.589delA	5	p.R197Efs*8	Frameshift	1	Novel
2	c.706C>T	6	p.Q236*	Nonsense	1	Reported
3	c.1620dupA	12	p.Q541Tfs*19	Frameshift	1	Reported
4	c.1690C>T	13	p.R564*	Nonsense	1	Reported
5	c.1779G>A	14	p.W593*	Nonsense	1	Reported
6	c.1861dupA	14	p.T621Nfs*13	Frameshift	1	Reported
7	c.2274delA	15	p.A759Pfs*2	Frameshift	1	Novel
8	c.2802_2805delTTAC	15	p.Y935Ifs*19	Frameshift	1	Reported
9	c.2804dupA	15	p.Y935*	Frameshift	1	Reported
10	c.2805_2815del11	15	p.Y935*	Frameshift	1	Novel
11	c.2805C>A	15	p.Y935*	Nonsense	1	Reported
12	c.2828C>G	15	p.S943*	Nonsense	1	Reported
13	c.3183_3187del5	15	p.Q1062*	Frameshift	3	Reported
14	3259_3260delCT	15	p.L1087Qfs*31	Frameshift	1	Novel
15	c.3298dupT	15	p.S1100Ffs*19	Frameshift	1	Novel
16	c.3682C>T	15	p.Q1228*	Nonsense	1	Reported
17	c.3815C>A	15	p.S1272*	Nonsense	1	Novel
18	c.3927_3931del5	15	p.E1309Dfs*4	Frameshift	7	Reported
19	c.4012C>T	15	p.Q1338*	Nonsense	1	Reported
20	c.4037C>G	15	p.S1346*	Nonsense	1	Novel

21	c.4202_4203delTT	15	p.I1401Sfs*7	Frameshift	1	Novel
22	c.4216C>T	15	p.Q1406*	Nonsense	1	Reported
23	c.4285C>T	15	p.Q1429*	Nonsense	1	Novel
24	c.4387_4394dup8	15	p. S1465Rfs*11	Frameshift	1	Novel
25	4446delT	15	p.P1483Qfs*24	Frameshift	1	Novel
26	c.4463T>G	15	p.L1488*	Nonsense	1	Reported
27	c.4529delG	15	p.S1510Tfs*13	Frameshift	1	Novel
28	c.4612_4613delGA	15	p.E1538Ifs*5	Frameshift	1	Reported
29	IVS14+1G>A	--	--	Splice site	1	Reported
30	Deletion of Exons 9-13	--	--	LGR	1	Novel
31	Duplication of promoter 1B	--	--	LGR	2	Novel

In 2 of the 3 AFAP cases, biallelic MUTYH mutations were identified. A homozygous MUTYH mutation E466X (now E480X) was identified in a South Indian Tamil AFAP patient with 40 polyps and CRC. Compound heterozygous MUTYH mutations R241W (inherited from mother) and G286E (inherited from father) were identified in a case (pedigree in Fig 3.9) with less than 100 polyps but no CRC.

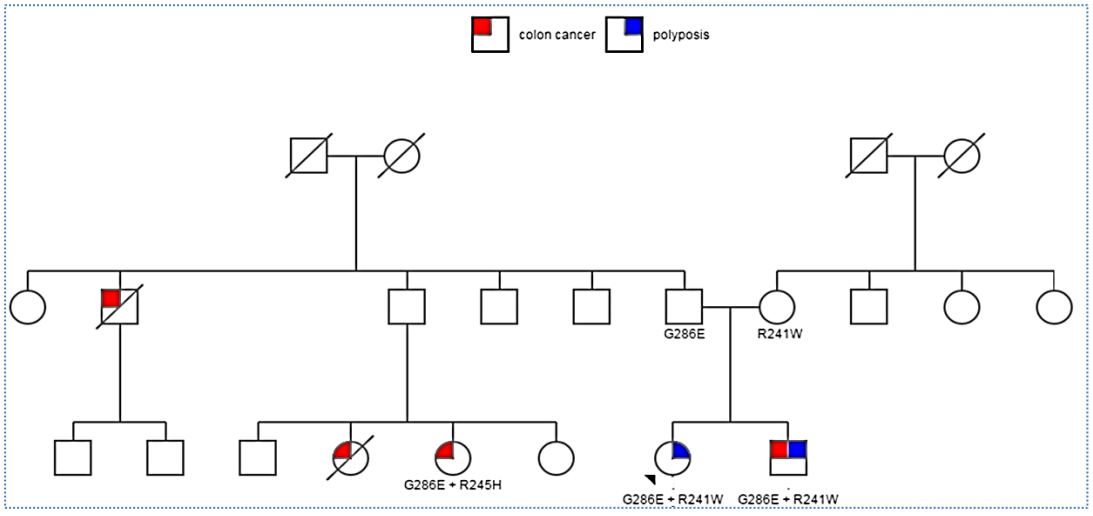


Fig 3.9 Pedigree of the family with compound heterozygous mutations in MUTYH gene

Proband's brother who had colon cancer at the age of 26 years also harbored the same combination of compound heterozygous mutation. However, one of the paternal cousins of proband who was affected with colon cancer at 45 years had different

compound heterozygous mutations; G286E which she must have inherited from her father and R245H which we suspect must have come from her mother.

In the 6 APC and MUTYH mutation negative cases with classical FAP phenotype, sequencing of the entire coding region of NTHL1gene and the exonuclease domain of POLD1 gene (exons 6–13) and POLE gene (exons 9–14) did not identify any mutation. Subsequently, in 4 of these 6 cases, targeted next generation sequencing Illumina panel of 86 genes was done at a commercial laboratory (Strand Genomics). We identified several VUS in many of these genes in each case, but no pathogenic mutation was identified in any of the 86 genes in the panel which included APC, MYH and all the MMR genes.

3.9.2 Phenotypic features and rare genotype-phenotype associations

Of the 54 APC mutation carriers, 27 had developed CRC at a mean age of 38.5 years (range 22–53 years) in a background of classical polyposis with hundreds to thousands of polyps in all but one case of AFAP with only 50 polyps. In 21 APC carriers, polyposis was diagnosed at a mean age of 29.5 years (range: 9–60 years) without CRC on endoscopic evaluation or histopathological examination of prophylactic procto-colectomy specimens. In the remaining 6 carriers, colonoscopy was yet to be performed or its details were not available. Six APC carriers developed extracolonic cancers with or without CRC. These included 2 cases with papillary thyroid cancer, 1 case with duodenal cancer, 1 case with intracranial germinoma, 1 case with papillary thyroid carcinoma and duodenal cancer, and 1 case with duodenal cancer and small intestine cancer. One or more benign extracolonic manifestations were identified in 25/54 APC mutation carriers. These included CHRPE (n = 9), desmoid tumor or fibromatosis (n = 10), upper GI polyps (n = 8) and osteomas (n = 3). Eight very rare FAP phenotypes or phenotypes rarely associated with mutations outside specific regions of the APC gene were observed. These include the second reported case of intracranial germ cell tumor in an APC carrier [138], absence of profuse polyposis and early onset CRC in 3 of the 7 codon 1309 mutation carriers as is classically described [114], attenuated phenotype with only 50 polyps at age 33 years in a codon 593 mutation carrier, desmoid tumor with codon 1228 mutation, papillary thyroid cancer with codon 1346 mutation and most interestingly CHRPE

with codon 1483 mutation [60, 62]. The APC mutation spectrum and novel genotype phenotype associations have been summarized in *Fig 3.10*.

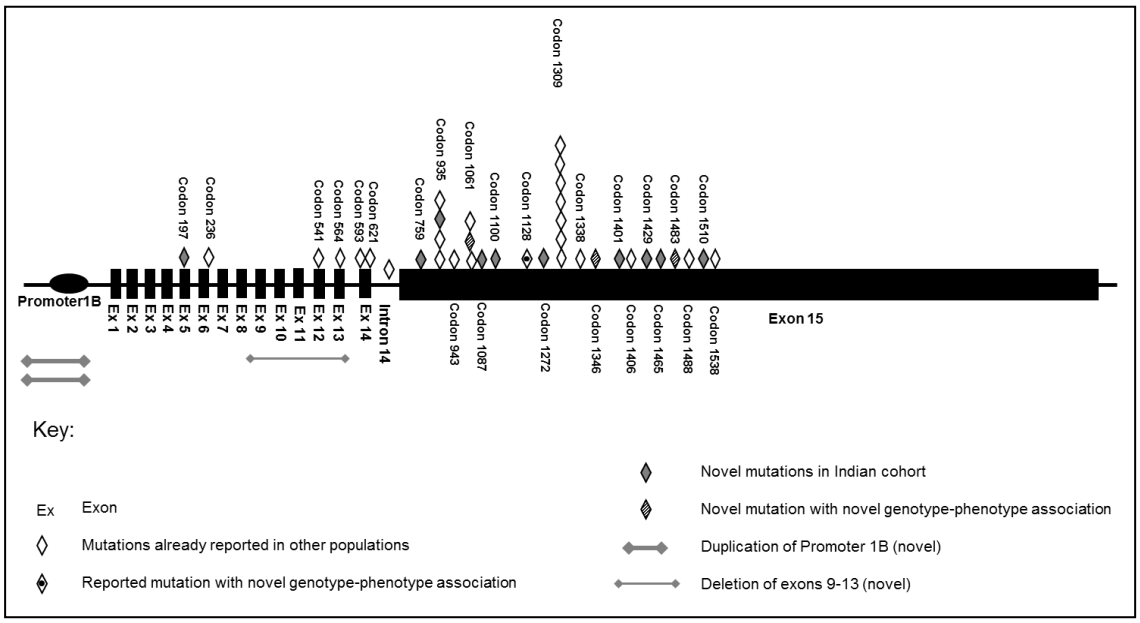


Fig 3.10 APC mutation spectrum & novel genotype-phenotype association in Indian FAP cohort. The mutation distribution shows clustering of two thirds of all APC mutations in proximal Exon 15, with three Indian mutational hotspots (codon 935, 1061 and 1309) contributing to one third of all APC mutations. Large number of novel APC mutations (n = 14) and few novel genotype phenotype associations for codon 1228, 1346 and 1483 mutations.

3.10 DISCUSSION

In FAP, the mutation spectrum of APC gene and genotype-phenotype correlations is well characterized for the Caucasian population and to some extent for the East Asian population [131-136]. Moreover, APC genotypes and genotype-phenotype associations rarely or never observed in Caucasian cohorts are now being increasingly reported from other geo-ethnic groups [135, 139-141]. This highlights the need to study different geo-ethnic groups to enrich the global APC mutational spectrum and expand our knowledge of phenotypic associations of distinct APC mutations. It is noteworthy that there exists no information on the characteristics of South Asian FAP cohorts. Also, comprehensive molecular characterization of all the 5 genes known to be associated with adenomatous polyposis has been performed in very limited number of cases, that too only in the Caucasian population. Our study is the first report of a South Asian cohort of 53 FAP families and the only non-Caucasian FAP cohort analysed for all the 5 adenomatous polyposis associated genes.

Using a combination of Sanger sequencing of APC and MYH gene with MLPA analysis to study large genomic rearrangements, we were able to identify deleterious germline APC or MUTYH mutations in 42 of 47 families, suggesting a mutation detection rate of 85%. This is comparable to the mutation detection rates observed for APC gene in other studies [129]. The mutation detection rate varies significantly between various populations and ranges from as low as 40-60% [132, 142, 143] to as high as 75-94% [129, 133, 144]. This variation can be partly explained by the difference in the stringency for making syndromic diagnosis and the methods used for mutation screening. The high mutation detection rate in our cohort reflects the appropriateness of our clinical characterization for making the syndromic diagnosis and the use of comprehensive genetic analysis.

Nearly all the mutations identified in our cohort were truncating mutation resulting in the loss of the β-catenin binding and degradation domains of APC protein, thus leading to the classical FAP phenotypes. This study has identified a new Indian mutational hotspot at codon 935 seen in 4 (10%) FAP families. In addition, the two other known hotspot mutations at codons 1309 and 1061 were seen in 18% and 9% families respectively. High frequency of codon 1309 and 1061 mutations worldwide [143] is a result of repetitive nucleotides in DNA sequence making it a mutational hotspot. Identification of APC LGR in 3 of the 10 families negative for APC point mutation or small indels and biallelic MUTYH mutation in 2 of the 7 families without APC mutation or LGR mandates its inclusion in comprehensive genetic analysis for south Asian FAP/AFAP cases.

The MUTYH mutation E466X (now E480X), previously described in 3 unrelated Indian families living in the UK [145] was identified as a homozygous mutation in one of our AFAP case from Tamil Nadu in south India. E466X may thus be a founder MUTYH mutation in Indians, possibly of Tamil ancestry. The founder effect of E466X needs to be confirmed with haplotyping studies and its population frequency can be established in a larger cohort. NTHL1, POLD1or POLE mutations were not identified in any of the 6 FAP probands negative for APC or MUTYH mutations. This is not surprising as none of these families fulfilled the salient features of PPAP or NAP as described in the literature [123, 127, 128].

Of the 35 distinct mutation identified in our cohort, 14 (45%) are novel and not previously reported in Caucasian or other geo-ethnic groups. This high frequency of novel mutations points towards a unique spectrum of APC mutations in Indian FAP cohort. The phenotypic characteristics and the genotype-phenotype correlations observed in our cohort largely concur with the Caucasian data with some notable exceptions. Eight very rare FAP phenotype or phenotypes rarely associated with mutations outside specific regions of the APC gene were identified. Interestingly, three of the cases with novel phenotypes or phenotypic associations harbor novel mutations. This suggests that novel mutation expands the knowledge on genotype and phenotypic spectrum. Also such deviations from the established genotype-phenotype correlations reiterates that these correlations, though useful for guiding genetic testing in some cases, are not absolute and must be used in combination with clinical data for taking important decisions about genetic testing, surveillance and treatment.

As our study is the first comprehensive report of APC and MYH gene mutation analysis, based on the mutational spectrum and hotspots identified, we propose a pragmatic stepwise genetic testing algorithm for FAP cases in south Asian countries where genetic testing is not routinely performed due to resource constraints (Fig. 3.11). Initial screening of three amplicons (15D–15F) harboring the mutational hotspot codons 1309, 1061 and 935 could identify 40% of all APC mutations and sequencing of additional 3 amplicons of exon 15 (15 C, 15 G, 15 H) could identify two thirds of all APC mutations. If no mutation is identified rest of the APC should be screened followed by LGR analysis and MUTYH gene sequencing. Extended testing of other adenomatous polyposis associated genes (NTHL1, POLD1 and POLE) may be considered but the yield is likely to be very low. The present study and few recent reports [146] highlight that a significant proportion of FAP cases do not harbor pathogenic mutations in the genes known to be associated with FAP, MAP, NAP, PPAP syndrome.

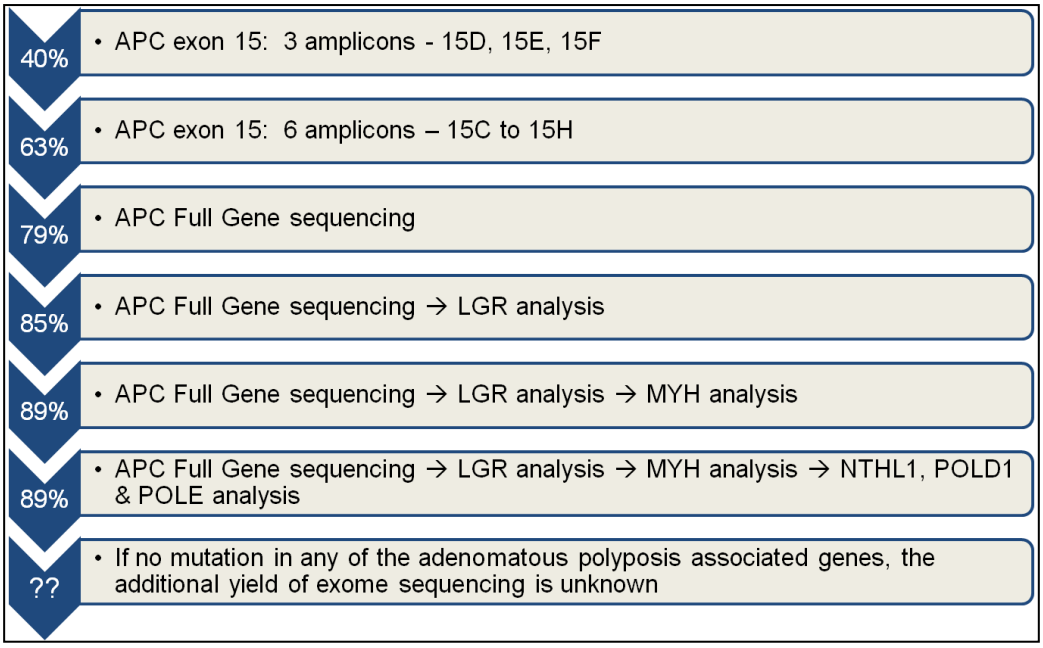


Figure 3.11 A pragmatic stepwise screening strategy to improve mutation detection rates in FAP patients.

Cumulative mutation detection rates with step wise screening of exons/genes most likely to be mutated in south Asian FAP cases. Arrows on left side shows the cumulative mutation detection rates in our cohort achieved after each step. In our cohort, the cumulative mutation detection rate did not change with NTHL1, POLD1 and POLE gene analysis it may increase the detection rate slightly in larger cohorts of APC and MUTYH negative adenomatous polyposis cases from different geo-ethnic background.

Germline exome sequencing in an adenomatous polyposis cohort has recently reported loss-of-function germline mutations in a few promising candidate genes (DSC2, PIEZO1, ZSWIM7) [146] and biallelic mutations in MSH3 gene [147]. However these recently identified adenomatous polyposis genes are likely to remain under-reported, unless they are tested as single genes or included in multi-gene next generation sequencing (NGS) panels. The currently used multi-gene panels may not be informative as they do not include NTHL1, POLD1and POLE genes. The lack of utility of commercial NGS panels for APC / MYH negative classical FAP cases is also demonstrated in our study which failed to identify pathogenic mutation in any of the 86 genes in the targeted NGS panel in 4 classical FAP cases. Therefore there is a need to conduct comprehensive genetic analysis of all the known adenomatous polyposis genes or exome sequencing studies in large pooled cohorts of APC and MUTYH negative adenomatous polyposis cases with detailed phenotypic and geo-ethnicity correlation.

CHAPTER 4

Study of Mutation spectrum of MMR genes in Indian Lynch Syndrome Patients

4.1 HISTORY OF LYNCH SYNDROME

The history of Lynch Syndrome (LS) began in 1895 with Alfred Scott Warthin, who was the first person to describe a family with history of multiple cancers, particularly of the colorectum, stomach and uterus. This was the family of his seamstress who lamented her inevitable death from cancer as had occurred with many of her family members earlier. Unfortunately, her fear came true and she died of endometrial cancer a few years later. Warthin drew her family tree and labeled it as Family G, as the family immigrated to USA from Germany (Fig 4.1- Family G pedigree). Pedigree of Family G remains one of the longest family cancer histories ever recorded. He also documented her medical history and the pathological finding of cancers in the family.

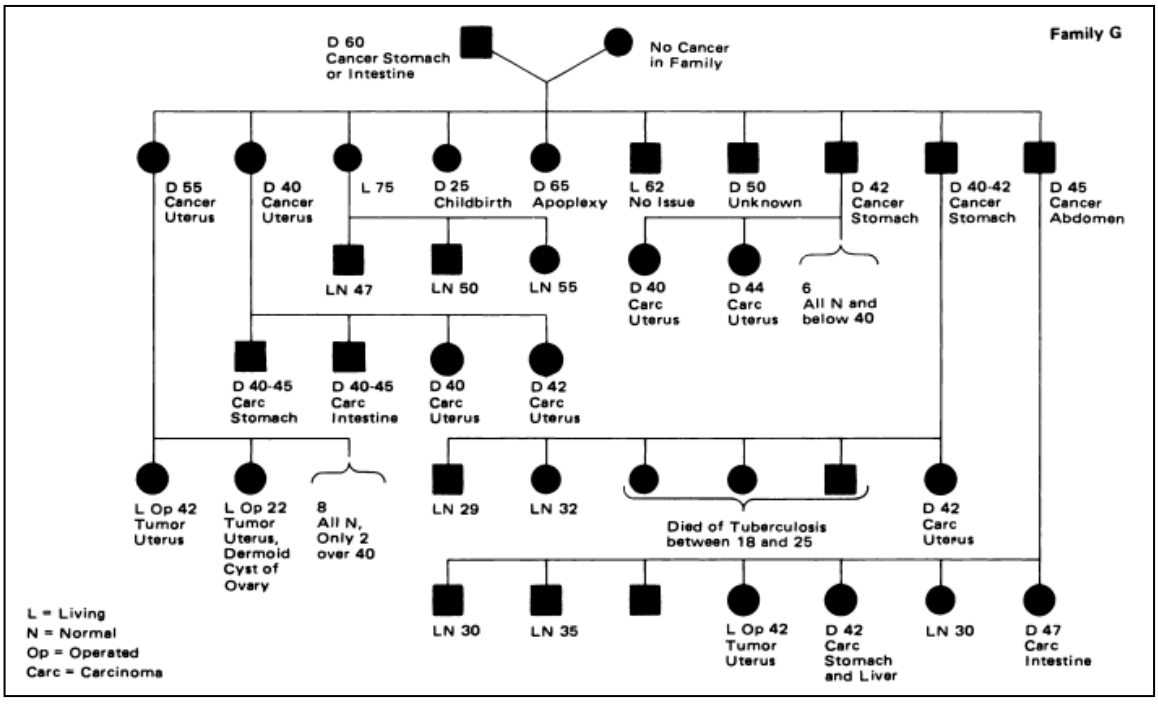


Fig 4.1 Pedigree of Family G (Taken from Ref. 148)

Being a pathologist at the University of Michigan, Warthin also studied nearly 1600 carcinomas of which 1000 carcinoma cases gave a history of cancer in their families. However, very detailed family history was available for only 4 of these cases including Family G which gave a clue about inherited susceptibility to cancer. Warthin published his findings in 1913 reporting one of the most comprehensive observations on familial clustering of cancer [148]. He also noted that transmission of the cancer phenotype within these families was consistent with Mendel's proposal of

autosomal dominant inheritance. Warthin died in 1931, later his colleagues Hauser and Weller published an update on Family G in 1936 [149].

In the next few decades, only occasional case reports about this phenotype were reported. Then in 1962, a resident doctor in University of Michigan Hospital, Henry Lynch came across a patient from Nebraska reporting similar family history as that of Warthin's seamstress. The proband, while recovering from delirium tremens, told Lynch that he drank because he was convinced that he would die of colorectal cancer (CRC), as "everybody" in the family died of this disease. After documenting a detailed family history, Lynch observed predominance of colorectal cancer transmitted through multiple generations in this family. Lynch thought of familial adenomatous polyposis (FAP) as the possible syndrome in this family as at that time this was the most favored diagnosis of CRC-prone families. However, as there was absence of florid polyposis, this assumption failed. This made Lynch contemplate if this could be an undescribed syndrome with a segregating pattern of CRC predisposition showing autosomal dominant inheritance but without the presence of the multiple colonic adenomas found in FAP. Later other cancers, particularly of the endometrium, were identified throughout this family, which Lynch labelled Family N (for Nebraska).

Marjorie Shaw of the University of Michigan also reported a family with clinical and pathological findings comparable to those of Family N; this family was labelled Family M (for Michigan). The pedigrees from both families were published in 1966 [150]. This study revived the interest in this syndrome which remained dormant since the work of Warthin. However, Lynch had to face a lot of skepticism when he proposed the concept of hereditary cancer syndrome, as at that time the exposure to same environmental factors was considered as the contributing factor for familial cancers. The fact that both families were part of Midwestern farming communities that experienced exposure to pesticides and other carcinogens seemed consistent with this concept.

The then chairperson of pathology at the University of Michigan, School of Medicine invited Lynch to study Warthin's Family G. Lynch organized a family union for Family G near Ann Arbor, MI. Along with Anne Krush, a medical social worker, Lynch conducted a detailed medical genetic investigation of the family, obtained data

on >650 family members (95 had developed cancers by this time) and found a predominance of cancers of the colon, uterus and stomach in the kindred. This iconic study was published in 1971as "Cancer Family 'G' Revisited" and the term 'Cancer Family Syndrome (CFS) was used for the first time in this manuscript [151]. The final update on Cancer Family G was published in 2005 with the data on 929 descendants of the ancestor and reported the specific germline mutation in this family [152].

The term 'Cancer Family Syndrome' was later replaced with the name 'Lynch Syndrome' (LS) in 1984 [153]. In the same year, the term Hereditary Non-Polyposis Colorectal Cancer (HNPCC) came into picture to differentiate this phenotype from FAP which is characterized by presence of multiple colorectal adenomatous polyps [154]. Subsequently when many extracolonic cancers were recognized as an integral part of this syndrome, the term HNPCC was considered a misnomer. Adding to the complexity was the reported incidence of adenomas in LS cases after which the use of term HNPCC for this syndrome seem inappropriate [155]. Therefore, by consensus, this syndrome is now referred to as **Lynch Syndrome**.

4.2 FEATURES OF LYNCH SYNDROME

Lynch Syndrome is an autosomal dominant colorectal cancer predisposition syndrome with high penetrance (70-85%). It is one of the most common inherited cancer syndromes and accounts for nearly 1-7% of all colorectal cancers. LS is caused due to germline mutations in one of the Mismatch repair (MMR) genes and is characterized by the early onset of colorectal and other extracolonic cancers. Lynch syndrome is divided into two types : Lynch Syndrome I which accounts for families with history of only colorectal cancers and Lynch Syndrome II which takes into consideration the families having history of extracolonic cancers along with colorectal cancers [156]. The diagnosis of LS can be made on the clinical grounds, by taking into consideration various clinico-pathological features and family history, or on molecular basis, by identification of pathogenic germline mutations.

4.2.1 Clinico-pathological features of LS:

Various clinical and histopathological features that are known to be associated with and aids in diagnosis of Lynch Syndrome are summarized in Table 4.1

Table 4.1 Clinicopathological features of Lynch syndrome

Clinical features
Early age at onset (45 years) as compared to general population (69 years)
Predominance of right sided tumors (proximal colon)
Increased occurrence of synchronous and metachronous cancers
Rapid adenoma to carcinoma progression compared with sporadic adenomas
Histological features
Poorly differentiated tumors
Increased mucin production
Host-lymphocytic infiltration and lymphoid aggregation around the tumor margin (Crohn's like reaction).
Higher proportion of signet ring cell carcinomas

Poor differentiation, a higher proportion of signet-ring cell, and increased mucin production, are considered as signs of an aggressive tumor behavior, whereas peritumoral lymphoid response and a Crohn's-like pattern might be indicative of a host defence mechanism, suggesting a favorable prognosis. Although most carcinomas in LS have aggressive histological features, it has been suggested that the prognosis of LS colorectal cancers in general is better than that of sporadic cases [157, 158].

4.2.2 Clinical criteria for LS diagnosis

Lynch syndrome became defined by an evolving series of clinical criteria (Table 4.2). These clinical criteria were introduced to standardize the inclusion criteria for patients. The first criteria, Amsterdam I Criteria were put forth by International Collaborative Group on Hereditary Non Polyposis Colorectal Cancer (ICG-HNPCC) in 1991 after a meeting of clinicians and researchers in Amsterdam, Netherlands [159]. Amsterdam I Criteria failed to acknowledge the contribution of extracolonic cancers and lead to underdiagnoses of the syndrome. So in 1999, the Amsterdam II criteria were formulated which takes into account the extracolonic cancers [160]. The discovery of Microsatellite Instability (MSI) in LS-associated tumors in 1993 revolutionized the diagnosis of LS patients. In recognition of the importance of MSI as a characteristic of LS tumors, Bethesda guidelines were proposed in 1996 [161]. In 2004, the Bethesda Guidelines were revised [162].

Table 4.2 Clinical criteria for diagnosis of LS

Amsterdam criteria I
At least three relatives should have histologically verified colorectal cancer
One should be a first-degree relative of the other two
At least two successive generations should be affected
At least one of the relatives should be diagnosed before the age of 50 years
Familial adenomatous polyposis should be excluded
Amsterdam criteria II
At least three relatives should have an histologically verified HNPCC-associated cancer (colorectal
At least two successive generations should be affected
One should be a 1st-degree relative of the other two
At least one of the relatives should be diagnosed before the age of 50 years
Familial adenomatous polyposis should be excluded in colorectal cancer case(s)
Bethesda Guidelines
Colorectal cancer in a patient who is less than 50 years of age
Synchronous or metachronous colorectal or other HNPCC-related tumor, regardless of age
Colorectal cancer with the MSI-H histology diagnosed in a patient who is less than 50 years of age.
Colorectal cancer or HNPCC-related tumor in one or more first degree relatives, with one of the cancers being diagnosed under age 50 years.
Colorectal cancer diagnosed in two or more first- or second-degree relatives with HNPCC-related tumors, regardless of age.
Revised Bethesda Guidelines
Colorectal cancer diagnosed in a patient who is less than 50 years of age.
Colorectal cancer with the MSI-H histology diagnosed in a patient who is less than 60 years of age.
Colorectal cancer and one or more first-degree relatives with an HNPCC-related tumor, with one of the cancers being diagnosed under age 50 years.
Colorectal cancer diagnosed in two or more first- or second-degree relatives with HNPCC-related tumors, regardless of age.

4.3 TUMOR SPECTRUM OF LYNCH SYNDROME

A significant enigma in LS (in common with a majority of familial cancer syndromes) is the specific spectrum of tumors in the germline MMR mutation carriers. Since the first description of this syndrome, colorectal and endometrial cancers were recognized as the LS associated tumors. In 2005, a study by Watson and Riley identified significantly increased frequencies of stomach, small bowel, hepatobiliary system,

upper urologic tract, ovarian cancers and brain tumors (particularly glioblastoma) in LS families [163]. In the recent years, evidence has been given for the increased incidence of pancreatic, prostate and the rare adrenocortical tumors in the background of Lynch Syndrome as compared to the general population risk [164-166]. Table 4.3 summarizes the comparative risks of various cancers in Lynch syndrome and general population with the mean age of onset of each cancer [167-169]. Whether or not, breast cancer is part of LS tumor spectrum is still controversial with studies both in favor and against it. The observed standardized incidence ratios of breast cancer vary from comparable to the average population [170-171] to significantly elevated [172-173], making it difficult to conclude whether or not it should be included in the LS spectrum. Phenotypic variants of Lynch syndrome is given in Fig 4.4

Table 4.3 Comparative risk of cancers in lynch syndrome

Cancer type	General population	Risk in Lynch	Mean age of onset in
Colon	5.5%	52-82%	44-61 years
Endometrium	2.7%	25-60%	48-62 years
Stomach	<1%	6-19%	56 years
Ovary	1.6%	4-12%	42.5 years
Hepatobiliary tract	<1%	2-7%	Not reported
Urinary tract	<1%	1-4%	55 years
Small bowel	<1%	3-6%	49 years
Pancreatic	1.5%	3-4%	Not reported
Brain/CNS tumors	<1%	1-3%	50 years
Sebaceous neoplasms	<1%	1-9%	Not reported

Table 4.4 Phenotypic variants of Lynch syndrome

Syndrome	Features
Muir-Torre syndrome	Autosomal dominant condition.
	Characterized by development of at least one sebaceous gland tumor or keratocanthoma that are associated with visceral malignancy including colorectal, endometrial, urological and upper GI tumors.
	MTS patients may have germline mutation in any of the MMR gene; though MSH2 mutations are identified more frequently.
Turcot syndrome	Autosomal dominant condition
	Characterized by coexistence of colorectal cancer and brain tumors (particularly glioblastoma)
	Germline mutation in MMR genes (PMS2)

4.4 MICROSATELLITE INSTABILITY: HALLMARK OF LS TUMORS

Microsatellites are type of simple sequence repeats present in human genome. It consists of multiple repeating units of upto1-6 nucleotides. One of the most important characteristics of microsatellites is that they are stably inherited and hence the length of repeats remains the same in all cells of the body [174]. Microsatellites are very prone to replication errors caused due to slippage of polymerase on DNA strand with repetitive nucleotides thereby leading to either deletion or addition of nucleotides [175]. MSI is defined as alterations in the lengths of microsatellites due to deletion or insertion of repeating units to produce novel length alleles in tumor DNA when compared with the normal/germline DNA from the same individual [176]. These errors are generally repaired by a proficient Mismatch Repair (MMR) system and hence presence of MSI is an indicative of deficient MMR system.

MSI was simultaneously reported in sporadic and familial colorectal cancers by three different groups in 1993 [177-179]. The molecular mechanism behind microsatellite instability is illustrated in Fig 4.2. During replication of repeat sequences, DNA strand denaturation may occur, resulting in strands misalignment. This may lead to the addition (or subtraction) of one or more nucleotides during replication. The extra nucleotide bulge is recognized by the mismatch repair (MMR) heterodimer MSH2–MSH6 which, together with MLH1–PMS2, promotes EXO1 mediated excision of the errant daughter strand. However, in the absence of MMR activity, the extra nucleotide remains. During the next round of DNA replication, the errant strand acts as a template strand, successful replication of which results in permanent fixation of the additional nucleotide and the generation of a new allele, thus leading to microsatellite instability.

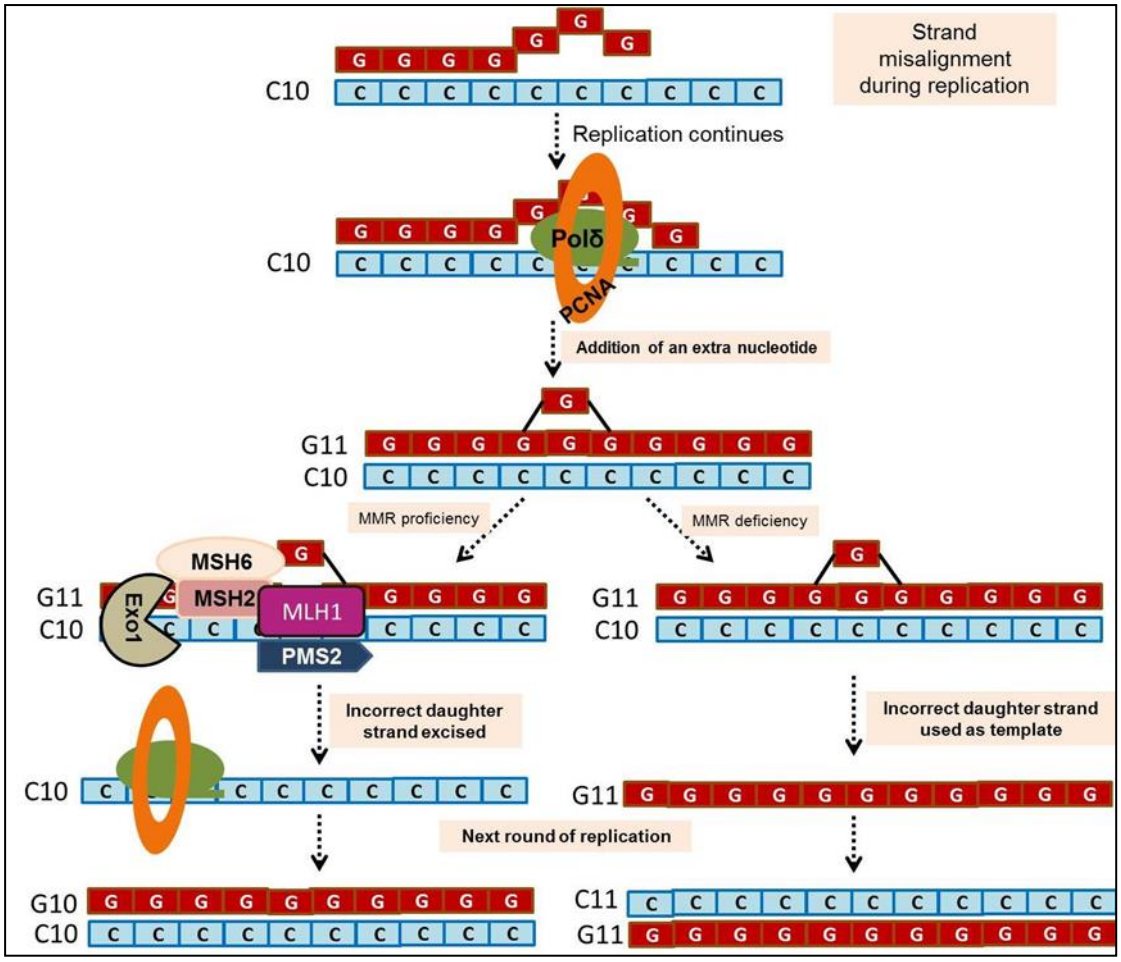


Fig 4.2 Molecular mechanism of microsatellite instability

EXO1 - exonuclease 1; PCNA - proliferating cell nuclear antigen; Polδ - DNA polymerase δ.

MSI is identified in over 85% of LS-associated CRCs and in 40-60% of LS-associated endometrial cancers and is therefore regarded as the hallmark of LS tumors [180]. MSI testing is therefore used as one of the diagnostic criteria for identifying LS cases and a prescreening technique for genetic testing in these cases. MSI testing is carried out by using a panel of microsatellite markers which are now studied by fragment analysis. In an attempt to standardize MSI analysis, a panel of five microsatellite markers known as Bethesda Panel was recommended by National Cancer Institute in 1997 [181]. This panel consisted of two mononucleotide markers (BAT-25 and BAT-26) and three dinucleotide markers (D2S123, D5S346 and D17S250). If two or more of the five loci show instability, the tumor is MSI-H; if only one locus is unstable, the tumor is MSI-L; and if all five loci are stable, the tumor is MSS [182]. It soon became apparent that mononucleotide markers were superior to the dinucleotide markers for assessment of MSI as it is easier to interpret. Moreover, the mononucleotide repeats were shown to be quasimonomorphic, meaning that normal tissue was not required

for MSI testing in most cases [183, 184]. However, the use of BAT26 alone is not recommended for diagnostic MSI screening because of the existence of polymorphisms in approximately 10 percent of the African population that can lead to false positive results for MSI [185].

To overcome these shortcomings, a comprehensive study was undertaken which proposed a pentaplex marker panel for MSI analysis comprising of five mononucleotide markers BAT25, BAT26, NR21, NR22 and NR24. The pentaplex panel which is now commercially available showed 100 percent sensitivity and 100 percent specificity for the detection of MSI and can be used without the need to test matching normal DNA. Several independent studies have found the pentaplex panel performs better in terms of sensitivity and specificity than the original NCI panel. By consensus, microsatellite status has been divided into three groups: microsatellite stable (MSS), with no instability seen; low-level instability (MSI-L), with less than 40% instability; and high-level instability (MSI-H), with more than 40% of microsatellite loci showing instability [186, 187]. Furthermore, the simultaneous assessment of just two markers, BAT26 and NR24, was shown to be as effective as the pentaplex panel for the diagnosis of MSI [188]. A detailed description of Bethesda panel and other markers is given in Table 4.5

Table 4.5 Microsatellite markers used in MSI analysis

Panel	Marker	MS repeat	Gene	Chromosomal location
Bethesda Panel	BAT25	A (25)	c-kit (Intron 16)	4q12
	BAT26	A (26)	MSH2 (Intron 5)	2p21
	D2S123	CA (n)	Linked to MSH2	2p16
	D5S346	CA (n)	Linked to APC	5q22-23
	D17S250	CA (n)	Linked to P53	17q12
Pentaplex Panel	BAT25	A (25)	c-kit (Intron 16)	4q12
	BAT26	A (26)	MSH2 (Intron 5)	2p21
	NR21	T (21)	SLC7A8 (5'UTR)	14q11
	NR24	T (24)	ZNF2 (3'UTR)	2q11
	NR27	A (27)	IAP-2 (5'UTR)	11q22

4.5 MOLECULAR GENETICS OF LYNCH SYNDROME

The phenomenon of repeat tract instability earlier described in lower organisms that had defects in their mismatch repair pathway led the basic scientists to focus on identifying a homologous pathway for mismatch repair in humans. Simultaneously, other groups were aiming to study the genetic basis of LS by traditional linkage analysis method in families following the Amsterdam I criteria. The first genetic locus linked to LS was mapped to chromosome 2p21 [189], chromosome 3p21-23 being the second locus mapped in families with MSI+ tumors [190]. However, not all the LS families known at that time showed linkage to these loci; thereby pointing towards further genetic heterogeneity associated with LS.

Taking into consideration that a human homologue of mismatch repair gene was a likely candidate, MSH2 gene (human homologue of bacterial mutS and yeast MSH genes) was identified as the underlying cause of LS [191]. Subsequently, MSH2 was mapped to the first LS-linked region, chromosome 2p21 using positional cloning [192]. An year later, two groups simultaneously cloned and mapped MLH1 gene (a homologue of yeast MUTL genes) to the second chromosomal region linked to LS, 3p21-23 [193, 194]. They also identified deleterious MLH1 mutations in many LS families. In the same year, association between germline mutation in PMS1 (2q31-33) and PMS2 (7p22) was reported [195]. However, the role of PMS1 in LS was challenged by subsequent studies [196]. Another important discovery was the identification of MSH6 gene mutation in a LS family not complying with the Amsterdam criteria I; as they had increased occurrence of extracolonic cancers and an age of onset over 50 years [197]. This study also for the first time proposed the concept of phenotypic heterogeneity associated with mutations in different MMR genes, a concept that has been corroborated through many subsequent studies. In 2000, another MMR gene, MLH3 was shown to be associated with Lynch syndrome [198]. However, very few MLH3 mutation carriers have been reported so far and there exist no conclusive evidence for its role in LS [199, 200].

4.5.1 Human Mismatch Repair Pathway

The integrity of DNA is constantly under threat by many genotoxic agents both within and outside the cell. It is estimated that the DNA in a typical mammalian cell is exposed to 10^4 to 10^5 lesions every day [201]; of which mismatches arising as a

replication error represent an important lesion. The DNA mismatch repair (MMR) process is therefore crucial for maintaining the stability of both prokaryotic and eukaryotic genomes by correcting the replication errors that might have escaped proof-reading by the replication complex. A replication error can result either from the misincorporation of a base or from strand slippage at repetitive sequences during the DNA replication process. If left undetected or unrepaired, such errors can lead to mutator phenotype, microsatellite instabilities or genetic defects [202]. The mismatch repair pathway is highly conserved from bacteria to humans [203].

