Molecular pathways in the origin of diverse tumors in individuals with germline TP53 mutation (Li Fraumeni Syndrome)

By

MD. MOQUITUL HAQUE
LIFE09201104005

Tata Memorial Centre, ACTREC

A thesis submitted to the
Board of Studies in Chemical Sciences/ Engineering Sciences/ Health Sciences/ Life Sciences/ Mathematical Sciences/ Physical Sciences/ Strategic Studies
(as applicable)
In partial fulfillment of requirements
for the Degree of

DOCTOR OF PHILOSOPHY
of
HOMI BHABHA NATIONAL INSTITUTE

September 2019

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.

Signature

Md. Moquitul Haque

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.

Signature

Md. Moquitul Haque

List of Publications arising from the thesis

Journal

1. Identification and characterization of TP53 gene Allele Dropout in Li-Fraumeni syndrome and Oral cancer cohorts". **Haque MM**, Kowtal P, Sarin R. Scientific Reports, 2018; 8(1): 11705 (doi10.1038/s41598-018-30238-7).

Peer reviewed conference proceedings

1. **Haque MM et al.** Genotype – Phenotype spectrum in a large Indian Li-Fraumeni Syndrome (LFS) cohort. *European Journal of Human Genetics,* (European Human Genetic Conference 2016, Barcelona), Vol-24, E-Supplement 1, 285 2016.
2. **Haque MM,** Puri A. & Sarin R. Contribution of germline *TP53* mutation in sporadic sarcoma cases, *European Journal of Cancer* (A Conference of New Ideas in Cancer - Challenging Dogmas 2016, TMC-Mumbai) Vol-54, Supplement 1, 552. 2016.

Chapters in books and lectures notes: NA

Conferences

1. **Poster Presentation**: Genotype – Phenotype spectrum in a large Indian Li-Fraumeni Syndrome (LFS) cohort. *European Human Genetics Conference*, Barcelona, Europe, 2016

2. **Poster Presentation**: Contribution of germline *TP53* mutation in sporadic sarcoma cases, *TMC Platinum jubilee Conference*, Mumbai, India 2016.

3. **Poster Presentation**: *TP53* germline mutation and its clinical correlation in a large Indian Li-Fraumeni syndrome (LFS) cohort, *IACR*, Jaipur, India 2015.

4. **Poster Presentation**: Genetic anticipation and phenotypic features in Indian Li-Fraumeni Syndrome (LFS) families with *TP53* mutation., *Indian Cancer Genetics Conference and Training Workshop*, ACTREC-TMC, Navi Mumbai, India, 2013.

Workshop

1. One day workshop on MLPA data analysis, 2014

Others

Workshop Instructor: Basic genetic and molecular technique:

1. 2nd Indian Cancer Genetics Conference and Workshop, ACTREC-TMC, Navi Mumbai 2014
2. 1st Indian Cancer Genetics Conference & Training Workshop, ACTREC-TMC, Navi Mumbai 2013

Participant in organizing conference

1. National Research Scholars' Meet in Life Sciences (NRSM) 2014
2. Indian Cancer Genetics Conference 2013 and 2014

Signature

Md. Moquitul Haque

This thesis is dedicated to my father Late Md. Mohsinul Haque, Mother, family members and All the Patients and families in whom I have done the study

ACKNOWLEDGEMENTS

"And if you would count the favours of Allah, never could you be able to count them. Truly, Allah is Oft-Forgiving, Most Merciful" [Quran, chapter 16: Verse 18].

The Almighty ALLAH has showered His mercy and blessed me with all happiness. I would like to thank Almighty Allah for the wonderful life, family, friends and opportunities that He has given me. May Allah bless me with even a stronger faith in Him and I shall abide by His orders. After Allah I would like to thank the last and the final Prophet Muhammad (Peace and blessings of Allah be upon him) who showed us the correct path of life and genuine path to achieve any goal in this world and hereafter.

I would like to thank my Mother as she is striving a lot only to make me achieve the highest possible academic degree so that I can spread the achieved knowledge and also to have a better future ahead not only of mine but the whole society of which I belong.

It gives me an enormous pleasure to express my gratitude to all of them who extended their help towards the successful completion of my PhD project. During this journey I met with many people, and all of them taught me a lot of things.

First and foremost, I would like to thank my Guide, Professor Rajiv Sarin. This work would not have been possible without his guidance, support and encouragement. I would like to thank Almighty for giving me an opportunity to work with him. He constantly provided with important inputs and arranged for any infrastructure whenever needed. He is very knowledgeable and sound in his field. He has a power of grasping things immediately which makes him sound in almost every field. He is an

institution in himself from whom you have a lot to learn and I tried and succeeded in absorbing only few skills from him. I would like to thank him, especially for supporting me financially after ending of my fellowship. May his teachings will help me in grooming and nurturing myself as an independent scientist.

I would like to thank Dr. Pradnya Kowtal (Co-Investigator, Sarin Lab). She taught all the basics of molecular biology and always helped with her valuable suggestions. She is very efficient and expert in molecular biology techniques and I am really grateful to her for such fruitful training. She taught all the good laboratory practices and helped me to understand the utilization of resources in an economical way. She is very sincere and dedicated towards work. Hope I may pursue a good career ahead with all her teachings and blessings.

I would like to express my sincerest gratitude to the Director of ACTREC Dr. S.V. Chiplunkar for providing such a wonderful infrastructure to work with. I am also thankful to Dr. Sudeep Gupta (Deputy Director) and Dr. Surekha Zingde (Former-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 Dr. Sorab Dalal who is the chairperson of my Doctoral Committee. He is always approachable for any discussion and suggestions. He not only gives his valuable suggestions in the doctoral committee meetings but is always available for his help, inputs and suggestions beyond that. I am thankful to my doctoral committee members, Dr. Tanuja Teni and Dr. Ajay Puri, for their valuable

suggestions throughout my Phd and encouragement to work. I am also thankful to my ex-Chairperson Dr. Rita Mulherkar and ex-member Dr. Sheshadri for their valuable suggestions and best wishes.

I would like to appreciate the hard work done by the Genetic Counselors at the Cancer Genetics Clinic to recruit the patients on which this thesis work is based. Without this integral help, this thesis would not have ever 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, Nidhi, Divya and Srika.

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. Thanks to Dr. Amin, Dr. Poonam Gera, Dr. Kshama, Dr. Rajshri, Sayee, Malcolm, Ranjeeta, Nishanth, Dr. Pooja, Ankush, Afrin, Neha and others for their help and co-operation.

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 with their friendship and bonding throughout my PhD career 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 and healthy environment provided by Sarin Lab members. I feel very lucky to be a part of this lab. 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 Nikhat, Jyoti, Disha, Priyanka, Roshan, Shujath, Pallavi, Amit, Arpita, Asmita, Kushal, Imran, Vasudha, Dr. Pallavi Shukla, Anuja, Prerana Dange, Afrin and Mr. Parag Madankar. I am thankful to trainees of the lab, Zeenat, Bhanu, Kaivalya,, Haritha, Tushar, Priyanka, Rucha, Himani, Satish, Aradhana, Gunjan, Nivedita, Flavia, Bhagyashree, Zishan, Ayesha, Tanishka, Trupti, Rushada, Shabnam and the list goes on.

I would like to mention three very important names Nikhat, Jyoti and Roshan, whose help cannot be defined in words. Throughout my PhD career, they helped me and extended their tremendous support. They always stand beside me whenever and wherever I was in any difficulty, whether it was professional or personal. Whatever I do for them, I will never be able to return back their help and support in any way.

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 her help related to all documentation work of academics and her suggestions and inputs..I would like to thank Sarwari also for her help related to administrative and academics. Mr. Nikhil Gadewal (BTIS) and Nilesh deserve special credit for helping

me in various Bioinformatics based problems. I would like to acknowledge the help of Smruti Mokal and Pallavi for their help in biostatistical analysis. I would like to thank Dr. Rohit for his kind help extended in compilation and formatting of my thesis. I would like to thank Pallavi Pandharipande and Vasudha for their generous help in the final phase of my thesis.

I would like to acknowledge my batch 2011. Every member of my batch was very helpful and supportive. I am blessed to have such batch mates. They are now part of life in both professional and personal world.

This whole work is incomplete without acknowledging the patients and their relatives for their cooperation and support. 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 am delighted to acknowledge my family for all the support they have extended. Pursuing PhD was never been possible without the help and motivation by my mother. She supported me in every possible way so that I can achieve all possible things in my life and can achieve all my dreams. My all family members always supported and motivated me for achieving goals in my life. My wife, who became the part of my life in the final phase of my PhD, suffered a lot as I was hardly able to give her time. I feel bound to my wife and my son as they sacrificed their important and special moments as I was not able to be with them.

TABLE OF CONTENTS

Synopsis

Homi Bhabha National Institute

SYNOPSIS OF Ph. D. THESIS

1. **Name of the Student:** Md. Moquitul Haque

2. **Name of the Constituent Institution:** Tata Memorial Centre, ACTREC

3. **Enrolment No. and Date of Enrolment:** LIFE09201104005; 25th July 2011

4. **Title of the Thesis:** Molecular pathways in the origin of diverse tumors in individuals with germline TP53 mutation (Li Fraumeni Syndrome)

SYNOPSIS

Introduction: Li Fraumeni syndrome (LFS) (OMIM#151623) is an autosomal dominant highly penetrant cancer predisposition syndrome first described by Li and Fraumeni in 1969 (1). In contrast to other inherited cancer syndromes characterized by site-specific cancers, LFS presents with a wide variety of tumor types. It is also called as SBLA syndrome as the most common cancer types are Sarcomas; Breast cancer, Brain tumors; Leukemia, and Adrenocortical carcinoma. LFS is defined by strict clinical criteria, described by Li and Fraumeni in 1969, as: Proband with sarcoma before 45yrs of age, a 1st degree relative with cancer before 45yrs of age, another 1° or 2° relative with any cancer under 45yrs of age or with sarcoma at any age. Relaxed cliical criteria by age of onset and tumour types are termed as Li-Fraumeni-like (LFL) syndrome (2, 3). Mutations in TP53 gene are detected in about

70% of LFS families and in 20-40% of LFL families. Association of *TP53* gene with the syndrome was first reported by David Malkin in 1990. (4)

TP53 gene is located on chromosome 17p13.1 which consists of 11 exons with the first exon being noncoding. *TP53* gene encodes for a tumor suppressor protein which plays an important role in various cellular mechanism like cell cycle arrest, apoptosis, and DNA repair etc and is also called the "guardian of the genome" (5). In contrast to normal p53 its mutant counterpart also has functions like gain of function, dominant negative function (6, 7). Mutations that deactivate p53 in cancer usually occur in the DNA binding domain (DBD), abrogating the ability of the p53 protein to bind to its target DNA sequences and prevents transcriptional activation of these genes. TP53 is mutated in 50% of tumors. LFS is associated with diverse type of tumors with varying severity and mutations in TP53 gene are found throughout the gene, with mutational hotspots in the DNA binding domain. *TP53* mutation spectrum is well characterized in other populations but so far there is no cohort based study from India.

Many studies have attempted to explain the correlation between the genotype and the phenotype to explain the mechanism behind the phenotypic variability seen in LFS families. The missense mutations in DBD of *TP53* are associated with higher incidence of cancers (particularly breast and CNS tumors) and earlier ages of onset than in families with mutations leading to truncated or non-functional proteins (8, 9). The role of low penetrance SNPs in certain genes in modulating the phenotype has also been examined by various studies. Individuals carrying G allele of the *MDM2* SNP309 and G allele of *TP53* P72R have early age of onset as compared to those with T allele and C allele of *MDM2* and *TP53* respectively in the LFS families (9). Many

common SNPs have been shown to be associated with cancer by genome-wide association (GWA) studies (10, 11).

Hypothesis: We hypothesize that the genotype – phenotype correlation in the Indian LFS families may be distinct from the well characterized Caucasian population (9, 12-15) and this knowledge can help in understanding the molecular pathways in the origin of diverse LFS tumors and devising population specific counseling and risk management.

AIM: To identify germline *TP53* gene mutation spectrum and its correlation with the tumour spectrum and age at diagnosis in a large Indian cohort of hereditary LFS and sporadic LFS associated cancers.

OBJECTIVES

1. Identifying LFS / LFL families and their detailed phenotypic characterization.
2. Comprehensive Genetic Analysis of *TP53* & MDM2 in LFS / LFL & sporadic sarcomas & establishing Genotype-Phenotype correlation
    A. Identification of germline point mutations and Large Genomic Rearrangements in *TP53* gene in LFS/LFL/suspected LFL families.
    B. *TP53* & *MDM2* polymorphisms and clinical phenotype in LFS / sporadic sarcoma
3. Establishing Lymphoblastoid cell lines from germline *TP53* mutation carriers for DNA repair studies & variant characterization.

METHODS

All patients were participants of studies approved by the Tata Memorial Centre-ACTREC Institutional Review Board and had provided written informed consent for

biobanking and genetic analysis. For minors, the written informed consent was provided by the parents. A cohort of cancer patients with personal or family history suggestive of hereditary LFS or LFL syndrome were enrolled in the Cancer Genetics Clinic of the Tata Memorial Hospital for genetic counseling and genetic testing. Germline *TP53* mutations were tested in a cohort of 500 LFS/LFL/Suspected LFL families registered in Cancer Genetics Clinic, Tata Memorial Hospital and under Tata International Sarcoma Kindred Study (TISKS). Of these, 197/500 families fulfilled the defined criteria of LFS or LFL, while the remaining 303/500 families did not fulfill the criteria for LFS or LFL but were tested as the proband or a family member had an LFS associated cancer. DNA was extracted using QiaAmp Blood Mini kit or by Phenol Chloroform method from the blood or Mouthwash samples of patients.

Mutation analysis of *TP53* gene: PCR amplification of the coding region of *TP53* gene was carried out using specific primers either designed or taken from literature (16). 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. MLPA analysis was carried out using SALSA MLPA kits from MRC-Holland; P056 (*TP53*) to detect large genomic rearrangements (LGRs) in cases where no point mutation or small indels were identified through sequencing. Data was analyzed using Coffalyser.Net software.

Next Generation Sequencing (NGS): Germline exome sequencing for LFS/LFL cases was carried out using Nextera rapid capture kit (Illumina) and sequenced on Hiseq2000 using manufacturer's protocol. Sequencing was done using standard v2 kit on Illumina MiSeq.

Genotyping by Restriction Fragment Length Polymorphism (RFLP): SNP 309 (rs2279744) in MDM2 was studied by RFLP using MspA1 restriction enzyme. An amplicons of 293bp digested into 193bp and 100bp in TT allele and 147, 100 and 46bp in GG allele. The amplicon has two restriction sites. One is created due to change of allele from T>G.

Establishing EBV transformed Lymphoblastoid cell lines (LCLS) and DNA damage repair studies: LCLs were established by EBV transformation of lymphocytes from the peripheral blood on patients with different types of TP53 mutations and also from healthy controls using standard. Two million cells from LCLs established from patients with TP53 mutation or healthy controls were given 2Gy ^{60}Co gamma radiation and fixed with 3% PFA at 1hr, 4hr, 24 hr and 48hr. Immuno-fluorescence images were captured in 63X, fluorescent intensity of individual nuclei was recorded and mean fluorescent intensity was calculated for 30-50 nuclei.

RESULT AND DISCUSSION

Objective 1: Identifying LFS / LFL families and their detailed phenotypic characterization.

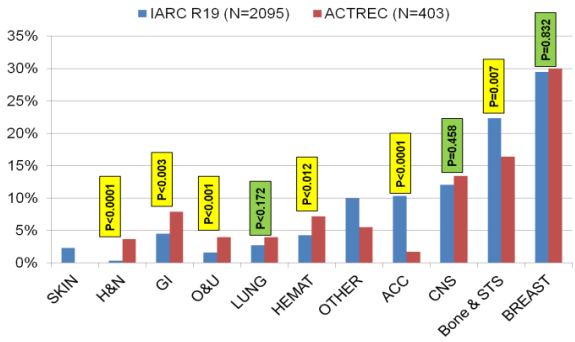


Figure 1: Comparison of Tumor spectrum

A total of 500 families with LFS, LFL or suspected LFL were registered in the Tata Memorial Centre (TMC) Cancer Genetics Clinic and under TISKS during the period 2003-2018. The frequency of several tumour types in our cohort (403 tumours in 356 individuals from TP53 mutation positive families) was significantly different from the frequency reported in the IARC

database which is predominantly a Caucasian / Brazilian database of TP53 +ve LFS families (Figure 1).

<u>Objective 2</u>: Comprehensive Genetic Analysis of *TP53* & *MDM2* in LFS / LFL and sporadic sarcomas with Genotype-Phenotype correlation

2A. Identification of TP53 point mutations and Large Genomic Rearrangements:

TP53 germline mutation analysis was done in 500 families. After excluding known polymorphisms & synonymous mutations, a total of 59 distinct mutations were identified in 79 families. Of these, 24 mutations were novel germline mutations. After further characterization based on co-segregation, *in-silico* prediction (AGVGD / SIFT) and reported transactivation activity (17), 22/24 mutations were classified as likely pathogenic while 2 mutations remained variant of unknown significance (VUS). Of these 79 mutations, 4 distinct LGRs were identified in 7 families. As expected, the mutation detection rate in families fulfilling the LFS criteria was significantly higher (12/13) than those fulfilling only the LFL criteria (54/184). The *TP53* mutation detection rate in families not fulfilling even the LFL criteria was (13/303). Surprisingly the mutation detection rate in families in which the proband has LFS spectrum tumor but not fulfilling the criteria is high (10/74), highlighting the need for further revision in the LFL criteria, which is often used to guide genetic analysis of *TP53* gene. Extended family screening in 119 members from these 79 families with likely pathogenic *TP53* mutation identified the mutation in 46 individuals. Ten of these individuals tested positive of family mutation have already developed LFS associated cancers and others are on intensive surveillance protocol. As in the IARC database, majority of *TP53* germline mutations in Indian LFS families are in the DNA Binding Domain. However, LGRs and frameshift mutations are more

common while missense mutations are less frequent in the Indian cohort as compared to the IARC database. R337H, the most frequent mutation in IARC database and a Brazilian founder mutation was not identified in our Indian cohort. Of the suspected LFL families 188 were of sporadic sarcoma cases and germline TP53 mutation was identified in two cases with a detection rate of 1%. Previous reports have shown the detection rate of 3-9% in sporadic sarcomas (18-20). A recently reported LFS associated 3'UTR region *TP53* variant (21) was not identified in the 120 LFS/LFL cases screened.

In 3 classic LFS families in whom no germline *TP53* mutations or LGRs were identified, we performed germline exome sequencing to identify driver mutations in other genes. Surprisingly a pathogenic *TP53* mutation was identified in two of these cases – one in exon 5 and the other in exon 7. Moreover, in four cases, a homozygous pathogenic mutation was identified in exon 7, which led us to suspect sequence dependent exon 7 allele drop out in these cases. In the third family where the proband and his son both had sarcoma, exome sequencing did not identify any *TP53* mutation but identified 128 frameshift indels, 228 inframe indels and 31 missense variants that were predicted to be deleterious by all the 7 *in silico* prediction tools. Interestingly, 2 of these 31 missense variants are in FGFR4 gene. Activating somatic mutation in FGFR4 gene has been reported earlier in RhabdoMyoSarcoma (RMS) (22) and merits further characterization as a germline inherited sarcoma susceptibility gene. In addition to the FGFR4 mutations, indels and missense variants predicted to be deleterious by at least one-four *in silico* prediction tools were identified in several genes, including genes associated with sarcomas like ERBB (Leiyomyosarcoma) and NSD1 (RMS).

Identification and characterization of *TP53* gene Allele Dropout in Li-Fraumeni syndrome

Allele Dropout (ADO) is a genotyping error arising from insufficient amplification of one of the two alleles, keeping it below the detection threshold by sequencing. Dropout of the mutant allele causes false negative result while dropout of Wild Type (WT) allele makes a heterozygous mutation appear homozygous (23) Sequence dependent ADOs arise from Primer Annealing Region (PAR) polymorphism (24), G-Quadruplexes, methylation or allele size differences (25, 26). Sequence independent ADOs arise from poor DNA quality as in forensics (27), Whole Genome Amplification of scanty starting DNA (27) or some unknown PCR conditions (23). *TP53,* which is the most frequently mutated gene in tumour tissues (Somatic) and its germline mutation cause LFS is a good candidate to study the frequency and causes of ADO. In two classical LFS families germline whole exome sequencing identified a TP53 mutation that was previously missed on Sanger Sequencing and later confirmed as ADO on repeat Sanger Sequencing. In four cases, ADO was suspected due to very unusual finding of germline homozygous TP53 mutation in exon 7. These mutations were later confirmed as heterozygous mutations on repeat sequencing with redesigned primers to avoid a common polymorphism (IVS7=92T>G) in the annealing region of exon 7 reverse primer. Repeat sequencing of exon 7 with redesigned primers in 150 LFL cases identified ADO in additional three cases. This highlights the need to suspect and confirm ADO in TP53 as a false negative genetic analysis report has major implications for the proband and the extended family.

Polymorphisms in *TP53* gene and the Primer annealing region (PAR): *TP53* gene has 6148 polymorphisms in *Homo sapiens*. In the PAR of Bodmer group primer set

that we used in our study, 58 polymorphic sites were identified. The exon 7 reverse PAR has a high minor allele frequency (MAF) of 0.17. The PAR of the commonly used IARC protocol primers has 83 polymorphisms with 4 having a MAF of >0.01. (http://p53.iarc.fr/download/*TP53*_directsequencing_iarc.pdf)

G-quadruplexes and methylation in *TP53* gene: The full *TP53* gene has 5931 regions including overlaps and 120 without overlaps G-rich sequences which can form quadruplex. The range of QGRS score (28) for these 5913 regions is 0-61. No CpG islands were found in any amplicon of *TP53* gene except exon 1 which is the promoter region

2B. *TP53* & MDM2 polymorphisms and the clinical phenotype in LFS and sporadic sarcoma

Germline single nucleotide polymorphisms (SNP) or other polymorphisms are known to have a modifier role in the pathogenesis of many human cancers (29) and may increase the risk of developing certain cancers or their aggressiveness (30). While several studies have examined the association between developing a cancer and its aggressiveness with the TP53 Arg72Pro - R72P (rs1042522), TP53 16bp duplication - PIN3 (rs17878362) and *MDM2* 309 SNP (rs2279744) polymorphisms, only few have examined their role in sarcomas.(31-34). Toffoli et al had shown that *MDM2* 309 T>G polymorphism increases the risk of osteosarcoma development in females and the prognostic value of *TP53* R72P SNP for overall survival in osteosarcoma (31). A meta analysis showed the association of PIN3 with cancer risk (35). With very few association studies of these SNPs in LFS or sporadic osteosarcoma, we studied their association in the first Indian LFS and sporadic sarcoma cohort. From an ongoing TISKS study, 311 sarcoma cases were analysed for the SNP genotype, phenotype and

clinical outcome. Of these, 200 had Osteosarcoma of the bone, 63 had other bone sarcomas and 48 had Soft Tissue Sarcoma (STS). Of these 265 cases received curative treatment at TMC with chemotherapy followed by surgery +/- radiotherapy. Distant metastases, mostly in lung were noted in 117 (37.5%) cases.

Age of sarcoma onset and genotype: The mean age of sarcoma onset in our cohort was 23.9 ±14.8 yrs (8- 89 yrs) with a significantly younger age of 21.3 (±12.5) yrs for bone sarcomas as compared to 38.2 (±18.3) yrs for STS ($p < 0.001$). In our cohort, females were diagnosed with sarcoma at a younger age as compared to males at a mean age of 22.4±15.2 yrs versus 24.6±14.6 years ($p=0.003$). While younger age of sarcoma onset in females is known, the gender specific influence of these SNPs on age of diagnosis has not been studied. Of the three polymorphisms studied, only R72P GG genotype showed a strong association between younger mean age of sarcoma onset in females at 18±10.3yrs as compared to males at 29.1±17.3yrs ($P<0.001$).

***Tumor Size and genotype*:** The association studies of these 3 polymorphism with the tumor size classified as <8 or >=8 cm revealed a significant association between R72P C allele with larger tumours in the entire sarcoma cohort of 311 cases ($P=0.038$ with all three genotypes and $P=0.052$ in dominant model) and in the bone sarcoma cohort of 263 cases ($P=0.024$ with all three genotypes and $P=0.035$ in dominant model). In the smaller osteosarcoma cohort of 200 cases, the trend of larger tumours with the C allele did not reach statistical significance ($p=0.089$). The MDM2 and PIN3 polymorphisms were not found to be associated with tumor size. This is the first study reporting significant association of R72P SNP with tumour size in any cancer.

***Association of genotype with chemotherapy response*:** No association was found between these polymorphism and >90% response to chemotherapy in the entire cohort and in osteosarcomas.

Overall and Metastases Free Survival and genotype: Kaplan Meier survival analysis was done for Overall survival (OS) and Metastasis Free Survival (MFS) was calculated for the whole group and in the OGS cases and the differences were compared using log rank test. The P53 Intron 3 duplication (PIN3) was associated with significantly worse OS at 3 years - 57.9% in 76 patients with heterozygous or homozygous duplication versus 82.6% in 122 patients with wild type genotype ($p<0.01$). The trend for better metastases free survival in wild type PIN3 did not reach statistical significance. This is the first reported association of worse survival with PIN3 duplication in Osteosarcoma.

<u>Objective 3:</u> Establishing LCL cell lines from LFS cases for DNA repair studies & variant characterization.

Lymphoblastoid Cell lines, a good source of genetic material are also useful for cytotoxic and DNA damage repair studies and show chromosomal aberration / genomic instability after irradiation (36). Altered DNA repair kinetics after irradiation of fibroblast cell lines from LFS cases has been seen (37) but it has not been studied in the LCLs established from LFS patients. We successfully established LCLs in 54/75 (72%) individuals with different hereditary cancer syndromes. Of these, 17 LCLs are from LFS patients of which 12 have a germline TP53 mutation. Gamma H2AX assay was performed on the LCLs made from 4 *TP53* mutation positive cases (2 missense, 1 frameshift and 1 splice site pathogenic *TP53* mutation) and 4 healthy controls. Significant difference between the DNA damage repair at 24 and 48 hrs after

2Gy radiation was seen between LCLs established from healthy control as compared to the 4 LCLs with heterozygous TP53 mutations (<0.001).

The utility of our LCLs was established through studies on characterization of a silent mutation (T125T, G>T substitution at the last base of exon 4) in the splice site region identified in our cohort. In silico prediction of such variants is hazardous as they will cause aberrant splicing only in the presence of a cryptic splice site and need experimental evidence for aberrant splicing (38) which requires constant source of fresh cells for repeated experiments. To express p53 in LCLs, they were treated with puromycin to block nonsense mediated RNA decay, given 2Gy radiation and whole RNA were extracted after 30 minutes. cDNA synthesis was performed and after PCR sequencing was performed. Transcript analysis showed that this mutation leads to aberrant splicing and skipping of 200bp of Exon 4 in the final transcript thus leading to a truncated protein.

We could successfully establish a large panel of LCLs from LFS patients with different germline *TP53* mutations to serve as a continuous source of genetic material, for DNA repair kinetics study and demonstrated its utility in characterizing a synonymous variant identified in our cohort. In future, these LCLs can be used for detailed functional studies comparing different types of *TP53* mutations and for cytotoxic studies.

SUMMARY AND CONCLUSIONS

This is the first report of tumor and mutation spectrum from an Indian LFS/LFL cohort and the largest Asian LFS cohort studied so far. Findings of several distinct difference in the tumor and mutation spectrum in the Indian cohort brings out the need

for detailed genotype – phenotype correlation in different populations. In this Indian cohort, ACC, skin, and sarcomas are less frequent but head and neck, GI, hematological and breast cancer are more frequent. Genotype-phenotype correlation reveals that the cases with dominant negative missense mutation and truncating type of mutations have significantly early age of onset with respect to other missense mutations. Screening of *TP53* gene mutation should not be restricted to the classification criteria of LFS/LFL as the mutation detection rate is around 13% in cases which do not fulfill the criteria of LFS/LFL. Our study raises an alert for all genetic testing laboratory to take in to account the possibility of allele dropouts which may lead to false negative results. In one of the largest sarcoma cohort reported from anywhere and the first from India cohort evaluating association of TP53 and MDM2 polymorphism with clinical phenotype and outcome, we identified several significant associations between polymorphisms with clinical phenotypes and survival outcome. This merits further studies in larger cohorts. Large numbers of LCLs were successfully established from this rare syndrome as a source of genetic material and characterization of *TP53* variants and other functional studies.

References:

1. Li FP, Fraumeni JF, Jr. Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? Annals of internal medicine. 1969;71(4):747-52.

2. Birch JM, Hartley AL, Tricker KJ, Prosser J, Condie A, Kelsey AM, et al. Prevalence and Diversity of Constitutional Mutations in the p53 Gene among 21 Li-Fraumeni Families. Cancer research. 1994;54(5):1298-304.

3. Chompret A, Abel A, Stoppa-Lyonnet D, Brugieres L, Pages S, Feunteun J, et al. Sensitivity and predictive value of criteria for p53 germline mutation screening. Journal of medical genetics. 2001;38(1):43-7.

4. Varley JM. Germline TP53 mutations and Li-Fraumeni syndrome. Human mutation. 2003;21(3):313-20.

5. Lane DP. Cancer. p53, guardian of the genome. Nature. 1992;358(6381):15-6.

6. Dittmer D, Pati S, Zambetti G, Chu S, Teresky AK, Moore M, et al. Gain of function mutations in p53. Nature Genetics. 1993;4:42.

7. Milner J, Medcalf EA. Cotranslation of activated mutant p53 with wild type drives the wild-type p53 protein into the mutant conformation. Cell. 1991;65(5):765-74.

8. Varley JM, Evans DG, Birch JM. Li-Fraumeni syndrome--a molecular and clinical review. British journal of cancer. 1997;76(1):1-14.

9. Bougeard G, Baert-Desurmont S, Tournier I, Vasseur S, Martin C, Brugieres L, et al. Impact of the MDM2 SNP309 and p53 Arg72Pro polymorphism on age of tumour onset in Li-Fraumeni syndrome. Journal of medical genetics. 2006;43(6):531-3.

10. Maher B. Personal genomes: The case of the missing heritability. Nature. 2008;456(7218):18-21.

11. Nagrani R, Mhatre S, Rajaraman P, Chatterjee N, Akbari MR, Boffetta P, et al. Association of Genome-Wide Association Study (GWAS) Identified SNPs and Risk of Breast Cancer in an Indian Population. Scientific reports. 2017;7:40963.

12. Gonzalez KD, Noltner KA, Buzin CH, Gu D, Wen-Fong CY, Nguyen VQ, et al. Beyond Li Fraumeni Syndrome: clinical characteristics of families with p53 germline mutations. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(8):1250-6.

13. Malkin D. p53 and the Li-Fraumeni syndrome. Cancer genetics and cytogenetics. 1993;66(2):83-92.

14. Bougeard G, Renaux-Petel M, Flaman JM, Charbonnier C, Fermey P, Belotti M, et al. Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2015;33(21):2345-52.

15. Formiga M, de Andrade KC, Kowalski LP, Achatz MI. Frequency of Thyroid Carcinoma in Brazilian TP53 p.R337H Carriers With Li Fraumeni Syndrome. JAMA oncology. 2017;3(10):1400-2.

16. Liu Y, Bodmer WF. Analysis of P53 mutations and their expression in 56 colorectal cancer cell lines. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(4):976-81.

17. Kato S, Han S-Y, Liu W, Otsuka K, Shibata H, Kanamaru R, et al. Understanding the function–structure and function–mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. Proceedings of the National Academy of Sciences. 2003;100(14):8424-9.

18. Evans DG, Birch JM, Thorneycroft M, McGown G, Lalloo F, Varley JM. Low rate of TP53 germline mutations in breast cancer/sarcoma families not fulfilling

classical criteria for Li-Fraumeni syndrome. Journal of medical genetics. 2002;39(12):941-4.

19. Mitchell G, Ballinger ML, Wong S, Hewitt C, James P, Young MA, et al. High frequency of germline TP53 mutations in a prospective adult-onset sarcoma cohort. PloS one. 2013;8(7):e69026.

20. Diller L, Sexsmith E, Gottlieb A, Li FP, Malkin D. Germline p53 mutations are frequently detected in young children with rhabdomyosarcoma. The Journal of clinical investigation. 1995;95(4):1606-11.

21. Macedo GS, Araujo Vieira I, Brandalize AP, Giacomazzi J, Inez Palmero E, Volc S, et al. Rare germline variant (rs78378222) in the TP53 3' UTR: Evidence for a new mechanism of cancer predisposition in Li-Fraumeni syndrome. Cancer genetics. 2016;209(3):97-106.

22. Taylor JGt, Cheuk AT, Tsang PS, Chung JY, Song YK, Desai K, et al. Identification of FGFR4-activating mutations in human rhabdomyosarcomas that promote metastasis in xenotransplanted models. The Journal of clinical investigation. 2009;119(11):3395-407.

23. Blais J, Lavoie SB, Giroux S, Bussieres J, Lindsay C, Dionne J, et al. Risk of Misdiagnosis Due to Allele Dropout and False-Positive PCR Artifacts in Molecular Diagnostics: Analysis of 30,769 Genotypes. The Journal of molecular diagnostics : JMD. 2015;17(5):505-14.

24. Silva FC, Torrezan GT, Brianese RC, Stabellini R, Carraro DM. Pitfalls in genetic testing: a case of a SNP in primer-annealing region leading to allele dropout in BRCA1. Molecular genetics & genomic medicine. 2017;5(4):443-7.

25. Stevens AJ, Taylor MG, Pearce FG, Kennedy MA. Allelic Dropout During Polymerase Chain Reaction due to G-Quadruplex Structures and DNA Methylation Is Widespread at Imprinted Human Loci. G3. 2017;7(3):1019-25.

26. Soulsbury CD, Iossa G, Edwards KJ, Baker PJ, Harris S. Allelic dropout from a high-quality DNA source. Conservation Genetics. 2007;8 (3):733-8.

27. Dolan SM, Goldwaser TH, Jindal SK. Preimplantation Genetic Diagnosis for Mendelian Conditions. Jama. 2017;318(9):859-60.

28. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic acids research. 2006;34(Web Server issue):W676-82.

29. Brookes AJ. The essence of SNPs. Gene. 1999;234(2):177-86.

30. Petitjean A, Achatz MI, Borresen-Dale AL, Hainaut P, Olivier M. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. Oncogene. 2007;26(15):2157-65.

31. Toffoli G, Biason P, Russo A, De Mattia E, Cecchin E, Hattinger CM, et al. Effect of TP53 Arg72Pro and MDM2 SNP309 polymorphisms on the risk of high-grade osteosarcoma development and survival. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15(10):3550-6.

32. Paulin FE, O'Neill M, McGregor G, Cassidy A, Ashfield A, Ali CW, et al. MDM2 SNP309 is associated with high grade node positive breast tumours and is in linkage disequilibrium with a novel MDM2 intron 1 polymorphism. BMC cancer. 2008;8:281.

33. Walsh CS, Miller CW, Karlan BY, Koeffler HP. Association between a functional single nucleotide polymorphism in the MDM2 gene and sporadic endometrial cancer risk. Gynecologic oncology. 2007;104(3):660-4.

34. Lum SS, Chua HW, Li H, Li WF, Rao N, Wei J, et al. MDM2 SNP309 G allele increases risk but the T allele is associated with earlier onset age of sporadic breast cancers in the Chinese population. Carcinogenesis. 2008;29(4):754-61.

35. Hu Z, Li X, Qu X, He Y, Ring BZ, Song E, et al. Intron 3 16 bp duplication polymorphism of TP53 contributes to cancer susceptibility: a meta-analysis. Carcinogenesis. 2010;31(4):643-7.

36. Schuler N, Palm J, Schmitz S, Lorat Y, Rube CE. Increasing genomic instability during cancer therapy in a patient with Li-Fraumeni syndrome. Clinical and translational radiation oncology. 2017;7:71-8.

37. Little JB, Nove J, Dahlberg WK, Troilo P, Nichols WW, Strong LC. Normal cytotoxic response of skin fibroblasts from patients with Li-Fraumeni familial cancer syndrome to DNA-damaging agents in vitro. Cancer research. 1987;47(15):4229-34.

38. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic acids research. 2009;37(9):e67.

Publications in Refereed Journal:

a. Published: Identification and characterization of *TP53* gene Allele Dropout in Li-Fraumeni syndrome and Oral cancer cohorts. (**Haque MM**, Kowtal P, Sarin R. *Scientific Reports*, 2018 doi10.1038/s41598-018-30238-7)

b. Accepted:

c. Communicated: Novel TP53 Genotype-Phenotype associations in a large south Asian cohort with modified criteria for Li-Fraumeni Like Syndrome. **Haque MM**, Kowtal P, Reddy R, Puri A, Sarin R. Pre-submission enquiry sent to American Journal of Human Genetics

Other Publications:

a. Book/Book Chapter
b. Conference/Symposium
   1. **Poster Presentation**: Genotype – Phenotype spectrum in a large Indian Li-Fraumeni Syndrome (LFS) cohort. *European Human Genetics Conference*, Barcelona, Europe, 2016
   2. **Poster Presentation**: Contribution of germline *TP53* mutation in sporadic sarcoma cases, *TMC Platinum jubilee Conference*, Mumbai, India 2016.
   3. **Poster Presentation**: *TP53* germline mutation and its clinical correlation in a large Indian Li-Fraumeni syndrome (LFS) cohort, *IACR*, Jaipur, India 2015.
   4. **Poster Presentation**: Genetic anticipation and phenotypic features in Indian Li-Fraumeni Syndrome (LFS) families with *TP53* mutation., *Indian Cancer Genetics Conference and Training Workshop*, ACTREC-TMC, Navi Mumbai, India, 2013.

Md. Mafuzul Haque

Signature of Student: Date: 27/8/18

Doctoral Committee:

S. No.	Name	Designation	Signature	Date
1.	Dr. Sorab Dalal	Chairman	S.N. Dalal	28/8/18
2.	Dr. Rajiv Sarin	Guide/ Convener	Rajiv Sarin	27/8/18
4.	Dr. Tanuja Teni	Member	TR Teni	27.08.18
5.	Dr. Ajay Puri	Member	Puri	27/8/18

Forwarded Through:

S.V. Chiplunkar 29/8/18

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

4/9/18

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

Prof. S. D. Banavali
Director - Academics, TMC
Mumbai - 400012.

Dr. S. V. Chiplunkar
Director
Advanced Centre for Treatment, Research & Education in Cancer (ACTREC)

Version approved during the meeting of Standing Committee of Deans held during 29-30 Nov 2018.

LIST OF FIGURES

LIST OF TABLES

Abbreviation	
ADO	Allele Drop Out
AGVGD	Align-Grantham Variation Grantham Deviation
AIDS	Acquired Immuno Deficiency Syndrome
ALV	Avian Leukemia Virus
BAM	Binary Alignment Map
BSA	Bovine Serum Albumin
CGC	Cancer Genetics Clinic
CMMRD	Constitutional Mismatch Repair Deficiency
CT	Computed Tomography
CTD	Craboxy Terminal Domain
DAPI	4′,6-diamidino-2-phenylindole
DBD	DNA Binding Domain
DMSO	Dimethyl sulfoxide
DNA	Deoxyribo Nucleic Acid
DNE	Dominant Negative Effect
DQ	Dosage Quotient
DSB	Double Strand Break
EBV	Epstein Barr Virus
EDTA	Ethylene Diamine Tetra Acetic acid
FATHMM	Functional Analysis through Hidden Markov Models
FBS	Fertal Bovine Serum
FH	Ficol Hypaque
FNCLCC	Fédération Nationale des Centres de Lutte Contre Le Cancer (French Federation of Cancer Centers)
GATK	Genome Analysis Tool Kit
HBOC	Hereditary Breast and Ovarian Cancer
HeLa	Henrietta Lacks
HNPCC	Hereditary Non Polyposis Colorectal Cancer
HR	Homologous Recombination
IARC	International Agency for Research on Cancer

LCL	Lymphoblastoid Cell Line
LFL	Li Fraumeni Like Syndrome
LFS	Li Fraumeni Syndrome
LGR	Large Genomic Rearrangements
MAF	Minor Allele Frequency
MDM	Mouse Double Minute
MEN2	Multiple Endocrine Neoplasia Type 2
MLPA	Multiples Ligation dependent Probe Amplification
MRI	Magnetic Resonance Imaging
NCBI	National Centre for Biotechnoly Information
NCCN	National Comprehensive Cancer Network
NEB	New England Biolabs
NGS	Next Generataion Sequencing
NHEJ	Non Homologous End Joining
NMD	Nonsense Mediated RNA Decay
OGS	Osteogenic Sarcoma
PAR	Primer Annealing Region
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
RBC	Red Blood Cells
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribo Nucleic Acid
RSV	Rous Sarcoma Virus
SAM	Sequence Alignment Map
SEER	Surveillance, Epidemiology, and End Results
SNP	Single Nucleotide Polymorphism
SS	Sanger Sequencing
STS	Soft Tissue Sarcoma
SV40	Simian Virus 40
TBE	Tris Boric acid EDTA buffer
TISKS	TMC International Sarcoma Kindred Study

TMC	Tata Memorial Centre
TP53	Tumor Protein 53
TSG	Tumor Suppressor Gene
UCSC	University of California, Santa Cruz
UV	Ultra Violet
VCF	Variant Call Format
VUS	Variant of unknown significance
WES	Whole Exome Sequencing
WHO	World Health Organization

Chapter 1

Introduction

Cancer Genetics and *TP53* gene

1.1 Genetics inheritance and Human diseases:

Genetics is the branch of science which deals with the study of material that helps in the inheritance and explains the similarities and differences among individuals. The foundation of modern genetics was laid by the work of Gregor Mendel, a monk who discovered how traits are inherited in discrete units in 1865. This discovery of Mendel remained unnoticed till 1900, when it was rediscovered by three scientists independently DeVries, Correns, and von Tschermak in the background of better understanding of cells and chromosomes (1). Meanwhile in 1869 Fredrick Miescher isolated Deoxyribonucleic Acid (DNA) from pus cells and called it "nuclein" (2). In 1902 Walter Sutton observed that the segregation pattern of chromosomes during meiosis matches with the segregation pattern of Mendel's traits (3). In the same year a British physician Archibald Garrod observed Mendelian autosomal recessive inheritance pattern for the disease alkaptonuria (4). While it was well known that characters, traits and diseases are passed on from one generation to other but the unit of this inheritance was not known. In 1909 Wilhelm Johannsen coined the term 'gene' for this Mendelian unit of inheritance. He also used the term genotype for genetic traits and phenotype for external appearance (5). These units of inheritance i.e. genes being carried in chromosomes was first reported by Thomas Hunt Morgan in 1911 and discovered genetic linkage through studies in the fruit fly *Drosophila Melanogaster* (6). The term DNA was coined as early as 1881 by Albrecht Kossel and he also isolated the building blocks of DNA and Ribonucleic Acid (RNA) i.e. Adenine (A), Cytosine (C), Guanine (G), Thymine (T) and Uracil (U) (7). However it took several decades to arrive at the structure of DNA from the gene and chromosome. William Astbury, a British scientist in 1943 first showed that DNA has

a regular periodic structure through X-ray diffraction pattern of DNA and suggested that the nucleotides are stacked upon each other (8). Later in 1953 James Watson and Francis Crick proposed the model of double helical structure of DNA based on their own work and previous findings reported in isolation by Phoebus Levene who proposed the polynucleotide nature of DNA in 1919, Erwin Chargaff's work that the amount of purines (A+G) are equal to pyrimidines (T+C), Rosalind Franklin and Maurice Wilikins' demonstrated X-ray diffraction pattern of DNA (9).

In 1928 Frederick Griffith first suggested the phenomenon of transformation by performing experiment on mice using two strains of *Streptococcus Pneumonia.* He showed that mice infected with the virulent smooth (S) strain bacteria died whereas the non virulent rough (R) strain did not kill the mice. The heat killed S strain did not kill the mice but when it was mixed with the R strain, it killed the mice. This suggested the presence of something in the killed virulent S strain that could transform the non virulent R strain to a virulent strain (10). The 'transforming principle' observed by Griffith was further confirmed by Avery, Macleod and McCarty in 1944 who showed that the DNA and not the protein, transforms the bacterial cell (11). Several lines of enquiries on the gene, DNA and proteins were going on simultaneously and it was not clear what the genes were made of. The breakthrough came in 1952 when Alfred Hershey and Martha Chase showed that, it is only the viral DNA which enters into a bacterial cell to infect it which supported that the genes are made of DNA (12). The field of cytogenetics evolved in the early 20^{th} century and after several conflicting reports, Joe Hin Tjio & Albert Levan reported correctly the existence of 46 chromosomes in 1956 (13).

In the second half of 20th century, there was a spurt of activity to understand how DNA is synthesized in the cell. Various models of DNA replication were explained like dispersive model, conservative model and semi-conservative model. Only semi-conservative model of replication was accepted after Matthew Meselson and Franklin Stahl proved it through an ingenious experiment which was published in 1958 (14). In parallel to the advancements in the knowledge regarding genes and DNA, their role in producing distinct clinical phenotypes and diseases started unraveling. In 1959, Jerome Lejeune and his colleagues discovered that Down syndrome is caused by the trisomy of chromosome 21 (15).

While several hereditary diseases were known since centuries, the use of laboratory methods for their diagnostic confirmation or screening was initiated in second half of 20th century. In 1961 Robert Guthrie developed a new method for the first time to test newborns for the metabolic defect phenylketonuria (16). Perhaps the most important discovery in the genetic research was DNA sequencing technique. It was first discovered in 1975 by two groups independently, Fredrick Sanger & colleague and Alan Maxam & Walter Gilbert. Sanger method is widely used even today (17, 18). Though sequencing was discovered in 1975 but at that time the disease related genes were identified by mapping studies as exemplified by the discovery of genetic marker of Huntington's disease in 1983 by Gusella et al (19). The most important discovery which revolutionized the field of molecular biology was the polymerase chain reaction (PCR) first described by Kary Mullis in 1983 (20). Through several studies of genes, their transcripts and the coded proteins, the central dogma of molecular biology was firmly established. The major advances in the concepts and techniques in

genetics and molecular biology formed the basis of the most ambitious Human Genome Project initiated in 1990 and completed in 2003.

With time, several major discoveries were made establishing the genetic basis of various human diseases. In the year 1986, for the first time three specific genes were shown to be the cause for human hereditary diseases. This was made possible by mapping studies using positional cloning in DNA samples from patients with Chronic Granulomatous Disease, Muscular dystrophy and Retinoblastoma (21-23). Since then there have been major advancements in the techniques of molecular genetic analysis which has provided newer insight in to the basic and mechanistic understanding of the genetic basis of a large number of human pathologies. Molecular genetic analysis now plays an important role in disease research and has led to the development of the new branch of medicine – Molecular Pathology. While the genetic nature is clearly explained and established for several diseases, it remains poorly understood for many diseases or families showing Mendelian inheritance pattern Moreover, the existing animal models, including the transgenic mice do not always recapitulate the human phenotypic features (24). This underscores the need for detailed genetic analysis and genotype – phenotype correlation in large and well characterized human cohorts with specific disease phenotype.

1.2 Cancer

Cancer is a disease which occurs when a cell is transformed, and acquires the ability to undergo uncontrolled growth, invade the surrounding tissues and metastasize to distant organs of the body through the blood stream. Cancer can arise in any organ of the body and from any cell or tissue type. Tumors can be benign or malignant in

nature. Benign tumors do not metastasize and are not life threatening except in the brain. Malignant tumors, which are also called cancer, has the ability to invade the surrounding organs and metastasize to distant organs, which eventually results in death of the patient. Cancer cells are different from normal cells in many ways and summarized as 'Hallmarks of Cancer' by Wienberg and Hanahan (25, 26). The cancer cell stops following the normal signaling pathway, is able to continuously proliferate, invade and metastasize; evade apoptosis and immune system; secures continuous supply of nutrients, oxygen and energy through tumor neo-angiogenesis and deregulating cell energetics (27).

Cancer is a major public health issue in both, the developed and the developing parts of world (28). It is a matter of concern that the burden of cancer is expected to grow worldwide in coming years. Factors responsible for the increasing cancer incidences include established risk factors like tobacco, alcohol, obesity, certain infections, change of reproductive patterns and exposure to environmental and occupational hazards including chemical carcinogens, ionizing and non-ionizing radiations (29).

In India, every year more than 1 million new cases of cancers are diagnosed. This number is expected to double by 2035 (30). Most common cancers reported in Indian men are oral, lung, stomach, colorectal, pharyngeal and esophageal cancers while in women, breast, cervix and colorectum cancers are more common (30). Report of other cancers is rare with an incidence of around five per 100000 individuals in both males and females (30). In the developed world, while the cancer incidence is higher, the relative mortality rate is decreasing across all age groups. However there is no such evidence of decreasing mortality in developing countries like India (28). This may be attributed to the contrasting etiology like recurrent infections, tobacco chewing habits,

limited access to early diagnosis, treatment and lack of knowledge on preventive measures and prophylactic intervention (30)

1.2.1 Types of cancer

There are more than 100 different types of human cancers and there are further subtypes within a specific organ. There are various classifications of cancer including those based on the tissue of origin, based on the primary site of tumor and more recently based on molecular characteristics of the cancer. Cancers can be grouped into nine major categories (31):

- Carcinoma: Cancer originating from the epithelial cell, which is the internal and external lining of the body, is termed as carcinoma. These are the most common type of cancers accounting for around 80-90% of cancers. Two major subtypes of carcinoma are adenocarcinoma which originates in an organ or gland and squamous cell carcinoma which originates from the squamous epithelium.

- Sarcoma: Cancers that originates from the supporting or connective tissues are called as sarcoma. These types of cancers are mainly sub classified into bone sarcomas and soft tissue sarcomas (STS). Bone sarcoma mainly occurs in the younger age. Osteosarcoma is the most common sarcoma which is a type of bone sarcoma. Chondrosarcoma classified under bone sarcoma originates from cartilage. Some examples of STS are fibrosarcoma, liposarcoma, leiomyosarcoma, synovial sarcoma etc. Sarcoma is discussed in detail in chapter 5.

- Myeloma: Cancer that originates in the plasma cells of bone marrow is termed myeloma.

- Leukemia: Cancer that originates from the blood forming tissue of the bone marrow is termed leukemia. It produces a large number of immature white blood cells and can be acute or chronic leukemia.

- Lymphoma: Cancers of lymphatic tissue are termed as lymphomas. These may develop in the lymph nodes or other organs rich in lymphatic tissue. These are of two types, Hodgkin lymphoma and Non Hodgkin lymphoma.

- Brain and spinal cord related: Tumors that originates in either brain or spinal cord comes under this category. Names of different types of brain and spinal cord tumors are based on the cell of origin.

- Melanoma: Cancer of melanocytes is termed as melanoma which is the melanin pigment producing cells. Melanoma originates in the skin but rarely, they arise in other organs also.

- Mixed Types: These have mixed features with components from different histological categories. These include adenosquamous carcinoma, mixed mesodermal tumor, carcinosarcoma, and teratocarcinoma.

- Other types: Some other type of tumors which have different cell of origin like Germ cell tumors occurring in germ cells, Neuroendocrine tumors occurring in cells that release hormones in response to signals from the nervous system, etc.

Hereditary, Familial and Sporadic cancers

- **Hereditary Cancer** Hereditary cancers show Mendelian autosomal pattern of inheritance of one or more types of cancer within a family. In absence of family history, hereditary cancer is considered if the patient has characteristic phenotypic feature of a hereditary cancer syndrome e.g. Familial Adenomatous Polyposis or Multiple Endocrine Neoplasia (MEN2B). In absence of Mendelian pattern of inheritance or characteristic syndromic features, cancer is sometimes suspected to be hereditary if the age of diagnosis is much younger for that cancer type, multiple primary or bilateral cancers in an individual or cancers of certain specific histological or molecular features. Hereditary cancers account for 2-5% of all cancers in general. However for certain cancers like retinoblastoma it accounts for 25% of all cases. These are monogenetic Mendelian disorders due to high penetrance germline mutation in one of the several dozen well characterized genes (32).

- **Familial Cancer**: This term is used when more cases of a specific type(s) of cancers are seen within a family than is statistically expected, but no specific pattern of inheritance is evident. In familial cancer the age of cancer diagnosis is variable. Familial cancers account for 10-20% of all cancers and may result from common genetic background, similar environment and or lifestyle factors (32).

- **Sporadic Cancer**: These are the most common cancers accounting for over 80% of all cancers. They are due to non-hereditary life style causes, which may or may not be evident. They do not have a family history of cancer and tend to occur at older age. These cancers may be due to the interaction of environmental or

endogenous carcinogen with low penetrance genetic variation such as polymorphisms in the xenobiotic metabolizing enzymes (33, 34).

1.2.2 Hereditary Cancer Syndromes

Highly penetrant germline mutations in one of the several known cancer associated genes can result in specific 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) or 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, Breast*, Brain, Blood, Adrenal gland	*TP53*
Familial Adenomatous Polyposis (FAP)/	Autosomal Dominant	Colorectum, Upper GI, Brain (medulloblastoma***), Papillary	*APC/ MUTYH*

Attenuated FAP (AFAP)		Thyroid***, Hepatoblastoma***	
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

1.3 Cancer- a genetic disease

The agents first reported to be responsible for transformation of cells were viruses. The first virus of this series was Rous Sarcoma Virus (RSV) which causes transformation in chickens (35). The cell free filtrate from the tumor of one chicken was able to develop tumor in a healthy chicken due to which Dr. Rous concluded that this filtrate causes cancer in chickens. This filtrate has viruses, later known as RSV in the name of Peyton Rous. Though the virus first discovered to be responsible for cancer was Avian Leukemia Virus (ALV) by two Danish scientists, Vilhelm Ellermann and Olaf Bang in 1908 ignored as leukemia was not believed to be cancer. ALV was not believed to be tumor virus for at least next 40 years (36-38). The discovery by Rous was also ignored for many decades as it was thought that the filtrate used by him may have got contaminated with the tumor cells and hence

produced tumors in chicken. The focus assay developed in 1958 revolutionized the field and helped in rapid advancement in virology (39). Many RSVs were characterized, and it was found that some strains of RSVs were able to transform cells and the other could not. By comparing the RNA of the transforming and non transforming strains of RSVs, it was first identified that the RNA of viruses that could transform cells was 20% larger in size than the RNA from the virus strains that could not transform cells (40). Bishop and Varmus first showed in 1976 that the cDNA part of the virus genome responsible for transformation was successfully hybridized with the normal chicken cells. This was a strong evidence that the part of virus genome responsible for transformation was actually a cellular gene (41). This led to the hypothesis that retroviral infection is not sufficient for transformation but a mutation in the proto-oncogene may lead to cancer. In 1979 Weinberg showed that transfection of DNA and chromatin from human tumor cells could also transform the NIH3T3 cells indicating that there is something in the DNA that can cause cancer(42, 43). The first human proto-oncogene was discovered in 1982 when different groups showed that a point mutation in *H-RAS* leads to its activation and gain of oncogenic potential (44-46). Proto-Oncogenes after acquiring gain of function mutations are called oncogene. More than 70 proto-oncogenes of cellular origin have been identified so far and most of these are related to either cell proliferation or apoptosis (47).

A DNA virus, Simian Virus 40 (SV40), was also shown to have transforming capabilities which depend on the expression of viral large tumor (T) antigen (48-50). In 1979 David Lane and colleagues using antibodies against the large T-antigen of SV40 showed that along with the SV40 large T-antigen, another cellular protein was also precipitated. This protein was called p53 as its molecular weight was

approximately 53 KDa (51, 52). This was the first evidence that the DNA viruses transform cells through interacting with the cellular proteins. As the p53 protein was overexpressed in the tumor cell, it was believed to be an oncogene. Later Vogelstein reported loss of heterozygosity of the p53 locus in human colorectal cancer and suggested that it is a Tumor Suppressor Gene (53). The Tumor Protein 53 (*TP53*) gene which codes for the p53 protein is the most commonly mutated gene in human cancer tissue. This paved way for the discovery of several other tumor suppressor genes in the next two decades.

These seminal studies spanning over 8 decades firmly established that cancer is a genetic disease occurring due to mutations in two types of cellular genes - proto-oncogenes and the Tumor Suppressor Genes (TSG).

1.3.1 Proto-oncogenes and Tumor Suppressor Genes

Accumulation of several genetic and epigenetic alterations in proto-oncogenes and TSGs which participates in cell proliferation are drivers for the development of cancer (54). Cells with these alterations have a growth and survival advantage over their normal counterparts and are naturally selected to evolve into cancer (55). Genes involved in the process of carcinogenesis are generally those that regulate cell proliferation, apoptosis or involved in DNA damage repair process.

1.3.1.1 Oncogenes

Proto-Oncogenes encode for proteins which regulate cell proliferation. Gain of function point mutations or amplification of these genes drives the cell for continued proliferation and results in the neoplastic phenotype. The mutations that convert a proto-oncogene to oncogene are dominant activating, gain-of-function mutations

which imply that mutation in one copy of the gene is sufficient to initiate carcinogenesis. These mutations includes: (i) point mutations, resulting in constitutively activated gene and its protein, (ii) gene duplications or amplification leading to its over-expression or (iii) chromosomal rearrangements that brings a growth-regulatory gene under the control of a different promoter resulting in altered gene expression (56). All these alterations confer growth and survival advantage to cells.

Based on the functional properties, oncogenes can be classified as: (i) Growth factors: Constitutive activation of growth factor genes can lead to malignant transformation. A well characterized example is Platelet-derived growth factor that stimulates proliferation of fibroblasts. (ii) Growth factor receptors: Alterations in growth factor receptor gene leads to their constitutive activation in the absence of ligand binding. The examples include Epidermal Growth Factor Receptor (EGFR) (iii) Signal transducers: Binding of ligand to its receptor leads to autophosphorylation of the kinase domain of the receptor which in turn activates several downstream signaling pathways. Many oncogenes encode members of signal transduction pathways such as SRC, ABL, AKT, RAF etc. (iv) Transcription factors: They are members of multigene families which regulate expression of several genes that control cell division. The examples include Fos and Jun (v) Chromatin remodelers: Modifications in chromatin structure and compaction play a critical role in the control of gene expression, replication and repair and any alterations within these can lead to neoplastic transformation. (vi) Apoptosis regulators: Various apoptotic regulator proteins like c-Myc, c-Fos/c-Jun and bcl2 are involved in either pro or anti-apoptotic regulation of cell (57, 58).

1.3.1.2 Tumor suppressor genes

Activation of oncogenes represents only one of the two distinct genetic alterations involved in tumor initiation and progression, the other is inactivation of TSGs (59). TSGs can be defined as genes which encode proteins that regulate cell proliferation by acting as the 'brakes' for cell division. Their normal function is to inhibit excessive cell proliferation (59). In cancer, these genes are either lost or inactivated, leading to removal of the brakes that act as negative regulators of cell proliferation resulting in abnormal cell proliferation and tumor development. Therefore, mutations in TSGs are termed as loss-of-function mutations. These have a recessive mode of inheritance at cellular level which implies that both the copies of the TSG needs to be altered in a 'two hit' fashion to initiate tumorigenesis (Knudson two hit hypothesis) (60, 61).

The first insight into the discovery of TSGs came from Henry Harris and his colleagues in 1969 during their somatic cell hybridization experiments, where they fused normal cells with tumor cells generating hybrid cells with chromosomes from both the parent cells. They observed that in most cases the hybrid cells were not capable of forming tumors suggesting that the inhibitory effect could be because of the normal copy of genes derived from normal parent cell that suppresses tumor development (59).

The first TSG to be identified was Rb through studies on retinoblastoma, a rare childhood eye cancer (62). This served as prototype for identification of several other TSGs including *TP53* as discussed later. Mutations in tumor suppressor genes are loss-of-function mutations. The loss of function mutations are generally truncating mutations such as nonsense or frameshift small insertion or deletions; splice site mutations; and less frequently missense mutations or large genomic rearrangements.

These mutations affect either the genes that directly regulate cell proliferation such as Rb and p53 or those that do not directly control cell proliferation, but mutations in these genes are compromised in their ability to repair DNA damage and thus can acquire mutations in other genes.

Kinzler and Bert Vogelstein in 1997 grouped the TSGs into two classes: "caretakers" and "gatekeepers" (63). A third classification of TSGs was proposed in 2004 as "landscaper" genes (64).

- ***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.

1.4 *TP53* gene and protein:

TP53 gene located on the short arm of chromosome 17 (17p13.1) spans approximately 20 kb of genomic DNA. It is composed of 14 exons which includes 10 coding exons (exon 2 to 11) which translates into the full length *TP53* protein, one non-coding exon (exon 1) and three alternative exons (exons 2/3, 9β, 9γ) (65). Interestingly, *TP53* gene has a 10kb long Intron 1 which constitutes 50% of the total genomic DNA of *TP53* gene (20kb). The gene has three promoters of which two promoters P1' (located 250bp upstream of exon 1) and P1 (located in intron 1) encode complete p53 and Δ40p53 respectively. The third promoter i.e. P3 lies in the intron 4 of *TP53* and encodes protein Δ133p53. These three promoters encode 12 different isoforms of p53 through alternative splicing and alternative use of these promoters (66).

Tumor protein 53 (*TP53*) also known as p53 or cellular tumor antigen p53 (Uniprot name), is a 393 amino acids long protein with several functional domains. The name p53 describes the apparent molecular mass; SDS-PAGE analysis indicates that it is a 53-kilodalton (kDa) protein. However, the actual mass of the full-length p53 protein (p53α) is only 43.7 kDa (67). This difference is due to the high number of proline residues in the protein, which slow its migration on SDS-PAGE. The p53 protein consists of an N-terminal transcriptional activation domain, proline-rich region (SH3 domain), DNA binding domain, tetramerization domain and C-terminal regulatory domain (Figure 1.1). Each domain has specific functions and *TP53* is therefore a multifunctional protein. P53 protein functions in a tetramer form and this assembly of proteins occur through the amino acid residues from 323 to 356 located in tetramerization domain of the protein (68). The protein consists of 5 conserved

domains locating in the N-terminal and central region of the protein. The first conserved region is located in the transactivation domain of the protein and the other four conserved domains of the protein is located in the DNA binding domain of the protein(69).

The DNA binding domain of the protein consists of two antiparellel β-sheets which is composed of four and five β-strands and a loop-sheet-helix motif is sandwiched in between these β-sheets. Two large loops at the end of the β sandwich are held by tetrahedrally coordinated zinc atom. DNA mainly interact with the loop-sheet-helix motif and the two loops at the end of DNA binding domain (70).

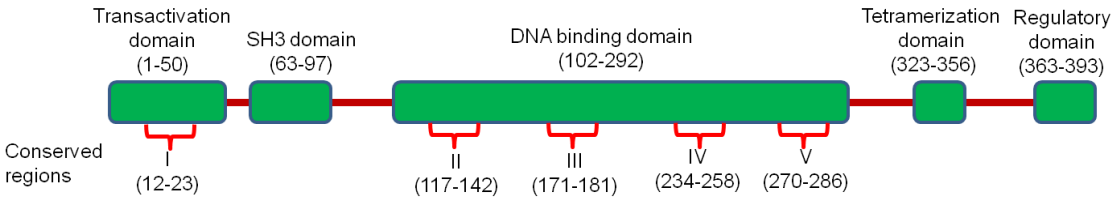


Figure 1.1: p53 protein domains and conserved regions. Roman numerals represent five conserved regions of protein. In bracket amino acid position are represented.

As many tumors were found to express higher levels of *TP53* encoded protein (71, 72), it was initially proposed that *TP53* is an oncogene. However, later it was shown that there was mutational inactivation of one *TP53* allele with loss of the second wild-type allele in two colorectal cancers studied (73). In the same year, Finlay et al showed that expression of wild-type *TP53* could inhibit oncogene driven cellular transformation (74). These studies collectively confirmed that *TP53* is a tumor suppressor gene and not an oncogene as it was thought initially.

1.4.1 p53 and mechanism of action

TP53 gene is transcribed whenever the cells encounter any stress condition. Initially Reich et al in 1984 have shown that on serum starvation the mRNA and protein level goes down but on stimulation with serum the level rises and reach to its highest prior to G1/S phase of cell cycle i.e. just before DNA replication so that if any damage occurs then the guardian of the genome is ready to show its parental care (75). c-Myc, WRAP53, C/EBPβ, RBP-Jκ etc are the reported proteins that participate in the transcription of *TP53* gene (76-78). The level of p53 in cell is also positively and negatively regulated by various miRNAs (79).

In unstressed cells, the levels of *TP53* protein are kept under check through a negative-regulatory feedback mechanism which is mediated by Mouse Double Minute 2 homolog (MDM2), an E3 ubiquitin ligase whose expression is induced by *TP53*. MDM2 interacts with the amino terminus of the p53 protein, then interacts with the DNA Binding Domain (DBD) of p53 and further the lysine rich region present in the C-terminal domain (CTD) are ubiquitinated for the p53 protein's proteosomal degradation (Figure 1.2). The biology of MDM2 interaction with CTD of p53 is not well understood (80). Therefore under normal conditions in the cell, *TP53* is very unstable, with a half-life ranging from 5 to 30 min (81). However, whenever the cell faces genotoxic stress, such as ionizing radiation or oxidative stress, *TP53* and MDM2 undergoes phosphorylation leading to a decrease in MDM2-*TP53* interaction. This decreased interaction further results in the reduced *TP53* degradation, allowing the accumulation of *TP53* protein inside the cell. *TP53* molecules then self-associate to form the functional tetramers which induce the expression of numerous downstream

target genes that regulate critical cellular processes, such as cell cycle arrest, apoptosis and DNA repair.

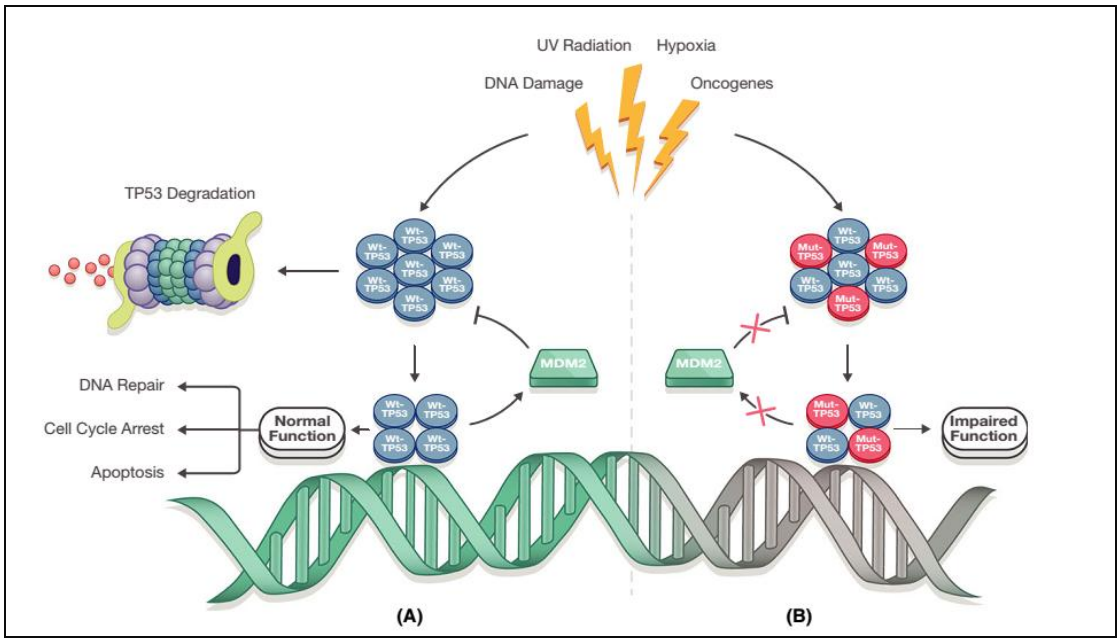


Figure 1.2: Functions of *TP53* (Taken from Ref (82))

(A) In response to cellular stress signals, wild-type *TP53* tetramers bind to DNA (in green) and trigger the activation of genes responsible for a myriad of functions including DNA repair, cell cycle arrest, apoptosis and senescence. During the course of these activities, *TP53* induces expression of MDM2, leading to *TP53* degradation through proteasomal degradation. (B) In contrast to wild-type *TP53*, mutant *TP53* proteins cannot bind effectively to DNA (in grey), resulting in impaired functions, including poor induction of MDM2 and hence *TP53* accumulation.

Activation of p53 depends on various types of stress. p53 after activation participates in several vital cellular functions like cell cycle arrest, DNA damage repair, apoptosis etc as shown in figure 1.3. p53 being a transcription factor activates many genes in response to the cellular stress to save the cell or to send the irreparable cell to apoptosis.

In 2006 Wei et al have produced a global map for the p53 binding site in human genome. They achieved this by performing chromatin immunoprecipitation assay to isolate the DNA fragment where p53 binds and further by paired end tag sequencing they identified around 542 binding loci with high confidence. Of these 542 loci, 98 genes were not known to be transactivated by p53 (83), highlighting the diverse role of the *TP53* gene.

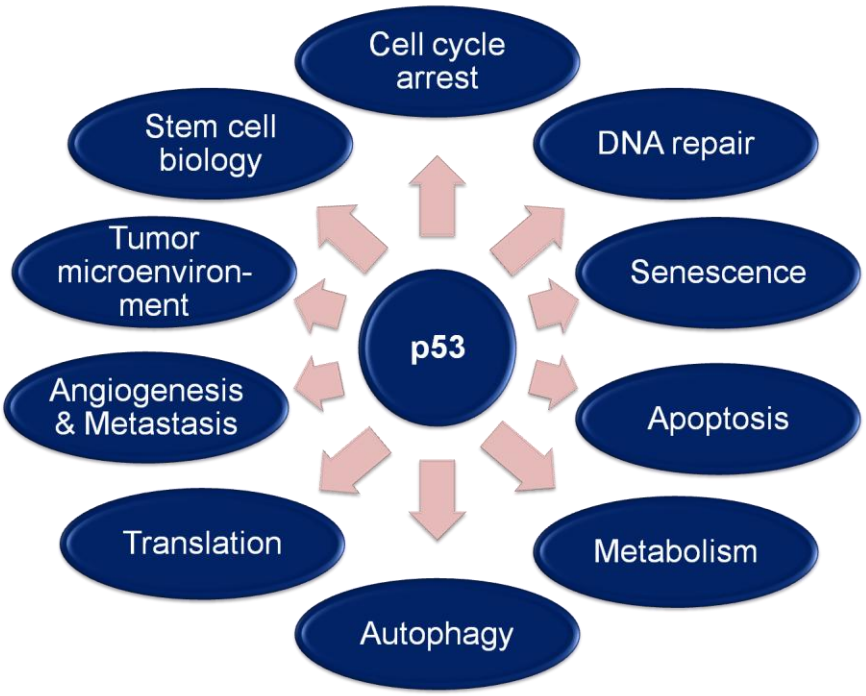


Figure 1.3: Diverse role of p53 (Guardian of the genome.)

1.4.2 *TP53* mutations and cancer

TP53 gene is found to be mutated in more than 50% of tumors (84). Germline mutation in *TP53* causes Li Fraumeni Syndrome (LFS) as described later in chapter-3 (85, 86). International Agency for Research on Cancer (IARC) maintains a

comprehensive database for all somatic and germline mutations in *TP53* gene and its clinical correlation.

1.4.2.1 Somatic mutations in *TP53*

Around 28000 somatic mutations are documented in IARC database. The figure 1.4 taken from IARC database shows the distribution of mutations codon (Figure 1.4A) and exon (Figure 1.4B) wise. All hot spot codon 175, 245, 248, 249, 273 and 282 lie in the conserved region (region III, IV, and V) of the p53 protein. Around 90% of the mutations are in these highly conserved regions of the gene.

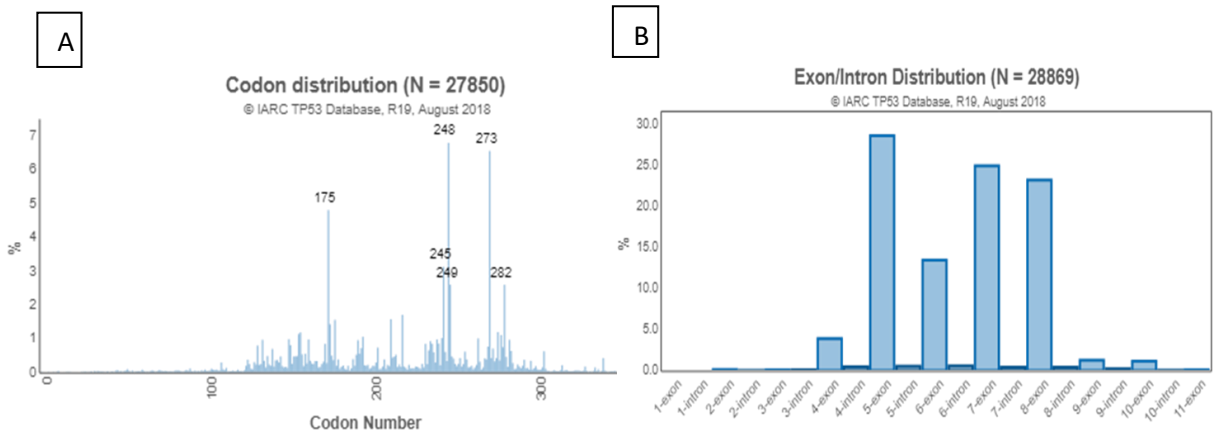


Figure 1.4: Distribution of somatic mutations reported in IARC database (R19). (A) Codon ; (B) exon / intron distribution

It is notable that 73% of the mutations reported in IARC R19 release are missense mutations (Figure 1.5A). Of the point mutations, 86% are missense mutations (figure 1.5B). *TP53* gene is very frequently mutated gene in almost all human cancers with the highest somatic mutation frequency in colorectal cancers and head and neck cancers (figure 1.6).

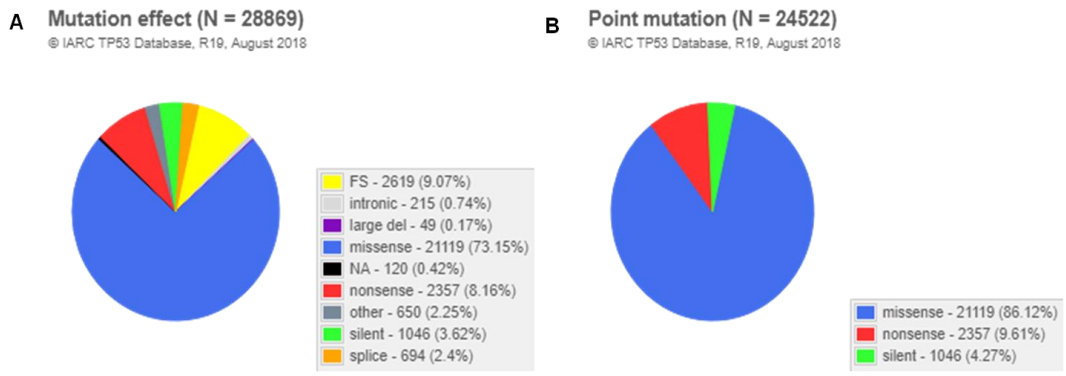


Figure 1.5: Types of somatic mutations reported in IARC R19. (A) Mutation Effect; (B) Point mutations

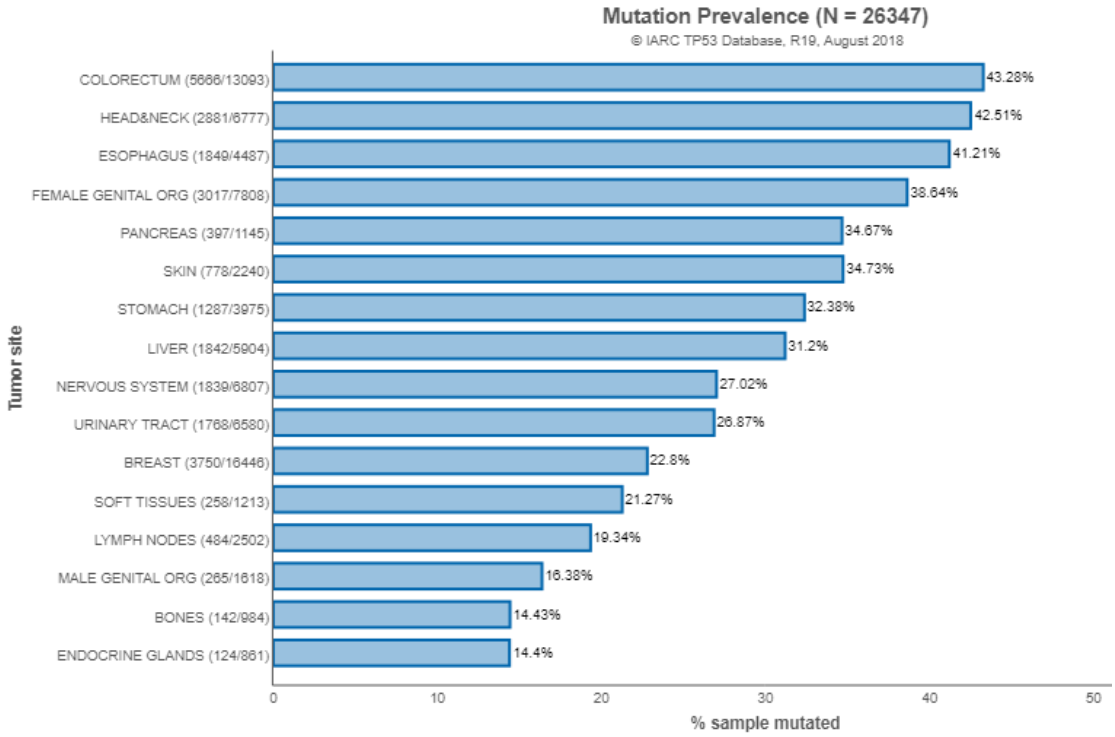


Figure 1.6: Mutation prevalence in different tumor types

1.4.2.2 Germline mutations in *TP53*

Germline mutations in *TP53* gene are responsible for Li Fraumeni Syndrome (LFS) / Li Fraumeni Like Syndrome (LFL). LFS is a highly penetrant cancer predisposition

syndrome in which diverse type of cancers are reported with the occurrence of sarcoma, breast cancer, brain cancer, leukemia, and adrenocortical cancers with very high frequency (87). The details of the syndrome described in chapter 3. Role of *TP53* as the gene responsible for LFS was first reported by David Malkin and Srivastava in 1990 (85, 86). IARC database documents the distribution of germline mutations as shown in figure 1.7.

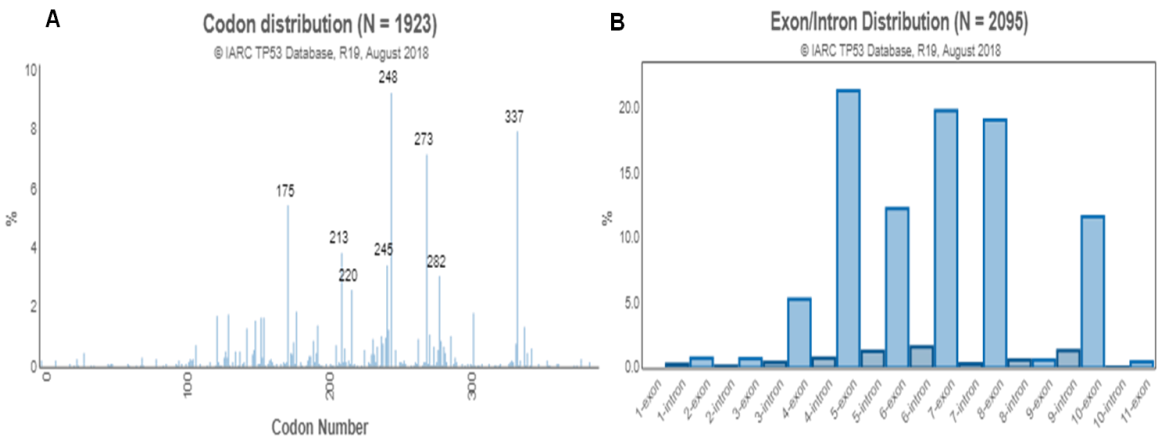


Figure 1.7: Distribution of germline mutations reported in IARC database (R19). (A) Codon ; (B) exon / intron distribution

There are some remarkable differences in the somatic mutations and germline mutations. Hotspot codons 175, 248, 273 and 282 are common in both types but the codon 249 mutation is not found to be a germline hot spot while codon 337 is only a germline hotspot and not a somatic one. Codon 337 is a germline founder mutation of Brazilian population. Amongst germline mutations the types are almost similar to somatic mutation with around 75% of mutations being missense (figure 1.8). Cancer distribution in carriers of germline TP53 mutation in the IARC database is shown in figure 1.9 and discussed later in chapter 3.

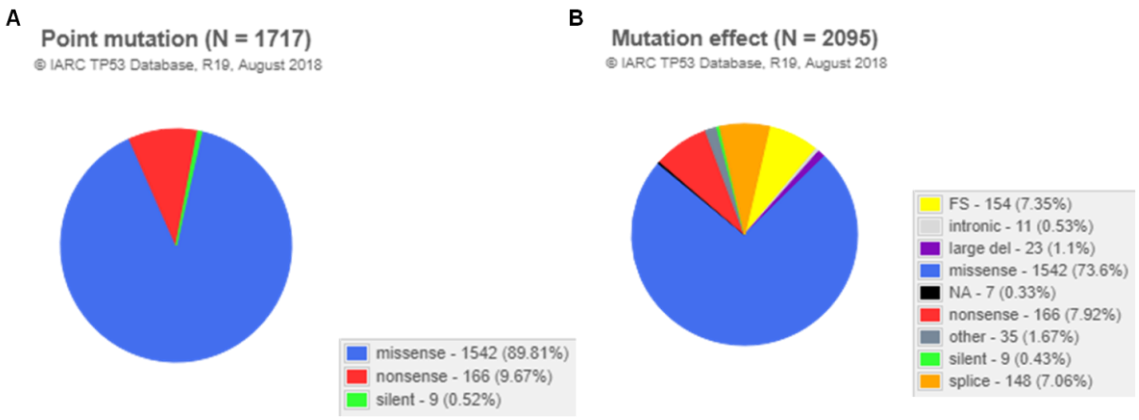


Figure 1.8: Types of germline mutations reported in IARC R19. (A) Mutation Effect; (B) Point mutations

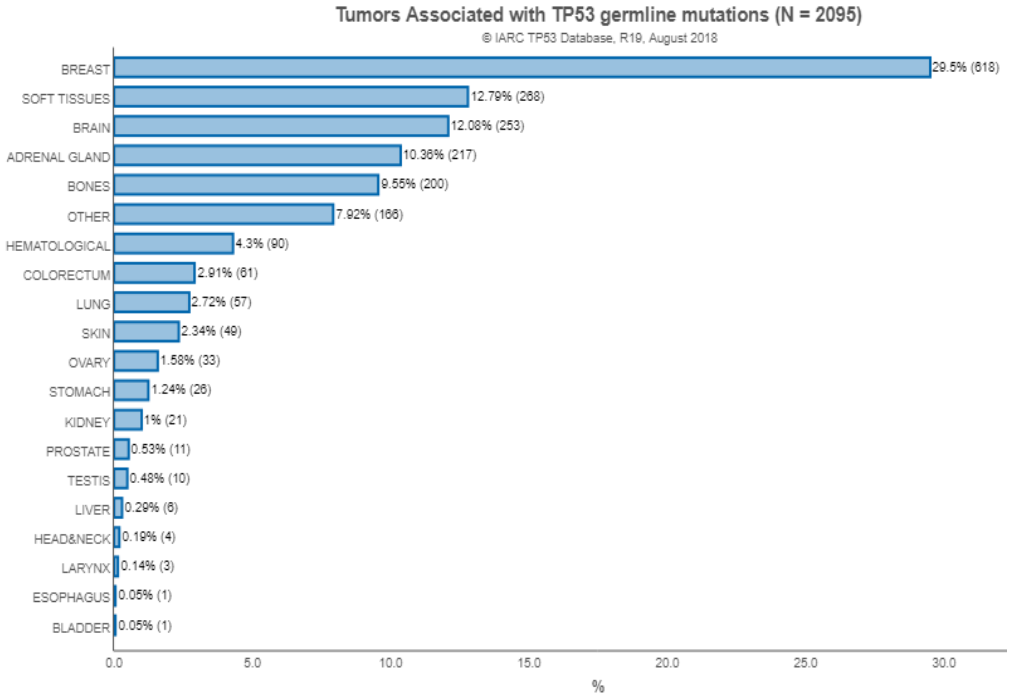


Figure 1.9: Tumors associated with germline *TP53* mutation. Tumors in carriers of *TP53* germline mutation in IARC database (R19)

1.4.3 Function of mutant *TP53*

Unlike other TSGs where loss of function truncating mutations are more frequent, majority of mutations in *TP53* gene are missense mutations. Some missense mutations give the p53 protein a functional advantage while other missense mutation leads to the

loss of function. Different functional consequences of *TP53* mutations can be broadly categorized into:

- Loss of function mutations: These mutations lead to the complete loss of function of the protein. These mutations are nonsense mutations, frameshift mutations, missense mutations and large genomic rearrangements (88, 89).

- Gain of function (GOF) mutations: These are the mutations that give an extra function to the protein. They may transactivate genes which are not otherwise transcribed by p53; may gain the ability to transcribe some proteins in greater abundance; or start interacting with the proteins which are not the interacting partner of normal p53 protein. Till now all the reported GOF mutations of *TP53* are pathogenic (90-93).

- Dominant negative effect (DNE) of mutations: As described earlier the p53 protein functions in a tetrameric form. Mutations which hamper the function of wild type protein by incorporating the mutant protein in the functional tetramer form are called as DNE mutations (67).

1.4.4 Polymorphisms in *TP53* and its negative regulator MDM2

TP53 is a highly polymorphic gene. *TP53* gene has 6148 polymorphisms in Homo sapiens as documented in dbSNP database. After specifying the variation class to Single Nucleotide Polymorphism (SNP), the total numbers of SNPs documented were 5311. But these variants are not filtered variants which give the exact count of SNPs whose frequency is higher than 1% in population. Referring to IARC database there are 439 validated SNPs reported in the R19 release of the IARC TP53 database whose frequency is >1% based on 1000G, gnomAD, or ESP6500. Of all the SNPs, the P72R

is the most studied SNP in *TP53* gene while a SNP in the intron 1 of MDM2 gene i.e. SNP309 is a widely studied intronic SNP. Polymorphism in the *TP53* and its negative regulator MDM2 has been shown to affect the function of the *TP53* gene. A large number of case control studies have examined the association of various *TP53* or MDM2 polymorphisms with cancer risk and its phenotype. Individual studies and their meta-analysis have shown conflicting results in most human cancer (94-101).

1.5 Model Systems for *TP53* study:

Earlier various cell lines were used as model system for studying the effect of mutant *TP53* (102). Later model organisms like mouse, Zebra fish, *Drosophila melanogaster*, or *Caenorhabditis elegans* were developed and used for studying this protein and its functions (102). Genetically the nearest model organism to humans is the mouse which is extensively used for studies. p53 participates mainly in cell cycle regulation, DNA damage repair pathway and apoptosis. All these aspects were studied using above mentioned models. Few earlier studies have shown the use of fibroblast cell lines and Lymphoblastoid Cell Lines (LCLs) established from Li Fraumeni Syndrome patients who carry the heterozygous *TP53* mutations to study cell cycle, DNA damage repair and apoptosis (103-106). However there was conflicting results related to cell cycle check points in these studies (105, 107, 108). LCLs established from LFS patients are a good model system as they can be a continuous source of DNA/RNA and can also be use for various functional studies like cytotoxic studies, DNA damage repair studies etc as described in detail in chapter 6.

1.6 Genetic testing

A medical test done to identify any change in genes, chromosomes or protein to confirm or rule out any suspected genetic condition in any individual is called genetic testing. It helps to identify if any individual is at risk of developing a genetic disease. Currently more than 1000 genetic tests are available. Various genetic tests include testing of either single gene of a short segment of DNA, alteration in which may lead to some genetic condition or chromosomal test for any chromosomal rearrangement which may lead to some genetic condition or biochemical genetic test for checking the activity and level of proteins which give an idea of change in the DNA that may result into any genetic condition (32).

1.6.1 Types of genetic testing

Genetic test can only be performed after explaining to the individual about the test during the counseling session and taking a written informed consent of the patient. Genetic tests can be performed at different stages of life and can be for different purposes. Based on the type and indication of genetic tests, they can be classified as (32):

I. **Preimplantation testing –** Preimplantation genetic test or preimplantation genetic diagnosis (PGD) is performed on in-vitro fertilization to select the embryo without the pathogenic mutation for implantation. This helps parents with high penetrance disabling or potentially lethal genetic disorders to have normal children.

II. **Prenatal testing –** Genetic testing performed during the pregnancy for the identification of any disorder in the fetus which may help parents to make a decision about the pregnancy.

III. **Newborn screening –** These tests are performed immediately after birth for diseases like phenylketonuria which leads to intellectual disability and congenital hypothyroidism if left untreated.

IV. **Diagnostic testing –** This test is performed to confirm the genetic condition by identifying pathogenic genetic alterations in the gene/s known to be associated with the condition. When phenotype of any individual reveals an underlying genetic condition, molecular genetic analysis performed to identify alterations in a gene could help in guidance and surveillance of the patient and the family.

V. **Predictive and presymptomatic testing –**In families with hereditary diseases and an identified mutation, healthy family members can be tested for the carrier status if this could help to institute preventive measures that can reduce the morbidity or mortality from that genetic condition. In hereditary cancers, carriers of pathogenic mutation can be kept on intensive surveillance for early detection or undergo preventive measures. This test can also be performed in individuals who belong to certain ethnic group in which there is a known high prevalence of germline mutations as exemplified by the Ashkenazi Jews (109).

VI. **Forensic testing –** This type of genetic testing is performed for legal purposes like paternity dispute, to identify a suspect from the samples available from the crime scene and to identify a dead when a situation warrants.

1.6.2 Genotyping assays and errors A variety of molecular genetic analysis methods are currently used in research as well as clinical laboratory settings. The nature of genotyping assay used depends on the type of genetic alterations expected, the sensitivity / specificity desired, cost and turnaround time. In clinical genetics settings,

accurate molecular genetic analysis provides very important information which is used for guiding the clinical management of the proband and reflex testing of family members. Hence these genotyping assays should have very high sensitivity and specificity. The sample used for extracting DNA for genetic analysis is peripheral blood, mouth wash or buccal swab. Extra care should be taken in laboratories that offer genetic testing at every step of sample collection, sample processing and report issuing. Various types of genotyping errors are known and are broadly classified as following:

1. Pre-analytical errors: These errors consist of wrong labeling, sample mix-up, cross contamination etc. These types of mistake should not happen in any kind of laboratory setup (110).

2. Analytical errors: These errors consist of various genotyping errors like Allele Drop Out (ADO) which may lead to false negative and false positive results (111). This is discussed in detail in chapter 4.

3. Post-analytical errors: Errors while reporting the genetic test results (110).

1.7 Rationale:

Worldwide 1221 families with LFS and germline *TP53* mutation have been reported in the August 2018 release R19 of the IARC germline *TP53* database of which only 6 are of Indian ethnicity. The IARC database shows differences in the *TP53* mutation spectrum and the tumor spectrum in different parts of the world. As there is paucity of data from India there is a need to establish *TP53* mutation spectrum and its clinical correlation in large and well characterized Indian LFS families. Comprehensive genetic analysis in populations not previously evaluated often results in identification

of several novel and uncharacterized variants. This requires development of model systems such as LCLs from patients with novel variants and from healthy individuals from similar genetic background as controls. Moreover, the role of low penetrance SNPs in *MDM2* and *TP53* P72R with clinical phenotype and response to treatment have been studied in relatively small Indian cohorts consisting of diverse cancers.

1.8 Hypothesis: We hypothesize that the genotype – phenotype correlation in the Indian LFS families may be distinct from the well characterized Caucasian population (88, 94, 112-114) and this knowledge can help in understanding the molecular pathways in the origin of diverse LFS tumors and devising population specific counseling and risk management.

1.9 AIM: To identify germline *TP53* gene mutation spectrum and its correlation with the tumor spectrum and age at diagnosis in a large Indian cohort of hereditary LFS and sporadic LFS associated cancers.

1.10 BJECTIVES

1. Identifying LFS / LFL families and their detailed phenotypic characterization
2. Comprehensive Genetic Analysis of *TP53* & MDM2 in LFS / LFL & sporadic sarcomas & establishing Genotype-Phenotype correlation.
    A. Identification of germline point mutations and Large Genomic Rearrangements in *TP53* gene in LFS/LFL/suspected LFL families
    B. *TP53* & *MDM2* polymorphisms and clinical phenotype in LFS / sporadic sarcoma
3. Establishing Lymphoblastoid cell lines from germline *TP53* mutation carriers for DNA repair studies & variant characterization.

Chapter 2

Materials and Methods

2.1 Patient cohort

Patients with LFS/LFL/ suspected LFL or sporadic LFS associated cancer were enrolled in this study and subjected to genetic analysis and genotype phenotype correlation as described.

All the patients were enrolled in between 2003-2018 as part of studies approved by the Tata Memorial Centre-ACTREC Institutional Review Board. Written informed consent was obtained from all subjects for biobanking and genetic analysis. For minors, the written informed consent was provided by the parents. All experiments were carried out in accordance with the approved guidelines and regulations. Germline *TP53* mutation analysis was performed in two cohorts. The first cohort was cancer patients with personal or family history suggestive of hereditary LFS or LFL syndrome or suspected to be LFL (SLFL) and enrolled in the Cancer Genetics Clinic (CGC), Tata Memorial Centre (TMC). Families enrolled under CGC were 300 for the study. The second cohort consisted of sarcoma patients with or without family history of cancer who were enrolled in an ongoing TMC International Sarcoma Kindred Study (TISKS). Cases enrolled under TISKS were 311 sarcoma cases with or without family history. All 311 sarcoma cases were taken for association study of clinical outcome with *TP53* and MDM2 polymorphisms. Germline *TP53* mutation analysis was performed in 500 families from these two cohorts. Of these, 197 families fulfilled the defined criteria of LFS or LFL, while the remaining 303 families did not fulfill the criteria for LFS or LFL as listed below.

The detailed inclusion criteria for the screening of *TP53* gene is geven below :

Group 1: Classical Li-Fraumeni Syndrome (LFS) cases: Using criteria originally proposed by Li and Fraumeni and described in the earlier version of the thesis (table 3.1, page 105).

Group 2: Li Fraumeni Like (LFL) cases: Using LFL criteria proposed by Birch 1994, Eeles 2001 or Chompret 2001, 2009 and 2015 and described in the earlier version of the thesis (table 3.1, page 105).

Group 3: Suspected Li Fraumeni Like (sLFL) cases: Individuals who do not fulfill the classical LFS or the existing LFL Birch, Eeles or Chompret criteria.

3A: Proband with early onset (<50 yrs) *Modified LFS spectrum cancer with 1st, 2nd or 3rd degree relative with any cancer at any age OR proband with multiple primary cancers of which at least one is a modified LFS spectrum cancer <50 years.

3B: Proband with early onset cancer (<50 years) other than modified LFS spectrum cancer and family history of *Modified LFS spectrum cancer at any age.

3C: Sporadic sarcoma except Emryonal RMS at any age.

Group 4: Probands with familial cancer at any age and do not fulfill criteria for LFS, LFL, suspected LFL (3A, 3B or 3C) or any other hereditary cancer syndrome.

*Modified LFS spectrum cancers: Includes *Haemato-Lymphoid malignancies and *Malignant Phyllodes in addition to the previously described LFS spectrum cancers - ACC, Sarcoma, CNS, Leukemia and Breast cancer

2.2 Patient Recruitment and Genetic Testing

2.2.1 Pre and pot-test counseling: A counseling session prior to the testing was held between the patients and their family members who accompanied them. During this session, a trained genetic counselor explained to them about the genetic aspects of the disease, what is genetic testing; expected outcomes of genetic analysis and its possible implications for the proband and the family. 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 LFS/LFL was made. Peripheral blood (3-6 ml) in Ethylene Diamine Tetra Acetic acid (EDTA) vacutainers or 50 ml of mouthwash samples in Normal Saline were collected from the patient after obtaining written informed consent. Germline *TP53* genetic testing was performed using methods described later. There were 14 mouth wash samples from 14 individuals from 12 families. For 9 of these cases, blood sample was also available. Of these 14 individuals, TP53 mutation was identified in 5 individuals. The mutation was confirmed in 4/5 cases in blood sample also but for one case blood sample was not available for confirmation of the mutation identified in mouth wash sample. In the post test counseling session, the genetic test results were explained to the patients and printed report was given. Based on the report, mutation carriers were advised appropriate surveillance and prophylactic measures were advised and importance of the screening of the family specific mutation in first and second degree relatives of mutation carriers was explained.

2.2.2 Comprehensive Germline *TP53* Genetic Analysis: Genomic DNA was extracted from peripheral blood lymphocytes or from mouth wash by Qiagen columns

(QIAamp DNA Blood Mini Kit; Catalogue number 51106) according to manufacturer's protocol or by conventional phenol chloroform method for some cases. All the exons of the *TP53* gene (1-11) were PCR amplified and sequenced by Sanger sequencing. If no germline *TP53* mutation was identified, analysis for Large Genomic Rearrangements (LGRs) was done by the Multiplex Ligation dependent Probe Amplification (MLPA) kit (MRC Holland) as per manufacturer's protocol. In three LFS/LFL cases without an identified *TP53* mutation on Sanger Sequencing and MLPA, germline exome sequencing was carried out.

2.3 Genomic DNA extraction

In the earlier part of the study the preferred DNA extraction method was the manual phenol chloroform method but since 2015, the DNA was extracted in almost all cases using kit method.

2.3.1 DNA extraction by kit method

DNA was extracted from 200µl of blood or 50 ml of Mouthwash samples using the QiaAmp Blood DNA Mini kit as per the manufacturer's protocol with some changes in incubation time for better yield. The step wise procedure was as follows: For mouth wash the cells from 50ml was pellet down and resuspended on 200 µl PBS and the same steps followed from step 3.

1. Microcentrifuge tube (1.5ml) taken and 20 µl QIAGEN Protease (or proteinase K) added into the bottom of the tube.

2. 200 µl blood sample to was added to the microcentrifuge tube.

3. 200 µl Buffer lysis buffer (AL) was added to the sample and mixed by pulse-vortexing for 10s.

4. Incubated at 56°C for 10-30 min.

5. The microcentrifuge tube briefly centrifuged to remove drops from the lid.

6. 200 µl of absolute ethanol was added to the sample, and mixed again by pulse-vortexing for 15 s. After mixing, the tubes were briefly centrifuged to remove drops from the lid.

7. The mixture from step 6 was carefully applied to the QIAamp Mini spin column (in a 2 ml collection tube) without wetting the rim. The cap closed, and centrifuged at 6000 x g (8000 rpm) for 1 min. The QIAamp Mini spin column placced in a clean 2 ml collection tube (provided), and the tube containing the filtrate was discarded.

8. The QIAamp Mini spin column was carefully opened and 500 µl of AW1 Buffer was added without wetting the rim. The tubes centrifuged at 6000 x g (8000 rpm) for 1 min. The QIAamp Mini spin column placed in a clean 2 ml collection tube (provided), and the collection tube discarded containing the filtrate.

9. The QIAamp Mini spin column carefully opened and 500 µl of AW2 Buffer was added without wetting the rim. Cap of the tube was closed and centrifuged at full speed (20,000 x g; 14,000 rpm) for 3 min.

10. It is recommended to give a dry spin by placing the QIAamp Mini spin column in a new 2 ml collection tube (not provided) after discarding the old collection tube with the filtrate. Centrifuged at full speed for 1 min.

11. The QIAamp Mini spin column transferred in a clean 1.5 ml microcentrifuge tube (not provided), and the collection tube containing the filtrate was discarded. The QIAamp Mini spin column was opened carefully and 200 µl of elution Buffer (AE) or distilled water was added. Incubated at room temperature (15–25°C) for 1 min (can be kept for longer time also for better yield), and then centrifuged at 6000 x g (8000 rpm) for 1 min.

12. DNA was quantified using Nanodrop spectrophotometer and by loading the DNA samples on 0.8% 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.

2.3.2 Extraction by conventional Phenol-Chloroform Method

Lymphocytes were separated from whole blood by lysing the Red Blood Cells (RBCs) using a buffer containing ammonium chloride, which has a minimal lysing effect on lymphocytes. The separated lymphocytes were washed with Phosphate Buffer Saline (PBS) to remove any traces of the salts before further processing. The composition of these reagents was as follows:

RBC Lysis Buffer:

1. Ammonium Chloride - 82.9 g

2. Ammonium Bicarbonate – 7.89g

3. EDTA – 0.37g

Dissolved in 1 litre of distilled water for 10X concentration and can be stored at room temperature. For use Diluted to 1 X concentration with distilled water.

Phosphate Buffer Saline : (PBS)

1. NaCL : 8 g

2. KCL : 0.2g

3. Na2 HPO4 : 1.44g

4. KH2PO4: 0.24g

First Dissolved in 800ml of distilled water and pH was adjusted to 7.4. then volume was make upto 1 litre with distilled water and autoclaved at 15 lb for 15 mins. Stored at room temperature.

Protocol :

1) Fresh blood collected in 3 ml EDTA vacutainer. Kept at room temperature until the plasma is separated.

2) The blood was transferred to 15ml falcon tubes.

3) Three volumes of RBC lysis buffer was added and mixed by rotating for 10 mins.

4) Tubes were centrifuged at 1500rpm for 15 mins.

5) Supernatant was discarded leaving about 1ml behind to prevent loss of cells.

6) Again three volumes of RBC lysis buffer was added and steps 3 to 5 were repeated until a clear supernatant and a clean white pellet is obtained (usually in 3 washes)

7) After final centrifuge, the supernatant was discarded completely and the pellet was re-suspended in 1 ml PBS.

8) The solution was transferred to two 1.5 ml fresh Eppendorf, distributing half the volume in each.

9) The tubes were centrifuged at 12,000rpm for 10 mins.

10) The supernatant was discarded and the pellets were re-suspended in 500µl of fresh PBS.

11) The tubes were centrifuged at 12000 rpm for 10mins and the supernatant was discarded completely.

12) Cell lysis buffer (400ul) was added and mixed gently by rotating.

13) The tubes were incubated at 55-56° C in a water bath for an hour.

14) 10 µl of Proteinase K was mixed (10mg/ml stock) gently into the viscous solution making sure no vortexing was done. .

15) The tubes were incubated overnight at 55-56 °C for proper lysis.

16) Equal volume of phenol (equilibrated with TRIS to get a pH above 7) was added to the tube and mixed well by inverting/rotating for ten mins.

17) Centrifuged at 12000 g for 10 mins.

18) The aqueous upper layer was transferred to a fresh tube containing half the volumes of phenol and chloroform: Iso-amyl alcohol (24:1), when transferring the aqueous phase, The aqueous layer was drawn into the pipette very slowly to avoid disturbing the material at the interface.

19) Mixed well by inverting/rotating, for 10 mins and then centrifuge for 10 mins at 12,000 g.

20) The supernatant was transferred to a fresh tube and equal volume of chloroform: iso-amyl alcohol (24:1 was added.

21) Mixed well by inverting, for 10 mins and then centrifuged at 12,000g for 10 mins.

22) The supernatant was transferred to a fresh tube.

23) Twice the volume of absolute alcohol was added and inverted gently a few times. The DNA appeared as thread like structures.

24) The tubes were incubated at -20° C for approximately 4 hours to overnight to allow for full precipitation of the DNA.

25) Two washes with 70% alcohol were given and the pellet was air dried.

The DNA was dissolved in appropriate amount of TE buffer and stored at 4° C.

2.4 Polymerase Chain Reaction (PCR)

PCR amplification of the entire coding region of *TP53* gene and flanking intronic region was carried out using specific primers designed for respective amplicons. PCR amplification was carried out in a 25µl reaction with 10pmol of primers (Sigma), 100ng of DNA and using 1 unit of Taq polymerase (Finnzyme). The composition of master mix is given in Table 2.1 and the cycling conditions are summarized in Table

2.2. The primer sequences and annealing temperatures for various amplicons of *TP53* gene is given in Table 2.3.

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)	0.5µl
10 pmoles/µl P (R)	0.5µl
20 ng/µl DNA template	5.0µl
5U/µl Taq polymerase	0.5 µl
MilliQ	15.0 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: **35 cycles**		

* annealing temperature variable for various primer pairs

Table 2.3 Primer sequences and annealing temperatures for *TP53* gene		
Amplicon	Primer	Annealing Temp
Exon 1F	CACAGCTCTGGCTTGCAGA	63.2°C
Exon 1R	AGCGATTTTCCCGAGCTGA	
Exon 2F	AGCTGTCTCAGACACTGGCA	63.2°C
Exon 2R	GAGCAGAAAGTCAGTCCCATG	
Exon 3+4-P1-F	AGACCTATGGAAACTGTGAGTGGA	58-51Touch Down
Exon 3+4-P1-R	GAAGCCTAAGGGTGAAGAGGA	
Exon 3+4-P2-F$^{\pm}$	AGACCTATGGAAACTGTGAGTGGA	68°C
Exon 3+4-P2-R$^{\pm}$	AGGAAGCCAAAGGGTGAAGAGG	
Exon 5+6F	CGCTAGTGGGTTGCAGGA	63.2°C
Exon 5+6R	CACTGACAACCACCCTTAAC	
Exon 7-P1-F	CTGCTTGCCACAGGTCTC	63.2°C
Exon 7-P1-R	TGGATGGGTAGTAGTATGGAAG	
Exon 7-P2-F$^{\pm}$	AGAATGGCGTGAACCTGGGC	66°C
Exon 7-P2-R$^{\pm}$	TCCATCTACTCCCAACCACC	
Exon 8+9F	GTTGGGAGTAGATGGAGCCT	63.2°C
Exon 8+9R	GGCATTTTGAGTGTTAGACTG	
Exon 10F	CTCAGGTACTGTGTATATACTTAC	57.8°C
Exon 10R$^{\pm}$	ATACACTGAGGCAAGAAT	
Exon 11F	TCCCGTTGTCCCAGCCTT	57.8°C
Exon 11R	TAACCCTTAACTGCAAGAACAT	
3'UTR_F$^{\pm}$	GCCAGCCAACTTTTGCAT G	60°C

3'UTR_R±	GTCTCCAGCCTTTGTTCCCC	

Table 1: ± represents newly designed primers. Rest of the primers were taken from reference (115)

2.5 Agarose Gel Electrophoresis

1. One gram of agarose was weighed to make a 1 % solution and was dissolved in 100 ml 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 (EtBr from 10mg/ml stock) was added to a final concentration of 0.5µg/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 Ultra Violet (UV) Transilluminator.
9. Pictures were taken and stored using Gel Documentation System (UVP products). A representative picture of gel with all the amplicons of *TP53* is shown in figure 2.1.

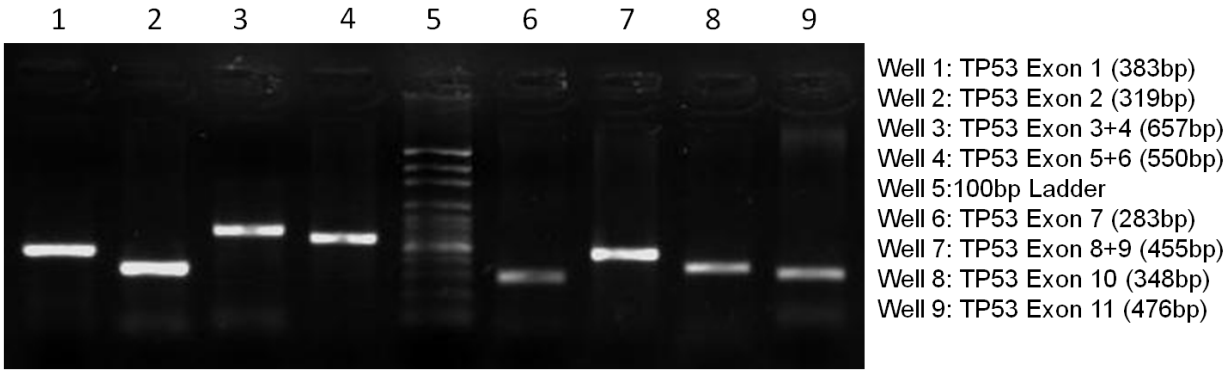


Figure 2.1: Amplicons of *TP53* gene: 1% agarose gel showing standardized PCR products of all the amplicons of *TP53* gene.

2.6 Purification of PCR products

The PCR products were diluted to a concentration of 30-50ng of template. 1.5 microliter of EXO-SAP IT [USB Affimetrix] was added to 7µl of diluted PCR product. This mix was incubated at 37ºC for 15 mins followed by incubation at 80ºC for 15 mins.

2.7 Sanger sequencing

Amplified products were cleaned with Exonuclease and Shrimp Alkaline phosphatase for sequencing to remove the unincorporated primers and dNTPs. 2ul of cleaned PCR products with 1ul of 1.5 pmol of primer was used for cycle sequencing reaction. Post cycle sequencing, products were sequenced with Big Dye Terminator kit version 2 (Applied Biosystems) on DNA sequencers 3500 Genetic Analyzer 8 capillary or 3730 DNA Analyzer 48 capillary (Applied Biosystems)

Chromatograms were analyzed using Chromas Lite and Sequencing Analysis Software v.5.3.1. The threshold for mixed base for detection of heterozygous mutations was set at a ratio of 0.3 of the wild type.

(https://tools.thermofisher.com/content/sfs/brochures/seq-quantification-app-note.pdf, accessed on 15th Mar 2018)

Step 1: Cycle sequencing

- 2µl of template and 1µl of primer (both in the desired concentration) were added in the 96-well sequencing plate.
- Cycle sequencing reaction master mix was prepared as:
  - MilliQ water - 4.75 µl
  - 5X Sequencing Buffer - 1.75 µl
  - Ready Reaction mix - 0.50 µl
- 7µl of mastermix was added to each sample
- Cycle sequencing reaction in the thermal cycler was set 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. 4ºC forever

Step 2: Post cycle sequencing cleanup

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

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

2.8 Mutation analysis

The chromatograms were analysed by using Chromas Lite software [Technelysium Pty Ltd]. The data was compared with the reference sequence of *TP53* gene taken from the NCBI (genomic DNA sequence) and University of California, Santa Cruz (UCSC) (c.DNA sequences) databases to identify variants. Locus specific database for *TP53* gene – IARC database (http://p53.iarc.fr/) was used to check if the variants were reported or not. All pathogenic or likely pathogenic germline mutations were confirmed by bi-directional sequencing. Also, the mutations were confirmed on a second independent sample, whenever available. The pathogenicity of the mutations was inferred based on their previous characterization and description in the most recent IARC germline database version R19. Novel variants not previously reported were characterized using multiple software tools for in silico prediction of pathogenicity. These included Align-Grantham Variation Grantham Deviation (AGVGD), Polyphen, Sorting Intolerant From Tolerant (SIFT) and Rare Exome Variant Ensemble Learner (REVEL).

2.9 Multiplex Ligation Dependent Probe Amplification (MLPA)

In patients without any mutation identified by Sanger sequencing, MLPA analysis was carried out for the large genomic rearrangements study. SALSA MLPA P056

TP53 Probemix kit from MRC Holland was used as per manufacturer's protocol described below.

2.9.1 MLPA for Large Genomic Rearrangement (LGR) analysis

Procedure

Step 1: Denaturation of DNA sample

1. 5 µl of DNA sample added to each tube.
2. The tubes were placed in a thermocycler and the MLPA thermocycler program was started (given at the end of protocol).
3. DNA denatured for 5 minutes at 98°C and cooled to 25°C.

Step 2: Hybridization of Probes to sample DNA

1. Hybridization master mix was prepared for each reaction: 1.5µl MLPA buffer + 1.5µl probemix. Hybridization master mix was well mixed by pipetting or vortexing.
2. After DNA denaturation, 3µl of the hybridization master mix was added to each sample tube. Mixed well by pipetting up and down.
3. thermocycler program continued: incubation for 1 minute at 95°C, then for 16 – 20 hours at 60°C.