The first step in the pathway is the recognition of the lesion which is mediated by a group of MutS homolog MSH proteins. In humans, atleast five MutS homologues (MSH2, MSH6, MSH3, MSH4 and MSH5) have been identified which acts as heterodimers in different combinations [204]. MSH2-MSH6 (MutSα) represents 80-90% of the cellular level of MSH2 and recognizes single base mismatches and small insertions/deletions [205]. It is believed that MSH6 is the subunit responsible for recognizing the mismatch [206]. MSH2-MSH3 (MutSβ) is the second major functional complexes in the cell which recognizes longer insertion/deletions up to 15 nucleotides respectively [207]. The two complexes are only partially redundant as they have different substrate specificities. For example, MSH6 (and therefore MutSα) has a higher affinity for binding GT mismatches because of which MSH6 was earlier also known as GTBP (GT binding protein) [208]. The identification of the newly synthesizing daughter strand harboring the mismatched base in bacteria is facilitated by the presence of nick due to unmethylation [209]. Similar mechanism that can direct the recognition of mismatches has not been defined in eukaryotes. However, it has been speculated that DNA termini that occur as natural intermediates during the replication (3' terminus on the leading strand; 3' and 5' termini on the Okazaki fragments on lagging strand) may suffice as strand signals to direct the correction of replication errors in eukaryotic cells [210]. Binding of the MutS complexes to the mismatch induces an ATP driven conformational change that transforms this complex into a sliding clamp which then recruits the MutL complex [210].

There are three MutL homolog complexes in humans: MutLα (MLH1-PMS2), MutLβ (MLH1-PMS1) and MutLγ (MLH1-MLH3). The most active complex is MutLα which participates in the repair of single-base mismatches and IDLs (initiated by MutSα). MutLγ mainly contribute to IDL repair whereas MutLβ does not seem to

participate in MMR [210]. Taking advantage of the sliding clamp ability of MutS complex, the MutS-MutL complex then translocates in either direction in search of a strand discontinuity. After the recognition of discontinuity, MLH1 that has an inherent endonuclease activity produces a nick 3' to the mismatched base, which acts as entry point for the excision and re-synthesis machinery [210].

The excision and re-synthesis machinery comprises of Exonuclease 1 (Exo1), Proliferating cell nuclear antigen (PCNA), RPA (replication protein A), RFC (replication factor C), DNA polymerases (Polδ and Polε) and DNA ligase I. Exo1 is a 5'-3' exonuclease that subsequently degrades several hundred nucleotides starting from the nick situated 5' from the mismatched base and travelling towards the mismatched base. The activity of Exo1 is stimulated by MutS complex, only after which the excision starts [211]. After excision, the single strand generated is stabilized by binding of the RPA which prevents further degradation by Exo1. The strand re-synthesis involves filling of gaps which is mediated mainly by Polδ in association with its cofactors PCNA and RFC; followed by ligation by DNA ligase I [211]. PCNA has an important role in MMR pathway as it helps in stabilizing the binding of MutS complex to the mismatched DNA by interacting with the MSH6 protein. Also, PCNA functions as a processivity factor for the DNA polymerase Polδ [211]. The process of eukaryotic MMR pathway is illustrated in Fig 4.3 and the details of MMR genes and protein are summarized in Table 4.6

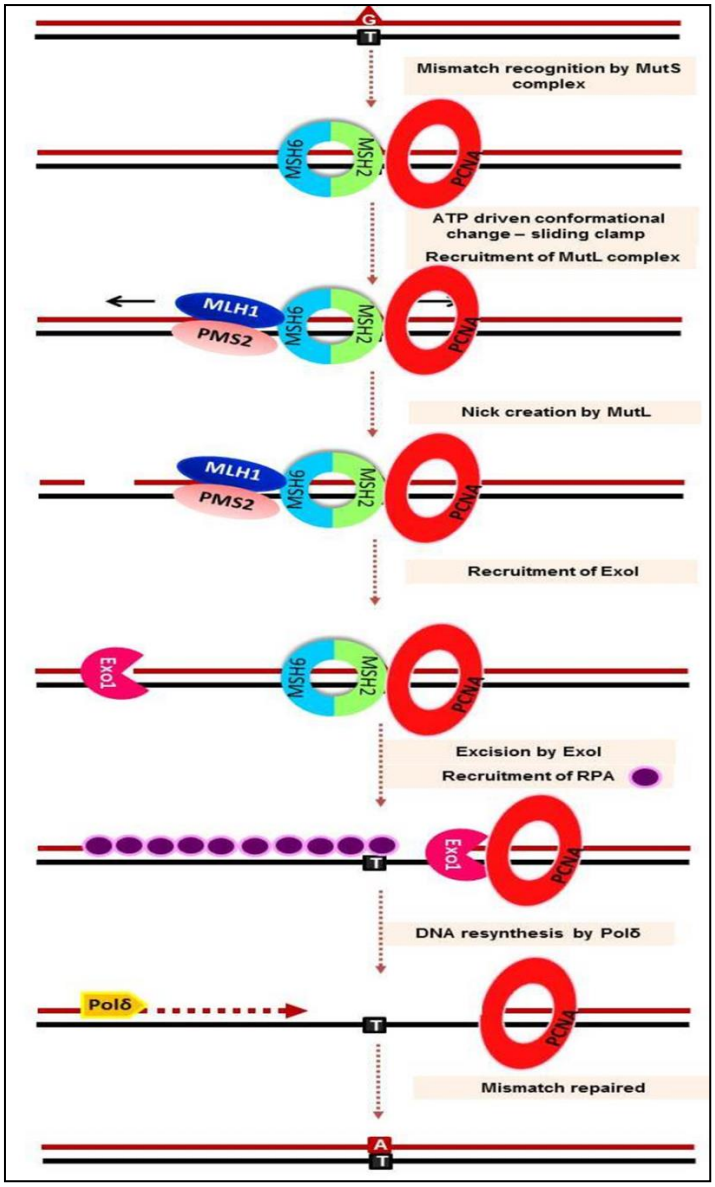


Fig 4.3 Mismatch repair in eukaryotes

Table 4.6 MMR genes and proteins

Gene/ Protein	Chromosomal location	No. of exons	Protein length
MLH1	**3p21**	**19**	**756 aa**
MSH2	**2p22**	**16**	**934 aa**
MSH6	**2p16**	**10**	**1360 aa**
PMS2	**7p22.1**	**11**	**863 aa**

4.6 MUTATION SPECTRUM OF MMR GENES IN LYNCH SYNDROME

The rate and nature of mutations detected in LS have reflected technological advances over the years. Exon by exon sequencing (including the exon –intron boundaries) have identified majority of the mutations over several years. Occasional sequence analysis of cDNA led to the detection of large exonic deletions within MLH1 and MSH2 [212]. A technique of conversion analysis in which the patient's diploid chromosomes are converted into haploid alleles and analyzed separately enabled the researchers to identify certain previously unidentified cryptic point mutations or large genomic rearrangements [213]. The discovery of MLPA technique facilitated identification of LGRs more efficiently and therefore led to increase in reporting of LGRs in LS families thereafter [214]. PMS2 mutation analysis which was earlier complicated due to the presence of pseudogenes was simplified by the use of Long range PCR, cDNA sequencing and MLPA [215-216]. Currently, many groups are undertaking mutation analysis in LS families using the targeted resequencing (gene panel testing) approach which has significantly increased the speed of genetic testing and detection rate in LS [217]. The mutation detection rate in LS families varies considerably from 40% to 88% depending upon the clinical criteria and pre-screening techniques employed for the diagnosis of Lynch syndrome [218-220]. Point mutations and small insertions and deletions are the predominant type of mutations identified. However, large genomic rearrangements too represent a significant proportion of all pathogenic mutations in mismatch repair genes in patients with Lynch syndrome. Many studies have demonstrated that genomic rearrangements represent 15% to 55% of all mutations in mismatch repair genes [221-222].

In 2004, a database of all LS associated mutations was established, which is curated and continuously updated by the International Society for Gastrointestinal and Hereditary Tumours (InSiGHT) [223]. While other databases for MMR variants exist (Universal Mutation Database and Woods MMR database), InSiGHT is the most comprehensive locus specific database which assembles data for almost every gene known to be associated with Lynch syndrome and other colorectal predisposition syndromes. As per the current data, MLH1, MSH2, MSH6 and PMS2 account for 40, 34, 18 and 8 %, respectively [224]. 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 [225].

Mutations are scattered throughout the MMR genes without any obvious hotspots; though there are a few exceptions. While the majority of mutations in MMR genes are truncating mutations (nonsense and frameshift); missense mutations also contribute significantly for all the 4 MMR proteins, MLH1, MSH2, MSH6 and PMS2 [226]. This high proportion of missense mutations prompted the InSiGHT to undertake massive scale study of MMR variants to infer the pathogenicity of all the variants using a 5-tiered classification system. The Class 5 and Class 4 variants are the "*pathogenic*" and "*likely pathogenic*" variants. Class 3 represents "*variants of unknown significance (VUS)"* whereas the Class 2 and Class 1 signify the "*likely benign"* and "*benign"* variants [224]. Pathogenic mutations (Class 5 and Class 4 mutations) constitute majority of variants reported in the InSiGHT database for all the MMR genes (except MSH6 where the proportion of VUS is significantly higher) while the benign variants (Class 2 and Class 1) account for smaller proportion of database variants. The percentage of VUS is 31% of all reported variants for MLH1, 28% for MSH2, 47% in MSH6 and 26% in PMS2 genes [224].

A mutation that arise *de novo* with high frequency is defined as recurrent. Mutations that occur once and are then passed on to succeeding generations are designated founder mutations and are typically limited to a certain geographic area or a certain ethnic group. Most germline mutations reported in the MMR genes are unique. However, recurrent and founder mutations in MMR genes have also been reported in many populations worldwide [227-228]. Furthermore, MMR gene mutations are inherited from either parent in most of the cases since LS is characterized by strong family history of cancers; whereas the rate of *de novo* mutations in MMR genes is very less (nearly 2.3%) [229].

Despite the improvements in genetic screening technologies, germ line mutations of the MMR genes remained undetected in a significant proportion of families with a clinical suspicion of LS. Constitutional epimutation of MLH1, which is the characterized by monoallelic methylation and transcriptional loss of expression throughout normal somatic tissues, represent one of the alternative cause for LS in

MMR gene mutation-negative case [230-231]. Epimutations of MLH1have been shown to be heritable but with distinct Mendelian and non-Mendelian patterns depending on the underlying mechanism [232]. The prevalence of MLH1 constitutional epimutations in colorectal cancers lacking MLH1 expression was reported to be 0 % among unselected cases and 16 % among cases fulfilling the revised Bethesda criteria, suggesting that testing for MLH1 epimutations should regularly be restricted to the latter group of patients [233]. Also, apart from the germline mutations in the canonical genes, deletion in EPCAM (also known as TACSTD1) gene that leads to transcriptional silencing of MSH2 gene (MSH2 epimutation) has been reported in LS [234]. EPCAM deletion-associated MSH2 epimutations vary a lot in frequency between populations depending on possible founder effects and may account for 10–40 % of families with absent MSH2 protein in tumors [235]. Such epimutations show regular Mendelian transmission along with EPCAM deletion in pedigrees [234].

4.7 GENOTYPE-PHENOTYPE CORRELATIONS IN LYNCH SYNDROME

Germline mutation carriers of different MMR genes confer different risk of developing colorectal and extracolonic cancers [236]. The age of onset and the aggressiveness of disease also vary depending upon which MMR gene is mutated [236]. Among the various cancers arising in MSH2 and MLH1 mutation carriers, the highest lifetime risk is for colorectal cancer, followed by endometrial cancer and other extracolonic cancers. Moreover, MSH2 mutations may be associated with higher risks of extracolonic cancers compared to MLH1 mutations [237]. Female carriers of MSH6 mutations are at a higher risk of endometrial than colorectal cancer [238]. Furthermore, MSH6 and PMS2 mutations show reduced age specific penetrance, resulting in higher average ages at onset of various cancers in MSH6 and PMS2 carriers compared to MSH2 or MLH1 mutation carriers, although family or mutation-specific variations exist [239, 240]. Also, heterozygous PMS2 mutation carriers develop excess of polyps but have a low risk of developing cancers [241]. EPCAM deletions have also been associated with lower risk of extracolonic cancers; although if the deletion is close to MSH2 gene, it confers a high risk of developing endometrial cancer in female carriers [242-243]. The MLH1 epimutation seems to have same phenotypic features as compared to the MLH1 gene mutation carriers [244]. Homozygous MMR gene mutation carriers (particularly PMS2), are at an increased

risk of developing hematological, colorectal, urinary tract and brain (glioblastoma) cancers, and neurofibromatosis at a very early age (pediatric) as part of the Constitutive Mismatch Repair Deficiency (CMMRD) syndrome. No clear-cut correlations have been observed between the type (e.g., truncating vs. missense) or location (e.g., relative to different functional domains) of a MMR gene mutation and clinical phenotype [245]. An exception to this is a study that reports a higher proportion of colorectal cancer and an earlier onset of the disease in individuals with large rearrangements in MLH1 and MSH2 genes [246].

4.8 IMMUNOHISTOCHEMISTRY IN LS DIAGNOSIS

Immunohistochemistry is a method to study the expression of four MMR genes, MLH1, MSH2, MSH6 and PMS2 in tumor samples; which serves as an ideal prescreening technique that guides genetic testing in Lynch syndrome by uncovering the gene that is likely to contain germline mutation. IHC has higher sensitivity and specificity over STR based MSI analysis for the molecular screening of LS [247]. A functional defect in MLH1 results in the degradation of both MLH1 and PMS2, whereas a defect in PMS2 results only in the degradation of PMS2. This is because; MLH1 acts as heterodimer with PMS2 and helps in the stabilization of PMS2 protein. In the absence of MLH1, PMS2 is not stabilized and hence gets degraded as MLH1 is an indispensable component of the complex. Nevertheless, when there is a loss of expression of PMS2, MLH1 is bound and stabilized by MLH3 which shares functional redundancy with PMS2. Consequently, loss of expression of MLH1 and PMS2 in CRC generally indicates an alteration in MLH1, either by somatic methylation of the MLH1 promoter region (sporadic cases) or by an MLH1 germline mutation (Lynch syndrome), and solitary loss of PMS2 expression generally indicates an underlying germline defect in PMS2. Similar phenomenon is seen with IHC of MSH2 and MSH6 proteins. Loss of staining of both proteins indicates germline mutation in MSH2 whereas solitary loss of MSH6 has been associated with MSH6 mutations. A very striking observation regarding IHC of MMR proteins is the identification of a germline MLH1 mutation in few cases showing solitary loss of PMS2 protein expression [248]. In one case, this phenomenon was explained by the presence of epitope positive truncating mutations in carboxyl terminal of MLH1 gene that resulted in the retention of the major part of MLH1 protein but loss of domains

involved in PMS2 binding and stabilization; thereby leading to solitary loss of PMS2 [249].

IHC is an important technique in LS diagnosis as it also represents an integral part of the universal tumor screening approach for LS diagnosis. The concept of reflex testing or universal tumor testing for lynch syndrome diagnosis which involves the evaluation of all colon and/or endometrial tumors at the time of diagnosis for evidence of MMR deficiency (MSI and IHC analysis) is the substance of debate in the current literature regarding Lynch syndrome. While few studies strongly proposes its use demonstrating the diagnosis of a significant fraction of LS cases that would otherwise have been missed [250], other studies discourage its application by outweighing its advantages over the limitations and challenges associated with the implementation of this concept [251].

4.9 THE INDIAN SCENARIO ON LS AND OBJECTIVE OF THIS STUDY

Lynch syndrome being the most common form of inherited CRC predisposition syndromes, is extensively studied and a plethora of literature is available on the clinical and molecular aspects of this syndrome. Most of these studies come from Caucasian cohorts [252-253]. However the Asian, particularly the South Asian LS cohorts are not well described. So far only 3 studies have been reported, describing only 28 Indian LS families with MMR gene mutations [254-256]. Of these 3 studies, the major study on 48 cases published recently used a comprehensive approach (including MLPA) for genetic dissection of the LS cases which identified MMR germline mutations in 24 families [254]. The other two studies includes a case study on an extended Indian family with R659X mutation in MLH1 gene [255] and a report of MLH1 and MSH2 mutations identified in 3 families after preliminary screening for MSI in 31 individuals [256]. One of the objectives of this project was therefore to characterize the spectrum of MMR gene mutations and study the phenotypic features in a cohort of 81 Lynch syndrome families from India.

The families were registered in the Cancer Genetics Clinic after making syndromic diagnosis of Lynch syndrome based on the clinical criteria like Amsterdam Criteria and Revised Bethesda Guidelines and other clinical features as per the medical records available. Majority of the cases in this study was preselected based on their IHC results that were retrieved from the pathology department of Tata Memorial

Hospital where IHC of all the four MMR proteins (MLH1, MSH2, MSH6, PMS2) were done on tumor samples of patients. 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 PCR amplify the coding and flanking intronic regions of the MMR gene as guided by IHC, in case of unavailability of IHC results, MLH1 gene was first studied followed by MSH2 and MSH6. PCR products were subjected to Sanger sequencing in order to identify the germline mutation. PMS2 analysis was not carried out in this series of LS cases.

4.10 RESULTS

This study was conducted on 205 individuals from 81 unrelated Indian Lynch Syndrome families. Of these 81 families, 44 followed the Amsterdam criteria of LS while the other 37 were diagnosed using Revised Bethesda Guidelines. This cohort represents families from diverse geographical regions of Indian subcontinent (Fig. 4.4) and belonging to Hindu, Muslim, Christian, Sikh, Jain and Sindhi religions.

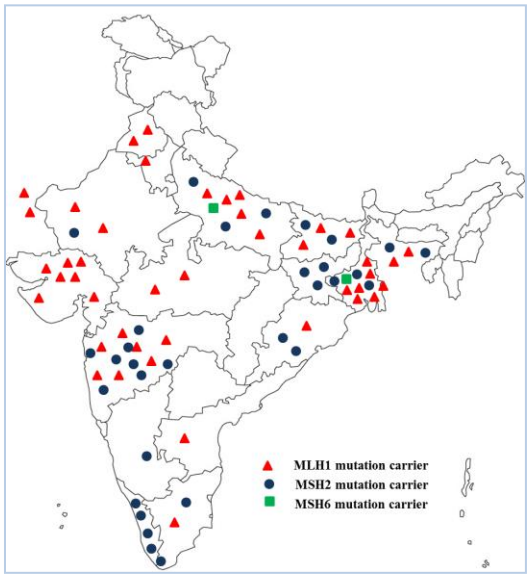


Fig 4.4 Geo-ethnic origins of MMR mutation carriers

From the 81 probands who received genetic counseling, 47 underwent MMR protein analysis by IHC while the other 34 were taken up for direct MMR germline genetic testing. All but 6 probands investigated in this study presented with family history of colorectal or other LS associated cancers in multiple members of family (representative pedigree in Fig 4.5). The 6 probands with no family history were

included in the study because of MMR deficiency in their tumors as detected by IHC and clinico-pathological features which raised the suspicion of LS due to de novo mutation in MMR gene (representative pedigree 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

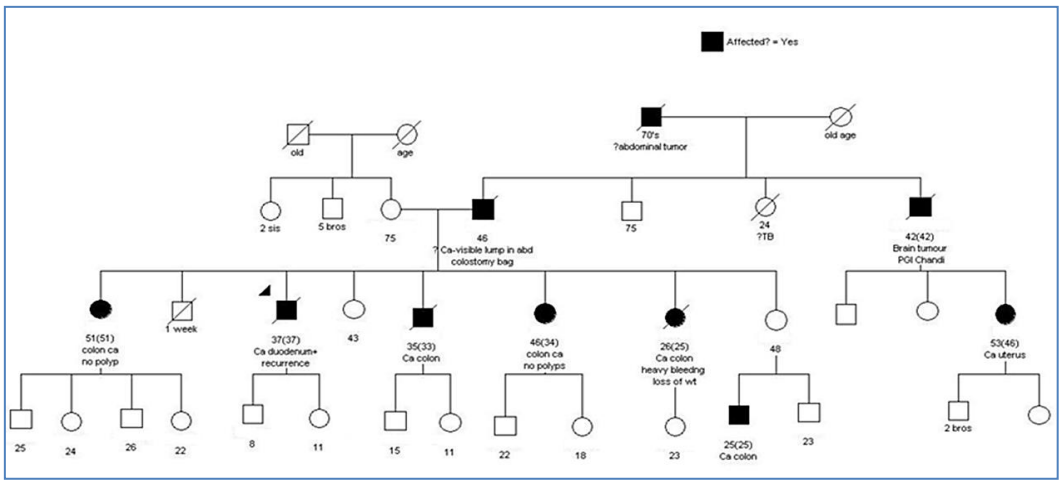


Fig 4.5 Classical Lynch syndrome family

This is a Hindu Baniya family from Punjab. A total of 10 members across 4 generations were affected with cancers of colon, duodenum, endometrium and brain between the ages of 25-70 years. Deleterious germline mutation in MLH1 gene was identified in the proband followed by testing of 10 at-risk individuals for the family specific mutation. Out of these 10, five were tested negative for the mutation while the other five were positive for FSM, of which 4 were unaffected so far and have been kept under surveillance.

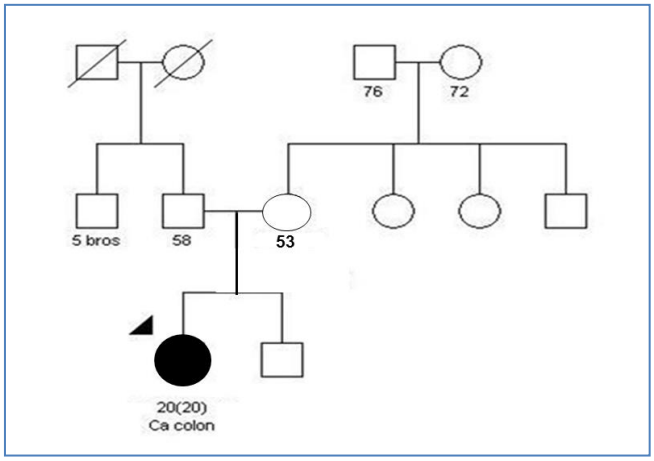


Fig 4.6 Suspected *de novo* Lynch Syndrome case

The proband is a 20 year old Hindu Marwadi female from Rajastahan who was diagnosed with poorly differentiated adenocarcinoma of transverse colon which showed signet ring cell morphology on histopathological analysis. The tumor also showed loss of expression of MSH2 and MSH6 proteins on IHC. She reported no history of cancers in her family but the early age of onset and the clinico-pathological features of tumors were suggestive of Lynch syndrome and therefore genetic testing was undertaken in this case which identified a deleterious nonsense mutation in MSH2 gene (*de novo* mutation).

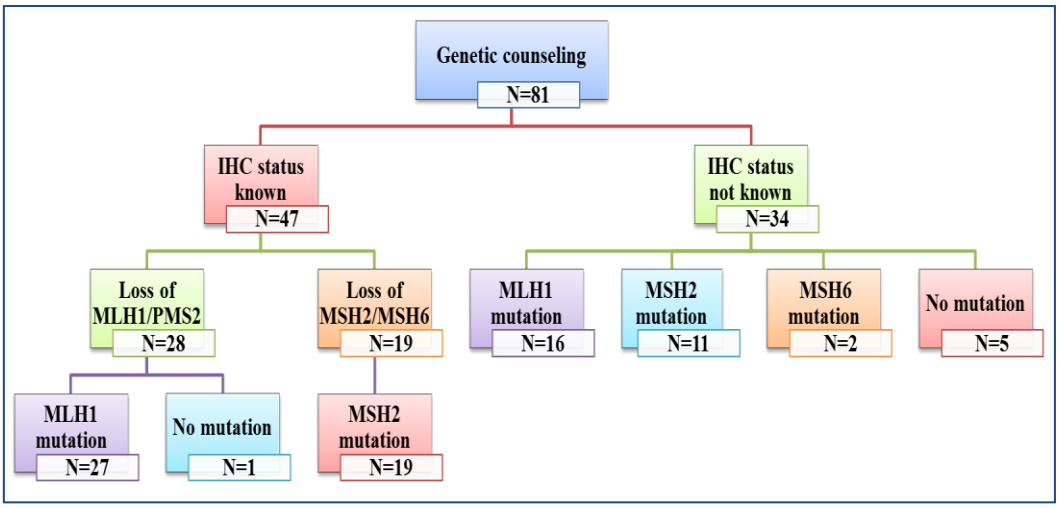


Fig 4.7 Summary of the genetic analysis approach and results in Indian LS patients

4.10.1 Mutation spectrum of MLH1 and MSH2 genes

Genetic analysis of MLH1, MSH2 and MSH6 genes in Indian LS families identified deleterious germline mutations in 75 out of 81 cases leading to a mutation detection rate of 93% in our cohort. MLH1 gene harbors 32 unique mutations in 43 families while 23 distinct mutations were identified in MSH2 gene in 29 LS families. Two MSH6 gene mutations were identified in two families while an EPCAM deletion was identified in one. Different types of mutation were prevalent in the two predominantly mutated genes, MLH1 and MSH2. Missense and splice site mutations predominate in the MLH1gene, while there is a preponderance of LGRs in the MSH2 gene (Fig 4.8).

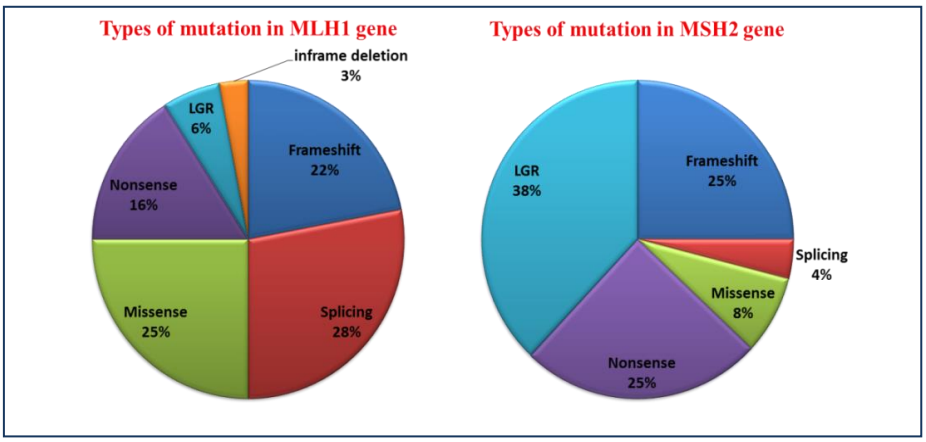


Figure 4.8 Frequency of different types of mutations in MLH1 and MSH2 genes

A total of 22 novel mutations were identified, of which 14 are MLH1, 7 MSH2 and 1 MSH6 gene mutation. The mutations in MLH1 gene is evenly scattered throughout

the gene with majority of the mutations being unique (Fig 4.9). Four mutations in MLH1 gene; c.46insG (Exon 1), c.156delA (Exon 2), c.199G>A (Exon 2) and c.1558+2insG (Intron 13) were found in more than family indicating the recurrent nature of these mutations. The MSH2 mutation spectrum shows the clustering of mutations (including LGRs) in the proximal part of the gene which encodes the DNA binding domain and MSH6/MSH3 interaction domain (Fig 4.10). All the mutations in MSH2 gene were identified in one family each; except two mutations, c.942+3A>T (Intron 5) and c.340G>T (Exon 2) that were identified in 6 and 2 families respectively. The details of MLH1 and MSH2 mutations are given in Table 4.7 and Table 4.8.

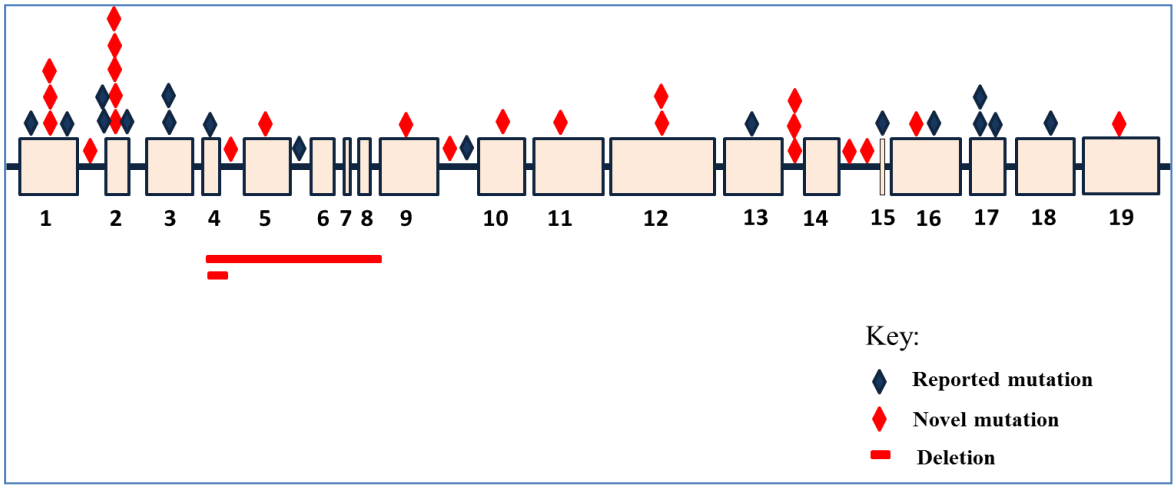


Fig 4.9 Exon-wise distribution of mutations in MLH1 gene in our cohort

The boxes represent the exons labelled by the numbers 1-19. Intervals between the boxes represent the intervening sequences (introns). The representation is in a non-proportional scale. Individual mutations are represented by a single diamond.

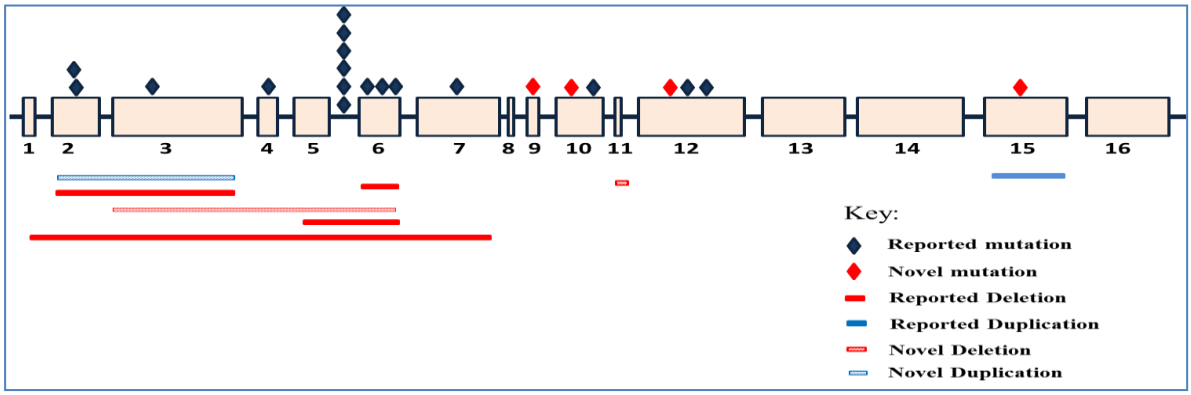


Fig 4.10 Exon-wise distribution of mutations in MSH2 gene in our cohort

The boxes represent the exons labelled by the numbers 1-16. Intervals between the boxes represent the intervening sequences (introns). The representation is in a non-proportional scale. Individual mutations are represented by a single diamond.

Table 4.7 MLH1 germline mutations in Indian LS patients

Sr. No.	Nucleotide change	Exon	Consequence	No. of families	Type of Mutation	Reported in InSiGHT database
1	c.2T>G	Exon 1	M1R	1	Missense	Reported 3 times
2	c.46insG	Exon 1	p.V16GfsX13	3	Frameshift	**Novel**
3	c.116G>T	Exon 1	C39F	1	Missense	Reported 1 time
4	IVS1-1G>C	Intron 1	-	1	Splice site	**Novel**
5	c.155_158delAAGA	Exon 2	p.L52Rfs*4	1	Frameshift	Reported 2 times
6	c.156delA	Exon 2	p.E53Rfs*4	5	Frameshift	**Novel**
7	c.199G>A	Exon 2	G67R	2	Missense	Reported 85 times
8	c.306G>T	Exon 3	p.E102D	2	Missense	Reported 14 times
9	c.350C>T	Exon 4	T117M	1	Missense	Reported 113 times
10	c.412delT	Exon 5	p.P139LfsX20	1	Frameshift	**Novel**
11	c.380+1G>T	Intron 4	-	1	Splice site	**Novel**
12	c.454-13A>G	Intron 5	-	1	Splice site	Reported 2 times
13	IVS9+1G>A	Intron 9	-	1	Splice site	Reported 40 times
14	IVS9+2T>C	Intron 9	-	1	Splice site	Reported 1 time
15	c.776T>G	Exon 9	L259X	1	Nonsense	**Novel**
16	c.879C>G	Exon 10	Y293X	1	Nonsense	**Novel**
17	c.991G>T	Exon 11	E331X	1	Nonsense	**Novel**
18	c.1389_1390delAC	Exon 12	p.P464Yfs*14	2	Frameshift	**Novel**
19	c.2041G>A	Exon 13	A681T	1	Missense	Reported 77 times
20	c.1731G>A	Exon 15	S576S	1	Missense	Reported 55 times
21	c.1838_1839delAG	Exon 16	p.E613Vfs*6	1	Frameshift	**Novel**
22	c.1852_1854delAAG	Exon 16	p.K618del	1	In-frame deletion	Reported 132 times
23	c.1790G>A	Exon 17	W597X	1	Nonsense	Reported 5 times
24	c.1975C>T	Exon 17	R659X	1	Nonsense	Reported 28 times
25	c.1976G>C	Exon 17	R659P	1	Missense	Reported 51 times
26	c.2059C>T	Exon 18	R687W	1	Missense	Reported 43 times
27	c.2258_2259dupTT	Exon 19	p.E754Lfs	1	Frameshift	**Novel**
28	c.1558+2insG	Intron 13	-	3	Splice site	**Novel**
29	IVS14-1G>C	Intron 14	-	1	Splice site	**Novel**
30	IVS14-2A>C	Intron 14	-	1	Splice site	**Novel**
31	Deletion of Exon 4	-	-	1	LGR	Reported 1 time
32	Deletion of Exon 4-8	-	-	1	LGR	Reported 1 time

Table 4.8 MSH2 gene mutations in Indian LS patients

Sr. No.	Nucleotide change	Exon	Consequence	No. of families	Type of Mutation	Reported in InSiGHT database
1	c.340G>T	Exon 2	p.E114X	2	Nonsense	Reported 1 time
2	c.478 C>T	Exon 3	Q160X	1	Nonsense	Reported 1 time
3	c.652C>T	Exon 4	Q218X	1	Nonsense	Reported 2 times
4	c.1009C>T	Exon 6	p.Q337X	1	Nonsense	Reported 5 times
5	c.1046C>T	Exon 6	p.P349L	1	Missense	Reported 1 time
6	c.1017_1018delAA	Exon 6	p.R340TfsX3	1	Frameshift	Reported 3 times
7	c.1147C>T	Exon 7	R383X	1	Nonsense	Reported 36 times
8	c.1490insA	Exon 9	p.I497Nfs*15	1	Frameshift	**Novel**
9	c.942+3A>T	Intron 5	-	6	Splicing	Reported 168 times
10	c.1552_1553delCA	Exon 10	p.Q518Vfs*10	1	Frameshift	Reported 15 times
11	c.1538_1539delTG	Exon 10	p.L513Rfs*14	1	Frameshift	**Novel**
12	c.1807G>A	Exon 12	D603N	1	Missense	Reported 26 times
14	c.1861C>T	Exon 12	R621X	1	Nonsense	Reported 16 times
15	c.1801_1805delCAGCT	Exon 12	p.Q601RfsX39	1	Frameshift	**Novel**
16	c.2504dupA	Exon 15	p.N835Kfs*2	1	Frameshift	**Novel**
17	Duplication of Ex 2-3	-	-	1	LGR	**Novel**
18	Deletion of Ex 5-6	-	-	1	LGR	Reported 5 times
19	Deletion of Ex 2-3	-	-	1	LGR	Reported 2 times
20	Deletion of Ex 11	-	-	1	LGR	**Novel**
21	Deletion of Exon 1-7	-	-	1	LGR	Reported 3 times
22	Deletion of Ex 3-6	-	-	1	LGR	**Novel**
23	Deletion of Ex 6	-	-	1	LGR	Reported 15 times
24	Duplication of Exon 15	-	-	1	LGR	Reported 1 time

4.10.2 Variants of unknown significance in MLH1 gene

We detected three variants in MLH1 gene in three families whose pathogenicity could not be established unequivocally after identification. This includes a novel mutation c.2258_2259dupTT in Exon 19, c.2T>G in Exon 1 and an intronic variant c.454-13A>G in Intron 5, reported three times and two times respectively in the InSiGHT database as Class 3 mutations. This was due to lack of sufficient evidence to conclude the pathogenicity of these mutations. We sought to understand the effect of these variants by studying the clinico-pathological features associated with them and characterizing the variant by transcript analysis in case of the intronic variant.

- ***c.454-13A>G (Intron 5 of MLH1)***

This variant was identified in a Gujarati Hindu female who was diagnosed with endometrial cancer at the age of 38 years. She had a family history of cancers with four first and second degree relatives affected with cancers of colon, prostate and brain between 36 to 70 years (pedigree in Fig 4.11). Two of the family members had two primary cancers.

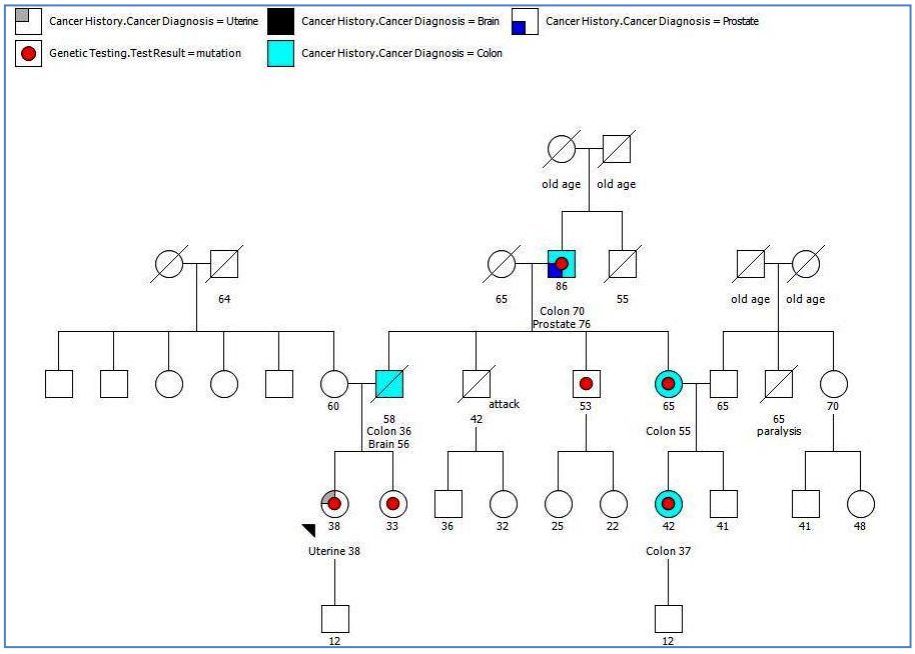


Fig 4.11 Pedigree of family with c.454-13A>G variant in MLH1

Genetic analysis of MMR genes identified c.454-13A>G variant in Intron 5 of MLH1 gene (fig 4.12). No other variant was identified in other genes in this case.

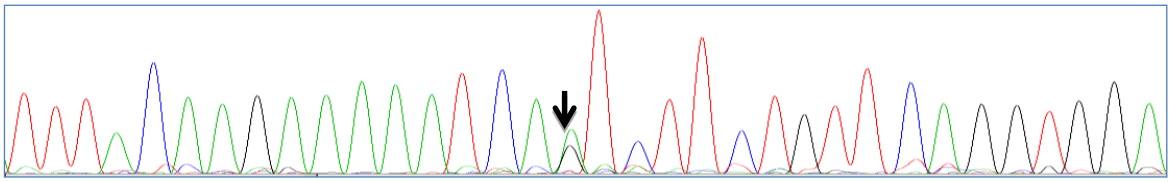


Fig 4.12 Chromatogram showing c.454-13A>G mutation in Intron 5 of MLH1 gene

To establish the pathogenicity of this variant, co-segregation study was undertaken in this family. Of the 9 members tested for FSM, 5 were positive for this mutation. The variant was found to segregate with the disease as three of the 5 mutation positive members were affected with LS associated cancers; the other two mutation carriers

were unaffected at the age of 53 and 40 years. MSI analysis was carried out on the colon cancer tumor sample of probands' paternal cousin using Bethesda panel markers which revealed the presence of MSI-H phenotype (instability at 3 out of 5 markers) in the tumor (Fig 4.13).

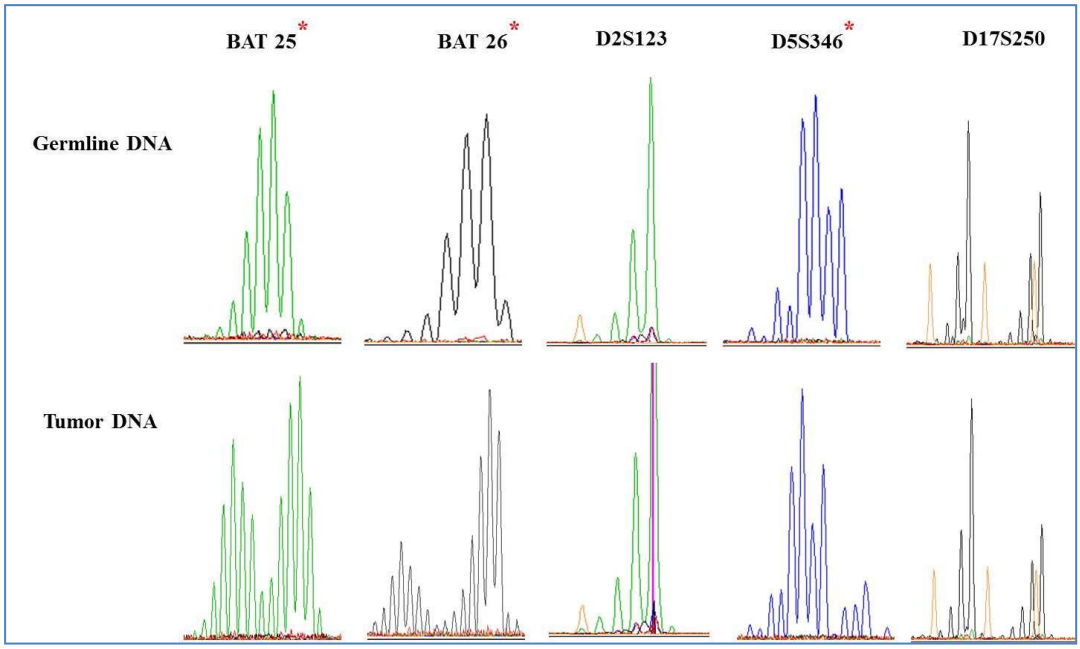


Fig 4.13 MSI-H in tumor of c.454-13A>G carrier
(* marker showing instability)

Further, RNA level studies were done to study the effect of this variant on the splicing of MLH1 gene. The methodology was to generate cDNA from RNA extracted from the lymphocytes of c.454-13A>G mutation carrier and a healthy control sample. The region between Exon 3 and Exon 8 of MLH1 was PCR amplified using cDNA specific primers. On agarose gel electrophoresis, two bands were observed as against one band observed for healthy control samples (Fig 4.14). Sequencing of the band corresponding to aberrant transcript revealed the skipping of Exon 6 from that transcript (Fig 4.15).