Step 3: Ligation of hybridized probes

1. Ligase-65 master mix was prepared by adding 25µl H_2O + 3µl Ligase buffer A + 3µl Ligase buffer B for each reaction. Then 1µl Ligase-65 enzyme was added. Mixed well by pipetting gently up and down. (Never vortex enzyme solutions).
2. The thermocycler program was paused at 54°C.
3. When the samples are at 54°C, 32 µl of the ligase master mix was added to each reaction tube. Gently mixed by pipetting up and down.

4. The thermocycler program was resumed: 15 minutes incubation at 54° (for ligation), followed by 5minutes at 98°C for heat inactivation of the Ligase-65 enzyme and then at 20°C for 10 minute.

Step 4: PCR amplification of Ligated Probes

1. A polymerase master mix was prepared by adding 7.5µl dH_2O + 2 µl SALSA PCR primer mix + 0.5 µl SALSA Polymerase for each reaction. Mixed well by pipetting up and down; vortexing should be avoided. Stored in ice until use.
2. **At room temperature**, 10 µl polymerase mix was added to each tube. Mixed by pipetting up and down. Thermocycler program of 35 cycles: 30 seconds at 95°C; 30 seconds at 60°C; 60 seconds at 72°C. Program ended with 20 minutes incubation at 72°C and then paused at 15°C.

Step 5: Fragment Separation by Capillary Electrophoresis

1. PCR products (1 µl), 0.15µl LIZ size standard and 10µl of Formamide were mixed.
2. The samples denatured at 95°C for 5minutes and then placed on ice rapidly to avoid renaturation.
3. Plate was loaded on the sequencer.

Table 2.4 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

2.9.2 MLPA Data Analysis

MLPA data was analysed for large genomic rearrangements 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 (table 2.4):

- *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.

2.10 Whole Exome Sequencing: Germline exome sequencing was done for three classical LFS cases in whome Sanger sequencing and MLPA had not identified any *TP53* mutation. The Nextera rapid capture kit (Illumina) was used for library preparation and sequencing was done on Hiseq 2000 following manufacturer's protocol. The number of targated exons covered in this kit was 214,405 and 98.3% of Refseq genome was covered.
(https://www.illumina.com/documents/products/datasheets/datasheet_nextera_rapid_capture_exome.pdf)

2.10.1 Steps of Exome Library preparation

1. Tagmentation of genomic DNA was done using Nextera transposomes. Tagmentation is a process which fragments gDNA & adds adapter sequence to their ends in a single step.

2. Index adapters were ligated to the ends of tagmented DNA & then amplified. Index adapters provide unique identity to each sample.

3. After amplification of tagmented DNA library, its cleanup was done to remove unwanted products which would otherwise interfere with downstream process.

4. The next step was pooling of all the DNA libraries at equal concentration into a single pool for exome enrichment using target specific probes. For this, libraries needs to be quantified accurately using fluorometric method such as QUBIT. Inaccurate quantification may lead to overrepresentation of some samples compared to others.

5. Each library was pooled at a concentration of 500 ng. Once the pooling was done, the probes were added to the library to bind to target region.

6. Next step was to capture hybridized probes using streptavidin magnetic beads & non specific binding were washed from the beads which gives enriched library. This process of probe hybridization and capture was done twice to ensure high specificity of captured regions in enriched libraries and removal of nonspecific bindings.

7. Further clean up of enriched library wass done using sample purification magnetic beads & then the purified & enriched library was amplified & purified to generate final Exome library to be used for the subsequent steps of cluster generation and sequencing.

8. Sample Library was taken for Quality control analysis using Agilent Technologies 2100 Bioanalyzer as well as 1% Agarose gel. The DNA library should have fragments ranging in size between 200 bp to 1 kbp.

2.10.2 Exome Data Analysis Pipeline

Data analysis was done using Genome Analysis Tool kit (GATK) standardized pipeline. The steps of data analysis included**:**

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

2. **Alignment with reference genome:** For analyzing sequence data, alignment to reference genome is necessary. Processed FastQ files were aligned to human reference genome sequence, extracted from UCSC. Alignment was done using Burrows-Wheeler Aligner (BWA) software/Mapping tool. After alignment .sai files were generated and converted into readable text files i.e. Sequence Alignment Map (SAM) files. Large SAM files are then compressed into Binary Alignment Map (BAM) files (Binary version of SAM files) using SAMtools command.

3. **Post-Alignment processing:** After alignment with reference genome, post alignment processing of BAM files was done which includes marking of PCR duplicates using PICARD tools, local realignment and quality score recalibration using GATK tools.

4. **SNP Calling:** This step produces raw SNP files from processed aligned reads. SNP calling was done using GATK Unified Genotyper which Calls SNPs & short InDels at the same time & gives well annotated Variant Call Format (VCF) file as an output.

Various filters were applied at this step based on quality and depth of coverage of alleles to filter out variants with low allele frequency.

5. **Variant Annotation and visualization:** VCF files were annotated using ANNOVAR tool with different database & an excel sheet was generated as an output summary which includes gene name, its chromosome location, function, rs ID, pathogenicity prediction score using different in silico prediction tool etc. Variants were further visualized using Integrated Genome Viewer (IGV) software.

Exome analysis was also performed by using Varscan where filtering was done on the basis of allele frequency. Only those variants were called whose allele frequency was greater than 20% as we have to look only for germline variants.

2.11 Germline Allele Drop Out (ADO) in *TP53* gene

A germline *TP53* ADO was suspected, if the *TP53* variant identified through exome sequencing was either not detected on Sanger sequencing or was detected with peak height ratio of <0.3 on the chromatogram. In addition, whenever a pathogenic germline mutation was detected in a homozygous state, we suspected a germline *TP53* ADO. Sanger sequencing was repeated for all suspected ADO samples in the discovery cohort and in a validation cohort of 150 LFS/LFL cases negative for *TP53* mutation on initial Sanger sequencing and MLPA. For exon 7 ADO, in which a polymorphism was identified in the annealing region of the primers initially used, sequencing was repeated with redesigned primers. Count of all *TP53* polymorphisms was searched in the National Centre for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/snp, accessed on 5th Apr 2018). Presence of polymorphisms in the binding region of the primers used in our study, was searched

in the UCSC genome browser (https://genome.ucsc.edu/, accessed on 14th Mar 2018) Figure 2.2. For each polymorphism identified, its minor allele frequency (MAF) was searched in the dbSNP database by their rs ids (https://genome.ucsc.edu/, accessed on 14th Mar 2018)

G-quadruplexes: For estimating G-quadruplexes in the *TP53* sequence, the Quadruplex forming G-Rich Sequences (QGRS) Mapper was used (116). Methylation status of *TP53* gene was checked in an available online tool named sequence manipulation suite (117).

Figure 2.2: UCSC genome browser showing primer annealing region polymorphisms: Sequence within the red box is the primer sequence and below black highlighted regions mark the polymorphisms with their rs-IDs.

2.12 Genotyping for *TP53* and MDM2 polymorphisms

Genotyping was done for 3 polymorphisms. The two *TP53* polymorphisms - P72R (rs1042522) and the 16bp duplication in intron 3 (PIN3-rs17878362) were genotyped during full gene sequencing of *TP53*. The MDM2 SNP309 was genotyped using

Restriction Fragment Length Polymorphism (RFLP). The PCR primers and temperature for MDM2 SNP309 is shown in table 2.5

Table 2.5 Primer sequence of MDM2SNP309 amplicon		
Primer	**Sequence**	**Annealing Temp**
MDM2SNP309_F	GTTTTGTTGGACTGGGGCTA	63.2°C
MDM2SNP309_R	CGGAACGTGTCTGAACTTGA	

- MspA1 restriction enzyme from New England Biolabs (NEB) was used for RFLP
- The amplicon has two restriction sites. One is created due to change of allele from T>G.
- 293bp digested into 193bp and 100bp in TT allele and 147, 100 and 46bp in GG allele as shown in figure 2.3.

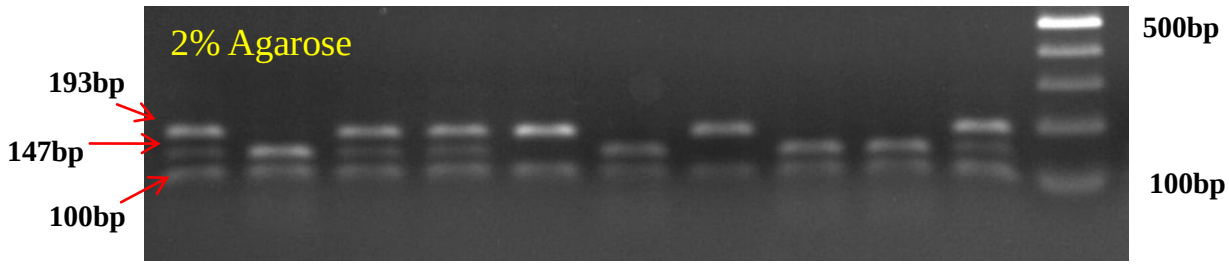


Figure 2.3: RFLP of MDM3 SNP309: 2% agarose gel electropherogram showing the heterozygous homozygous and wild type genotypes of MDM2 SNP 309; (bp-base pair).

PROCEDURE:

1. The required volume of PCR product was taken and diluted with MilliQ 20-30 ng/ul.

2. 10µl of diluted PCR product was taken and 2ul of Buffer was added (provided with the restriction enyme from NEB and the reaction volume was made up to 20ul by MilliQ.

3. Restriction Enzyme (MspIA1) was added 0.1U/reaction, and incubated at 37 ^{0} C for 30mins to 1 hour. (Note: Incubation time and temperature varies with Restriction Enzyme used).

4. The digested products were visualized on 2% agarose gel electrophoresis followed by confirmation of 5% genotyping results on Sanger Sequencing.

2.13 Establishment of Epstein Barr Virus transformed Lymphoblastoid Cell Lines:

Epstein Barr Virus (EBV) transformed LCLs were established 3 ml freshly collected peripheral blood from cancer patients with or without germline *TP53* mutations and from healthy controls with the following protocol.

Reagents

1. Media

2. Normal Saline

3. Ficol Hypaque (FH)

4. EBV soup

Plasticware: 15ml falcon tubes, Pasture pipettes, culture flasks

Medium Preparation

Incomplete media: 1 packet of Iscove's Modified Dulbecco's Medium (IMDM) from Gibco dissolved in 1lt of fresh MQ + sodium bicarbonate as mentioned in the packet (0.3% of $NaHCO_3$). It is then filtered using Millipore Express Plus Filters of 0.22um. Incomplete media kept for 48 hours to check for any contamination. This incomplete media is used for preparation of complete media. The composition of complete media is given in table 2.6 below.

Table 2.6 Composition of complete medium

Components	**Volume**
Incomplete media	83.5 ml
Fetal Bovine Serum (FBS)	15.0ml
Glutamine	1.0ml
penicillin, streptomycin, gentamycin and amphotericin B (PSGA) Antibiotic	0.5ml
Total	100.0ml

Freezing Media: Used while freezing the established cell line.

FBS: Dimethyl Sulphoxide (DMSO) was used in a ratio of 9:1 (900ul+100ul) as freezing media

Normal Saline (NS) preparation

9 gm of NaCl was dissolved in 1litre of water, autoclaved and stored at room temperature.

Ficol-Hypaque:

Composition and preparation

Ficol- 21.6gm of Ficol powder was dissolved in 80ml of Milli-Q.

Hypaque (Sodium Ditrizoate) 33.33 gm was dissolved in 200ml of Milli-Q. Both the solutions mixed together after which the density of the solution was adjusted to 1.077. This solution is stored at 4 ^{0}C until use.

Epstein Barr Virus Supernatant (Sup) Preparation:

- Initially the cell line B95.8 Monkey Marmoset was cultured in a minimal volume of 5 mL medium in a 25 mL flask. On reaching the required density of viral particles the culture was transferred to a 100 ml flask for further growth and enrichment of viral particles.

Procedure:

- Three tubes were labeled, 3ml blood taken in one tube and volume made upto 10 ml by adding NS, FH (2.5ml) taken in another tube. The diluted blood was slowly added over FH, after which it was centrifuged at 2000rpm for 20 min.
- The peripheral blood mononuclear cells (PBMNCs) separate out at an interphase as a white ring, which is collected in 10 ml saline (figure 2.4). The PBMNCs were washed twice with NS.
- The supernatant was discarded completely, then 1 ml of EBV sup and 1 ml complete media was added on the pellet and incubated in a CO_2 incubator in a 25ml T-flask at 37ºC.

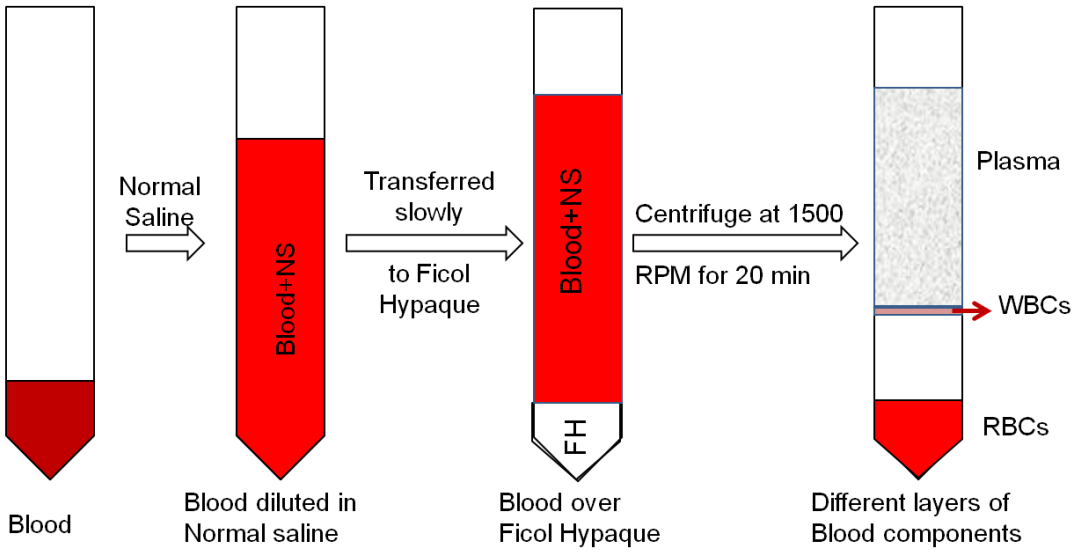


Figure 2.4: WBC separation by Ficol Hypaque density gradient centrifugation: FH-Ficol Hypaque; NS-Normal Saline; WBC-White Blood Cells; RBC-Red Blood Cells.

Freezing of Lymphoblastoid cell lines

Freezing of cell lines: Approximately half of the medium in the flask of the cell culture was removed, discarded and replaced with fresh medium 24 hours prior to freezing of the cell line. The cell suspension was gently mixed and centrifuged at 1500rpm for 10mins. The cell pellet was re-suspended in precooled freezing medium (90% FBS+ 10% DMSO) while keeping the cell density at around 1 million cells/ml. The cell suspension was then transferred to cryotubes and immediately frozen at -80°C. Cryo vials containing the cells were transferred to liquid nitrogen after approximately 24-48 hours.

2.14 DNA damage repair study- Gamma H2AX assay

EBV transformed lymphoblastoid cell lines were exposed to gamma radiation and the Gamma H2AX foci were estimated at baseline and at different time points after irradiation to estimate the DNA damage repair. The following protocol was used.

Requirements

Plastic wares and Others: Culture plates, 25ml T-Flask, 15ml falcon tubes, Radiation Source,

Reagents:

1. Incomplete and Complete media- Prepared as described above
2. NS and PBS- composition and preparation described above. All the components dissolved in 800 ml of milli-Q and pH is adjusted to 7.4 and the final volume made upto 1litre.
3. Para Formaldehyde (PFA): For fixation 3% PFA was made. 3gm of PFA dissolved in 100ml of PBS.
4. FACS Buffer for processing cell line: Buffer is prepared by adding 1ml FBS in 99ml of PBS and 0.02 gm of Sodium Azide.
5. Triton-X-100: For perforation of cell 0.3% of Triton-X-100 was used. 300 ul of Triton-X-100 was added in the 100 ml of buffer
6. Tween-20: Tween-20 was used at a concentration of 0.1% for antibody dilution. 100 ul of tween 20 added in 200ml of buffer for use.
7. Bovine Serum Albumin (BSA): 5% BSA prepared in PBS
8. Primary and Secondary Antibody: Anti-phospho-Histone H2A.X (Ser139), clone JBW301 primary antibody was used from Merck Millipore (cat # 05-636, Lot # 2250524). Secondary antibody used was from invitrogen, by Life technologies.

Catalogue # A11029 lot 1252783, 0.5ml, Alexa Fluor 488 goat anti mouse IgG (H+L).

9. 4′,6-diamidino-2-phenylindole (DAPI)- 0.5 ug/ml concentration of DAPI was used for counter staining.

Revival of Lymphoblastoid cell line

1. Cryovials from liquid nitrogen cylinder were taken out and immediately kept in water bath previously set at 37ºC
2. Culture from the cryovials was transferred to 15ml falcon with 10ml of incomplete media for wash. Centrifuged at 1500 RPM for 10 mins. Supernatant discarded.
3. After wash cultures were transferred to 25ml T-flask with 2ml of complete medium

After successful revival the cells were split in two parts, one for freezing and the other for experiment. Experimenting cells were again split into two parts and after reaching the count of 10 million in each flask, cells were taken for irradiation.

Irradiation of cell lines

1. After achieving the 10 million count cells were pelleted down, old media discarded and fresh complete media was added.
2. Cells were distributed into four culture plates from each flask with 2.5 million counts in each plate for four different time points (1hr, 4hr, 24hr and 48hr). One

set of plates was used for irradiation and the other set for no radiation control. All steps were done simultaneously for LCLs from healthy controls.

3. Irradiation set were taken for both cases and controls. Source of radiation used was ^{60}Co and cells were exposed to 2Gy of radiation with a dose rate of 1.9 Gy/min at 80 cm Source Skin Distance (SSD) on Bhabhatron Telecobalt unit.

4. After irradiation cells were kept at 37ºC in CO_2 incubator and fixed at different time points.

Post irradiation processing

1. Cells were taken at appropriate time for processing and fixing.

2. Cells from culture plates were transferred to 1.5ml eppendorf tube and centrifuged at 1500 rpm for 10 mins. Supernatant discarded and 1ml of FACS buffer was added and pellet dislodged and pulse vortexed. Centrifuged at 1500 rpm for 10 mins and supernatant was discarded.

3. 1ml of 3% PFA was added for fixation at appropriate time. Incubated for 15 mins at 4ºC. Centrifuged after incubation at 1500 RPM for 10 mins and supernatant discarded.

4. These steps were repeated at every time point. Cells were taken for staining at the same time or kept at 4ºC and taken for staining after fixing the cells from all time points.

Post fixation processing

1. Fixed cells were taken and a wash given by adding 1ml FACS buffer. Mixed and centrifuged at 1500 rpm for 10mins and supernatant discarded.

2. Tap mixed and resuspended, then 500 ul of triton-x-100 added. Mixed by vortexing for 10 sec and then 10 min incubation at room temperature.

3. Mixed well after incubation and centrifuged at 1500 rpm for 10 min and the supernatant discarded carefully.

4. Tap mixed and 700 ul of FACS buffer was added. Again tap mixed and then centrifuged at 1500 rpm for 10 mins.

5. Supernatant discarded. Mixed well. 50 ul of 1:200 diluted (1:200-Antibody-diluted in 5% BSA in Tween 20) primary antibody was added.

6. Kept for overnight incubation in 4°C.

7. After overnight incubation, 700 µl of FACS buffer was added (mix well). Centrifuged at 1500 rpm for 10 min, supernatant discarded.

8. Secondary antibody added in 1:200 dilution (diluted in 5% BSA in tween 20) in dark room. Mixed and incubated at room temperature in dark for 1-2 hr.

9. FACS buffer (700ul) was added, mixed well and centrifuged at 1500 rpm for 10 min. Supernatant discarded and tap mixed (maintain darkness)

10. Pre cut coverslips were cleaned with methanol and properly wiped with tissue paper.

11. Sample volume of 20 ul were transferred from each tube to cover slip. Properly spread and kept for air drying.

12. After air drying DAPI diluted to working concentration (0.5ug/ml) from stock solution. DAPI (15ul) put on the parafilm. Coverslip was put over the drop by keeping cell smear downwards. Kept for 10-15 mins.

13. PBS washes given four times by keeping the coverslip in 2ml PBS in stericups and by shaking it (cell smear upside). Pipette out the PBS.

14. Then a methanol wiped slide is taken and a drop of mounting media (vecta shield) were put on the slide. Coverslip from stericups kept over the slide keeping cell smear downward over the mounting media drop.

15. Extra mounting media removed by putting tissue paper over the cover slip and by rubbing it gently.

16. Coverslips sealed using transparent nail polish. Slide ready for imaging

Image interpretation and analysis

1. Images were captured in confocal microscope at 63X of at least 5-10 fields of each coverslip covering around 30-50 nuclei for each time point.

2. Data was analysed using Zies software. Fluorescent intensity of each nuclei was recorded and Mean Fluorescent Intensity was calculated for at least 30-50 nuclei.

3. A higher mean fluorescent intensity indicated higher number of cells with unrepaired double stranded DNA damage.

2.15 Use of Lymphoblastoid cell lines for variant characterization

In-silico prediction of variant was performed by Human Splice Finder (HSF). HSF predicted it to be deleterious and thus we took it for functional characterization (http://www.umd.be/HSF3/). LCLs of the case with *TP53* variant and a healthy control were revived using method described earlier. After achieving the count of 2.5 million cells, they were exposed to 2Gy gamma irradiation to achieve full expression of p53 in the LCL. One hour prior to irradiation the cells were treated with Puromycin for blocking Nonsense mediated decay (NMD) of RNA. Radiation was given for puromycin treated as well as puromycin untreated cells. LCLs without radiation exposure (0 Gy) were kept as controls. After irradiation, cells were taken for RNA extraction after 3 hours.

2.15.1 RNA extraction

RNA was extracted using conventional TRIzol method as described below

1. Cells were taken in 1.5 ml eppendorf and centrifuged at 4ºC and supernatant discarded. 1 ml of TRIzol reagent was added and mixed well by pipetting.

2. Chilled chloroform (200ul) added to each tube and vortex at maximum speed for 1 min.

4. The vortexed mixture was kept on ice for 15 mins.

5. The Eppendorf tube was then centrifuged at 12000 rpm at 4°C for 15 mins to ensure extraction of RNA into the aqueous layer. The interphase layer is the DNA of the tissue and the organic phase is the protein and lipid content of the tissues.

6. The aqueous phase was slowly pippetted out in one go without disturbing the interphase to avoid DNA contamination and placed in fresh Eppendorf tube.

7. To the separated aqueous phase, isopropanol was added in 1:1 ratio and the mixture was vortexed for 10 seconds.

8. Incubated the Eppendorf on ice for 15 mins to ensure complete precipitation of RNA by isopropanol.

9. After 10 mins of incubation, the Eppendorf tubes were centrifuged at 12000 rpm for 15 mins at 4°C and the supernatant carefully discarded.

10. Transparent RNA pellet in the bottom of the tube.

11. The RNA pellet was washed twice with 70% ethanol (prepared in DEPC water) at 4°C and 12000 RPM for 10 mins. After the wash with absolute alcohol, the pellet was air dried in the hood for 10 mins. Care was taken that the pellet is not kept for drying for more than 10 mins to prevent degradation of RNA.

12. The air dried pellet was now suspended in 20 µl of DEPC water.

13. The RNA was quantified by Nanodrop and integrity checked by 1% agarose gel electrophoresis after which it is stored at -80 degrees.

2.15.2 First strand synthesis and PCR

First strand synthesis was done using Amersham Biosciences First-Strand cDNA Synthesis Kit (Catalogue 27-9261-01). cDNA synthesis or Reverse Transcriptase PCR (RT-PCR) is a PCR where the starting material is total RNA. Here, RNA is

transcribed into complementary DNA (cDNA) by reverse transcriptase enzyme followed by its amplification.

Template: Total RNA instead of mRNA was used. The purity of RNA is important as the contaminants (DNA, proteins and phenol) may affect the reaction.

Primers: For cDNA synthesis random hexamer primers provided with the kit were used. These primers anneal to the mRNA to facilitate reverse transcription.

Steps:

1. 1-5µg RNA in a 20 µl reaction volume (RNAse free water) was heated at 65 degrees waterbath for 10 mins and then chilled on ice.

2. The bulk first-strand cDNA reaction mix was mixed properly by pipetting to obtain a uniform suspension. The appropriate volume of the bulk first strand cDNA reaction mix (11µl) was added to a sterile 1.5 ml microcentrifuge tube. To this tube 1µl of DTT solution and 1µl of the random hexamer primers were added to heat denatured RNA. Pipetted up and down several times to mix and incubated at 37°C for 1 hr. The first strand cDNA was synthesized and ready for next steps of PCR.

PCR and gel loading

PCR is performed using the cDNA specific primers given in table 2.7 to amplify the cDNA of *TP53* to reveal any aberrant transcript.

Table 2.7 Primers for cDNA of *TP53* gene covering the DNA binding Domain

Primer	**Sequence**	**Properties**	**Annealing Temp**
*TP53*_cDNA_F	GATGATTTGATGC	Present at Exon 4	55ºC

	TGTCC		
*TP53*_cDNA_R	GGATCTGAAGGG TGAAATATTC	Junctional primer Exon 9-10	

2.15.3 Sequencing of Aberrant transcript

The aberrant transcript visualized in gel was cut and cleaned using NucleoSpin Extract kit II of Macherey-Nagel using the given protocol along with the kit. Cleaned products were taken for sequencing as described in section 2.2.6.

2.16 Statistical methods:

All the statistical analysis was performed in Statistical Package for the Social Sciences (SPSS) software version 21.0. Different statistical methods were used for various genotype–phenotype correlations. For various association analysis of polymorphisms with various clinical data and phenotypic data, chi square test was used. To compare mean age at diagnosis in different genotypes, mean comparison t-test or non parametric test was used depending on the distribution. The survival probability was estimated using Kaplan Meier method and the differences were compared using the log rank test. For subgroup analysis instead of $p<0.05$, a p-value of <0.01 was considered as statistically significant in sarcoma cohort. For rest of the analysis $p<0.05$ was considered as significant.

Chapter 3

Comprehensive genetic analysis of *TP53* and MDM2 in the Indian Li-Fraumeni Syndrome cohort

3.1 Li-Fraumeni Syndrome an introduction

Li Fraumeni Syndome (LFS; Mendelian Inheritance in Man [MIM] #151623) is a clinically and genetically heterogeneous inherited cancer predisposition syndrome. LFS is characterized by an autosomal dominant inheritance pattern, diagnosis of various tumor types at a young age, multiple primary tumors within an individual, and multiple affected family members(118). LFS is distinct from other hereditary cancer syndromes in that it is associated not with one (or a few) specific type(s) of cancer, but with a broad spectrum of tumors, including several uncommon or rare cancers, across all ages.

LFS was initially described in 1969 by Frederick Li and Joseph Fraumeni. Using a classic epidemiological approach, they retrospectively evaluated 280 medical charts and 418 death certificates of children diagnosed with RhabdoMyoSarcoma (RMS) in the United States from 1960 to 1964 (119). Four families were identified in which siblings or cousins had a childhood sarcoma. In addition, a high frequency of diverse cancer types was also observed among first and second degree relatives in these 4 families. Along with soft tissue sarcomas and premenopausal breast cancers, carcinomas of the lung, skin, pancreas or adrenal cortex, leukemia, and various brain tumors were observed. Li and Fraumeni suggested that the occurrence of diverse neoplasms in these families might represent a previously undescribed familial cancer syndrome, with transmission suggestive of an autosomal dominant gene. Subsequent prospective studies confirmed a high risk in family members to develop tumor types that comprise LFS (120, 121). This syndrome was once referred to as SBLA syndrome, an acronym derived from the tumor spectrum in this hereditary syndrome:

sarcoma, breast and brain tumors, leukemia, laryngeal and lung cancer, and adrenal cortical carcinoma (122, 123).

The second milestone in the history of LFS was the simultaneous identification of germline *TP53* mutations as the genetic basis of LFS by two independent groups (85, 86). Based on earlier observations that somatic mutations of the p53 tumor suppressor gene were observed in more than half of sporadic human cancers and that p53 mutant transgenic mice developed a wide spectrum of tumors, p53 was examined in the constitutional DNA of LFS kindreds identifying germline mutations in all the families studied (85, 86).

3.1.1 Clinico-Pathological criteria for diagnosis of LFS

The first criteria for the diagnosis of classic LFS syndrome was established in 1988 based on the studies on 24 families (124). Since its initial description, the French LFS working group has modified the LFS criteria to include the expanding tumor spectrums associated with germline *TP53* mutations and to facilitate its clinical diagnosis (125-128). The criteria proposed by Birch et al in 1994, Eeles in 1995 and Chompret in 2001 is more inclusive and termed as Li Fraumeni–like syndrome (LFL) which has features of the Li Fraumeni syndrome but the strict definition of LFS is not fulfilled. The criteria proposed by Chompret are more restrictive in terms of the cancer types and age of cancer diagnosis in the proband, but it allow for the possibility of absence of family history. Various criteria for the diagnosis of LFS and LFL are outlined in Table 3.1

Table 3.1 Clinical criteria for Li-Fraumeni and Li-Fraumeni-like syndromes.

Criteria	Description
Classical LFS (124) **Year 1969**	• Sarcoma diagnosed in childhood/young adulthood (≤ 45 years) **and** • First-degree relative with any cancer in young adulthood (≤ 45 years) **and** • First- or second-degree relative with any cancer diagnosed in young adulthood (≤ 45 years) **or** sarcoma diagnosed at any age.
LFL – Birch (125) **Year 1994**	• Any childhood cancer (at any age) or sarcoma, CNS tumor, or ACC in young adulthood (≤ 45 years) **and** • First- or second-degree relative with LFS-spectrum cancer (sarcoma, BC, CNS tumor, ACC, leukemia) at any age **and** • First- or second-degree relative with any cancer diagnosed at age < 60 years.
LFL – Eeles (129) **Year 2001**	• At least 2 first- or second-degree relatives with LFS-spectrum cancer (sarcoma, BC, CNS tumor, ACC, leukemia, melanoma, prostate cancer, pancreatic cancer) diagnosed at any age; **or** • Sarcoma diagnosed at any age **and** At least 2 other tumors diagnosed in one or more first- or second-degree relatives: BC at age < 50 years; CNS tumor, leukemia, ACC, melanoma, prostate cancer, pancreatic cancer at age < 60 years; **or** sarcoma at any age.
LFL – Chompret (126) **Year 2001**	• Diagnosis of sarcoma, CNS tumor, BC, ACC at age < 36 years **and** • First- or second-degree relative with any of the above cancers (except BC if proband had BC) or relative with multiple primary tumors at any age **or** • Multiple primary tumors, including two of the following: sarcoma, CNS tumor, BC, or ACC, with the first tumor diagnosed at age < 36 years regardless of family history; **or** • ACC at any age, regardless of family history.
LFL – Modified Chompret (88, 127, 128) **Year 2009 and 2015**	• Index case with LFS-spectrum cancer (sarcoma, BC[^], CNS[*] tumor, ACC[$], leukemia, bronchioloalveolar carcinoma) occurring at age < 46 years **and** • First- or second-degree relative with LFS-spectrum cancer occurring at age < 56 years (except BC if the index case has BC as well), **or** multiple tumors; **or** • Index patient with multiple tumors, at least two of which are in the LFS spectrum, the first occurring at age < 46 years; **or** • ACC or choroid plexus carcinoma occurring at any age or BC occurring at age < 36 years without BRCA1or BRCA2 mutations. • RMS of embryonal anaplastic subtype, irrespective of family history **or** Early-onset breast cancer < 31 yr included in 2015.

[^]BC-Breast Cancer; [*]CNS- Central Nervous System related; [$]ACC-Adrenocortical Cancer

Various studies have identified pathogenic or likely pathogenic germline *TP53* mutations in nearly 70% of classic LFS cases and in 20-40% of LFL families (87).

3.1.2 Cancer spectrum in LFS/LFL syndromes

A wide spectrum of cancers is observed in the setting of LFS or LFL syndrome. Breast cancer and sarcomas (bone and soft-tissue sarcoma), are the most common cancers reported in LFS families. Sarcoma represent around 25% (130) and breast cancer account for approximately 27% (131) of all LFS-related cancers. Other core cancers associated with LFS include brain tumors, adrenocortical carcinoma (ACC) and leukemia (131). In addition to these cancers, the incidence of lymphoma, melanoma, lung, pancreas, prostate and ovarian cancers is also increased in LFS families (118). The most common LFS associated childhood cancers are ACC, RMS, choroid plexus carcinoma and CNS tumors like gliomas and medulloblastoma (124, 130, 132, 133). In addition, gastric and colorectal cancers can occur in LFS patients at an early age, although the increased occurrence of these cancers in individuals from LFS families with respect to the general population remains unclear (134, 135). Similarly, lung, testicular, laryngeal, and head and neck cancers might be more frequent in LFS than the general population (136); however, whether these tumor types arise sporadically in LFS and/or LFL syndrome families or are truly part of LFS cancer spectrum is not clear at present.

LFS patients are also at increased risk of developing multiple synchronous or metachronous primary tumors. The risk of additional cancers is inversely correlated with the age at diagnosis of the first malignancy (137). The risk of second primary tumor according to the age of first diagnosis is described in Table 3.2. In a large

clinical cohort of 525 consecutive patients with germline *TP53* mutations, half of the identified carriers had two or more primary cancers, compared with 32% in patients without *TP53* mutations (138). Moreover, the mean age of onset of first malignancy was 21.9 years in the *TP53* mutation carriers versus 31.6 years in those without mutations in this gene (138). Similarly, another prospective study of 559 sarcoma cases reported that 47% of the carriers of *TP53* mutations had multiple cancers whereas only 15% of mutation negative patients developed multiple cancers.

Table 3.2 Risk of second primary tumor in LFS/LFL families

Age at diagnosis of 1st primary tumor (years)	Relative risk of a second primary tumor (95%CI)
0-19	83.0 (36.9-87.6)
20-44	9.7 (4.9-19.2)
≥45	1.5 (0.5-4.2)
All ages	5.3 (2.8-7.8)

This table is taken from reference (139)

3.2 Molecular basis of LFS and the role of *TP53*

LFS is associated with heterozygous germline mutations in the *TP53* gene, which codes for a transcription factor implicated in cell proliferation, apoptosis, and genomic stability. Germline *TP53* mutations seem to be highly penetrant in the setting of classic LFS; *TP53*-associated cancer eventually develops in 73% of men and almost 100% of women who carry such mutations, with the higher penetrance in the latter predominantly attributable to breast cancer (140, 141).

3.2.1 Carcinogenesis due to *TP53* mutations

Using the model of sporadic and hereditary retinoblastoma due to the *RB1* tumor suppressor gene (TSG), Knusdosn first described how two hits are required to inactivate a TSG. The *TP53* gene is also a TSG and follows the Knudson's two hit

model, in sporadic cancers, two hits in the gene are acquired after birth. In contrast, in the hereditary LFS syndrome, a person is born with the first hit in *TP53* gene and the second hit acquired in the life time of individual making them more susceptible to cancer (Figure 3.1). In the hereditary LFS syndrome, the key features of the Knudson's two hit model are clearly evident in terms of earlier age of onset, bilateral and multiple primary cancers. These hits (mutations) in the gene leads to the loss of function (LoF) (88, 89) or conversely, the acquisition of 'dominant negative' (DN) (67) or 'gain of function' (GoF) (90-93) effects as defined in chapter 1. The loss of tumor suppressor function is caused by reduced *TP53* expression due to genomic deletions or altered *TP53* function due to missense mutations in the DBD. The missense mutations produce mutant p53 proteins with impaired DNA binding and thus altered regulation of downstream target genes. Regardless of the mechanism of action, *TP53* mutations lead to aberrant transcription or altered MDM2-*TP53* interactions, mutant p53 accumulation, deregulated cell cycle progression and impaired mechanisms of DNA repair.

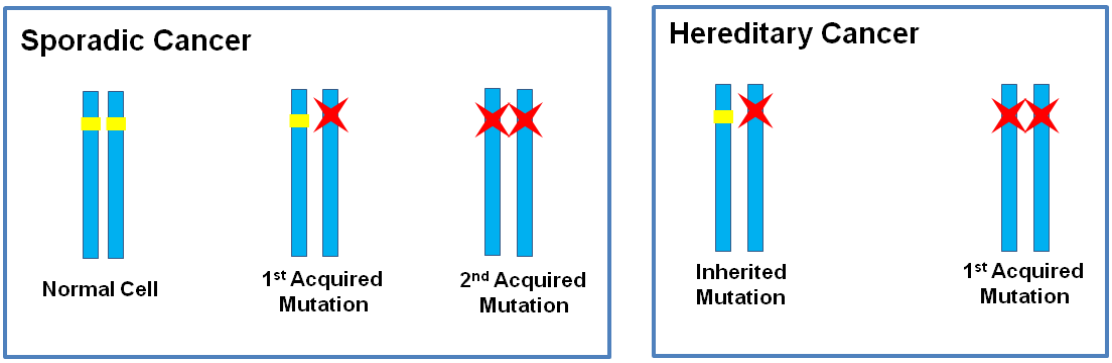


Figure 3.1: Knudson two hit hypothesis. Two independent hits required for sporadic cancer while only one hit can damage the function of gene. Yellow highlighted region-Normal gene; Red cross-damaged gene.

3.2.2 LFS associated germline *TP53* mutations

Since the first two reports of germline *TP53* mutations in LFS families by Malkin et al and Srivastav et al in 1990, several cases reports, case series and cohorts of LFS and LFL families have described the *TP53* genotype and its associated phenotype from different regions of the world (88, 112, 114, 127, 136, 139, 142, 143). Similar to the somatic mutations, germline *TP53* mutations are also clustered in the DNA Binding Domain (Exon 4-8) of *TP53*. This region harbours six recurrent 'hotspot' germline mutations involving codons 175 (R175H), 220 (Y220C), 245 (G245S), 248 (R248Q, R248W), 273 (R273H) and 282 (R282W) (144). Functional studies have revealed that majority of the deleterious mutations analysed so far exert their effect through DNE (145). Several mutations in the intronic or regulatory regions of *TP53* are reported but their functional significance or pathogenicity is not yet clearly established.

The International Agency for Research on Cancer (IARC) located in Lyon, France has collated and curated information from all publications and databases to create the most comprehensive database of somatic and germline *TP53* mutations (146) as discussed in chapter-1. The IARC database provides detailed description for each *TP53* mutation including its pathogenicity as predicted by 4 in silico tools (AGVGD, SIFT, Polyphen2 and REVEL) and its functional characterization, if known. It also provides important and detailed phenotypic information on each individual with germline *TP53* mutation in these LFS/LFL families . This included geo-ethnicity, pedigree information, maternal or paternal inheritance, whether affected with cancer, age at cancer diagnosis, cancer site and its histology. The recently released August 2018 R19 version of IARC germline *TP53* database (http://p53.iarc.fr) includes data

for 456 unique *TP53* mutations identified in 1221 families. Among the germline *TP53* mutations included, missense alterations constitute 73.6%, nonsense 7.9%, splice site 7%, frameshift 7.3%, silent 0.4%, large deletions 1.1%, intronic 0.5% and other mutations (complex rearrangement, insertions, deletions) 1.7% (Figure 1.8). The higher frequency of missense mutations in *TP53* is in stark contrast with other tumor suppressor genes which are frequently inactivated by truncating (frameshift and nonsense) mutations.

3.3 Genotype-Phenotype correlation in LFS/LFL

Several studies have evaluated the *TP53* genotype-phenotype correlation in LFS / LFL families and found some association between the location and the type of germline *TP53* mutation with the tumor type and age of cancer diagnosis. An early study (147) on 34 LFS families reported higher penetrance and significantly younger age at diagnosis for several cancers in families with DNA binding region missense mutations as compared to families with nonsense or truncating mutations or when no mutation was identified. More recently, Bougeard et al (2015) observed a significantly earlier age of cancer diagnosis in patients with missense mutations compared to those with truncating mutations(88). The difference was even greater when missense mutations were compared to genomic rearrangements. Additionally, patients with mutations exerting a dominant negative effect were associated with an even earlier age at cancer diagnosis. Interestingly, ACC appear to be the only LFS-associated cancer that is consistently associated with germline mutations outside the DNA-binding domain, with the majority of patients carrying mutations that are not dominant-negative missense mutations (88).

3.4 LFS associated *TP53* Genotype-Phenotype differences across various populations

We examined the R19 release of the germline *TP53* IARC database for difference in the LFS tumor spectrum and the *TP53* mutation spectrum between Caucasians, Central / South Americans and the Asian population. Majority of the families in this IARC database are Caucasian and account for 826/1221 families with germline *TP53* mutation. South / Central American countries and Asian countries are underrepresented. Surprisingly, only 6 families are from south Asia which includes India. A comparative data on the tumor and mutation spectrum in LFS is described in Figure 3.2 and 3.3 and Table 3.3 and 3.4

Table 3.3 Geo-ethnic differences in tumor spectrum in families with *TP53* mutation (Data from R19, IARC)

Tumor	Central & South America (%)	Caucasian (%)	Asia (%)	Chi-square (df*=2) pvalue
ACC	90 (25%)	136 (5%)	14 (3%)	<0.001
Sarcoma (Bone & STS)	34 (9%)	497 (20%)	63 (14%)	<0.001
CNS	37 (10%)	355 (14%)	38 (9%)	<0.01
Breast	70 (19%)	706 (28%)	177 (40%)	<0.001
Lung	14 (4%)	85 (3%)	17 (4%)	<0.8
Hemato Lymph	10 (3%)	127 (5%)	27 (6%)	<0.08
Gynaecological	7 (2%)	81 (3%)	9 (2%)	<0.2
GI including	45 (13%)	161 (6%)	58 (13%)	<0.001

Hepatobilliary				
Head Neck	11 (3%)	37 (1%)	9 (2%)	<0.8
Skin	7 (2%)	52 (2%)	3 (1%)	<0.14
Others	31 (9%)	146 (6%)	16 (4%)	-
Unknown site	3 (1%)	124 (5%)	10 (2%)	-
	359 (100%)	2507 (100%)	441 (100%)	

*Degree of freedom

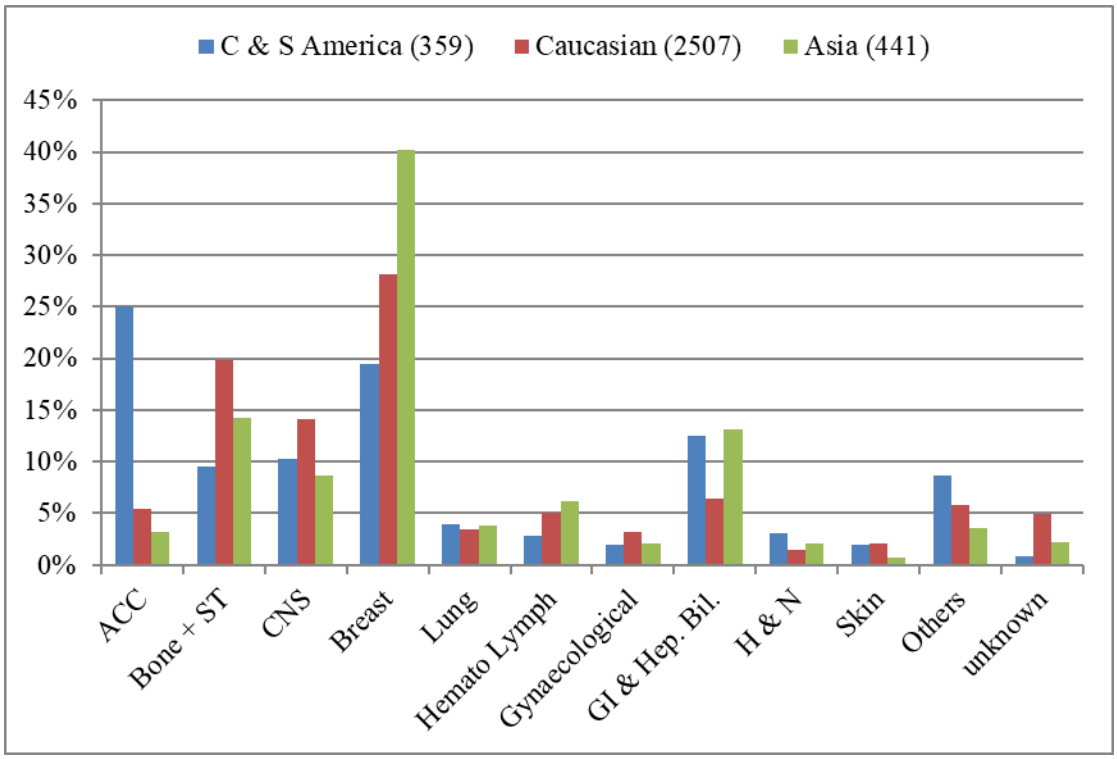


Figure 3.2 Geo-ethnic differences in tumor spectrum in families with *TP53* mutation (IARC R19)

Table 3.4 Geo-ethnic differences in *TP53* mutation (Data from R19, IARC)

	South & Central America		Caucasian		Asia		
	No. of individuals	**No. of carrier families**	**No. of individuals**	**No. of carrier families**	**No. of individuals**	**No. of carrier families**	**Chi-square (df*=2) pvalue**
MS	354 (83%)	115	1840 (73%)	591	366 (80%)	125	<0.0001
FS	9 (2%)	2	192 (8%)	57	45 (10%)	15	<0.0001
NS	13 (3%)	1	201 (8%)	74	37 (8%)	7	<0.01
Splice	18 (4%)	3	258 (10%)	68	6 (1%)	4	<0.0001
LGR	30 (7%)	4	38 (2%)	12	2 (0.4%)	2	<0.0001
Total	424	125	2529	802	456	153	

*** Degree of Freedom**

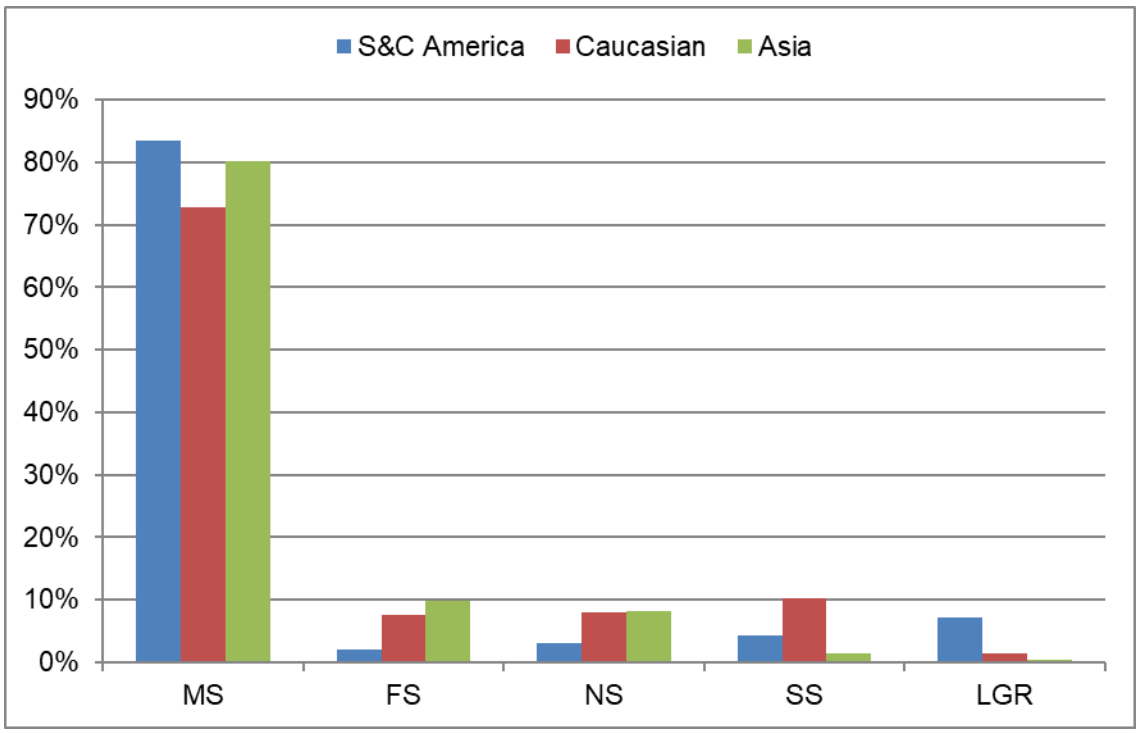


Figure 3.3 Comparison of mutation spectrum in various populations (IARC R19). MS-Missense; FS-Frame Shift; NS-Non Sense; SS-Splice Site; LGR- Large Genomic Rearrangement

3.5 Role of other genes in LFS

Several studies have failed to identify *TP53* germline mutations in a proportion of classic LFS and LFL families, thus raising the possibility of involvement of other genes. Various studies investigated the contribution of germline mutations in genes of the P53 pathway, apoptosis, or cell cycle control such as p63, BCL10, CDKN2A, PTEN, and CHEK1/CHEK2 in LFS(148-154). However, none of these studies could identify genes other than *TP53* in classic LFS families. Few studies reported possible role of germline mutations in CHEK2 gene which encodes a kinase that phosphorylates Cdc25c and p53 in 3 LFS/LFL families (152, 155, 156). However, one of the reported mutation in CHEK2 1422delT, was subsequently shown to be on a duplicated exon and the other two reported germline mutations Ile157Thr and 1100delC were found to be reported as both pathogenic and variant of unknown significance (VUS) in ClinVar database. Taken together, these data argue against a major role of CHEK2 in LFS. Another locus that was reported to be linked to LFS was chromosome 1q23, but the association of this locus to LFS remains to be confirmed by further studies (157). The syndromic features of LFS and LFL partly overlap with few other syndromes such as HBOC syndrome, constitutive mismatch repair deficiency (CMMRD) syndrome and Neurofibromatosis type 1. Hence in some *TP53* negative LFS/LFL families, the germline mutation may be in genes like BRCA1, BRCA2, PMS2, MLH1, MSH2, MSH6 or NF1. In CMMRD syndrome the presentation of multiple early onset and childhood cancers may mimic LFS but the distinguishing factor is the homozygous mutation in PMS2 or other mismatch repair genes and presence of café-au-lait spots (158).

3.6 Genetic modifier effect of *TP53* and MDM2 polymorphisms in LFS

The phenotypic variability in terms of type of cancers and age at diagnosis in carriers of specific *TP53* mutation within and between families has supported the notion that the larger genetic background of a germline *TP53* mutation carrier influences the phenotype to some extent. To date, a number of polymorphisms have been identified as having possible roles in determining the clinical outcome in carriers of germline *TP53* mutations. Further recognition of such genetic modifiers may prove helpful in defining the surveillance and management strategies for families with *TP53* mutations.

In LFS cases with germline pathogenic *TP53* mutation, the MDM2 gene SNP309 (NM_002392.3:c.14+309T>G; SNP309; rs2279744) has been shown to be associated with an earlier age of tumor onset (94, 159, 160) and multiple primary cancers (159). In a study conducted on 88 LFS patients, an earlier age of cancer diagnosis was observed in association with MDM2 SNP309 (159, 161). Specifically, individuals with the heterozygous G/T or homozygous G/G genotype developed STS at an age 12 years earlier and breast cancer 10 years earlier than those carrying the homozygous wild-type T/T genotype. Also, the number of primary tumors was higher in STS patients with SNP309 G/T or G/G genotype. Studies conducted to understand the mechanism behind this correlation revealed that the G allele for SNP309 results in increased expression of MDM2 protein by increasing the binding of Sp1 transcription factor. This eventually leads to increased degradation of *TP53* and thus defective *TP53*-mediated apoptosis (159, 161). The modifier effect of MDM2SNP309 polymorphism is further enhanced by the presence of an additional SNP in codon 72 of *TP53* (NM_000546.5 (*TP53*): c.215C>G; rs1042522) in which the proline (P) at

codon 72 is replaced by arginine (R). The presence of MDM2 SNP309 and *TP53* R72 was shown to have a cumulative effect and associated with earlier age of cancer onset (87). This effect can be explained by the observation that as compared to the *TP53* containing P72, *TP53* protein with R72 binds MDM2 with higher affinity, resulting in enhanced degradation. Another MDM2 polymorphism, SNP285G>C (rs117039649), was reported to be strongly linked with the SNP309G allele in a cohort of 195 LFS patients. An earlier age of onset was observed in LFS cases with the MDM2 285–309 G–G haplotype as compared to other haplotypes (162). However, the interaction between the MDM2 285–309 G–G haplotype and the *TP53* codon 72 polymorphism was not examined.

Other polymorphisms within *TP53* have also been reported to modify the penetrance of germline mutations in *TP53*. This include a 16bp duplication in intron 3 (PIN3; rs17878362) that was shown to be particularly important in determining the cancer risk in Brazilian carriers of the Arg337His (R337H) *TP53* mutation. The age of onset of cancer in the R337H carriers was observed on average 19 years earlier in individuals homozygous for the non-duplicated A1 allele compared with those with the A1A2 PIN3 genotype, who were heterozygous for the A2 allele containing the 16bp duplication (163). Another polymorphisms with modifier effect include a SNP within a microRNA known as miR-605 (rs2043556) (164) and germline DNA copy number variants (87).

3.7 Germline *TP53* mutations in sarcoma or breast cancer cases not fulfilling LFS/LFL criteria

A small number of studies have investigated the frequency of germline pathogenic *TP53* mutation in patients with the two most common LFS associated cancer in

absence of family history of cancer or a family history which does not fulfill the criteria for LFS/LFL syndrome. Mitchell et al (2013) identified *TP53* mutation in 7/465 (1.5%) sporadic adult onset sarcomas (99). Evans et al (2002) identified *TP53* mutation in 1/21 (5%) breast cancer cases with a family history of sarcoma but not fulfilling the classical LFS criteria (165). Diller et al in 1995 identified germline *TP53* mutation in 3/33 (10%) children with rhabdomyosarcoma in absence of family history or multiple primary cancers (166).

These studies suggest that LFS associated caners, especially sarcoma and breast cancers, when presenting as adult onset sporadic cancers or with some family history which does not fulfill the LFS/LFL criteria, germline *TP53* mutation may be identified in <5% cases. Identification of germline *TP53* mutation has major implications in subsequent surveillance of the proband and also for the genetic counseling and screening of the family. However due to limited number of studies, the role of *TP53* genetic testing in such cases is not clearly established. This highlights the need to establish the prevalence of *TP53* germline mutation in larger cohorts of sarcoma and other LFS associated cancers with or without family history of cancer that does not fulfill the current LFS/LFL criteria.

3.8 Study objectives

The geo-ethnic variation in the *TP53* genotype-phenotype correlation in different population is well known but the Indian population has not been studied. We hypothesize that the genotype – phenotype correlation in the Indian families may be distinct from the well characterized Caucasian population and other Asian population and this knowledge can help in developing population specific risk estimation, counseling and risk management strategies. This study aims to identify germline *TP53*

gene mutation spectrum and its correlation with the tumor spectrum and age at diagnosis in a large Indian cohort of hereditary LFS / LFL and sporadic LFS associated cancers.

3.9 Results

Germline *TP53* mutation analysis was conducted on 500 LFS/LFL/Suspected LFL families registered in Cancer Genetics Clinic, Tata Memorial Hospital and under Tata International Sarcoma Kindred Study (TISKS) during the period 2003 to 2018. Of these, 197/500 families fulfilled the defined criteria of LFS or LFL. Of the remaining 303 cases, in 281 cases we suspected LFL based on expanded LFL spectrum cancers by including Malignant Phyllodes and any Haematolymphoid malignancy and by relaxing age cut off from <46 years (Chompret, 2015) to <50 years and described in methods section (page 68). Of these 303 non LFS / LFL cases, 22 cases were screened even though they were not on our suspected LFL group as they had familial cancers not fulfilling criteria for suspected LFL (3A, 3B or 3C) or any other hereditary cancer syndrome. *TP53* mutation frequency in each of these groups is shown in Figure 3.4 and described in table 3.8. An example of the classic LFS, LFL and sLFL families are given in Figure 3.5, 3.6 and 3.7a and 3.7b. Deleterious germline *TP53* mutations were identified in 79 out of the 500 probands studied resulting in an overall mutation detection rate of 15.8%.