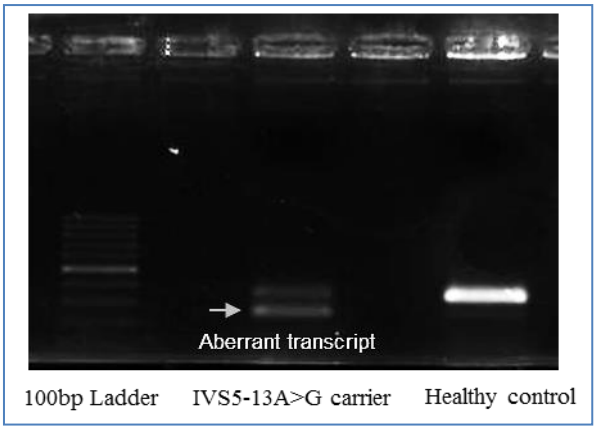


Fig 4.14 2% agarose gel electrophoresis of amplified cDNA

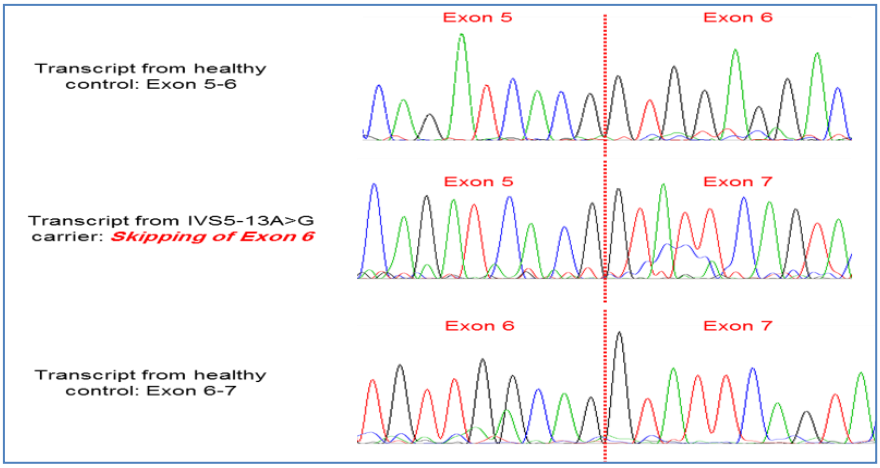


Fig 4.15 Chromatograms showing skipping of Exon 6 in c.454-13A>G carrier

The results of co-segregation studies, MSI analysis and transcript analysis confirms the pathogenicity of this intronic variant which we suggest should be reclassified as a Class 5 mutation.

- ***c.2T>G (Exon 1 of MLH1)***

This variant was identified in a 60 years old Rajasthani Jat male who was diagnosed with synchronous cancer of ascending and sigmoid colon at the age of 53 years. IHC analysis showed loss of expression of MLH1 and PMS2 proteins in his tumor samples. He reported a history of colon cancers in two of his brothers at the age of 41 years and 51 years (who had recurrence at 57 years). His maternal aunt was also diagnosed with endometrial cancer at the age of 64 years (pedigree Fig 4.16).

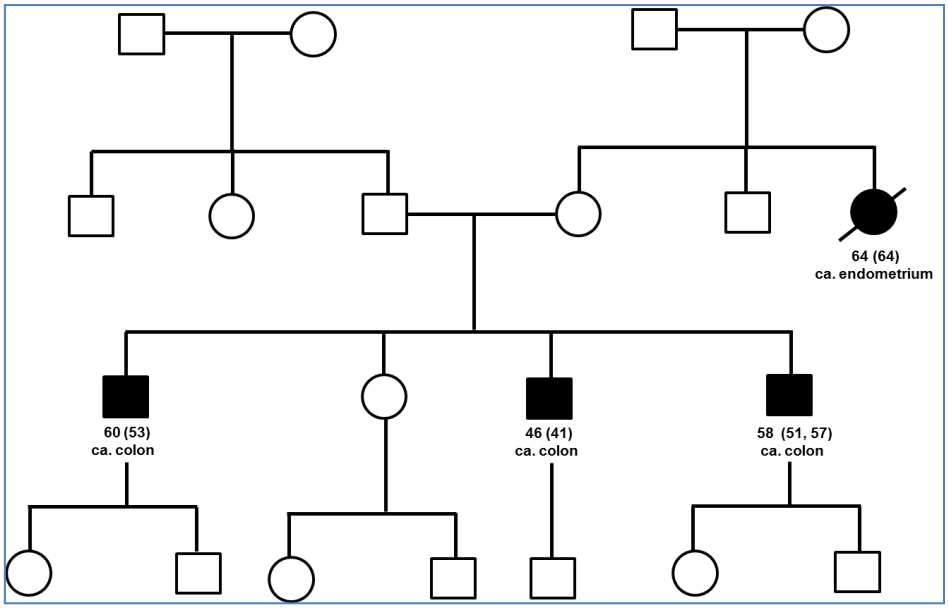


Fig 4.16 Pedigree of family with c.2T>G mutation

MLH1 gene mutation analysis revealed the presence of c.2T>G variant which results in the start codon (M1R) mutation that may have an effect on the initiation of MLH1 protein translation. This same variant was identified in his brother affected with colon cancer at 51 years and had a recurrence at 57 years. However, no other samples were available from this family for co-segregation studies and no further functional studies could be carried out. Nevertheless, taking into consideration the IHC results, associated clinical features, presence of this variant in one affected relative and the absence of other deleterious germline mutation, we conclude that this variant must be reclassified as a "likely pathogenic" mutation (Class 4).

- ***c.2258_2259dupTT (Exon 19 of MLH1)***

This variant was identified in a Maharashtrian female hailing from Malvan district; who was diagnosed with moderately differentiated adenocarcinoma of colon with mucinous histology at the age of 59 years. She had a family history of colon cancers in 2 sisters and brother between ages 52-57 years. Also, one of her maternal aunt was affected with endometrial cancer at the age of 65 years (pedigree in Fig. 4.17).

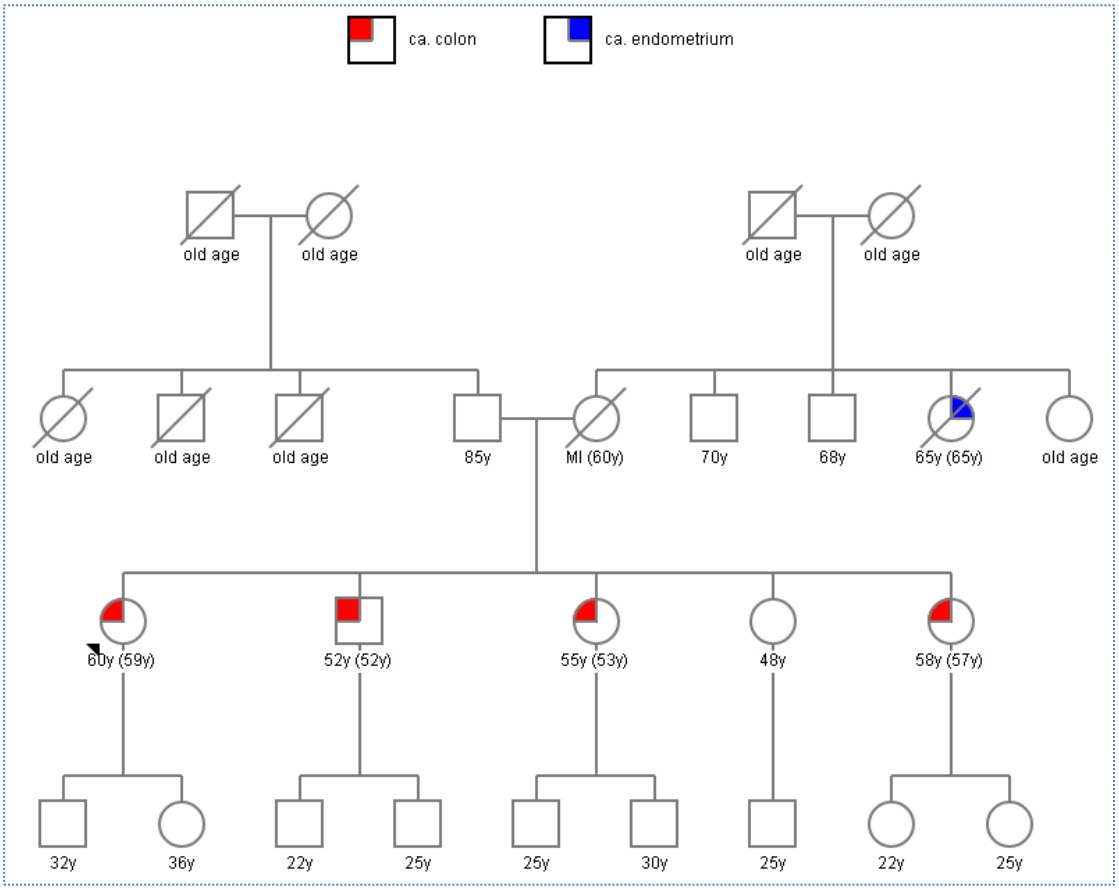


Fig 4.17 Pedigree of family with c.2258_2259dupTT mutation

This is the only variant identified in the family after screening of all MMR genes. This variant was also identified in her brother who was also affected with colon cancer at the age of 52 years. It is a novel mutation that is predicted to cause a frameshift that leads to the loss of stop codon thus resulting in the addition of the extra 26 amino acids in the translated protein. This prediction was made by using in silico analysis tool, ExPASy-Translate hosted by Swiss Institute of Bioinformatics (SIB). The result of ExPASy is given in Fig 4.18

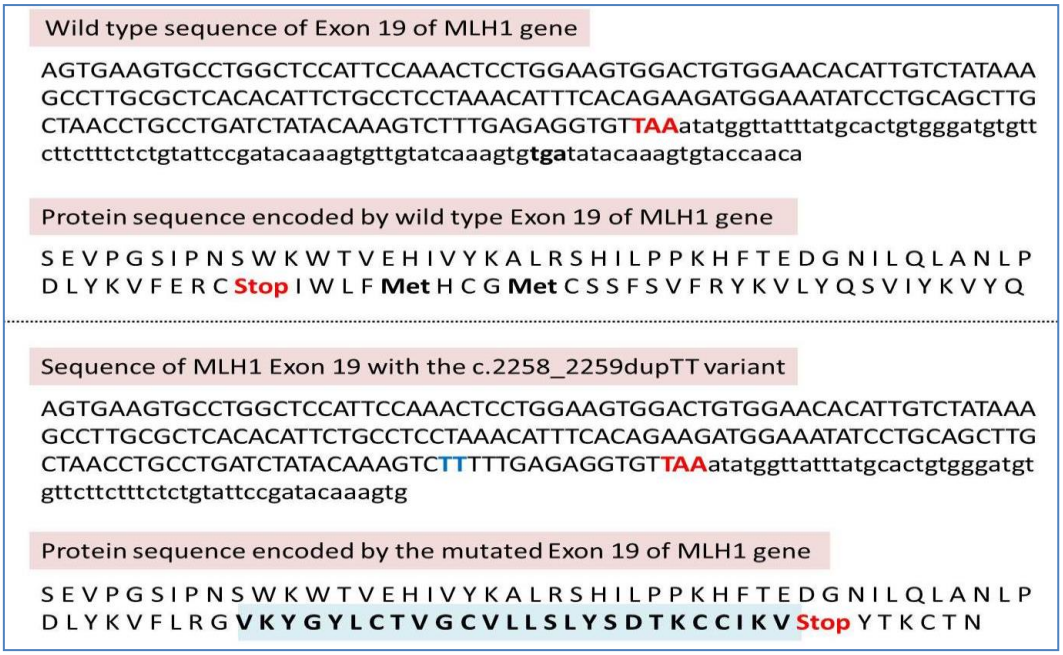


Figure 4.18 ExPASy translate analysis of c.2258_2259dupTT variant in Exon 19 of MLH1 gene

TAA in the wild type sequence represent the stop codon shown as **Stop** in the wild type protein sequence. The nucleotides after codon TAA, in the small alphabets represent the 3' UTR region of MLH1 gene. In the mutant sequence, **TT** represents the c.2258_2259dupTT variant which leads to the addition of extra 26 amino acids highlighted by blue box in the mutant protein sequence.

This mutation leads to the formation of an elongated protein of 782 amino acids instead of the wild type protein of 756 amino acids. This may disrupt the C-terminus of the MLH1 protein which is involved in interaction with PMS2/PMS1/MLH3. Further evidence regarding the possibility of this variant being pathogenic comes from two studies who report similar findings. In one study, a frameshift mutation was identified in Exon 19 of MLH1 gene in a Portuguese family that was also predicted to lead to an elongated protein of 782 amino acids like our variant. The authors concluded this variant to be pathogenic based on the fact that MLH1 C-terminus is highly conserved among species. Another study reported a 4-bp insertion in MLH1 Exon 19 that leads to elongation of MLH1 protein by 34 amino acids. This was a functional study in which the author showed that the elongated protein was unable to bind to PMS2 protein thus interfering with the mismatch repair function of MLH1. In light of the above evidences, we infer this variant c.2258_2259dupTT to be pathogenic.

4.10.3 Phenotypic characterization of MMR mutation carriers

The study included 152 proven MMR mutation carriers from 75 families, of which 28.75% were cancer free. This included 91 MLH1, 56 MSH2 and 5 MSH6 gene mutation carriers. In addition, after taking into account the first and second degree cancer affected relatives from these 75 families who were not tested for the family specific mutation but in whom a mutation positive status was assumed, the phenotype data was available for 347 affected members in which a total of 379 cancers were recorded.

Tumor frequencies and spectrum

Among 379 tumors, CRC was the most common malignancy, accounting for 73% versus 59% of tumors in MLH1 as opposed to MSH2 mutation carrier families (P=0.0053) (Table 4.9). Proportion of CRC in females (34%) was lower than in males (66%) and this trend was similar in both MLH1 and MSH2 mutation families. MLH1 families showed higher proportion of synchronous and metachronous CRCs (7.5%) as compared to MSH2 mutation families (3.5%), though this difference was not statistically significant.

The mean tumor burden per family was nearly the same in both MLH1 and MSH2 mutation positive families (Fig 4.19) with MLH1 mutation carrier families showing 2-18 tumors per family (mean – 5.91) whereas the MSH2 carrier families showed 1-32 tumors per family (mean – 5.33). The number of cancers per person ranged between 1 and 4, with nearly 87% of the individuals being diagnosed with only 1 cancer.

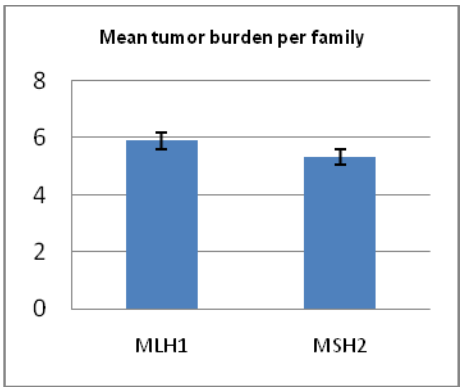


Fig 4.19 Mean tumor burden per family

Table 4.9 Distribution of Tumor Occurrences in MLH1 and MSH2 Mutation Families

Tumor localization	MLH1 (n=228)	MSH2 (n=151)
Colorectum	72.82%	58.94%
Endometrium	7.48%	15.27%
Stomach	5.72%	3.31%
Breast	1.76%	5.29%
Ovary	0.87%	0.66%
Small Intestine	1.75%	0.66%
Skin	1.31%	1.98%
Hepato-biliary tract[#]	2.19%	0.66%
Lung	0.88%	1.32%
Pancreas	0.44%	1.99%
Urinary tract[*]	0.88%	6.62%
Brain & CNS	2.60%	1.32%
Bone	1.30%	1.32%
Cervix	-	0.66%
Total	**100%**	**100%**

Include cancers of liver, bile duct, ampulla and gall bladder

* Include cancers of ureter, bladder, kidney and prostate

Extracolonic cancers were more common in MSH2 (41%) as compared to the MLH1 (27%) mutation positive families (p-value=0.053). Endometrial cancer was the most frequent extracolonic cancer in both MLH1 and MSH2 carrier families. Apart from these; breast, stomach, hepato-biliary tract, urinary tract, brain, bone and skin cancers were also reported in both MLH1 and MSH2 carrier mutation families. As seen in Fig. 4.20, endometrial cancers, breast cancers, skin cancers and cancers of the urinary tract were more common in MSH2 positive families while stomach and brain & CNS tumors were more common in MLH1 mutation carrier families (Table 4.9).

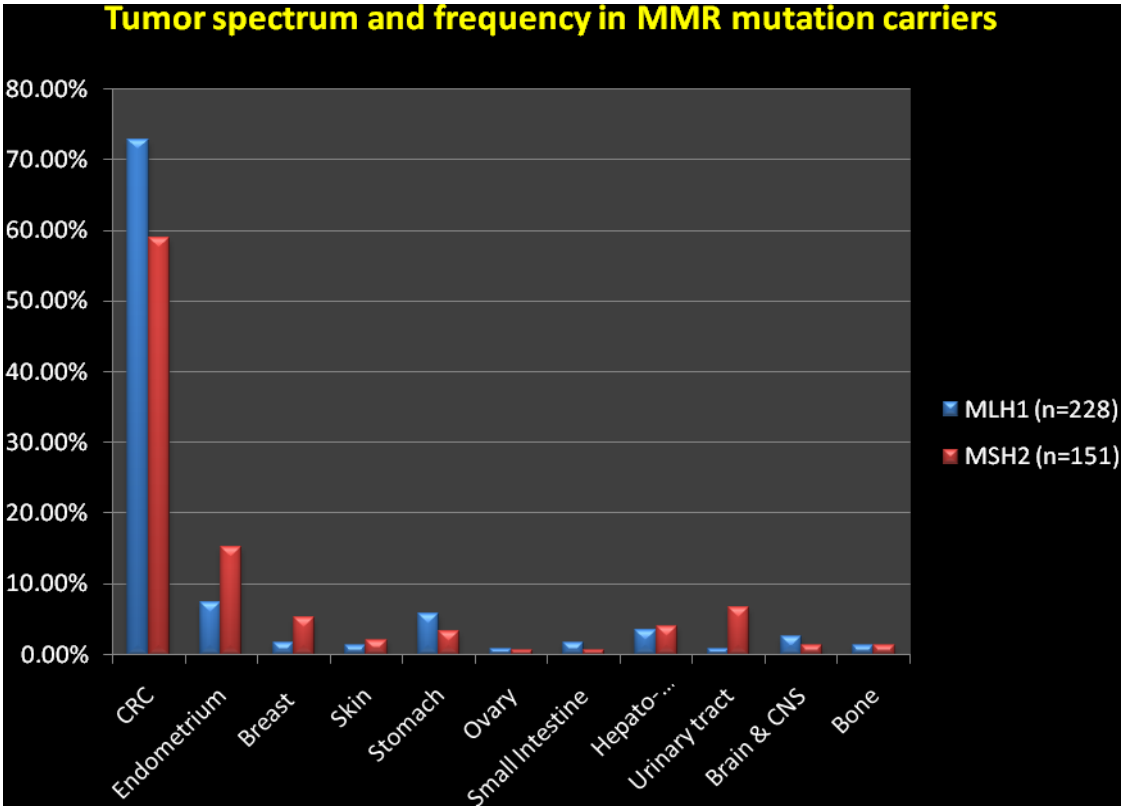


Fig 4.20 Tumor spectrum in MLH1 and MSH2 carrier families

Breast cancers were seen in 12 females from both MLH1 (n=4) and MSH2 (n=8) mutation carrier families. Prostate cancer was seen in 3 males (MLH1 – 2; MSH2 – 1).

Age at diagnosis

The mean age at first diagnosis (any cancer) was 47.4 years (Range: 8y-84y) for MLH1 and 49.3 years (Range: 22y-74y) for MSH2 mutation carriers. There was a significantly lower age of onset of CRC in males (mean age of diagnosis - 46y) as compared to females (mean age of diagnosis - 50y) in MLH1 mutation positive families (p=0.036). No such trend was observed in MSH2 carrier families (p=0.356).

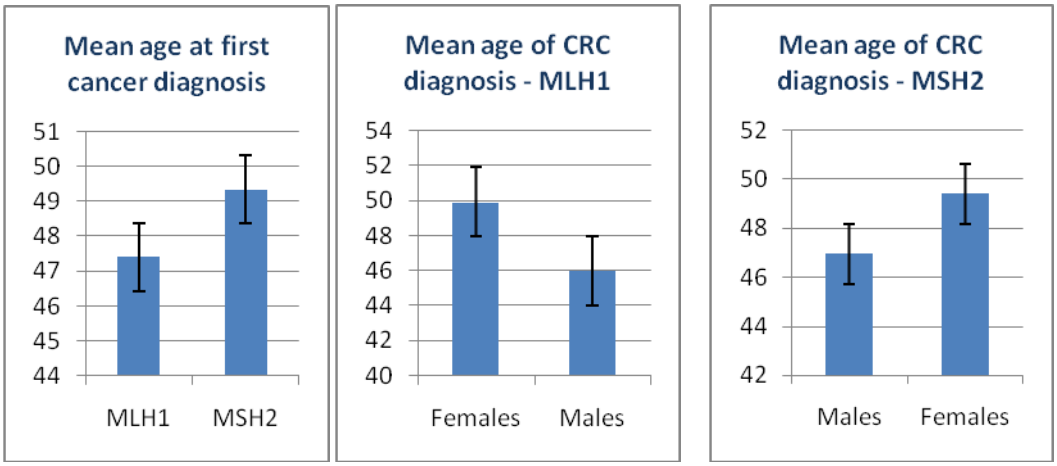


Fig 4.21 Mean age at diagnosis in MLH1/MSH2 mutation carriers

We found no significant differences in the ages at diagnosis between the two genes for the extracolonic cancers [4.21]. However, the extracolonic cancer showed later age of diagnosis for extracolonic cancers as opposed to CRC in MSH2 mutation families (p=0.0176). This observation was not noted in MLH1mutation positive families (p=0.228).

4.11 DISCUSSION

Lynch syndrome is the most common hereditary colorectal cancer predisposition syndrome characterized by a high lifetime risk of developing colorectal and other extracolonic cancers [156]. Lynch Syndrome is genetically heterogenous and is caused due to germline mutations in one of the MMR genes. As there are only three studies reporting MMR mutations in Indian LS patients [254-256] as opposed to a large number of studies in Caucasian population [252-253], we undertook a comprehensive mutation analysis of MMR genes in an Indian cohort of 81 LS families.

A combination of Sanger sequencing to study point mutations and small indels, and MLPA to study LGRs resulted in the identification of MMR gene mutations in 75 out of the 81 families studied thereby leading to a mutation detection rate of 93% which is a fairly high mutation detection rate. The reported mutation detection rates for MMR genes vary from as low as 7% [257] to as high as 88% [220]. This variation is due to the use of different clinical criteria for syndromic diagnosis, use of prescreening techniques and the approach used for mutation screening. Previous studies have shown that the different clinical criteria used for diagnosis of Lynch syndrome differ

in their sensitivity and specificity [252]. Therefore, these criteria must be used with great caution because the stringent criteria like Amsterdam Criteria will miss many of the mutation carriers while relying on relaxed criteria like Revised Bethesda Guidelines will lead to low mutation detection rates. In this sense, use of prescreening techniques like IHC and MSI in combination with clinical criteria serves as efficient method to guide genetic testing and have been widely used in various studies [258]. The high mutation detection rate achieved in our cohort is because the syndromic diagnosis relied on the presence of significant family history of cancers along with clinico-pathological features typical of Lynch syndrome. In absence of family history, loss of expression of MMR proteins was used as a criterion to suspect Lynch syndrome which helped select more appropriate cases with increased yield of genetic testing.

MLH1 and MSH2 gene mutations were identified in 72 out of the total 75 mutation carriers in our cohort. MSH6 gene mutations were identified in 2 families and an EPCAM deletion was identified in one case. This observation is in line with other studies which report that MLH1 and MSH2 mutations together account for nearly 90% of all the MMR mutation carriers [166]. The mutation spectrum is diverse in MLH1 gene and the mutations are scattered throughout the gene while MSH2 mutations are spread up to Exon 12 of the gene. Missense and splice site mutations predominate in MLH1 gene whereas MSH2 mutation spectrum shows a large frequency of large genomic rearrangements. High frequency of LGRs in MSH2 is explained by the presence of Alu repeat sequences in MSH2 gene [259]. The mutation spectrum of MMR genes in our cohort 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. Most of the mutations identified in MLH1 and MSH2 genes were unique while 6 mutations in MLH1 and 2 mutations in MSH2 were identified in two or more unrelated families. The most recurrent mutation in our cohort is the c.942+2A>T mutation in Intron 5 of MSH2 gene which was identified in 7 families. This particular mutation is also the most recurrent mutation worldwide. It is considered to be arising *de novo* in various populations [260], without any proven founder effect except in Newfoundland where it was identified as a founder mutation [261].

In our study, phenotype data was available for 347 affected members from the 75 mutation carrier families in which a total of 379 cancers were recorded. The phenotypic characteristics and the genotype-phenotype correlations studied in this cohort largely concur with the known reports with some notable exceptions. CRC was the most common diagnosed in both MLH1 and MSH2 carriers. In keeping with the known literature, extracolonic cancers were more common in MSH2 carriers with endometrial cancer being the most common extracolonic in women from both MLH1 and MSH2 carrier families [171]. Whether breast cancer is part of Lynch syndrome tumors remains controversial [170-173]. Diagnosis of breast cancers in a total of 12 females (nearly 6%) from both MLH1 and MSH2 mutation carriers in our cohort suggest that that breast cancer may be included in the tumor spectrum of Lynch syndrome.

CHAPTER 5

Founder mutations in Indian Lynch Syndrome Patients

5.1 FOUNDER MUTATIONS: AN INTRODUCTION

Inherited syndromes caused due to germline mutations follows two rules in general. First, different pathogenic mutations in the same gene generally lead to expression of same phenotype in diseased individuals. As a result, different families affected by the same disease usually have different mutations responsible for that disease. Secondly, in a few genetic diseases, the same mutation in the same gene is observed repeatedly. For example, more than 1000 unique DNA variants have been reported for each of the two major LS genes, MLH1 and MSH2, plus several hundred for MSH6 and PMS2 after two decades of mutation analysis [262]. However, the number of mutation carrier families are six nearly fold higher which implies that mutations tend to recur among the populations [262].

Two possible explanations exist for this recurrence of mutations among unrelated families. On the one hand, the so-called hotspot mutations, which has the tendency to recur in populations arise due to sequence specific characteristics that can predispose to an abnormal event at meiosis (e.g. c.942+3A>T in MSH2) [261, 263]. On the other hand, a founder mutation arise in a single individual whose offspring each have a 50% or 25% chance of inheriting the mutation depending upon the dominant or recessive mode of transmission; who further pass on this mutation to their progenies and thus this mutation is successively passed on to subsequent generations. In many of the genetic diseases, the germline mutation carriers die before they reach their reproductive age, thus preventing the mutant genes to reach the future generations. However, founder mutations are the mutations in which the carriers reproduce and consequently they are spread from the ancestor to his or her descendants. The fate of the mutation in the following generations will depend on two main factors; natural selection and chance events like genetic drift which are the enrichment mechanisms for founder mutations.

***Natural selection*:** This mechanism usually predominates in large populations. Some founder mutations could be positively selected through natural selection; whereby the mutations that confers reproductive and genetic qualities that prove advantageous to survival prevail into future generations [264]. Natural selection mainly operates by differential reproductive success (fitness) of individuals and gives rise to populations that have evolved to succeed in specific environments.

***Genetic drift*:** This is a random event that occurs by chance in nature which influences or changes allele frequency within a population [265]. Unlike natural selection, genetic drift is not driven by environmental or adaptive pressures [265]. The effect of genetic drift is larger in small populations and smaller in large populations. The incidence of founder mutations can be increased at population bottlenecks (described in Box 5.1).

Box 5.1: Bottleneck effect

Bottleneck effect occurs when there is a sharp decline in a population's size typically due to environmental factors (natural disasters such as earthquakes or tsunamis; epidemics that can decimate the number of individuals in the population, predation or habitat destruction, etc.). It is a random event, in which some genes are extinguished from the population. This results in a drastic reduction of the total genetic diversity of the original gene pool. The small surviving population is considerably farther from the original one in its genetic makeup.

For example, suppose a mutation arise *de novo* in an individual from a mixed population such that its frequency is 1:1000 (low incidence) and if 10 people of this migrate to certain geographically isolated region of whom 1 is the mutation carrier, then a 100-fold enrichment of this mutation will occur at the founding. In this new population, genetic drift will determine if the frequency of this mutation will increase, decrease or remain unchanged in the founding population. Because genetic drift is enhanced in small populations, most founder mutations have been described in populations that have remained isolated while growing rapidly. We therefore find numerous examples of founder mutations in relatively isolated regions (e.g. Quebec, Newfoundland, Tenerife Island in Spain), countries (e.g. Finland, Iceland, Netherlands) or ethnic groups such as the Ashkenazi Jews (AJ). Founder mutations have been also discovered in large, genetically heterogeneous populations (e.g. Europe, North-America) [228].

Founder mutations are mostly reported in recessive disorders such as beta thalassemia [266], cystic fibrosis [267] or Xeroderma pigmentosum [268] among others. In dominant predisposing diseases, founder mutations usually exist when the age of onset of the disease is past the reproductive age. This is the case of some founder mutations in BRCA1 and BRCA2 genes causing HBOC syndrome [269, 270].

Founder mutations are identified in both stable populations as well as in populations with a migratory history. In stable populations, the mutation may arise within the population itself or may be brought into a population by an immigrant. In migratory populations, the mutation may be introduced by a founder whose descendants either reproduce with other immigrants or with a local population.

5.2 STUDY OF FOUNDER MUTATIONS

The unique feature of founder mutations is that the mutation will always be present in a larger DNA stretch which is identical in all the mutation carriers and identical to that of the founder. This phenomenon is called as "identity by descent". This entire shared region of DNA is called as a haplotype. Thus the haplotype can be defined as the combination of alleles on a single chromosome that are inherited together. Thus, if you share a haplotype, you share an ancestor (the founder) [271].

In order to prove that a particular mutation is a founder, it is necessary to haplotype several markers [single nucleotide polymorphisms (SNPs) or microsatellites] surrounding the mutation in both carriers and non-carriers. If all carrier individuals from the different families share a common haplotype not frequently present in non-carriers, we can conclude that most probably the mutation originated in a single founder individual who spread the mutation [271]. The age of founder mutation can be estimated by determining the length of the haplotype because we expect that the shorter the haplotype, the older is the mutation [271]. The original founder haplotype is actually the entire chromosome that contains the mutation. The founder passes on that chromosome to offspring which then exchange their DNA segments with the chromosome from mate in a recombination event during meiosis. Carrier offspring thus inherit a newly mixed chromosome that contains the mutation along with other parts of the founder's haplotype. The mutation will still be embedded in a very long section of the founder's version of DNA after only one recombination event. The haplotype that includes the mutated gene thus gets whittled down with each subsequent recombination. The concept of founder mutation and haplotype is illustrated in Fig 5.1

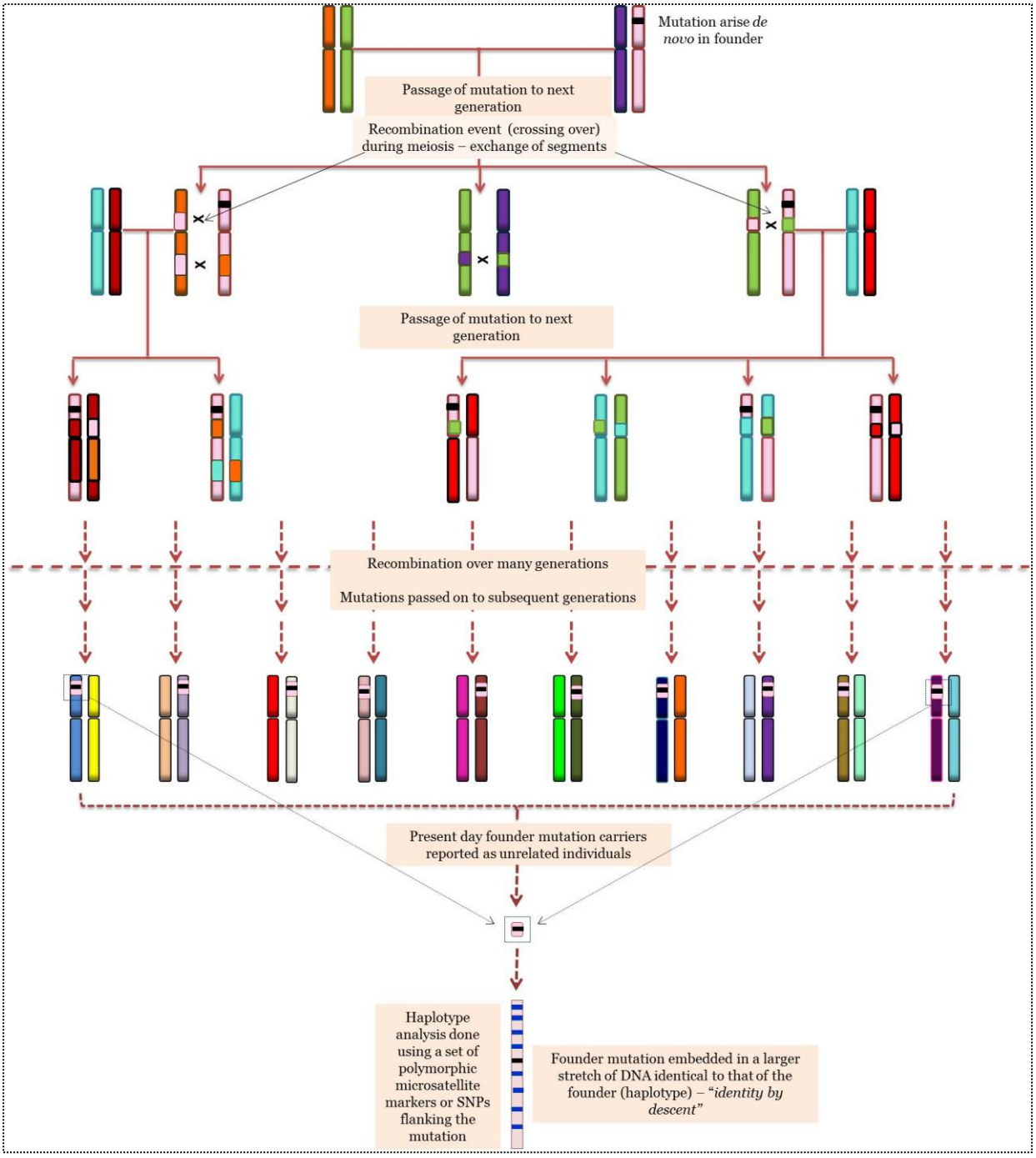


Fig 5.1 Founder mutation mechanism and haplotype analysis

X – Represent the exchanged segments between the homologous chromosomes after recombination. As seen in the figure, the present day mutation carriers who report themselves as unrelated to each other are actually descendants of a common ancestor, the founder. All of them carry the mutation surrounded in a haplotype, studied using a panel of polymorphic markers.

5.3 FOUNDER MUTATIONS IN MMR GENES

The first evidence of existence of founder mutations in Lynch syndrome was reported in Finnish population in 1996 [272]. After this, more than 50 founder mutations have been reported so far in the MMR genes in Lynch syndrome families from all over the world [273]. Some of these mutations have been reported in only few unrelated families while some founder mutations contribute to the majority of mutation carriers in the population. Most evident example of the latter is the presence of two founder mutations (c.454-1G>A and exon 16 deletion) in MLH1 gene that together accounts for nearly half of all Finnish LS families [274]. Founder mutations can be present at a very high frequency in certain populations which can change the generally accepted distribution of mutations among the MMR genes. For example, the presence of two founder mutations in MSH2 gene in the Spanish population doubles the rate of mutation in this gene as compared to the MLH1 gene; which is actually presumed to have a slightly higher or equal mutation rate as MSH2 [273]. Founder mutation in the MMR genes encompasses all types of mutations including the frameshift mutations, missense, splicing mutations and large genomic rearrangements [273]. Epimutations in the MLH1 gene and EPCAM deletions have also been proved to be a founder mutation in certain populations [273].

Interestingly, few mutations that have been reported in the MMR genes at a relatively higher frequency in different populations show founder effect in certain populations whereas they represent recurrent mutations in other populations lacking the presence of common haplotype in the mutation carriers. A notable example of this is the c.942+3A>T mutation in Intron 5 of MSH2 gene which is the most recurrent mutation in MSH2 gene that have been reported to arise *de novo* in various populations worldwide [275]. However, this mutation was shown to be a founder mutation in Newfoundland where it represents one fourth of all the mutation carriers that were shown to share a common haplotype absent in the carriers from other populations [261]. Another example of this observation is the c.388_389del mutation in MSH2 gene that was proved to be a founder mutation in Portuguese LS patients [276]. This same mutation was also identified in families from Germany, Scotland, England and Argentina but the carriers from these populations did not share the Portuguese haplotype which led the authors to propose that may be this mutation is a recurrent mutation [273]. Another striking observation in Lynch syndrome is the presence of

same founder mutations in different population; with the mutation occurring on unique haplotypes in distinct populations. Example of this include a MLH1 splicing mutation, c.589-2A>G that was identified in 10 unrelated American families and also in 3 Italian families [277]. However, this mutation occurred in the background of different haplotypes in the two populations. Most of the LS founder mutations have been described in isolated populations and communities like the Ashkenazi Jews (AJ); three founder mutations (c.1906G>C in MSH2 and c.3959_3962del and c.3984_3987dup mutations in MSH6) account for nearly three quarters of all MMR mutations in a cohort of AJ from Israel [278 - 280]. However, not all LS founder mutations are limited to specific, relatively isolated regions or communities; they also occur rarely in outbred populations. A remarkable example is presence of several founder mutations in the Americans which is a large heterogenous population [277, 281].

Identification of founder mutations in MMR genes in LS has enabled the researchers to design cost-effective molecular diagnostic approaches. The most evident case is found in Finland where screening of two founder mutations identifies nearly half of all the MMR mutation carriers [272]. Apart from this, founder mutation studies in different and far-away geographical regions has allowed scientists to retrace the main migratory patterns that have involved large populations in the past centuries. In Hong Kong, 10 families have been shown to share an MSH2 founder mutation c.1452_1455del in a common haplotype. Interestingly, they all originated from the Guangdong province of southern China, which is the origin of the most Hong Kong inhabitants. Given that during the 19th and 20th centuries there were major emigrations from Hong Kong and Guangdong province, this mutation is interesting not only for its founder effect in China, but also for Chinese communities worldwide [282].

Table 5.1 Founder mutations in MLH1 gene (Adapted from Ref. 273)

Mutation	Population/ Region	No. of families	Haplotype studies	Age of mutation (years)
c.85G>T	Caucasians	5	Yes	-
c.112A>C	Netherlands	6	Yes	-
c.306+5G>A	Spain	17	Yes	~1879
c.392C>T	Republic of Macedonia	3	Yes	-
c.454-1G>A	Finland	5	Yes	125–525
c.545+3A>G	Italy	1	Yes	-
	Quebec	1	Yes	-
	Brazil	3	No	-
c.589-2A>G	US	10	Yes	~450
	Italy	3	Yes	-
c.731G>A	Italy	3	Yes	-
c.793C>T	Taiwan	13	Yes	-
c.1381A>T	North America	3	Yes	-
c.1489dup	Germany	21	Yes	-
c.1558+1G>T	Italy	2	Yes	-
c.1667+2_1667+8del7ins4	Denmark	16	Yes	-
c.1731G>A	Italy	2	Yes	-
c.1758dup	Korea	11	Yes	-
c.1831delAT	Quebec	2	Yes	-
c.1865T>A	Spain	12	Yes	-
c.2142G>A	Swiss	1	Yes	>200
c.2195_2198dup	Quebec	5	Yes	-
c.2252_2253del	Italy	11	Yes	~1550
c.2269dup	Italy	4	Yes	-
	Argentina	1	No	-
Exon 11 deletion	China	2	Yes	-
Exon 12 deletion	Quebec	6	Yes	-
Exon 12–13 duplication	Colombia	2	No	-
Exon 16 deletion	Finland	14	Yes	400–1075
Exon 17–19 MLH1 + Exon 26–29 LRRFIP2 deletion	Portugal	14	Yes	~283

Table 5.2 Founder mutations in MSH2 gene and EPCAM gene (Adapted from Ref. 273)

Mutation	Population/ Region	No. of families	Haplotype studies	Age of mutation (years)
MSH2				
c.388_389del	Portugal	16	Yes	-
c.942+3A>T	Newfoundland	11	Yes	-
c.1165C>T	Quebec	3	Yes	-
c.1452_1455del	China (Guangdong)	10	Yes	550-2575
c.1788_1790del	Denmark	5	Yes	-
c.1906G>C	Ashkenazi Jews	16	Yes	200-500
c.2063T>G	Spain (Tenerife Island)	5	Yes	-
Exon 1–6 deletion	North America	41	Yes	~500
Exon 1–6 deletion	Italy	3	Yes	-
Exon 4–8 deletion	Spain	4	Yes	-
Exon 7 deletion	Spain	3	Yes	-
Exon 8 deletion	Italy (Sardinia Island)	7	Yes	-
Exon 8 deletion	Italy (Sardinia Island)	2	Yes	-
EPCAM				
Exon 8–9 deletion	Spain	3	Yes	-
Exon 8–9 deletion	Denmark	3	Yes	-

Table 5.3 Founder mutations in MSH6 gene (Adapted from Ref. 273)

Mutation	Population/ Region	No. of families	Haplotype studies	Age of mutation (years)
c.10C>T	Quebec	11	Yes	~513
c.467C>G	Netherlands	7	Yes	-
c.650dupT	Netherlands	7	Yes	-
c.1346T>C	Sweden	5	No	>300
c.1614_1615delinsAG	Netherlands	3	Yes	-
c.2931C>G	Sweden	2	Yes	> 200
c.2983G>T	Finland	2	No	-
c.3959_3962del	Ashkenazi Jews	8	Yes	~1425
c.3984_3987dup	Ashkenazi Jews	14	Yes	~1325

Table 5.4 Founder mutations in PMS2 gene (Adapted from Ref. 273)

Mutation	Population/ Region	No. of families	Haplotype studies	Age of mutation (years)
c.137G>T	US (Caucasian)	10	Yes	-
c.736_741delins11	US (British, Swedish)	12	Yes	~1425
c.903G>T	US (Caucasian)	3	Yes	-
c.989-1G>T	Norway	>2	No	-
Exon 10 deletion	US (Caucasian)	5	Yes	-

5.4 GENESIS OF INDIAN POPULATION

India being the second most populated country after China, contributes to one sixth of the world's total population. India is a land of diverse religions and languages; with each of the 29 states practicing their own traditions and cultures. Apart from this ethnic diversity, Indians are also racially and genetically diverse people. This diversity is the consequence of the multiple waves of migration and gene flow that took place in historic and prehistoric times.