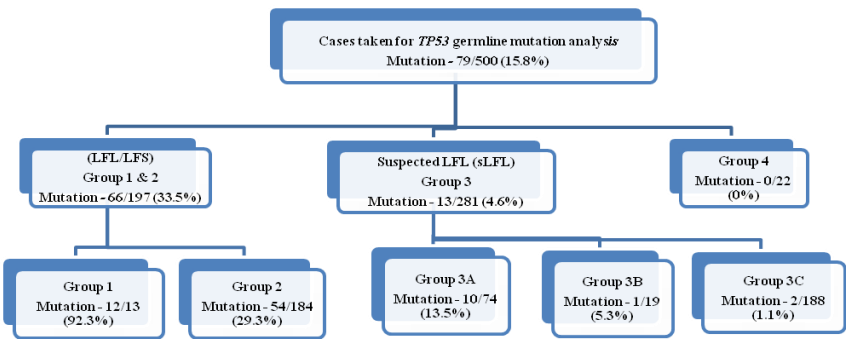


Figure 3.4 *TP53* screening groups and mutation frequency.

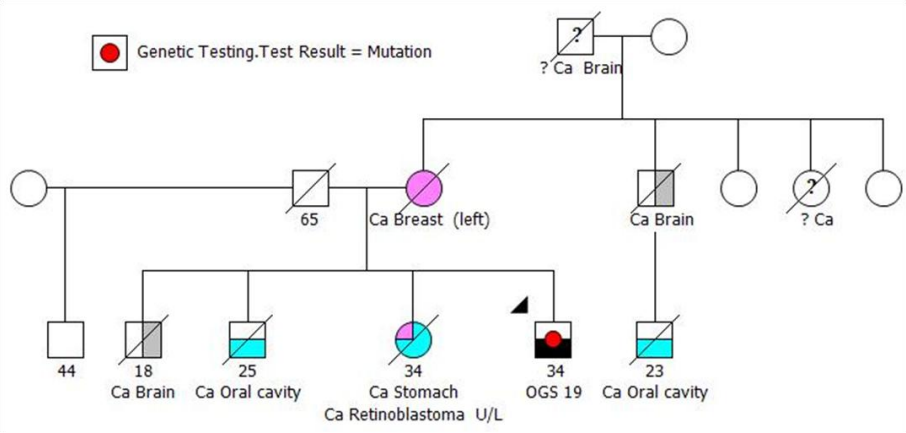


Figure 3.5 Pedigree of a classical LFS (Group 1) family with germline *TP53* mutation (c.272G>A)

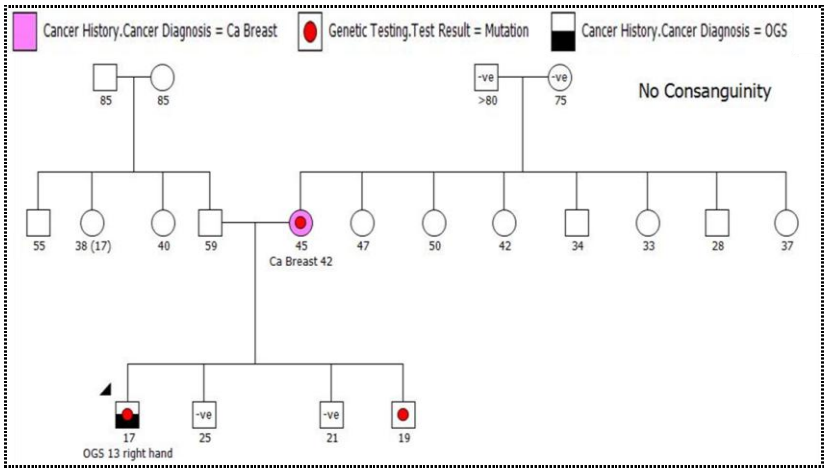


Figure 3.6 Pedigree of a LFL family (group 2) with germline *TP53* mutation (c.672+1G>T)

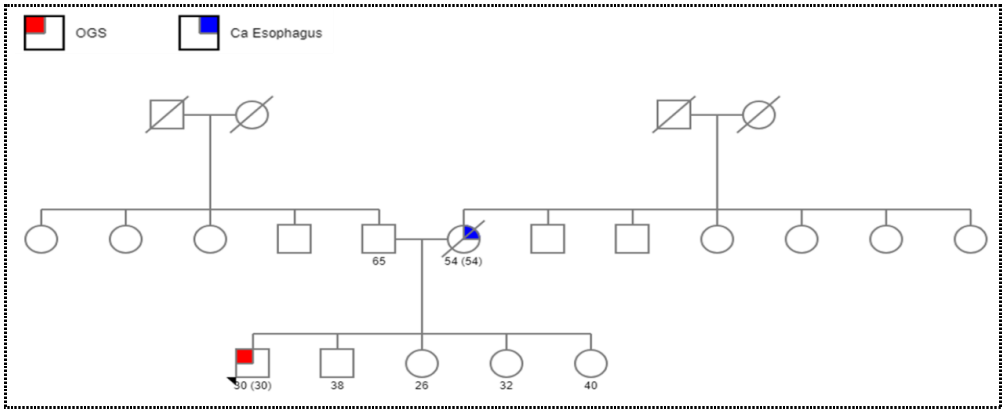


Figure 3.7A: Pedigree of a suspected LFL (Group 3A) family with germline *TP53* mutation (Promoter duplication)

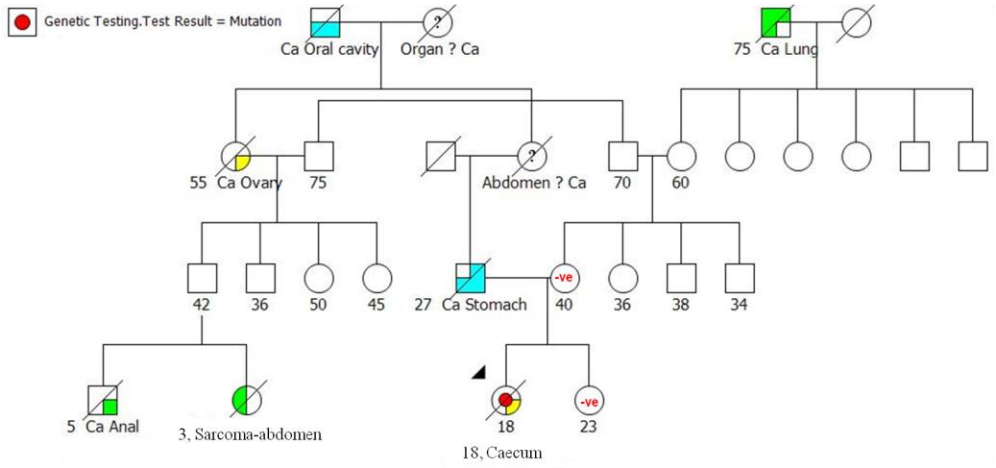


Figure 3.7B: Pedigree of a suspected LFL (Group 3B) family with germline *TP53* mutation (c.817C>T)

3.9.1 Tumor spectrum in the Indian cohort with *TP53* mutation

Phenotypic characterization was carried out in 356 individuals from the 79 *TP53* mutation positive families. As in the IARC database, we included tumors not only in confirmed *TP53* mutation carriers but also in the obligate carriers or other untested members in these LFS/LFL families with an identified *TP53* mutation. A total of 403 cancers were recorded in these families with breast cancer being the most common cancer type. Sarcoma was the second most common cancer followed by CNS tumors identified in 16% and 14% of the cases respectively. As expected, breast cancer accounted for 121/246 (49.2%) of all female cancers in these 79 families. Multiple primary tumors were identified in 40/356 (11%) individuals. In males, sarcoma was the most common cancer and accounted for 34 (21.7%) of the 157 cancers in males. The mean age of cancer diagnosis in the entire cohort was 32.1 years (range: 1-85 years). The frequency of various tumors types identified in our cohort is given in Figure 3.10 and 3.11.

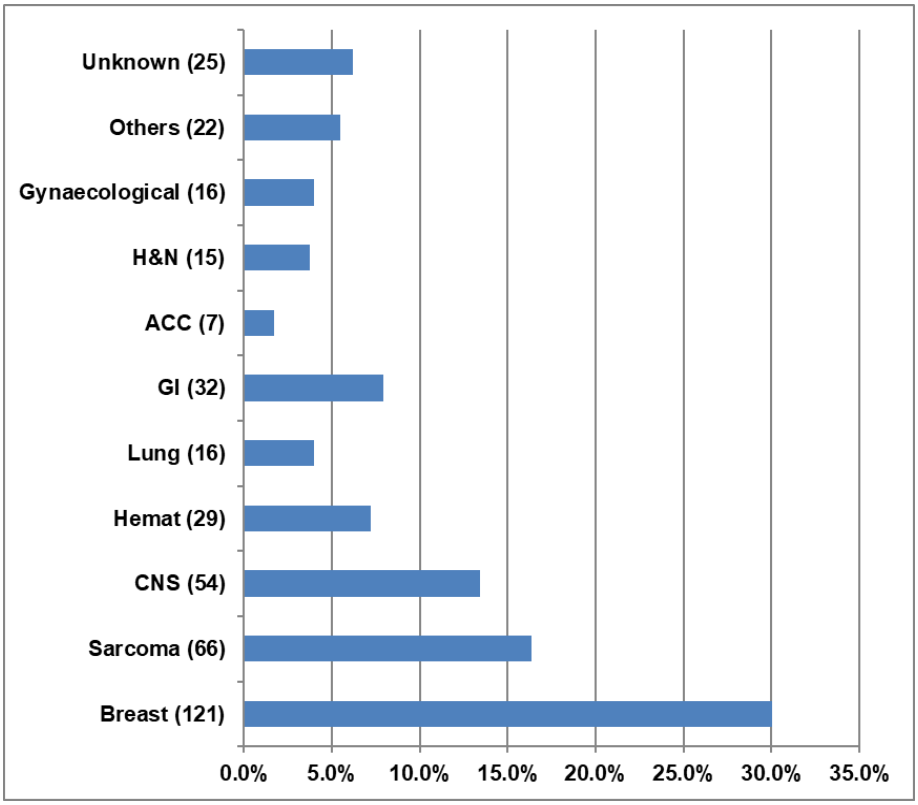


Figure 3.8 Spectrum of 403 cancers in 356 individuals from 79 *TP53* mutation positive familes

We also analysed our data for tumor spectrum in mutation positive members and obligate carriers. Of the 79 mutation positive families 132 members were either mutation positive or the obligate carriers. Of these 132 cases 34 are unaffected and 98/132 members were affected with 132 cancers with maximum frequency of breast cancer (38%) followed by bone and soft tissue cancers (32.6%) followed by haemato-lymphoid tumors (6.1%) and so on as shown in figure 3.9 and 3.10. Of the 132 carrier individuals (either mutation or obligate) 84 were females and 48 were males. Of the 84 female carriers 72 were affected with 101 cancers and of 48 males, 26 were affected with 31 cancers. In females 50/101 (50%) cancers were breast cancers as shown in figure 3.11.

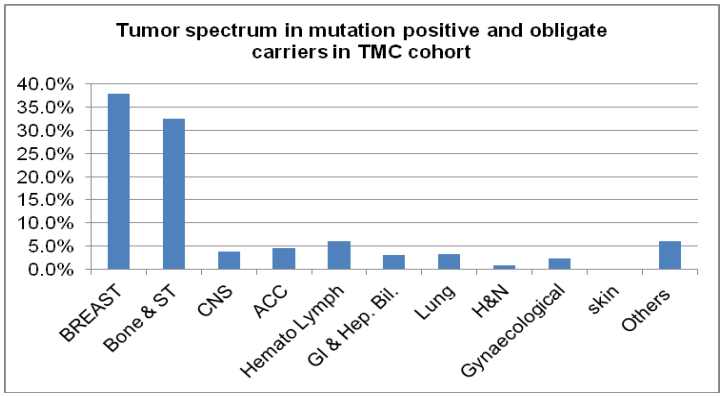


Figure 3.9: Spectrum of 132 cancers in 103 individuals with confirmed *TP53* mutation and obligate carriers.

Tumor spectrum of *TP53* mutation carriers and obligate carriers of TMC cohort were compared with IARC data of all mutation and obligate carriers (N=2095) as shown in figure 3.10. Tumor spectrum varies significantly for bone and soft tissues cancers p<0.01), CNS tumors (p<0.01) and adrenocortical cancers (p<0.05) in comparison to IARC database.

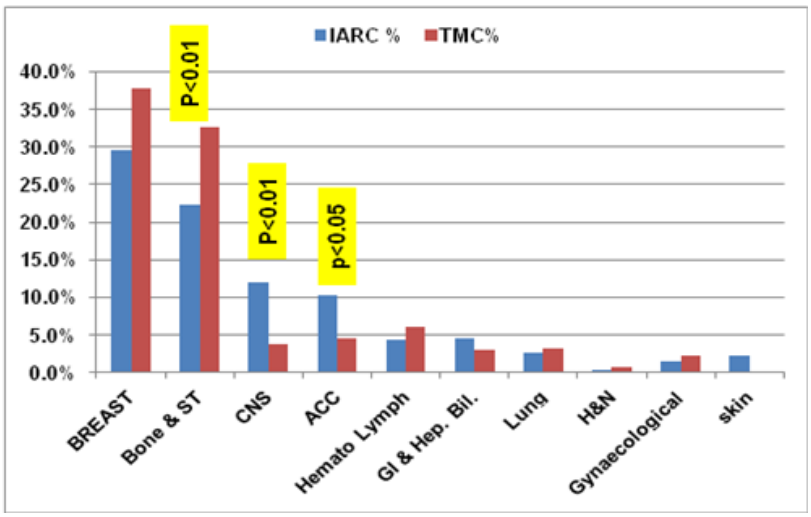


Figure. 3.10: Comparison of cancer spectrum between IARC (n=2095 tumours) & TMC (n=132 tumours) in confirmed carriers of *TP53* mutation and obligate carriers

Table 3.5: Comparison of tumor spectrum of mutation carriers and obligate carriers Between IARC and TMC cohort				
	IARC	IARC %	ACTREC	TMC%
BREAST	618	29.5%	50	37.9%
Soft Tissue & bone	468	22.3%	43	32.6%
CNS	253	12.1%	5	3.8%
ACC	217	10.4%	6	4.5%
Hemato-Lymphoid cancers	90	4.3%	8	6.1%
GI & hep.Bil.	94	4.5%	4	3.0%
Lung	57	2.7%	4	3.2%
Others	209	10.0%	8	6.1%
H&N	7	0.3%	1	0.8%
Gynecological	33	1.6%	3	2.3%
Skin	49	2.3%	0	0.0%
Total	2095		132	

As compared to the tumor spectrum reported in the IARC database (R19), we found significant difference in the frequency of various tumor types between our cohort and various other populations (Figure 3.12). In our cohort no skin cancers were seen and ACC were rare, accounting for only 7/403 (~1.7%) tumors. This is in stark contrast to the Central and South American cohort where ACC account for 90/359 (25%) of all cancers and the Caucasian population where they account for 136/2507 (5%) tumors (Figure 3.12).

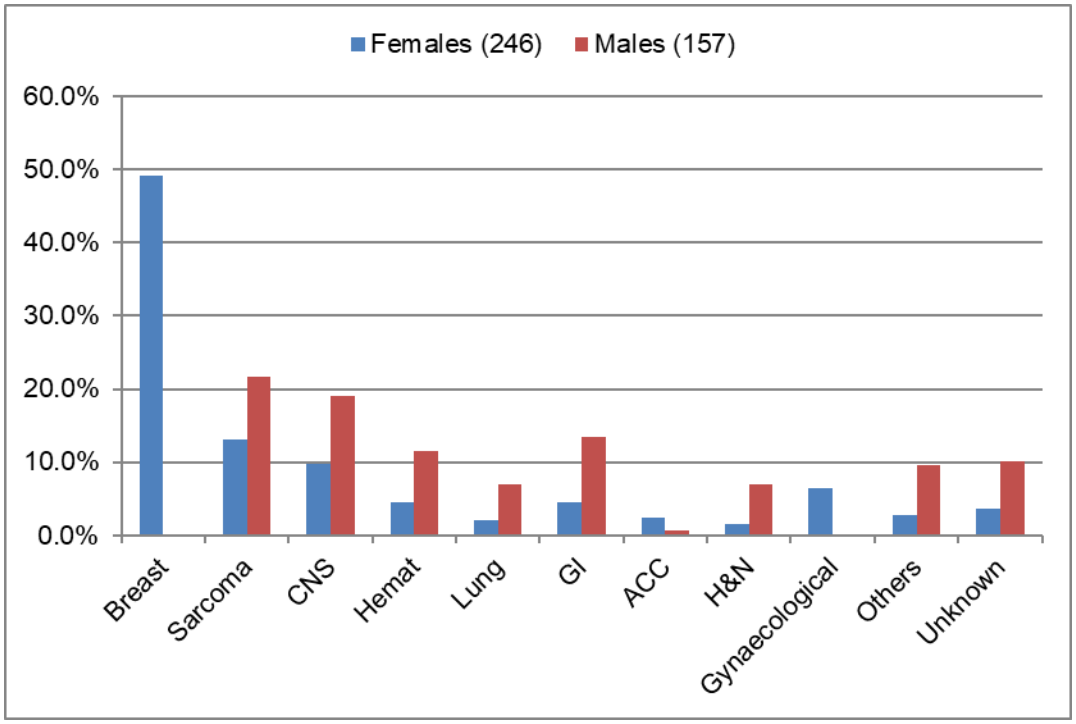


Figure 3.11: Frequency of tumors gender wise in TMC-LFS cohort

The other cancers which show significant differences in comparison to other populations includes sarcoma, head & neck cancers, gastro-intestinal cancers (Figure 3.12; table 3.6). In our cohort, there were 5 cases of malignant phyllodes of the breast in confirmed *TP53* mutation carriers whom we included in the breast cancer group for analysis. Of the 9 cases with Malignant Phyllode tumor of the breast we tested, *TP53* mutation was identified in 5 cases, all were affected with cancer under the age of 30 years.

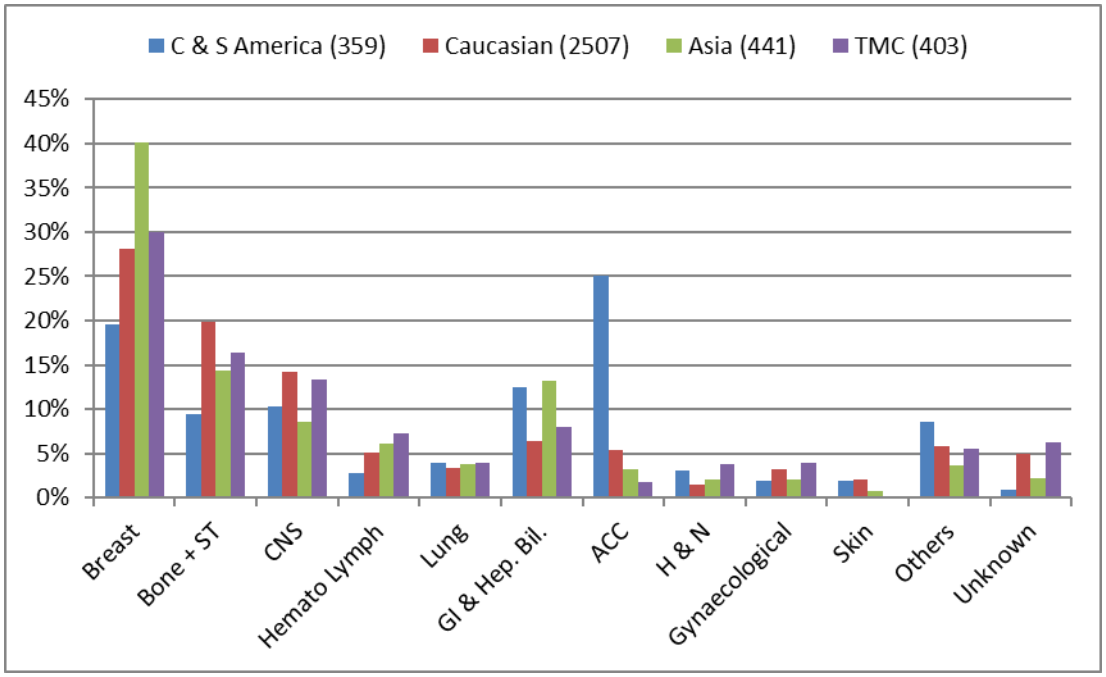


Figure 3.12 Comparison of tumor spectrum in different populations in IARC R19 and TMC

Table 3.6 Comparison of tumors of different populations along with TMC cohort					
Tumor	**Central & South America (%)**	**Caucasian (%)**	**Asia (%)**	**TMC (%)**	**Chi-square (df*=3) pvalue**
ACC	90 (25%)	136 (5%)	14 (3%)	7 (1.7%)	<0.00001
Bone + ST	34 (9%)	497 (20%)	63 (14%)	66 (16.4%)	<0.00001
CNS	37 (10%)	355 (14%)	38 (9%)	54 (13.4%)	<0.006
Breast	70 (19%)	706 (28%)	177 (40%)	121 (30%)	<0.00001
Lung	14 (4%)	85 (3%)	17 (4%)	16 (4%)	<0.8
Hemato Lymph	10 (3%)	127 (5%)	27 (6%)	29 (7.2%)	<0.04
Gynaecological	7 (2%)	81 (3%)	9 (2%)	16 (4%)	<0.2
GI+ Hep billiary	45 (12%)	161 (6%)	58 (13%)	32 (7.9%)	<0.00001
H & N	11 (3%)	37 (1%)	9 (2%)	15 (3.7%)	<0.008

***df-Degree of Freedom**

3.9.2 Germline *TP53* mutation spectrum

TP53 gene mutation analysis was carried out in the 500 LFS/LFL/SLFL cases using PCR-Sanger sequencing approach. A total of 59 unique mutations were identified in 79 families leading to an overall mutation detection rate of 15.8% in our cohort. Mutation in fourteen families out of these 79 was identified earlier and were re-sequenced and reanalyzed for the mutation and included in the 79 count. Out of the 197 LFS/LFL cases analysed, deleterious germline *TP53* mutations were identified in 68 families. The 79 families with germline *TP53* mutations hailed from various parts of south Asia and with different geo-ethnic backgrounds as shown in Figure 3.13

The point mutations and frameshift mutations were mostly between Exon 4 to 10 while LGRs were identified across the *TP53* gene. Majority of *TP53* germline mutations were in the DNA Binding Domain (Figure 3.14) which is comparable to the IARC data. As compared to the IARC R19 database, missense mutations were significantly less frequent and LGRs more frequent in our cohort (Figure 3.15). Five known hotspot mutations at codon 175, 220, 248, 273 and 282 were found to be recurrent mutations in our cohort also and identified in a total of 19 families. As expected, the Brazilian founder mutation R337H was not identified in our Indian cohort.

Similar to other populations, with the exception of the Brazilian population with an exon 10 founder mutation, 58/79 (73.4%) mutations in our cohort were in the DNA binding domain spanning exon 4 – 8 in (Figure 3.14). Frequencies of various types of mutations were compared with IARC database is shown in figure 3.15.

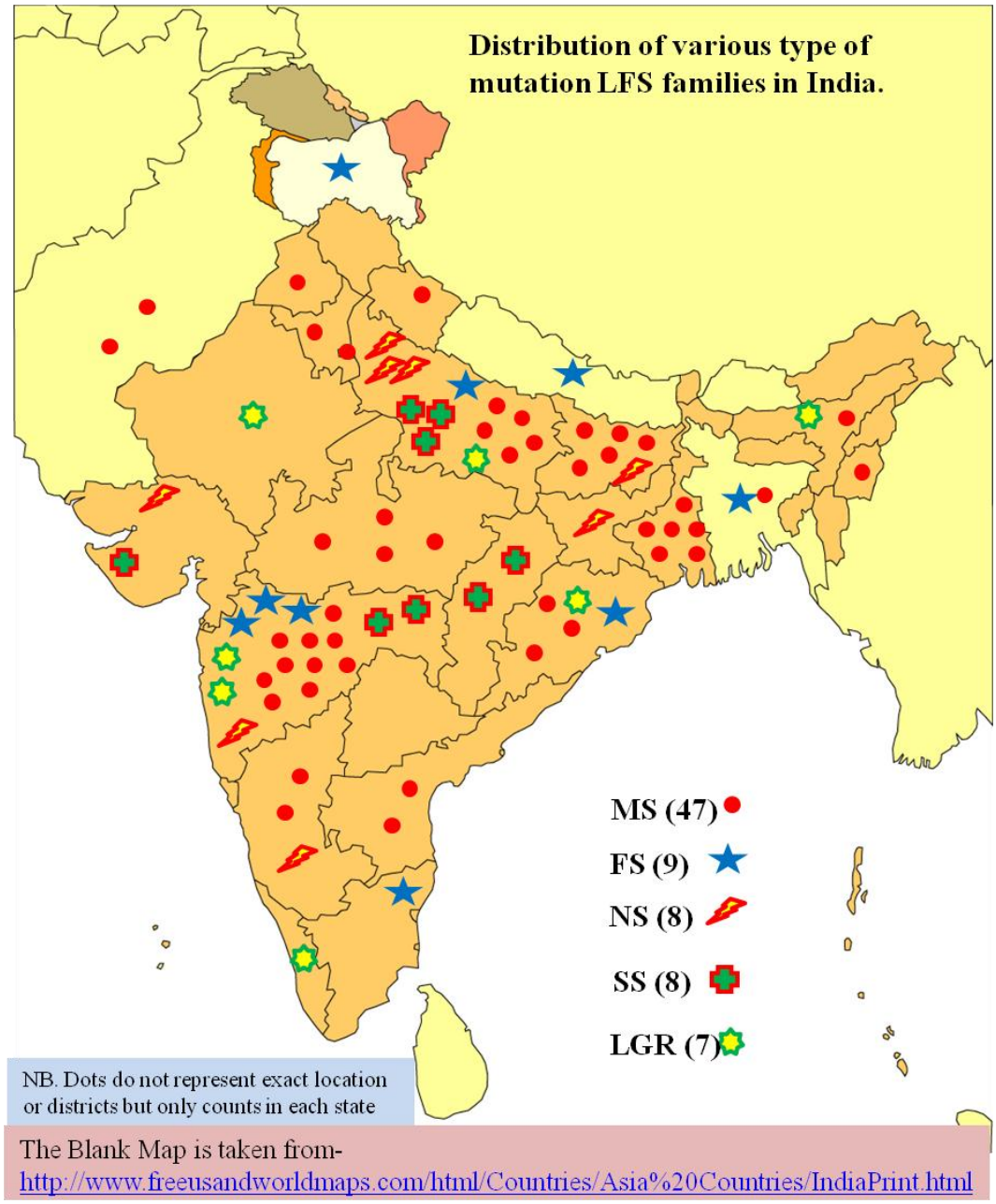


Figure 3.13 Geo-ethnic origin of 79 *TP53* mutation carrier families in south Asia. MS-Missense; FS-Frame Shift; NS-Non Sense; SS-Splice Site; LGR- Large Genomic Rearrangement

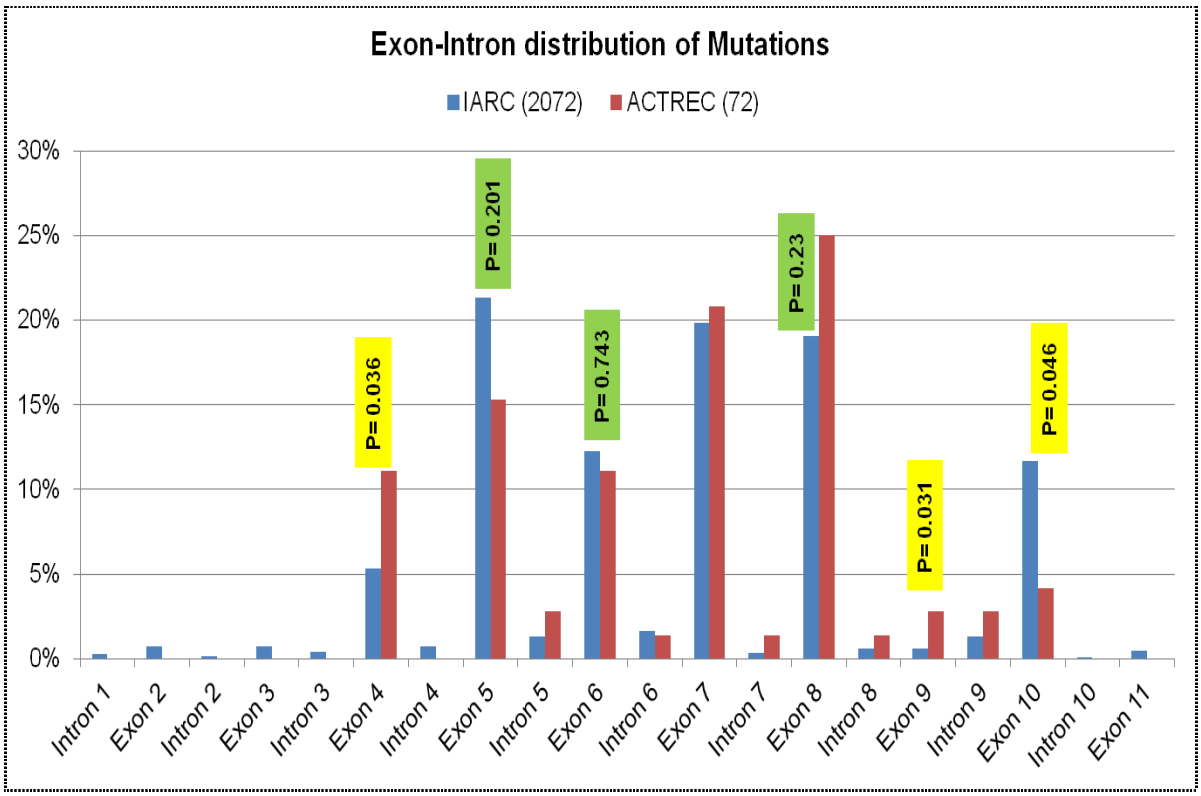


Figure 3.14 Distribution of *TP53* mutations in IARC R19 and our cohort

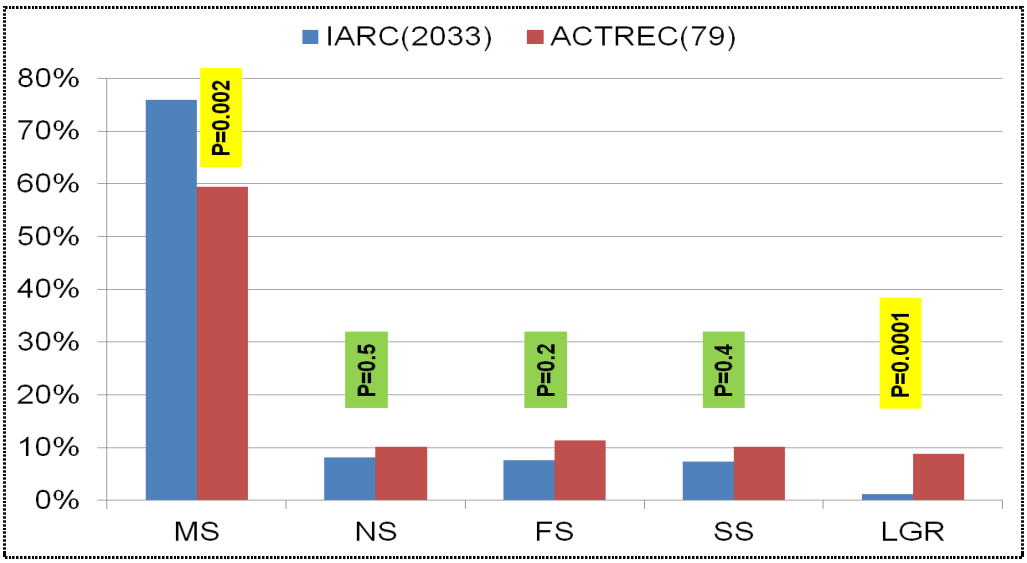


Figure 3.15 Frequency of different types of *TP53* mutation in TMC cohort and IARC R19.

MS-Missense; FS-Frame Shift; NS-Non Sense; SS-Splice Site; LGR- Large Genomic Rearrangement

Of the 59 distinct mutations that we identified, 24 were novel germline mutations. After further characterization based on co-segregation, *in-silico* prediction (AGVGD / SIFT) and the reported transactivation activity (17), 22/24 novel germline mutations were classified as likely pathogenic while 2 mutations remained VUS. Type of mutations compared with different populations with the TMC cohort which showed significant differences (Figure 3.16 and table 3.7). In addition, a variant was identified in the 3'UTR region of *TP53* gene in a suspected LFL case and considered as a VUS as discussed later.

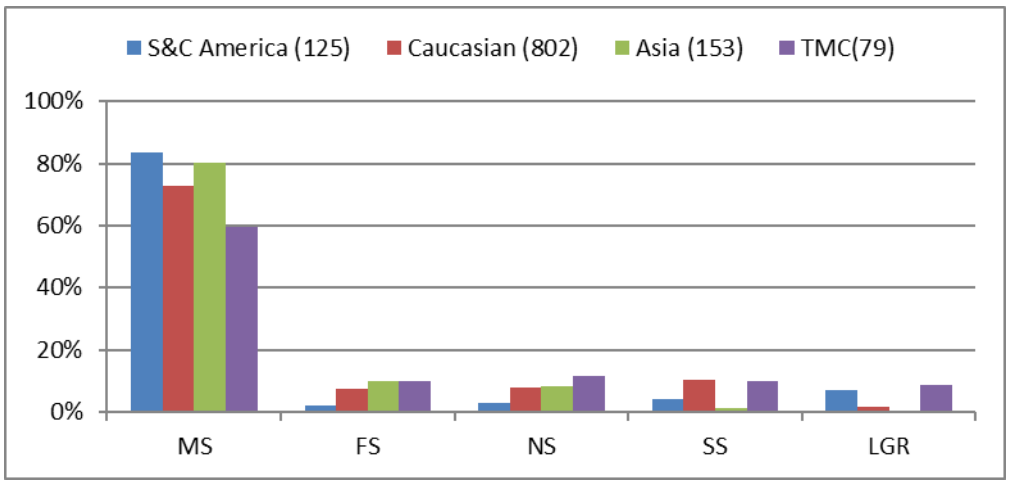


Figure 3.16: Comparison of types of mutations of TMC cohort with other populations. MS-Missense; FS-Frame Shift; NS-Non Sense; SS-Splice Site; LGR- Large Genomic Rearrangement

Table 3.7 Comparison of types of mutation of TMC cohort with various populations as given in IARC R19

	S&C America		Caucasian		Asia		TMC		
	Individuals	**Families**	**Individuals**	**Families**	**Individuals**	**Families**	**Individuals**	**Families**	**Chi-square(df=3) pvalue**
MS	354 (83%)	115	1840 (73%)	591	366 (80%)	125	75 (58%)	47	<0.0001
FS	9 (2%)	2	192	57	45	15	13	9	<0.0001

			(8%)		(10%)		(10%)		
NS	13 (3%)	1	201 (8%)	74	37 (8%)	7	14 (10.8%)	8	<0.002
Splice	18 (4%)	3	258 (10%)	68	6 (1%)	4	15 (11.6%)	8	<0.001
LGR	30 (7%)	4	38 (2%)	12	2 (0.4%)	2	12 (9.3%)	7	<0.001
	424	125	2529	802	456	153	129	79	

MS-Missense; FS-Frame Shift; NS-Non Sense; SS-Splice Site; LGR- Large Genomic Rearrangement

Mutations were classified as dominant negative based on their transactivation activity in previous studies (Kato et al) and as mentioned in the IARC database. Extended family screening in 119 members from these 79 families with likely pathogenic *TP53* mutation identified mutation in additional 50 individuals. Of these 50 individuals, 16 individuals with family specific mutation have already developed cancer. The remaining 34 carriers are as yet unaffected and on intensive surveillance protocol. Detailed description of *TP53* mutations identified in our cohort is given in Table 3.8

Table 3.8 *TP53* mutations in TMC LFS cohort

Mutation	c.DNA	No. of fam.	Exon/ Intr.	Domain	Ti or Tv	Novel/ Report	SIFT result	Trans-activation*
Missense mutations with DNE								
p.R175H	c.524G>A	3	5	DBD	Ti	Reported	Deleterious	NF
p.H179R	c.536A>G	1	5	DBD	Ti	Novel germline	Deleterious	NF
p.C242F	c.725G>T	1	7	DBD	Tv	Novel germline	Deleterious	NF
p.G244D	c.731G>A	1	7	DBD	Ti	Reported	Deleterious	NF
p.G245S	c.733G>A	2	7	DBD	Ti	Reported	Deleterious	NF

p.R248W	c.742C>T	3	7	DBD	Ti	Reported	Deleterious	NF
p.R248Q	c.743G>A	2	7	DBD	Ti	Reported	Deleterious	NF
p.R273C	c.817C>T	4	8	DBD	Ti	Reported	Deleterious	NF
p.R273H	c.818G>A	1	8	DBD	Ti	Reported	Deleterious	NF
Missense mutations with moderate DNE								
p.C141Y	c.422G>A	1	5	DBD	Ti	Reported	Deleterious	NF
p.H193R	c.578A>G	1	6	DBD	Ti	Reported	Deleterious	NF
p.Y220C	c.659A>G	3	6	DBD	Ti	Reported	Deleterious	NF
p.Y236C	c.707A>G	1	7	DBD	Ti	Reported	Deleterious	NF
p.R267W	c.799C>T	1	8	DBD	Ti	Reported	Deleterious	NF
p.R282W	c.844C>T	3	8	DBD	Ti	Reported	Deleterious	NF
Missense mutations with DNE status not available								
p.P98R	c.293C>G	1	4	Outside	Tv	Novel	Deleterious	PF
p.F109S	c.326T>C	1	4	DBD	Ti	Novel germline	Deleterious	NF
p.R110L	c.329G>T	1	4	DBD	Tv	Reported	Deleterious	NF
p.L111P	c.332T>C	1	4	DBD	Ti	Novel germline	Deleterious	NF
p.A119G	c.356C>G	1	4	DBD	Tv	Novel germline	Deleterious	PF
p.T125R	c.374C>G	1	4	DBD	Tv	Reported	Deleterious	NF
p.H168L	c.503A>T	1	5	DBD	Tv	Novel germline	Neutral	NF
p.R175C	c.523C>T	1	5	DBD	Ti	Reported	Deleterious	PF
p.C176G	c.526T>G	1	5	DBD	Tv	Novel germline	Deleterious	NF
p.R196Q	c.587G>A	1	6	DBD	Ti	Reported	Deleterious	PF
p.R213L	c.638G>T	1	6	DBD	Tv	Novel germline	Deleterious	NF
p.C242R	c.724T>C	1	7	DBD	Ti	Reported	Deleteri	NF

							ous	
p.M246L	c.736A>C	1	7	DBD	Tv	Novel germline	Deleterious	NF
p.R249T	c.746G>C	1	7	DBD	Tv	Novel germline	Deleterious	NF
p.E258A	c.773A>C	1	7	DBD	Tv	Novel germline	Deleterious	NF
p.E287K	c.859G>A	1	8	DBD	Ti	Novel germline	Neutral	F
p.K291R	c.872A>G	1	8	DBD	Ti	Novel germline	Deleterious	F
p.R342P	c.1025G>C	2	10	Outside	Tv	Reported	Deleterious	NF
Nonsense mutations								
p.W91*	c.272G>A	1	4	Outside	Ti	Novel germline	NA	NA
p.Y163*	c.489C>G	1	5	DBD	Tv	Novel germline	NA	NA
R196*	c.586C>T	1	6	DBD	Ti	Reported	NA	NA
p.R213*	c.637C>T	1	6	DBD	Ti	Reported	NA	NA
p.R306*	c.916C>T	2	8	Outside	Ti	Reported	NA	NA
p.Q331*	c.991C>T	1	9	Outside	Ti	Reported	NA	NA
p.R342*	c.1024C>T	1	10	Outside	Ti	Reported	NA	NA
Frameshift mutations								
p.147delG fs169*	c.439delG	1	5	DBD	NA	Novel germline	NA	NA
p.250delC fs344*	c.748delC	1	7	DBD	NA	Novel germline	NA	NA
p.264delC fs344*	c.790delC	2	8	DBD	NA	Novel germline	NA	NA
p.282delG fs344*	c.846delG	1	8	DBD	NA	Novel	NA	NA
p.281insC fs305*	c.844_845insC	1	8	DBD	NA	Novel germline	NA	NA
p.301insC fs305*	c.902_903insC	1	8	Outside	NA	Novel germline	NA	NA
	10bp del	1	9	Outside	NA	Novel germline	NA	NA
	37bp del	1	5	DBD	NA	Novel germline	NA	NA

Splice site mutations								
T125T (Silent)	c.375G>T	1	EX 4	NA	Tv	Reported	NA	NA
-	c.560-1G>A	2	IN 5	NA	Ti	Reported	NA	NA
-	c.672+1G>T	1	IN 6	NA	Tv	Reported	NA	NA
-	c.783-2A>G	1	IN 7	NA	Ti	Reported	NA	NA
-	c.919+1G>A	1	IN 8	NA	Ti	Reported	NA	NA
-	c.993+1G>A	1	IN 9	NA	Ti	Reported	NA	NA
-	c.994-1G>C	1	IN 9	NA	Tv	Reported	NA	NA
Large Genomic Rearrangements								
Full gene	-	1	-	NA	NA	-	-	-
Duplication of 4 exons	-	2	-	NA	NA	-	-	-
Deletion of 5 exons	-	2	-	NA	NA	-	-	-
Promoter Duplication	-	2	-	NA	NA	-	-	-

*Based on transactivation studies done (167); F: Functional protein; NF: non-functional protein; PF: partially functional protein; Ti: transition; Tv: tansversion; DBD: DNA Binding Domain; NA: Not applicable. DNE classifications was taken from the IARC database which is based on the functional validation of the variants.

3.9.3 Large genomic rearrangements in *TP53* gene

MLPA analysis was carried out in 202 cases where no mutation was identified in *TP53* gene through Sanger sequencing. LGRs were identified in 7 families (4 distinct LGRs). The frequency of LGR in our cohort is relatively higher as compared to the IARC database. The identified LGRs include a deletion of entire *TP53* gene in a

LFS/LFL family; duplication of 5 exons and deletion of 4 exons in 2 families each and promoter duplication in 2 families.

LGR 1: Deletion of entire gene

A large deletion was in a classic LFS case (pedigree in Figure 3.17). Proband is a 17 year old girl who was diagnosed with Osteogenic Sarcoma (OGS) at the age of 16 years. The LGR identified is about a 72kb deletion on chromosome 17p encompassing entire *TP53* gene (except Exon 1) and its adjoining genes; ATPB1 and a part of SHBG (Figure 3.18 and Figure 3.19)

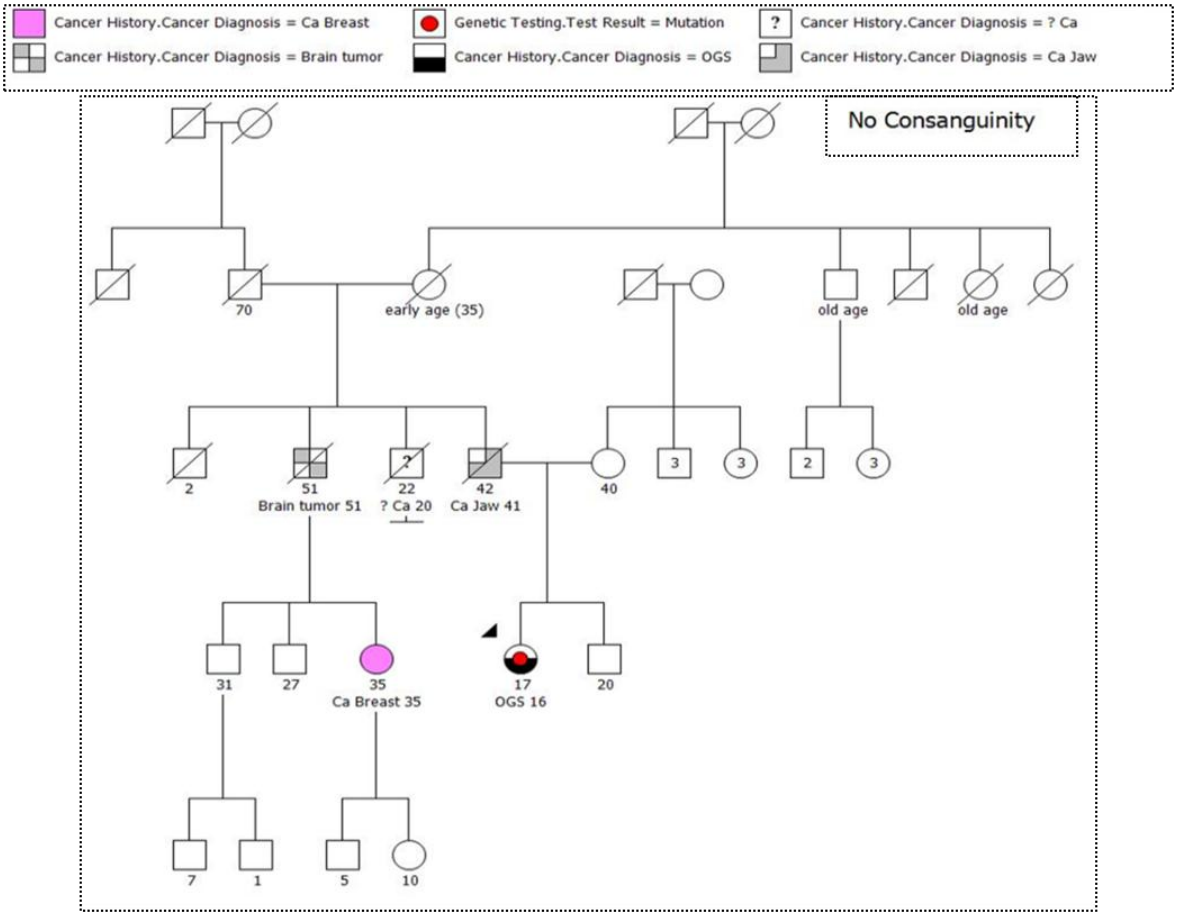


Figure 3.17 Pedigree of classic LFS family with deletion of *TP53* gene

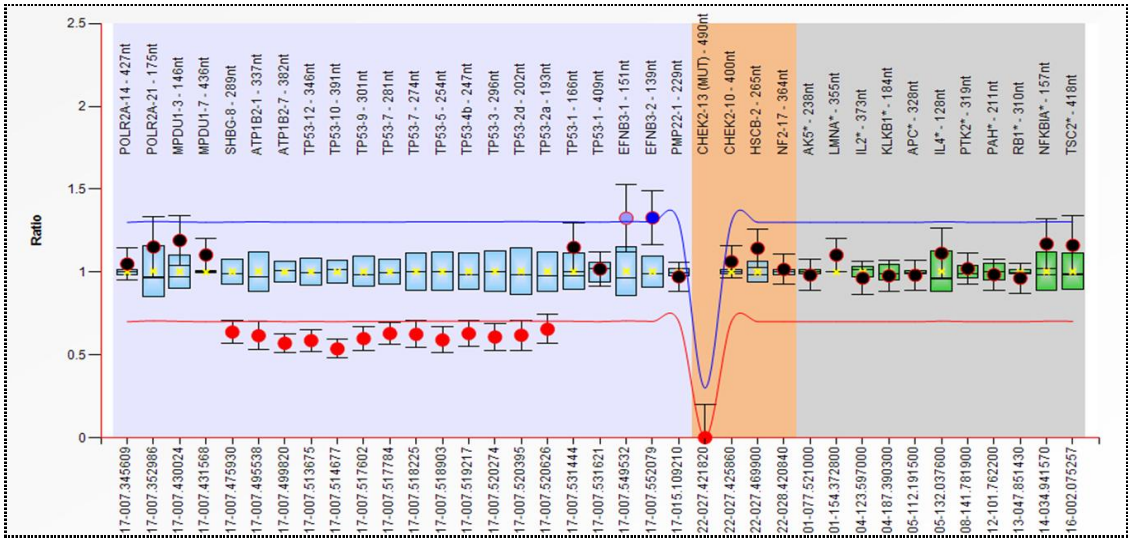


Figure 3.18 Coffalyser output showing reduced copy number for probes of *TP53* Exon 2-11; ATBP1 and SHBG genes

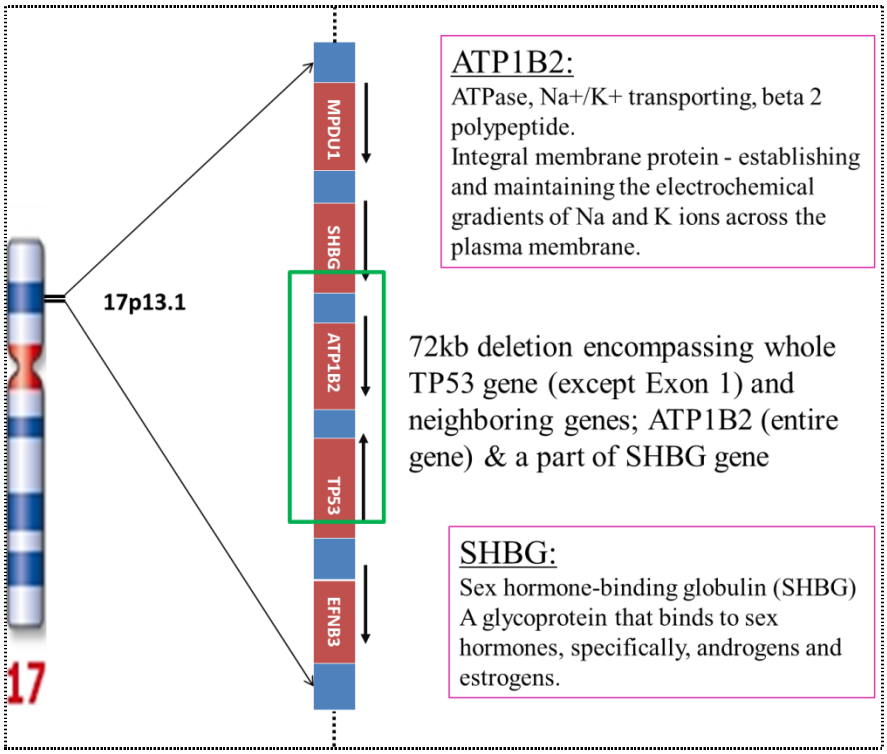


Figure 3.19: Schematic representation of LGR showing the arrangement of deleted genes on chromosome 17p.

LGR 2: Duplication of Exon 2-6 of *TP53*

This duplication LGR (Figure 3.20) was identified in two cases. The first case was diagnosed with rhabdomyosarcoma of the right elbow at the age of 6 years. His father was diagnosed with brain tumor at the age of 26 years (Pedigree in Figure 3.21). The other case with this duplication is a girl from an unrelated family diagnosed with ACC at the age of 13 years. She also had first and second degree relatives with diverse cancers and a diagnosis of LFL syndrome was made in this family (Pedigree in Figure 3.22).