❖ ***Origins of Indian Population***

India has served as a major corridor for the dispersal of modern humans [283]. The exact timing of the entry of modern humans in India remains largely unknown. However, a crude estimate is that humans appear to have spread to many parts of India during the middle of Paleolithic period some 50000-20000 years before present (BP) [284]. A recent genome-wide study on 367 Indians representing geographic, linguistic and ethnic diversities revealed four major ancestries in mainland India [285]. This is in contrast with an earlier study based on limited population sampling which proposed two major ancestries; Ancestral North Indian (ANI - genetically close to Middle Easterners, Central Asians, and Europeans) and Ancestral South Indian (ASI - not closely related to groups outside the South Asian subcontinent) [286]. As per the new study, the four ancestries contributing to the genomic structure of mainland Indian population include the previously inferred ANI and ASI and the recently suggested Ancestral Austro-Asiatic (AA language speakers) and Ancestral Tibeto-Burman (TB language speakers). The authors proposed that the absence of significant resemblance of ASI and the AAA with any of the neighboring populations is indicative of them being early settlers in India, possibly arriving on the "southern exit" wave out of Africa. Differentiation between the ASI and the AAA possibly took

place after their arrival in India. The ANI and the ATB can clearly be rooted to the Central South-Asians and East-Asians respectively and they likely entered India through the northwest and northeast corridors, respectively. Ancestral populations therefore seem to have occupied geographically separated habitats.

- ***Admixture to Endogamy in Indian population***

Extant Indian population possess multi-ancestral components and some geographical displacements which points towards extensive admixture between the ANI and ASI populations in an asymmetric manner [285, 286]. Evidence for this admixture comes from genetic studies, archaeological studies and linguistic studies done to trace the history of Indian subcontinent. Archaeological evidences have assumed that major mixture between populations in India occurred 1,900–4,200 years BP, well after the establishment of agriculture in the subcontinent [287]. Indo-European languages including Sanskrit and Hindi (primarily spoken in northern India) are part of a larger language family that includes the great majority of European languages. In contrast, Dravidian languages including Tamil and Telugu (primarily spoken in southern India) are not closely related to languages outside of South Asia [288]. Evidence for long-term contact between speakers of these two language groups in India is evident from the fact that there are Dravidian loan words (borrowed vocabulary) in the earliest Hindu text (the Rig Veda, written in archaic Sanskrit) that are not found in Indo-European languages outside the Indian subcontinent [288, 289]. The genetic studies showed that the ANI ancestry ranges from 39-71% in India, and is higher in traditionally upper caste and Indo-European speakers [289]. Groups with only ASI ancestry may no longer exist in mainland India. However, the Andamanese are an ASI-related group without ANI ancestry, showing that the peopling of the islands must have occurred before ANI-ASI gene flow on the mainland. This shows that different Indian groups have inherited different proportions of ancestry from ANI and ASI, however reliable estimates of these proportions in the current Indians is not possible [289].

Genetic studies have shown that India was a “relatively” pan-mixing society that embraced endogamy between 1,900 and 4,200 years [285, 289]. Evidence was recently provided that gene flow ended abruptly with the defining imposition of some social values and norms. The reign of the ardent Hindu Gupta rulers, known as the age of Vedic Brahminism, was marked by strictures laid down in Dharmashastra—the

ancient compendium of moral laws and principles for religious duty and righteous conduct to be followed by a Hindu—and enforced through the powerful state machinery of a developing political economy. These strictures and enforcements resulted in a shift to endogamy. More recent admixture among the Maratha is in agreement with the known history of the post-Gupta Chalukya (543–753 CE) and the Rashtrakuta empires (753–982 CE) of western India, which established a clan of warriors (Kshatriyas) drawn from the local peasantry. In eastern and northeastern India, populations such as the West Bengal Brahmins (WBR) and the TB populations continued to admix until the emergence of the Buddhist Pala dynasty during the 8th to 12th centuries CE. As pointed out earlier, the admixture is asymmetrical because males from dominant populations, possibly upper castes, with high ANI component, mated outside of their caste, but their offspring were not allowed to be inducted into the caste; however, vice versa was not seen. This is consistent with the theory of elite dominance and patriarchy. In summary, the current day Indian population is derived from extensive admixture between four ancestral populations (predominantly ANI and ASI) followed by an era where endogamy was embraced that led to the cessation of gene flow events [285, 289].

5.5 THE EXTANT INDIAN POPULATION

Contemporary populations of India are linguistically, geographically, and socially stratified and are largely endogamous with variable degrees of porosity. The people of India are culturally classified as tribals, who constitute less than 10% of the total population and non-tribals [290]. There are nearly 450 tribal communities in India who speak 750 dialects that can be classified into one of the following three language families: Austro-Asiatic (AA), Dravidian (DR) and Tibeto-Burman (TB) [285, 290]. Most of the present day non-tribal populations of India belong to the Hindu religious fold and are hierarchically arranged in four main caste classes, namely, Brahmin (priestly class), Kshatriya (warrior class), Vysya (business class) and Sudra (menial labor class). In addition, there are several religious communities, who practice different religions; Islam, Christianity, Sikhism, Judaism and so on. The non-tribals predominantly speak languages that belong to the Indo-European (IE) or Dravidian families [290].

5.6 FOUNDER MUTATIONS: INDIAN SCENARIO

India now has a unique anthropological structure, caste system and practices; all of which have affected the genetic composition of the existing Indian population [291]. Allele frequency differences between diverse ethno-linguistic groups in India are larger than in Europe, reflecting strong founder effects whose signatures have been maintained for thousands of years. This is because endogamous and consanguineous marriages have become conventional in India owing to the evolving socio-religious activities of the subsects and castes; which further prevent founder events to be erased from the gene flow. Consequently, the frequency of founder mutations and the burden of heritable genetic disorders rose significantly [291]. Specifically, the incidence of certain autosomal-recessive disorders is relatively high in select Indian subpopulations and communities that share common recent ancestry. This is consistent with the theory of Reich et al who predicted that there will be an excess of recessive diseases in India, different in each group, which should be possible to screen and map genetically [286]. Also, Haldane appropriately stated 50 years ago that recessive characteristics will become rarer in India when inter-caste marriages become common [292].

Rough estimates are that more than 50 million individuals in India are affected with single-gene disorders (monogenic disorders) [293]. However, the current situation in India shows lack of widespread awareness about genetic disorders both in the general population & medical professionals and lack of affordable genetic tests; further increasing the burden on Indian society [293]. Therefore, in a vast country like India, where the disease burden is tremendous and resources are limited, low-cost diagnostic tests that target specific common mutations or founder mutations in select communities or subpopulations hold significance. The pace of research and expanded clinical services in genetics has picked up in India in recent years with the efforts of national agencies and funding bodies like ICMR [291]. Over these years, research has uncovered many founder and common mutations in Indian subpopulations associated with a variety of recessive diseases. Most of the founder mutations have been identified in Agrawal community in various diseases [291, 294]. On the contrary, very few founder mutations have been reported in cancer predisposition syndrome associated genes in the Indian population. A summary of all founder mutations in Indian population is given in Table 5.5

Table 5.5 Founder mutations in Indian population (Adapted from Ref. 291)

Disease	Gene	Mutation	Region/Sect
Hallervorden–Spatz syndrome	PANK2	c.216dupA	Agrawals
Heme oxygenase 1 deficiency	HMOX1	c.130C>T	Agrawals
LGMD2A	CAPN3	c.2338G>C	Agrawals
		c.2051–1G>T	Agrawals
Megalencephalic leukodystrophy with cysts	MLC1	c.135dupC	Agrawals of North India
Werner syndrome	WRN	c.561A>G	Kerala, Tamil Nadu
Tay Sachs disease	HEXA	c.1385A>T	Gujarat
Von Willebrand disease	VWF	c.2908delC	Kachi Modh Ghanchi community (Gujarat)
		c.5335C>T	Gaderia community, Uttar Pradesh
Oculocutaneous albinism	TYR	c.832C>T	Tili group; West Bengal
Hydatidiform mole	NLRP7	c.2078G>C	North India
		c.2738A>G	North-west India
Hypohidrotic ectodermal dysplasia	EDAR	c.1144G>A	Not known
Tricho-HepatoEnteric syndrome	TTC37	c.2808G>A	Gujarat
Progressive pseudorheumatoid dysplasia	WISP3	c.1010G>A	South India
Combined pituitary hormone deficiency	PROP1	c.112124 del13	Indian subcontinent
Haim–Munk syndrome	CTSC	c.2127A>G	Jews from Cochin
Growth hormone deficiency	GHRHR	c.214G>T	West India
Infantile systemic hyalinosis	CMG2	c.1A>G	Mali community (Jodhpur), Rajasthan
Hereditary Fructose Intolerance	ALDOB	c.324+1G>A	Agrawals
Xeroderma Pigmentosa	XPA	p.F112SfsX2	Chavan (Maratha); Maharashtra

5.7 HYPOTHESIS AND OBJECTIVE OF THE STUDY

Atleast 55 founder mutations have been identified in MMR genes in various populations worldwide. Also, numerous founder mutations have been reported in various genes in the Indian population, particularly the Agrawal community. However, no founder mutations have been reported so far in MMR genes in the Indian Lynch syndrome patients so far. As discussed in previous chapter, our study on the mutation spectrum of MMR genes in Indian LS patients have identified five mutations in MLH1 gene that were present in 2 or more unrelated families belonging to specific geo-ethnic communities of Indian subcontinent. We hypothesize that these five mutations are founder mutations in MLH1 gene which accounts for a fraction of Indian LS patients. The objective of this part of the project was therefore to study the founder effect of these mutations by haplotyping analysis.

5.8 RESULTS

This study was conducted to prove the founder effect of five recurrent mutations in MLH1 gene whose details are given in Table 5.6. Three of these mutations are frameshift mutation that leads to truncated protein formation; one is a splice site mutation while the last one is a missense mutation. Carriers of these mutations belong to specific geo-ethnic groups of India.

Table 5.6 Recurrent mutations in MLH1 gene

Sr. No.	Mutation	Consequence	Exon	No. of families	Community/ Ethnic origin	Geographic origin
1	c.156delA	p.E53Rfs*4	2	6	Shia Momin	Patan district, Gujarat
2	c.46insG	p.V16Gfs*13	1	3	Hindu	Bankura district, West Bengal
3	c.306G>T	p.E102D	3	2	Sindhi	Pakistan
4	c.1389_1390delA	p.P464Yfs*14	12	2	Hindu	Bihar
5	c.1558+2insG	-	13	3	Maratha	Sindhudurg, Maharashtra

Haplotyping was carried out using a panel of 13 tightly linked polymorphic microsatellite markers to prove the founder effect of all five suspected founder mutations in MLH1 gene. These markers (Telomere-D3S3564-D3S1298-D3S3623-

D3S1007-D3S1611-D3S1561-D3S2411-D3S3718-D3S3512-D3S4153-D3S3518-D3S3936-D3S2432-Centromere) spans a length of 10MB on Chromosome 3p flanking both sides of the mutations on MLH1 gene. The methodology was to PCR amplify the microsatellite markers using the fluorescently labeled primers (only one primer was labeled) followed by capillary electrophoresis (fragment analysis) of these products to identify the allele sizes of these markers for each sample. Results were analysed using GeneMapper software from Thermo Fisher.

1. c.156delA mutation

This novel frameshift mutation c.156delA in Exon 2 (Fig 5.2) was first identified in 2 unrelated families from Shia Momin community hailing from Patan district of Gujarat which raised the suspicion of founder effect of this mutation which was partly confirmed when one of the mutation carriers reported that GI cancers are very common in his community. In order to prove the founder effect of this mutation, we first genotyped 400 healthy individuals from this community to establish the frequency of c.156delA mutation in this community. We identified a total of 6 healthy individuals from 3 families to harbor this deleterious mutation. Later, this mutation was also identified in 2 individuals from a Hindu family from Saurashtra region of Gujarat. The pedigrees of the 5 families are given in Fig 5.3 – 5.7 (pedigree of one family was not available as they did not follow up with CGC even after many call sessions).

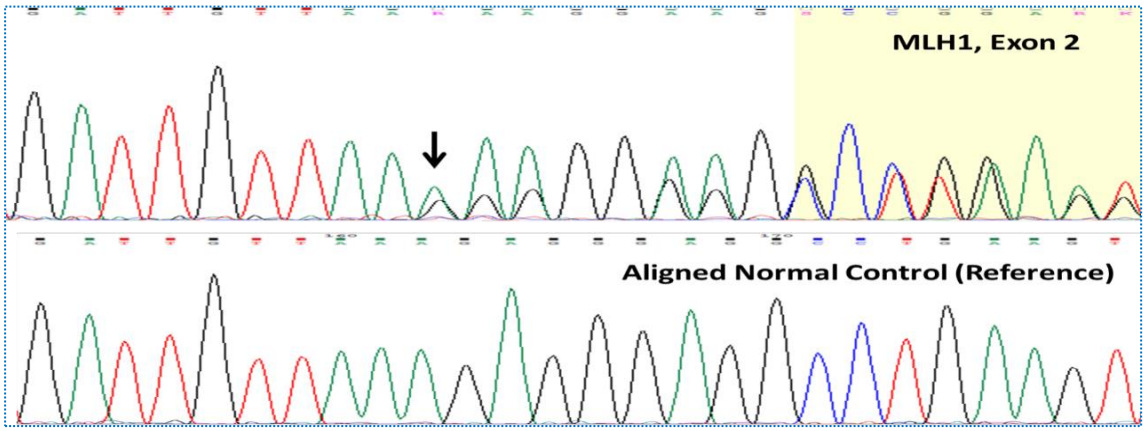


Fig 5.2 Chromatogram showing c.156delA mutation

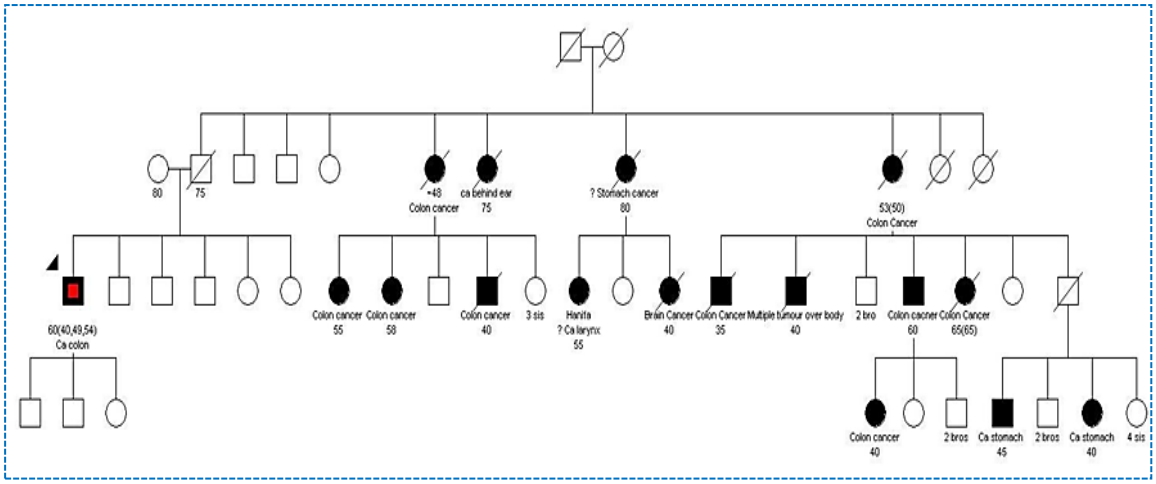


Fig 5.3 Pedigree of Shia Momin Family from Patan, Gujarat (Family ID: SM-1)

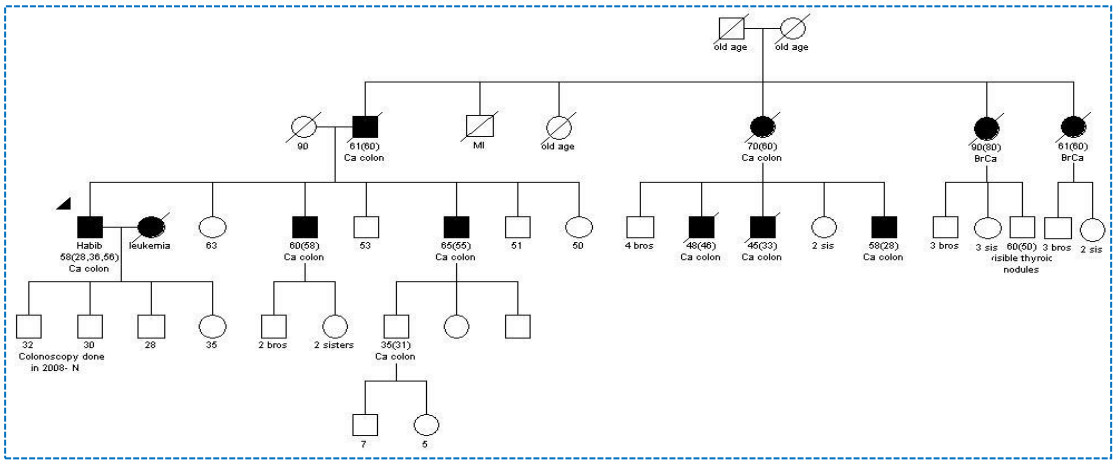


Fig 5.4 Pedigree of Shia Momin Family from Patan, Gujarat (Family ID: SM-2)

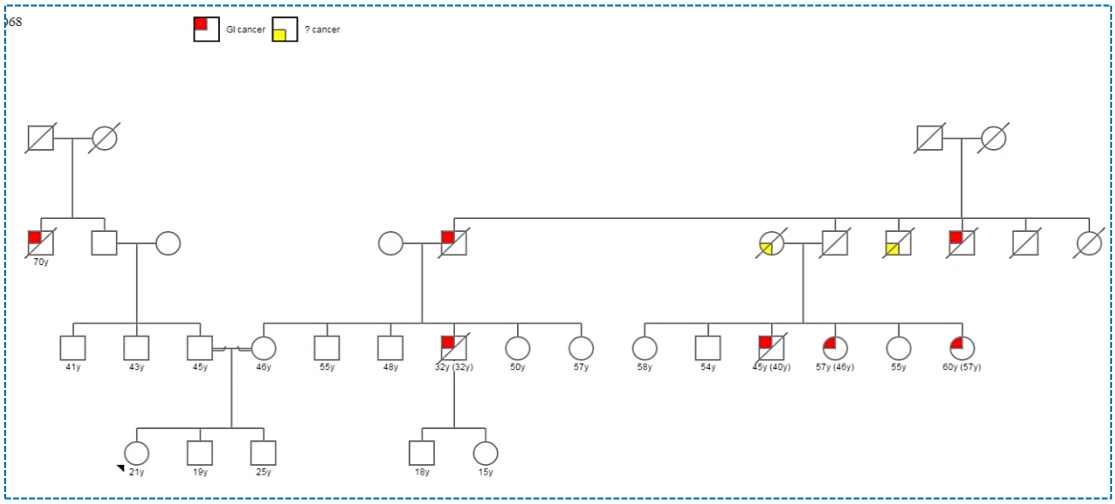


Fig 5.5 Pedigree of Shia Momin Family from Patan, Gujarat (Family ID: SM-3)

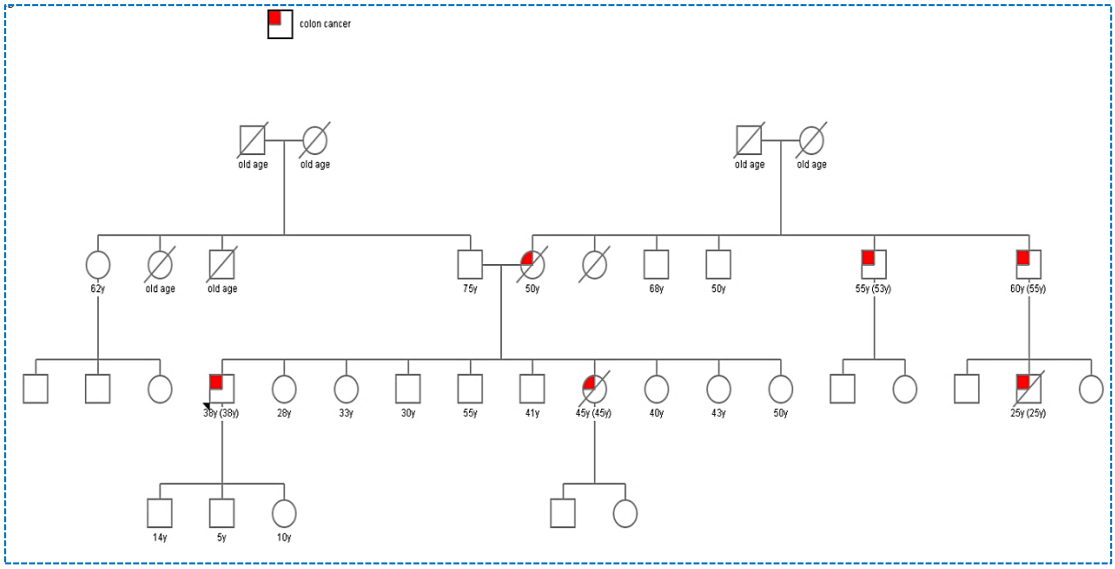


Fig 5.6 Pedigree of Shia Momin Family from Patan, Gujarat (Family ID: SM-4)

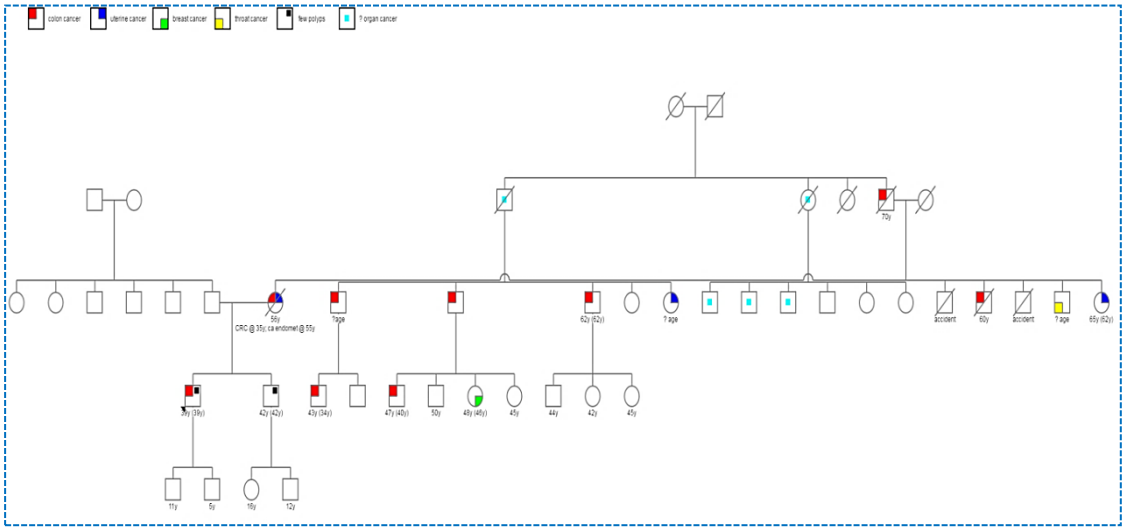


Fig 5.7 Pedigree of Hindu Family from Saurashtra, Gujarat (Family ID: SM-6)

In total, 16 carriers from 6 families (5 Shia Momin and 1 Hindu) were found to carry this deleterious germline mutation. Haplotyping analysis using the 12 markers revealed a conserved haplotype of nearly 5 Mb encompassing 8 microsatellite markers. This haplotype was found to segregate with the c.156delA mutation being identified in all the 16 mutation carriers and in none of the 80 controls samples, thus confirming the founder of this mutation. The allele sizes of the different markers of the conserved haplotype and their frequency in the controls has been summarized in Table 5.7 and Fig 5.8.

Table 5.7: Haplotype associated with c.156delA mutation

Marker	Physical location (Mb)	Sample BK-1	Sample BK-2	Sample BK-3	Sample BK-4	Sample BK-5	Sample BK-6	Inferred haplotype
		Allele sizes (bp)						
D3S3564	**42.37**	205 - 211	203 – 203	205 - 209	203 - 205	209 - 215	211 - 213	-
D3S1298	**38.01**	203 – 207	205 - 205	203 - 211	199 - 203	199 - 205	209 - 215	-
D3S3623	**37.40**	213 – 219	217 – 217	213 - 217	213 - 215	211 – 219	217 - 219	-
D3S1007	**37.03**	**66** – 66	**66** – 66	**66** - 72	**66** - 76	**66** – 72	**66** - 66	**66bp**
D3S1611	**37.02**	**251** – 251	**251** – 251	**251** - 257	**251** -261	**251** – 251	**251** - 261	**251bp**
D3S1561	**36.44**	**223** - 227	**223** - 251	**223** - 223	**223** - 243	**223** - 237	**223** - 241	**223bp**
D3S2411	**36.31**	**225** - 231	**225** -235	**225** - 221	**225** - 223	**225** – 231	**225** - 231	**225bp**
D3S3718	**36.12**	**151** - 151	**151** - 151	**151** – 157	**151** - 151	**151** – 157	**151** - 161	**151bp**
D3S4153	**34.35**	**270** - 270	**270** - 270	**270** - 270	**270** - 270	**270** – 270	**270** - 270	**270bp**
D3S3512	**33.55**	**130** - 130	**130** - 130	**130** - 136	**130** - 134	**130** – 142	**130** -132	**130bp**
D3S3936	**33.0**	**115** - 115	**115** - 115	**115** - 115	**115** - 115	**115** – 115	**115** - 115	**115bp**
D3S2432	**32.13**	129 – 147	127 – 127	125 – 141	121 - 129	123 – 133	127 - 133	-

** Markers shaded in grey in bold fonts are markers showing conserved haplotype

Allele sizes in bold are the common allele sizes shared by carriers

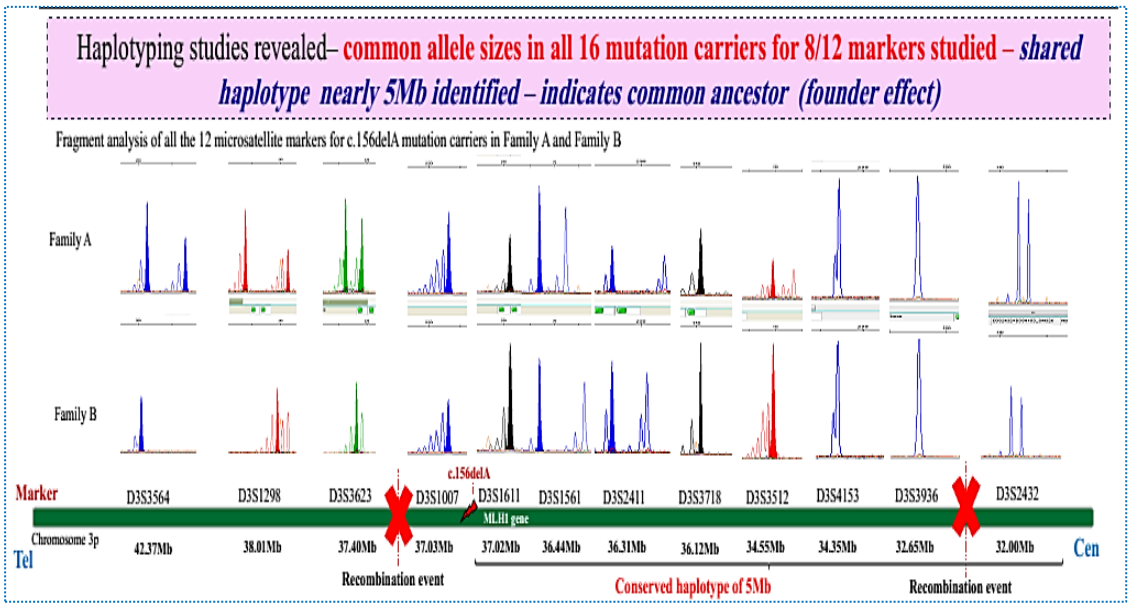


Fig 5.8 Conserved haplotype associated with c.156delA mutation

2. c.46insG mutation:

This novel frameshift mutation (Fig 5.9) was identified in 3 unrelated families of Hindu ethnicity hailing from Bankura district of West Bengal. The pedigrees of these families are given in *Fig 5.10 – Fig 5.12*.

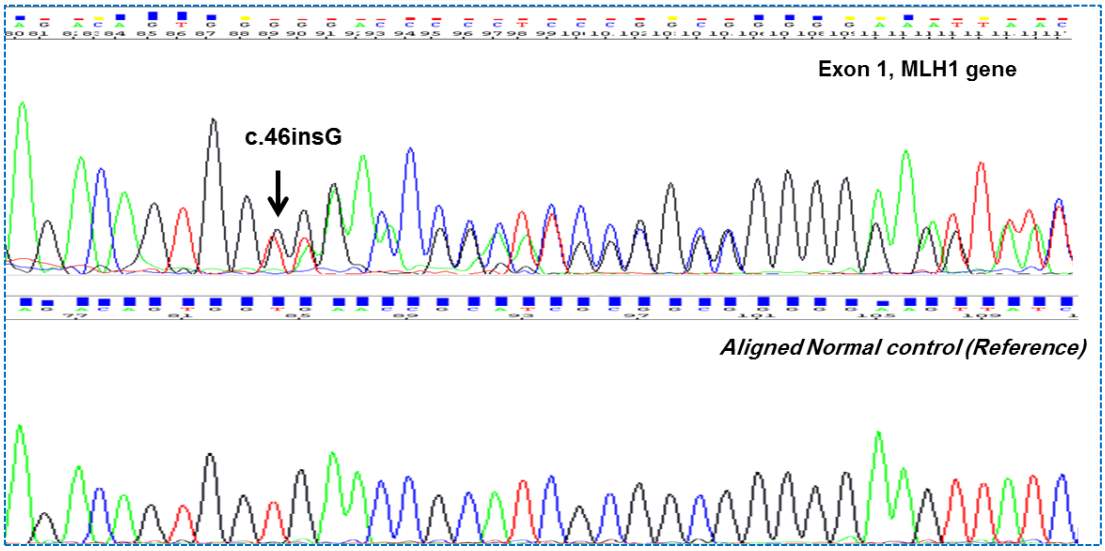


Fig 5.9 Chromatogram showing c.46insG mutation

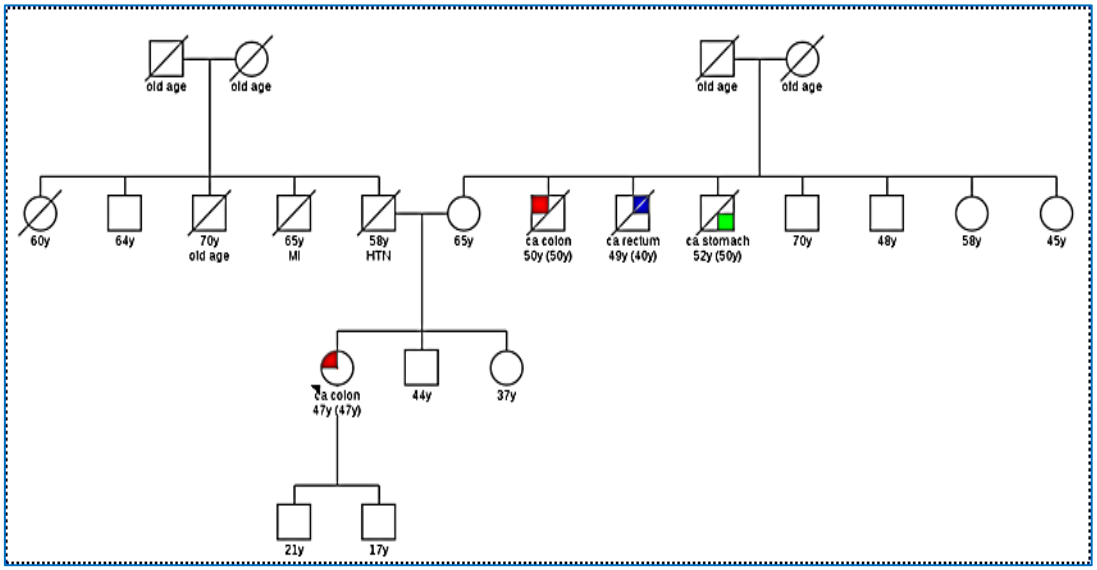


Fig 5.10 Pedigree of Hindu Teli family from Bankura (Family ID: BK-1)

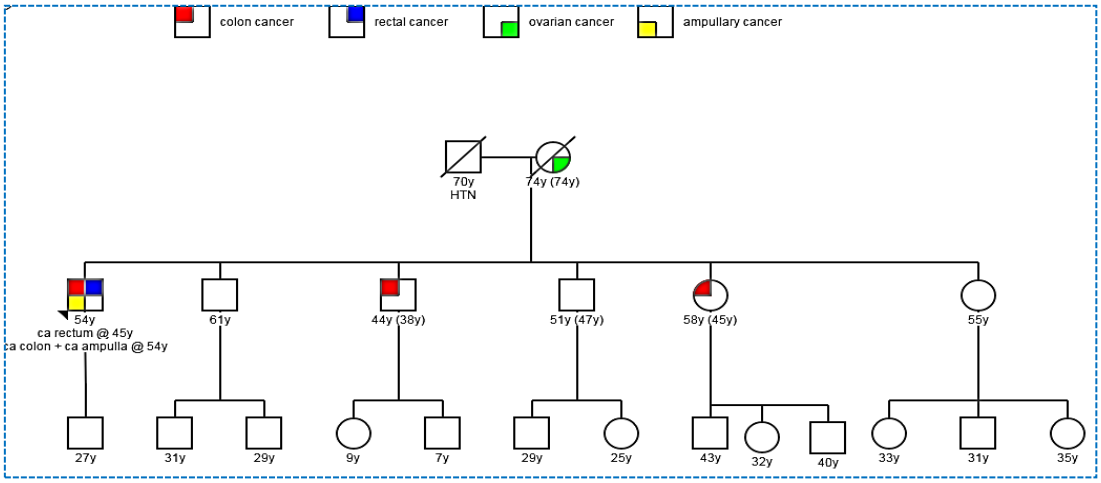


Fig 5.11 Pedigree of Hindu Kshatriya family from Bankura (Family ID: BK-2)

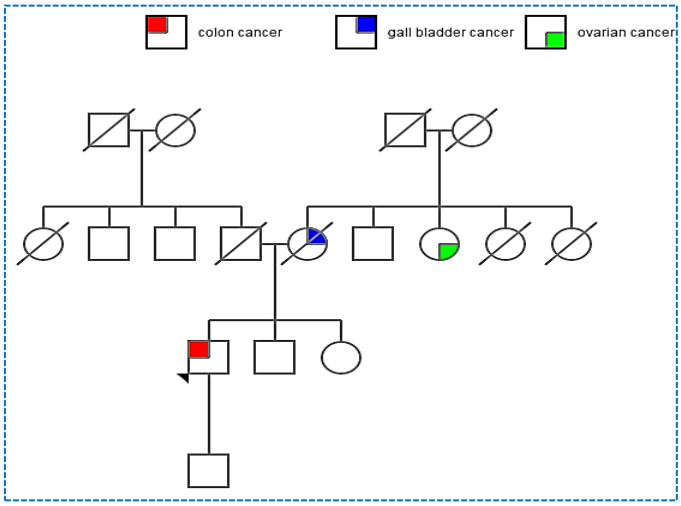


Fig 5.12 Pedigree of Hindu Nagrishi family from Bankura (Family ID: BK-3)

In total, 8 individuals from these 3 families were found to harbor this deleterious mutation. A conserved haplotype was shared by all the 8 carriers of c.46insG mutation which was also identified in 1 out of the 80 control samples haplotyped, thus confirming the founder effect of this mutation. The allele sizes of the different markers of the conserved haplotype and their frequency in the controls has been summarized in Table 5.8

Table 5.8: Haplotype associated with c.46insG mutation

Marker**	Physical location (Mb)	Sample BK-1	Sample BK-2	Sample BK-3	Inferred haplotype
		Allele sizes (bp)			
D3S3564	**42.37**	**193** – 203	**193** – 205	**193** – 203	**193bp**
D3S1298	**38.01**	**202** – 202	**202** – 198	**202** – 194	**202bp**
D3S3623	**37.40**	**216** – 214	**216** – 218	**216 –** 214	**216bp**
D3S1007	**37.03**	**76** – 76	**76** – 72	**76** – 74	**76bp**
D3S1611	**37.02**	**261** – 261	**261** – 257	**261** – 259	**261bp**
D3S1561	**36.44**	**222 –** 220	**222** – 224	**222** – 220	**222bp**
D3S2411	**36.31**	**221** – 221	**221** – 223	**221** – 221	**221bp**
D3S3718	**36.12**	**155** – 157	**155** – 151	**155** – 157	**155bp**
D3S3512	**34.55**	**130** – 130	**130** – 146	**130** – 132	**130bp**
D3S4153	**34.35**	**270** – 270	**270** – 270	**270** – 270	**270bp**
D3S3518	**33.65**	158 – 160	152 – 158	154 – 160	-
D3S3936	**33.0**	115 – 115	115 – 115	115 – 115	-
D3S2432	**32.13**	119 – 145	127 – 139	123 – 147	-
Haplotype length: 8.02 Mb					
Haplotype frequency in controls: 1.25%					

** Markers shaded in grey in bold fonts are markers showing conserved haplotype

Allele sizes in bold are the common allele sizes shared by carriers

3. c.306G>T:

This missense mutation (E102D) arises due to substitution of G to T at the last base of the exon 3. Therefore, it is predicted to affect the splicing process. This particular mutation has been reported 14 times in the InSiGHT database. This mutation (Fig 5.13) was identified in 3 unrelated Indian Sindhi Lynch Syndrome families who report their ancestral origin from Sindh region of Pakistan. The pedigrees of these cases are given in *Fig 5.14 – Fig 5.15*.

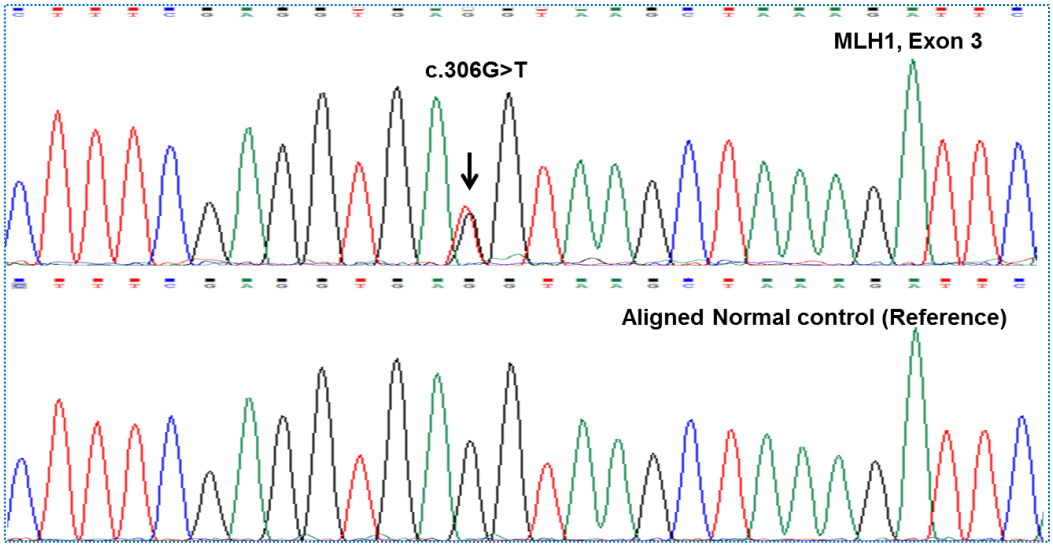


Fig 5.13 Chromatogram showing c.306G>T mutation

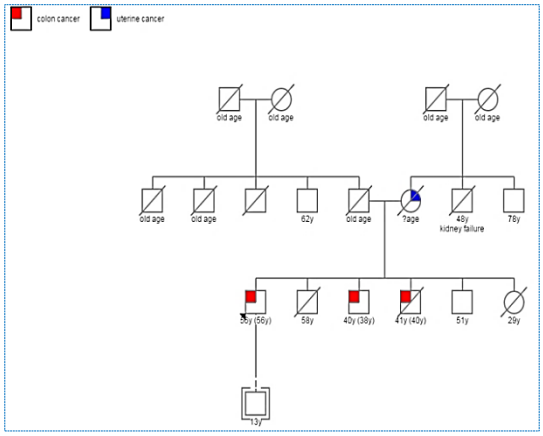


Fig 5.14 - Pedigree of Sindhi family (Family ID: SP-1)

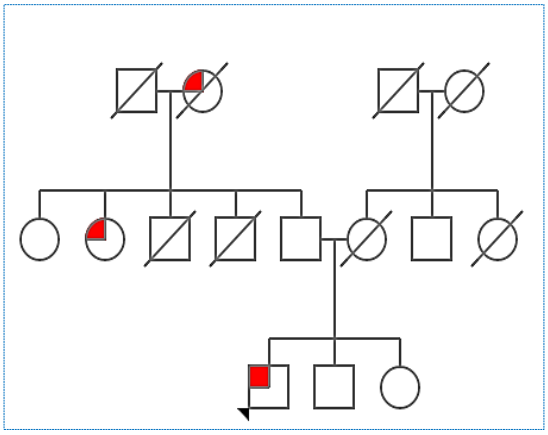


Fig 5. 15- Pedigree of Sindhi family (Family ID: SP-2)

In total, 4 individuals from these 2 families were found to harbor this deleterious mutation. A conserved haplotype was shared by all the 4 carriers of c.306G>T mutation which was not identified in any of the control samples, thus confirming the founder effect of this mutation. The allele sizes of the different markers of the conserved haplotype and their frequency in the controls has been summarized in Table 5.9

Table 5.9 Haplotype associated with c.306G>T mutation

Marker	Physical location (Mb)	SP-1	SP-2	Inferred haplotype
D3S3564	42.37	200 – 210	202 – 212	-
D3S1298	38.01	208 – 210	196 – 204	-
D3S3623	37.40	218 – 222	215 – 215	-
D3S1007	37.03	**76** – 72	**76** – 76	**76bp**
D3S1611	37.02	**261** – 257	**261** – 261	**261bp**
D3S1561	36.44	**222** – 212	**222** – 238	**222bp**
D3S2411	36.31	**221** – 223	**221** – 229	**221bp**
D3S3718	36.12	159 – 167	149 – 155	-
D3S3512	34.55	126 – 126	132 – 132	-
D3S4153	34.35	270 – 270	270 – 270	-
D3S3518	33.65	152 – 162	158 – 158	-
D3S3936	33.0	114 – 114	114 – 114	-
D3S2432	32.13	143 - 147	133 – 139	-
Haplotype length: 0.72 Mb				
Haplotype frequency in controls: 0%				

** Markers shaded in grey in bold fonts are markers showing conserved haplotype

Allele sizes in bold are the common allele sizes shared by carriers

4. c.1558+2insG mutation:

This novel germline mutation (Fig 5.16) was identified in 3 unrelated Lynch Syndrome cases from Sindhudurg region of Maharashtra. The pedigrees of these cases are given in *Fig 5.17 – Fig 5.19.*

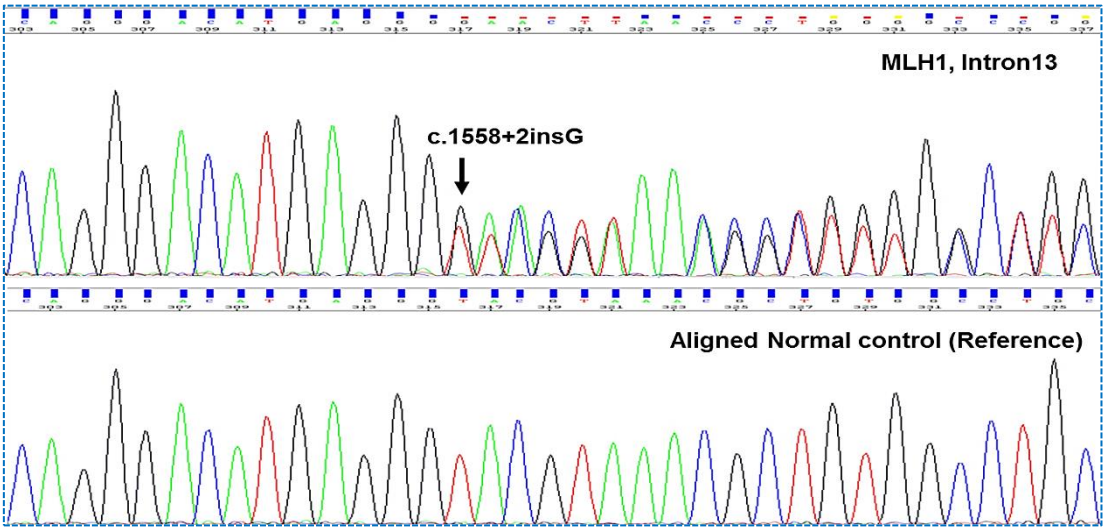


Fig 5.16 Chromatogram showing c.1558+2insG mutation

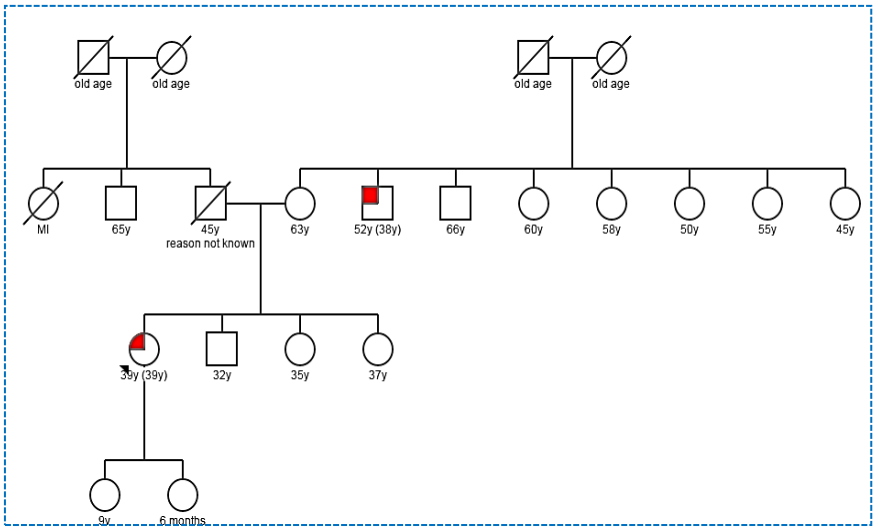
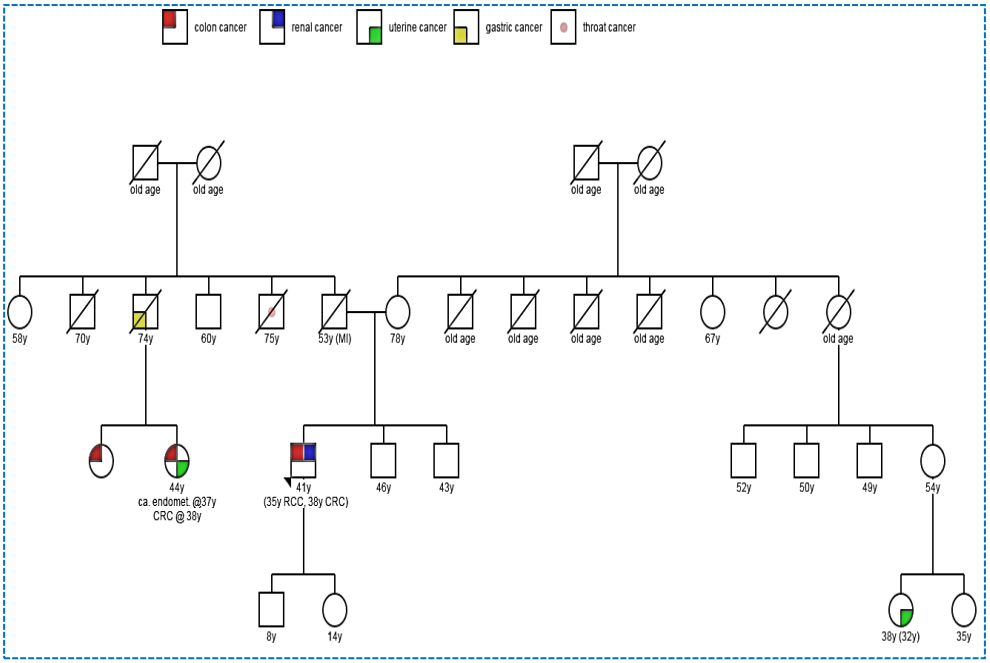


Fig 5.17 Pedigree of Maratha family from Sindhudurg (Family ID: MH-1)



5.18 Pedigree of Maratha family from Sindhudurg (Family ID: MH-2)

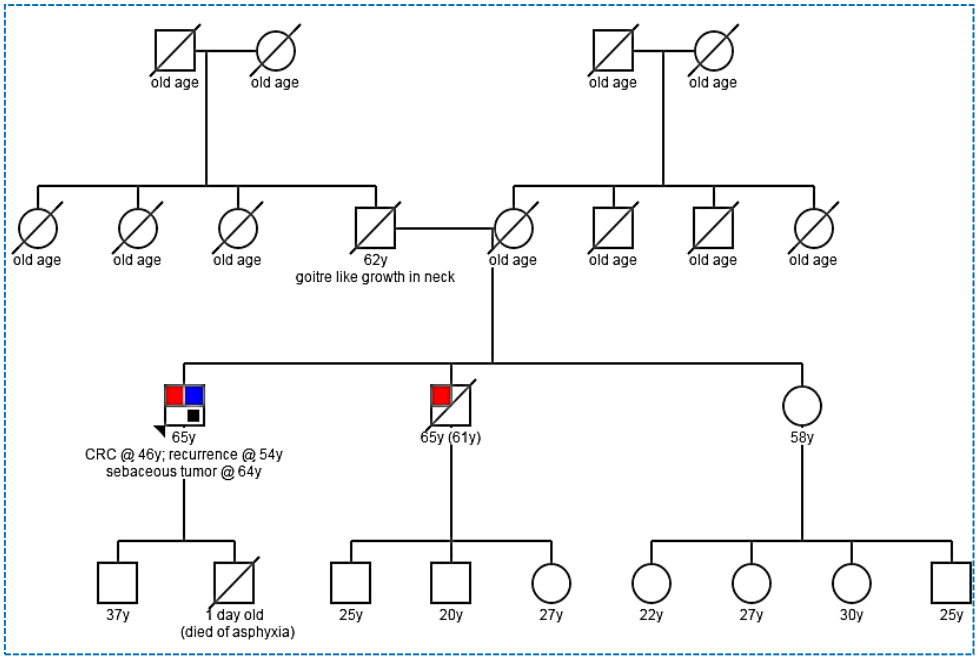


Fig 5.19 Pedigree of Maratha family from Sindhudurg (Family ID: MH-3)

Samples of other family members were not available for genetic testing from these 3 families. Therefore, haplotype analysis was carried out on only the 3 proband cases which revealed the presence of a common haplotype in all the 3 mutation carriers and was not identified in any of the 80 control samples haplotyped.

Table 5.10 Haplotype associated with c.1558+2insG mutation

Marker	Physical location (Mb)	MH-1	MH-2	MH-3	Inferred haplotype
		Allele sizes (bp)			
D3S3564	42.37	202 – 214	202 – 202	192 – 203	-
D3S1298	38.01	202 – 202	202 – 206	202 – 194	-
D3S3623	37.40	213 – 221	215-221	216 - 222	-
D3S1007	37.03	**72** – 76	**72** – 70	**72** – 76	**72**
D3S1611	37.02	**257** - 261	**257** – 257	**257** - 261	**257**
D3S1561	36.44	**222 –** 224	**222** – 236	**222** – 232	**222**
D3S2411	36.31	**223** – 231	**223** – 221	**223** – 223	**223**
D3S3718	36.12	**157** – 161	**157** – 143	**157** – 151	**157**
D3S3512	34.55	**132** – 134	**132** – 130	**132** – 132	**132**
D3S4153	34.35	**270** - 270	**270** – 270	**270** – 270	**270**
D3S3518	33.65	**158** – 152	**158** – 158	**158** – 160	**158**
D3S3936	33.0	**115** – 115	**115** – 115	**115** – 115	**115**
D3S2432	32.13	**139** – 135	**139** – 143	**139** – 143	**139**
Haplotype length: 4.9 Mb					
Haplotype frequency in controls: 0%					

** Markers shaded in grey in bold fonts are markers showing conserved haplotype

Allele sizes in bold are the common allele sizes shared by carriers

5. c.1389_1380delAC mutation:

This novel germline mutation (Fig 5.20) was identified in 2 unrelated Lynch Syndrome cases from Bihar. The pedigrees of these cases are given in *Fig 5.21 – Fig 5.22*. Haplotype analysis revealed a shared haplotype between these two cases (Table 5.11)

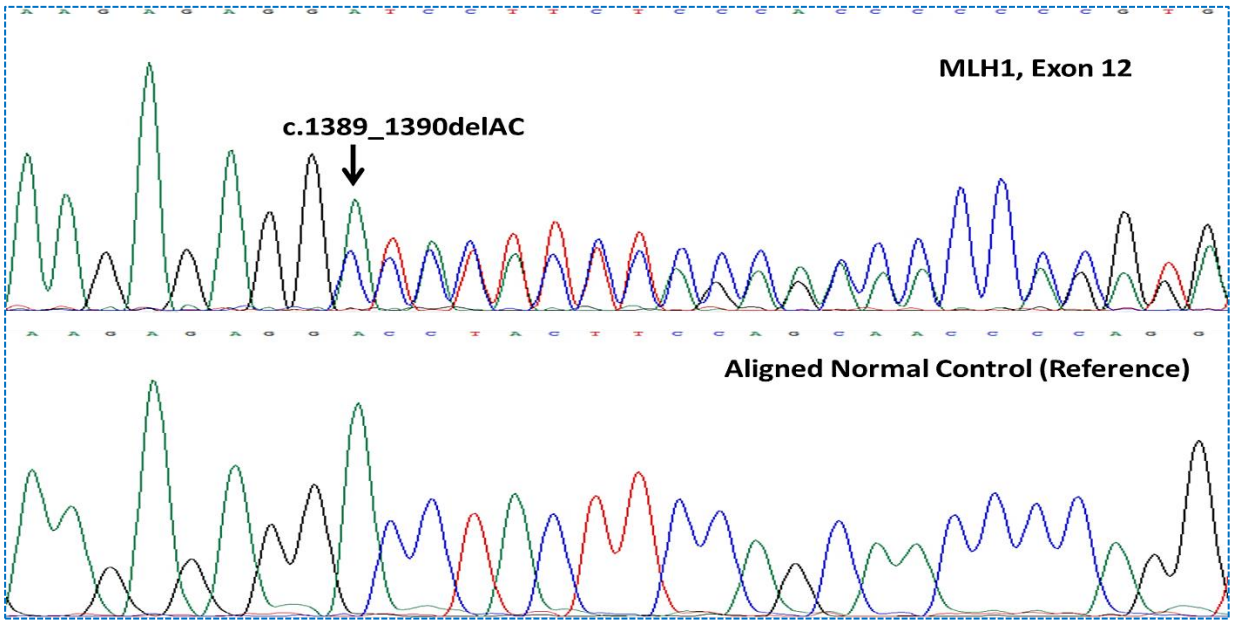


Fig 5.20 Chromatogram showing c.1389_1390delAC mutation

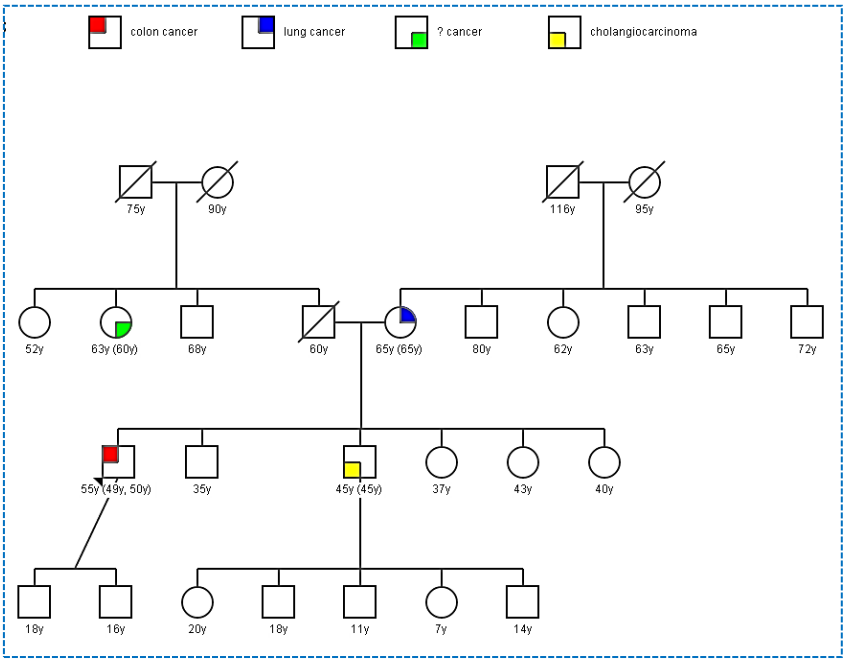


Fig 5.21 Pedigree of Hindu family from Bihar (Family ID: BH-1)

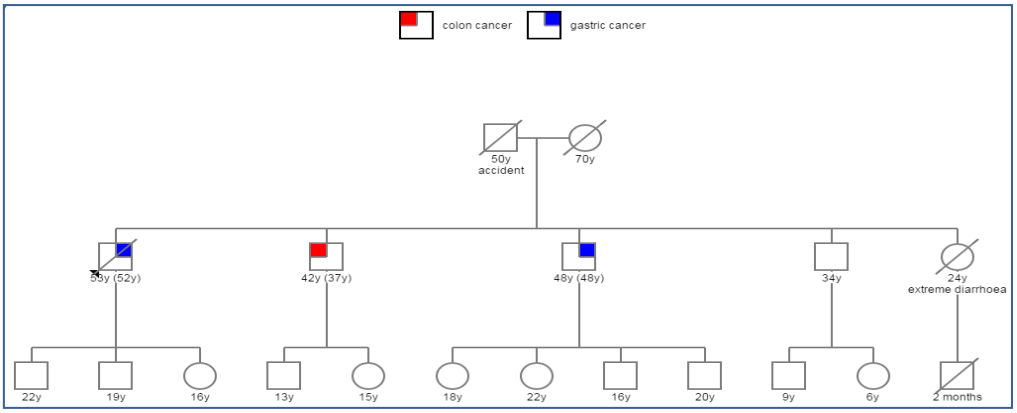


Fig 5.22 Pedigree of Hindu family from Bihar (Family ID: BH-2)

Table 5.11 Haplotype associated with c.1389_1390delAC mutation

Marker	Physical location (Mb)	BH-1	BH-2	Inferred haplotype
		Allele sizes (bp)		
D3S3564	42.37	200 - 202	204 - 204	
D3S1298	38.01	202 - 204	206 – 206	
D3S3623	37.40	**222** - 216	**222** – 214	**222bp**
D3S1007	37.03	**72** - 76	**72** – 72	**72bp**
D3S1611	37.02	**257** - 261	**257** – 251	**257bp**
D3S1561	36.44	**233** - 223	**233** – 251	**233bp**
D3S2411	36.31	**222** - 231	**222** – 233	**222bp**
D3S3718	36.12	**137** - 143	**137** - 137	**137bp**
D3S3512	34.55	**130** - 130	**130** – 134	**130bp**
D3S4153	34.35	**270** - 270	**270** – 270	**270bp**
D3S3518	33.65	150 – 158	156 – 162	-
D3S3936	33.0	115 – 115	115 – 115	-
D3S2432	32.13	138 - 148	130 – 152	-
Haplotype length: 3 Mb				
Haplotype frequency in controls: 0%				

** Markers shaded in grey in bold fonts are markers showing conserved haplotype

Allele sizes in bold are the common allele sizes shared by carriers

5.9 DISCUSSION

Founder mutations in specific populations are common in several Mendelian disorders. They are shared by seemingly unrelated families which inherited the founder mutation from a common ancestor several hundred to thousands years ago. Founder populations have always been instrumental in the study of molecular basis of specific Mendelian disorders and they are of great interest to molecular geneticists. Several founder populations have been studied extensively to understand the implications of founder events and mutations, such as the Finnish founder populations and Ashkenazi Jewish populations in which several founder mutations have been reported [272, 278].

Founder mutations have a profound impact in molecular diagnostics strategies in specific populations, where they can be assessed as the first screening step and, if positive, avoid further expensive genetic screening. An example of this is seen in Finnish population where the screening of the two founder mutations as a first step in the mutation analysis identifies nearly half of the mutation carriers in LS families [272]. In Lynch syndrome (LS) more than 50 founder pathogenic mutations have been described so far in the mismatch repair (MMR) genes (MLH1, MSH2, MSH6 and PMS2) however no MMR founder mutations are reported from Indian population. This is due to the limited number of studies conducted on the Indian LS patients that have only reported 28 MMR mutation carrier families so far [254-256].

In an attempt to characterize the mutation spectrum of MMR genes in a cohort of 81 Indian LS patients, we identified 5 MHL1 gene mutations that were present in two or more unrelated families. All five mutations were identified in families sharing the same geo-ethnic background which is why we speculated these mutations to be founder mutations in Indian population. Haplotyping analysis was conducted to prove the effect of these mutations which revealed that all the five mutations occur on a unique chromosomal background (haplotype) that is shared by all the carriers of that mutation and not present or rarely seen in the control population.

A frameshift mutation in Exon 2 of MLH1 gene (c.156delA) was identified as a founder mutation in the Shia Momin community from Patan district of Gujarat. Interestingly, this same mutation was identified in a Hindu family from Gujarat and

was found to share the same haplotype as the Shia Momim carriers of c.156delA mutation. This can be explained by going back to the history of Shia Momin community which showed that they are the progenies of Hindu ancestors who accepted Islam some 500 years ago through the preaching of some Sayyadnas and Pirs [294]. This leads to a theory that c.156delA mutation must have arose in a Hindu individual who spread this mutation over many generations to a number of Hindu and Shia Momin families who now report to be unrelated to each other.

Another mutation in Exon 3 of c.306G>T was identified in two ostensibly unrelated Indian Sindhi families who reported their ancestors to be from Sindh region of Pakistan. This mutation has been reported 14 times in the InSiGHT database as a Class 4 (likely pathogenic) mutation. We propose that this mutation must be reclassified as a Class 5 mutation as the mutation was found to segregate with the disease in 4 individuals from two families. Also, IHC analysis on tumor of one carrier showed loss of expression of MLH1 and PMS2 proteins. The geoethnic origin of the mutation carriers (InSiGHT database) was reported in only 2 cases (one reported as Asian and one as Indian) while for the others the geo-ethnicity is not known. There can be two possible scenario: Either the other 12 carriers are also Sindhi (Pakistan origin) thus going with the founder effect seen for this mutation in our study or this mutation is a recurrent mutation in other populations while it is a founder mutation in Sindhi Lynch syndrome patients. Similar observations has been reported for some mutations in MMR genes like c.942+3A>T in MSH2 gene is a recurrent mutation in various populations worldwide but it is a founder mutation in Newfoundland [275]. Another example includes c.388_389del mutation in MSH2 which is a founder mutation in Portuguese while it is a hotspot mutation in other population like Germany, Scotland and Argentina [276]. Furthermore, it is tempting to hypothesize that there will also be a fraction of LS patients of Sindhi ancestry in Pakistan that may harbor this same mutation with the unique haplotype found to be associated with c.306G>T mutation in our cohort.

One of the mutations in MLH1 Exon 12, c.1389_1390delAC was identified in three individuals from two Hindu Bihari families in our cohort. Though the haplotype analysis has proved the founder effect of this mutation, we refer this mutation as “putative founder mutation” as this mutation has been identified in only two families so far. The founder status of this mutation can be confirmed after this mutation is

identified in few more LS families particularly from the Bihar region. Similar approach was used by another study in which 2 mutations with proven founder effect that was identified in only 2 families each were defined as "putative founder mutations" [296].

The other two founder mutations identified in our study were also identified in specific geo-ethnic groups of India. Identification of MLH1 gene founder mutations in the Indian cohort of LS patients shows that founder mutations can occur not only in genetically isolated founder populations but also in geographically localized subsets of larger populations which have earlier been demonstrated by few other studies as well [277]. The frequency of founder mutations in Indian population is assumed to be high owing to the practice of consanguineous and endogamous marriages because of which specific mutations are identified in select Indian subgroups and ethnicities. In our cohort, altogether the five founder mutations accounted for 21% (16 out of 75) of all the MMR mutation carriers which makes it important to further take up such studies to identify founder mutations prevalent in Indian population. Based on this, the idea of establishing a pan-India founder mutation panel test is attractive which will reduce the time, efforts, and most importantly the cost of testing in various genetic diseases.

CHAPTER 6

Study of Stepwise Carcinogenesis in Hereditary Colorectal Cancer

6.1 Introduction to Colorectal Carcinogenesis

6.1.1 Overview of mechanism of carcinogenesis

Cellular processes like growth, apoptosis, differentiation, cell–cell interactions and cell–extracellular matrix interactions are strictly regulated by orchestration of defined signals that are released at specific rate at specific times inside the cell. Tumorigenesis is a complex process that involves acquisition of multiple sequential genetic events which allow escape from the tight constraints that control normal cells. The nature of these events includes gene mutations, loss of heterozygosity (LOH), promoter hypermethylation, gene amplification, gain of function mutations, transcriptional silencing or up-regulation and translocations that generate chimeric proteins with oncogenic activity**.**

Somatic mutations are acquired randomly. A mutation that provides survival advantage to the cell will be selected for and result in clonal expansion of the cell in which it arises. The clonal expansion depends on the rate of cell proliferation taking over the rate of cell death. Apoptosis here becomes rate limiting step at this particular stage of tumor evolution and the mutations that help the cell to evade apoptosis will be positively selected at this stage. Similarly, larger increases in tumor bulk without adequate vascularization would place evolutionary pressure for mutations stimulating angiogenesis. Consequently, the sequence of acquired mutations, or genetic pathway, in any developing tumor is a reflection of the constraints that become rate limiting at different stages in the evolution of the tumor.

6.1.2 Colorectal tumorigenesis

Colorectal carcinogenesis has been a subject of extensive research since many years with its various aspects being studied in detail. It is believed that carcinogenesis in the colorectum is driven by the presence of chemical carcinogens in the food, potentially harmful microbes, by the production of carcinogens generated by microbes, and by the induction of inflammation and modulation of the immune system. These factors are believed to create genotoxic stress to promote genetic and epigenetic alterations leading to cancer. Also, the total proliferation rate is 3-10 billion colonocytes per day which makes the colonic mucosa the organ with the highest proliferation rate of all organs in mammals [297]. This constant proliferation rate is the result of a fine balance between the functions of proto-oncogenes that increase proliferation and

tumor suppressor genes that decrease proliferation. The rapid turnover and immense number of mitoses in the colon result in accumulation of tens of thousands of mutations in the normal colonic mucosa per day [298]. Most of these mutations are repaired by the DNA repair systems inside the cells, yet a few mutations slip from these repair systems and lead to the accumulation of these mutations. Thus, studies on various aspects of colorectal carcinogenesis merge at the common conclusion that CRCs arise through stepwise accumulation of genetic and epigenetic alterations.

Colorectal cancers (CRCs) develop through an ordered series of events beginning with the transformation of normal colonic epithelium to an adenomatous intermediate and then ultimately adenocarcinoma, the so-called "adenoma-carcinoma sequence". This histological progression of colorectal cancer from adenoma to carcinoma is paralled by genetic pathway which parallels this transition was proposed for the first time by Fearon and Vogelstein in 1990 [299]. Since 1990, distinct molecular pathways underlying the malignant transformation of advanced adenomatous polyps into cancerous lesions have been studied [300]. These different pathways are based on independent genomic events leading to the loss of key cellular regulatory mechanisms causing proliferation, invasion and metastasis.

Baseline mutation rates are insufficient to account for the multiple mutations that are required for cancer to develop. The rate of mutations per nucleotide base pair is estimated to be as low as 10^{-9} per cellular generation [301]. Therefore, cancer cells must acquire some form of intrinsic genomic instability for the cancer to progress [302, 303]. Genomic instability is now universally recognized as an essential feature that drives colorectal tumorigenesis. Based on the type of genetic instability involved and the mutations in specific genes, there are three pathogenetic pathways that are currently known to be implicated in the development of CRCs [304] which include:

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

6.1.3 Chromosomal Instability Pathway (CIN)

Chromosomal instability refers to an accelerated rate of gains or losses of whole or large portions of chromosomes that result in karyotypic variability from cell to cell. The consequence of this is an imbalance in chromosome number (aneuploidy), subchromosomal genomic amplifications, and a high frequency of loss of heterozygosity (LOH). The combination of the mutations in tumor suppressor genes and oncogenes along with chromosomal instability is historically designated as the *"Chromosomal Instability Pathway" (CIN)* for colorectal carcinogenesis. This pathway is also sometimes referred to as the "*Canonical pathway"* and was earlier also stated as "Suppressor pathway", however, the use of this term is now being discouraged as tumor suppressor genes are also mutated in other pathways as well. CIN is the most common type of genomic instability occurring in nearly 80-85% of CRCs [305]. CIN pathway is essentially based on the adenoma-carcinoma sequence model of colorectal carcinogenesis that was put forth by Fearon and Vogelstein in 1990. The genetic changes found in CRC arising via the CIN pathway include APC mutations, KRAS mutation, TP53 mutations, and allelic loss of 18q [299]. The CIN pathway has been traditionally associated with CRC arising in adenomatous polyps [306].

❖ **Fearon and Vogelstein model**

This model was derived based on the studies on a total of 80 colorectal adenomas at various stages (early, intermediate and late adenoma) and 92 carcinoma samples taken from FAP and non-FAP patients [307]. The model stated that CRC tumorigenesis proceed through accumulation of a series of genetic alterations involving oncogenes (ras) and tumor suppressor genes (particularly those on chromosome 5q, 18q and 17p). Fig 6.1 summarizes the genetic model of colorectal tumorigenesis.

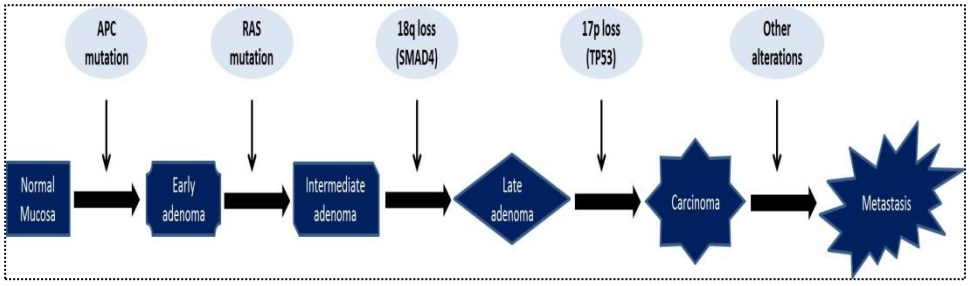


Fig 6.1 Genetic model of colorectal tumorigenesis (Fearon and Vogelstein Model)

Salient features of FV model:

1. Colorectal tumors appear to arise as a result of the mutational activation of oncogenes coupled with the mutational inactivation of tumor suppressor genes; the latter changes predominate.

2. Mutations in at least four to five genes are required for the formation of a malignant tumor.

3. Although the genetic alterations often occur according to a preferred sequence, the total accumulation of changes, rather than their order with respect to one another, is responsible for determining the tumor's biologic properties.

4. In some cases, mutant tumor suppressor genes appear to exert a phenotypic effect even when present in heterozygous state; thus, some tumor suppressor genes may not be "recessive" at the cellular level.

❖ **Key players in CIN pathway**

APC/CTNNB1

The activation of the Wnt signaling pathway through the genetic disruption of APC is regarded as the initiating event in colorectal cancer [308, 309]. Somatic APC mutations have been reported in 5% of dysplastic ACF, 30%–70% of sporadic adenomas, and in as many as 72% of sporadic tumor [310-312]. Gain-of-function mutations in β-catenin (CTNNB1) have been identified in as many as 50% of colon tumors with intact APC, reflecting the importance of the Wnt pathway [313, 314]. Without the APC/CTNNB1 mutation the adenoma-carcinoma pathway is unlikely to take place.

KRAS

KRAS is mutated in 30%–50% of CRCs [315, 316]. Point mutations in codon 12 and 13 account for nearly 90% of all the KRAS mutations identified in CRC [317]. Other than these, single nucleotide changes at codon 61 and codon 146 are also reported in CRC though at a lower frequency [318]. These mutations lock the enzyme in the GTP-bound, activated form, leading to constitutive activation of RAS downstream signalling. Activated RAS regulates multiple cellular functions through well-

described effectors. The best characterized effector is the Raf-MEK-ERK pathway [319]. RAS-GTP also binds the catalytic subunit of type I PI3Ks, which translocate to the plasma membrane to generate phosphoinositol lipids. Of note, PIK3CAis mutated in almost 20% of CRC (hotspot mutations in exons 9 and 20) [320, 321].

TP53

TP53 mutations are present in nearly all human cancers and CRC is no exception [322]. TP53 gene mutations are identified in up to 25% of adenomas, 50% of adenomas with invasive foci, and in 50%–75% of CRCs, defining its role in the transition from an adenoma to carcinoma [323]. TP53 is an important gene for maintaining genome stability and therefore it is regarded as "Guardian of the Genome" [324]. When replication errors or mutations occur, p53 stops or slows down the cell cycle in G1/S phase (before S-phase) and points out the DNA damage to the caretakers for repair [325]. If DNA damage is too extensive to be repaired, p53 induces apoptosis through the caspase pathway by shutting down mitochondrial function [326]. Most of the mutations in TP53 gene are missense mutations that essentially are clustered in the Exon 4-8 encoding for the DNA Binding Domain of the p53 protein [327].

18q loss

Allelic loss at chromosome 18q is one of the important genetic alterations in the CIN pathway (FV model) and is reported in about 70% of primary colorectal tumors [299]. The gene Deleted in Colorectal Carcinoma (DCC) was initially proposed to encode a colorectal tumor suppressor at this locus, but it encode for a cell surface receptor that is not associated with CRC development as shown by studies in mice [328, 329]. Other candidate genes at this locus include SMAD4 which is a component of TGFβ signaling pathway and has been shown to be implicated in CRC pathogenesis [330].

CIN: Cause or Consequence

One of the unanswered questions regarding CIN pathway is whether CIN is a cause or a consequence of the malignant process. The stage of tumorigenesis at which the CIN phenotype arises is controversial. Some reports have demonstrated that CIN, measured as allelic imbalance in specific chromosomal regions, occurs at very early stages of tumor development suggesting that CIN initiates tumorigenesis [331, 332].

However, others have contended that CIN is an acquired phenomenon and has a role in maintaining the tumorigenic process [333]. APC gene mutations have been proposed as a potential initiator of CIN [334]. In addition to its central role in Wnt pathway, APC has a role in cytoskeletal regulation through its ability to bind the plus ends of cytoplasmic and spindle microtubules and centrosomes through an EB1-binding domain [335].

6.1.4 Microsatellite Instability Pathway (MIN)

The microsatellite instability pathway represents another key milestone in the natural history of CRC. Microsatellite instability (MSI) results from a failure of the mismatch repair system (MMR) to correct base errors and maintain genomic stability as cells with abnormally functioning MMR accumulate errors rather than correcting them. About 15% of sporadic colorectal cancers are MSI-H and follow the MIN pathway [336]. There are two main ways in which a cell may fail to synthesize a mismatch-repair protein: 'two hit' inactivation (mutation or loss) of the gene and failure to express the gene due to promoter hypermethylation. In more than 90% of cases of sporadic MSI-H cancers, loss of mismatch-repair proficiency is due to inactivation of hMLH1 [337]. In summary, MSI in Lynch Syndrome is due to a germline mutation within a mismatch-repair gene followed by a second hit, whereas in sporadic colorectal cancer there is loss of expression of the gene (most commonly hMLH1).

Many genetic differences between MSI-positive and MSI-negative colorectal cancers have also been reported. Loss of heterozygosity is seen less frequently in MSI-H cancers, as are mutations of the tumor suppressor genes APC and p53, and the oncogene KRAS [338, 339]. By contrast, mutations within the polyA repeat of the transforming growth factor β type II receptor (TGFβRII) are very common in MSI-H cancers but uncommon in MSS cancers [340]. Moreover, accumulation of mutations in the mononucleotide repeats in the coding region of tumor suppressor genes like TGFβRII, IGFIIR, MSH3, MSH6, and BAX have been shown to be important in MIN pathway mediated carcinogenesis [341]. A list of genes with microsatellite instability in their coding sequences is given in Table 6.1.

While a fraction of sporadic CRC and Lynch Syndrome tumors both follows the MIN pathway for tumorigenesis, there exist various genetic differences between these two subgroups. The frequency of genetic alterations in Lynch Syndrome differs from that

in sporadic MSI-H cancers. For example, KRAS mutation occurs more frequently in Lynch Syndrome [342]. Also, promoter-region methylation is the major cause of hMLH1 inactivation in sporadic MSI-H colorectal cancer whereas hypermethylation of a mismatch-repair gene promoter is less frequently seen in Lynch Syndrome [343]. The apparent lack of APC mutations in MSI-H cancers is compensated by an increased rate of β-catenin mutation in Lynch Syndrome, but not in sporadic MSI-H tumors [344]. One study also demonstrated the presence of distinct patterns of KRAS mutations in cancers according to hMLH1 methylation status and germ line DNA MMR defects [342]. BRAF mutations (V600E) have been more frequently detected in sporadic MSI-H tumors than in Lynch Syndrome tumors [345]. Altogether these differences strongly suggest that MIN mediated carcinogenesis proceeds through alternative pathways in the sporadic and hereditary colorectal cancer (Lynch Syndrome).

Table 6.1 Genes with MSI in their coding sequence (adapted from Ref. 305)

Gene	Function	Coding repeat
AXIN-2	Wnt signaling	(A) 6*2, (G) 7, (C) 6
BAX	Proapoptotic factor	(G) 8
Caspase-5	Proapoptotic factor	(A) 10
CDX2	Homeobox transcription factor	(G) 7
FAS	Proapoptotic factor	(T) 7
MLH3	MMR	(A) 9
MSH3	MMR	(A) 8
MSH6	MMR	(C) 8
PTEN	Cell cycle	(A) 6*2
RAD-50	Response to DNA damage	(A) 9
BCL-10	Proapoptotic factor	(A) 8
BLM	Response to DNA damage	(A) 9
hG4-1	Cell cycle	(A) 8
IGFIIR	Growth factor receptor	(G) 8
MBD-4	DNA glycosylase and methyl CpG binding protein	(A) 10
AIM2	IFN inducible	(A) 10

6.1.5 CpG Island Methylator Phenotype Pathway (CIMP)

CpG sequences cluster in islands in the 5'region of many human genes. Methylation of cytosine within these islands leads to loss of gene expression. This type of widespread hypermethylation at CpG islands across various loci is referred to as CpG Island Methylator phenotype and this constitutes one of the important alternative pathways of colorectal carcinogenesis [346]. There is a loss of expression due to such hypermethylation of mismatch-repair genes as well as many other genes involved in colorectal cancer like p16 (cell cycle regulator), ER (growth suppressor), THBS1 (angiogenesis inhibitor), and MGMT (DNA repair gene) [347]. CpG-island methylation can occur at many sites. Two distinct types of methylation are seen in colorectal cancer, depending on the tissues affected. Type A (age-related) CpG-island methylation is seen in normal colonic mucosa as well as carcinomas and increases with age. It does not seem to be related to colorectal carcinogenesis [348]. Type C methylation occurs exclusively in colorectal cancer and is very probably involved in tumorigenesis [348].

Currently there are two main methylation marker panels. Depending on the markers used, 24%-51% of all CRCs are CIMP-positive [349, 350]. The first panel that was proposed, and the one that many studies have used, includes analysis of the promoter regions of the genes MLH1, p16, MINT 1, 2 and 3 [351]. The main limitation of the originally proposed CIMP panel was its inability to reliably classify cancers into well-defined subsets. Therefore, an alternative panel of markers (CACNA1G, CDKN2A, CRABP1, IGF2, MLH1, NEUROG1, RUNX3, and SOCS1) was proposed [352]. In this new panel, methylation is defined quantitatively and cancers are defined according to the percentage methylation ratio (PMR), with CIMP+ CRCs having a PMR of > 10 at 3 or more of the 5 sites [353]. Tumors with high-frequency type C methylation are labelled CIMP-positive tumors, and those with rare methylation are grouped with CIMP-negative tumors [354].

One study showed a positive correlation between MSI status and widespread CpG-island methylation, which is now considered as a characteristic of sporadic MSI-H colorectal cancer but not Lynch Syndrome [354]. MSI is infrequently seen in premalignant adenomas, except for those from individuals with Lynch Syndrome, in which the early development of MSI is associated with a germline mutation in a

mismatch-repair gene. About 45% of CIMP-positive colorectal cancers have MSI, but 100% of CIMP positive cancers in which hMLH1is methylated show MSI [348]. CIMP positivity seems to be an early event leading to methylation at multiple promoter regions and inactivation of certain genes. Methylation and inactivation of many tumor-suppressor genes, but not hMLH1, will result in CIMP-positive MSI-L or MSS cancers [348]. Few studies have reported distinct genetic differences between tumors that are positive and negative for CIMP [347]. Most CIMP+ cancers contain a BRAF mutation, but those that don't have a BRAF mutation often have a mutation in KRAS [353]. BRAF and KRAS mutations occur in a mutually exclusive fashion, suggesting that a pathway common to both is critical to the development of these cancers [353]. The RAS-RAF-MEK-ERK signaling pathway is important in apoptosis and in particular *anoikis*, the process of apoptosis following loss of the epithelial connection to the basement membrane [355]. Failure of *anoikis* is important in the development of hyperplastic polyps and serrated adenomas, which are the postulated precursors of CIMP+ CRCs [355]. Therefore it is hypothesized that the serrated polyps are precursors to CIMP+ CRCs. Also, p53 mutations are infrequent in CIMP+ tumors as compared to the CIMP- tumors [347]. Methylation of p16 is also exclusively in CIMP positive tumors (50% versus 0%) [347]. All these features suggest that CIMP mediated carcinogenesis is a genetically distinct pathway that involves alterations and mechanisms different from the CIN and MIN pathway.

6.1.6 Common and divergent pathways in colorectal tumorigenesis

The CIN and MIN tumors develop along a different genetic pathway though there is a little degree of overlap between the two pathways. Since environmental constraints are same for both the pathways, it is likely that although different sets of genes may be targeted by mutation, there will be an overlap of the signaling pathways affected. For example, inactivation of transcription growth factor-β (TGF-β) signaling may occur via mutation within the poly-A tract in the TGFβ receptor II (TGFβRII) in MIN tumours whilst mutation of the downstream signaling molecule SMAD4 may be the most efficient means to inactivate this pathway in CIN tumours. Similarly, TP53mutations are apparently less common in MIN tumours. However, a p53-induced effector of apoptosis, Bax, may be a common target in MIN tumours and thus may provide an alternative to TP53 mutations.