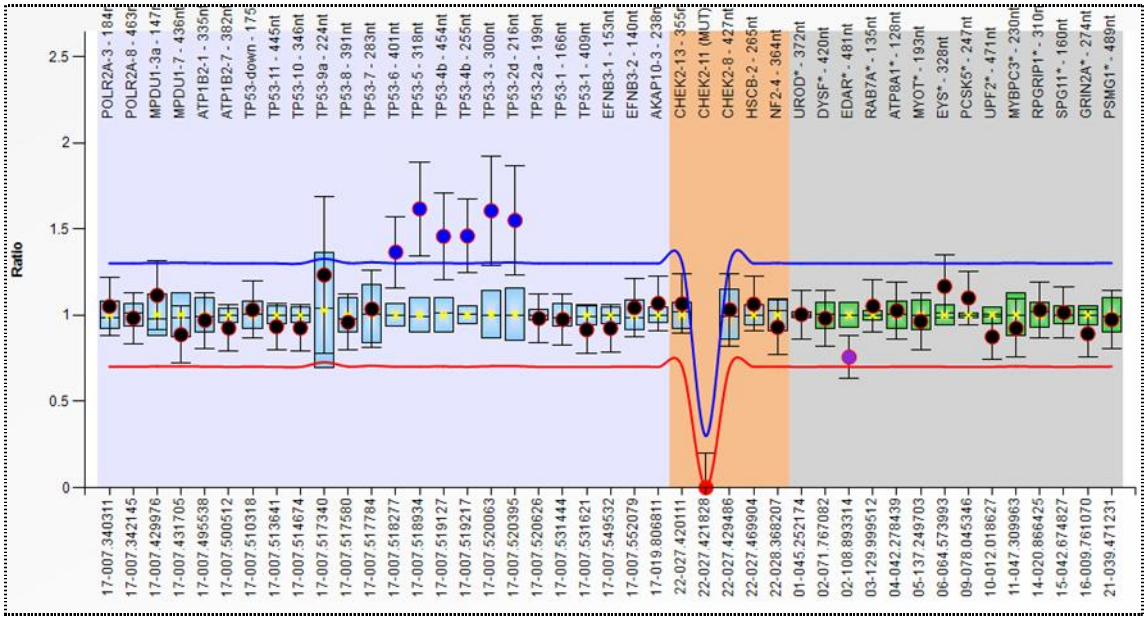


Figure 3.20: Coffalyser output showing duplication of Exon 2-6 of *TP53* gene

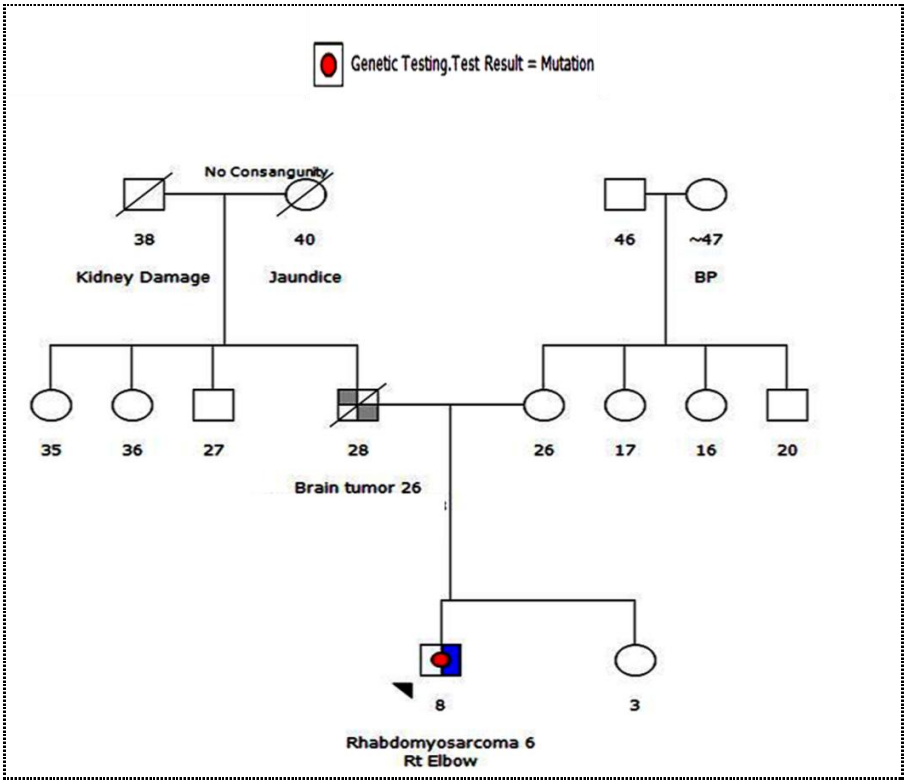


Figure 3.21 Pedigree of the LFL family with duplication of exons 2-6 of *TP53* gene

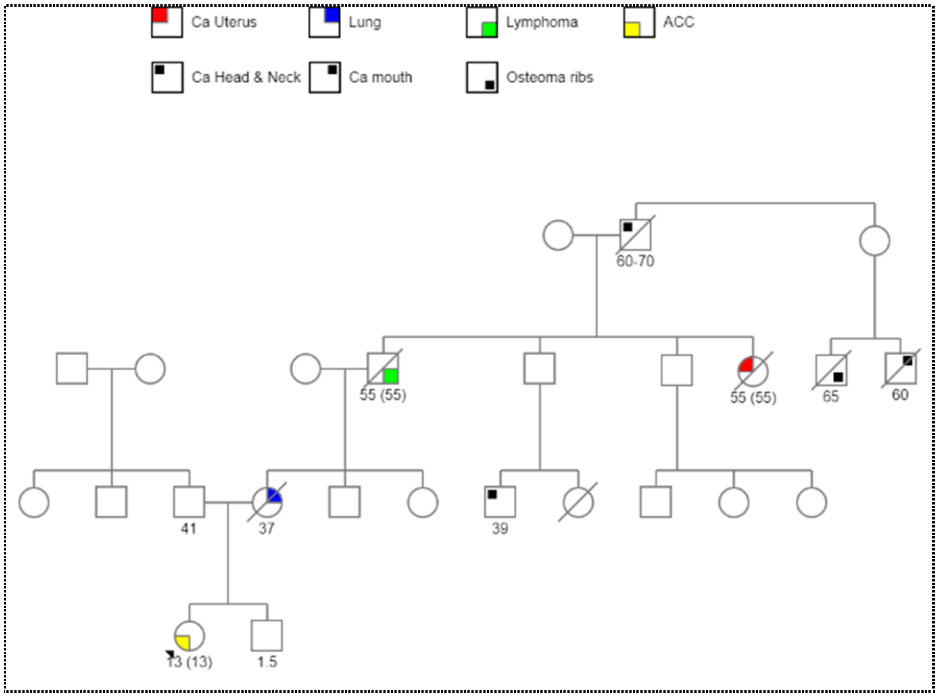


Figure 3.22 Pedigree of LFL family 2 with duplication of exons 2-6 of *TP53* gene

LGR 3: Deletion of Exon 2-5 of *TP53*

This 4 exon deletion was identified in two unrelated families. In the first case, proband was a 39 year old female who was diagnosed with breast cancer at the age of 38 years. She had a strong family history of LFS associated cancer (Figure 3.23). The other family with this LGR is a LFL family in which the proband was diagnosed with breast cancer at the age of 32 years (Pedigree in Figure 3.24). The Coffalyser output showing the Dosage Quotient for this LGR is given in Figure 3.25

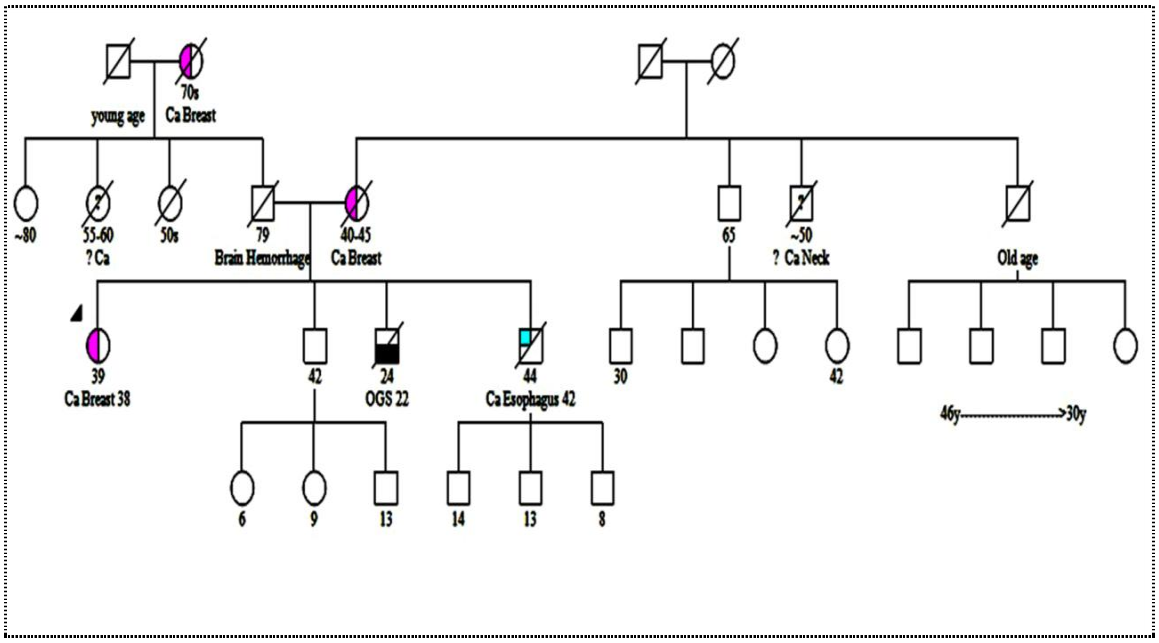


Figure 3.23: Pedigree of the LFL family with deletion of exons 2-5 of *TP53* gene

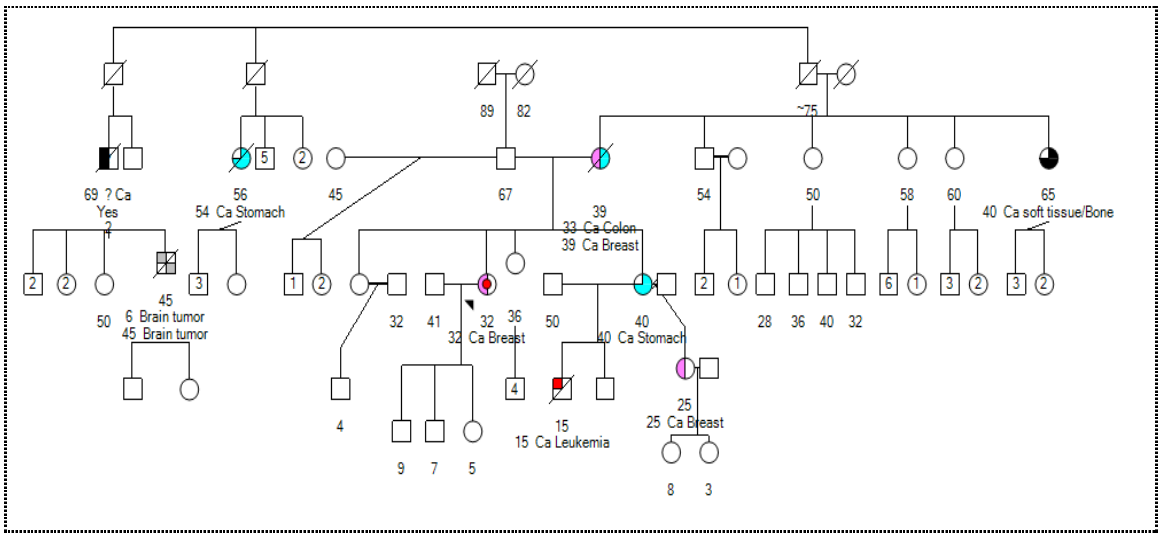


Figure 3.24: Pedigree of the LFL family 2 with deletion of exons 2-5 of *TP53* gene

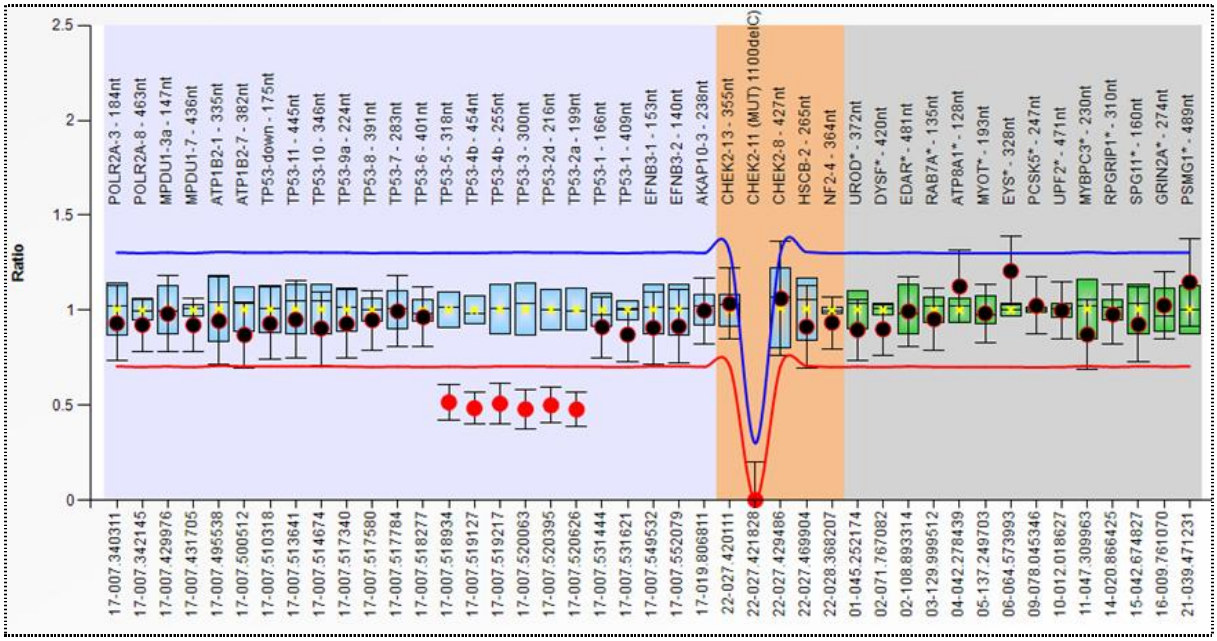


Figure 3.25: Coffalyser output showing deletion of Exon 2-5 of *TP53* gene

LGR 4: Duplication of *TP53* promoter

Promoter duplication was identified in 2 LFL/LFS families (Pedigrees in Figure 3.26 and 3.27). This Coffalyser output is given in Figure 3.28

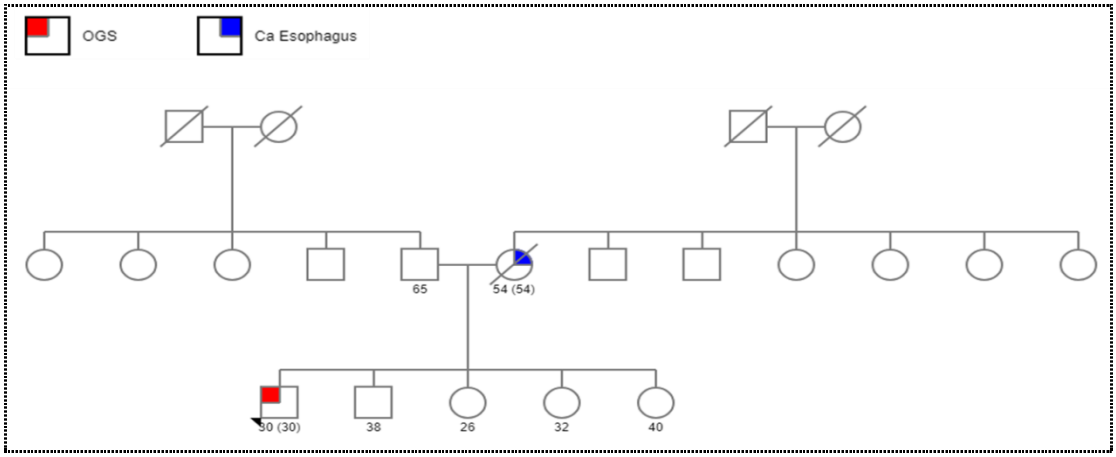


Figure 3.26: Pedigree of the Family 1 (Group 3A) with *TP53* Promoter duplication

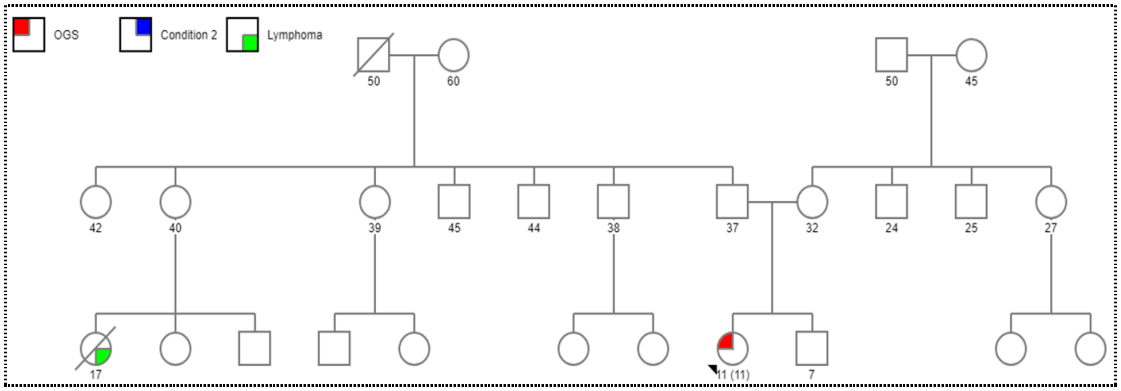


Figure 3.27: Pedigree of the Family 2 (Group 3A) with *TP53* Promoter duplication

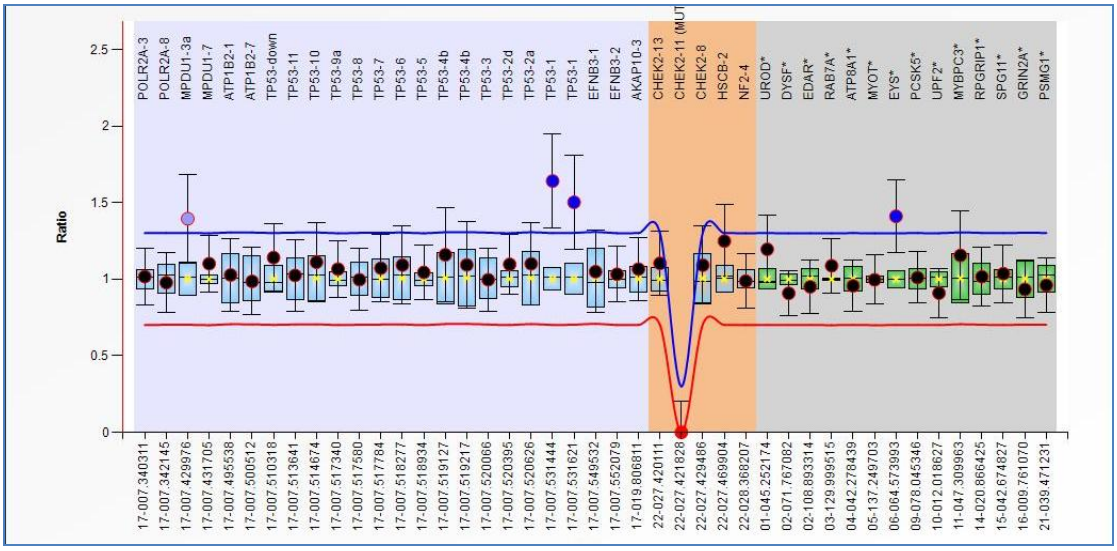


Figure 3.28: Coffalyser output of promoter duplication

3.9.4 Germline *TP53* mutations in sporadic sarcoma cases

Out of the 281 suspected LFL cases who underwent germline *TP53* analysis, 188 cases had sporadic sarcoma. We identified 2 distinct mutations in 2 sarcoma cases leading to a mutation detection rate of 2/188 (1.1%) in sporadic sarcoma. Both the mutations identified in sarcoma cases are codon 196 mutations; p.R196Q and p.R196* (Figure 3.29 & 3.30).

p.R196Q mutation was identified in a 42 year old male with chondrosarcoma

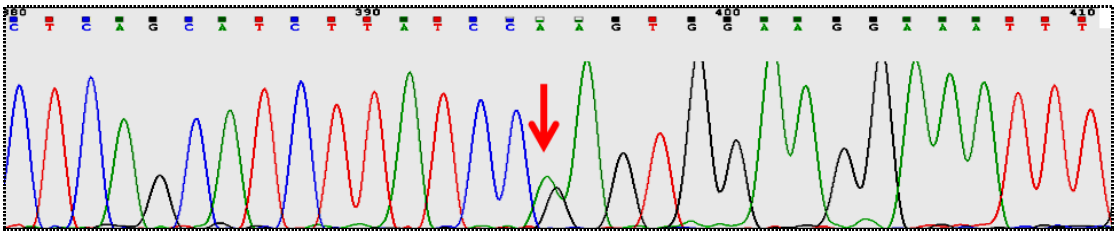


Figure 3.29: Chromatogram showing c.587G>A (R196Q) mutation in Exon 6 of *TP53* gene

The p.R196* mutation was identified in a 20 year old male diagnosed with osteosarcoma

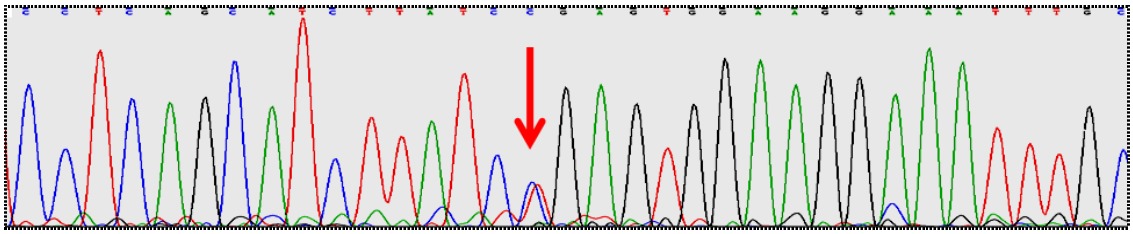


Figure 3.30: Chromatogram showing c.586C>T (R196*) mutation in Exon 6 of *TP53* gene

3.9.5 *TP53* Mutation detection rates in different subgroups

As expected, the *TP53* mutation detection rate varied significantly between the different groups of patients. Of the 13 classic LFS families (group 1), germline mutations were identified in 12 (92%) families. In the 184 families that fulfilled the LFL criteria (group 2), mutations were identified in 54/184 (29.35%) families. In the 281 families who did not fulfill the LFL criteria but were considered as suspected LFL or sLFL (group 3), germline pathogenic *TP53* mutation was detected in 13 (4.6%) families. Within the sLFL cohort of 281 families, *TP53* mutation were identified in 10/74 (13.5%) families in group 3A where the proband had a modified LFS associated tumor below 50 years of age. Table 3.9 describes the mutation detection rates in various groups studied in this study.

Table 3.9 *TP53* mutation detection rates in TMC LFS / LFL / sLFL cohort

Group	Subgroup	No. of families tested	No. of *TP53* mutation carrier families	Mutation detection rate (%)
Classic LFS Group 1	-	13	12	92.3%
LFL families Group 2	-	184	54	29.35.%

Suspected LFL Group 3	3A	74	10	13.5%
	3B	19	1	5.3%
	3C	188	2	1%
Others Group 4	-	22	0	0%

3.9.6 Genotype – Phenotype Correlations in Indian LFS/LFL cohort

The phenotypic clinical data of 356 cancer affected individuals from the 79 mutation carrier families was analysed for its correlation with the *TP53* genotype. The mutations were classified as DNE as previously described, other missense mutation (OMS), truncating mutations, splice site mutation and LGRs. Missense mutations which have been described to have moderate dominant negative effect in the IARC databases were included in the OMS group. Association between the type of mutation and the age at cancer diagnosis and the tumor type was examined.

Age at diagnosis

In 356 individuals with 403 tumors, information regarding age of diagnosis was known for 379 tumors. The mean age of cancer diagnosis in carriers of DNE missense mutations was 30.6 years as compared to 35 years for carriers of other missense mutations (p=0.053) and shown in figure 3.31. No difference was observed in the age of cancer diagnosis between carriers of DNE missense mutations and truncating mutations (30.6 years vs 30.9 years). The mean age of cancer diagnosis was 36.5 years for carriers of large genomic rearrangement carriers versus 30.6 years for DNE and 30.9 years for truncating mutation but the difference did not reach statistical significance.

Tumor type

Correlation between tumor type and the mutation type was made only for cases with female breast cancer (n=121) as for other cancer sites, number of cases for each mutation type were too few for statistical analysis. Breast cancer as a proportion for all cancers was highest in females

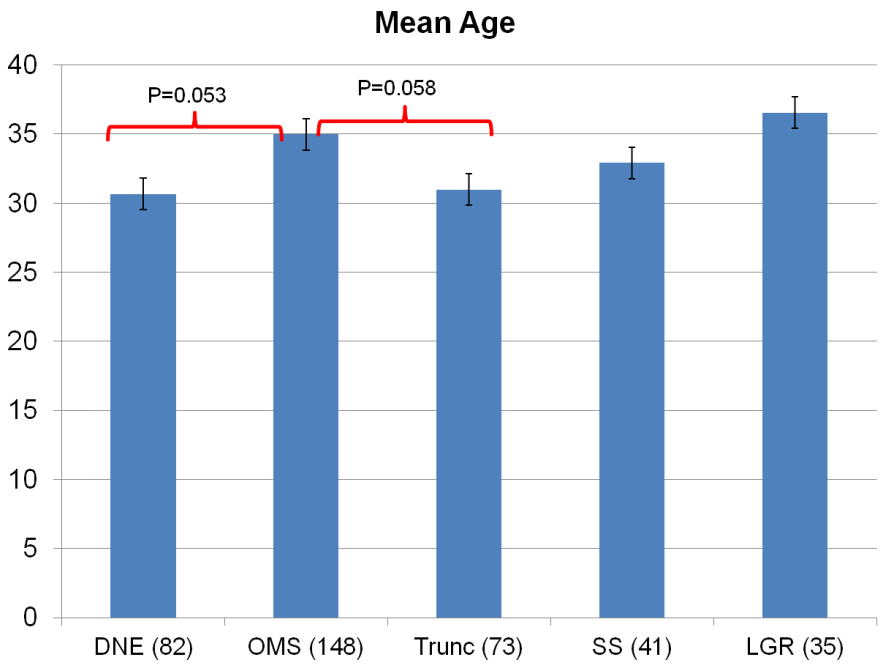


Figure 3.31 Comparison of mean age of cancer onset across all mutation types. MS-Missense; FS-Frame Shift; NS-Non Sense; SS-Splice Site; LGR- Large Genomic Rearrangement

with truncating mutation - 34/56 (60%) as compared to those with DNE – 17/43 (39.5%) and other missense mutation – 42/94 (44.6%). The differences when compared across all three groups was of borderline significance (p= 0.07) but the difference between truncating versus DNE mutation was significant (p=0.04) as shown in figure 3.32.

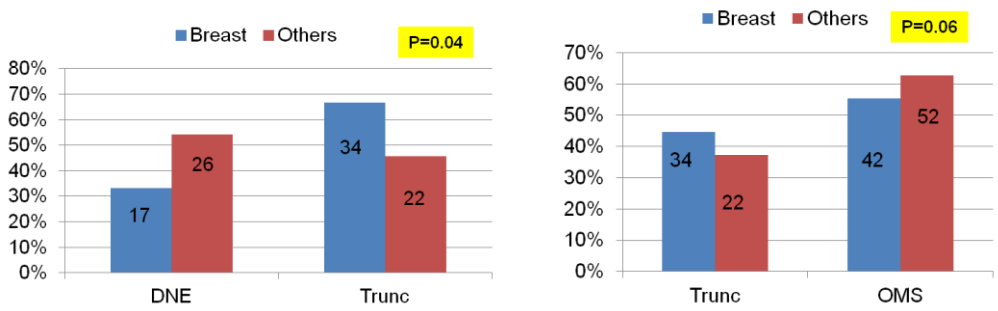


Figure 3.32 Breast cancer versus other cancers in different mutation groups

3.9.7 Association of *TP53* and MDM2 polymorphisms with tumor occurrence and age of cancer diagnosis in carriers of germline *TP53* mutations

Association of three polymorphisms (P72R, PIN3 and MDM2SNP309) was studied with respect to occurrence of cancer and age of cancer diagnosis in carriers of germline *TP53* mutation. Polymorphisms data was available for 93 mutation positive individuals for P72R and 94 individuals for the other two polymorphisms. Of these 93 individuals with P72R genotype data, 73 were affected with cancers; of the 94 individuals with PIN3 genotype data, 74 were affected with cancer and of 94 with MDM2 SNP genotype data, 72 were affected with cancer. No significant association was identified for occurrence of cancer or the age of diagnosis of cancer with the genotype of these three polymorphisms (Table 3.10 and figure 3.33).

Table 3.10: Association of polymorphisms with the occurrence of cancer in *TP53* mutation carriers

R72P SNP	CC	CG	GG	Chi-square (df=2)
Cancer affected (73)	15	36	22	0.138
Cancer unaffected (20)	7	5	8	
PIN3 Duplication	Wt	Htz	Dup	0.855
Cancer affected (74)	52	17	5	

Cancer unaffected (20)	13	5	2	
MDM2 SNP309	TT	TG	GG	0.684
Cancer affected (72)	18	39	15	
Cancer unaffected (22)	5	14	3	

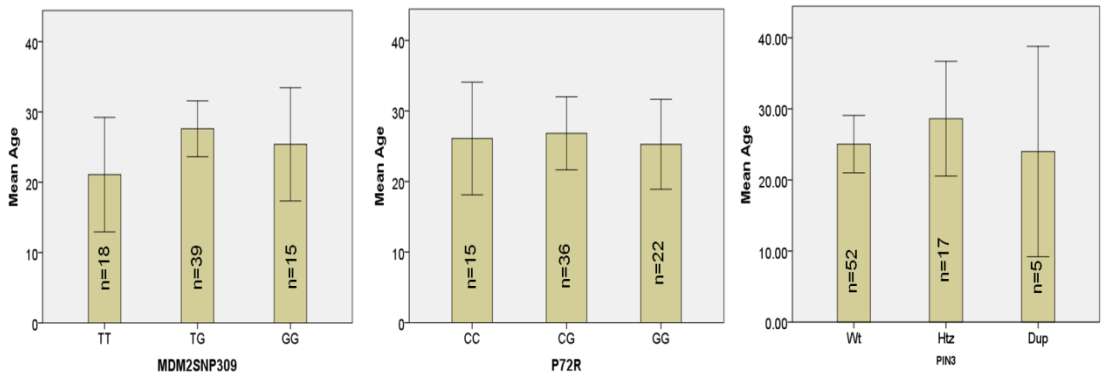


Figure 3.33: Mean age of cancer diagnosis in *TP53* mutation carriers for various SNP genotypes

3.9.8 Exome sequencing to identify driver mutation in classic LFS cases

No pathogenic *TP53* germline mutations or LGRs was identified in initial Sanger sequencing or MLPA in 3 classical LFS families (Figure 3.34 –3.36) Hence germline whole exome sequencing (WES) was performed in the probands of these 3 LFS families to identify driver mutations in other genes that could explain LFS associated cancers. WES was carried out at 30X coverage on Illumina Hi-Seq 1500 platform using Nextera Rapid Capture Exome Kit. The detailed procedure is described in the Materials and Methods section.

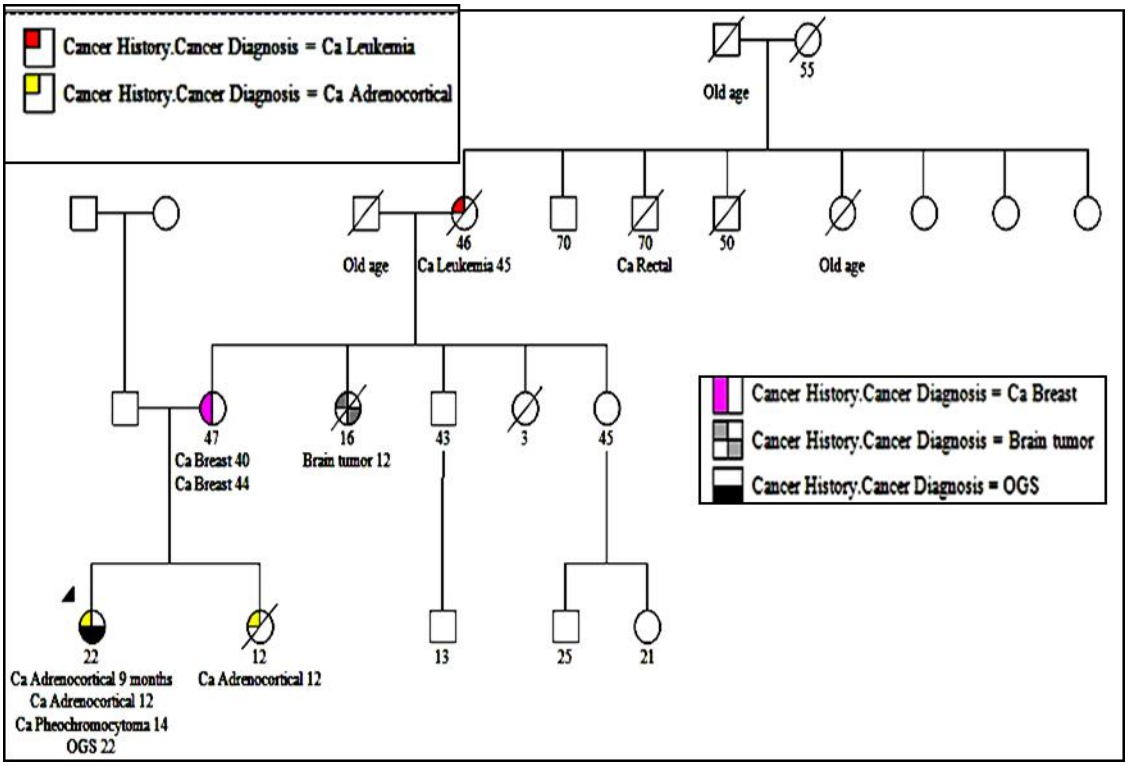


Figure 3.34 Pedigree of classic LFS family #1 (Group 1) for WES

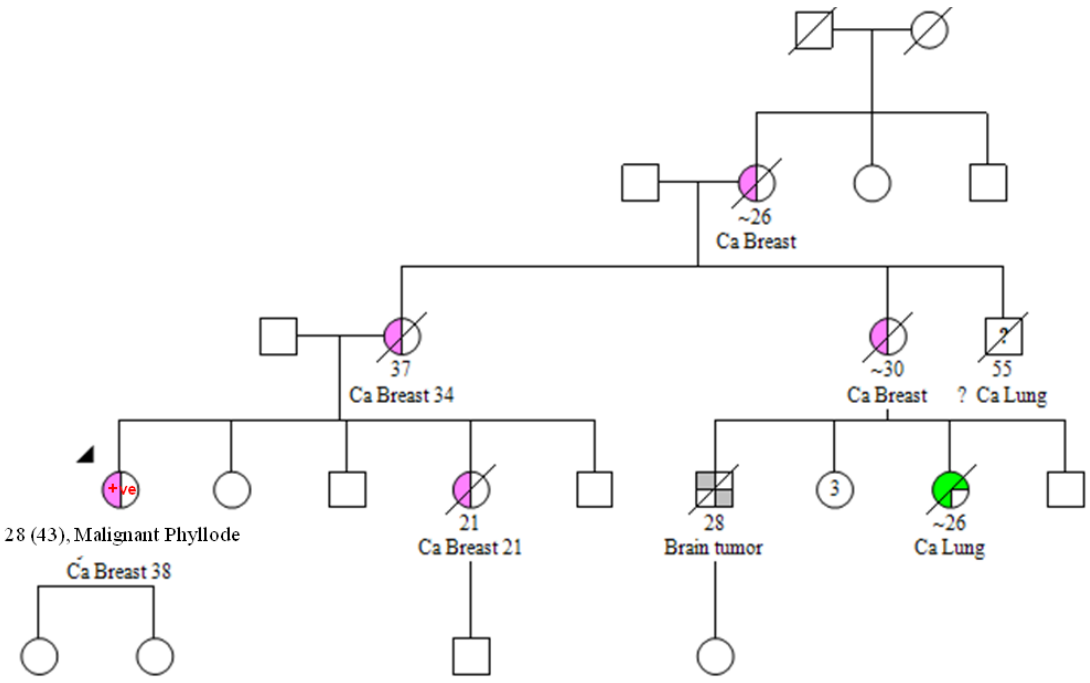


Figure 3.35 Pedigree of classic LFS family #2 (Group 1) for WES

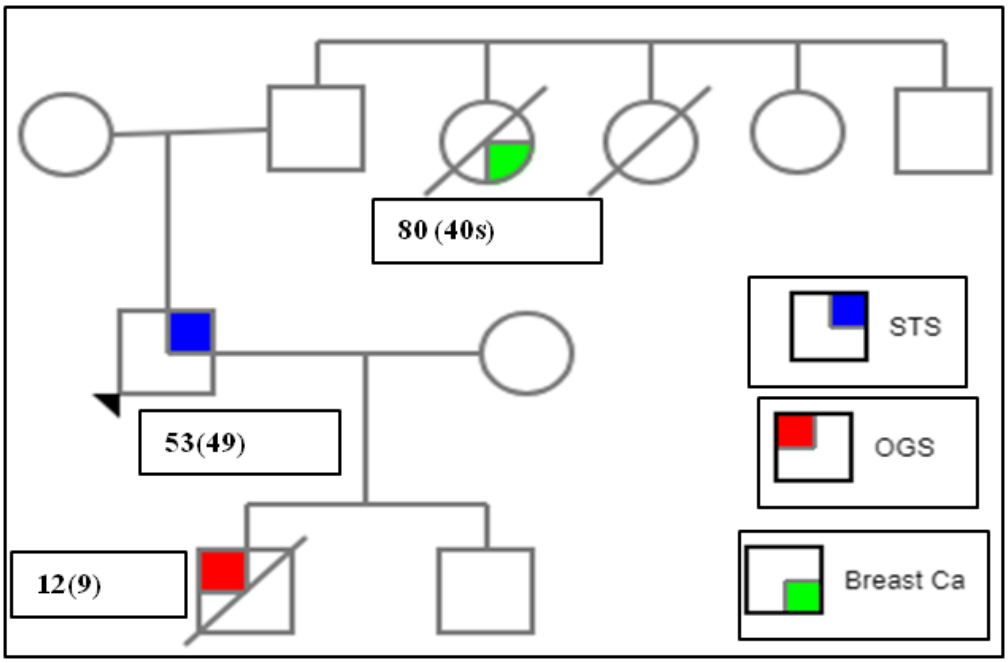


Figure 3.36 Pedigree of classic LFS family #3 (Group 1) for WES

Bio-informatic analysis using tools and pipeline described earlier in chapter 2, we identified pathogenic *TP53* mutation in two cases, indicating dropout of this mutant allele in the initial Sanger sequencing. The investigation into the causes of allele dropout and its characterization is discussed in chapter 4. The mutations identified were p.R248Q and p.Y163X. The IGV image showing the *TP53* mutations in these cases is given in Figure 3.37 and 3.38.

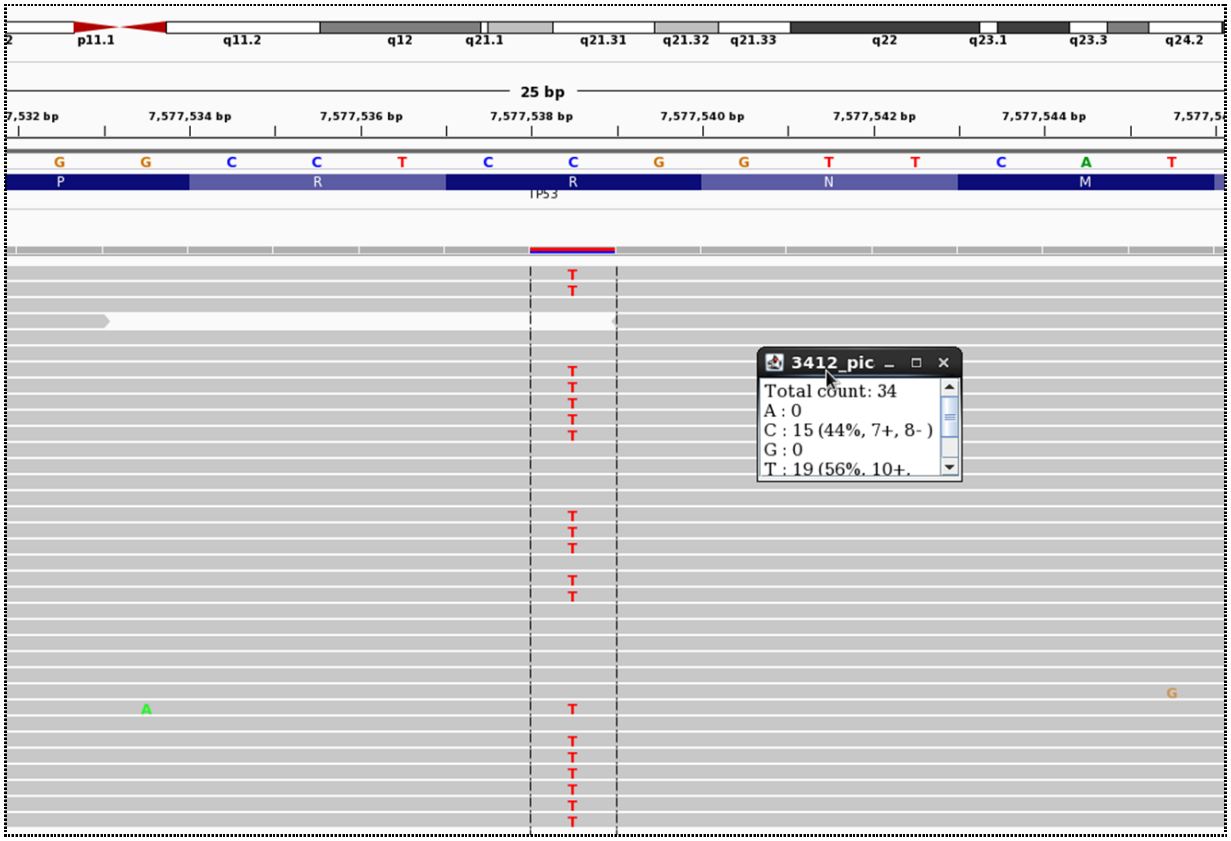


Figure 3.37 IGV image showing c.743G>A change (p.R248Q) in Exon 7 of *TP53*

In the third classic LFS family, a total of 30260 germline variants were identified in different genes but not in *TP53*. After variant annotation and filtering, we identified a total of 128 frameshift indels, 228 in-frame indels (Table 3.11) and 31 missense variants that were predicted to be deleterious by all the 7 *in-silico* prediction tools (Table 3.12). The prediction tools used were SIFT, Polyphen2, LRTpred, MutationTester, MutationAssessor, Functional Analysis through Hidden Markov Models (FATHMM) and PROVEAN.

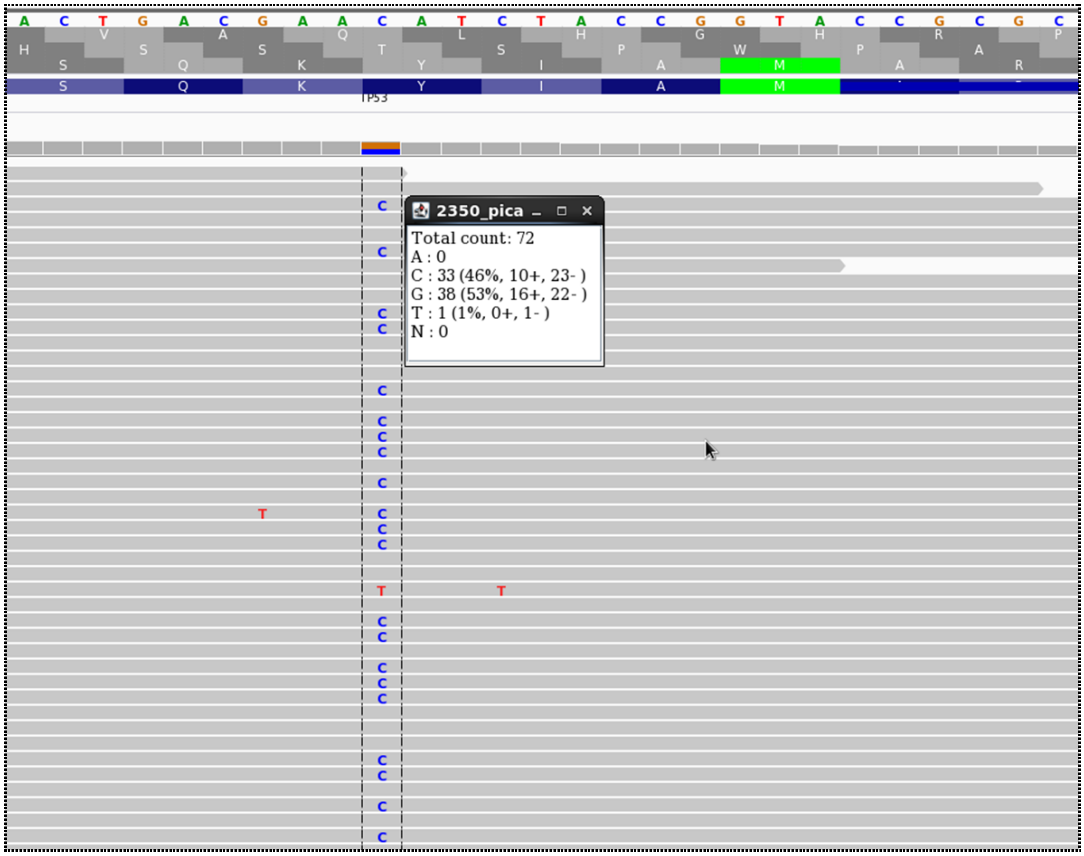


Figure 3.38 IGV image showing c.489C>G change (p.Y163*) in Exon 5 of *TP53*

Table 3.11 Variants identified through WES in the classic LFS family 3

Variants	N
Single nucleotide variants (N = 29823)	
Nonsynonymous SNVs	12266
- SNVs predicted as deleterious with at least 3 *in silico* tools	292
- SNVs predicted as deleterious with all 7 *in silico* tools	31
Indels (N = 356)	
- Frameshift deletion	67
- Frameshift insertion	61
- In-frame deletion	121
- In-frame insertion	107
Stop-gain mutations	**1**
Unknown (effect not predicted)	**80**

SNVs: Single nucleotide variants

Out of the 31 non-synonymous missense variants predicted to be deleterious by all the 7 *in-silico* prediction tools, 2 were in FGFR4 gene (table 3.12). FGFR4 has been reported to be associated with certain cancers like prostate cancer, neuroma and rhabdomyosarcoma (168-170).

Table 3.12 Nonsynonymous missense variants predicted as pathogenic by all 7 tools

Chr.	Ref	Alt	Func.refGene	Gene.refGene	ExonicFunc.refGene	Disease association
chr1	C	A	Exonic	EXTL1	nonsynonymous SNV	Hereditary Multiple Exostoses and Exostosis
chr2	G	T	Exonic	CAD	nonsynonymous SNV	Epileptic Encephalopathy, Early Infantile, 50
chr5	**G**	**T**	**Exonic**	**FGFR4**	**nonsynonymous SNV**	**Prostate Cancer, Neuroma, Rhabdomyosarcoma**
chr5	**G**	**T**	**Exonic**	**FGFR4**	**nonsynonymous SNV**	**Prostate Cancer, Neuroma, Rhabdomyosarcoma**
chr6	G	T	Exonic	ALDH5A1	nonsynonymous SNV	Succinic Semialdehyde Dehydrogenase Deficiency and Gamma-Amino Butyric Acid Metabolism Disorder
chr6	C	A	Exonic	GRM1	nonsynonymous SNV	Spinocerebellar Ataxia, Autosomal Recessive 13 and Chondromyxoid Fibroma
chr8	T	A	Exonic	ADAM18	nonsynonymous SNV	No disease reported
chr9	A	G	Exonic	AQP7	nonsynonymous SNV	Chronic Closed-Angle Glaucoma and Meniere's Disease
chr9	C	A	Exonic	LHX3	nonsynonymous SNV	Pituitary Hormone Deficiency, Combined, 3 and Lhx3-Related Combined Pituitary Hormone Deficiency.
chr12	G	C	Exonic	LRP1	nonsynonymous SNV	Keratosis Pilaris Atrophicans and Atrophoderma Vermiculata

chr12	T	C	Exonic	KMT5A	nonsynonymous SNV	No disease reported
chr14	G	C	Exonic	KLHDC2	nonsynonymous SNV	Osteochondrosis and Ischemic Bone Disease
chr17	T	C	Exonic	MYH10	nonsynonymous SNV	May-Hegglin Anomaly and Lymphangioleiomyomatosis
chr17	G	A	Exonic	KCNJ12; KCNJ18	nonsynonymous SNV	Smith-Magenis Syndrome and Andersen Syndrome
chr17	C	A	Exonic	KCNJ12; KCNJ18	nonsynonymous SNV	Smith-Magenis Syndrome and Andersen Syndrome
chr17	G	A	Exonic	KCNJ12; KCNJ18	nonsynonymous SNV	Diseases associated with KCNJ12 include Smith-Magenis Syndrome and Andersen Syndrome
chr20	C	T	Exonic	PPP1R16B	nonsynonymous SNV	No disease reported
chr22	G	A	Exonic	MLC1	nonsynonymous SNV	Megalencephalic Leukoencephalopathy With Subcortical Cysts and Mlc1-Related Megalencephalic Leukoencephalopathy With Subcortical Cysts
chrX	C	A	Exonic	ATP11C	nonsynonymous SNV	Congenital Hemolytic Anemia.

3.10 Discussion

We report the first Indian cohort and the largest Asian LFS/LFL cohort of 79 families with germline pathogenic or likely pathogenic *TP53* mutation. Based on detailed phenotypic characterization, pedigree analysis and comprehensive genetic analysis including LGR analysis we could draw important conclusions. This large cohort provides important insight on the LFS/LFL genotype–phenotype correlation in the South Asian population, which has not been studied earlier. With a wide referral base of rare cases to our apex cancer centre, these 79 mutation positive families hail from

different regions and communities of the Indian sub-continent. Thus, our cohort is fairly representative of the south Asian population.

The tumor spectrum in our cohort was different from other populations reported in the IARC database, especially for ACC, sarcoma, skin neoplasms, breast cancers and head & neck cancers. ACC is a rare paediatric cancer, the age standardized population incidence of adrenocortical cancer is not described in the Indian population. However, the lower incidence of ACC in our germline TP53 cohort may be true as the Brazilian Founder mutation (R337H) which has the highest incidence of ACC was not identified in our cohort. Head and neck cancer incidence is indeed higher in the Indian population and when we compared incidence of H&N cancer in TP53 mutation positive families irrespective of mutation confirmation in each case, we noted significantly higher incidence of H&N cancer in our Indian cohort as compared to the IARC global cohort (table 3.5, page 125). However when we analysed cancer spectrum only in confirmed mutation carriers and obligate carriers, the incidence of H&N cancers in our cohort is not significantly different from IARC global cohort (shown in figure 3.9 and 3.10, page 124).

The *TP53* mutation spectrum in our cohort is also significantly different from other populations in certain aspects. In our cohort, missense mutations accounted for 59.5% of all unique mutations identified, which is significantly lower than the global database (IARC R19) (Figure 3.15). The 8.8% frequency of LGRs in our study is also significantly higher than the 1.1% frequency in the combined global cohort as shown in figure 3.15 (IARC R19). This is unlikely to be a true difference but due to infrequent use of LGR analysis in earlier studies. This highlights the importance of LGR analysis as part of comprehensive *TP53* genetic analysis. It is important to note

that Next Generation Sequencing (NGS) multigene panels may not identify or report *TP53* LGRs. In a large French cohort reported recently, LGRs accounted for 9/133 distinct mutations (82, 88). Hence in LFS/LFL families in whom NGS panel testing has not identified any pathogenic or likely pathogenic mutation in a cancer associated gene, *TP53* LGR analysis should be done.

While we recommend comprehensive *TP53* genetic analysis, the reality of resource constraint setting of developing world and access to genetic testing cannot be ignored. Not surprisingly, only 7 *TP53* mutation positive families from south Asia and Africa are reported in the large IARC R19 database. In such settings, all families fulfilling the LFS or LFL criteria should have sequencing for at least the 4 exons of the DNA binding domain which is expected to identify 75-85% of all *TP53* mutations (84).

Another striking finding of our study was a very high rate of novel germline pathogenic or likely pathogenic *TP53* mutations. The IARC R19 release has 456 distinct germline *TP53* mutations but 24/59 distinct germline mutations identified in this Indian cohort have never been reported as germline mutation in any study or database. This underscores the need for studying each population to expand the germline *TP53* mutation spectrum. The known *TP53* germline hotspot mutations were also observed in our cohort and accounted for 24% of all *TP53* positive families in our cohort. The frequency of the hotspot mutations is in accordance with published studies (144). As expected, the R337H hotspot mutation was not identified in our cohort as it is a Brazilian founder mutation (171).

The significantly younger age of cancer diagnosis in carriers of dominant negative missense mutations in our cohort is in accordance with findings from a French LFS cohort (88). But in our cohort the age of diagnosis is almost same for DNE mutations

and truncating mutations and the age of diagnosis in both DNE and truncating mutations are less than other missense mutations with borderline significance. However earlier studies have shown that the age of diagnosis is earlier in missense mutations and DNE missense mutations with respect to truncating mutations (88).

Our observation of significantly higher proportion of breast cancers in females carriers of truncating *TP53* mutations as compared to DNE missense mutations has not been previously reported. The significant differences in the mutation spectrum and genotype-phenotype correlation between the Indian cohort with other population brings out the need for clinical and genetic characterization of this syndrome in every population. This could help in developing population specific genetic counselling, genetic testing and surveillance guidelines for *TP53* mutation carriers. No significant association was identified with the various polymorphisms along with the *TP53* mutation in cancer occurrence and at age of diagnosis of cancer.

The very high mutation detection rate of ~90% in classical LFS families and ~29.35% in LFL families in our cohort reflects the detailed phenotypic characterization and family information used in our cohort for making syndromic diagnosis. The mutation detection rate in our study is in accordance with previous studies which showed high mutation detection rates of around 70% in classic LFS cases and 20-40% in LFL cases (87).

Our cohort also included 303 patients who could not be classified as LFL by the existing criteria but were categorized as suspected LFL (sLFL-281) and familial cancers (22) as described in page 68 and subjected to comprehensive *TP53* genetic analysis. While the overall mutation detection rate in the sLFL group was 13/281 (4.6%), in the group 3A of 74 probands with a modified LFS spectrum tumor below

50 years of age and family history of any cancer at any age, *TP53* mutation was identified in 10/74 (13.5%) cases. These cases do not fulfill the classical LFS or the existing LFL Birch, Eeles or Chompret criteria but were Suspected Li Fraumeni Like (sLFL) cases if proband has early onset (<50 yrs) *Modified LFS spectrum cancer with 1st, 2nd or 3rd degree relative with any cancer at any age OR proband has multiple primary cancers of which at least one is a modified LFS spectrum cancer <50 years. *Modified LFS spectrum cancers: Includes *Haemato-Lymphoid malignancies and *Malignant Phyllodes in addition to the previously described LFS spectrum cancers - ACC, Sarcoma, CNS, Leukemia and Breast cancer. By expanding the LFL spectrum cancers to include Malignant Phyllodes and any Haematolymphoid malignancy and by relaxing the age cut off from <46 years (Chompret) to <50 years, we could identify TP53 mutation in 13.5% cases who would have been missed even by the relaxed LFL Eeles and Chompret 2015 criteria.

While this needs to be confirmed in larger independent cohorts, previous studies by Birch et al have shown that in *TP53* mutation positive families, the greatest increase in the risk relative to the general population was for Malignant Phyllodes ((136)). In the same study it was found that all other LFS spectrum cancer were strongly associated but Leukemia when taken alone was only weakly associated. It is possible that if all haematolymphoid malignancies are considered as one entity a stronger association may be seen. Moreover, LFS spectrum cancers is not limited to acute leukemia and there is overlap between diagnostic features and management of some chronic leukemias and lymphomas e.g. Chronic Lymphocytic Leukemia and Small Lymphocytic Lymphoma.

In suspected LFL category 3B, 3C and Group 4 (familial cancers not fulfilling criteria of any hereditary cancer syndrome) the TP53 mutation detection rate was only 5.3%, 1% and 0% respectively. Hence we do not recommend these to be included in modified LFL criteria. Based on such high mutation detection rate in this group, we suggest that the current LFL criteria could be revised to include patients with modified LFS spectrum tumor below 50 years of age along with a family history of any cancer at any age. Based on our data we also suggest that similar to other rare tumors such as adrenocortical carcinoma and choroid plexus tumor, malignant phyllodes under the age of 50 years can also be considered as LFS associated cancer and acriteria for defining LFL. This recommendation is suggested based on identification of *TP53* mutation in 5/9 cases with malignant phyllodes of breast in our cohort and all were <30 years of age. The IARC R19 release has 14 cases of phyllodes tumor of the breast with germline *TP53* mutation and all cases for whom the age information was available were <30 years of age.

All the *TP53* LGRs identified in our study originated from intron 1 which is around 10kb long. Intron 1 of *TP53* is reported to be the hot spot for LGRs. Earlier reports have shown frequent occurrence of LGR originating from *TP53* intron 1 in sporadic osteosarcoma (172, 173). Similar to the high frequency of LGR in our study, Bougeard et al in 2015 also reported that LGRs accounted for 9/133 (6.8%) of all *TP53* mutations in their cohort (88).

We identified germline *TP53* mutations in only 2/188 cases with sporadic sarcoma in our cohort. Similar to our findings, Mitchell et al (2013) identified *TP53* mutation in only 7/465 (1.5%) sporadic adult onset sarcomas (99). Diller et al in 1995 identified germline *TP53* mutation in 3/33 (10%) sporadic sarcoma cases but these were all

children and the sarcoma was rhabdomyosarcoma (166). These findings suggest that in sporadic adult onset sarcoma, routine *TP53* genetic analysis may not be required in low income or developing countries.

After a recent report of LFS associated variant in the 3'UTR region of *TP53* (174), we also screened this region but did not identify this 3'UTR variant in the 120 LFS/LFL cases we screened. Instead, another variant, 15bp downstream of the reported variant was identified. While it is possible that the 3'UTR variant we identified may interfere with the transcription of *TP53* gene, as its functional effect was not studied, we consider it as a VUS.

Germline exome sequencing in 3 classic LFS cases without an identified *TP53* mutation on initial round of Sanger sequencing led to the serendipitous finding of *TP53* allele dropout in 2 cases. The nature of this genotyping error and the corrective actions are discussed in chapter 4. In the third case with classic LFS, exome sequencing did not identify any *TP53* mutation but variants in certain other genes that have been previously reported to be associated with sarcoma (Table 3.12). This includes 2 missense variants in FGFR4 gene which were predicted to be pathogenic by all 7 in silico tools (SIFT, Polyphen2, LRTpred, MutationTester, MutationAssessor, FATHMM and PROVEAN). Activating somatic mutation in FGFR4 gene has been reported earlier in rhabdomyosarcoma (170) and merits further characterization as a germline inherited sarcoma susceptibility gene.

Table 3.13 Nonsynonymous missense variants predicted as pathogenic by 2 or more tools in genes reported previously to be associated with sarcoma predisposition

Reported predisposing genes	Function of the gene	Associated Sarcoma	Related genes identified	Function of the identified gene
RECQLA	DNA repair	Osteosarcoma	REC8 (exon8:c.681_682insGAA)	Structural maintenance of Chromosome
			CDKN2AIP (exon3:c.722_724del)	DNA repair
HRAS	GTPase, signal transduction	RMS	RASGEF1C (exon3:c.181G>T; p.A61S)	Activates HRAS, KRAS, NRAS
NSD1	Chromatin organization	RMS	NSD1 (exon5:c.2176T>C; p.S726P)	Chromatin organization
APC	WNT signaling pathway	Desmoids	APC (exon14:c.5411T>A; p.V1804D)	WNT signaling pathway
ERBB3	RTK; signal transduction, cell proliferation	Leiomyosarcoma	ERBB3 (exon27:c.A3355T:p.S1119C)	RTK; signal transduction, cell proliferation

A lower threshold of considering variants as pathogenic even if only two of the 7 in-silico prediction tools predicted them to be deleterious, identified 5 other genes which are associated with sarcoma (table 3.13). There is need to validate the findings of exome sequencing in a larger cohort of *TP53* negative sarcoma cases and undertake functional characterization.