6.1.7 High-throughput approach to address heterogeneity in CRC

Next-generation sequencing technologies have revolutionized cancer genomics research by providing fast and accurate information thus facilitating study of somatic mutation profile of individual tumors. As most of the cancer-associated mutations are located in protein-coding regions, exome sequencing serves as a very useful technique for mutation discovery in cancer tissues. Indeed, several studies have successfully described the mutational background of different types of tumors by using this approach [356, 357].

Next-generation sequencing has also been employed to study the mutational landscapes in human colorectal cancers. The Cancer Genome Atlas (TCGA) characterized the genomes of 276 sporadic colorectal cancers, focusing almost exclusively on invasive cancers and metastatic tumours [358]. Other studies have analysed the exomes of microsatellite-instable (MSI) primary cancers [359], while Nikolaev et al [360] reported a detailed and comprehensive analysis of 24 sporadic adenomas. Despite these major initiatives to analyze sporadic colorectal cancers, little is known about the somatic mutational landscape of tumours from patients with hereditary forms of the disease which are associated with very high life time risk of developing CRC.

While colorectal cancer is the common manifestation of germline variants in various genes associated with the hereditary colorectal cancer syndromes, they initiate tumorigenesis in different ways, meaning that the landscape of somatic mutations, the genes that are mutated and the paths to malignancy, are likely to differ. Moreover, the phenotypes and responses to therapy in patients with germline mutations can vary, suggesting a complex interplay between the germline genetics of each patient and the somatic landscape of mutations acquired by the adenomas and tumours that they develop. It therefore becomes imperative to study somatic mutational landscape of tumors from hereditary CRC syndromes using high throughput approaches like whole exome sequencing.

6.2 GENESIS OF THE HYPOTHESIS OF THIS STUDY

Vast majority of CRC were assumed to arise from the well accepted Fearon and Vogelstein model of CRC carcinogenesis which is mono-directional. This model considered as the genetic paradigm of CRC tumorigenesis, implied that if some CRCs that cannot be explained by this model exist, they account for a significantly small proportion to warrant further consideration. Therefore, there was a failure to recognize that FV model cannot be the sole pathway that can explain carcinogenesis in all CRC tumors as they arise in the setting of different driver mutations, clinical and histopathological features. In the past decade however, many studies were taken up which showed that CRC is a heterogenous disease involving different genetic pathways giving rise to distinct subgroups of CRCs with specific clinical, pathological and molecular features [361, 362].

Inferences from the existing knowledge on colorectal carcinogenesis over this past decade have highlighted that there exists three alternative pathways for colorectal cancer development. Most of the sporadic CRC and FAP tumors develop along the CIN pathway (which essentially follows the FV model), nearly 15-20% of sporadic CRC and Lynch tumors follows the MIN pathway while CIMP constitutes the third pathway which explains carcinogenesis in a small subset of sporadic CRCs arising from conventional adenoma and a large fraction of CRCs arising from serrated/hyperplastic adenomas. However, it must be noted that most of these inferences are drawn from studies conducted on the sporadic colorectal cancers. Hereditary colorectal cancers however show certain notable differences in the molecular features of sporadic and hereditary CRCs even if they are assumed to follow the same pathway, as has been described above.

In the light of aforementioned points, we hypothesized that the hereditary colorectal cancers with known driver mutations follow distinct genetic pathways that are more heterogenous and complex to be explained by the FV model of stepwise carcinogenesis and the other alternative pathways described. One of the objectives of this project was therefore to understand the key genetic alterations in stepwise colorectal carcinogenesis using the unique model system of paired blood, normal mucosa, polyps and cancer samples from clinically characterized FAP patients with germline pathogenic APC mutations in all but 1 case.

6.3 METHODS

6.3.1 Patients and Samples

This study was conducted on paired samples from a total of 20 FAP patients after taking written informed consent from them. Of the 20 cases, fresh frozen tumor and normal tissue samples were available for 8 FAP cases while FFPE (Formalin Fixed Paraffin Embedded) tissue samples were available. A deleterious germline APC mutation was identified in 19 of the 20 cases studied here. Details of the FAP cases and the samples available from these cases are summarized in Table 6.2

Table 6.2 Details of the sample from FAP cases used in this study

Case ID	Germline APC mutation	Blood	Normal mucosa	Adenoma (polyp)	Tumor
Cases whose fresh frozen samples were available					
RS001	No germline APC mutation	+	+	+	+
RS002	c.2274delA	+	+	+	+
RS003	c.589delA	+	+	+	+
RS004	c.2805_2815del11	+	+	+	+
C5	c.4037C>G	+	+	+	-
C6	c.4612_4613delGA	+	+	+	-
C7	c.706C>T	+	+	-	+
C8	c.1861dupA	+	+	-	+
Cases whose FFPE blocks were available					
F1	c.1958+2T>C	+	-	+	+
F2	c.3183_3187del5	+	-	+	+
F3	c.3927_3931del5	+	-	+	+
F4	c.3259_3260delCT	+	-	+	-
F5	c.4216C>T	+	-	+	-
F6	c.3815C>A	+	-	+	-
F7	c.1620dupA	+	-	+	+
F8	c.3298dupT	+	-	+	-
F9	c.3183_3187del5	+	-	+	-
F10	c.3358 G>T	+	-	+	-
F11	c.4529delG	+	-	+	+
F12	c.4463T>G	+	-	+	-
Total (n)		**20**	**8**	**18**	**11**

Two parallel approaches were used to investigate the genetic alterations involved in the stepwise carcinogenesis in the FAP patients. This included:

1. Whole exome sequencing of the 4 FAP cases with complete paired samples (RS001, RS002, RS003 and RS004); whose blood, normal mucosa, adenoma and carcinoma samples were available for study.

2. Sanger sequencing of the key genes implicated in the Fearon-Vogelstein model which included KRAS, CTNNB1, TP53 and SMAD4 genes. In the 8 FAP cases (RS001, RS002, RS003, RS004, C5, C6, C7 and C8); entire coding region of KRAS, TP53, SMAD4 and exon 3 of CTNNB1 genes were analyzed. In the 12 FAP cases whose FFPE blocks were available (F1-F12); Exon 2 of KRAS (encompassing the most mutated codon 12 and 13), DNA Binding Domain (DBD) of TP53 (exon 4 to exon 8) and exon 3 of CTNNB1 genes were analyzed.

6.3.2 Sample preparation

For the fresh frozen tissue samples, genomic DNA was extracted using the PAXgene Tissue DNA Kit (Qiagen) as per the manufacturer's protocol. Genomic DNA from blood samples were extracted using the QiaAmp Blood DNA Mini Kit (Qiagen). For the FFPE blocks, FFPE DNA extraction kit from Qiagen was used. The extracted DNA was quantified using Nanodrop spectrophotometer and Qubit in some samples as per the standard protocol. The quality of DNA was also checked by loading the DNA on 1% agarose gel. Based on the DNA concentration, DNA was diluted to appropriate working concentration as per the requirement of various experiments.

6.3.3 Exome sequencing pipeline

Genomic DNA from the 16 samples (4 samples from 4 FAP cases) was sequenced in the National Institute of Biomedical Genomics (NIBMG), Kolkata, India using the Illumina HiSeq-1500 platform. Exome capture was performed with the Nextera Rapid Capture Expanded Exome Kit (which delivers 62 Mb of genomic content, including exons, UTRs, and miRNA) from Illumina. Tumor and polyp exomes were sequenced at 60X coverage (2x100 bp reads), and exomes from blood and normal mucosa were sequenced at 30X (2x100 bp reads).

FastQC software was used to assess the quality of the sequences (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc). BWA (Burrows Wheelers Aligner) software was used to align sequences over the human reference genome

hg19. To refine data, reads that were PCR or optical duplicates were discarded using Mark Duplicates command line from Picard tools package (http://picard.sourceforge.net/). Local realignment around indels defined in dbSNP were carried out using GATK tools (https://software.broadinstitute.org/gatk/) following which the quality score recalibration was carried out using GATK tools. Variant calling was executed with Varscan 2 software (varscan.sourceforge.net/) from the mpileup files generated for each samples. Somatic variants were called as Varscan runs comparative analysis between the tumor sample file and matched normal (germline) sample file and works by subtracting the germline variants form the tumor sample file.

All the somatic variants called by Varscan 2 were annotated using ANNOVAR software (annovar.openbioinformatics.org/). The variants that were present in the intronic or intergeneic regions 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 1000 genomes databases were also excluded. Further, the non-synonymous (missense) variants that was predicted to be benign by two or more in silico analysis tools amongst the various tools used to assess the pathogenicity of missense variants (SIFT, PolyPhen2, MutationAssessor, MutationTaster) were also excluded from the analysis.

6.3.3 Sanger Sequencing of key genes in FV model

The approach was to PCR amplify the entire coding region or the respective exons of KRAS, CTNNB1, TP53 and SMAD4 genes as outlined above using specific primers designed for these genes. PCR products were loaded on 1% agarose gel to check the amplification. The amplicons were then purified by EXO-SAP IT [USB] and Sanger sequencing of these purified products was then carried out using Big Dye Terminator kit [Thermo Fisher] on ABI 3500 and ABI 3730 machines [Thermo Fisher]. Data was analyzed using Chromas Lite software.

6.4 RESULTS

6.4.1 WES analysis of paired samples from FAP patients

We performed WES of 16 paired samples collected from 4 patients diagnosed with FAP with all but one having confirmed germline APC alterations (Tables 6.2). 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 the mean read length being 93bp for all samples. The FASTQC report image for one of the samples is given in Fig 6.2

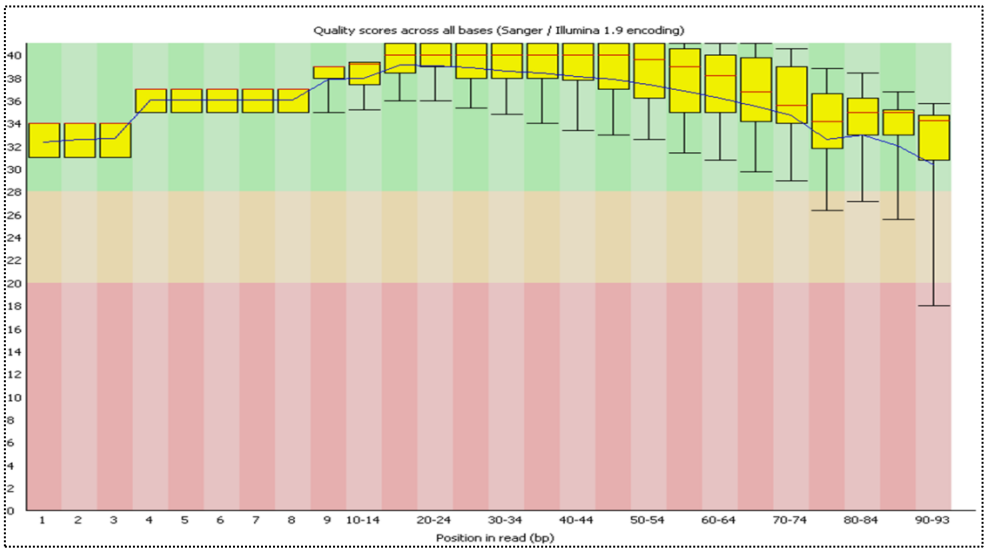


Fig 6.2 FASTQC analysis of raw data shows quality scores above 30 across all bases

After mapping the reads to the reference human genome hg19, nearly 76.70%–88.08% of reads were mapped to targets, and the mean target coverage was between 25X and 40X for the blood and mucosa samples whereas it was between 81X to 98X for the adenoma (polyp) and carcinoma samples (tumor). Detailed description of the exome coverage statistics are given in Table 6.3. More than 50% of the bases were covered at 20X in case of blood and the normal mucosa samples while more than 80% of the bases had coverage of 20X in case of the adenoma (polyp) and the carcinoma (tumor) samples.

Table 6.3 Coverage statistics for WES samples

Sr. No.	Sample	Total no. of reads	No. of uniquely mapped reads	Mapping rate	Mean target coverage	Percent bases covered at 10X	Percent bases covered at 20X
1	RS001B	97916976	75104909	76.70%	27.7	76.40%	57.40%
2	RS001N	81910210	64267559	78.40%	25.1	73.70%	52.70%
3	RS001P	230912716	181991162	78.81%	86.3	87.40%	80.20%
4	RS001T	194186216	169031114	87.04%	88.5	86.10%	81.10%
5	RS002B	98486126	83748781	85.03%	34.5	85.65%	69.86%
6	RS002N	92845952	76812662	82.73%	30.5	78.50%	61.80%
7	RS002P	243097466	214125560	88.08%	93.1	84.20%	80.50%
8	RS002T	236662010	200653467	84.78%	96.4	86.70%	82.90%
9	RS003B	88891496	71835408	80.81%	26.4	74.70%	54.60%
10	RS003N	80895412	65453221	80.91%	26.4	74.30%	54.20%
11	RS003P	189410336	163570851	86.35%	81.3	86.10%	81.10%
12	RS003T	212348920	184714812	86.98%	97.6	86.90%	82.70%
13	RS004B	122218922	100856556	82.53%	40.05	87.23%	74.25%
14	RS004N	121819110	95114641	78.07%	40.12	85.99%	72.30%
15	RS004P	253260950	216740034	85.57%	98.3	87.30%	83.50%
16	RS004T	256030258	211136352	82.46%	97.2	92.35%	89.03%

B: Blood, N: Normal mucosa, P: Polyp, T: Tumor

Exome analysis revealed a total of 20630 somatic variants within 12 samples analyzed (4 pairs of mucosa-adenoma-tumors) with the average number of variants per sample being 1716 (range: 1368-2231). Most of the variants were intergenic and intronic variants and the exonic variants accounted for only 3% of the total variants (635 exonic variants), of which 486 were single nucleotide variants (SNVs) and 149 were indels. Of the total 486 exonic SNVs present among the 12 samples, 241 were synonymous changes and the other 245 were non-synonymous variants including the splicing variants and nonsense variants. The overall mean frequency of exonic variants in all samples was 53 (range: 32-80). The mean number of exonic variants in the 4 polyp samples was 46 while it was 55 for the tumor samples. 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

Sr. No.	Sample	Total variants	Exonic variants	Exonic SNVs		Exonic Indels
				Synonymous	Non-synonymous	
1	RS001N	1512	39	11	18	10
2	RS001P	1778	54	26	17	11
3	RS001T	1696	64	28	26	10
4	RS002N	1409	36	18	7	11
5	RS002P	1771	41	15	10	16
6	RS002T	1469	42	13	16	13
7	RS003N	1368	81	37	32	12
8	RS003P	1627	58	28	14	16
9	RS003T	1552	53	18	21	14
10	RS004N	2120	76	23	38	15
11	RS004P	2231	32	13	10	9
12	RS004T	2097	59	11	36	12
Total		**20630**	**635**	**241**	**245**	**149**

While inspecting the somatic mutation spectrum in the normal mucosa, adenoma and carcinoma samples, we found that C:G>T:A transitions were the most significant changes in all the samples of all cases. The somatic mutation signatures by the type of nucleotide changes is given in Fig 6.3

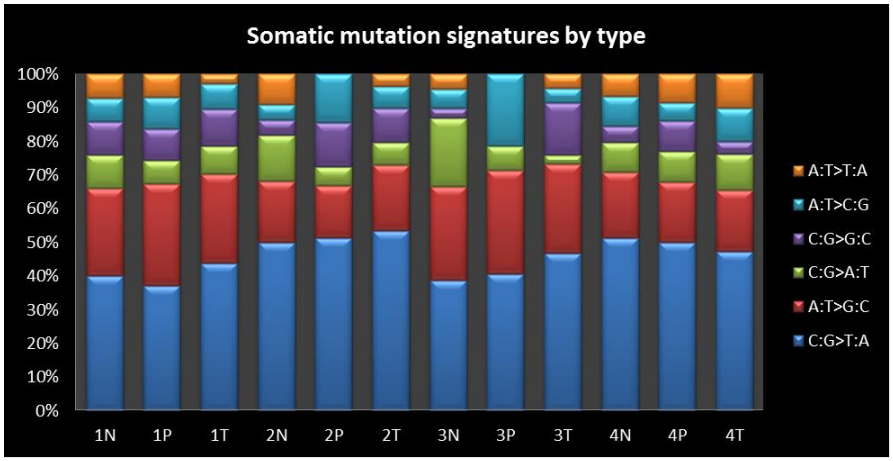


Fig 6.3 Somatic mutation spectrum by type of nucleotide change

Exonic variants were analyzed after carrying out the filtering steps, removing all the intronic and intergeneic sequences. Amongst the exonic variants, synonymous variants and the non-synonymous variants that were reported at high frequency in dbSNP, 1000g and ESP6500 databases were removed. Non-synonymous missense variants predicted to be benign by atleast two *in silico* tool were also filtered out. The potentially pathogenic indels nonsynonymous variants identified in each sample is given in Table 6.5 to Table 6.8

Table 6.5: Pathogenic variants in all samples of RS001 case

Sr. No.	Gene	Cytoband	cDNA change	Protein change	Type of mutation	rsID	Prediction by in silico tool*
RS001N							
1	TARBP1	1q42.2	c.3095delA	p.K1032fs	Frameshift	-	-
2	KIAA1211	4q12	c.924delT	p.R308fs	Frameshift	-	-
3	SPEG	2q35	c.1499G>A	p.R500H	Missense	-	S, P, L, MT
4	DYNC1LI1	3p22.3	c.709G>T	p.G237C	Missense	-	S, P, L, MT
5	SYNPR	3p14.2	c.638A>G	p.N213S	Missense	-	S, P, L, MT
6	KDM3B	5q31.2	c.5059G>T	p.A1687S	Missense	-	S, P, L, MT
7	TMEM63B	6p21.1	c.124C>G	p.P42A	Missense	-	S, P, L, MT
8	KMT2C	7q36.1	c.3340T>C	p.C1114R	Missense	rs200559566	S, P, L, MT
RS001P							
1	KRTAP9-1	17q21.2	c.457insA	p.C153X	Nonsense	-	-
RS001T							
1	NOTCH1	9q34.3	c.688insA	p.G230fs	Frameshift	-	-
2	MST1L	1p36.13	c.1925C>T	p.P642L	Missense	rs11545933	S, P, L, MT
3	CDHR1	10q23.1	c.2216T>G	p.V739G	Missense	-	S, P, L, MT
4	KLHL35	11q13.4	c.508C>G	p.L170V	Missense	-	S, P, L, MT
5	PARP4	13q12.12	c.3176A>G	p.Q1059R	Missense	rs77269056	S, P, L, MT
6	TRPV2	17p11.2	c.447G>T	p.Q149H	Missense	-	S, P, L, MT

*** S – SIFT, P – PolyPhen2, L – LRT, MT – MutationTaster**

Table 6.6: Pathogenic variants in all samples of RS002 case

Sr. No.	Gene	Cytoband	cDNA change	Protein change	Type of mutation	rsID	Prediction by in silico tool*
RS002N							
1	DNAJC11	1p36.31	c.343_344del	p.E115fs	Frameshift	rs374290353	-
2	ANTXR2	4q21.21	c.1069delG	p.A357fs	Frameshift	-	-
3	NADSYN1	11q13.4	c.1054delT	p.F352fs	Frameshift	-	-
4	LRRN2	1q32.1	c.G120A	p.W40X	Nonsense	-	-
5	VILL	3p22.2	c.301G>T	p.E101X	Nonsense	-	-
6	CC2D2A	4p15.32	c.3737C>A	p.P1246H	Missense	-	S, P, MT
7	KMT2C	7q36.1	c.851G>A	p.R284Q	Missense	rs2951027	S, P, MT
8	OR13C5	9q31.1	c.235C>T	p.P79S	Missense	rs7025570	S, P, MT
9	CCDC87	11q13.2	c.2120C>T	p.S707F	Missense	-	S, P, MT
10	TLK2	17q23.2	c.1802C>A	p.V601E	Missense	-	P, L, MT
RS002P							
1	CAMK2B	7p13	c.1553dupC	p.P518fs	Frameshift	-	-
2	A3GALT2	1p35.1	c.427C>A	p.L143I	Missense	-	S, P, MT
RS002T							
1	CNTRL	9q33.2	c.1328dupA	p.E443fs	Frameshift	-	-
2	NUP62	19q13.33	c.834delC	p.T278fs	Frameshift	-	-
3	PCMTD1	8q11.23	c.754C>T	p.R252X	Nonsense	rs75748152	-
4	SPAG8	9p13.3	c.426C>A	p.S142R	Missense	-	P, L, MT
5	PARP4	13q12.12	c.3116T>C	p.I1039T	Missense	rs73172125	S, P, MT

*** S – SIFT, P – PolyPhen2, L – LRT, MT – MutationTaster**

Table 6.7: Pathogenic variants in all samples of RS003 case

Sr. No	Gene	Cytoband	cDNA change	Protein change	Type of mutation	rsID	Prediction by in silico tool*
RS003N							
1	UBXN11	1p36.11	c.1462_1463del	p.G488fs	Frameshift	rs376181141	-
2	RBM45	2q31.2	c.31dupG	p.G10fs	Frameshift	rs146365140	-
3	CCNT1	12q13.11	c.497-2insT	-	Splicing	-	-
4	RAI1	17p11.2	c.837_838del	p.Q279fs	Frameshift	rs35068024	-
5	DPYD	1p21.3	c.2194G>A	p.V732I	Missense	rs1801160	S, P, L, MT
6	C2orf42	2p13.3	c.934+1G>T	-	Splicing	-	-
7	CATIP	2q35	c.625A>G	p.T209A	Missense	-	S, P, L, MT
8	PBRM1	3p21.1	c.236G>T	p.R79I	Missense	-	P, L, MT
9	PAK2	3q29	c.383A>G	p.K128R	Missense	rs78043821	S, P, L, MT
10	OTOP1	4p16.3	c.167A>G	p.K56R	Missense	rs78657691	S, P, L, MT
11	HLA-DRB1	6p21.32	c.100+2T>C	-	Splicing	rs200079869	
12	SPATA31A5	9q12	c.937G>C	p.E313Q	Missense	rs3873863	P, L, MT
13	ANKRD20A4	9q21.11	c.370G>A	p.A124T	Missense	-	S, P, MT
14	BRINP1	9q33.1	c.92A>G	p.D31G	Missense	-	S, P, MT
15	JMJD1C	10q21.3	c.2723C>A	p.P908H	Missense	-	S, P, MT
16	KRTAP5-4	11p15.5	c.335G>T	p.G112V	Missense	rs184758001	P, L, MT
RS003P							
1	DPYD	1p21.3	c.2194G>A	p.V732I	Missense	rs1801160	S, P, L, MT
2	MUC4	3q29	c.11897C>T	p.A3966V	Missense	-	S, P, MT
RS003T							
1	UBXN11	1p36.11	c.1462_1463del	p.G488fs	-	rs376181141	-
2	CAPN12	19q13.2	c.1459_1472ins	p.R487fs	-	-	-
3	DPYD	1p21.3	c.2194G>A	p.V732I	Missense	rs1801160	S, P, L, MT
4	TTBK1	6p21.1	c.1946C>G	p.P649R	Missense	rs35175743	S, P, L, MT
5	TMEM191C	22q11.21	c.494T>C	p.L165P	Missense	rs74899254	S, L, MT

*** S – SIFT, P – PolyPhen2, L – LRT, MT – MutationTaster**

Table 6.8: Pathogenic variants in all samples of RS004 case

Sr. No	Gene	Cytoband	cDNA change	Protein change	Type of mutation	rsID	Prediction by in silico tool*
RS004N							
1	HNRNPDL	4q21.22	c.203deG	p.R68fs	Frameshift	-	-
2	SDCCAG3	9q34.3	c.184-2insT	-	Splicing	-	-
3	SPNS3	17p13.2	c.7dupG	p.A2fs	Frameshift	-	-
4	UBXN11	1p36.11	c.1486G>T	p.G496C	Missense	rs193142354	S, P, L, MT
5	ASH1L	1q22	c.6547A>G	p.M2183V	Missense	-	S, P, L, MT
6	GNL1	6p21.33	c.1821C>A	p.C607X	Nonsense	-	-
7	PSPH	7p11.2	c.395C>A	p.A132D	Missense	-	S, P, L, MT
8	IFT140	16p13.3	c.4208G>A	p.R1403Q	Missense	rs113216558	S, P, L, MT
9	HOXB3	17q21.32	c.1022A>T	p.Y341F	Missense	-	S, P, L, MT
10	NME1	17q21.33	c.428A>G	p.H143R	Missense	-	S, P, L, MT
11	ANKRD24	19p13.3	c.813G>T	p.Q271H	Missense	-	S, P, L, MT
12	CKM	19q13.32	c.653+1G>A	-	Splicing	-	-
RS004P							
1	RAB6B	3q22.1	c.14delG	p.G5fs	Frameshift	-	-
2	SEC24D	4q26	c.1708-2insT	-	Splicing	rs35951660	-
3	UBXN11	1p36.11	c.1486G>T	p.G496C	Missense	rs193142354	S, P, L, MT
RS004T							
1	DCAF6	1q24.2	c.656A>G	p.Y219C	Missense	-	S, P, L, MT
2	PCDHB15	5q31.3	c.472C>T	p.R158W	Missense	-	S, P, L, MT
3	ARSI	5q32	c.971G>A	p.R324Q	Missense	rs147972506	S, P, L, MT
4	LPA	6q25.3	c.4973+1G>A	-	Splicing	rs79242985	-
5	GET4	7p22.3	c.803G>A	p.R268Q	Missense	-	S, P, L, MT
6	EGFR	7p11.2	c.2856G>T	p.M952I	Missense	-	S, P, L, MT
7	ARHGEF10	8p23.3	c.1496C>T	p.T499I	Missense	-	S, P, L, MT
8	FOXF1	16q24.1	c.166C>T	p.L56F	Missense	-	S, P, L, MT
9	INSM1	20p11.23	c.913A>C	p.S305R	Missense	-	S, P, L, MT
10	INSM1	20p11.23	c.920C>A	p.P307Q	Missense	-	S, P, L, MT
11	INSM1	20p11.23	c.925A>T	p.N309Y	Missense	-	S, P, L, MT
12	WWC3	Xp22.2	c.1934C>A	p.P645H	Missense	-	S, P, L, MT

*** S – SIFT, P – PolyPhen2, L – LRT, MT – MutationTaster**

Most of the variants identified in the samples from FAP cases were private events not shared between the cases. Furthermore, vast majority of the shared variants between the samples of one case and between the cases were intronic or intergeneic variants or exonic variants that were not predicted to be pathogenic. Genes harboring potentially pathogenic variants that were observed in more one sample included -

1. **KMT2C**

KMT2C also known as MLL3 is a histone lysine methyltransferase that is involved in transcriptional co-activation. Two different likely pathogenic missense variants were identified in the mucosa samples of RS001 and RS002.

2. PARP4

This gene encodes poly(ADP-ribosyl)transferase-like 1 protein, which is capable of catalyzing a poly(ADP-ribosyl)ation reaction. Two different likely pathogenic missense variants were identified in two tumor samples, RS001 and RS002.

3. UBXN11

UBX domain containing protein 11, also known as colorectal tumor associated antigen 1 (COA1). This protein is thought to be involved in the reorganization of actin cytoskeleton mediated by RND1, RND2 and RND3. A frameshift variant in UBXN11 gene was identified in our study in the mucosa and carcinoma samples of RS003 and a missense variant was identified in the mucosa and adenoma samples of RS004.

One of the significant observations of this study is that we did not identify any mutation in the known genes (KRAS, TP53, SMAD4) implicated in CRC tumorigenesis in the Fearon Vogelstein model.

6.4.2 Sanger sequencing of key genes in FV model in paired samples (FAP cases)

We sequenced the key genes implicated in CRC tumorigenesis in the FV model in a total of 8 normal mucosa, 18 adenoma and 11 carcinoma samples with paired blood samples taken from 20 FAP cases. The details of the samples used and the sequencing approach have been discussed above. A known recurrent mutation in KRAS gene G12V was identified in one carcinoma sample. Three deleterious (one frameshift and two likely pathogenic missense) mutations in TP53 genes were identified in 3 carcinoma samples. No mutation was identified in the CTNNB1 and SMAD4 genes. A summary of all the pathogenic and likely pathogenic variants identified in various genes in this study is given in Table 6.9

Table 6.9 Summary of pathogenic and likely pathogenic variants identified in this study

Gene	Blood	Normal mucosa	Adenoma	Carcinoma
KRAS	0/8	0/8	0/18	1/11
CTNNB1	0/8	0/8	0/18	0/11
TP53	0/8	0/8	0/18	3/11
SMAD4	0/8	0/8	0/18	0/11
UBXN11	0/4	2/4	1/4	1/4
PARP4	0/4	0/4	0/4	2/4
KMT2C/MLL3	0/4	2/4	0/4	0/4

6.5 DISCUSSION

Carcinogenesis is a multi-step process that involves the accumulation of numerous genetic and epigenetic alterations in cell that affects a limited number of pathways. According to the most widely accepted genetic model for colorectal cancer carcinogenesis, sequential accumulation of specific genetic alterations in various tumor suppressor genes (e.g. APC, SMAD4, or TP53) and oncogenes (e.g. CTNNB1, KRAS) are necessary for CRC development [299]. This model suggests that APC and KRAS mutations are an early event in CRC tumorigenesis leading to the formation of adenoma followed by TP53 alterations that marks the progression of adenoma into adenocarcinoma [299]. Though many studies have suggested the existence of alternative genetic pathways in colorectal cancer development [362, 363], it is important to note that Fearon Vogelstein's model that is the basis of CIN pathway is still considered as major pathway implicated in carcinogenesis of most sporadic CRC and FAP [363]. Moreover, the genetic pathways involved in CRC development are thought to be the same for hereditary and sporadic colorectal cancers. However, with clue from some previous studies [339, 342, 345], we hypothesized that the hereditary colorectal cancers with known driver mutations follow distinct genetic pathways that are more heterogenous and complex to be explained by the FV model of stepwise carcinogenesis. We used two parallel approaches to investigate the stepwise carcinogenesis in FAP using paired samples from these cases.

Exome sequencing of paired blood, mucosa, adenoma, and carcinoma samples taken from same patient with germline mutations in APC gene was carried out in 4 such FAP cases. WES for tumor and adenoma samples were carried out at 60X coverage while sequencing of blood and mucosa samples was done at a lower coverage of 30X.

This is because the tumor and adenoma samples are known to harbor a large heterogeneity and hence sequencing at a higher coverage is necessary to identify variants presents in smaller fractions of cells in tumor (rare variants). This is not the case in blood and mucosa samples and hence sequencing at lower coverage can be justified.. Whole exome sequencing revealed the presence of an average of 53 exonic variants per sample (range 32-80) which shows that FAP tumors belong to non-hypermutated CRCs. This is in accordance with the frequency of somatic variants reported in other studies for the non-hypermutated tumors [364, 365]. The hypermutated tumor shows a very high frequency of somatic variants and includes the tumors carrying germline mutations in the MMR, POLE and MUTYH genes [358]. All the samples showed an enrichment of C:G>T:A changes in our study which is a common feature seen in most of the colorectal cancers [366, 367]. As DNA methylation frequently occurs within the context of CpG sites, the enrichment of C:G->T:A transitions might be indicative of DNA hypermethylation being involved in tumorigenesis in these cases. The normal mucosa samples in all the 4 cases showed a high frequency of mutations which suggests that early genetic alterations starts accumulating in the normal mucosa which then leads to dysplastic changes later in life. This observation justifies the importance of prophylactic procto-colectomy in at-risk individuals.

Most of the variants identified in paired samples from FAP cases were private events which indicate the presence of tumor specific mutational landscapes. Most of the shared variants between samples of same case or between samples of different cases were located in the intronic and intergenic regions. The genes which had exonic variants that were shared between the samples included UBXN11, PARP4, and KMT2C. However, we could not check the frequency of mutations in these genes in other paired samples used for Sanger sequencing study due to the very low yield and poor quality of the DNA extracted from FFPE block samples. Nevertheless, these genes need to be investigated in a larger pool of paired samples from FAP to understand their involvement in colorectal carcinogenesis in hereditary and sporadic CRC. UBXN11 also known as COA-1 encodes for a colorectal antigen which has been shown to be involved in T-cell mediated immune response in colorectal cancer [368]. The role of this gene in CRC tumorigenesis remains elusive. As UBXN11

variants were identified in two out of the four mucosa samples studied, we speculate that this is an early event in colorectal tumorigenesis.

PARP4 is involved in DNA repair process but its exact mechanism of action in repair process is not known. PARP4 mutations have also been reported earlier in WES studies on sporadic colorectal cancers but it was not found to be recurrently mutated in CRCs [369]. Somatic mutations in PARP4 gene has been reported in 17 out of the 960 breast cancers reported in the TCGA database. Recently, a study has identified germline PARP4 mutations in 6 out of 14 cases who presented with two primary cancers of thyroid and breast (with wild type PTEN gene) [370]. These evidences suggest that PARP4 is an important tumor suppressor gene though its role in CRC tumorigenesis yet needs to be investigated. As PARP4 variants were identified in 2 out of the 4 carcinoma samples, we propose that PARP4 variants may be involved in progression of colorectal cancer.

KMT2C also known as MLL3 is a known tumor suppressor gene which has been reported to be involved in tumorigenesis of pancreatic cancers and glioblastomas [371]. Studies attempted to elucidate role of KMT2C in CRC pathogenesis have shown different results. While one study failed to establish its role in CRC [372], another study has shown that KMT2C is implicated in carcinogenesis of MSI-H CRCs [373]. It would therefore be interesting and important to investigate the role of KMT2C in carcinogenesis of APC driven CRCs. Of the four mucosa samples studied by WES KMT2C mutations were identified in 2 mucosa samples which suggests that may be it is an early event in colorectal tumorigenesis.

One of the striking observations of the WES analysis in this study is the absence of any TP53, KRAS and SMAD4 mutations in any of the samples of the 4 FAP cases studied. This is very intriguing, as variants in these have been described as key genetic alterations in many sporadic colorectal cancers and FAP both of which follow the same pattern of histological progression from adenoma to carcinoma and hence is assumed to follow the same genetic pathway. Furthermore, the other parallel approach used to study the key genetic alterations of FV model using Sanger sequencing also showed the rarity of KRAS mutations (identified in 1/11 carcinoma) and TP53 mutations (identified in 3/11 carcinoma samples) and absence of SMAD4 mutations.

KRAS mutations are defined as one of the most common alterations being detected in more than 50% of the late stage adenomas in sporadic and some forms of hereditary colorectal cancers [299, 317, 342]. Exome sequencing (WES) data in the 4 FAP cases showed that there are no KRAS, HRAS and NRAS gene mutations in any case at any stage. Taking together the WES and Sanger sequencing data, we show that KRAS mutation was not identified in any of the 18 adenoma samples studied. This observation is noteworthy in case of FAP as there are conflicting reports regarding the frequency of KRAS mutations in adenoma and carcinoma from FAP cases. Some studies report a frequency of KRAS mutations in accordance with the frequency proposed in sporadic cases [374, 375], while the others have shown a low frequency (6-18%) or absence of KRAS mutations in adenoma or carcinoma samples from FAP patients harboring germline APC mutations [376-378]. Absence of KRAS mutations in the adenoma samples and rarity in carcinoma samples raise a speculation that may be KRAS in not a key player in germline APC driven colorectal tumorigenesis.

All these reports signify a large heterogeneity between the sporadic and hereditary forms of colorectal cancer. The current concept of CRC tumor evolution through the three pathways seems insufficient as these pathways are based on a set of specific genetic or epigenetic alterations that are accumulated which explains deregulation of a defined set of pathways. This three pathway classification system failed to acknowledge the possibility of occurrence of combination of these pathways that may be instrumental in CRC tumorigenesis. Recently, researchers have turned their attention into the exploration of this possibility. Indeed, one recent study has addressed the issue of heterogeneity within the FAP tumors and showed that KRAS mutation frequency in FAP tumors vary according to the level of CIMP (intermediate or low) [379].

CHAPTER 7

Summary and Conclusions

Summary and Conclusion

Hereditary colorectal cancer syndromes account for 5-10% of all CRC cases with Lynch syndrome and Familial Adenomatous Polyposis being the most common forms of HCRC. These syndromes are widely studied in the Western population and most of the data on mutation spectrum of associated genes, clinical features, and the genotype-phenotype correlations seen in these syndromes is derived from Caucasian population. There has been a growth in research in this field in some East Asian population as well. On the contrary, these syndromes have rarely been investigated in the South Asian countries like India where the knowledge of molecular features can have serious implications in the risk management and genetic testing of these patients. No FAP cohort was reported from India so far. On the other hand, only 28 MMR gene mutation carrier families were reported by three studies from India. As the mutation spectrum of APC and MMR gene remains poorly defined in Indian LS and FAP patients, there exist no data on the recurrent and founder mutations in these genes that are prevalent in the Indian population. The present thesis work therefore worked towards the characterization of mutation spectrum of APC and MMR genes, and the associated clinical features seen in Indian FAP and LS patients.

The comprehensive investigation of all the five adenomatous polyposis associated genes in a well characterized Indian cohort of 47 FAP families confirmed the high frequency of APC mutations in classical FAP, MUTYH in AFAP cases and absence of NTHL1, POLD1 and POLE mutations in cases not showing syndromic features of PPAP or NAP. This is the first report of mutation analysis of adenomatous polyposis associated genes in Indian FAP cohort. Therefore, based on our data we proposed a pragmatic stepwise approach for molecular analysis of APC and other genes which can improve uptake of genetic testing for FAP in south Asian countries. Identification of a large number of novel APC mutations and genotype phenotype associations that are rare in the Caucasian population highlights the need for comprehensive phenotypic characterization and genetic analysis in large FAP cohorts from diverse geo-ethnic backgrounds in order to expand the current knowledge on these syndromes.

The genetic and clinical characterization of Indian Lynch syndrome patients have revealed a high mutation detection rate and identified novel and recurrent mutations

prevalent in Indian population. This study has shown that clinical criteria for diagnosis of Lynch Syndrome must be used in combination with a prescreening technique like IHC of MMR proteins that may increase the mutation detection rate and reduce the cost and time taken for genetic testing of MMR genes. The phenotypes in the MMR mutation carriers in our cohort largely concur with other studies reported on LS. One important conclusion regarding the phenotype in LS from this study however is that breast cancer may be considered as a part of LS tumor spectrum. Moreover, it is important to realize that further studies on even larger cohorts are necessary to create the evidence for effective screening methodologies and expand the existing knowledge on genotypic and phenotypic features of this syndrome.

In this study, we identified five recurrent mutations in MLH1 gene which was shown to be founder mutation with haplotyping studies. Together these founder mutations accounted for nearly 20% of all the mutation carrier families in our cohort underscoring the importance of characterization of mutation spectrum of genes associated with HCRC in distinct population. Our study has reiterated that founder mutations along with their implications in genetic testing, is also of great interest to evolutionary biologists and enables the scientists to trace the ancestry, migration, and growth of specific human populations over time as has been demonstrated earlier by previous studies. Finally, in the era of high-throughput technologies for molecular diagnostics of cancer, genetic testing for founder mutations can complement next generation sequencing (NGS) to most efficiently identify MMR gene mutations in any population.

Colorectal carcinogenesis has been shown to be a very complex and heterogenous process that is currently thought to be mediated by one of the three pathways described as CIN, MIN and CIMP. Most of the sporadic CRCs and FAP tumors are assumed to follow the CIN pathway which is largely based on the adenoma-carcinoma sequence model proposed by Fearon and Vogelstein in 1990. The whole exome sequencing and the Sanger sequencing data in our study however raised a possibility that the adenoma-carcinoma sequence model proposed by Fearon and Vogelstein does not explain a large fraction of FAP tumorigenesis. Also, this study provided evidence that KRAS alterations may not be a key player in CRC development in the setting of FAP syndrome. Furthermore, this study identified three recurrently mutated genes, UBXN11, KMT2C and PARP4. While the latter is

speculated to be involved in progression of carcinoma, alterations in the other two genes may be regarded as an early event in germline APC driven colorectal carcinogenesis. However, owing to the relatively smaller sample size which is one of the major limitations of this study, no definite conclusions could be made in this part of study. This project merits further studies with a large number of paired samples from well characterized FAP patients and using multi-modal approaches that explore the genetic and epigenetic alterations so that a clear picture of stepwise carcinogenesis in FAP tumors can be perceived.

CHAPTER 8

Bibliography

1. Cooper J P et al. Balancing cell growth and death. Curr Opin Cell Biol. 24 (6), 802-803 (2012).
2. Hunter, T. Signaling – 2000 and beyond. Cell 100, 113-127 (2000).
3. Ferlay J et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136, E359–E386 (2015).
4. Yancik R. Population aging and cancer: a cross-national concern. Cancer J 11, 437–41 (2005)
5. Torre L A et al. Global Cancer Statistics. Ca Cancer J Clin 65, 87-108 (2015).
6. Mallath M K et al. The growing burden of cancer in India: epidemiology and social context. Lancet Oncology Series Cancer burden and health systems in India 1 (2014).
7. Hanahan D et al. Hallmarks of Cancer: The Next Generation. Cell 144, 646-674 (2011)
8. Vogelstein B et al. Cancer genome landscapes. Science 339, 1546-1558 (2013).
9. Stratton M R et al. The cancer genome. Nature 458 (7239), 719–724 (2009).
10. Vogelstein B et al. Cancer genes and the pathways they control. Nature Medicine 10 (8), 789–799 (2004).
11. Croce C M. Oncogenes and cancer. NEJM 358 (5), 502–511 (2008)
12. Pylayeva-Gupta Y et al. RAS oncogenes: weaving a tumorigenic web. Nature Reviews Cancer 11 (11), 761–774 (2011).
13. Knudson A G. Mutation and cancer: statistical study of retinoblastoma. PNAS USA 68(4), 820-823 (1971).
14. Santarosa M et al. Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way. Biochimica et biophysica acta. 1654 (2), 105–122 (2004).
15. Berger et al. A continuum model for tumour suppression. Nature 476 (7359), 163-169 (2011).
16. Kinzler K W et al. Cancer susceptibility-genes. Gatekeepers and caretakers. Nature 386 (6627), 761-763 (1997).
17. Michor F et al. Dynamics of cancer progression. Nature Reviews Cancer 4, 197-205 (2004)
18. Jemal A et al. Cancer statistics, 2008. CA: a cancer journal for clinicians 58 (2), 71–96 (2008).
19. Butterworth A S et al. Relative and absolute risk of colorectal cancer for individuals with a family history: a metaanalysis. Eur J Cancer 42, 216-227 (2006).
20. Johns L E et al. A systematic review and meta-analysis of familial colorectal cancer risk. Am J Gastroenterol 96, 2992-3003 (2001)
21. Tsilidis K K et al. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. BMJ 350, g7607 (2015).