Chapter 4

Identification and characterization of *TP53* gene Allele Dropout in Li-Fraumeni syndrome cohort

4.1 Allele Drop Out

Accurate molecular genetic analysis is important in research setting as well as in clinical genetic testing. Genotyping errors in germline or somatic mutation testing could have major clinical consequences for cancer patients and their families. Genetic analysis should be highly sensitive as well as specific. False negative or false positive results due to genotyping errors may occur due to several reasons. These include (i) Pre-Analytical errors due to sample mix up (175), poor quality or degraded DNA; (ii) Analytical error arising from lack of standardization or validation of the assay, use of less sensitive or specific assay, variability in the samples, and rarely due to Allele Drop Out; (iii) Post analytical errors of incorrect interpretation arising from human interpretative errors or use of inappropriate bioinformatics approaches or *in-silico* tools or databases. Almost all genotyping studies are based on Polymerase Chain Reaction (PCR) for initial amplification of the target region of DNA followed by Sanger sequencing, RFLP or other genotyping assays. In spite of its robustness, PCR has its own limitations which may lead to genotyping errors. Sometimes one of the allele of dsDNA does not get amplified either due to sequence dependent reasons or sequence independent reasons. This phenomenon where one of the allele of target region is not amplified and not detected on sequencing is known as Allele Drop Out (ADO).

There are several known experimental and other factors which may cause ADO and summarized in table 4.1. These include the DNA sequence, sample quality, reagents, equipment and human factors (111). In current clinical molecular diagnostics, the pre and post analytical errors have been greatly reduced with good laboratory practices and accreditation of laboratories. The analytical errors have also been reduced with

careful design and validation of genotyping assays and external quality assurance (QA) programme (110). However, ADO remains an important analytical error in genotyping. ADO arises from insufficient amplification of one of the two alleles and the dropped allele remains below the detection threshold of sequencing. Dropout of the mutant allele causes false negative result while dropout of the Wild Type (WT) allele makes a heterozygous mutation appear homozygous. The sequence dependent ADO occurs due to certain features within the sequence of the DNA being amplified. These include variants in the annealing region of the primers (176-178), presence of tertiary structures like G-Quadruplexes and i-motifs, (179, 180) methylation (180) or allele size differences (181). The sequence independent ADOs arise from poor DNA quality as in forensics (111), Whole Genome Amplification of scanty starting DNA as used in Single Cell Sequencing or Preimplantation Genetic Diagnosis (182) and from unknown PCR conditions (181, 183).

Table 4.1 Factors and causes of ADO

Factors		**Cause**	**Mechanism**
Sequence dependent	Primer Annealing Region (PAR)	SNP or variant in PAR	Prevents annealing of the primer and subsequent amplification of one allele
	DNA sequence	Heterozygous insertion or deletion	Different size of allele may hamper amplification of one allele
		Different GC content in heterozygous	Less amplification of higher GC content allele due to less efficient denaturation.
Sequence	Sample quality	Low quality or quantity of DNA	Amplification of shorter allele or only one allele

independent		Contamination	Amplification of contaminant allele
	Reagents	Buffer, Taq, equipment etc	Due to presence of inhibitor or Taq slippage only one allele is amplified
	Human Errors	Mislabeling Experimental error	All these leads to genotyping errors

ADOs as a cause of incorrect genotyping has been highlighted in diverse molecular diagnostic contexts (110, 178, 183-185), but have not been systematically evaluated in oncology. It assumes greater importance in oncology as genetic analysis is being increasingly used for prognostication, precision medicine, hereditary risk assessment and cancer prevention. In the first systematic study of ADO in any cancer related gene, we have examined *TP53* gene which harbours a large number of well annotated germline and somatic mutations that are catalogued in the IARC *TP53* database. *TP53* is the most frequently mutated gene in diverse cancer tissues (186) and germline mutation in *TP53* is responsible for LFS or LFL syndrome (85)

4.2 Method

The genomic DNA was extracted from peripheral blood lymphocytes by Qiagen columns (QIAamp DNA Blood Mini Kit; Cataloque number 51106) according to manufacturer's protocol or in some cases by phenol chloroform method as described in chapter 2. The entire coding region of the *TP53* gene was sequenced by Sanger sequencing. If no germline *TP53* mutation was identified, Large Genomic Rearrangement analysis was done by the Multiplex Ligation dependent Probe Amplification (MLPA) kit (MRC Holland) as per manufacturer's instructions. In

selected LFS/LFL cases without an identified *TP53* mutation on Sanger Sequencing and MLPA, germline exome sequencing was done (n=3). Detailed methodology is given in chapter-2

4.3 Results:

4.3.1 Discovery Set:

Germline *TP53* mutations were tested in a cohort of 500 families. Of these, 197 families fulfilled the defined criteria of LFS or LFL(87), while the remaining 303 families did not fulfil the criteria for LFS or LFL but were tested as either the proband had an LFS associated cancer below 50 years with any cancer in family (281/303) or the family had various cancers which do not fit into any criteria (22/303) (Table 3.7). A total of 79 probands in this *TP53* tested cohort were found to carry a germline heterozygous mutation in *TP53.* In two classical LFS families germline whole exome sequencing was done as previous Sanger Sequencing and MLPA had not identified any *TP53* mutation. In both these cases, a deleterious *TP53* germline mutation was identified on whole exome sequencing and later confirmed as ADO on repeat Sanger Sequencing (figure 4.1 #G1-2). One of the cases had mutation in exon 5, missed earlier due to low peak height and the other had a mutation in exon 7 that was detected by redesigned primers.

In four cases (#G3–6), ADO was suspected due to mutation homozygosity and confirmed on repeat sequencing with redesigned primers for exon 7 to avoid a common polymorphism as shown in figure 4.2. Therefore a total of 6 cases in the

discovery set suspected of ADO were confirmed to have a heterozygous mutation on Sanger Sequencing.

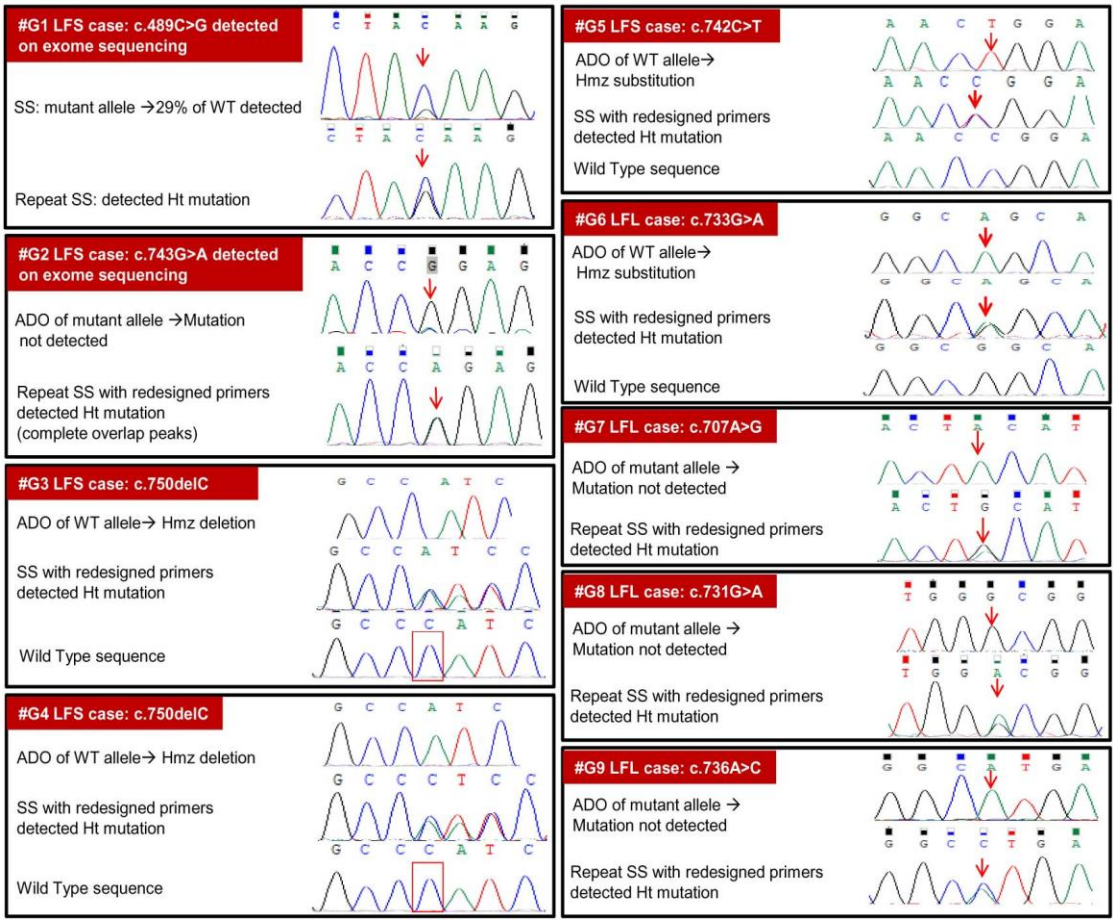


Figure 4.1 Germline *TP53* ADO in nine LFS/LFL cases: #G-Germline case number and syndromic diagnosis. #G1-6: ADO in discovery Set; #G7-9: ADO found in validation cohort; WT-Wild Type; Ht-Heterozygous; Hmz-Homozygous; LFS-Li Fraumeni Syndrome; LFL-Li Fraumeni like Syndrome; ADO-Allele Drop Out; SS-Sanger Sequencing. #G1- Exon 5 and #G2-G9- Exon 7.

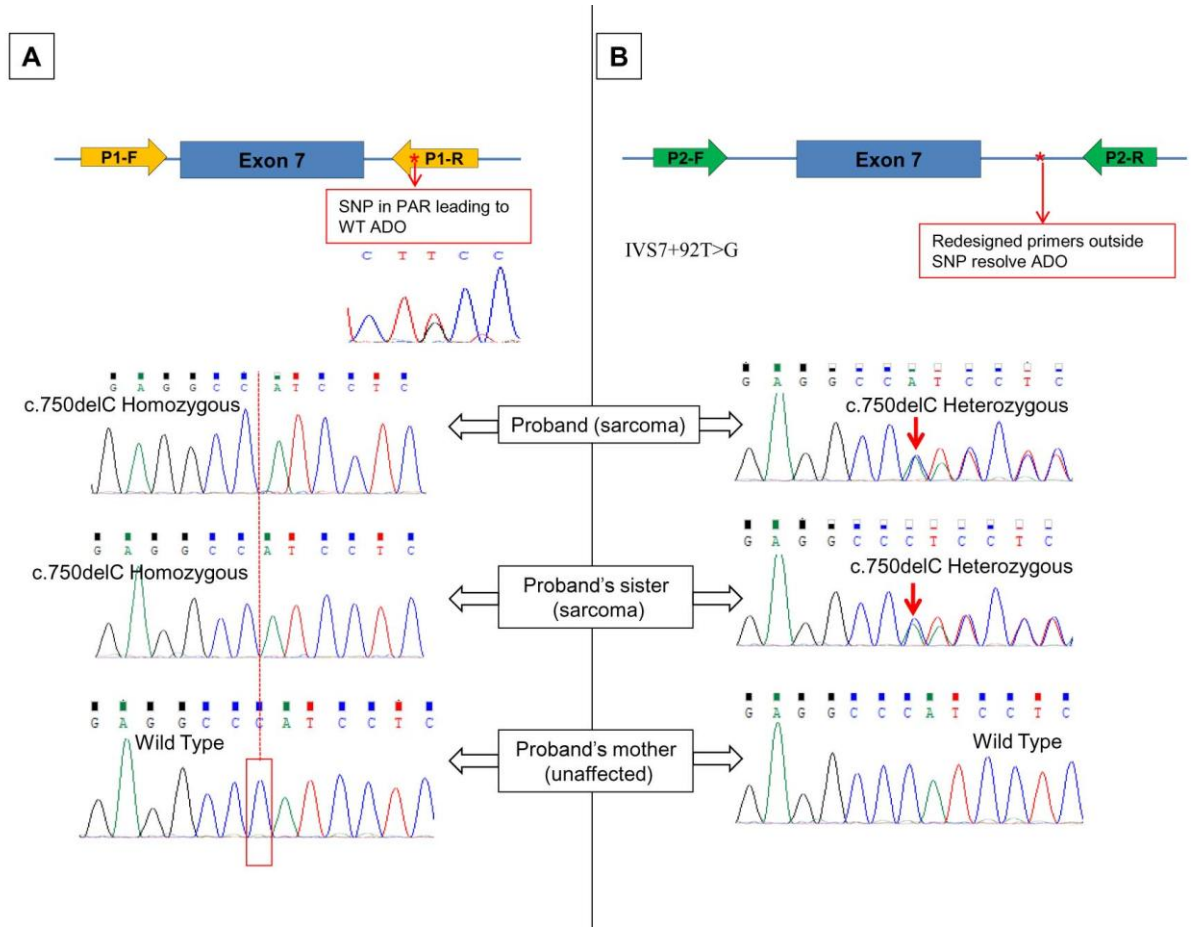


Figure 4.2 *TP53* Germline Allele dropout in an LFS family due to annealing region SNP: (A) Homozygous deletion in two sarcoma affected siblings. Mother unaffected and WT. Father who died of cancer was not tested. ADO of WT allele suspected and SNP (rs12951053) detected in the reverse primer (P1R) binding region. Chromatogram of the SNP is shown. The mutant allele is in the trans to the T allele of IVS7+92T>G polymorphism and cis to the G allele.(B) Resequencing with redesigned primers (P2F and P2R) avoiding the SNP region confirmed ADO and detected heterozygous deletion mutation in the two siblings.

4.3.2 Validation cohort: This consisted of 150 cases fulfilling the defined criteria of LFS or LFL in whom *TP53* full gene Sanger sequencing and MLPA had not identified any pathogenic mutation. In these 150 cases, chromatograms of the initial Sanger sequencing were read again to identify any possible variant that was not called earlier because the variant peak height ratio was <0.3 or it was >0.3 with background noise. Reanalysis of the chromatograms identified 28 such suspected variants. However on

repeat Sanger sequencing using same conditions, none of these suspect variants could be detected and were considered to be artefacts (Figure 4.3).

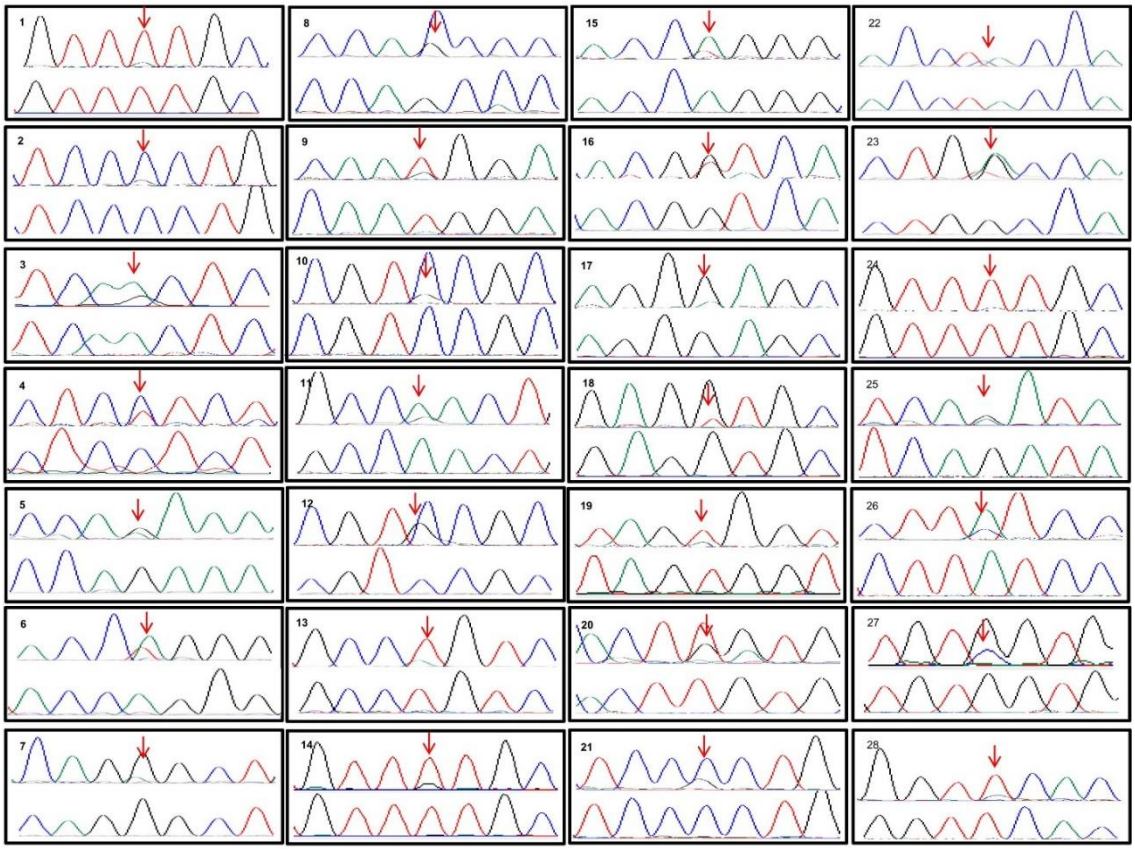


Figure 4.3: Suspected variants (N=28) in Sanger Sequencing in various exons of *TP53:* Repeat Sanger sequencing was done using same conditions. None of these suspected variants could be detected and considered as artifacts.

For exon 7, sequencing was repeated with redesigned primers in these 150 cases. Three additional germline ADOs (#G7-9) were identified. For exon 4-9 which has the DNA Binding Domain and harbours about 85% of all germline *TP53* mutations (146), sequencing was repeated using same primers and conditions in 50 cases from the validation cohort. No additional ADO was identified in these 50 cases. A total of 9 germline ADO were detected, 6 from discovery set and 3 from validation set. The detail of all the cases in which ADO was identified is given in table 4.2.

Table 4.2: The details of all the cases in which ADO was identified				
Case No.	**Clinical details**	**Family history (Syndromic diagnosis)**	**Reason for suspecting Allele dropout on SS**	**Allele dropout detected (method)**
1	22 year female with 4 primaries at 9 months, 12 years, 14 years and 22 years	Family history of breast, brain and blood cancer	*TP53* Sanger sequencing of full gene and MLPA negative in a classical LFS case. Later mutation detected (p.R248Q; c.743G>A)	Exome sequencing to identify any other gene mutation detected allele dropout of the *TP53* exon 7 mutant allele and reconfirmed by SS using different primers
2	43 Year female with two primaries, Sarcoma and Breast cancer at 22 and 38 years	Seven family members affected with breast, lung and brain cancers	*TP53* Sanger sequencing of full gene and MLPA negative in a classical LFS case.	Exome sequencing to identify any other gene mutation detected allele dropout of the *TP53* exon 5 mutant allele and reconfirmed by SS
3	20 year male with Osteosarcoma at 20 years.	Family history of Sarcoma and stomach cancer. Classic LFS	Homozygous deletion mutation (p.P250delCfs353*; c.750delCfs353*) and mother (37) not carrier of this variant	Repeat SS with different primers revealed that an Exon 7 WT allele was dropped earlier.
4	15 year female With	Family history of Sarcoma and	Homozygous deletion mutation (p.P250delCfs353*;	Repeat SS with different primers revealed that the WT

	Osteosarcoma at 13 years.	stomach cancer. Classic LFS	c.750delCfs353*) and mother not carrier of this variant	allele was dropped earlier.
5	17 year Male with Lipoblastoma at 17 years.	Family history of brain tumor.	Homozygous missense mutation (p.R248W; c.742C>T)	Repeat SS with different primers revealed that the WT allele was dropped earlier.
6	7 year male with RMS at 18 months.	No Family history	Homozygous missense mutation (p.G245S; c.733G>A)	Repeat SS with different primers revealed that the WT allele was dropped earlier.
7	30 years Female with Breast cancer at 26 and 30 years	11 members in family affected with various cancers like breast cancer, brain tumor etc	Resequencing of exon 7 with different primers in a cohort of 150 LFS/LFL cases in whom no *TP53* mutation was identified earlier on full gene SS & MLPA. Allele dropout of *TP53* Exon 7 mutant allele (p.Y236C; c.707A>G) detected	
8	55 year Male with two primaries sarcoma & Lung cancer at 53 and 55 years	10 members affected with various cancers like Breast cancer, leukemia, brian, uterine, etc	Resequencing of exon 7 with different primers in a cohort of 150 LFS/LFL cases in whom no *TP53* mutation was identified earlier on full gene SS MLPA. Allele dropout of *TP53* Exon 7 mutant allele (p.G244D;c.731G>A) detected	
9	44 year	8 members	Resequencing of exon 7 with different	

	female with breast cancer at 44 years	affected with various cancers like Lung, Breast, Bone cancers etc.	primers in a cohort of 150 LFS/LFL cases in whom no *TP53* mutation was identified earlier on full gene SS MLPA. Allele dropout of *TP53* Exon 7 mutant allele (p.M246L; c.736A>C) detected

4.3.3 Polymorphisms in *TP53* gene and primer annealing region

In the dbSNP polymorphism database (NCBI), the *TP53* gene has 6148 polymorphisms in *Homo sapiens*. After specifying the variation class to SNP, a total of 5311 SNP were listed.

In the annealing region of the Bodmer group primer set that were used in our study, 58 polymorphic sites were identified. High minor allele frequency (MAF) of >0.01 was noted for one polymorphism in reverse primer of exon 7. In comparison, the annealing region of the most commonly used IARC protocol primer set (http://p53.iarc.fr/download/*TP53*_directsequencing_iarc.pdf, accessed on 12th Mar 2018) harbours 83 polymorphisms and 4 of these have a MAF of >0.01 and affect five primers as shown in Table 4.3.

Table 4.3: Comparison of polymorphisms in primer annealing region of Bodmer group Primers and IARC primers and their MAFs			
		Primers used in the study (Bodmer group primers)	IARC primers
		Number of Polymorphisms in PAR (MAF range)^	Number of polymorphisms in PAR (MAF range)^
1F		1 (0.0007)	--
1R		3 (0.00003)	--
2F		2 (0.00003-0.0001)	--
2R		7 (0.00003-0.0002)	--
2-3F		--	2 (0.00002-0.00003)
2-3R		--	7 (0.000008-0.0853*) rs17883323
3+4F		3 (0.000008-0.0012)	--
3+4R		2 (0.00003)	--
4F		--	7 (0.000008-0.0113*) rs35117667
4R		--	2 (0.000008-0.0002)
2nd set	4F	--	7 (0.000008-0.0853*) rs17883323
	4R	--	1 (0.0001)
5F		--	4 (0.000009-0.0002)
5R		--	6 (0.000008-0.0002)
5+6F		2 (0.0018-0.0073)	4 (0.00001-0.00003)
5+6R		10 (0.00001-0.0034)	8 0.000009-0.0088)
6F		--	3 (0.00002)

6R		--	8 (0.000009-0.0088)
7F		4 (0.0001-0.0002)	3 (0.000008-0.0002)
7R		3 (0.0001-0.178*) rs12951053	2 (0.00003)
2nd set	7F	--	3 (0.00002-0.0127*) rs17880604
	7R	--	5 (0.000008-0.0025)
8F		--	4 (0.00002-0.0013)
8R		--	3 (0.000008-0.0002)
8-9F		2 (0.0002)	1 (0.0002)
8-9R		2 (0.000008-0.00002)	3 (0.000008-0.0002)
9F		--	5 (0.000008-0.0002)
9R		--	3 (0.00002)
10F		6 (0.00001-0.0002)	0
10R		6 (0.00003-0.00007)	1 (0.0016)
11F		1 (0.0002)	3 (0.00002-0.0112*) rs17881850
11R		4 (0.00007-0.0001)	3 (0.0002-0.0014)
Total		58	98

Note: * indicates the polymorphisms with MAF > 0.01

4.3.4 G-quadruplexes and methylation in *TP53* gene: The full *TP53* gene has 5931 regions including overlaps and 120 without overlaps G-rich sequences which can form quadruplex. The range of QGRS score for these 5913 regions is 0-61. The coding DNA sequence of *TP53* gene has 192 regions including overlaps and 5 without overlaps having QGRS score ranging from 3-33. G-rich sequences in the different amplicons of *TP53* gene is described in the table 4.4. The maximum QGRS

score for a 30bp long sequence can be achieved upto 105 (116). No CpG islands were found in any amplicon of *TP53* gene except exon 1. Exon 1 being the promoter region has the CpG island.

Table 4.4: G-quadruplex in *TP53* gene					
		Without overlaps	**QGRS score range**	**Including overlaps**	**QGRS score range**
Full gene (19149 ntds)		120	5-61	5931	0-61
CDS (1182 ntds)		5	3-33	192	3-33
Amplicon	1 (439ntds)	2	18	75	2-18
	2 (319ntds)	3	16-21	122	1-21
	3+4 (651ntds)	4	8-40	2482	0-40
	5+6 (550ntds)	4	11-41	322	0-41
	7 (283ntds)	1	14	2	14
	8+9 (455ntds)	2	16	15	10-16
	10 (348ntds)	2	18-33	200	5-33
	11 (476ntds)	2	17-21	100	0-21

Table 4.4: G-Quadruplex in *TP53* gene: Calculated through Quadruplex forming G-Rich Sequence (QGRS) Mapper (http://bioinformatics.ramapo.edu/QGRS/index.php). Amplicons according to Bodmer group primers. CDS- Coding DNA Sequence, ntds- nucleotides.

4.4 Discussion

The risk of wrong genotyping due to ADO exists in all amplification based genotyping methods. These include Fragment Length Polymorphism analysis, Sanger sequencing and NGS based assays which are done after target amplification (187). In

several fields such as forensics, population genetics, molecular ecology, single cell sequencing and prenatal genetic diagnosis, the possibility of misgenotyping due to ADO is always considered and remedial actions are taken to the extent possible. In forensics and molecular ecology, the risk of ADO is due to the poor DNA quality or trace amounts of DNA with possible admixture of DNA from other individuals. In forensics there is a debate on the vagaries of Low Copy Number typing and its implications (188, 189). Ewens et al have proposed three main strategies to identify genotyping errors - checking for Mendelian inheritance discrepancies, replicate analysis of samples and independent allele calling (190). Using this strategy in the first systematic study of germline ADO in a cancer related gene, we report that ADOs resulted in missing ~7% pathogenic *TP53* germline mutations in Li Fraumeni syndrome cohort.

In contrast to several systematic studies of somatic ADOs in oncogenes or tumor suppressor genes (185, 191-193), very scanty information exists regarding germline ADOs in genes responsible for hereditary cancers. Worldwide, a million or more cancer patients would have undergone germline genetic testing in the last two decades. However our systematic literature review could identify only 7 cases of germline ADO in any cancer predisposing gene (176, 178, 194, 195). This extreme rarity of reported germline ADOs in cancer genes is possibly a reflection that germline ADOs are usually not suspected and consequently not not confirmed and reported. It is important to note that the seemingly high 7% false negative rate for germline *TP53* mutation in our study was established only through a systematic ADO evaluation in a large cohort of the monogenetic LFS/LFL syndrome and the presence of an exon 7 PAR SNP (IVS7+92T>G) with a high MAF of 0.16 in our cohort as discussed later (Table 4.5). More importantly, these ADO would have remained

unnoticed without the serendipitous finding of homozygous deleterious *TP53* mutations in few families and Mendelian inheritance discrepancy in one family (Figure 4.2).

Table 4.5: Minor Allele Frequency of IVS7+92T>G (rs12951053) in TMC cohort (N=112)

Genotype	**Count**		**Allele**	**Calculation**	**MAF**
TT	80		T	80*2+28=188	
TG	28		G	4*2+28=36	0.16
GG	4		Total	224	

So far only two studies have systematically examined germline ADOs in molecular diagnostics (110, 183). In the multi-centre eMERGE-PGx study, genotyping errors for SNPs in 6 genes of pharmacogenomic relevance (*VKORC1, TMPT, SLCO1B1, DPYD, CYP2C9* and *CYP2C19*) were determined in 1792 cases. Each sample was genotyped independently in the participating research laboratories using NGS panel PGRNSeq at a mean depth of 496x and in clinical laboratories using orthogonal genotyping platforms like commercial ADME panels, Sanger sequencing or TaqMan or some other assays. All clinical laboratories were Clinical Laboratory Improvement Amendments (CLIA) approved. The overall genotyping discordance between research labs and the CLIA approved clinical labs was 2.8%. The research laboratories performing next generation sequencing on the Illumina Hiseq 2000 or 2500 platform had no analytical errors or ADO. However in these labs, wrong genotyping occurred in 11/1792 (0.06%) samples due to pre-analytical errors like sample switching. In the CLIA approved clinical laboratories, wrong genotyping occurred in 26/1702 (1.5%)

samples and 24 of these were due to ADO caused by polymorphism in the binding region of commercial genotyping assays (110).

Another large Canadian study of patients with various hereditary disorders studied the prevalence of germline ADOs in 30769 genotyping assays using Allele Specific Oligonucleotide (ASO) PCR for 8 specific mutations in *CFTR*, *CHE*, *FAH* and *SLC12A6* genes (183). These ASO PCR assays were carefully designed and validated as per College of American Pathologist (CAP) guidelines. While no ADO was observed during external quality assurance (EQA) with CAP, allele dropout or dropin later occurred in 135/30769 (0.44%) genotype assays. Unlike the eMERGE-PGx study, 94% ADOs were due to sequence independent factors and only 6% were due to sequence dependent factors like polymorphism in the PAR. The rarity of sequence dependent ADO in this study is probably due to the allele specific oligonucleotide based PCR used for genotyping.

In our LFS/LFL cohort, of the 79 pathogenic germline *TP53* mutations identified so far, 5 mutations were initially missed and 4 heterozygous mutations were incorrectly genotyped as homozygous due to ADOs. The redesigned exon 7 primers resolved 8/9 ADOs by avoiding the polymorphism IVS7+92T>G which was found to have a high minor allele frequency of 0.16 (N=112; TT=80;TG=28 & GG=4) in our cohort and 0.17 in the 1000 genome database. The P1-7R primers we used, were initially described by Bodmer's group (115, 196) and have been widely used, including the Children Oncology Group study correlating *TP53* mutations with sarcoma outcomes (197), St Jude's Children Hospital glioma study (198) and several other studies (199-203). As compared to the Bodmer group primers we used in our study, the annealing

regions of the widely used IARC protocol primers (204-207) has a higher number of polymorphisms and polymorphisms with MAF of >0.01(Table 4.3 and Table 4.6 and 4.7). However, none of IARC primers had a MAF of >0.1 unlike the MAF of 0.17 for the SNP rs12951053 in exon 7 primer of the Bodmer group. Unfortunately the primer sequence is not described in majority of the recent publications and in commercial assays (110). Hence it is difficult to estimate ADO probability in cohorts of individuals tested by different laboratories or institution and institute corrective measures.

Table 4.6: Specific polymorphisms and their minor allele frequencies in the Bodmer group primers and redesigned primers

Primer	**rs-Id**	**Primer sequence (5'→3')**	**Minor allele frequency**
1F	rs542205900	CACAGCTCTGGCTTGC*A*GA	-=0.0006/3 (1000 Genomes) -=0.0007/21 (TOPMED)
1R	rs759135662	AGCGATT*T*TCCCGAGCTGA	NA
	rs950173336	AGCGATTTTC*C*CGAGCTGA	NA
	rs934950040	AGCGATTTTCCCGA*G*CTGA	T=0.00003/1 (TOPMED)
2F	rs946807916	AGCTGTCTCAGA*C*ACTGGCA	A=0.00003/1 (TOPMED)
	rs1044211651	AGCTGTCT*C*AGACACTGGCA	T=0.0001/3 (TOPMED)
2R	rs1051425969	*G*AGCAGAAAGTCAGTCCCATG	NA
	rs747908393	GA*G*CAGAAAGTCAGTCCCATG	A=0.000008/1 (ExAC)

			A=0.00003/1 (TOPMED)
	rs910296194	GAGCA*G*AAAGTCAGTCCCATG	NA
	rs769791585	GAGCAGAAAGT*C*AGTCCCATG	G=0.000008/1 (ExAC)
	rs943163238	GAGCAGAAAGTCAGT*C*CCATG	NA
	rs368444215	GAGCAGAAAGTCAGTCCC*A*TG	G=0.00002/2 (ExAC) G=0.0002/2 (GO-ESP) G=0.00003/1 (TOPMED)
	rs745951235	GAGCAGAAAGTCAGTCCCAT*G*	A=0.00003/3 (ExAC) A=0.00003/1 (TOPMED)
3+4F	rs748527030	AGACCT*A*TGGAAACTGTGAGTGGA	C=0.000008/1 (ExAC)
	rs1800369	AGA*C*CTATGGAAACTGTGAGTGGA	A=0.0006/73 (ExAC) A=0.0012/6 (1000 Genomes)
	rs786201754	*A*GACCTATGGAAACTGTGAGTGGA	NA
3+4R	rs1794286	GAAGCC*T*AAGGGTGAAGAGGA	G=0.00003/1 (TOPMED)
	rs888713050	GAAGC*C*TAAGGGTGAAGAGGA	T=0.00003/1 (TOPMED)
5+6F	rs145153611	CGCTAGTGGGTTGCAGG*A*	A=0.0018/9 (1000 Genomes)

			A=0.0017/49 (TOPMED)
	rs35850753	C*G*CTAGTGGGTTGCAGGA	T=0.0056/28 (1000 Genomes) T=0.0073/213 (TOPMED)
5+6R	rs766876306	CA*C*TGACAACCACCCTTAAC	T=0.00001/1 (ExAC)
	rs1014274675	CACT*G*ACAACCACCCTTAAC	NA
	rs763417136	*C*ACTGACAACCACCCTTAAC	T=0.00001/1 (ExAC)
	rs751098447	CACTG*A*CAACCACCCTTAAC	G=0.00001/1 (ExAC)
	rs17884607	CACTGA*C*AACCACCCTTAAC	T=0.0019/179 (ExAC) T=0.0034/17 (1000 Genomes) T=0.00007/2 (TOPMED)
	rs752280122	CACTGACAACCAC*C*CTTAAC	A=0.00001/1 (ExAC)
	rs755937328	CACTGACAACCACCCT*T*AAC	A=0.00001/1 (ExAC)
	rs777498958	CACTGACAACCACCCTT*A*AC	T=0.00001/1 (ExAC)
	rs749063728	CACTGACAACCACCCTTA*A*C	G=0.00001/1 (ExAC) G=0.00003/1 (TOPMED)

	rs587778001	CACTGACAACCACCCTTAA*C*	A=0.00001/1 (ExAC) A=0.00003/1 (TOPMED)
7F	rs559230724	CTGCTTGCCAC*A*GGTCTC	C=0.000008/1 (ExAC) C=0.0002/1 (1000 Genomes)
	rs866409111	CTGCTTGCCA*C*AGGTCTC	NA
	rs949871368	CTGCTTGC*C*ACAGGTCTC	A=0.0001/3 (TOPMED)
	rs530001801	CTG*C*TTGCCACAGGTCTC	A=0.0002/1 (1000 Genomes)
7R	rs753873488	TGGATGG*GTA*GTAGTATGGAAG	NA
	rs12951053	TGGATGGGTAGTAGTATGG*A*AG	**C=0.1783/893 (1000 Genomes) C=0.1012/2947 (TOPMED)**
	rs767106291	TGGATGGGTAGTAGTATGGAA*G*	T=0.0001/3 (TOPMED)
8+9F	rs189582361	GTTGGGAGTAGATGGA*G*CCT	T=0.0002/1 (1000 Genomes) T=0.00003/1 (TOPMED)
	rs546697065	*G*TTGGGAGTAGATGGAGCCT	NA
8+9R	rs749446092	GGCATTTTG*A*GTGTTAGACTG	G=0.00002/2 (ExAC)
	rs757347555	GGCATTTTGAGTGTTAGAC*T*G	A=0.000008/1

			(ExAC))
10F	rs774628036	CTCAGGTACTGTGTATA*T*ACTTAC	G=0.00002/2 (ExAC)
	rs2856754	CTCAGGTACTGTG*T*ATATACTTAC	NA
	rs759914394	CTCAGGTACT*G*TGTATATACTTAC	T=0.000010/1 (ExAC)
	rs555791463	CTCAGGTA*C*TGTGTATATACTTAC	A=0.00003/3 (ExAC) A=0.0002/1 (1000 Genomes)
	rs868745759	CTCAGG*T*ACTGTGTATATACTTAC	NA
	rs753236134	C*T*CAGGTACTGTGTATATACTTAC	C=0.00001/1 (ExAC) C=0.00003/1 (TOPMED)
10R	rs993124613	*A*TACACTGAGGCAAGAAT	NA
	rs951972108	*AT*ACACTGAGGCAAGAAT	=0.00007/2 (TOPMED)
	rs984707200	ATA*C*ACTGAGGCAAGAAT	G=0.00007/2 (TOPMED)
	rs750659817	ATAC*A*CTGAGGCAAGAAT	NA
	rs1026153616	ATACAC*T*GAGGCAAGAAT	NA
	rs910384618	ATACACTGAGGCAAGAA*T*	C=0.00003/1 (TOPMED)
11F	rs562130162	TCCC*G*TTGTCCCAGCCTT	T=0.0002/1 (1000 Genomes) T=0.0001/3 (TOPMED)
11R	rs771579509	TAACCCTTAAC*T*GCAAGAACAT	NA

	rs1027267843	TAACCCTTAAC*T*GCAAGAACAT	C=0.00007/2 (TOPMED)
	rs897923545	TAACCCTTAACTG*C*AAGAACAT	NA
	rs953090502	TAACCCTTAACTGCAA*G*AACAT	A=0.0001/3 (TOPMED)
Redesigned primers			
3+4-P2-F	rs1057517593	AGACCTATGGAAACTGTGAGTGG*A*	NA
	rs776933919	AGACCTATGGAAACTGTGAGTG*G*A	T=0.000008/1 (ExAC)
	rs748527030	AGACCT*A*TGGAAACTGTGAGTGGA	C=0.000008/1 (ExAC)
	rs1800369	AGA*C*CTATGGAAACTGTGAGTGGA	A=0.0006/73 (ExAC) A=0.0012/6 (1000 Genomes)
	rs786201754	*A*GACCTATGGAAACTGTGAGTGGA	NA
3+4-P2-R	rs888713050	AGGAAGC*C*AAAGGGTGAAGAGG	T=0.00003/1 (TOPMED)
	rs1794286	AGGAAGCC*A*AAGGGTGAAGAGG	G=0.00003/1 (TOPMED)
	rs781734539	AGGAAGCCAAAGGGTGA*A*GAGG	G=0.0001/3 (TOPMED)
7-P2-F	rs1008179602	AGAATGGCGTGAACCTGGG*C*	C=0.00003/1 (TOPMED)
	rs187553272	AGAATGGC*G*TGAACCTGGG	T=0.0014/7 (1000 Genomes) T=0.0014/41 (TOPMED)

7-P2-R	rs986535183	TCCATCTACTCCCAACCAC*C*	T=0.00003/1 (TOPMED)
	rs546697065	TCCATCTACTCCCAA*C*CACC	NA

Table 4.6: Specific polymorphisms and their minor allele frequencies in the Bodmer group primers and redesigned primers: Nucleotides in between two '*' is the polymorphic site. Rows designated with "¶" indicates the MAF of polymorphism is >0.01 from any one of the database.

A false negative test for a high penetrance germline mutation or an actionable oncogenic mutation could have major clinical implications (185) as exemplified in our cohort. As a consequence of undetected germline *TP53* ADO, five of our families would not have been offered LFS screening for the probands and reflex testing for the extended family. It is therefore imperative to minimize the possibility of ADO during the design, validation and quality assurance of the genotyping assays. It is important to note that since ADOs originate during amplification process, some of the NGS assays which employ amplification based target capture are as prone to sequence dependent ADOs as Sanger sequencing (187). NGS assays which use non-amplification methods like hybrid capture could minimize ADO probability and as in our study, they are useful for confirming sequence dependent ADO. Use of primer tiling with overlapping primers could minimize sequence dependent ADOs (187) . However, a recent report revealed that primer tiling without primer trimming had resulted in missing 2/174 germline *BRCA1/BRCA2* mutations (195). Rarely, variants outside primer binding site can also cause ADO (195, 208, 209). As ADO could arise with a wide range of PCR based Sanger sequencing and NGS assays, bio-informatic flagging for homozygosity of rare variants is recommended to raise ADO alert (187).

Table 4.7: Specific polymorphisms in the annealing region of IARC protocol primers			
Primer	**rs-Id**	**Primer sequence (5'→3')**	**Minor allele frequency**
2-3F	Rs759055064	TCTCATGCTGGATCCC*C*ACT	T=0.00002/2 (ExAC)
	rs922854842	TCTCATG*C*TGGATCCCCACT	A=0.00003/1 (TOPMED)
2-3R	rs757827281	AGTCAGAGGACCAGGTCCT*C*	- =0.000008/1 (ExAC)
	rs1052483026	AGTCAGAGGACCA*G*GTCCTC	A=0.0001/3 (TOPMED)
	rs754228616	AGTCAGAGGA*C*CAGGTCCTC	T=0.000008/1 (ExAC)
	rs764517800	AGTCAGAGG*A*CCAGGTCCTC	G=0.000008/1 (ExAC)
	rs17883323	AGTCAGAGGACCAGGTC*C*TC ¶(same polymorphism as in 2nd 4F primer)	T=0.0659/7954 (ExAC) T=0.0777/389 (1000 Genomes) T=0.0853/2483 (TOPMED)
	rs200989844	AGTCAG*A*GGACCAGGTCCTC	T=0.000008/1 (ExAC) T=0.0002/1 (1000 Genomes)
	rs772244589	AG*T*CAGAGGACCAGGTCCTC	C=0.000008/1 (ExAC)
4F	rs786203749	TGCTCTTTTCACCCATCTA*C*	NA
	rs746791390	TGCTCTTTTCACCCATCT*A*C	A=0.000008/1 (ExAC)
	rs768373702	TGCTCTTTTCACCCATC*T*AC	C=0.000008/1

			(ExAC)
	rs35117667 ¶	TGCTCTTTTCACCCAT*C*TAC	A=0.0025/303 (ExAC) A=0.0076/38 (1000 Genomes) A=0.0113/330 (TOPMED)
	rs202217267	TGCTCTTTTCACC*C*ATCTAC	A=0.0001/16 (ExAC) A=0.0002/2 (GO-ESP) A=0.00007/2 (TOPMED)
	rs769697802	TGCTCTTTTCA*C*CCATCTAC	A=0.00002/2 (ExAC)
	rs374547451	TGCTC*T*TTTCACCCATCTAC	G=0.000008/1 (ExAC) G=0.00008/1 (GO-ESP)
4R	rs530718177	ATACGGCCAGG*C*ATTGAAGT	T=0.0001/16 (ExAC) T=0.0002/1 (1000 Genomes)
	rs759403255	ATA*C*GGCCAGGCATTGAAGT	A=0.000008/1 (ExAC)
2nd 4F	rs772244589	TGAGGACCTGGTCCTCTG*A*C	C=0.000008/1 (ExAC)
	rs200989844	TGAGGACCTGGTCC*T*CTGAC	T=0.000008/1 (ExAC) T=0.0002/1 (1000 Genomes)
	rs17883323	TGAGGACCTGGTC*C*TCTGAC	T=0.0659/7954

	¶ Same as in 2-3R		(ExAC) T=0.0777/389 (1000 Genomes) T=0.0708/921 (GO-ESP) T=0.0853/2483 (TOPMED)
	rs764517800	TGAGGACCTGG*T*CCTCTGAC	G=0.000008/1 (ExAC)
	rs754228616	TGAGGACCTG*G*TCCTCTGAC	T=0.000008/1 (ExAC)
	rs1052483026	TGAGGAC*C*TGGTCCTCTGAC	A=0.0001/3 (TOPMED)
	rs757827281	*TG*AGGACCTGGTCCTCTGAC	-=0.000008/1 (ExAC)
2nd 4R	rs781734539	*A*GAGGAATCCCAAAGTTCCA	G=0.0001/3 (TOPMED)
5F	rs747705704	TTCAACTCTGTCTCCTTCC*T*	C=0.000009/1 (ExAC)
	rs376713749	TTCAACTCTGTCT*C*CTTCCT	A=0.00004/4 (ExAC) A=0.0002/2 (GO-ESP)
	rs756417643	TTCAACTCTGTC*T*CCTTCCT	A=0.000009/1 (ExAC)
	rs773029752	TTCAACT*C*TGTCTCCTTCCT	A=0.00002/2 (ExAC)
5R	rs748298015	CA*G*CCCTGTCGTCTCTCCAG	A=0.00002/2 (ExAC) T=0.00003/1 (TOPMED)
	rs778145407	CAGCCCT*G*TCGTCTCTCCAG	C=0.00002/2

			(ExAC)
	rs56181208	CAGCCCTGTC*G*TCTCTCCAG	A=0.0002/26 (ExAC) A=0.00008/1 (GO-ESP) A=0.00007/2 (TOPMED)
	rs774831915	CAGCCCTGTCG*T*CTCTCCAG	C=0.000008/1 (ExAC)
	rs547244762	CAGCCCTGTCGTCT*C*TCCAG	T=0.00006/7 (ExAC) T=0.0002/1 (1000 Genomes)
	rs772637453	CAGCCCTGTCGTCTCT*C*CAG	G=0.000008/1 (ExAC)
6F	rs76962868	GCCTCTGATTCCTCAC*T*GAT	NA
	rs869158731	GCCTCTGAT*T*CCTCACTGAT	NA
	rs757029874	GCCTCT*G*ATTCCTCACTGAT	T=0.00002/2 (ExAC)
6R	rs755937328	T*T*AACCCCTCCTCCCAGAGA	A=0.00001/1 (ExAC)
	rs777498958	TT*A*ACCCCTCCTCCCAGAGA	T=0.00001/1 (ExAC)
	rs749063728	TTA*A*CCCCTCCTCCCAGAGA	G=0.00001/1 (ExAC) G=0.00003/1 (TOPMED)
	rs587778001	TTAA*C*CCCTCCTCCCAGAGA	A=0.00001/1 (ExAC) A=0.00003/1 (TOPMED)
	rs778894743	TTAAC*C*CCTCCTCCCAGAGA	G=0.00002/2

			(ExAC)
	rs34949160	TTAACCCC*T*CCTCCCAGAGA	C=0.0088/880 (ExAC) C=0.0050/25 (1000 Genomes) C=0.0037/48 (GO-ESP) C=0.0041/118 (TOPMED)
	rs200372146	TTAACCCCT*C*CTCCCAGAGA	T=0.0002/1 (1000 Genomes) T=0.0002/5 (TOPMED)
	rs762077893	TTAACCCCTCCTCCCAGA*G*A	C=0.000009/1 (ExAC)
5-6F	rs576912263	TGTTCACTTGTGCCC*T*GACT	G=0.00001/1 (ExAC)
	rs770686190	TGTTCACTTG*T*GCCCTGACT	G=0.00001/1 (ExAC)
	rs994417264	TGTTCA*C*TTGTGCCCTGACT	A=0.00003/1 (TOPMED)
	rs774330271	TGT*T*CACTTGTGCCCTGACT	C=0.00001/1 (ExAC)
5-6R	(same as 6R)	TTAACCCCTCCTCCCAGAGA	
7F	rs559230724	CTTGCCAC*A*GGTCTCCCCAA	C=0.000008/1 (ExAC) C=0.0002/1 (1000 Genomes)
	rs866409111	CTTGCCA*C*AGGTCTCCCCAA	NA
	rs530001801	*C*TTGCCACAGGTCTCCCCAA	A=0.0002/1 (1000 Genomes)
7R	rs1642786	AGGGGTCAG*A*GGCAAGCAGA	NA

	rs1036988722	AGGGGTCAGA*G*GCAAGCAGA	A=0.00003/1 (TOPMED)
2nd 7F	rs769521263	AGGCGCACTG*G*CCTCATCTT	T=0.00003/3 (ExAC)
	rs17880604 ¶	AGGC*G*CACTGGCCTCATCTT	G=0.0127/1527 (ExAC) G=0.0050/25 (1000 Genomes) G=0.0102/296 (TOPMED)
	rs374907737	AGG*C*GCACTGGCCTCATCTT	A=0.0003/31 (ExAC) A=0.0002/1 (1000 Genomes) A=0.0004/5 (GO-ESP)
2nd 7R	rs756510290	*T*GTGCAGGGTGGCAAGTGGC	C=0.000008/1 (ExAC)
	rs201930255	TGTGCA*G*GGTGGCAAGTGGC	T=0.00006/7 (ExAC) C=0.0002/1 (1000 Genomes) T=0.0001/4 (TOPMED)
	rs17880172	TGTGCAGGGT*G*GCAAGTGGC	A=0.0006/71 (ExAC) A=0.0014/7 (1000 Genomes) A=0.0018/52 (TOPMED)
	rs17881780	TGTGCAGGGTGGCAA*G*TGGC	A=0.00007/8 (ExAC)

			A=0.0002/6 (TOPMED)
	rs200277687	TGTGCAGGGTGGCAAGT*G*GC	A=0.00002/2 (ExAC) A=0.0002/1 (1000 Genomes)
8F	rs113302588	TTCCTTAC*T*GCCTCTTGCTT	G=0.0002/14 (ExAC) G=0.0013/37 (TOPMED)
	rs376988747	TTCCTTA*C*TGCCTCTTGCTT	A=0.00002/1 (ExAC) A=0.00007/2 (TOPMED)
	rs776659879	TTCCTT*A*CTGCCTCTTGCTT	C=0.00007/4 (ExAC) C=0.0001/4 (TOPMED)
	rs761687865	TTCCT*T*ACTGCCTCTTGCTT	NA
8R	rs144496254	AGGCATAAC*T*GCACCCTTGG	A=0.00002/2 (ExAC) C=0.0002/1 (1000 Genomes)
	rs945036538	AGGCATAACT*G*CACCCTTGG	A=0.00007/2 (TOPMED)
	rs747431888	AGGCATAACTGCACC*C*TTGG	G=0.000008/1 (ExAC)
8-9F	rs189582361	TTGGGAGTAGATGGA*G*CCT	T=0.0002/1 (1000 Genomes) T=0.00003/1 (TOPMED)
8-9R	rs749446092	*A*GTGTTAGACTGGAAACTTT	G=0.00002/2

			(ExAC)
	rs757347555	AGTGTTAGAC*T*GGAAACTTT	A=0.000008/1 (ExAC)
	rs149983651	AGTGTTAGACTGGAAA*C*TTT	T=0.00002/3 (ExAC) G=0.0002/1 (1000 Genomes)
9F	rs768950740	GACAAGAAGCGGT*G*GAG	T=0.000008/1 (ExAC)
	rs376079415	GACAAGAAGC*G*GTGGAG	T=0.00007/8 (ExAC) T=0.0002/6 (TOPMED)
	rs549731874	GACAAGAAG*C*GGTGGAG	A=0.000008/1 (ExAC) A=0.0002/1 (1000 Genomes) A=0.00003/1 (TOPMED)
	rs936119541	*G*ACAAGAAGCGGTGGAG	-=0.00003/1 (TOPMED)
	rs769211481	*G*ACAAGAAGCGGTGGAG	A=0.000008/1 (ExAC)
9R	rs769817504	*C*GGCATTTTGAGTGTTAGAC	NA
	rs567351657	CGGC*A*TTTTGAGTGTTAGAC	NA
	rs749446092	CGGCATTTTG*A*GTGTTAGAC (same polymorphism as in 8-9R)	G=0.00002/2 (ExAC)
10F	No polymorphism	CAATTGTAACTTGAACCATC	
10R	rs375905156	GGATGAGAATGGAATCCT*A*T	G=0.0008/4 (1000 Genomes) G=0.0016/46 (TOPMED)

11F	rs750853458	AGACCCTCTCA*C*TCATGTGA	A=0.00002/2 (ExAC)
	rs17881850 ¶	AGACC*C*TCTCACTCATGTGA	A=0.0112/1359 (ExAC) A=0.0076/38 (1000 Genomes) A=0.0064/185 (TOPMED)
	rs767047551	AGA*C*CCTCTCACTCATGTGA	A=0.00002/3 (ExAC)
11R	rs375913211	TGA*C*GCACACCTATTGCAAG	T=0.0002/6 (TOPMED)
	rs35919705	TGAC*G*CACACCTATTGCAAG	A=0.0014/7 (1000 Genomes) A=0.0014/40 (TOPMED)
	rs886053513	TGACGCACAC*C*TATTGCAAG	NA

Table 4.7: Specific polymorphisms and their minor allele frequencies in IARC protocol primers: (http://p53.iarc.fr/download/*TP53*_directsequencing_iarc.pdf): Nucleotides in between two '*' is the polymorphic site. Rows designated with "¶" indicates the MAF of polymorphism is >0.01 from any one of the database.

With regards to the *TP53* gene, there are several factors which could influence the probability of ADO. While the relatively smaller size of the gene and minimal repetitive sequences in *TP53* would have a lower ADO risk, it may be more than offset by the large number of polymorphisms (Table 4.3) and an abundance of G-Quadruplexes (Table 4.4) in *TP53*. Hence the prevalence and nature of ADOs in *TP53* may not be applicable for other genes. Nevertheless, our findings should prompt systematic large studies of ADO in *TP53* and other genes in diverse cancer cohorts, genotyped with different methods. This will provide better insight on the various

clinical, genetic and technological contexts where ADOs could be a special consideration or require systematic corrective actions. ADO should be suspected whenever a homozygous germline mutation is identified in a Mendelian Dominant condition. Similarly, a false negative report due to ADO should be suspected in patients with classical syndromic diagnosis in whom no likely pathogenic mutation is identified in the relevant genes on amplification based genotyping and MLPA. Such cases or families, whether tested now or in the past, may be recalled for retesting by appropriate methods. Retrospective retesting of individuals with redesigned primers should be considered whenever any polymorphism with a significant allele frequency is identified in the annealing region of the primers used. The *TP53* exon 7 primers used in several studies, including ours, should never be used for somatic or germline *TP53* analysis.

In conclusion, we show that germline allele dropouts in *TP53* are not extremely rare and this may be true in other cancer genes. Considering the major clinical implications of a false negative genetic analysis report due to ADOs, a systematic evaluation of ADOs in different clinical, genetic and technological contexts with appropriate remedial actions or retesting is warranted.

Chapter 5

TP53 & *MDM2* polymorphisms and clinical phenotype in sporadic sarcoma

5.1 Sarcoma

Sarcoma is a rare type of cancer of the connective tissues predominantly involving bone, cartilage, fat, muscles and ligaments. Two major types of sarcoma include the bone sarcoma and soft tissue sarcomas, each of which are further classified into many subtypes. Bone sarcoma accounts for only about 0.2% of all cancers yet it is the third leading cause of death in adolescents (210).

5.1.1 History

The term sarcoma is originated from the Greek word "*sarx"*, which means flesh. Initially 'Fungus' a term derived from the Latin word *Spongos* or Sponge which means an abnormal spongy outgrowth was also used for sarcoma but the word sarcoma became more popular (211). Till the 19^{th} century sarcoma lesions were studied along with other cancers as there were no techniques to distinguish between them. With the advent of cellular pathology, different classification of tumor terminology started to appear from the 19^{th} century (211, 212). Alexis Boyer from Paris was the first to differentiate between tumors like exostosis, gumma of bone, spina ventosa and osteosarcoma and was the first one to coin the term osteosarcoma in 19^{th} century. Rapid progress was made after the work of Jean Cruveilhier and J. C. A. Recamier who published the gross anatomy and pathology of bone and soft tissue sarcomas and differentiation of bone primary and metastatic tumors respectively. Samuel Weissel Gross and Albert Ferguson had major contribution in the field of sarcoma and its treatment including limb amputation for cure (211, 212). Subsequent advancement in biological understanding of different sarcomas, diagnostic and therapeutic techniques, the classification system and the treatment outcome of sarcoma improved over the decades.

5.1.2 Epidemiology

Bone and soft tissue sarcomas are uncommon malignancies in both genders as reported by the IARC Globocan in 2012 and the American Cancer Society in 2015 report (American Cancer Society. *Global Cancer Facts & Figures* 3rd Edition. Atlanta: American Cancer Society; 2015). However there are some notable exceptions. In children, after haemato-lymphoid and central nervous system malignancies, bone and soft tissue sarcomas are the third most common cancers as seen in Figure 5.1. The other striking difference is in the geographical differences in the prevalence of Kaposi Sarcoma. Being an Acquired Immuno Deficiency Syndrome (AIDS) defining cancer, Kaposi Sarcoma is the most common cancer in certain countries in East Africa where HIV AIDS incidence is very high. Excluding the Kaposi Sarcoma, soft tissue sarcomas are more common than bone sarcoma with their prevalence being 73% and 27% respectively in the USA as reported recently by USA Surveillance, Epidemiology, and End Results (SEER) Program (https://seer.cancer.gov/csr/1975_2015/).

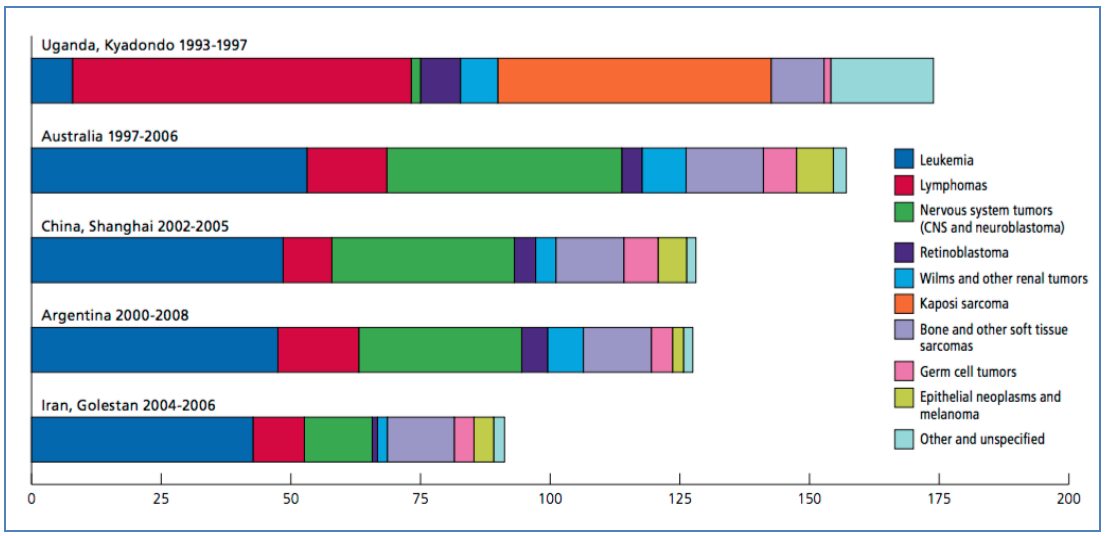


Figure 5.1: Distributions of Cancer in Children 0-14 Years of Age in Selected Populations (age-standardized incidence rates per million). Reproduced from Global Cancer Facts & Figures, 3rd Edition.

Of the bone tumors, osteosarcoma is the most frequently occurring tumor followed by chondrosarcoma (213, 214). Both STS and osteosarcomas have bimodal peak incidence. The incidence of STS, especially the embronal type, is high in early childhood upto the age of 5 years, then it dips and starts rising again from middle age. In contrast, the incidence of bone tumors increases during adolescence and then again in old age (figure 5.2).

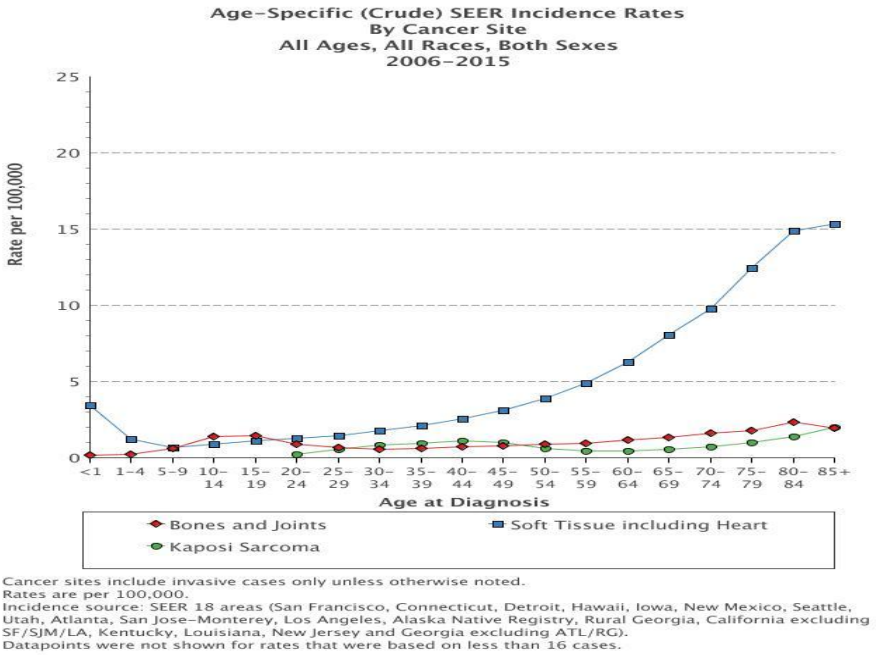


Figure 5.2: Incidence of bone and joints cancer, soft tissue cancer and Kaposi sarcoma in USA. Taken from (https://seer.cancer.gov/csr/1975_2015/)

Incidence of osteosarcoma below 24 years of age is same worldwide which is around 3-5 per million in males and 2-4 per million in females (213). Incidence of osteosarcoma in males is maximum during the age of 15-19 year and in females during the age of 10-14year. Osteosarcoma is more common in males than in females (213).

There are few reports from India which show trends in the incidence rate over time. In a 1998 report by Yeole and Jussawala from the Bombay Cancer Registry, bone cancer accounted for 0.9% of all the cancers, with Ewing sarcoma being the most common bone tumor (215). In 1996 Rao et al reported that in the Dakshina Kannada district of Karnataka, of the 523 bone tumors, 39% were malignant and of these malignant tumors 45.7% were Osteogenic Sarcoma (OGS) and 19.4% were Ewing sarcoma (216). In this Indian study, the incidence of bone tumors was maximum in the second and third decade of life with a higher incidence in males (216). Gulia et al have recently reported on patients registered with bone or soft tissue tumors registered from 1st January 2010 to 31 December 2010 at the Tata Memorial Hospital. They found that 65% of the patients were males and 35% were females. During this period, bone tumors were more frequent (60%) as compared to the soft tissue tumors (36%) which are in contrast to SEER database where the soft tissue tumors are more prevalent. This is possibly due to the referral bias in a hospital based report. A recent review by Ramaswamy et al covering most of the publication from India on sarcoma revealed that OGS is the most common bone tumor followed by Ewings'sarcoma and chondrosarcoma (210).

5.1.3 Classification

Sarcoma is broadly classified into bone sarcoma and soft tissue sarcoma. The most common subytpes of bone sarcomas include osteosarcoma, chondrosarcoma and Ewing's sarcoma while fibrosarcomas, chordomas, and undifferentiated pleomorphic sarcoma (UPS) being the other rare subtypes. Soft tissue tumors have multiple classification of which the most common type of soft tissue sarcomas are spindle cell sarcoma, synovial sarcoma, leiomyosarcoma, Liposarcoma etc. The prevalence of

different soft tissue sarcoma varies according to age with Rhabdomyosarcoma being more common in children (217). World Health Organization (WHO) classification of soft tissue and bone tumors published in 2002 and modified in 2013, classified soft tissue tumors into 9 sub categories and bone tumors in to 12 subcategories (218). A brief outline of WHO classification of bone and soft tissue tumors along with few examples of malignant/benign type of each category are given in the table 5.1.

Table 5.1: WHO classification of Bone and Soft tissue tumors:

Bone Tumors			**Soft Tissue Tumors**		
1	Cartilage tumors	Chondrosarcoma	1	Adipocytic tumors	Dedifferentiated liposarcoma
					Myxoid liposarcoma
					Round cell liposarcoma
2	Osteogenic tumors	chondroblastic Osteosarcoma	2	Fibroblastic/ myofibroblastic Tumors	Adult fibrosarcoma
		fibroblastic Osteosarcoma			Myxofibrosarcoma
		Osteoblastic Osteosarcoma			Low grade fibromyxoid sarcoma
3	Fibrogenic tumors	Fibrosarcoma	3	So-called fibrohistiocytic tumors	Pleomorphic ‘MFH’ / Undifferentiated pleomorphic sarcoma
					Giant cell ‘MFH’ / Undifferentiated pleomorphic sarcoma with giant cells
					Inflammatory ‘MFH’ / Undifferentiated pleomorphic sarcoma

					with prominent inflammation
4	Fibrohistiocytic tumors	Malignant fibrous histiocytoma	4	Smooth muscle tumors	Leiomyosarcoma (excluding skin)
5	Ewing sarcoma/primitive Neuroectodermal tumor	Ewing sarcoma	5	Pericytic (perivascular) tumors	Glomus tumor (and variants)
6	Haematopoietic tumors	Malignant lymphoma	6	Skeletal muscle tumors	Embryonal rhabdomyosarcoma
					Alveolar rhabdomyosarcoma
					Pleomorphic rhabdomyosarcoma
7	Giant cell tumor	Malignant giant cell tumor	7	Vascular tumors	Epithelioid haemangioendothelioma
					Angiosarcoma of soft tissue
8	Notochordal tumors	Chordoma	8	Chondro-osseous tumors	Mesenchymal chondrosarcoma
					Extraskeletal osteosarcoma
9	Vascular tumors	Angiosarcoma	9	Tumors of uncertain Differentiation	Synovial sarcoma
					Epithelioid sarcoma
					Alveolar soft part sarcoma
10	Myogenic,	Leiomyosarcom			

	Lipogenic, Neural, and Epithelial Tumors	a	
		Liposarcoma	
		Neurilemmoma	
		Adamantinoma	
11	Tumors of Undefined Neoplastic Nature	Fibrous dysplasia	
		Osteofibrous dysplasia	
		Langerhans cell histiocytosis	
12	Congenital and Inherited Syndromes Associated with Bone and Soft Tissue Tumor		Li-Fraumeni syndrome
			Beckwith-Wiedemann syndrome
			Bloom syndrome
			Maffucci syndrome (sporadic)
			Paget disease of bone, familial
			Retinoblastoma
			Rothmund-Thomson syndrome
			Rubinstein-Taybi syndrome
			Werner syndrome

5.1.4 Tumor distribution

Sarcomas are distributed throughout the body. A large study with soft tissue sarcoma revealed that 40% of tumors originated in the extremities and 38% in visceral or retroperitoneal areas (219). One third of extremity and trunk sarcomas are superficial with a median diameter of 5cm and the remaining two third are deep seated with a median diameter of 9cm (218). Osteosarcoma occurs more frequently in younger age group and in this age group 80% of the tumors occur in the long bones. Incidence rate of tumor in bone extremity in the older age group is one third of the younger age

group (218). Chondrosarcoma are also more common in long bones with other common sites being pelvis and ribs. Ewing's sarcoma has similar epidemiological features as osteosarcoma but tends to occur in the diaphysis of long bones as compared to metaphyseal areas for OGS

5.1.5 Etiology

Exact cause of bone and soft tissue sarcomas are not known though there are many risk factors associated with these types of tumors (220-223). Various risk factors include genetic and environmental factors such as exposure to ionizing radiation, infection and immune deficiency. While few studies have shown the association of chemical carcinogens with sarcoma occurrence (218, 224, 225), their results have not been replicated. About 0.1 to 1% of sarcoma cases are due to prior radiation exposure during treatment and occur after a median time of 10 years after exposure (218, 226). Immune deficient individuals such as those with HIV AIDS are prone to sarcoma, specifically Kaposi sarcoma (227). While majority of sarcoma cases are sporadic in nature, they are also associated with various hereditary syndromes like Li Fraumeni Syndrome (22%), Retinoblastoma (less than 15%), Rothmund Thompson syndrome (7-32%), neurofibromatosis (17-67%), Costello syndrome (20%) etc (228). In Li-Fraumeni and Li-Fraumeni like syndrome (LFS/LFL), sarcoma is the second most common cancer in females after breast cancer (146) and most common in males.

5.1.6 Diagnosis, grading and staging of sarcomas

Exact diagnosis of any cancer is very important for their proper treatment. Sarcoma being very diverse in nature, needs to be diagnosed and staged accurately. Important parameters for diagnosis are tumor location, size, type of matrix, and nature of tumor

whether it is benign or malignant. Age is also an important consideration in making the diagnosis of sarcomas as certain sarcomas are age specific. The Radiological features also helps to suggest possible histological type and tumor aggressiveness (218).

Due to nonspecific clinical features of bone tumors, early diagnosis is difficult. The initial presentation may be pain and swelling. Along with age, location and size the radiological features of type of cortical destruction and periosteal reaction is an important parameter for the diagnosis of bone tumors. It is important to find out the axis of the lesion while diagnosing the bone sarcomas. . National Comprehensive Cancer Network (NCCN) provide the guidelines for cases below 40 years of age with abnormal radiograph to be referred to orthopedic oncologist and individuals above 40 years should undergo an investigation for bone metastasis (229, 230). European Society for Medical Oncology (ESMO) guidelines recommends an Magnetic Resonance Imaging (MRI) followed by radiograph and CT scan if MRI results are not equivocal (231).

MRI is the recommended technique for the initial diagnosis of soft tissue sarcoma as it is superior to other imaging modalities in terms of differentiating between normal and abnormal tissue and in visualizing neuro-vascular structures for the purpose of surgical planning. For the diagnosis of sarcoma, biopsy is necessary for identifying the specific histology and its grading (218).

Histological grading of tumor is done for predicting the behavior of a malignant tumor based on its histological feature. The French Federation of Cancer Centers Sarcoma Group (FNCLCC) system and the National Cancer Institute (NCI) system are the most widely used grading system. Tissue differentiation, tumor necrosis and

mitotic activity are considered in the FNCLCC grading system (232) and histology, location and necrosis are considered in NCI grading system (233).

For bone tumors various staging systems are used. Tumor-Node-Metastasis (TNM) classification recommended by American Joint Cancer Committee and Union for International Cancer Control (AJCC/UICC) (234) and surgical staging system by Musculoskeletal Tumor Society (MTS) (235) are followed by NCCN.

5.1.7 Treatment

Most sarcomas require multimodality treatment combining surgery, radiotherapy and chemotherapy in different sequence based on the tumor size, location, and histological subtype and patient's age. Traditionally, extremity sarcomas were treated with amputation. However with better techniques of radiotherapy and surgical refinements and better endo-prosthesis, it is often possible to do limb sparing surgery for sarcomas followed by radiotherapy. Primary line of treatment for low grade sarcomas is wide excision with or without radiotherapy. For OGS neoadjuvant chemotherapy with high dose methotrexate is used followed by surgery (229). Histological response to chemotherapy is an important prognostic factor in OGS. For,high grade soft tissue sarcomas, in addition to surgery and radiotherapy, adjuvant chemotherapy may be used (230). For childhood rhabdomyosarcoma, chemotherapy is the mainstay of treatment along with appropriate local treatment with surgery or radiotherapy.

5.2 Genetics of sarcoma

Genetic basis of Sarcoma is not well defined. Recently, Ballinger M. et al have reported the genetic landscape in 1162 sarcoma patients enrolled in the International Sarcoma Kindred Study (ISKS) from various parts of the world. The ISKS study

identified germline monogenic and polygenic variants in 55% sarcoma patients, highlighting the role of genetic factors in sarcoma genesis (236). Variant in these genes were correlated with earlier diagnosis of disease. The reported variants included pathogenic and likely pathogenic variants as well as VUS. In addition to mutations in the tumor suppressor genes like *TP53*, *BRCA2* and *ATM*, mutations in the *ERCC2* gene were observed with high frequency. The ERCC2 gene encodes a helicase helps in nucleotide excision repair.

Over the last few decades, a large number of studies have reported conflicting results of the association of several polymorphic gene variants with cancer risk (97, 98, 237-240). Several studies have also evaluated the modifier effect of low penetrance polymorphic variants on clinical outcome of some cancers (145, 240). While several *TP53* polymorphisms like Arg72Pro polymorphism (rs1042522) and MDM2 309 SNP (rs2279744) have been extensively studied in different cancers (145) only few studies have examined their role in sarcomas. Toffoli et al have shown the prognostic value of *TP53* Arg72Pro SNP for event free and overall survival in osteosarcoma and the risk association of MDM2 SNP309 for osteosarcoma development in females (241). Pauline et al have reported risk association of G allele of MDM2 309SNP with higher grade and node positivity in breast cancer in the Scottish population (242). Lum et al found association of T allele of MDM2 309SNP with early onset of breast cancer in the Chinese population (243). Polymorphism of Intron 3 (PIN3), a 16 base pair duplication *TP53* polymorphism has been studied in various cancers. Risk association of PIN3 with breast cancer development has been reported in two meta-analysis in 2010 and 2013 (237, 239). However the risk association of the PIN3 polymorphism with sarcoma risk or the clinical phenotype or clinical outcome and survival in

sarcoma has not been reported. Overall, very few studies have evaluated the impact of *TP53* and MDM2 polymorphisms on the clinical outcome in sarcoma patients. In the Indian population, the *TP53* and MDM2 polymorphisms have not been studied for sarcoma risk or its clinical phenotype and outcome. We have therefore undertaken the first study of *TP53* and MDM2 polymorphism in a large cohort of sporadic and hereditary sarcoma patients. .

5.3 Aim: To study the impact of common *TP53* and MDM2 polymorphisms in Indian sarcoma patients.

5.4 Methodology: Sarcoma cases were enrolled from the TMC International Sarcoma Kindred Study (TISKS) at the Tata Memorial Centre. The cases were enrolled between June 2010 to September 2016 and follow-up data available uptil January 2018. All demographic, clinical, histopathological and treatment details were noted and blood sample was collected after taking written informed consent. Response to chemotherapy and disease status during follow up was recorded. Genotyping was performed on the genomic DNA extracted from blood for 3 polymorphisms (*TP53* R72P, *TP53* PIN3 and MDM2 SNP309) as per the methodology described in chapter 2. Genotype of these 3 polymrphisms was correlated with age at sarcoma diagnosis, maximum tumor dimension, response to chemotherapy, metastases free and overall survival. Different statistical methods were used for various genotype– phenotype correlations. To compare mean age at diagnosis in different genotypes, mean comparison t test or non parametric test was used depending on the distribution. For response to chemotherapy (≥90% or <90% response) and maximum tumor size (≥ or < 8 cm), chi square test was used. The metastases free and overall survival probability was estimated using Kaplan Meier method and the differences were compared using

the log rank test. As subgroup analysis was done, instead of $p<0.05$, a p value of <0.01 was considered as statistically significant.

5.5 Results

A total of 311 sarcoma cases enrolled in the ongoing TMC International Sarcoma Kindred Study (TISKS) in between December 2010 till July 2018 were genotyped for 3 polymorphisms (*TP53* R72P, *TP53* PIN3 and MDM2 SNP309). These genotypes were correlated with various clinico-pathological characteristics and the clinical outcome including survival in these sarcoma patients treated with multimodality treatment at the Tata Memorial Hospital. The demographic detail of these 311 cases is shown in table 5.2. Of the 311 cases enrolled, 302 are from India, 8 from Bangladesh and 1 from Nepal. Majority (N=127) of the cases were from Maharashtra (n=127), followed by 43 cases Uttar Pradesh (n=43), West Bengal (n=35) and rest of India as shown in figure 5.3 . Of the 311 cases, 100 were female and 211 were males.

Table 5.2 Demographics and clinical details

Country	Frequency	Percent
Bangladesh	8	2.5
India	302	97.1
Nepal	1	0.3
Gender		
Female	100	32.2
Male	211	67.8
Family History		
No	282	90.4
Yes	29	9.3
Tumor Type		
Bone	263	84.3
Soft Tissue	48	15.4
Religion		
Hindu	246	78.8
Muslim	56	17.9

Others	9	2.9
Tumor Size*		
<8	61	19.6
≥8	238	76.3
Not Available	12	3.8
Tumor response to chemotherapy[+]		
<90	73	23.5
≥90	81	26
Not Available	157	50.5
Intent of Treatment		
Curative	265	84.9
Paliative	26	8.3
Not Available	20	6.4
Metastasis at presentation and followup		
No	168	53.8
Yes	117	37.5
No Record	26	8.3

*Tumor size is represented as the maximum dimension of the tumor and the scale is in cm. [+] **Response to chemotherapy is shown in percentage (%) of tumor shrinked.**

Family history of cancer in a first or second degree relative was noted in 29 cases but none of these had a strong family history of a classical Li-Fraumeni Syndrome. Bone sarcomas accounted for 84.5% (n=263) cases with only 15.4% (n=48) cases having STS. Around 76% (n= 238) of tumors had the maximum tumor dimension of ≥8cm and 20% (n=61) had <8 cm. Tumor size was not known for 12 cases. Majority of these cases had curative treatment at the Tata Memorial Hospital. Distant metastasis was noted in 57 cases at presentation. Of the remaining 254 cases, 168 cases remained metastases free till their last follow up, 60 developed metastases during follow up and the metastatic status was not known in the remaining 26 cases.

Overall mean age of sarcoma diagnosis was 23.9yrs (±14.8) for all sarcomas, 21.3yrs (±12.5) for bone sarcomas and 38.2yrs (±18.3) for STS. The mean age for OGS was 18.8yrs (±8.8). The mean age of sarcoma diagnosis was 22.4yrs (±15.3) in females as

compared to 24.6yrs (±14.6) in males. Similarly for OGS, the mean age of diagnosis in females was 17.3yrs (±9.5) as compared to 19.4yrs (±8.4) in males.

The trend of early age of sarcoma diagnosis in females did not reach statistical significance

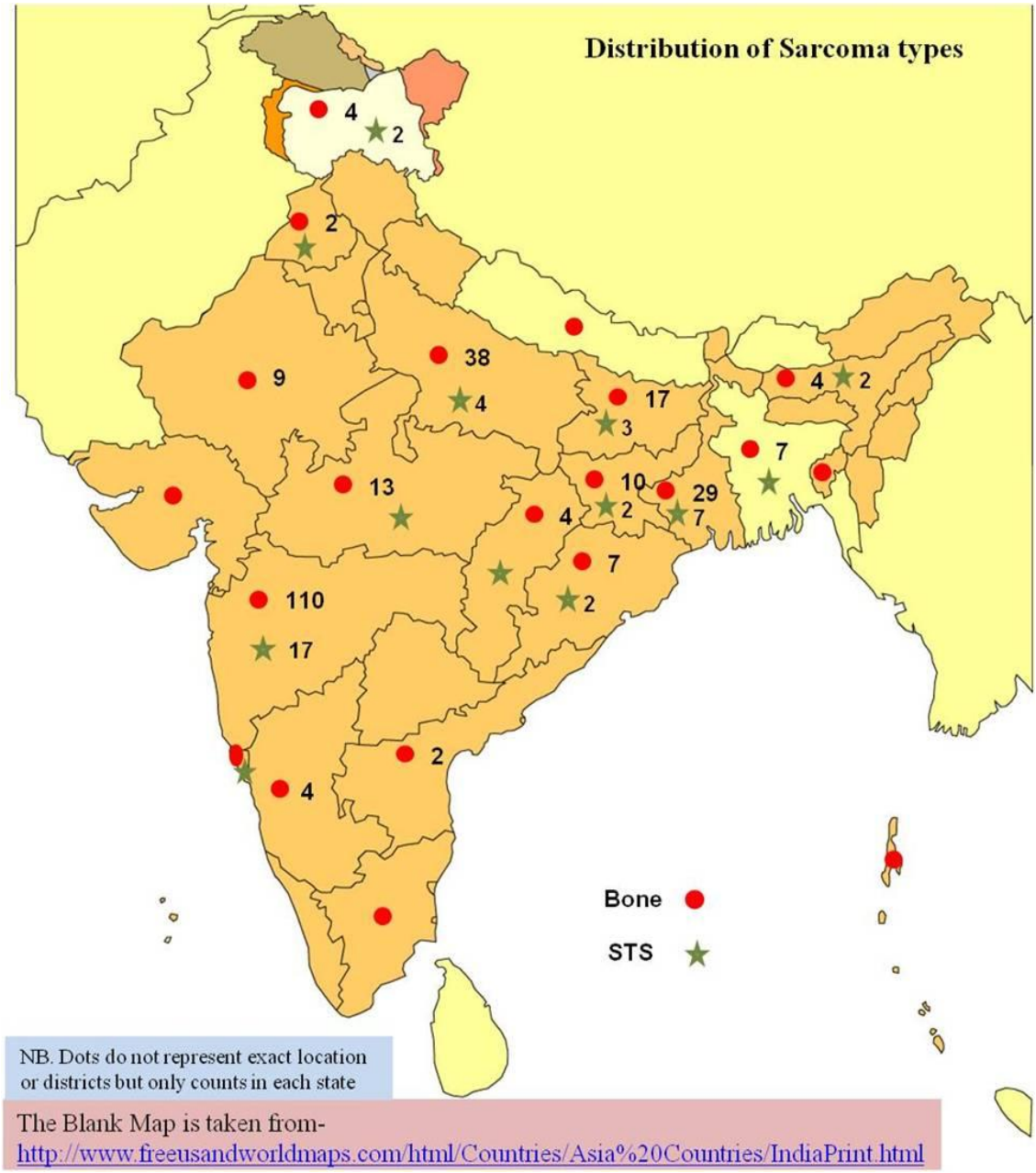


Figure 5.3: Distribution of Sarcoma Type in India

5.5.1 Allele frequency of three polymorphisms:

The genotyping results in 311 cases were analysed to estimate the allele frequency of , Arg72Pro, PIN3 and *MDM2*SNP 309 polymorphism as shown in Table 5.3. All the polymorphisms were found to be in Hardy Weinberg Equilibrium.

Table 5.3 Genotypic frequency of three polymorphisms		
***TP53* Arg72Pro**	Counts (N= 311)	Allele frequency
CC	65 (20.8%)	C=47.59%
GC	166 (53.2%)	G=52.41%
GG	80 (25.6%)	
MDM2 SNP309		
GG	85 (27.2%)	G=50.96%
TG	147 (47.1%)	T=49.04%
TT	79 (25.3%)	
TP53 PIN3 (**Intron 3 Duplication)**		
WT	197 (63.1%)	A1*=80.22%
Htz	105 (33.7%)	A2*=19.77%
DUP	9 (2.9%)	

Table 2: *A1 represents single 16 bp stretch (not duplicated and *A2 represents duplication of 16bp

5.5.2 Polymorphisms and age

The mean age at diagnosis of all sarcomas (n=311) or only OGS (n=200) was correlated with different genotypes as shown in Table 5.4. There was a trend for early age of sarcoma diagnosis in cases with PIN3 duplication but did not reach

Table 5.4: Age of sarcoma diagnosis with various genotypes in entire sarcoma cohort

TP53 R72P	Age at diagnosis	
GG (80)	Mean	25.49
	Median	18.00
	Std. Deviation	16.399
GC (167)	Mean	23.29
	Median	18.00
	Std. Deviation	14.135
CC (64)	Mean	23.63
	Median	18.00
	Std. Deviation	14.580
Combined Genotype		
GG (80)	Mean	25.49
	Std. Deviation	16.399
GC+CC (231)	Mean	23.39
	Std. Deviation	14.229
CC (64)	Mean	23.57
	Std. Deviation	14.473
CG+GG (247)	Mean	24.02
	Std. Deviation	14.936

MDM2SNP309	Age at diagnosis	
TT (79)	Mean	22.58
	Median	17.00
	Std. Deviation	14.025
TG (147)	Mean	24.61
	Median	20.00
	Std. Deviation	15.264
GG (85)	Mean	23.99
	Median	18.00
	Std. Deviation	14.840
Combined Genotype		
TT (79)	Mean	22.53
	Std. Deviation	13.946
TG+GG (232)	Mean	24.41
	Std. Deviation	15.108
TT+TG (226)	Mean	23.90
	Std. Deviation	14.843
GG (85)	Mean	23.99
	Std. Deviation	14.840

PIN3	Age at diagnosis	
WT (197)	Mean	24.03
	Median	18.00
	Std. Deviation	14.904
Htz (105)	Mean	24.14
	Median	19.00
	Std. Deviation	14.848
DUP (9)	Mean	19.11
	Median	17.00
	Std. Deviation	13.186
Combined Genotype		
WT (197)	Mean	24.03
	Std. Deviation	14.904
Htz+Dup (114)	Mean	23.75
	Std. Deviation	14.734
Wt+Htz (302)	Mean	24.07
	Std. Deviation	14.860
Dup (9)	Mean	19.11
	Std. Deviation	13.186

significance due to small number of cases with duplication (n=9). Correlation between age of sarcoma diagnosis and different genotype is shown in table 5.4 and figure 5.4.

5.5.3 Polymorphisms and tumor size

Information on the maximum tumor size was available for 299 cases and was categorized as <8 cm or ≥8 cm. For the R72P SNP, the GC and GC+CC genotype showed a trend for larger tumors with a p value of 0.038 and 0.052.

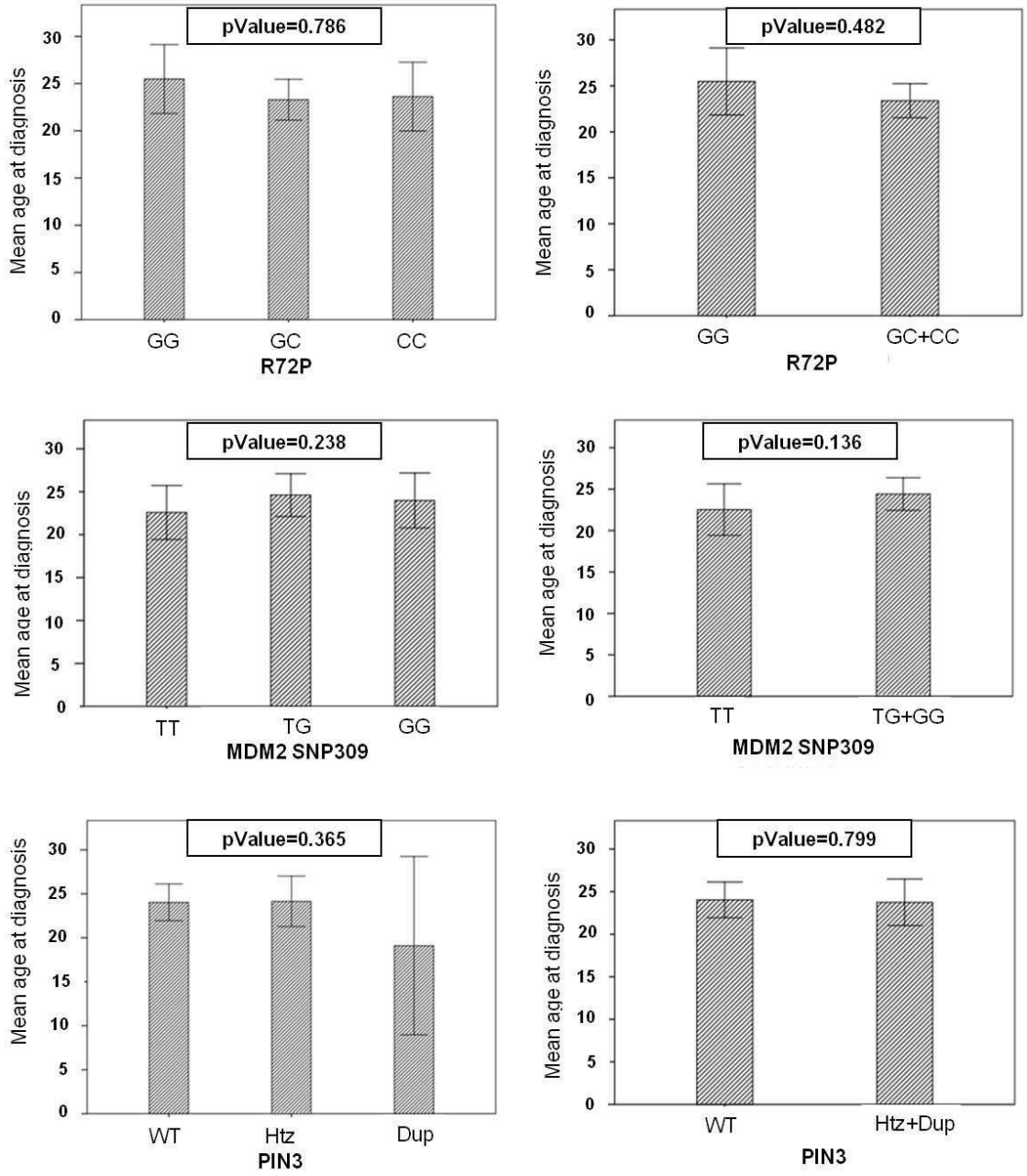


Figure 5.4: Mean age of Sarcoma diagnosis

Table 5.5: Association of Tumor size in entire sarcoma cohort with various polymorphisms

Tumor size in entire sarcoma cohort									
	P72R			Chi-square (df=2)	P72R		Chi-square (df=1)	P72R		Chi-square (df=1)
	CC	GC	GG		CC	CG+GG		CC+CG	GG	
<8 (n=61)	16	24	21	P=0.038	16	45	P=0.156	40	21	P=0.052
≥8 (n=238)	46	137	55		46	192		183	55	
	MDM2SNP309				SNP309			SNP309		
	TT	TG	GG		TT	TG+GG		TT+TG	GG	
<8 (n=61)	16	27	18	P=0.822	16	45	P=0.467	43	18	P=0.397
≥8 (n=238)	58	116	64		58	180		174	64	
	PIN3				PIN3			PIN3		
	WT	Htz	DUP		WT	Htz+Dup		Wt+Htz	Dup	
<8 (n=61)	40	20	1	P=0.764	40	21	P=0.416	60	1	P=0.421
≥8 (n=238)	150	80	8		150	88		230	8	

The MDM2SNP309 and PIN 3 polymorphisms did not show any significant association with the tumor size in the entire sarcoma cohort and in the OGS cohort.

Table 5.6: Association of tumor size in OGS with various genotypes

Tumor size in OGS cohort									
	P72R			Chi-square (df=2)	P72R		Chi-square (df=1)	P72R		Chi-square (df=1)
	CC	GC	GG		CC	CG+GG		CC+CG	GG	
<8 (n=32)	8	13	11	P=0.151	8	24	P=0.341	21	11	P=0.089
≥8 (n=164)	33	96	35		33	131		129	35	
	MDM2SNP309				SNP309			SNP309		
	TT	TG	GG		TT	TG+GG		TT+TG	GG	
<8 (n=32)	8	16	8	P=0.981	8	24	P=1.0	24	8	P=1.0
≥8 (n=164)	43	78	43		43	121		121	43	
	PIN3				PIN3			PIN3		
	WT	Htz	DUP		WT	Htz+Dup		Wt+Htz	Dup	
<8 (n=32)	20	12	0	P=0.602	20	12	P=0.406	32	0	P=0.544
≥8 (n=164)	101	58	5		101	63		159	5	

5.5.4 Polymorphism and chemotherapy response

Histological response of ≥90 or <90 seen in the tumor after chemotherapy was analaysed for 134 OGS cases for whom this information was available. No significant association was seen with any polymorphisms.

Table 5.7 Tumor response in OGS cohort

Tumor response in OGS cohort(N=134)									
	P72R			**Chi-square (df=2)**	**P72R**		**Chi-square (df=1)**	**P72R**		**Chi-square (df=1)**
	CC	**GC**	**GG**		**CC**	**CG+GG**		**CC+CG**	**GG**	
≥90 (n=72)	16	41	15	P=0.504	16	56	P=1.0	57	15	P=0.317
<90 (n=62)	14	30	18		14	48		44	18	
	MDM2SNP309				**SNP309**			**SNP309**		
	TT	TG	GG		TT	TG+GG		TT+TG	GG	
≥90 (n=72)	18	34	20	P=0.683	18	54	P=0.562	52	20	P=1.0
<90 (n=62)	19	25	18		19	43		44	18	
	PIN3				**PIN3**			**PIN3**		
	WT	**Htz**	**DUP**		**WT**	**Htz+Dup**		**Wt+Htz**	**Dup**	
≥90 (n=72)	45	25	2	P=0.889	70	2	P=1.0	45	27	P=.858
<90 (n=62)	40	21	1		61	1		40	22	

5.5.5 Polymorphism and distant metastasis

Information on distant metastasis was available for 285 cases. Of these 117/285 who had metastasis 57 had metastasis at presentation and 60 patients developed metastases at follow up. Association of all these polymorphisms were analysed with the occurrence of metastasis in the entire sarcoma cohort (table 5.8) and in OGS cohort.

No significant association was identified. Metastases free survival calculated with Kaplan Meir method also did not show any significant differences between different genotype (Figure 5.6).

Table 5.8: Association of metastasis occurrence with polymorphisms

SNP association with metastasis									
	P72R			**Chi-square (df=2)**	**P72R**		**Chi-square (df=1)**	**P72R**		**Chi-square (df=1)**
Mets	**CC**	**GC**	**GG**		**CC**	**CG+GG**		**CC+CG**	**GG**	
Yes (n=117)	23	68	26	P=0.512	23	94	P=0.327	91	26	P=0.486
No (n=168)	38	86	44		38	130		124	44	
	MDM2SNP309				**SNP309**			**SNP309**		
	TT	TG	GG		TT	TG+GG		TT+TG	GG	
Yes (n=117)	37	47	33	P=0.237	37	80	P=0.075	84	33	P=0.786
No (n=168)	37	87	44		37	131		124	44	
	PIN3				**PIN3**			**PIN3**		
	WT	**Htz**	**DUP**		**WT**	**Htz+Dup**		**Wt+Htz**	**Dup**	
Yes (n=117)	73	41	3	P=0.977	73	44	P=1.0	114	3	P=1.0
No (n=168)	105	58	5		105	63		163	5	

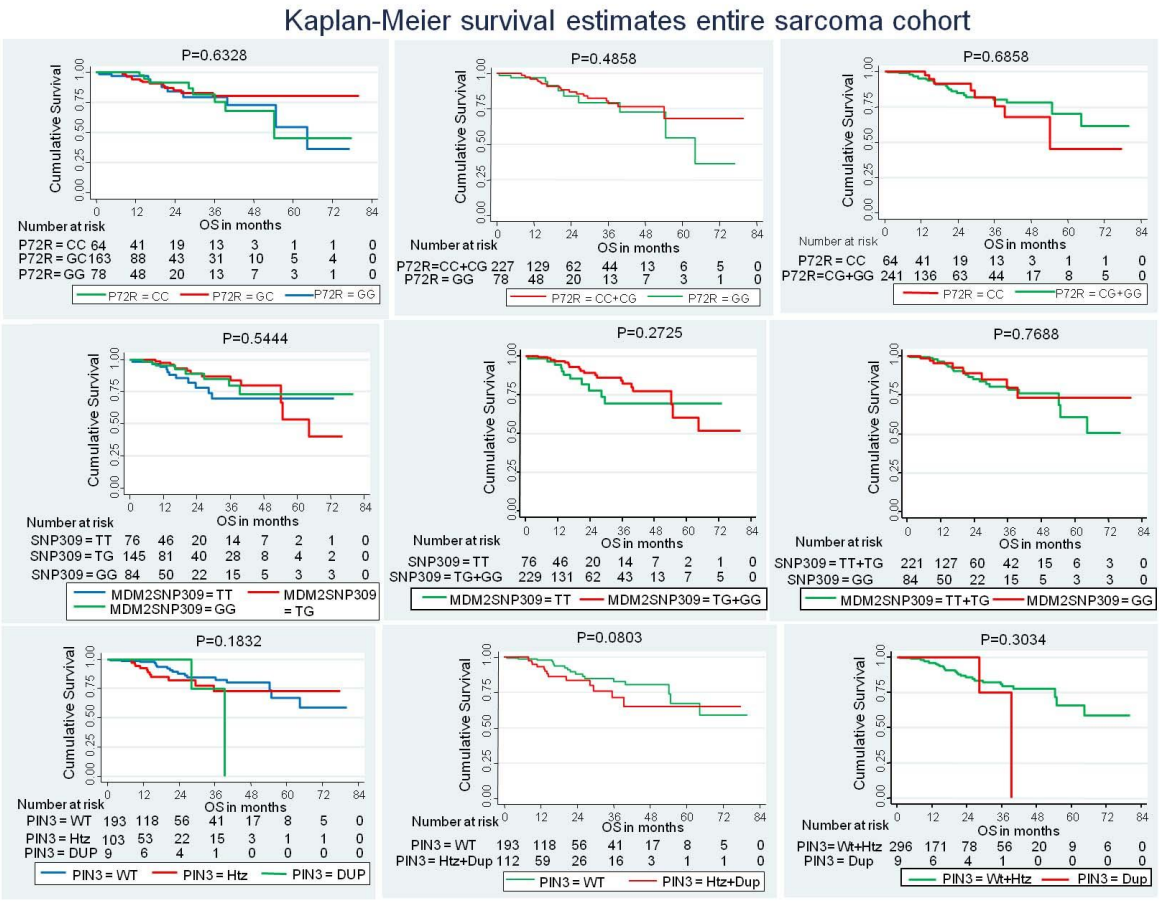


Figure 5.5: Overall survival with three polymorphisms: Survival curves are shown in 3 rows and 3 coloumns. First row is for P72R, second for MDM2 SNP309 and the third for PIN3. First column is for all 3 gennotypes, 2nd and 3rd for recessive and dominant models of the three polymorphisms.

5.5.6 Polymorphisms and overall survival

Information on survival was available for 305 cases in whom date of death or date of last follow up was recorded. Association of these polymorphisms with overall survival was studied for the entire sarcoma cohort and for OGS group. P72R and MDM2SNP309 were not found to be associated with Overall Survival (figure 5.5).

The PIN3 duplication polymorphism was associated with a highly significant inferior overall survival in OGS cases and a trend for inferior survival was seen in the entire sarcoma cohort. In 76 OGS cases with a Homoygous or Heterozugous intron 3

duplication, the 3 year probability of survival was 57.9% as compared to 82.6% for the 122 OGS cases wild type for this polymorphism (p=0.009). No significant association was identified with MFS but similar trend was seen (figure 5.6).

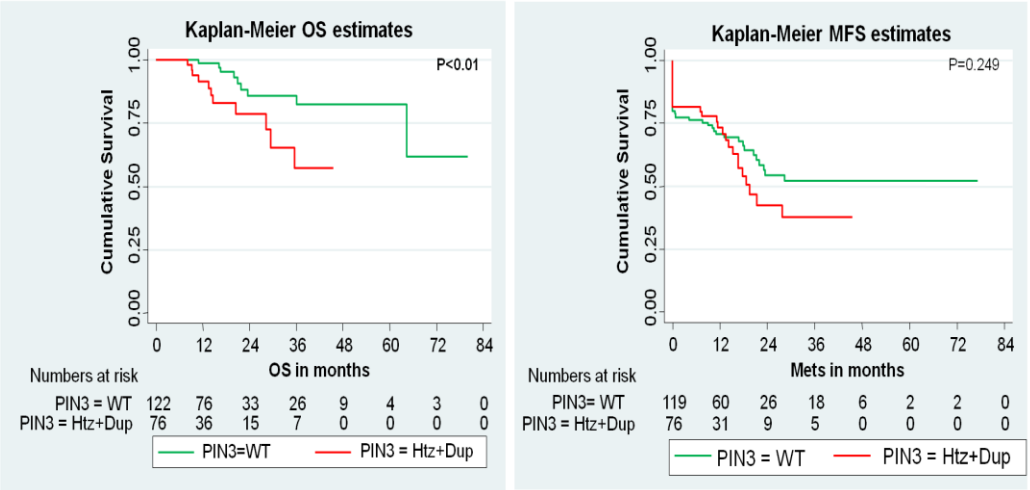


Figure 5.6: Overall and metastatic free survival in OGS cases. OS- Overall Survival; MFS-Metastatic Free Survival

5.6 Discussion

In our study, we did not genotype healthy controls, association of these three *TP53* / MDM2 polymorphisms with the risk of sarcoma development could not be examined. However, this is one of the largest sarcoma cohort where the modifier effect of these polymorphism on the clinical phenotype (age at diagnosis and tumor size), chemotherapy response and survival outcome has been studied. While a trend towards association between certain genotypes and clinical phenotypes were observed in our sarcoma cohort, they did not reach the predefined significance level of $p<0.01$, except for the PIN3 duplication. This is the first reported association of highly significant worse overall survival with PIN3 duplication in Osteosarcoma.

Jian-Ying Ru et al examined 5 *TP53* SNPs for risk association and survival in Chinese osteosarcoma patients. They reported a shorter median survival time of 19 months in patients with P72R GG genotype as compared to 28 months in patients with CC genotype with a hazard rate of 1.94 (95% confidence interval 1.03-3.65). However the PIN3 duplication polymorphism was not studied by them (100). Another study by Ohnstad HO et al. did not find any association of P72R and MDM2 with the survival (96). In our cohort also we did not find any significant association of P72R and MDM2SNP309 with the survival of sarcoma cases.

Though PIN3 polymorphism has never been studied in sarcoma, there are reports with its risk association with other cancers (244, 245). Functional studies on this polymorphism show that the duplication polymorphism affects the expression of p53 protein. More expression of full transcript of *TP53* protein was observed in the presence of wild type allele (246). Duplication of PIN3 also leads to haploinsufficiency of *TP53*. Due to addition of 16bp the length of intron 3 increases by about 16% which stabilize the formation of G-quadruplexes (247). This results into the less expression of the protein which may explain the worst survival in the duplication polymorphism. A meta-analysis also supports the finding that the duplication polymorphism is associated with increased risk of cancer. Geographically this association was found in Indian, Mediterranean and Northern Europe populations but not in the Caucasian population of the United States. In this meta-analysis with cancer type, the increased risk was found with breast cancer and colorectal cancer but not with lung cancer (248). All these studies supports our finding and to further confirm this finding there is a need for detailed functional studies of the effect of PIN3 duplication in the OGS cell lines, sarcospheres and animal models.

The observations from our study highlight the need of further studies in larger cohort of patients with different types of sarcoma and with longer follow up. This could help in confirming the progrnostic role of PIN3 duplication polymorphism in OGS, other types of sarcomas and other cancers.

Chapter 6

Establishing lymphoblastoid cell lines from LFS *TP53* mutation carriers for DNA repair studies and characterization of variants

6.1 Introduction

Establishment of human cell lines in mid twentieth century is considered as a milestone in biological research as it allowed mechanistic and functional studies on human pathologies. One of the first human cell line to be successfully established was in 1951 when George Gey at Johns Hopkins University established a cell line from the cervical cancer tissue of a patient named Henrietta Lacks. Though the donor of the tumor tissue was never remembered but her cell line was named as HeLa (Henrietta Lacks). The HeLa cell line growing all over the world is perhaps the most extensively used cell lines for cancer research (249). After the establishment of HeLa cell line various other tumor cell lines were established. In 1964, Epstein and Barr for the first time found that a herpes like virus was present in the lymphoblast cultures established from Burkitt's lymphoma patient (250). Subsequently, this particle was seen in other studies with Burkitt lymphoma tumors (251-253). This particle was later known as EBV, which is antigenically different from other herpes viruses. The frequent association of this virus with lymphoblastoid cells in long term cultures raised the possibility of the lymphoproliferative nature of this virus. Co-culture experiment with irradiated Burkitt's lymphoma cells with leukocytes provided evidence, that these viruses have some stimulatory effect on growth (254). In 1968 Pope et al showed the transformation ability of these viruses by transforming foetal human leukocyte (255) and in 1969 Gerper et al showed the transformation of normal human leukocyte cultures (256). These studies established the fact that EBV has transforming capabilities of EBV on human B-lymphocytes and that it can be used for establishing transformed lymphoblastoid cell lines.

EBV has a structure similar to that of herpes viruses (Figure 6.1A). It is round structured with surface glycoproteins, a nuclear envelop, viral tegument, nucleocapsid, and genetic material. It belongs to the γ herpesvirus subgroup. The genetic material is linear dsDNA as found in other herpes viruses. Its genome is about 170kb long with around 94 genes. In the latent phase it remains in the circular form as shown in figure 6.1B. Important genes of this virus are EBNA, LMP, EBER, BARF.

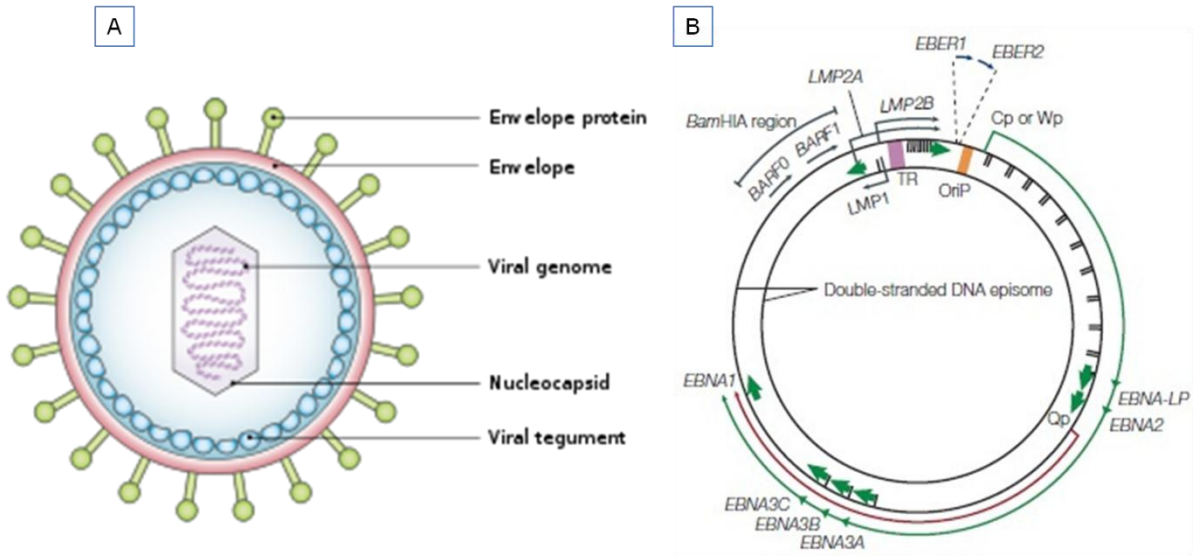


Figure 6.1: EBV particle (A) and its genome(B) image taken from-(https://www.genome.jp/dbget-bin/www_bget?refseq:NC_007605)

These viruses can infect both B-Lymphocytes and epithelial cells but the underlying mechanisms to invade both the cells are different. To infect B lymphocytes the viral glycoprotein gp350 encoded by BLLF1 gene binds with the C21 B-cell receptor. After this interaction gp42 a membrane protein binds with MHC class II molecules which trigger the fusion of viral particle with the B-cell allowing the viral genome to enter the B-cell. In CD21 negative cells gp350/220 interacts with CD35 which is also

known as complement receptor1 to invade the cells (257). EBV invades epithelial cells by a different set of proteins. EBV surface protein BMRF2 interacts with β1 integrins then viral glycoprotein H encoded by the gene BXLF2 and glycoprotein L encoded by BKRF2 interacts with host cell integrins which results in the fusion of the viral particle with the host epithelial cell membrane. This results in the release of viral genome into the epithelial cell. After the entry into the host cell the viral capsid gets dissolved and the dsDNA of virus gets transported to the nucleus of host cell (257).

There are various strains of this virus, of which a strain isolated from a patient with mononucleosis due to transfusion has been used to immortalize monkey blood leukocytes in vitro (258). The virus produced from this cell line was called B95-8 strain and the cell line is marmoset cell line (B95-8 cell line). DNA profiling later revealed that this cell line has most probably originated from Cotton-Top Tamarian Monkey's peripheral blood lymphocytes (259). This cell line is a large producer of EBV and currently used for production of the virus for transforming and establishing Lymphoblastoid cell lines. Lymphoblastoid cell lines can be established with standard protocols as described in the methodology chapter. These cell lines have been extensively used in a wide variety of research, including cancer research.

6.2 Utility of Lymphoblastoid cell lines

Lymphoblastoid cell line (LCL) became the choice for various types of studies as they are relatively easy to establish and maintain and serve as a continuous source of precious genetic material including RNA. In the last few decades these cell lines have been extensively used for genetic studies, functional studies, cytotoxicty assays and for genotype phenotype correlation (260). In a review of published literature from 2001 – 2010, Hussain et al showed that over this period of 10 years, LCLs was used

as a model system in 65% research studies, and in the remaining studies they were used for understanding basic virus biology (Figure 6.2A). During the same period 15% of the studies were related to DNA damage and repair, 29% for cytotoxic studies and the remaining 56% for other functional studies (Figure 6.2B) (261). Recently, Satish et.al. have shown the utility of LCLs for generating induced pluripotent stem cells (262).

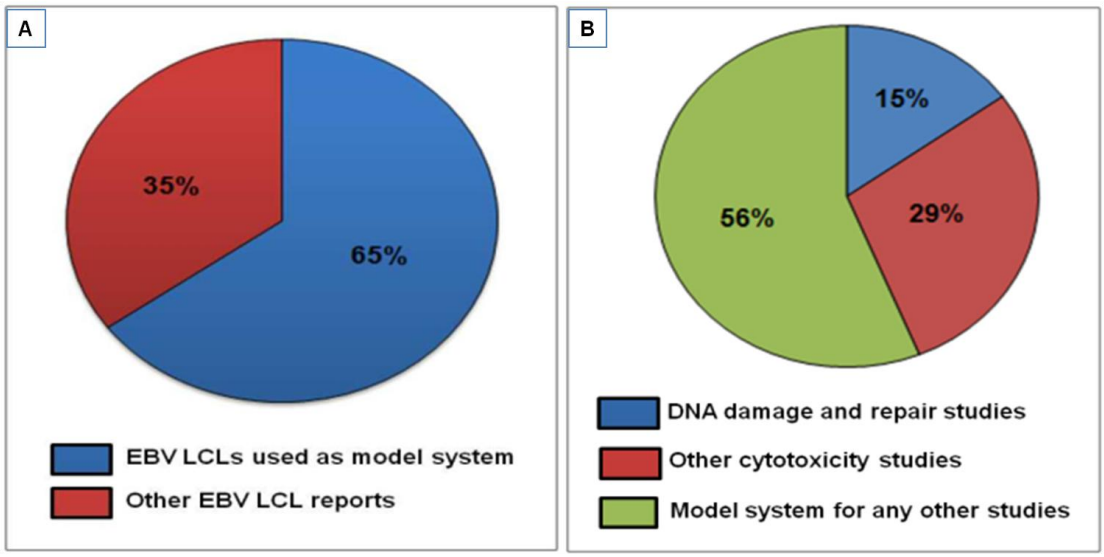


Figure 6.2: Use of Lymphoblastoid cell lines in various studies. Taken from Reference (261)

6.3 DNA damage and repair

The integrity of DNA, the store house of genetic information, has to remain intact. Alteration in DNA could result in changes in the cellular make up and function. DNA is a very dynamic molecule in cell which undergoes replication in every cell division, and all expressing genes in the respective cell synthesize their RNA from the DNA molecule keeping it functioning every time. This chemical entity is prone to get damaged by various means and the DNA of a single cell may be affected by one

million changes per day (263). In every cell division the DNA makes a copy of itself by the replication process and this process is not error free. Despite several proof reading and repair mechanisms, few residual mutations may remain in the DNA. The mutation rates vary for different genes and in different organisms (264).

The DNA damage can be divided into two types (i) Endogenous damage due to intracellular events e.g. replication errors and damages due to Reactive Oxygen Species. (ii) Exogenous damage- Damage occurs due to external agents like UV light, ionizing radiation and various chemicals which forms adducts with DNA.

The DNA damage repair process takes place in a stepwise manner. Firstly, the cell senses the damage in the DNA and transcribes various proteins to halt the cell cycle. Secondly proteins at repair site get phosphorylated and various signaling cascades start to recruit various other proteins at the damaged site which helps in preparing the endogenous cellular environment for repair to occur. Finally the damaged DNA is repaired. If the damage is irreparable then the signaling cascade of apoptosis starts and ultimately leads to the death of the cell (265). There are various repair mechanisms which deals with different types of damages. These include, mismatch repair, base excision repair, nucleotide excision repair and Double Strand Break (DSB) repair. Different DSB repair pathways include the homologous recombination (HR) repair, Non Homologous End Joining repair (NHEJ) and Microhomology mediated End Joining (MMEJ) repair. The choice of DSB repair mechanism depends on various factors. HR mostly takes place in the S phase during the replication process as the cell cannot tolerate error while replication. In the remaining phases, NHEJ takes place. There are other factors also like complexity of the break influences which repair mechanism takes place (266).

Study of DNA damage and its repair has allowed researchers to explore the function of many proteins and their mutant counterparts. In these experiments, first the damage is induced in the cell and the various assays are used to study the kinetics of repair and function of any repair protein.