22. Boyle T et al. Physical activity and risks of proximal and distal colon cancers: a systematic review and meta-analysis. JNCI 104, 1548-1561 (2012).

23. Ma Y et al. Obesity and risk of colorectal cancer: a systematic review of prospective studies. PLoS One 8, e53916 (2013).

24. Kim E et al. Review of the association between meat consumption and risk of colorectal cancer. Nutr Res 33, 983-994 (2013).

25. Song M et al. Nutrients, foods, and colorectal cancer prevention. Gastroenterology 148, 1244-1260 (2015).

26. Ferrari P et al. Lifetime and baseline alcohol intake and risk of colon and rectal cancers in the European prospective investigation into cancer and nutrition (EPIC). Int J Cancer 121, 2065-2072 (2007).

27. Compton C C et al. The staging of colorectal cancer: 2004 and beyond. CA: a cancer journal for clinicians 54 (6), 295–308 (2004).

28. Taylor D P et al. Population-based family history–specific risks for colorectal cancer: a constellation approach. Gastroenterology138, 877–85 (2010).

29. Kerber R A et al. Frequency of familial colon cancer and hereditary nonpolyposis colorectal cancer (Lynch syndrome) in a large population database. Fam. Cancer 4, 239–44 (2005).

30. Lynch H T et al. Hereditary colorectal cancer. NEJM 348, 919–32 (2003).

31. Rustgi A K. The genetics of hereditary colon cancer. Genes Dev 21, 2525–38 (2007).

32. Winawer S J et al. The advanced adenoma as the primary target of screening. Gastrointest Endosc Clin N Am 12, 1-9, v (2002).

33. Fearon E R. Molecular genetics of colorectal cancer. Annual Rev. Pathol. Mech. Dis. 6, 479–507 (2011).

34. Lynch H T et al. Phenotypic and genotypic heterogeneity in the Lynch syndrome: diagnostic, surveillance and management implications. EJHG 14 (4), 390–402 (2006).

35. Kerr S E et al. APC Germline Mutations in Individuals Being Evaluated for Familial Adenomatous Polyposis. A Review of the Mayo Clinic Experience with 1591 Consecutive Tests. J Mol Diagnostics 15(1), 31-43 (2013).

36. Friedl W et al. Familial Adenomatous Polyposis: Experience from a Study of 1164 Unrelated German Polyposis Patients. Hereditary Cancer in Clinical Practice 3, 95–114 (2005).

37. Goldberg Y et al. Genetic features of Lynch syndrome in the Israeli population. Clinical Genetics 87, 549–553 (2015).

38. Mangold E et al. Spectrum and frequencies of mutations in MSH2and MLH1identified in 1,721 german families suspected of hereditary nonpolyposis colorectal cancer. Int. J. Cancer 116, 692–702 (2005)

39. Maharaj R et al. The impact of hereditary colorectal cancer on the Indian population. Indian J Cancer 51(4), 538-541 (2014).

40. Muto T et al. The evolution of cancer of the colon and rectum. Cancer 36 (6), 2251-2270 (1975).

41. Fearon E et al. A genetic model for colorectal tumorigenesis. Cell 61, 759-767 (1990)

42. Olschwang S et al. Alternative genetic pathways in colorectal carcinogenesis. PNAS USA 94, 12122–27 (1997)

43. Ionov Y et al. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 363, 558–61 (1997).

44. Fodde R et al. APC, signal transduction and genetic instability in colorectal cancer. Nature Reviews Cancer 1 (1), 55–67 (2001).

45. Pino M S et al. The chromosomal instability pathway in colon cancer. Gastroenterology 138, 2059-2072 (2010).

46. Grady W M. Genomic instability and colon cancer. Cancer metastasis reviews 23, 11–27 (2004).

47. Barber T D et al. Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. PNAS U.S.A 105 (9), 3443–3448 (2005).

48. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 487 (7407), 330–337 (2012).

49. Jass J R. HNPCC and sporadic MSI-H colorectal cancer: a review of the morphological similarities and differences. Familial Cancer 3 (2), 93–100 (2004).

50. Issa J P. CpG island methylator phenotype in cancer. Nature Reviews Cancer 4, 988-993 (2004).

51. Haydon A M M et al. Emerging pathways in colorectal cancer development. Lancet Oncology 3, 83-88 (2002).

52. Nazemalhosseini M E et al. The CpG island methylator phenotype (CIMP) in colorectal cancer. Gastroenterol Hepatol Bed Bench 6, 120-128 (2013).

53. Weisenberger D J et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nature Genetics 38, 787-793 (2006).

54. Bulow S et al. The history of familial adenomatous polyposis. Familial Cancer 5, 213–220 (2006).

55. Gardner E J. A genetic and clinical study of intestinal polyposis, a predisposing factor for carcinoma of the colon and rectum. AJHG 3, 167–76 (1951).

56. Bodmer W F et al. Localization of the gene for familial adenomatous polyposis on chromosome 5. Nature 328, 614–6 (1987).

57. Leppert M et al. The gene for familial polyposis maps to the long arm of chromosome 5. Science 238, 1411–1413 (1987).

58. Groden J et al. Identification and characterization of the familial adenomatous polyposis gene. Cell 66, 589–600 (1991).

59. Kinzler K W et al. Identification of FAP locus genes from chromosome 5q21. Science 253, 661–665 (1991).

60. Nieuwenhuis M H et al. Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): A review of the literature. Critical Reviews in Oncology/Hematology 61, 153-161 (2007).

61. Crail HW. Multiple primary malignancies arising in the rectum, brain and thyroid. Report of a case. US Navy Med Bull 49, 123–8 (1949).

62. Groen E J et al. Extra-Intestinal Manifestations of Familial Adenomatous Polyposis. Annals of Surgical Oncology 15(9), 2439–2450 (2008).

63. Kingston J E et al. Association between hepatoblastoma and polyposis coli. Arch Dis Child 58, 959–62 (1983).

64. Turcot J et al. Malignant tumors of the central nervous system associated with familial polyposis of the colon: report of two cases. Dis Colon Rectum 2, 465–468 (1959).

65. Spigelman A D et al. Upper gastrointestinal cancer in patients with familial adenomatous polyposis. Lancet 2, 783-785 (1989).

66. Bulow S et al. Study Group: Duodenal adenomatosis in familial adenomatous polyposis. Gut 53, 381-386 (2006).

67. Wallace M H et al. Upper gastrointestinal disease in patients with familial adenomatous polyposis. British J Surgery 85, 742-50 (1998).

68. Giardiello F M et al. Increased risk of thyroid and pancreatic carcinoma in familial adenomatous polyposis. Gut 34, 1394–1396 (1993).

69. Smith T G et al. Adrenal masses are associated with familial adenomatous polyposis. Dis Colon Rectum 43, 1739–1742 (2000).

70. Heiskanen I et al. Management of duodenal adenomas in 98 patients with familial adenomatous polyposis. Endoscopy 31, 412-416 (1999).

71. Burke C A et al. The utility of capsule endoscopy small bowel surveillance in patients with polyposis. American J Gastroenterology 100, 1498-1502 (2005).

72. Traboulsi E I et al. Prevalence and importance of pigmented ocular fundus lesions in Gardner's syndrome. NEJM 316, 661–667 (1988).

73. Half E et al. Familial Adenomatous Polyposis. Orphanet J Rare Diseases 4, 22-44 (2009).

74. Knudsen A L et al. Desmoid tumour in familial adenomatous polyposis. A review of literature. Fam Cancer 1, 111–119 (2001).

75. Sturt N J et al. Current ideas in desmoid tumors. Fam Cancer 5, 275-285 (2006).

76. Eccles D M et al. Hereditary desmoid disease due to a frameshift mutation at codon 1924 of the APC gene. AJHG 59, 1193-1201 (1996).

77. Wijn M A et al. Oral and maxillofacial manifestations of familial adenomatosis polyposis. Gardner's syndrome. Ned Tijdschr Tandheelkd 112, 340–344 (2005).

78. Kubo K et al. Widespread radiopacity of jaw bones in familial adenomatosis coli. J Craniomaxillofac Surg 17, 350–3 (1989).

79. Burger B et al. Prevalance of skin lesions in familial adenomatous polyposis: a marker for presymptomatic diagnosis? Oncologist 16(12), 1698-1705 (2011)

80. Giardiello F M et al. Nasopharyngeal angiofibroma in patients with familial adenomatous polyposis. Gastroenterology 105, 1550–1552 (1993).

81. Lynch H T et al. Attenuated familial adenomatous polyposis (AFAP). A phenotypically and genotypically distinctive variant of FAP. Cancer 76(12), 2427-2433 (1995).

82. Joslyn G et al. Identification of deletion mutations and three new genes at the familial polyposis locus. Cell 66, 601–13 (1991).

83. Beroud C et al. APC gene: database of germline and somatic mutations in human tumors and cell lines. Nucleic Acid Research 24, 121-124 (1996).

84. Fearon E R. Molecular genetics of colorectal cancer. Annual Reviews of Pathology: Mechanism of Disease 6, 479-507 (2011).

85. Dihlmann S et al. Dominant negative effect of the APC1309 mutation: A possible explanation for genotype-phenotype correlations in familial adenomatous polyposis. Cancer Research 59, 1857-1860 (1999).

86. Seeling J M et al. Regulation of beta-catenin signaling by the B56 subunit of protein phosphatase 2A. Science 283, 2089-2091 (1999).

87. Kawasaki Y et al. Asef, a link between the tumor suppressor APC and G-protein signaling. Science 289, 1194–1197 (2000).

88. Munemitsu S et al. Regulation of intracellular beta-catenin levels by the adenomatous polyposis coli (APC) tumor-suppressor protein. PNAS USA 92, 3046-3050 (1995).

89. Rubinfeld B et al. Loss of beta-catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. Cancer Research 57, 4624-4630 (1997).

90. Hart M J et al. Downregulation of beta-catenin by human Axin and its association with the APC tumor suppressor, beta-catenin and GSK3beta. Current Biology 8, 573-581 (1998).

91. Smith K J et al. Wild-type but not mutant APC associates with the microtubule cytoskeleton. Cancer Research 54, 3672-3675 (1994).

92. Su L K et al. APC binds to the novel protein EB1. Cancer Research 55, 2972-2977 (1995).

93. Ishidate T et al. The APC-hDLG complex negatively regulates cell cycle progression from G0/G1 to S phase. Oncogene 19, 365-372 (2000).

94. Henderson B R. Nuclear-cytoplasmic shuttling of APC regulates β-catenin subcellular localization and turnover. Nat Cell Biology 2, 653-660 (2000).

95. Narayan S et al. Role of APC and DNA mismatch repair genes in the development of colorectal cancers. Molecular cancer 2, 41-56 (2003).

96. Rosin-Arbesfeld R et al. The APC tumour suppressor has a nuclear export function. Nature 406, 1009-1012 (2000).

97. Fagotto F et al. Nuclear localization signal-independent and importin/karyopherin-independent nuclear import of beta-catenin. Current Biology 8, 181-190 (1998).

98. Neufeld K L. Nuclear APC. Adv Exp Med Biol 656, 13-29 (2009).

99. Schneikert J et al. The canonical wnt signalling pathway and its APC partner in colon cancer development. Gut 56, 417-425 (2007).

100. Nelson S et al. Interactions and functions of the adenomatous polyposis coli (APC) protein at a glance. Journal of Cell Science 126, 873–877 (2013).

101. Sieber O M et al. The adenomatous polyposis coli (APC) tumour suppressor – genetics, function and disease. Molecular Medicine Today 6, 462-469 (2000).

102. Fearnhead N S et al. The ABC of APC. Human Molecular Genetics 10(7), 721-733 (2001).

103. Kroboth K et al. Lack of adenomatous polyposis coli protein correlates with a decrease in cell migration and overall changes in microtubule stability. Mol Biol Cell 18, 910-918 (2007).

104. Ishidate T et al. The APC–hDLG complex negatively regulates cell cycle progression from the G0/G1 to S phase. Oncogene 19, 365–372 (2000).

105. Tetsu O et al. Beta-catenin regulates expression of cyclin D1in colon carcinoma cells. Nature 398, 422-426 (1999).

106. Morin P J et al. Apoptosis and APC in colorectal tumorigenesis. PNAS USA 93, 7950-7954 (1996).

107. Galiatsatos P et al. Familial adenomatous polyposis. American Journal of Gastroenterology 101, 385–398 (2006).

108. Kohler E M et al. Functional definition of the mutation cluster region of adenomatous polyposis coli in colorectal tumours. Human Molecular Genetics 17(13), 1978-1987 (2008).

109. Fostira F et al. Mutational spectrum of APC and genotype-phenotype correlations in Greek FAP patients. BMC Cancer 10, 389-395 (2010).

110. Half E et al. Familial Adenomatous Polyposis. Orphanet J Rare Diseases 4, 22-44 (2009).

111. Lamlum H et al. The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis. Nature Medicine 5, 1071-1075 (1999).

112. Aretz S et al. Frequency and parental origin of de novo APC mutations in familial adenomatous polyposis. EJHG 12(1), 52-58 (2004).

113. Nagase H et al. Correlation between the location of germ-line mutations in the APC gene and the number of colorectal polyps in familial adenomatous polyposis patients. Cancer Research 52, 4055–4057 (1992).

114. Caspari R et al. Familial adenomatous polyposis: mutation at codon 1309 and early onset of colon cancer. Lancet 343, 629–632 (1994).

115. Soravia C et al. Genotype–phenotype correlations in attenuated adenomatous polyposis coli. AJHG 62, 1290–301 (1998).

116. Al-Tassan N et al. Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. Nature Genetics 30, 227-232 (2002).

117. Papp J et al. Contribution of APC and MUTYH mutations to familial adenomatous polyposis susceptibility in Hungary. Familial Cancer 15 (1), 85-97 (2016).

118. Sieber O M et al. Multiple colorectal adenomas, classic adenomatous polyposis, and germline mutations in MUTYH. NEJM 348,791–799 (2003).

119. Farrington S M et al. Germline susceptibility to colorectal cancer due to base excision repair gene defects. AJHG 77, 112–119 (2005).

120. Vogt S et al. Expanded extracolonic tumor spectrum in MUTYH-associated polyposis. Gastroenterology 137(6), 1976-1985 (2009).

121. Palles C et al. Germline mutations in the proof-reading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nature Genetics 45(2), 136-144 (2013).

122. Valle L et al. New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis. Human Molecular Genetics 23(13), 3506-3512 (2014).

123. Bellido F et al. POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. Genetics in Medicine 18(4), 325-332 (2016).

124. Spier I et al. Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas. International Journal of Cancer 137, 320-331 (2015).

125. Weren R D et al. A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. Nature Genetics 47(6), 668-671 (2015).

126. Kuiper R P et al. NTHL1 defines novel cancer syndrome. Oncotarget 6(33), 34069-34070, (2015).

127. Rivera B et al. Biallelic NTHL1 mutations in a woman with multiple primary tumors. NEJM 373, 1985-1986 (2015).

128. Belhadj S et al. Delineating the phenotypic spectrum of the NTHL1-associated polyposis. Clinical Gastroenterology and Hepatology (2016).

129. Cao X et al. Singapore Familial Adenomatous Polyposis (FAP) Patients with Classical Adenomatous Polyposis but Undetectable APC Mutations Have Accelerated Cancer Progression. American Journal of Gastroenterology 101, 1-8 (2006).

130. Mongin C et al. Unexplained polyposis: a challenge for geneticists, pathologists and gastroenterologists. Clinical Genetics 81(1), 38-46 (2012).

131. Friedl W et al. Familial Adenomatous Polyposis: Experience from a Study of 1164 Unrelated German Polyposis Patients. Hereditary Cancer in Clinical Practice 3, 95–114 (2005).

132. Plawski A et al. APC gene mutations causing familial adenomatous polyposis in Polish patients. Journal of Applied Genetics 49(4), 407–414 (2008).

133. Rivera B et al. Clinical and genetic characterization of classical forms of familial adenomatous polyposis: a Spanish population study. Annals of Oncology 22, 903–909 (2011).

134. Chiang J M et al. Mutation analysis of the APC gene in Taiwanese FAP families: low incidence of APC germline mutation in a distinct subgroup of FAP families. Familial Cancer 9(2), 117-124 (2010).

135. Kim D W et al. Mutation Spectrum of the APC Gene in 83 Korean FAP Families. Human Mutation 26, 281-292 (2005).

136. Sheng J Q et al. APC gene mutations in Chinese familial adenomatous polyposis patients. WJG 16(12), 1522-1526 (2010).

137. Savithri H S et al. Predictive testing for familial adenomatous polyposis in a rural South Indian community. Clinical Genetics 58(1), 57-60 (2000).

138. Lima B R et al. Germinoma presenting as a fourth cranial nerve palsy in a patient with adenomatous polyposis coli (APC) gene mutation. Journal of American Association for Pediatric Ophthalmology 15(1), 71–73 (2011).

139. Patel N et al. A novel APC mutation defines a second locus for Cenani-Lenz syndrome. Journal of Medical Genetics 52(5), 317–321 (2015).

140. Tao H et al. Identification of 5 novel germline APC mutations and characterization of clinical phenotypes in Japanese patients with classical and attenuated familial adenomatous polyposis. BMC Research Notes 3, 305–313 (2010).

141. Torrezan G T et al. A novel SYBR-based duplex qPCR for the detection of gene dosage: detection of an APC large deletion in a familial adenomatous polyposis patient with an unusual phenotype. BMC Medical Genetics 13, 55–61 (2012).

142. Gavert N et al. Molecular Analysis of the APC Gene in 71 Israeli Families: 17 Novel Mutations. Human Mutation 19, 664–670 (2002).

143. Gomez-Fernandez N et al. Molecular analysis of the APC and MUTYH genes in Galician and Catalonian FAP families: a different spectrum of mutations? BMC Medical Genetics 10, 57–68 (2009).

144. De Rosa M et al. The Mutation Spectrum of the APC Gene in FAP Patients from Southern Italy: Detection of Known and Four Novel Mutations. Human Mutation 21, 655–656 (2003).

145. Jones S et al. Biallelic germline mutations in MUTYH predispose to multiple colorectal adenoma and somatic G:C–>T:A mutations. Human Molecular Genetics 11, 2961–2967 (2002).

146. Spier I et al. Exome sequencing identifies potential novel candidate genes in patients with unexplained colorectal adenomatous polyposis. Familial Cancer 15, 281–288 (2016).

147. Adam R et al. Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. American Journal of Human Genetics 99(2), 337–351 (2016).

148. Warthin A S. 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. Arch. Intern. Med. 12, 546–555 (1913).

149. Hauser I J et al. A further report on the cancer family of Warthin. Am J Cancer. 27, 434-449 (1936).

150. Lynch H T et al. Hereditary factors in cancer: study of two large Midwestern kindreds. Arch. Intern. Med. 117, 206–212 (1966).

151. Lynch H T et al. Cancer family "G" revisited: 1895–1970. Cancer 27, 1505–1511 (1971)

152. Douglas J A, et al. History and molecular genetics of Lynch syndrome in Family G: a century later. JAMA 294, 2195–2202 (2005).

153. Boland C R et al. Familial colonic cancer without antecedent polyposis. Ann. Intern. Med. 100, 700–701 (1984).

154. Boland C R. Evolution of the nomenclature for the hereditary colorectal cancer syndromes. Familial Cancer 4, 211–218 (2005).

155. Kalady M F et al. Adenomas in Lynch syndrome: don't be fooled by the "non" in hereditary nonpolyposis colorectal cancer. Familial Cancer 12(Suppl. 2), 53 (2013).

156. Lynch H T et al. Hereditary nonpolyposis colorectal cancer (Lynch syndromes I & II). Genetics, pathology, natural history, and cancer contro, Part I. Cancer Genetics and Cytogenetics 53(2), 143-160 (1991).

157. Watson P et al. Colorectal carcinoma survival among hereditary nonpolyposis colorectal carcinoma family members. Cancer 83(2), 259–266 (1998).

158. Stigliano V et al. Survival of hereditary non-polyposis colorectal cancer patients compared with sporadic colorectal cancer patients. J Exp Clin Cancer Res 27, 39-, (2008).

159. Vasen H F A et al. The International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer (ICG-HNPCC). Dis. Colon Rectum 34, 424–425 (1991).

160. Vasen H F A et al. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative Group on HNPCC. Gastroenterology 116, 1453–1456 (1999).

161. Rodriguez-Bigas M A et al. A National Cancer Institute workshop on hereditary nonpolyposis colorectal cancer syndrome: meeting highlights and Bethesda Guidelines. JNCI 89, 1758–1762 (1997).

162. Umar A et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. JNCI 96, 261–268 (2004).

163. Watson P et al. The tumor spectrum in the Lynch syndrome. Familial Cancer 4, 245-248 (2005).

164. Ryan S et al. Risk of prostate cancer in Lynch syndrome: a systematic review and meta-analysis. Cancer Epidemiol Biomarkers Prev. 23(3):437–449 (2014).

165. Kastrinos F et al. Risk of pancreatic cancer in families with Lynch syndrome. JAMA 302(16), 1790–1795 (2009).

166. Bonadona V et al. Cancer Risks Associated With Germline Mutations in MLH1, MSH2, and MSH6 Genes in Lynch Syndrome. JAMA 305 (22), 2304-2310 (2011).

167. Wolf A I et al. Historical review of Lynch syndrome. Journal of Coloproctology 33(2), 95-100 (2013).

168. Zhang T et al. Mismatch repair gone awry: Management of Lynch syndrome. Critical Reviews in Oncology/Hematology 93, 170-179 (2015).

169. Cohen S A et al. The genetic basis of Lynch syndrome and its implications for clinical practice and risk management. The Application of Clinical Genetics 7, 147-158 (2014).

170. Aarnio M et al. Cancer risk in mutation carriers of DNA mismatch repair genes. Int J Cancer 81, 214–218 (1999).

171. Watson P et al. The risk of extracolonic, extra-endometrial cancer in the Lynch syndrome. Int J Cancer 123, 444–449 (2008).

172. Engel C et al. Risks of less common cancers in proven mutation carriers with Lynch syndrome. Journal of Clinical Oncology 30, 4409–441 (2012).

173. 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 30, 958–964 (2012).

174. Naidoo R et al. The application of microsatellites in molecular pathology. Pathol Oncol Res 4(4), 310-315 (1998).

175. Farabaugh P J et al. Genetic studies of the lac repressor. VII. On the molecular nature of spontaneous hotspots in the lacI gene of Escherichia coli. J. Mol. Biol. 126, 847–857 (1978).

176. Murphy K M et al. Comparison of the Microsatellite Instability Analysis System and the Bethesda Panel for the Determination of Microsatellite Instability in Colorectal Cancers. Journal of Molecular Diagnostics 8(3), 305-311 (2006).

177. Aaltonen L A et al. Clues to the pathogenesis of familial colorectal cancer. Science 260, 812–816 (1993).

178. Ionov Y M et al. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 363, 558–561 (1993).

179. Thibodeau S N et al. Microsatellite instability in cancer of the proximal colon. Science 260, 816–819 (1993).

180. Aarnio M. Clinicopathological Features and Management of Cancers in Lynch Syndrome. Patholog Res Int. 2012, 350309 (2012).

181. 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 58, 5248–5257 (1998).

182. Kawakami H et al. MSI testing and its role in the management of colorectal cancer. Curr Treat Options Oncol. 16(7), 30 (2015).

183. Zhou X P et al. Determination of the replication error phenotype in human tumors without the requirement for matching normal DNA by analysis of mononucleotide repeat microsatellites.Genes Chromosomes Cancer 21, 101–7 (1998).

184. Hoang J M et al. BAT-26, an indicator of the replication error phenotype in colorectal cancers and cell lines. Cancer Research 57, 300–303 (1997).

185. Pyatt R et al. Polymorphic variation at the BAT-25 and BAT-26 loci in individuals of African origin. Implications for microsatellite instability testing. Am J Pathol 155, 349–353 (1999).

186. Bacher J W et al. Development of a fluorescent multiplex assay for detection of MSI-high tumors. Dis Markers 20, 237–250 (2004).

187. Murphy K M et al. Comparison of the microsatellite instability analysis system and the Bethesda panel for the determination of microsatellite instability in colorectal cancers. J Mol Diagn 8, 305–311 (2006).

188. Xicola R M et al. Performance of different microsatellite marker panels for detection of mismatch repair-deficient colorectal tumors. JNCI 7: 244–52.

189. Peltomäki P et al. Genetic mapping of a locus predisposing to human colorectal cancer. Science 260, 810–812 (1993).

190. Lindblom A et al. Genetic mapping of a second locus predisposing to hereditary nonpolyposis colorectal cancer. Nature Genet. 5, 279–282 (1993).

191. Fishel R et al. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 75, 1027–1038 (1993).

192. Leach F S et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell 75, 1215–1225 (1993).

193. Papadopoulos N et al. Mutation of a mutL homolog in hereditary colon cancer. Science 263, 1625–1629 (1994).

194. Bronner C E et al. Mutation in the DNA mismatch repair gene homologue hMLH1is associated with hereditary nonpolyposis colon cancer. Nature 368, 258–261 (1994).

195. Nicolaides, N C et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature 371, 75–80 (1994).

196. Liu T et al. The role of hPMS1 and hPMS2 in predisposing to colorectal cancer. Cancer Res. 61, 7798–7802 (2001).

197. Miyaki M et al. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. Nature Genet. 17, 271–272 (1997).

198. Lipkin S M et al. MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. Nature Genet. 24, 27–34 (2000).

199. Liu H X et al. The role ofhMLH3in familial colorectal cancer. Cancer Res. 63, 1894–1899 (2003).

200. Korhonen M K et al. The first functional study of MLH3 mutations found in cancer patients. Genes Chromosomes Cancer 47, 803–809 (2008).

201. Saini N. The journey of DNA Repair. Trends in Cancer 1(4), 215-216 (2015).

202. Loeb L A et al. Errors in DNA replication as a basis of malignant changes. Cancer Res. 34, 2311–2321 (1974).

203. Hsieh P et al. DNA mismatch repair: Molecular mechanism, cancer, and ageing. Mech Ageing Dev. 129(7-8), 391–407 (2008).

204. Li G M. Mechanisms and functions of DNA mismatch repair. Cell Res 18, 85–98 (2008).

205. Jiricny J. Mediating mismatch repair. Nature Genetics 24(1), 6-8 (2000).

206. Iaccarino I et al. hMSH2 and hMSH6 play distinct roles in mismatch binding and contribute differently to the ATPase activity of hMutSalpha. EMBO Journals 17(9), 2677-2686 (1998).

207. Strand M et al. Mutations in the MSH3 gene preferentially lead to deletions within tracts of simple repetitive DNA in Saccharomyces cerevisiae. PNAS USA 92, 10418–10421 (1995).

208. Palombo F et al. GTBP, a 160-kilodalton protein essential for mismatch binding activity in human cells. Science 268, 1912–1914 (1995).

209. Iyer R R et al. DNA mismatch repair: functions and mechanisms. Chem Rev. 106(2), 302-23 (2006).

210. Modrich P. Mechanisms in eukaryotic mismatch repair. J Biol Chem. 281(41), 30305-30309 (2006).

211. Langston L D et al. DNA Polymerase δ Is Highly Processive with Proliferating Cell Nuclear Antigen and Undergoes Collision Release upon Completing DNA. J Biol Chem. 283(43), 29522–29531 (2008).

212. Marshall B et al. Insertion of a short Alu sequence into the hMSH2 gene following a double cross over next to sequences with chi homology. Gene 174, 175–179 (1996).

213. Yan H et al. Conversion of diploidy to haploidy: individuals susceptible to multigene disorders may now be spotted more easily. Nature 403, 723–724 (2000).

214. Gille J J et al. Genomic deletions of MSH2and MLH1 in colorectal cancer families detected by a novel mutation detection approach. Br. J. Cancer 87, 892–897 (2002).

215. Clendenning M et al. Long-range PCR facilitates the identification of PMS2-specific mutations. Human Mutation 27, 490–495 (2006).

216. Wimmer K et al. PMS2 gene mutational analysis: direct cDNA sequencing to circumvent pseudogene interference. Methods Mol. Biol. 1167, 289–302 (2014).

217. Pritchard C C et al. ColoSeq provides comprehensive Lynch and polyposis syndrome mutational analysis using massively parallel sequencing. J. Mol. Diagn. 14, 357–366 (2012).

218. Nystrom-Lahti M et al. DNA mismatch repair gene mutations in 55 verified or putative kindreds with hereditary non-polyposis colorectal cancer. Hum Mol Genet 5, 763–769 (1996).

219. Wijnen J et al. Hereditary nonpolyposis colorectal cancer families not complying with the Amsterdam criteria show extremely low frequency of mismatch repair gene mutations. Am J Hum Genet 61, 329–335 (1997).

220. 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. AJHG 72, 1088-1100 (2003).

221. Van der Klift H et al. Molecular characterization of the spectrum of genomic deletions in the mismatch repair genes MSH2, MLH1, MSH6, and PMS2 responsible for hereditary nonpolyposis colorectal cancer (HNPCC). Genes Chromosomes Cancer 44:123–138 (2005).

222. Taylor C F et al. Genomic Deletions in MSH2 or MLH1 Are a Frequent Cause of Hereditary Non-Polyposis Colorectal Cancer: Identification of Novel and Recurrent Deletions by MLPA. Human Mutation 22, 428-433 (2003).

223. Peltomäki P et al. Mutations associated with HNPCC predisposition — update of ICG-HNPCC/INSiGHT mutation database. Dis. Markers 20, 269–276 (2004).

224. Thompson B A et al. Application of a five-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants lodged on the InSiGHT locus-specific database. Nature Genetics 46, 107–115 (2014).

225. Clark N et al. MutS Homologues hMSH4 and hMSH5: genetic variations, functions, and implications in human diseases. Current Genomics 14, 81–90 (2013).

226. Plazzer J P et al. The InSiGHT database: utilizing 100 years of insights into Lynch syndrome. Familial Cancer 12, 175–180 (2013).

227. Desai D C et al. Recurrent germline mutation in MSH2 arises frequently de novo. J Med Genet 37:646–652 (2000).

228. Lynch H T et al. A founder mutation of the MSH2 gene and hereditary nonpolyposis colorectal cancer in the United States. JAMA 291, 718–724 (2004).

229. Win A K et al. Determining the frequency of de novo germline mutations in DNA mismatch repair genes. J Med Genet 48:530–534 (2011).

230. Hitchins M et al. MLH1germline epimutations as a factor in hereditary nonpolyposis colorectal cancer. Gastroenterology 129, 1392–1399 (2005).

231. 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 15:25–35 (2013).

232. Hitchins M P et al. Inheritance of a cancer-associated MLH1 germ-line epimutation. NEJM 356:697–705 (2007).

233. Castillejo A et al. Prevalence of constitutional epimutations as a cause of Lynch syndrome in unselected versus selected consecutive series of patients with colorectal cancer. J Med Genet 52, 498–502 (2015).

234. 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. Nature Genetics 41, 112–117 (2009).

235. Niessen R C et al. Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. Genes Chromosomes Cancer 48:737–744 (2009).

236. Bandipalliam B et al. Clinical presentation correlates with the type of mismatch repair gene involved in hereditary nonpolyposis colon cancer. Gastroenterology 126, 936–937 (2004).

237. Vasen H F A et al. MSH2 mutation carriers are at higher risk of cancer than MLH1 mutation carriers: a study of hereditary nonpolyposis colorectal cancer families. J Clin Oncol 19, 4074–4080 (2001).

238. Baglietto L et al. Risks of Lynch syndrome cancers for MSH6 mutation carriers. J Natl Cancer Inst 102:193–201 (2010).

239. Ten Broeke S W et al. Lynch syndrome caused by germline PMS2 mutations: delineating the cancer risk. J Clin Oncol 33:319–325 (2014).

240. Hendriks Y M C et al. Cancer risk in hereditary nonpolyposis colorectal cancer due to MSH6 mutations: impact on counselling and surveillance. Gastroenterology 127, 17–25 (2004).

241. Lynch H T et al. Milestones of Lynch syndrome: 1895–2015. Nature Reviews Cancer 15, 181-194 (2015).

242. Kempers M J E et al. Risk of colorectal and endometrial cancers in EPCAM deletion-positive Lynch syndrome: a cohort study. Lancet Oncol. 12, 49–55 (2011).

243. Lynch H et al Lynch syndrome associated extracolonic tumors are rare in two extended families with the same EPCAM deletion. Am. J. Gastroenterol. 106, 1829–1836 (2011).

244. 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. 15(1), 25–35 (2013).

245. Peltomaki P. Update on Lynch syndrome genomics. Familial Cancer 2016.

246. Smith M J et al. The Contribution of Whole Gene Deletions and Large Rearrangements to the Mutation Spectrum in Inherited Tumor Predisposing Syndromes. Human Mutation 37(3), 250-256 (2016).

247. Shia J et al. Lynch syndrome-associated neoplasms: a discussion on histopathology and immunohistochemistry. Familial Cancer 12(2), 241–260 (2013).

248. Rosty C et al. Germline mutations in PMS2 and MLH1 in individuals with solitary loss of PMS2 expression in colorectal carcinomas from the Colon Cancer Family Registry Cohort. BMJ Open 6, 1-7 (2016).

249. Zighelboim I et al. Epitope-positive truncating MLH1 mutation and loss of PMS2: implications for IHC-directed genetic testing for lynch syndrome. Familial Cancer 8, 501-504 (2009).

250. Hampel H et al. Feasibility of screening for Lynch syndrome among patients with colorectal cancer. J. Clin. Oncol. 26, 5783–5788 (2008).

251. Shia J. Immunohistochemistry versus Microsatellite Instability Testing For Screening Colorectal Cancer Patients at Risk For Hereditary Nonpolyposis Colorectal Cancer Syndrome. J Mol Diagn. 10(4): 293–300 (2008).

252. Mangold E et al. Spectrum and frequencies of mutations inMSH2and MLH1identified in 1,721 german families suspected of hereditary nonpolyposis colorectal cancer. International Journal of Cancer 116, 692-702 (2005).

253. Kurzawski G et al. Germline MSH2 and MLH1 mutational spectrum including large rearrangements in HNPCC families from Poland (update study). Clinical Genetics 69, 40-47 (2006).

254. Bhasyam M et al. Evidence for presence of mismatch repair gene expression positive Lynch syndrome cases in India. Molecular Carcinogenesis 54 (12): 1807-1814 (2015).

255. Rajender S et al. R659X mutation in the MLH1 gene in hereditary non-polyposis colorectal cancer (HNPCC) in an Indian extended family. IJMR 131: 64-70 (2010).

256. Rajkumar T et al. Mutation analysis of hMSH2 and hMLH1 in colorectal cancer patients in India. Genetic Testing 8(2): 157-62 (2004).

257. Loizidou M A et al. The Mutational Spectrum of Lynch Syndrome in Cyprus. Plos One 9 (8), e105501 (2014).

258. Pinol V et al. Accuracy of Revised Bethesda Guidelines, Microsatellite Instability, and Immunohistochemistry for the Identification of Patients With Hereditary Nonpolyposis Colorectal Cancer. JAMA 293(16), 1986-1994 (2005).

259. 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). Human Mutation 27, 388 – 392 (2006).

260. Desai D C et al. Recurrent germline mutation in MSH2 arises frequently de novo. J Med Genet 37, 646–52 (2000).

261. Froggatt N J et al. A common MSH2 mutation in English and North American HNPCC families: origin, phenotypic expression, and sex specific differences in colorectal cancer. J Med Genet 36, 97–102 (1999).

262. Fokkema I F et al. LOVD v.2.0: the next generation in gene variant databases. Human Mutation 32, 557–563 (2011).

263. Froggatt N J et al. A frequent hMSH2 mutation in hereditary non-polyposis colon cancer syndrome. Lancet 345, 727 (1995).

264. Ashley-Koch A et al. Sickle hemoglobin (HbS) allele and sickle cell disease: a HuGE review. Am J Epidemiol 151, 839–845 (2000).

265. Slatkin M. A Population-Genetic Test of Founder Effects and Implications for Ashkenazi Jewish Diseases. AJHG 75(2), 282-293 (2004).

266. Lopez-Escribano H et al. Balearic archipelago: three islands, three beta-thalassemia population patterns. Clin Genetics 83, 175–180 (2013).

267. Morral N et al. The origin of the major cystic fibrosis mutation (delta F508) in European populations. Nature Genetics 7, 169–175 (1994).

268. Tamura D et al. Founder mutations in xeroderma pigmentosum. J Invest Dermatol 130, 1491–1493 (2010).

269. Janavicius R. Founder BRCA1/2 mutations in the Europe: implications for hereditary breast-ovarian cancer prevention and control. EPMA J. 1(3), 397–412 (2010).

270. Levy-Lahad E et al. Founder BRCA1 and BRCA2 mutations in Ashkenazi Jews in Israel: frequency and differential penetrance in ovarian cancer and in breast-ovarian cancer families. AJHG 60(5), 1059–1067 (1997).

271. Drayna D. Founder mutations. Scientific American 293(4), 78-85 (2005).

272. Moisio A L et al. Age and origin of two common MLH1 mutations predisposing to hereditary colon cancer. AJHG 59, 1243–1251 (1996).

273. Ponti G et al. Mismatch repair genes founder mutations and cancer susceptibility in Lynch syndrome. Clin Genet 87, 507–516 (2015).

274. Nystrom-Lahti M et al. Founding mutations and Alu-mediated recombination in hereditary colon cancer. Nat Med 1, 1203–1206 (1995).

275. Desai D C et al. Recurrent germline mutation in MSH2 arises frequently de novo. J Med Genet 37, 646–652 (2000).

276. Pinheiro M et al. The MSH2 c.388_389del mutation shows a founder effect in Portuguese Lynch syndrome families. Clin Genet 84, 244–250 (2013).

277. Tomsic J et al. An American founder mutation in MLH1. Int J Cancer 130, 2088–2095 (2012).

278. Foulkes W D et al. The founder mutation MSH2*1906G→C is an important cause of hereditary nonpolyposis colorectal cancer in the Ashkenazi Jewish population. AJHG 71, 1395–1412 (2002).

279. Raskin L et al. Characterization of two Ashkenazi Jewish founder mutations in MSH6 gene causing Lynch syndrome. Clin Genet 79, 512–522 (2011).

280. Goldberg Y et al. Lynch Syndrome in high risk Ashkenazi Jews in Israel. Fam Cancer 13, 65–73 (2014).

281. Clendenning M et al. Origins and prevalence of the American Founder Mutation of MSH2. Cancer Res 68, 2145–2153 (2008).

282. Chan T L et al. MSH2 c.1452-1455delAATG is a founder mutation and an important cause of hereditary nonpolyposis colorectal cancer in the southern Chinese population. AJHG 74, 1035–1042 (2004).

283. Cann R L Genetic clues to dispersal in human populations: Retracing the past from the present. Science 291(5509), 1742–1748 (2001).

284. Misra V N. Stone age in India: An ecological perspective. Man and Env.14, 17–64 (1992).

285. Basu A et al. Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. PNAS 113(6), 1594 – 1599 (2016).

286. Reich D et al. Reconstructing Indian Population History. Nature 461(7263), 489–494 (2009).

287. Boivin N. Anthropological, historical, archaeological and genetic perspectives on the origins of caste in South Asia. In: Petraglia, MD.; Allchin, B., editors. The evolution and history of human populations in South Asia. Springer 341-362 (2007).

288. Witzel M. Substrate languages in Old Indo-Aryan (Rgvedic, Middle and Late Vedic). Electronic J. Vedic Studies 5, 1–67 (1999).

289. Moorjani P et al. Genetic Evidence for Recent Population Mixture in India. AJHG 93, 422–438 (2013).

290. Basu A et al. Ethnic India: A Genomic View, With Special Reference to Peopling and Structure. Genome Research 13:2277–2290 (2003).

291. Ankala A et al. Clinical Applications and Implications of Common and Founder Mutations in Indian Subpopulations. Human Mutation 36, 1–10 (2015).

292. Haldane J B. A defense of beanbag genetics. Perspect Biol Med 7, 343–359 (1964).

293. Singh J R et al. Directory of Human Genetic Services in India—2007. Int J Hum Genet 10, 187–192 (2010).

294. Ankala A et al. Ancestral founder mutations in calpain-3 in the Indian Agarwal community: historical, clinical, and molecular perspective. Muscle Nerve 47, 931–937 (2013).

295. Das R. Momins: Community of India. In: Singh N K and Khan A M, editors. Encyclopaedia of the World Muslims: Tribes, Castes and Communities, Volume 1. Global Vision Pub House, 1017-1027 (2001).

296. Chong G et al. High Frequency of Exon Deletions and Putative Founder Effects in French Canadian Lynch Syndrome Families. Human Mutation 30, 797-812 (2009).

297. Potten CS et al. Measurement of in vivo proliferation in human colorectal mucosa using bromodeoxyuridine. Gut 33, 71-78 (1992).

298. Barnes DE, Lindahl T. Repair and genetic consequences of endogenous DNA base damage in mammalian cells. Annu Rev Genet. 2004;38:445–476

299. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 61, 759-767 (1990).

300. Tomlinson I et al. A comparison of the genetic pathways involved in the pathogenesis of three types of colorectal cancer. J Pathol.184,148–152 (1998).

301. Albertini R et al. In vivo somatic mutations in humans: measurement and analysis. Annu Rev Genet. 24, 305–326 (1990).

302. Loeb LA et al. Multiple mutations and cancer. PNAS U S A. 100, 776–781 (2003).

303. Lengauer C et al. Genetic instabilities in human cancers. Nature 396, 643–649 (1998).

304. Shen L et al. Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. PNAS U S A. 104, 18654–18659 (2007).

305. Moran A et al. Differential colorectal carcinogenesis: Molecular basis and clinical relevance. World J Gastrointest Oncol 2(3), 151-158 (2010).

306. Jass J R et al Emerging concepts in colorectal neoplasia. Gastroenterology 123, 862–876 (2002).

307. Vogelstein B et al. Genetic alterations during colorectal-tumor development. NEJM 319, 525–532 (1988).

308. Powell S et al. APC mutations occur early during colorectal tumorigenesis. Nature 359, 235–237 (1992).

309. Kinzler K W and Vogelstein B. Lessons from hereditary colorectal cancer. Cell 87, 159–170 (1996).

310. Otori K et al. Infrequent somatic mutation of the adenomatous polyposis coli gene in aberrant crypt foci of human colon tissue. Cancer 83, 896–900 (1998).

311. Miyaki M et al. Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors. Cancer Res. 54, 3011–3020 (1994).

312. Cottrell S et al. Molecular analysis of APC mutations in familial adenomatous polyposis and sporadic colon carcinomas. Lancet 340, 626–630 (1992).

313. Sparks A et al. Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. Cancer Res. 58, 1130–1134 (1998).

314. Samowitz W et al. Beta-catenin mutations are more frequent in small colorectal adenomas than in larger adenomas and invasive carcinomas. Cancer Res. 59, 1442–1444 (1999).

315. Santini D, Loupakis F, Vincenzi B, et al. High concordance of KRAS status between primary colorectal tumors and related metastatic sites: implications for clinical practice. Oncologist 13, 1270–1275 (2008).

316. Vaughn C P et al. Frequency of KRAS, BRAF, and NRAS mutations in colorectal cancer. Genes Chromosomes Cancer 50(5), 307-12 (2011).

317. Zhang J et al. Molecular spectrum of KRAS, NRAS, BRAF and PIK3CA mutations in Chinese colorectal cancer patients: analysis of 1,110 cases. Scientific Reports 5, 18678-18685 (2015).

318. Loupakis F et al. KRAS codon 61, 146 and BRAF mutations predict resistance to cetuximab plus irinotecan in KRAS codon 12 and 13 wild-type metastatic colorectal cancer. Br J Cancer 101(4), 715–721 (2009).

319. Pruitt K et al. Ras and Rho regulation of the cell cycle and oncogenesis. Cancer Letters 171, 1–10 (2001).

320. Samuels Y et al. High frequency of mutations of the PIK3CA gene in human cancers. Science. 2004; 304:554.

321. Samuels Y et al. Oncogenic mutations of PIK3CA in human cancers. Cell Cycle 3, 1221–1224 (2004).

322. Hollstein M et al. p53 Mutations in Human Cancers. Science 253(5015), 49-53 (1991).

323. Leslie A et al. The colorectal adenoma-carcinoma sequence. Br J Surg. 89, 845–860 (2002).

324. Lane D P. p53, guardian of the genome. Nature 358, 15-16 (1992).

325. Khoo K H. Cell fate outcomes after p53 induction: apoptosis, cell cycle arrest or senescence? Nat. Rev. Drug Discovery 13, 217-236 (2014).

326. Amaral JD, Xavier JM, Steer CJ, Rodrigues CM. The role of p53 in apoptosis. Discov Med 9, 145-152 (2010).

327. Beroud C et al. The UMD-p53 database: New mutations and analysis tools. Human Mutation 21, 176–181 (2003).

328. Keino-Masu K et al. Deleted in Colorectal Cancer (DCC) encodes a netrin receptor. Cell 87, 175–185 (1996).

329. Fazeli A et al. Phenotype of mice lacking functional Deleted in colorectal cancer (Dcc) gene. Nature 386, 796–804 (1997).

330. Woodford-Richens K L et al. SMAD4 mutations in colorectal cancer probably occur before chromosomal instability, but after divergence of the microsatellite instability pathway. PNAS USA 98, 9719-9723 (2001).

331. Stoler D et al. The onset and extent of genomic instability in sporadic colorectal tumor progression. PNAS U S A. 96, 15121–15126 (1999).

332. Shih I et al. Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. Cancer Res. 61, 818–822 (2001).

333. Hadjihannas M et al. Aberrant Wnt/beta-catenin signaling can induce chromosomal instability in colon cancer. PNAS U S A. 103, 10747–10752 (2006).

334. Fodde R et al. Mutations in the APC tumour suppressor gene cause chromosomal instability. Nat Cell Biol. 3, 433–438 (2001).

335. Kaplan K et al. A role for the Adenomatous Polyposis Coli protein in chromosome segregation. Nat Cell Biol. 3, 429–432 (2001).

336. de la Chapelle A et al. Clinical Relevance of Microsatellite Instability in Colorectal Cancer. Journal of Clinical Oncology 28(20), 3380–3387 (2010).

337. 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. 57, 808–811 (1997).

338. Eshleman J R et al. Chromosome number and structure both are markedly stable in RER colorectal cancers and are not destabilized by mutation of p53. Oncogene 17, 719-725 (1998).

339. Jass J R et al. Distinction between familial and sporadic forms of colorectal cancer showing DNA microsatellite instability. Eur J Cancer 38, 858-866 (2002).

340. Markowitz S et al. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science 268, 1336-1338 (1995).

341. Yamamoto H et al. Frameshift somatic mutations in gastrointestinal cancer of the microsatellite mutator phenotype. Cancer Res 57, 4420-4426 (1997).

342. 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 13, 2303-2311 (2004).

343. Veigl M L et al. Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. PNAS U S A. 95, 8698–8702 (1998).

344. 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. 59, 4506–4509 (1999).

345. Capper D et al. BRAF V600E-specific immunohistochemistry for the exclusion of Lynch syndrome in MSI-H colorectal cancer. Int. J. Cancer 133, 1624–1631 (2013).

346. Kawasaki T et al. CpG island methylator phenotype-low (CIMP-low) colorectal cancer shows not only few methylated CIMP-high-specific CpG islands, but also low-level methylation at individual loci. Mod Pathol. 21(3), 245-55 (2008).

347. Toyota M et al. Distinct genetic profiles in colorectal tumors with or without the CpG island methylator phenotype. PNAS USA 97, 710–715 (2000).

348. Haydon A MM et al. Emerging pathways in colorectal cancer development. THE LANCET Oncology 3, 83-88 (2002).

349. Toyota M et al. CpG island methylator phenotype in colorectal cancer. PNAS USA 96, 8681-8686 (1999).

350. van Rijnsoever M et al. Characterisation of colorectal cancers showing hypermethylation at multiple CpG islands. Gut 51, 797-802 (2002).

351. Toyota M et al. Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. Cancer Res 59, 2307-2312 (1999).

352. Rashid A et al CpG island methylation in gastroenterologic neoplasia: a maturing field. Gastroenterology 127, 1578-1588 (2004).

353. Weisenberger D J et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat Genet 38, 787-793 (2006).

354. Worthley D L et al. Colorectal carcinogenesis: Road maps to cancer. World J Gastroenterology 13(28), 3784-3791 (2007).

355. Erhardt P et al. B- Raf inhibits programmed cell death downstream of cytochrome c release from mitochondria by activating the MEK/Erk pathway. Mol Cell Biol 19, 5308-5315 (1999).

356. Agrawal N et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science 333, 1154-1157 (2011).

357. Jia D et al. Exome sequencing of hepatoblastoma reveals novel mutations and cancer genes in the Wnt pathway and ubiquitin ligase complex. Hepatology 2014.

358. TCGA. Comprehensive molecular characterization of human colon and rectal cancer. Nature 487, 330-337 (2012).

359. Kim T M et al. The Landscape of Microsatellite Instability in Colorectal and Endometrial Cancer Genomes. Cell 155(4), 858-868 (2013).

360. Nikolaev S I et al. A single-nucleotide substitution mutator phenotype revealed by exome sequencing of human colon adenomas. Cancer Res 72, 6279–6289 (2012).

361. Olschwang S et al. Alternative genetic pathways in colorectal carcinogenesis. PNAS USA 94 (22), 12122–12127 (1997).

362. Smith G et al. Mutations in APC, Kirsten-ras, and p53—alternative genetic pathways to colorectal cancer. PNAS USA 99 (14), 9433–9438 (2002).

363. Salahshor S et al. Colorectal Cancer With and Without Microsatellite Instability Involves Different Genes. Genes, Chromosomes and Cancer 26, 247–252 (1999).

364. Sanz-Pamplona et al. Exome sequencing reveals AMER1 as a frequently mutated gene in colorectal cancer. Clin Cancer Res. 21(20), 4709-4718 (2015).

365. Borras E et al. Genomic Landscape of Colorectal Mucosa and Adenomas. Cancer Prev. Res. 9(6), 417-427 (2016).

366. Lee S H et al. Whole-exome sequencing identified mutational profiles of high-grade colon adenomas. Oncotarget 8(4), 6579-6588 (2017).

367. Zhou D et al. Exome capture sequencing of adenoma reveals genetic alterations in multiple cellular pathways at the early stage of colorectal tumorigenesis. PLoS One 8, e53310 (2013).

368. Maccalli C et al. Induction of Both CD8+ and CD4+ T-Cell–Mediated Responses in Colorectal Cancer Patients by Colon Antigen-1. Clin Can Res 14(22), 7292-7303 (2008).

369. Yu J et al. Novel recurrently mutated genes and a prognostic mutation signature in colorectal cancer. Gut 64(4), 636-645 (2015).

370. Ikeda Y et al. Germline PARP4 mutations in patients with primary thyroid and breast cancers. Endocr. Relat. Cancer 23(3), 171-179 (2016).

371. Balakrishnan A et al. Novel somatic and germline mutations in cancer candidate genes in glioblastoma, melanoma, and pancreatic carcinoma. Cancer Res. 67(8), 3545-3550 (2007).

372. Ahn C H et al. Absence of MLL3 mutations in colorectal carcinomas of Korean patients. APMIS 115(7), 859-860 (2007).

373. Watanabe Y et al. Frequent Alteration of MLL3 Frameshift Mutations in Microsatellite Deficient Colorectal Cancer. PLoS One 6(8), e23320 (2011).

374. Miyaki M et al. Genetic Changes and Histopathological Types in Colorectal Tumors from Patients with Familial Adenomatous Polyposis. Cancer research 50, 7166-7173 (1990).

375.Miyaki M et al. Both BRAF and KRAS mutations are rare in colorectal carcinomas from patients with hereditary nonpolyposis colorectal cancer. Cancer Letters 211, 105–109 (2004).

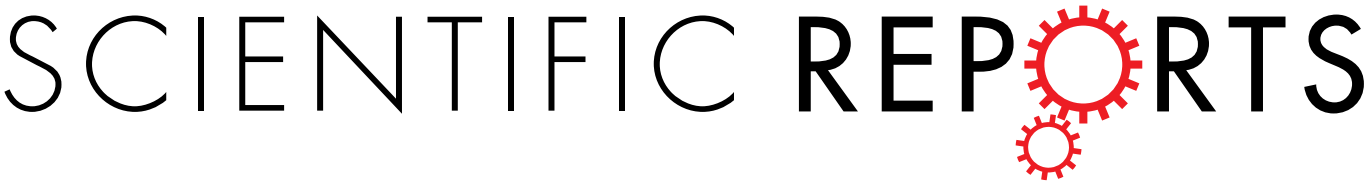
376. Lamlum et al. APC mutations are sufficient for the growth of early colorectal adenomas. PNAS 97 (5): 2225-2228 (2000).

377. McLellan E A et al. High frequency of K-ras mutations in sporadic colorectal adenomas. Gut 34: 392-396 (1993).

378. Kucherlapati R et al. Mouse models for human familial adenomatous polyposis. Seminars in Cancer Biology 11: 219-225 (2001).

379. Takane K et al. Two subtypes of colorectal tumor with distinct molecular features in familial adenomatous polyposis. Oncotarget 7(51), 84003-84016 (2016).

Reprint of Publication



OPEN

Novel mutations and phenotypic associations identified through *APC*, *MUTYH*, *NTHL1*, *POLD1*, *POLE* gene analysis in Indian Familial Adenomatous Polyposis cohort

Received: 3 November 2016
Accepted: 10 April 2017
Published online: 22 May 2017

Nikhat Khan[1,2], Anuja Lipsa[1,2], Gautham Arunachal[3], Mukta Ramadwar[4] & Rajiv Sarin[1,2]

Colo-Rectal Cancer is a common cancer worldwide with 5–10% cases being hereditary. Familial Adenomatous Polyposis (FAP) syndrome is due to germline mutations in the *APC* or rarely *MUTYH* gene. *NTHL1*, *POLD1*, *POLE* have been recently reported in previously unexplained FAP cases. Unlike the Caucasian population, FAP phenotype and its genotypic associations have not been widely studied in several geoethnic groups. We report the first FAP cohort from South Asia and the only non-Caucasian cohort with comprehensive analysis of *APC*, *MUTYH*, *NTHL1*, *POLD1*, *POLE* genes. In this cohort of 112 individuals from 53 FAP families, we detected germline *APC* mutations in 60 individuals (45 families) and biallelic *MUTYH* mutations in 4 individuals (2 families). No *NTHL1*, *POLD1*, *POLE* mutations were identified. Fifteen novel *APC* mutations and a new Indian *APC* mutational hotspot at codon 935 were identified. Eight very rare FAP phenotype or phenotypes rarely associated with mutations outside specific *APC* regions were observed. *APC* genotype-phenotype association studies in different geo-ethnic groups can enrich the existing knowledge about phenotypic consequences of distinct *APC* mutations and guide counseling and risk management in different populations. A stepwise cost-effective mutation screening approach is proposed for genetic testing of south Asian FAP patients.

Inherited predisposition is seen in 5–10% of all colorectal cancers (CRC). Major forms of hereditary colorectal cancer include the non-polyposis Lynch syndrome and the Familial Adenomatous Polyposis (FAP) syndrome[1]. Colorectal polyposis syndromes are characterized by multiple adenomatous or hamartomatous polyps and account for about 1% of all CRC cases. The adenomatous polyposis syndromes with high risk of colorectal cancer include the autosomal dominant Familial Adenomatous Polyposis (FAP [MIM: 175100]); and the autosomal recessive *MUTYH* associated polyposis (MAP [MIM: 608456]) syndrome. Recently two new entities have been described – the autosomal recessive *NTHL1* associated polyposis (NAP [MIM: 616415])[2] and the autosomal dominant polymerase proofreading-associated polyposis (PPAP) syndrome due to mutations in *POLD1* [MIM: 174761] and *POLE* [MIM: 174762] genes[3,4].

FAP is characterized by the early onset of hundreds to thousands of adenomatous polyps throughout the colon and rectum with over 90% risk of development of carcinoma in one or more of the polyps[1]. FAP is caused by germline mutation in the *APC* gene[5]. *APC* is an integral part of the wnt-signalling mechanism and regulates the proliferation of colonic epithelial cells[6]. *APC* mutation carriers also have an increased risk of developing small bowel, upper gastrointestinal and papillary thyroid carcinoma as well as childhood medulloblastoma and

[1]Sarin Lab, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC)-Tata Memorial Centre, Navi Mumbai, India. [2]Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai, 400085, India. [3]Clinical Genetics Unit, Christian Medical College and Hospital, Vellore, India. [4]Department of Pathology, Tata Memorial Hospital-Tata Memorial Centre, Mumbai, India. Correspondence and requests for materials should be addressed to R.S. (email: rsarin@actrec.gov.in)

hepatoblastoma[7]. Benign manifestations like congenital hypertrophy of the retinal pigment epithelium (CHRPE), desmoid tumors, osteomas and dental anomalies are also common[7]. Correlation between the location of mutations in *APC* gene (genotype) and the clinical phenotype in terms of the number of polyps, age of onset of polyps and CRC and distinct extracolonic manifestations is well described[8]. An attenuated variant of FAP (AFAP) due to mutations in 5′ or 3′ end of the *APC* gene, is characterized by polyps not exceeding 100 and late age of onset[1,5]. Up to 10% of FAP cases in whom *APC* mutation is not identified, there is bi-allelic germline mutation in the *MUTYH* gene[5]. Unlike FAP, the *MUTYH* associated polyposis has a lower polyp burden which rarely exceeds 100[5,9]. Comprehensive genetic analysis of *APC* and *MUTYH* fails to identify underlying gene mutation in 10–20% of FAP cases[2,4,10,11] and only a small proportion of these are explained by the recently described NAP and PPAP syndromes[2,4].

Current knowledge regarding the spectrum of *APC* gene mutation, mutational hotspots and the genotype phenotype correlations is derived mainly from studies in Caucasian cohorts[5,8,12]. In recent years, studies from other geo-ethnic groups have identified several novel *APC* genotypes, phenotypes and genotype-phenotype associations[10,13–16]. The underlying reason for differences in phenotypic associations has not been investigated but may be due to difference in the underlying genetic background or dietary habits[17,18]. *APC* genotype-phenotype association studies in different geo-ethnic groups can enrich the existing knowledge about phenotypic consequences of distinct *APC* mutations and guide counseling and risk management in different populations. This is the first FAP cohort being reported from South Asia and the only non-Caucasian cohort with comprehensive molecular genetic analysis of all the five adenomatous polyposis associated genes (*APC*, *MUTYH*, *NTHL1*, *POLD1* and *POLE*).

Results

The 53 unrelated Indian FAP families reported here represent the diverse regions and religions of the Indian subcontinent with 15 hailing from northern, 15 from eastern, 14 from western and 9 from southern states of India and belonging to Hindu (46), Muslim (2), Christian (3) and Jain (2) religions. Of the 53 probands, 25 had no family history of polyposis or cancer suggesting a *de novo* mutation. The remaining 28 probands reported a family history of polyposis with or without CRC or other extracolonic manifestations. All the probands had classical polyposis except three AFAP cases with <100 adenomatous polyps. Through Sanger sequencing and MLPA of *APC* and *MUTYH* genes, 45 families were found to harbor deleterious germline mutation in the *APC* gene (35 distinct mutations) and 2 families with bi-allelic *MUTYH* gene mutation. With extended testing of family members, a total of 60 carriers of *APC* mutation and 4 carriers of bi-allelic *MUTYH* mutations were identified. In a combined analysis in 60 *APC* mutation carriers and their 58 untested relatives with FAP associated cancer or benign manifestation, the phenotypic features observed were 79 CRC, 5 upper GI cancers, 3 thyroid cancer, 2 brain tumors, 13 desmoid tumors/fibromatosis. CHRPE was noted in 14/34 *APC* mutation carriers for whom fundus examination details were available.

Mutation spectrum. Of the 35 distinct *APC* mutations described in Table 1 and Fig. 1, 15 (43%) were *novel* mutations not previously described in the literature or the InSiGHT database. Vast majority of the mutations were truncating (17 frameshift & 14 nonsense), 2 splice site and 2 large genomic rearrangements (LGR). All the mutations were between codons 197 to 1538. The proximal exon 15 harbored 24 (69%) of all the mutations. A 5 base pair deletion at codon 1309 (c.3927_3931delAAAGA) was the most frequent mutation, identified in 7 unrelated families. A 2 base pair deletion affecting the same codon 1309 (c.3925_3927delGA) was identified in an additional family. Codon 1061 mutation (c.3183_3187delACAAA) was identified in 4 families. Interestingly, 4 distinct truncating mutations at codon 935 occurred due to 4 different nucleotide alterations (c.2804dupA, c.2805_2815del11, c.2805 C > A and c.2802_2805delTTAC) in 4 families. The remaining 28 mutations were rare and identified in one family each. The *APC* LGRs identified were a duplication of the Promoter1B identified in two families and deletion of exons 9–13 in one family. In 2 of the 3 AFAP cases, biallelic *MUTYH* mutations were identified. A homozygous *MUTYH* mutation E466X (now E480X) was identified in a South Indian Tamil AFAP patient with 40 polyps and CRC. Compound heterozygous *MUTYH* mutations R241W and G286E were identified in a case with less than 100 polyps. In the 6 *APC* and *MUTYH* mutation negative cases with classical FAP phenotype, sequencing of the entire coding region of *NTHL1* gene and the exonuclease domain of *POLD1* gene (exons 6–13) and *POLE* gene (exons 9–14) did not identify any mutation.

Phenotypic features and rare genotype-phenotype associations. Of the 60 *APC* mutation carriers, 31 had developed CRC at a mean age of 38.3 years (range18–53 years) in a background of classical polyposis with hundreds to thousands of polyps in all but one case of AFAP with only 50 polyps. In 23 *APC* carriers, polyposis was diagnosed at a mean age of 32 years (range: 9–60 years) without CRC on endoscopic evaluation or histopathological examination of prophylactic procto-colectomy. In the remaining 6 carriers, colonoscopy was yet to be performed or its details were not available. Six *APC* carriers developed extracolonic cancers with or without CRC. These included 2 cases with papillary thyroid cancer, 1 case with duodenal cancer, 1 case with intracranial germinoma, 1 case with papillary thyroid carcinoma and duodenal cancer, and 1 case with duodenal cancer and small intestine cancer. One or more benign extracolonic manifestations were identified in 27/60 *APC* mutation carriers. These included CHRPE (n = 14), desmoid tumor or fibromatosis (n = 13), upper GI polyps (n = 8) and osteomas (n = 3). Eight very rare FAP phenotypes or phenotypes rarely associated with mutations outside specific regions of the *APC* gene were observed. These include the second reported case of intracranial germ cell tumor in an *APC* carrier[19], absence of profuse polyposis and early onset CRC in 3 of the 7 codon 1309 mutation carriers as is classically described[20], attenuated phenotype with only 50 polyps at age 33 years in a codon 593 mutation carrier, desmoid tumor with codon 1228 mutation, papillary thyroid cancer with codon 1346 mutation and most interestingly CHRPE with codon 1483 mutation[7,8].

Sr. No	Nucleotide change	Exon	Consequence	Type of mutation	No. of families with this mutation	Reported in InSiGHT® database or novel
1	c.589delA	5	p.R197Efs*8	Frameshift	1	Novel
2	c.706 C > T	6	p.Q236*	Nonsense	1	Reported
3	c.694 C > T	6	p.R232*	Nonsense	1	Reported
4	c.1620dupA	12	p.Q541Tfs*19	Frameshift	1	Reported
5	c.1690C > T	13	p.R564*	Nonsense	1	Reported
6	c.1779G > A	14	p.W593*	Nonsense	1	Reported
7	c.1861dupA	14	p.T621Nfs*13	Frameshift	1	Reported
8	c.2274delA	15	p.A759Pfs*2	Frameshift	1	Novel
9	c.2802_2805delTTAC	15	p.Y935Ifs*19	Frameshift	1	Reported
10	c.2804dupA	15	p.Y935*	Frameshift	1	Reported
11	c.2805_2815del11	15	p.Y935*	Frameshift	1	Novel
12	c.2805 C > A	15	p.Y935*	Nonsense	1	Reported
13	c.2828 C > G	15	p.S943*	Nonsense	1	Reported
14	c.3183_3187del5	15	p.Q1062*	Frameshift	4	Reported
15	3259_3260delCT	15	p.L1087Qfs*31	Frameshift	1	Novel
16	c.3298dupT	15	p.S1100Ffs*19	Frameshift	1	Novel
17	c.3358 G > T	15	p.G1120*	Nonsense	1	Reported
18	c.3682 C > T	15	p.Q1228*	Nonsense	1	Reported
19	c.3815 C > A	15	p.S1272*	Nonsense	1	Novel
20	c.3925-3926delGA	15	p.E1309Lfs*5	Frameshift	1	Reported
21	c.3927_3931del5	15	p.E1309Dfs*4	Frameshift	7	Reported
22	c.4012 C > T	15	p.Q1338*	Nonsense	1	Reported
23	c.4037 C > G	15	p.S1346*	Nonsense	1	Novel
24	c.4202_4203delTT	15	p.I1401Sfs*7	Frameshift	1	Novel
25	c.4216 C > T	15	p.Q1406*	Nonsense	1	Reported
26	c.4285 C > T	15	p.Q1429*	Nonsense	1	Novel
27	c.4387_4394dup8	15	p. S1465Rfs*11	Frameshift	1	Novel
28	4446delT	15	p.P1483Qfs*24	Frameshift	1	Novel
29	c.4463 T > G	15	p.L1488*	Nonsense	1	Reported
30	c.4529delG	15	p.S1510Tfs*13	Frameshift	1	Novel
31	c.4612_4613delGA	15	p.E1538Ifs*5	Frameshift	1	Reported
32	IVS14 + 1 G > A	—	—	Splice site	1	Reported
33	IVS14 + 2 T > C	—	—	Splice site	1	Novel
34	Deletion of Exons 9-13	—	—	LGR	1	Novel
35	Duplication of promoter 1B	—	—	LGR	2	Novel

Table 1. Spectrum of *APC* mutations in Indian FAP cohort.

Discussion

In FAP, the mutation spectrum of *APC* gene and genotype-phenotype correlations is well characterized for the Caucasian population[12, 21–25], and to some extent for the East Asian population[15, 26–30]. Also, comprehensive molecular characterization of all the 5 known genes has been performed in very limited number of cases, that too only in the Caucasian population. Our study is the first report of a South Asian cohort of 53 FAP families and the only non-Caucasian FAP cohort analysed for all the 5 adenomatous polyposis associated genes.

The wide variation in the reported frequency of germline *APC* or *MUTYH* mutations in FAP cohorts from as low as 40–60%[23, 31, 32] to as high as 75–94%[10, 24, 33] is due to the stringency in making a syndromic diagnosis or lack of comprehensive genetic analysis. The high mutation detection rate of 89% in our cohort reflects the appropriateness of our clinical characterization for making the syndromic diagnosis and the comprehensive genetic analysis for *APC* and *MUTYH* including MLPA.

This study has identified a new Indian mutational hotspot at codon 935 seen in 4 (9%) FAP families. In addition, the two other known hotspot mutations at codons 1309 and 1061 were seen in 18% and 9% families respectively. High frequency of codon 1309 and 1061 mutations worldwide[32] is a result of repetitive nucleotides in DNA sequence making it a mutational hotspot. Identification of *APC* LGR in 3 of the 11 families negative for *APC* point mutation or small indels and biallelic *MUTYH* mutation in 2 of the 8 families without *APC* mutation or LGR mandates its inclusion in comprehensive genetic analysis for south Asian FAP/AFAP cases. The *MUTYH* mutation E466X (now E480X), previously described in 3 unrelated Indian families living in the UK[34] was identified as a homozygous mutation in one of our AFAP case from Tamil Nadu in south India. E466X may thus be a founder *MUTYH* mutation in Indians, possibly of Tamil ancestry. The founder effect of E466X needs to be confirmed with haplotyping studies and its population frequency can be established in a larger cohort. *NTHL1*, *POLD1* or *POLE*

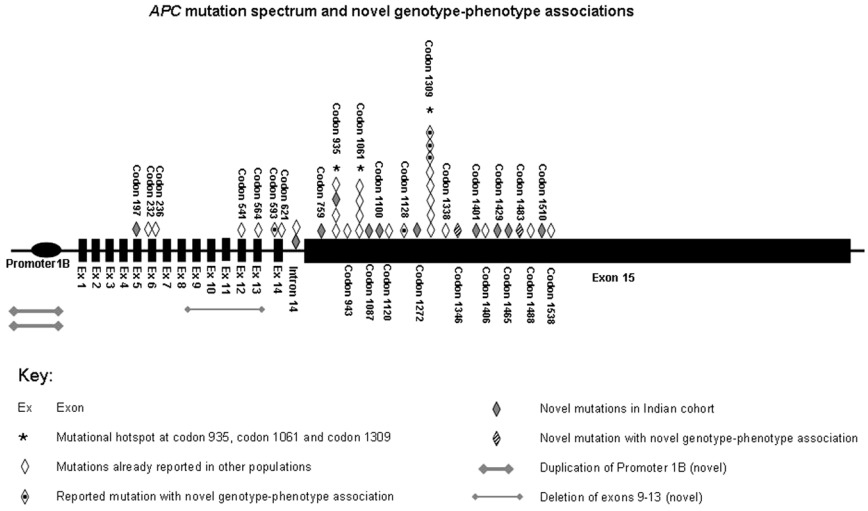


Figure 1. *APC* mutation spectrum and novel genotype-phenotype associations. The mutation distribution shows clustering of two thirds of all *APC* mutations in proximal Exon 15, with three Indian mutational hotspots (codon 935, 1061 and 1309) contributing to one third of all *APC* mutations. Large number of novel *APC* mutations (n = 15) and few novel genotype phenotype associations for codon 1228, 1346 and 1483 mutations.

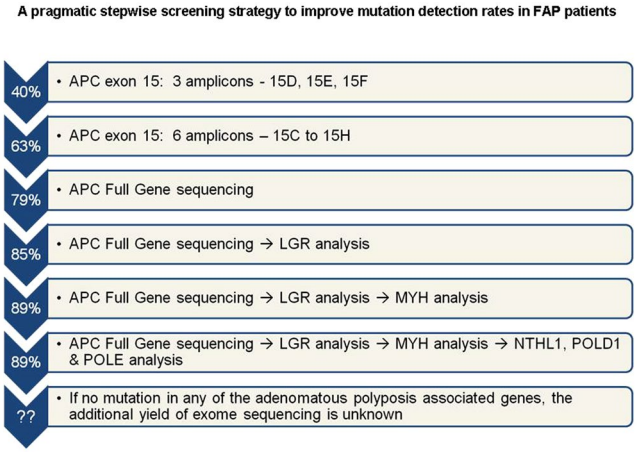


Figure 2. A pragmatic stepwise screening strategy to improve mutation detection rates in FAP patients. Cumulative mutation detection rates with step wise screening of exons/genes most likely to be mutated in south Asian FAP cases. Arrows on left side shows the cumulative mutation detection rates in our cohort achieved after each step. In our cohort, the cumulative mutation detection rate did not change with *NTHL1*, *POLD1* and *POLE* gene analysis it may increase the detection rate slightly in larger cohorts of *APC* and *MUTYH* negative adenomatous polyposis cases from different geo-ethnic background.

mutations were not identified in any of the 6 FAP probands negative for *APC* or *MUTYH* mutations. This is not surprising as none of these families fulfilled the salient features of PPAP or NAP as described in the literature[4,35].

Of the 35 distinct mutation identified in our cohort, 15 (43%) are novel and not previously reported in Caucasian or other geo-ethnic groups. Moreover eight very rare FAP phenotype or phenotypes rarely associated with mutations outside specific regions of the *APC* gene were identified. *APC* genotypes and genotype-phenotype associations rarely or never observed in Caucasian cohorts are being increasingly reported from other geo-ethnic groups[10,14–16,29]. This highlights the need to study different geo-ethnic groups to enrich the global *APC* mutational spectrum and expand our knowledge of phenotypic associations of distinct *APC* mutations.

Based on the mutational spectrum and hotspots identified, a pragmatic stepwise genetic testing algorithm is proposed for FAP cases in south Asian countries where genetic testing is not routinely performed due to resource constraints (Fig. 2). Initial screening of three amplicons (15D–15F) harboring the mutational hotspot codons 1309, 1061 and 935 could identify 40% of all *APC* mutations and sequencing of additional 3 amplicons of exon 15 (15 C, 15 G, 15 H) could identify two thirds of all *APC* mutations. If no mutation is identified rest of the *APC* should be screened followed by LGR analysis and *MUTYH* gene sequencing. Extended testing of other adenomatous polyposis associated genes (*NTHL1*, *POLD1 and POLE)* may be considered but the yield is likely to be very low. The present study and few recent reports[36] highlight that a significant proportion of FAP cases do not harbor

pathogenic mutations in the genes known to be associated with FAP, MAP, NAP, PPAP syndrome. Germline exome sequencing in an adenomatous polyposis cohort has recently reported loss-of-function germline mutations in a few promising candidate genes (*DSC2*, *PIEZO1*, *ZSWIM7*)[36] and biallelic mutations in *MSH3* gene[37]. However these recently identified adenomatous polyposis genes are likely to remain under-reported, unless they are tested as single genes or included in multi-gene next generation sequencing (NGS) panels. The currently used multi-gene panels may not be informative as they do not include *NTHL1*, *POLD1* and *POLE* genes. Therefore there is a need to conduct comprehensive genetic analysis of all the known adenomatous polyposis genes or exome sequencing studies in large pooled cohorts of *APC* and *MUTYH* negative adenomatous polyposis cases with detailed phenotypic and geo-ethnicity correlation.

In conclusion, the comprehensive investigation of all the five adenomatous polyposis genes in a well characterized Indian FAP cohort confirms the high frequency of *APC* mutations in classical FAP, *MUTYH* in AFAP cases and absence of *NTHL1*, *POLD1* and *POLE* mutations in cases not showing syndromic features of PPAP or NAP. The pragmatic stepwise approach proposed can improve uptake of genetic testing for FAP in south Asian countries. Identification of a large number of novel *APC* mutations and genotype phenotype associations that are rare in the Caucasian population highlights the need for comprehensive phenotypic characterization and genetic analysis in large FAP cohorts from diverse geo-ethnic backgrounds.

Methods

Patients and Phenotype characterization. The study was conducted on 53 FAP families recruited through Cancer Genetics Clinic at Tata Memorial Centre, Mumbai and Christian Medical College, Vellore; India. The study was approved by the Hospital Ethics Committee of the Tata Memorial Hospital and all participating subjects provided written informed consent. All experiments were carried out in accordance with the approved guidelines and regulations. Syndromic diagnosis of FAP or AFAP was based on the number of adenomatous polyps in the colorectum with or without colorectal cancer. Further phenotypic characterization was done based on colonoscopy, esophago-gastro-duodenoscopy (EGD), computerized tomography of abdomen, thyroid ultrasound and ophthalmic examination. Detailed family history and medical records were taken from all the families reported in this study. Genetic testing was extended on first and second degree relatives if a deleterious germline mutation was identified in the proband. Blood sample was collected from 112 members from these 53 families.

PCR and Sequencing. For germline mutation analysis the complete coding sequence of the *APC*, *MUTYH* and *NTHL1* genes and the exonuclease domain of *POLD1* gene (exons 6–13) and *POLE* gene (exons 9–14) were amplified by Polymerase Chain Reaction (PCR). Primer sequences and annealing temperatures for PCR used are given in the supplementary Tables S1–S4. PCR products were purified with ExoSAP-IT [USB products, Affymetrix] and sequenced using an ABI 310 Avant, 3500 and 3730 DNA sequencer (Applied Biosystems). All mutations were confirmed by bidirectional sequencing. For most of the cases, the mutations were further reconfirmed on a second independent sample collected after the identification of mutation. InSiGHT database (LOVD) and available literature was used to check if the mutations identified was reported or novel. The mutations identified in our cohort are submitted in the InSiGHT database (www.insight-group.org).

MLPA analysis. If no *APC* mutation was identified on sequencing, large genomic rearrangement (LGR) in *APC* and *MUTYH* gene were evaluated with Multiplex ligation-dependent probe amplification (MLPA) using the SALSA MLPA *APC* P043 kit [MRC-Holland] as per the instructions provided by the company. The data was analyzed with Coffalyser software. All deletions or duplications identified and all uncertain results were confirmed in at least two independent MLPA reactions.

References

1. Stoffel, E. M. *et al.* Hereditary Colorectal Cancer Syndromes: American Society of Clinical Oncology Clinical Practice Guideline Endorsement of the Familial Risk–Colorectal Cancer: European Society for Medical Oncology Clinical Practice Guideline. *Journal of Clinical Oncology* **33**, 209–217 (2015).
2. Weren, R. D. *et al.* A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. *Nature Genetics* **47**, 668–671 (2015).
3. Palles, C. *et al.* Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nature Genetics* **45**(2), 136–144 (2013).
4. Bellido, F. *et al.* POLE and POLD1 mutations in529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. *Genetics in Medicine* **18**(4), 325–332 (2016).
5. Grover, S. *et al.* Prevalence and phenotypes of APC and MUTYH mutations in patients with multiple colorectal adenomas. *JAMA* **308**, 485–92 (2012).
6. Narayan, S. & Roy, D. Role of APC and DNA mismatch repair genes in the development of colorectal cancers. *Molecular Cancer* **2**, 41–55 (2003).
7. Groen, E. J. *et al.* Extra-Intestinal Manifestations of Familial Adenomatous Polyposis. *Annals of Surgical Oncology* **15**, 2439–2450 (2008).
8. Nieuwenhuis, M. H. & Vasen, H. F. Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): A review of the literature. *Critical Reviews in Oncology/Hematology* **61**, 153–161 (2007).
9. Vogt, S. *et al.* Expanded extracolonic tumor spectrum in MUTYH-associated polyposis. *Gastroenterology* **137**, 1976–1985 (2009).
10. Cao, X., Hong, Y., Eu, K. W., Loi, C. & Cheah, P. Y. Singapore Familial Adenomatous Polyposis (FAP) Patients with Classical Adenomatous Polyposis but Undetectable *APC* Mutations Have Accelerated Cancer Progression. *American Journal of Gastroenterology* **101**, 1–8 (2006).
11. Mongin, C. *et al.* Unexplained polyposis: a challenge for geneticists, pathologists and gastroenterologists. *Clinical Genetics* **81**, 38–46 (2012).
12. Friedl, W. & Aretz, S. Familial Adenomatous Polyposis: Experience from a Study of 1164 Unrelated German Polyposis Patients. *Hereditary Cancer in Clinical Practice* **3**, 95–114 (2005).
13. Patel, N. *et al.* A novel APC mutation defines a second locus for Cenani-Lenz syndrome. *Journal of Medical Genetics* **52**(5), 317–321 (2015).
14. Tao, H. *et al.* Identification of 5 novel germline APC mutations and characterization of clinical phenotypes in Japanese patients with classical and attenuated familial adenomatous polyposis. *BMC Research Notes* **3**, 305–313 (2010).
15. Kim, D. W. *et al.* Mutation Spectrum of the APC Gene in 83 Korean FAP Families. *Human Mutation* **26**, 281–292 (2005).

16. Torrezan, G. T. *et al*. A novel SYBR-based duplex qPCR for the detection of gene dosage: detection of an *APC* large deletion in a familial adenomatous polyposis patient with an unusual phenotype. *BMC Medical Genetics* **13**, 55–61 (2012).
17. Crabtree, M. D. *et al*. Analysis of candidate modifier loci for the severity of colonic familial adenomatous polyposis, with evidence for the importance of the N-acetyl transferases. *Gut* **53**(2), 271–276 (2004).
18. Cruz-Correa, M. *et al*. Combination treatment with curcumin and quercetin of adenomas in familial adenomatous polyposis. *Clinical Gastroenterology and Hepatology* **4**(8), 1035–1038 (2006).
19. Lima, B. R., Schoenfield, L. & Rychwalski, P. J. Germinoma presenting as a fourth cranial nerve palsy in a patient with adenomatous polyposis coli (APC) gene mutation. *Journal of American Association for Pediatric Ophthalmology* **15**(1), 71–73 (2011).
20. Caspari, R. *et al*. Familial adenomatous polyposis: mutation at codon 1309 and early onset of colon cancer. *Lancet* **343**(8898), 629–32 (1994).
21. Bisgaard, M. L., Ripa, R. S. & Bulow, S. Mutation Analysis of the Adenomatous Polyposis Coli (APC) Gene in Danish Patients with Familial Adenomatous Polyposis (FAP). *Human Mutation* **23**, 522–526 (2004).
22. Torrezan, G. T. *et al*. Mutational spectrum of the APC and MUTYH genes and genotype–phenotype correlations in Brazilian FAP, AFAP, and MAP patients. *Orphanet Journal of Rare Diseases* **8**, 54–65 (2013).
23. Plawski, A. & Slomski, R. APC gene mutations causing familial adenomatous polyposis in Polish patients. *Journal of Applied Genetics* **49**, 407–414 (2008).
24. De Rosa, M. *et al*. The Mutation Spectrum of the APC Gene in FAP Patients from Southern Italy: Detection of Known and Four Novel Mutations. *Human Mutation* **21**, 655–656 (2003).
25. Vandrovcova, J., Stekrova, J., Kebrdlova, V. & Kohoutova, M. Molecular Analysis of the APC and MUTYH Genes in Czech Families Affected by FAP or Multiple Adenomas: 13 Novel Mutations. *Human Mutation* **23**, 397–404 (2004).
26. Song, G., Yuan, Y., Zheng, F. & Yang, N. Novel insertion mutation p.Asp610GlyfsX23 in APC gene causes familial adenomatous polyposis in Chinese families. *Gene* **516**, 204–208 (2013).
27. Chiang, J. M. *et al*. Mutation analysis of the APC gene in Taiwanese FAP families: low incidence of APC germline mutation in a distinct subgroup of FAP families. *Familial Cancer* **9**, 117–124 (2010).
28. Sheng, J. Q. *et al*. APC gene mutations in Chinese familial adenomatous polyposis patients. *World Journal of Gastroenterology* **16**, 1522–1526 (2010).
29. Liao, D. X. *et al*. Two Chinese pedigrees for adenomatous polyposis coli: new mutations at codon 1309 and predisposition to phenotypic variations. *Familial Cancer* **13**(3), 361–368 (2014).
30. Wei, S. C. *et al*. Genetic Analysis of the APC Gene in Taiwanese Familial Adenomatous Polyposis. *Journal of Biomedical Sciences* **11**, 260–265 (2004).
31. Gavert, N. *et al*. Molecular Analysis of the APC Gene in 71 Israeli Families: 17 Novel Mutations. *Human Mutation* **19**, 664–670 (2002).
32. Gomez-Fernandez, N. *et al*. Molecular analysis of the APC and MUTYH genes in Galician and Catalonian FAP families: a different spectrum of mutations? *BMC Medical Genetics* **10**, 57–68 (2009).
33. Rivera, B. *et al*. Clinical and genetic characterization of classical forms of familial adenomatous polyposis: a Spanish population study. *Annals of Oncology* **22**, 903–909 (2011).
34. Jones, S. *et al*. Biallelic germline mutations in MUTYH predispose to multiple colorectal adenoma and somatic G:C–>T:A mutations. *Human Molecular Genetics* **11**, 2961–2967 (2002).
35. Rivera, B., Castellsague, E., Bah, I., van Kempen, L. C. & Foulkes, W. D. Biallelic NTHL1 Mutations in a Woman with Multiple Primary Tumors. *New England Journal of Medicine* **37**, 1985–1986 (2015).
36. Spier, I. *et al*. Exome sequencing identifies potential novel candidate genes in patients with unexplained colorectal adenomatous polyposis. *Familial Cancer* **15**, 281–288 (2016).
37. Adam, R. *et al*. Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. *American Journal of Human Genetics* **99**(2), 337–351 (2016).

Acknowledgements

We thank the Indian Council of Medical Research for funding the project and ACTREC for providing fellowship to Nikhat Khan. We thank the genetic counselors, Nina Bhatnagar, Payal Manek, Gouri Pandit, Ravindra Reddy and Vandana Kembhavi at Cancer Genetics Clinic, TMC. We also appreciate the help extended from Dr. Pradnya Kowtal, Jyoti Patel and Moquitul Haque from Sarin Lab, ACTREC; Sharda Haralkar and Naresh Mahida from Genomics Facility, ACTREC. We acknowledge the cooperation of members of the Gastro-Intestinal Disease Management Group of TMH for referring the patients. We are thankful to all the patients and their families for participating in the study.

Author Contributions

R.S. and N.K. designed the project. N.K. and A.L. performed experiments. R.S., G.A. and M.R. assisted with sample collection and clinical evaluation. N.K. and R.S. wrote the manuscript. All authors approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-02319-6

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.