6.4 Gamma H2AX assay

H2AX is a variant of histone protein H2A. This variant of H2A has a unique motif called SQ motif at the C-terminal end and this motif is highly conserved from plants to humans (267). In 1998 it was first reported by Emmy et al that the serine residue, Ser 139 gets immediately phosphorylated after the DNA damage in human cells (268). After phosphorylation it is called as γH2AX as it was first observed by irradiating the cells with γ-irradiation. After the damage H2AX is phosphorylated by PI3-Kinases like ATM, ATR, DNA-PK depending on the type of damage (268-270). Phosphorylation of H2AX is one of the earliest events after DNA damage which starts with in minutes and is at its highest level 30 minutes after the damage (267). Using γH2AX as a target of research many important findings have been made like the phosphorylation of H2AX at the damaged site and subsequently hundreds to thousands γH2AX molecules surrounds the DSB to form foci which further help in the recruitment of various proteins for the repair process. After the discovery of γH2AX , it was used to discover various other proteins involved in the DNA repair process, through interaction and colocalization.

This assay was used to study DNA repair defects in cell lines with different p53 status. Van Oorsschot et al have reported that the colorectal cancer cell line RC10.1 which is p53 null shows delayed repair with greater number of residual H2AX foci in

comparison to the p53 wild type RKO cell line (271). Many studies have shown the effect of mutant p53 in the cell cycle checkpoint post irradiation (103-105, 108). These studies were done in either fresh peripheral blood lymphocytes or fibroblasts from LFS cases. LCLs from LFS patients have been used in only one mechanistic study, where the *TP53* mutation status did not correlate with G1 arrest (108). However, other studies have shown the effect of *TP53* mutation on cell cycle check points (103, 107). The effect of *TP53* mutation on radiosensitivity of cell lines have been reported in few studies using γH2AX assay or 53BP1 foci formation post irradiation (106, 272, 273). However, the kinetics of DNA damage repair in LCLs established from *TP53* mutation carriers has not been studied. We undertook a pilot study to see the difference in the repair kinetics of *TP53* mutant LCLs and the LCLs established from the healthy individuals post irradiation.

6.5 Purpose of establishing LCLs in LFS cases

LCLs derived from patients with rare and lethal hereditary syndromes could be especially useful if a novel gene variant is later identified in those patients and the subject is not available for obtaining repeated fresh blood samples. Our lab focuses on various hereditary cancers which led us to establish LCLs from different cancer syndromes. These cell lines are a continuous source of genetic material for further studies like characterizing any novel variant and for DNA repair studies.

6.6 Results

6.6.1 LCL establishment

Detailed protocol for LCLs establishment is given in materials and methods chapter from the peripheral blood of 75 inherited cancer syndrome cases and unaffected healthy individuals, 54 lymphoblastoid cell lines were successfully established, with a success rate of 72%. Of these 54 cell lines, 17 LCLs were from LFS families of which 12 had a germline *TP53* mutation and 5 were not carrying any *TP53* germline mutation. The LCL culture at different time points is shown in figure 6.3. Details of the LFS/LFL cases from whom these cell lines were established are provided in table 6.1. Other LCLs successfully established were from families with HBOC, MEN2, HNPCC, Peutz–Jeghers syndrome, Von Hippel–Lindau (VHL) syndromes.

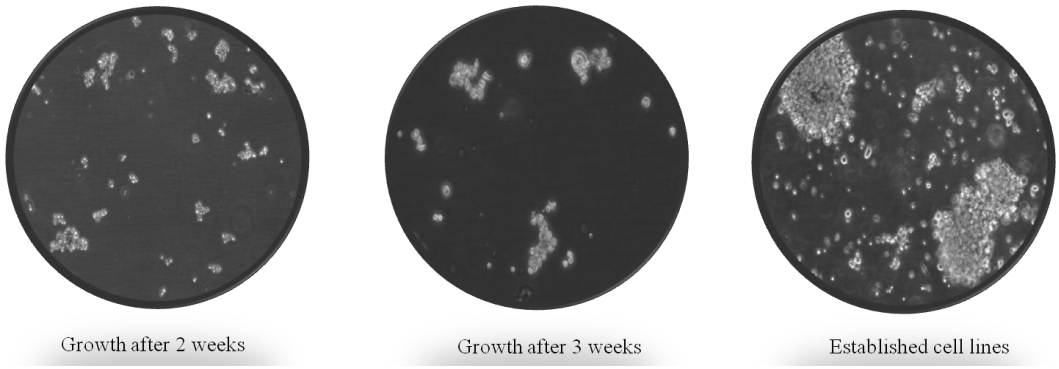


Figure 6.3: EBV transformed LCLs

The LCLs grow in suspension in loose aggregates. The protocol used for establishing the LCLs took 20 days to 40 days for successful establishment of the cell lines. The average population doubling time observed for the cells was 24 hours ranging from 12-36 hours.

Table 6.1: Details of LFS/LFL cases from whom LCLs were established					
Sr. No.	Case	Tumor (Age)	Gender	Mutation in *TP53*	Mutation status
1	1	Osteosarcoma (16)	Female	around 72kb Deletion	LGR¶
2	2	Osteosarcoma (13)	Male	c.672+1G>T	Splice
3	2A	Breast (42)	Female	c.672+1G>T	Splice
4	3	Rhabdomyosarcoma (5)	Male	Duplication of 4 Exons	LGR
5	4	Renal cell carcinoma (41)	Male	Negative	
6	5	Sarcoma (3)	Female	c.846delG fs;282delG fs 344*	Novel Frameshift
7	6	Breast (30)	Female	Negative	
8	6A	Adreno Cortical Carcinoma (4)	Female	Negative	
9	7	Breast (26)	Female	Y236C	Missense Moderate DNE
10	8	Breast (34)	Female	c.902_903insC;	Frameshift

				301insC fs 305*	
11	9	Osteosarcoma (20)	Male	R196*	Nonsense
12	10	Breast (38), Breast (38)	Female	R175H$^{\pm}$	Missense (DNE)
13	11	Ovary (18)	Female	R273C$^{\pm}$	Missense (DNE)*
14	12	Acute lymphoid Leukemia (15), Breast (33)	Female	T125T (c.375G>T)	Silent/Splice Germline
15	12A	Unaffected relative of case #12 (11)	Male	T125T	Silent/Splice
16	12B	Unaffected relative of case #12 (23)	Female	Negative	
17	13	Acute Myeloid Leukemia (14)	Male	Negative	

±DNE- Dominant Negative effect; ¶ LGR- Large Genomic Rearrangement

6.6.2 DNA damage repair study

DNA damage repair study was conducted on LCLs established from four cases (case #5, #10, #11, and #12 in table 6.1) with different germline *TP53* mutation - missense mutations in two, frame shift mutation and splice site mutation in one each. LCLs from 4 healthy individuals were used as controls.

The 4 *TP53* mutant and 4 healthy control LCLs were exposed to 2 Gy irradiation (2Gy), and then fixed at different time points of 1hr, 4hr, 24hr and 48hr. For every time point, LCLs not exposed to radiation (0Gy) were taken as controls. Fluorescent intensity of 30-50 nuclei was recorded from 6-7 fields and their mean intensity was calculated. A significant difference in the repair kinetics between the mutation carriers and the controls was observed. In the *TP53* mutant LCLs the DNA repair was delayed with significantly higher mean floursecent intensity at 24 and 48 hours. Figure 6.4 showed the representative image of case and control. Residual foci are clearly visible in LCLs with *TP53* mutation even after 48 hour.

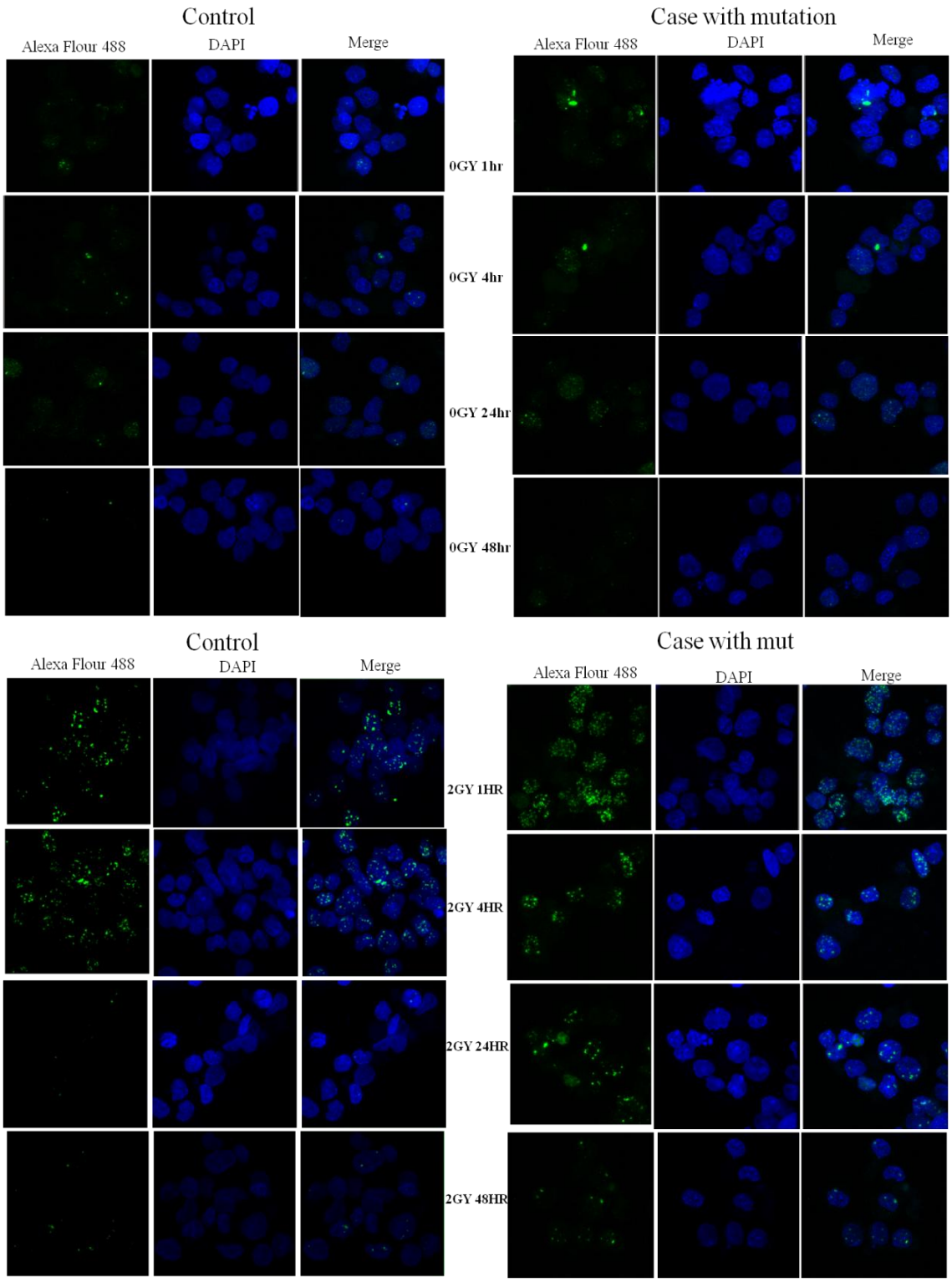


Figure 6.4: Gamma H2AX assay of a case and control at different time points. The case has the mutation R273C. Upper panel-without radiation (0Gy). Lower panel- with γ-irradiation (2Gy).

The data was analyzed by combining the MFI of all the 4 *TP53* mutant LCLs and compared with the 4 *TP53* wild type LCL controls. At 24 and 48 hour, there is significant higher MFI in the *TP53* mutatnt LCLs indicating delayed DNA damage repair (Figure 6.5). At 24 hour the MFI of cases is around 24 and in controls is around 16 and the difference is significant (P<0.001). At 48 hour the MFI of cases is around 23 and of controls is around 15 showing the presence of more fluorescence in cases representing residual foci with a significant difference (P<0.0001) as shown in figure 6.5 and the table of standard error of mean in table 6.2.

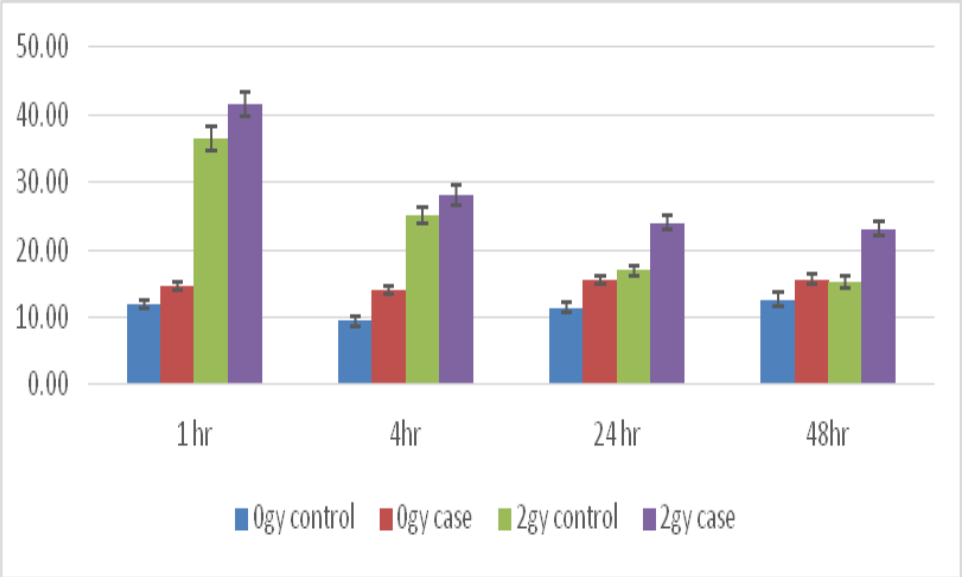


Figure 6.5: Mean fluorescent intensity at different time points.

Table 6.2: The value of standard errors:

Standard Errors				
	1 hr	**4hr**	**24 hr**	**48hr**
0gy control	0.6458	0.7252	0.7341	1.0121
0gy case	0.6275	0.5446	0.5123	0.6935
2gy control	1.8599	1.1909	0.8309	0.8829
2gy case	1.7819	1.4411	1.1063	1.0838

6.6.3 Characterization of a germline*TP53* variant (T125T)

In our LFS/LFL cohort we screened a family which fulfilled the Chompret criteria of LFL for *TP53* mutation. Multiple members of the family were affected with various types of tumors and the proband was affected with two primary cancers, one Acute Lymphoid Leukemia (ALL) at the age of 15 and then breast cancer at the age of 31 years (figure 6.6). With such a strong history of LFS spectrum cancer the family was tested for *TP53* mutation.

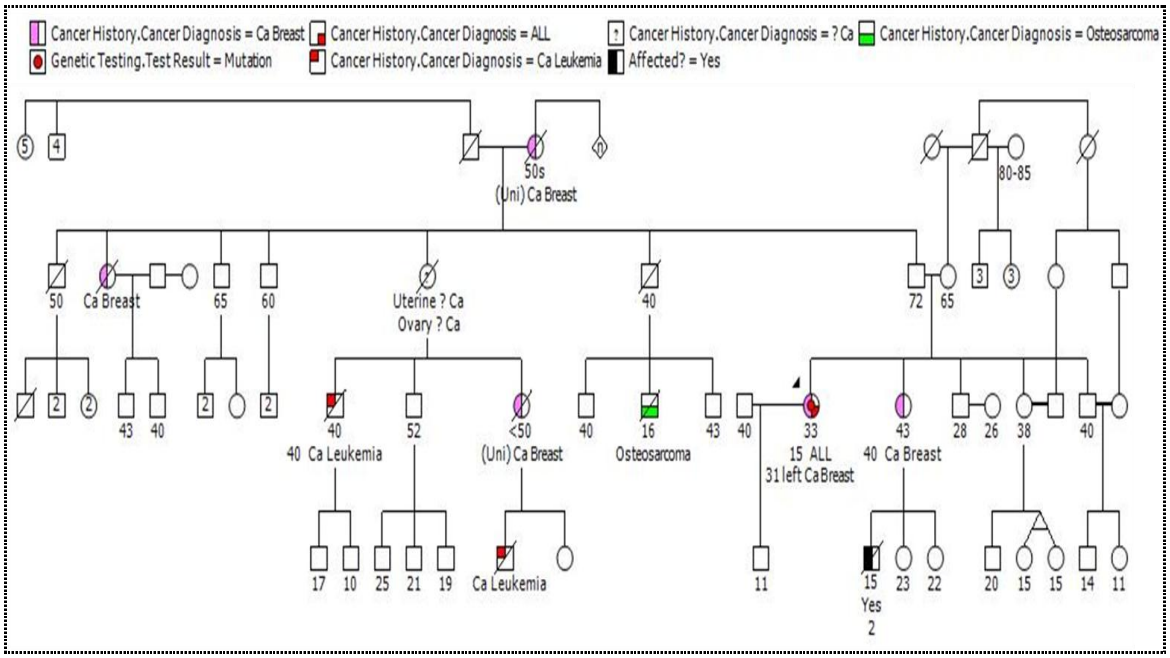
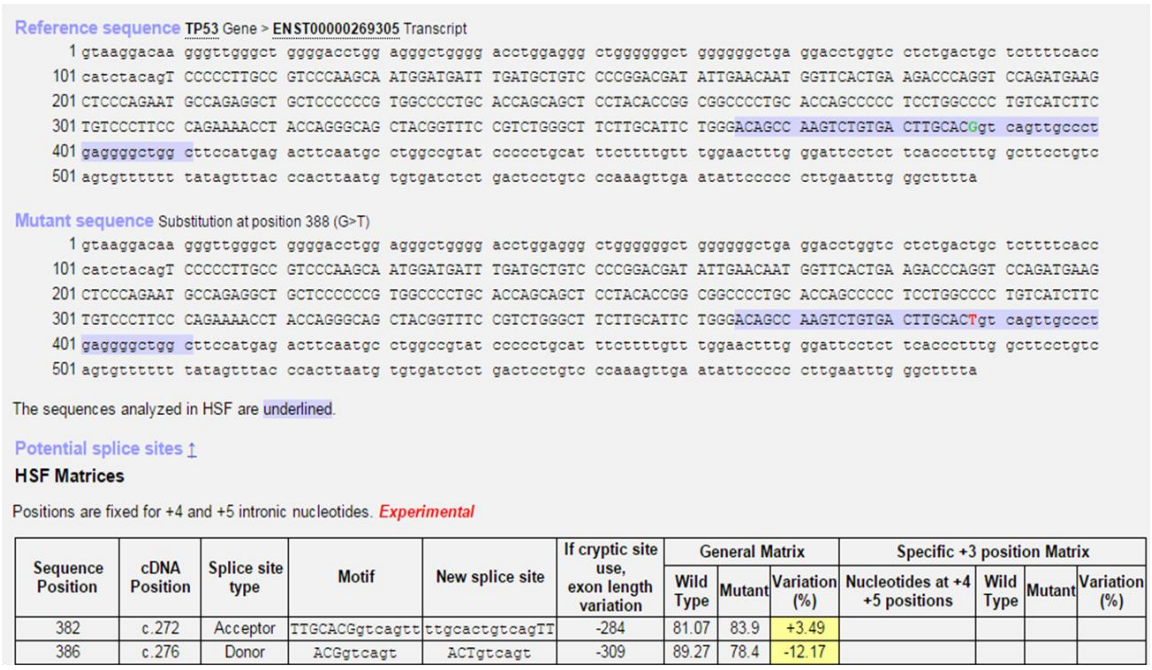


Figure 6.6: Pedigree of T125T mutation carrying family (LFL-Group 2)

A variant T125T (c.375G>T), *TP53* was identified in this family. This variant is a silent mutation. No other mutations were identified in the *TP53* gene. This is a reported mutation and a study characterized the variant to affect splicing (274). Using the LCLs available from T125T mutation carrier we independently characterized this variant to see if it causes aberrant splicing and is pathogenic.. The variant was also

found to be co-segregating with the disease in this family. To address this, first in silico prediction of the variant by Human Splice Finder (271) was carried out. The prediction scoring showed that the variant may affect the splicing of the *TP53* transcript (Figure 6.7). To confirm the finding of in silico prediction, the variant was functionally characterized.



Reference sequence TP53 Gene > ENST00000269305 Transcript

```
  1 gtaaggacaa gggttgggct ggggacctgg agggctgggg acctggaggg ctgggggggct gggggggctga ggacctggtc ctctgactgc tcttttcacc
101 catctacagT CCCCCTTGCC GTCCCAAGCA ATGGATGATT TGATGCTGTC CCCGGACGAT ATTGAACAAT GGTTCACTGA AGACCCAGGT CCAGATGAAG
201 CTCCCAGAAT GCCAGAGGCT GCTCCCCCCG TGGCCCCTGC ACCAGCAGCT CCTACACCGG CGGCCCCTGC ACCAGCCCCC TCCTGGCCCC TGTCATCTTC
301 TGTCCCTTCC CAGAAAACCT ACCAGGGCAG CTACGGTTTC CGTCTGGGCT TCTTGCATTC TGGGACAGCC AAGTCTGTGA CTTGCACGgt cagttgccct
401 gaggggctgg cttccatgag acttcaatgc ctggccgtat cccctgcat ttcttttgtt tggaactttg ggattcctct tcaccctttg gcttcctgtc
501 agtgtttttt tatagtttac ccacttaatg tgtgatctct gactcctgtc ccaaagttga atattccccc cttgaatttg ggctttta
```

Mutant sequence Substitution at position 388 (G>T)

```
  1 gtaaggacaa gggttgggct ggggacctgg agggctgggg acctggaggg ctgggggggct gggggggctga ggacctggtc ctctgactgc tcttttcacc
101 catctacagT CCCCCTTGCC GTCCCAAGCA ATGGATGATT TGATGCTGTC CCCGGACGAT ATTGAACAAT GGTTCACTGA AGACCCAGGT CCAGATGAAG
201 CTCCCAGAAT GCCAGAGGCT GCTCCCCCCG TGGCCCCTGC ACCAGCAGCT CCTACACCGG CGGCCCCTGC ACCAGCCCCC TCCTGGCCCC TGTCATCTTC
301 TGTCCCTTCC CAGAAAACCT ACCAGGGCAG CTACGGTTTC CGTCTGGGCT TCTTGCATTC TGGGACAGCC AAGTCTGTGA CTTGCACTgt cagttgccct
401 gaggggctgg cttccatgag acttcaatgc ctggccgtat cccctgcat ttcttttgtt tggaactttg ggattcctct tcaccctttg gcttcctgtc
501 agtgtttttt tatagtttac ccacttaatg tgtgatctct gactcctgtc ccaaagttga atattccccc cttgaatttg ggctttta
```

The sequences analyzed in HSF are underlined.

Potential splice sites ↕

HSF Matrices

Positions are fixed for +4 and +5 intronic nucleotides. *Experimental*

Sequence Position	cDNA Position	Splice site type	Motif	New splice site	If cryptic site use, exon length variation	General Matrix			Specific +3 position Matrix			
						Wild Type	Mutant	Variation (%)	Nucleotides at +4 +5 positions	Wild Type	Mutant	Variation (%)
382	c.272	Acceptor	TTGCACGgtcagtt	ttgcactgtcagTT	-284	81.07	83.9	+3.49				
386	c.276	Donor	ACGgtcagt	ACTgtcagt	-309	89.27	78.4	-12.17				

Figure 6.7: Insilico prediction for aberrant splicing by Human Splice Finder (275).

To study the aberrant splicing due to this variant we revived the LCLs of this case and a healthy control. In normal condition the p53 expression may be low. Hence we irradiated the cells with Gamma irradiation from ^{60}Co. One hour prior to irradiation the cells were treated with puromycin to block the NMD (276). Irradiated cells were incubated in CO_2 incubator for 3 hours and then RNA was extracted as the accumulation of p53 in the cell after any stress is reported to be maximum from 1-24 hour (277, 278). The quality and quantity of RNA was checked on Nano drop and

agarose gel electrophoresis as shown in left gel of figure 6.8. About 2.5 µg of RNA was taken for first strand synthesis of cDNA using Amersham 1st strand synthesis kit .

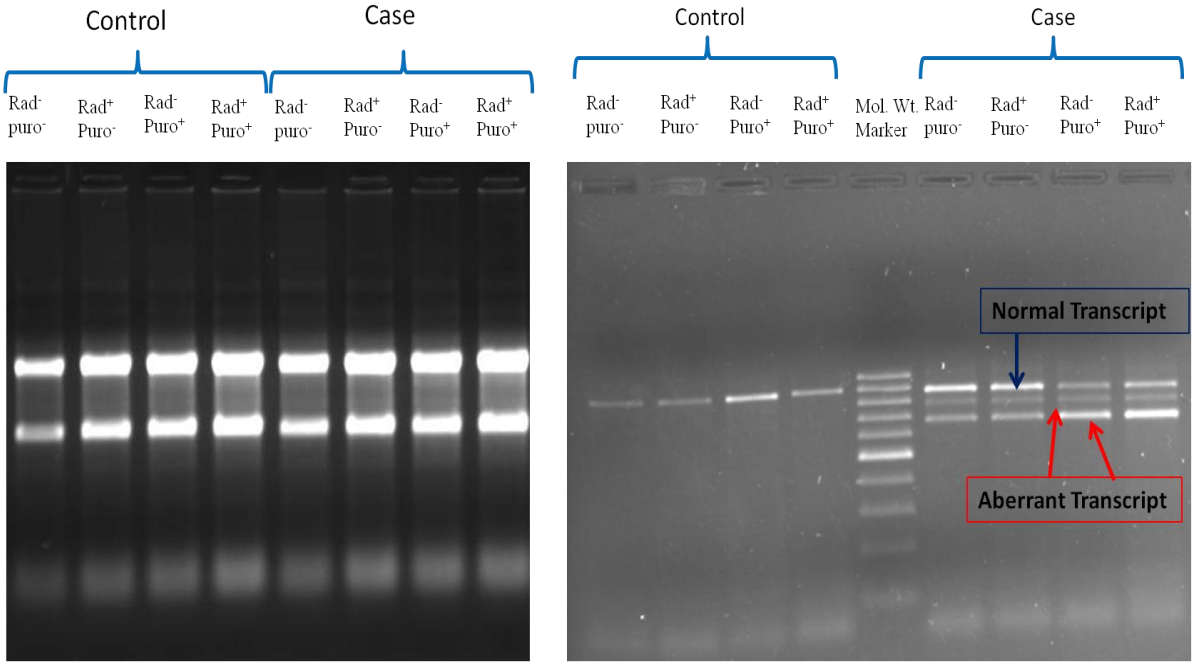


Figure 6.8: Agarose gel showing RNA yield (Left gel) **and aberrant spliced product** (Right gel)

The cDNA was amplified by using a primer set (sequence given in Materials and Methods) designed to amplify exons 4-10. The forward primer was designed at the start of exon 4 and the reverse primer was located at the junction of exon 9 and 10, amplifying a cDNA product of size 877 bp. The amplified products shown in right gel of figure 6.8 indicates three bands of cDNA which are the normal cDNA and the smaller sized aberrantly spliced cDNA between exons 4-10 of *TP53*.

The PCR amplicons electrophoresed on agarose gels were cut, eluted and sequenced to confirm the splicing defect. Sequencing results showed a deletion of 200 bp in the

aberrant transcript. Detailed analysis revealed that due to mutation in the last base of exon the splice donor site was lost and a cryptic splice site was created (red font) in figure 6.9 from which the alternative

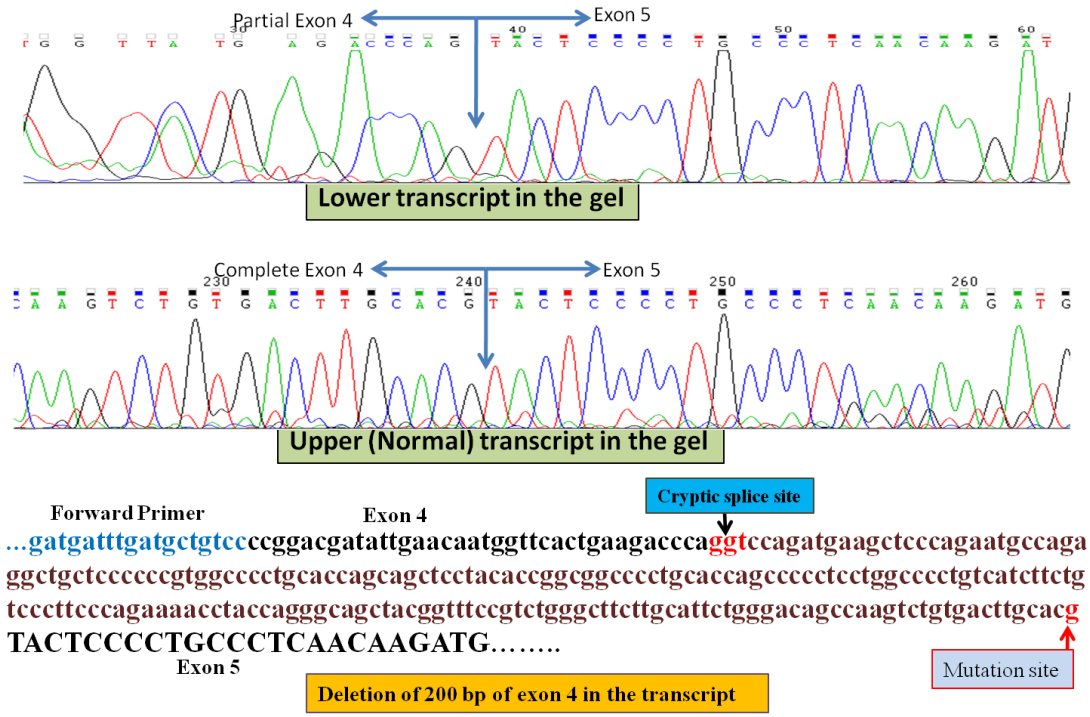


Figure 6.9: Sequencing data to confirm aberrant splicing

splicing occurred leading to a loss of 200bp of exon 4 in the transcript. *In silico* analysis of the aberrant sequence was carried by Expert Protein Analysis System (ExPASy) translate tool (www.web.expasy.org). In silico translation result showed a changed amino acid sequence after the cryptic splice site and protein truncation due to a premature stop codon 44 amino acids downstream of the cryptic splice site (figure 6.10).

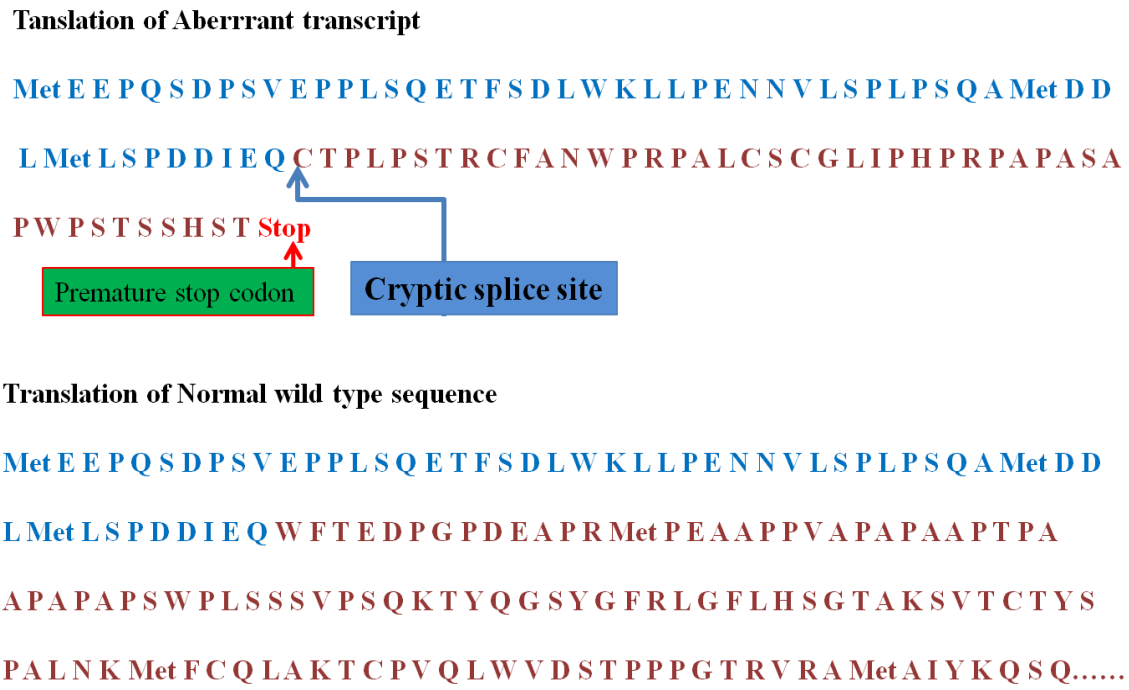


Figure 6.10: Translation of the aberrant transcript using ExPASy translate tool (https://web.expasy.org/translate/)

6.7 Discussion

In this study, a large panel of LCLs from healthy controls and LFS patients with different germline *TP53* mutations was successfully established. Of several methods that could be used for characterizing novel variants identified in cases with rare syndromes, best results could be obtained by performing functional studies on growing cells or on mRNA which not only has the same uncharacterized variant but also the same genetic background. LCLs derived from patients with rare and lethal hereditary syndromes could be especially useful if a novel gene variant is later identified in those patients. Moreover, these LCLs could be used for other mechanistic studies on *TP53*, cytotoxicity assays and other aspects of cancer biology. While genomic mutations are reported to accumulate in cultured cells, LCLs are found to be relatively stable for mutation accumulation (277-281). Due to their

relatively stable nature LCLs were used in various large genomic studies like Human Genome Diversity Project, Hap Map Project (282) and 1000 genome project (http://www.internationalgenome.org/category/cell-lines/). Of the LFS cases from whom we could establish LCLs, one individual harboured novel germline *TP53* variants. The novel variant in the LFS case was a frameshift mutation where in silico prediction of a truncated protein is sufficient to characterize it as a likely pathogenic mutation. However the second germline variant (c.375G>T)was a point mutation resulting in a synonymous change. This variant was in the splice site region and in silico analysis predicted that it may affect *TP53* splicing as reported earlier (274). Further evidence from functional studies was desirable as the prediction of aberrant splicing may have been hazardous since aberrant splicing would require activation of a cryptic splice site (275). We could clearly establish aberrant splicing caused by this variant by cDNA sequencing. This required repeated studies over a period of one year which would have been difficult without the LCLs from this case as a source of RNA. Earlier studies have used fibroblasts from LFS patients to study chromosomal aberration, genomic instability and DNA repair. For making fibroblasts cell lines, invasive skin biopsy is required where as LCLs can be easily established from peripheral blood lymphocytes (PBL). Fresh PBLs from LFS cases have also been used for various studies (272). However, unlike LCLs, fresh PBLs have much limited utility and can be used for very limited experiments conducted soon after their collection (106, 272).

In this pilot study we found the significant difference between the DNA repair kinetics of LCLs with germline *TP53* mutations as compared to LCLs from healthy controls with wild type *TP53*. The model system established through this study can

be used for detailed investigation of *TP53* mutation status with cell cycle check points, apoptosis and DNA damage repair mechanisms. This is particularly relevant as earlier reports have shown conflicting results. Using LCLs from LFS cases, Gilchrist et al did not find correlation between the *TP53* mutation status and cell cycle arrest. However Williams et al found *TP53* mutation results in cell cycle arrest. Boyle et al reported that lesser percentage of *TP53* mutation positive fibroblast cells from LFS patients underwent G1 arrest in comparison to fibroblast cell line from healthy controls. These conflicting results may be due to different type of mutations of *TP53* gene. It is well known that different *TP53* missense mutations affect the gene function differently. Dominant negative mutations have more severe effects in comparison to other type of missense mutations. In 2017 Suchankova J et al have shown that the accumulation of 53BP1 at the damage site varies with different types of *TP53* mutations (273). They have shown that in R273C mutants the accumulation occurs during 10-30 minutes post irradiation, in R282W it occurs immediately after radiation and in L194F mutants it occurred during 60-70 minutes. We have established LCLs from various types of mutations like missense, nonsense, Frameshift, Splice site and LGRs in *TP53*gene and can be used as model systems to study several questions related to cell cycle check points, DNA damage repair kinetics and apoptosis. The effect of different *TP53* mutation can be examined individually and previously uncharacterized variants in the *TP53* gene can be subjected to detailed functional studies, some of which may result in establishing their pathogenicity.

Chapter 7

Summary and Conclusion

Comprehensive genotype-phenotype correlation in this first Indian cohort and the largest Asian LFS/LFL cohort and comparison with other populations allow us to draw important conclusions on this syndrome and the TP53 germline mutation. The tumor spectrum as well as the mutation spectrum in the Indian LFS cohort is quite distinct from all three major populations in the IARC databases – Caucasians, Central and South American and rest of Asia.

Of the 59 distinct TP53 mutations identified in this Indian cohort, 40% have never been reported as germline mutation in any other study or database. When compared to the global IARC database, the Indian cohort has significantly higher frequency of LGR mutations and relative abundance of certain cancers like head neck, gynecological, and hematological cancers and absence or rarity of skin and adrenocortical cancers. A clear association between dominant negative mutations and earlier age of cancer diagnosis was seen in our cohort. This underscores the need for studying each population as the mutation as well as the tumor spectrum may be very different.

This study also brings out the need for comprehensive genetic analysis including large genomic rearrangements assays and when to suspect and confirm allele drop outs. The findings of this study highlight the need for carefully designing primers and avoiding polymorphisms with high allelic frequency in their annealing region.

The very high *TP53* mutation detection rate of ~90% in classical LFS and ~30% in LFL families is a reflection of the detailed phenotypic and family data obtained in the cohort and strict criteria used for classification of the LFS and LFL syndrome. The major contribution of this study is the identification of a group of patients who do not fulfill any LFL criteria but have a mutation detection rate of >10%. In our 74

suspected LFL group 3A cases the mutation detection rate was 13.5%**.** Suspected Li Fraumeni Like (sLFL) Group 3A cases included proband with early onset (<50 yrs) *Modified LFS spectrum cancer with 1st, 2nd or 3rd degree relative with any cancer at any age OR a proband with multiple primary cancers, of which at least one is a *Modified LFS spectrum cancer <50 years. *Modified LFS spectrum cancers: Includes *Haemato-Lymphoid malignancies and *Malignant Phyllodes in addition to the previously described LFS spectrum cancers - ACC, Sarcoma, CNS, Leukemia and Breast cancer. By expanding the LFL spectrum cancers to include Malignant Phyllodes and any Haematolymphoid malignancy and by relaxing the age cut off from <46 years (Chompret) to <50 years, we could identify TP53 mutation in 13.5% cases who would have been missed even by the relaxed LFL Eeles and Chompret 2015 criteria. Similarly we identified a high mutation detection rate of 55% in the Malignant Phyllodes tumour of the breast. Based on our study we suggest that the current LFL criteria could be revised to include patients with an modified LFS spectrum tumour at an age below 50 years with a family history of any cancer. Similar to other rare tumours such as adrenocortical carcinoma and choroid plexus tumour which are considered in the 2015 definition of LFL by Bougeard et al, we suggest that malignant phyllodes under the age of 50 years could be considered as LFS associated cancer and criteria for defining LFL.

Another major finding of our study was the strong adverse impact of the *TP53* intron 3 duplication polymorphism on survival in sporadic sarcoma. This study highlight the utility of Lymphoblastoid cell lines established from LFS cases with different germline *TP53* mutations. These could serve as a model system to study several questions related to *TP53* biology and for functional characterization of novel *TP53* mutations.

Chapter 8

Bibliography

1. Wilson BE. Theory Change in Science: Strategies from Mendelian Genetics. Lindley Darden. Philosophy of Science. 1994;61(1):153-5.
2. Dahm R. Friedrich Miescher and the discovery of DNA. Developmental biology. 2005;278(2):274-88.
3. Crow EW, Crow JF. 100 years ago: Walter Sutton and the chromosome theory of heredity. Genetics. 2002;160(1):1-4.
4. Garrod AE. The incidence of alkaptonuria: a study in chemical individuality. 1902 [classical article]. The Yale journal of biology and medicine. 2002;75(4):221-31.
5. Roll-Hansen N. The holist tradition in twentieth century genetics. Wilhelm Johannsen's genotype concept. The Journal of physiology. 2014;592(11):2431-8.
6. Morgan TH. Random Segregation Versus Coupling in Mendelian Inheritance. Science. 1911;34(873):384.
7. Jones ME. Albrecht Kossel, a biographical sketch. The Yale journal of biology and medicine. 1953;26(1):80-97.
8. Hall K. William Astbury and the biological significance of nucleic acids, 1938-1951. Studies in history and philosophy of biological and biomedical sciences. 2011;42(2):119-28.
9. Watson JD, Crick FH. A structure for deoxyribose nucleic acid. 1953. Nature. 2003;421(6921):397-8; discussion 6.
10. Griffith F. The Significance of Pneumococcal Types. The Journal of Hygiene. 1928;27(2):113-59.
11. Avery OT, MacLeod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. INDUCTION OF TRANSFORMATION BY A D<span

class="sc">ESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. 1944;79(2):137-58.

12. Hershey AD, Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. The Journal of general physiology. 1952;36(1):39-56.
13. Tjio JH. The chromosome number of man. American journal of obstetrics and gynecology. 1978;130(6):723-4.
14. Meselson M, Stahl FW. The Replication of DNA in Escherichia Coli. Proceedings of the National Academy of Sciences of the United States of America. 1958;44(7):671-82.
15. Gautier M. [Fiftieth anniversary of the trisomy 21: return on a discovery]. Medecine sciences : M/S. 2009;25(3):311-5.
16. Guthrie R. Blood screening for phenylketonuria. Jama. 1961;178(8):863-.
17. Maxam AM, Gilbert W. A new method for sequencing DNA. 1977. Biotechnology. 1992;24:99-103.
18. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of molecular biology. 1975;94(3):441-8.
19. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, et al. A polymorphic DNA marker genetically linked to Huntington's disease. Nature. 1983;306(5940):234-8.
20. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor symposia on quantitative biology. 1986;51 Pt 1:263-73.

21. Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM, et al. A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature. 1986;323(6089):643-6.

22. Leiding JW, Holland SM. Chronic Granulomatous Disease. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, et al., editors. GeneReviews((R)). Seattle (WA)1993.

23. Menon AG, Klanke CA, Su YR. Identification of disease genes by positional cloning. Trends in cardiovascular medicine. 1994;4(3):97-102.

24. Shanks N, Greek R, Greek J. Are animal models predictive for humans? Philosophy, ethics, and humanities in medicine : PEHM. 2009;4:2.

25. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.

26. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.

27. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nature medicine. 2013;19(11):1423-37.

28. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer. 2015;136(5):E359-86.

29. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018.

30. Mallath MK, Taylor DG, Badwe RA, Rath GK, Shanta V, Pramesh CS, et al. The growing burden of cancer in India: epidemiology and social context. The Lancet Oncology. 2014;15(6):e205-12.

31. De_Vita VT, Lawrence TS, Rosenberg SA. Cancer : principles & practice of oncology. Philadelphia: Wolters Kluwer Health : Lippincott Williams & Wilkins; 2015.

32. Munjal YP, Association of Physicians of I. API textbook of medicine. Mumbai: Association of Physicians of India; 2012.

33. Kotnis A, Kannan S, Sarin R, Mulherkar R. Case-control study and meta-analysis of SULT1A1 Arg213His polymorphism for gene, ethnicity and environment interaction for cancer risk. British journal of cancer. 2008;99(8):1340-7.

34. Kotnis A, Sarin R, Mulherkar R. Genotype, phenotype and cancer: role of low penetrance genes and environment in tumour susceptibility. Journal of biosciences. 2005;30(1):93-102.

35. Rous P. A Sarcoma of the Fowl Transmissible by an Agent Separable from the Tumor Cells. The Journal of experimental medicine. 1911;13(4):397-411.

36. Javier RT, Butel JS. The history of tumor virology. Cancer research. 2008;68(19):7693-706.

37. Newman JH, Zloza A. Infection: a Cause of and Cure for Cancer. Current pharmacology reports. 2017;3(6):315-20.

38. Rubin H. The early history of tumor virology: Rous, RIF, and RAV. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(35):14389-96.

39. Temin HM, Rubin H. Characteristics of an assay for Rous sarcoma virus and Rous sarcoma cells in tissue culture. Virology. 1958;6(3):669-88.

40. Duesberg PH, Vogt PK. Differences between the ribonucleic acids of transforming and nontransforming avian tumor viruses. Proceedings of the National Academy of Sciences of the United States of America. 1970;67(4):1673-80.

41. Stehelin D, Varmus HE, Bishop JM, Vogt PK. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. Nature. 1976;260(5547):170-3.

42. Krontiris TG, Cooper GM. Transforming activity of human tumor DNAs. Proceedings of the National Academy of Sciences of the United States of America. 1981;78(2):1181-4.

43. Shih C, Shilo BZ, Goldfarb MP, Dannenberg A, Weinberg RA. Passage of phenotypes of chemically transformed cells via transfection of DNA and chromatin. Proceedings of the National Academy of Sciences of the United States of America. 1979;76(11):5714-8.

44. Reddy EP, Reynolds RK, Santos E, Barbacid M. A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. Nature. 1982;300(5888):149-52.

45. Tabin CJ, Bradley SM, Bargmann CI, Weinberg RA, Papageorge AG, Scolnick EM, et al. Mechanism of activation of a human oncogene. Nature. 1982;300(5888):143-9.

46. Taparowsky E, Suard Y, Fasano O, Shimizu K, Goldfarb M, Wigler M. Activation of the T24 bladder carcinoma transforming gene is linked to a single amino acid change. Nature. 1982;300(5894):762-5.

47. Butel JS. Viral carcinogenesis: revelation of molecular mechanisms and etiology of human disease. Carcinogenesis. 2000;21(3):405-26.

48. Brugge JS, Butel JS. Role of simian virus 40 gene A function in maintenance of transformation. Journal of virology. 1975;15(3):619-35.

49. Martin RG, Chou JY. Simian virus 40 functions required for the establishment and maintenance of malignant transformation. Journal of virology. 1975;15(3):599-612.

50. Tegtmeyer P. Function of simian virus 40 gene A in transforming infection. Journal of virology. 1975;15(3):613-8.

51. Lane DP, Crawford LV. T antigen is bound to a host protein in SV40-transformed cells. Nature. 1979;278(5701):261-3.

52. Linzer DI, Levine AJ. Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. Cell. 1979;17(1):43-52.

53. Vogelstein B, Fearon ER, Kern SE, Hamilton SR, Preisinger AC, Nakamura Y, et al. Allelotype of colorectal carcinomas. Science. 1989;244(4901):207-11.

54. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell. 2018;174(4):1034-5.

55. Wintersberger U. The selective advantage of cancer cells: a consequence of genome mobilization in the course of the induction of DNA repair processes? (Model studies on yeast). Advances in enzyme regulation. 1984;22:311-23.

56. Lodish HF. Molecular cell biology. New York, NY: Freeman : Macmillan Higher Education; 2016.

57. Teng CS. Protooncogenes as mediators of apoptosis. International review of cytology. 2000;197:137-202.

58. Kufe DW, Holland JF, Frei E, American Association for Cancer R. Holland-Frei cancer medicine 7 Hamilton, Ont.; Lewiston, NY [distributor]: BC Decker; 2006. Available from: http://online.statref.com/document.aspx?fxid=72&docid=1.

59. Cooper GMHRE. CELL : a molecular approach. [S.l.]: OXFORD UNIV PRESS; 2018.

60. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proceedings of the National Academy of Sciences of the United States of America. 1971;68(4):820-3.

61. Berger AH, Knudson AG, Pandolfi PP. A continuum model for tumour suppression. Nature. 2011;476(7359):163-9.

62. Dyson NJ. RB1: a prototype tumor suppressor and an enigma. Genes & development. 2016;30(13):1492-502.

63. Kinzler KW, Vogelstein B. Cancer-susceptibility genes. Gatekeepers and caretakers. Nature. 1997;386(6627):761, 3.

64. Michor F, Iwasa Y, Nowak MA. Dynamics of cancer progression. Nature reviews Cancer. 2004;4(3):197-205.

65. Hainaut P, Pfeifer GP. Somatic TP53 Mutations in the Era of Genome Sequencing. Cold Spring Harbor perspectives in medicine. 2016;6(11).

66. Khoury MP, Bourdon JC. p53 Isoforms: An Intracellular Microprocessor? Genes & cancer. 2011;2(4):453-65.

67. Soussi T, May P. Structural aspects of the p53 protein in relation to gene evolution: a second look. Journal of molecular biology. 1996;260(5):623-37.

68. Wang P, Reed M, Wang Y, Mayr G, Stenger JE, Anderson ME, et al. p53 domains: structure, oligomerization, and transformation. Molecular and cellular biology. 1994;14(8):5182-91.

69. Soussi T, Caron de Fromentel C, May P. Structural aspects of the p53 protein in relation to gene evolution. Oncogene. 1990;5(7):945-52.

70. Cho Y, Gorina S, Jeffrey PD, Pavletich NP. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science. 1994;265(5170):346-55.

71. Rotter V. p53, a transformation-related cellular-encoded protein, can be used as a biochemical marker for the detection of primary mouse tumor cells. Proceedings of the National Academy of Sciences of the United States of America. 1983;80(9):2613-7.

72. Eliyahu D, Raz A, Gruss P, Givol D, Oren M. Participation of p53 cellular tumour antigen in transformation of normal embryonic cells. Nature. 1984;312:646.

73. Baker SJ, Fearon ER, Nigro JM, Hamilton SR, Preisinger AC, Jessup JM, et al. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. Science. 1989;244(4901):217-21.

74. Finlay CA, Hinds PW, Levine AJ. The p53 proto-oncogene can act as a suppressor of transformation. Cell. 1989;57(7):1083-93.

75. Reich NC, Levine AJ. Growth regulation of a cellular tumour antigen, p53, in nontransformed cells. Nature. 1984;308(5955):199-201.

76. Levine AJ. p53, the cellular gatekeeper for growth and division. Cell. 1997;88(3):323-31.

77. Farnebo M. Wrap53, a novel regulator of p53. Cell cycle. 2009;8(15):2343-6.

78. Reisman D, Takahashi P, Polson A, Boggs K. Transcriptional Regulation of the p53 Tumor Suppressor Gene in S-Phase of the Cell-Cycle and the Cellular Response to DNA Damage. Biochemistry research international. 2012;2012:808934.

79. Liu J, Zhang C, Zhao Y, Feng Z. MicroRNA Control of p53. Journal of cellular biochemistry. 2017;118(1):7-14.

80. Poyurovsky MV, Katz C, Laptenko O, Beckerman R, Lokshin M, Ahn J, et al. The C terminus of p53 binds the N-terminal domain of MDM2. Nature structural & molecular biology. 2010;17(8):982-9.

81. Moll UM, Petrenko O. The MDM2-p53 interaction. Molecular cancer research : MCR. 2003;1(14):1001-8.

82. Valdez JM, Nichols KE, Kesserwan C. Li-Fraumeni syndrome: a paradigm for the understanding of hereditary cancer predisposition. British Journal of Haematology. 2017;176(4):539-52.

83. Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, et al. A global map of p53 transcription-factor binding sites in the human genome. Cell. 2006;124(1):207-19.

84. Olivier M, Hollstein M, Hainaut P. TP53 mutations in human cancers: origins, consequences, and clinical use. Cold Spring Harbor perspectives in biology. 2010;2(1):a001008.

85. Malkin D, Li FP, Strong LC, Fraumeni JF, Jr., Nelson CE, Kim DH, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science. 1990;250(4985):1233-8.

86. Srivastava S, Zou ZQ, Pirollo K, Blattner W, Chang EH. Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome. Nature. 1990;348(6303):747-9.

87. Malkin D. Li-fraumeni syndrome. Genes & cancer. 2011;2(4):475-84.

88. Bougeard G, Renaux-Petel M, Flaman JM, Charbonnier C, Fermey P, Belotti M, et al. Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2015;33(21):2345-52.

89. Miller M, Shirole N, Tian R, Pal D, Sordella R. The Evolution of TP53 Mutations: From Loss-of-Function to Separation-of-Function Mutants. Journal of cancer biology & research. 2016;4(4).

90. Lozano G. The oncogenic roles of p53 mutants in mouse models. Current opinion in genetics & development. 2007;17(1):66-70.

91. Donzelli S, Biagioni F, Fausti F, Strano S, Fontemaggi G, Blandino G. Oncogenomic Approaches in Exploring Gain of Function of Mutant p53. Current genomics. 2008;9(3):200-7.

92. Brosh R, Rotter V. When mutants gain new powers: news from the mutant p53 field. Nature reviews Cancer. 2009;9(10):701-13.

93. Olivier M, Petitjean A, Marcel V, Petre A, Mounawar M, Plymoth A, et al. Recent advances in p53 research: an interdisciplinary perspective. Cancer gene therapy. 2009;16(1):1-12.

94. Bougeard G, Baert-Desurmont S, Tournier I, Vasseur S, Martin C, Brugieres L, et al. Impact of the MDM2 SNP309 and p53 Arg72Pro polymorphism on age of tumour onset in Li-Fraumeni syndrome. Journal of medical genetics. 2006;43(6):531-3.

95. Wu CC, Krahe R, Lozano G, Zhang B, Wilson CD, Jo EJ, et al. Joint effects of germ-line TP53 mutation, MDM2 SNP309, and gender on cancer risk in family studies of Li-Fraumeni syndrome. Human genetics. 2011;129(6):663-73.

96. Ohnstad HO, Castro R, Sun J, Heintz KM, Vassilev LT, Bjerkehagen B, et al. Correlation of TP53 and MDM2 genotypes with response to therapy in sarcoma. Cancer. 2013;119(5):1013-22.

97. Gao J, Kang AJ, Lin S, Dai ZJ, Zhang SQ, Liu D, et al. Association between MDM2 rs 2279744 polymorphism and breast cancer susceptibility: a meta-analysis based on 9,788 cases and 11,195 controls. Therapeutics and clinical risk management. 2014;10:269-77.

98. Ma Y, Bian J, Cao H. MDM2 SNP309 rs2279744 polymorphism and gastric cancer risk: a meta-analysis. PloS one. 2013;8(2):e56918.

99. Mitchell G, Ballinger ML, Wong S, Hewitt C, James P, Young MA, et al. High frequency of germline TP53 mutations in a prospective adult-onset sarcoma cohort. PloS one. 2013;8(7):e69026.

100. Ru JY, Cong Y, Kang WB, Yu L, Guo T, Zhao JN. Polymorphisms in TP53 are associated with risk and survival of osteosarcoma in a Chinese population. International journal of clinical and experimental pathology. 2015;8(3):3198-203.

101. Samuel N, Id Said B, Guha T, Novokmet A, Li W, Silwal-Pandit L, et al. Assessment of TP53 Polymorphisms and MDM2 SNP309 in Premenopausal Breast Cancer Risk. Human mutation. 2017;38(3):265-8.

102. Lane D, Levine A. p53 Research: the past thirty years and the next thirty years. Cold Spring Harbor perspectives in biology. 2010;2(12):a000893.

103. Little JB, Nove J, Dahlberg WK, Troilo P, Nichols WW, Strong LC. Normal cytotoxic response of skin fibroblasts from patients with Li-Fraumeni familial cancer syndrome to DNA-damaging agents in vitro. Cancer research. 1987;47(15):4229-34.

104. Camplejohn RS, Perry P, Hodgson SV, Turner G, Williams A, Upton C, et al. A possible screening test for inherited p53-related defects based on the apoptotic response of peripheral blood lymphocytes to DNA damage. British journal of cancer. 1995;72(3):654-62.

105. Boyle JM, Spreadborough A, Greaves MJ, Birch JM, Varley JM, Scott D. The relationship between radiation-induced G(1)arrest and chromosome aberrations in Li-Fraumeni fibroblasts with or without germline TP53 mutations. British journal of cancer. 2001;85(2):293-6.

106. Mirzayans R, Severin D, Murray D. Relationship between DNA double-strand break rejoining and cell survival after exposure to ionizing radiation in human fibroblast strains with differing ATM/p53 status: implications for evaluation of clinical

radiosensitivity. International journal of radiation oncology, biology, physics. 2006;66(5):1498-505.

107. Williams KJ, Heighway J, Birch JM, Norton JD, Scott D. No defect in G1/S cell cycle arrest in irradiated Li-Fraumeni lymphoblastoid cell lines. British journal of cancer. 1996;74(5):698-703.

108. Gilchrist R, Lomax ME, Camplejohn RS. The need for dynamic methods for measuring cell cycle perturbations: a study in radiation-treated lymphoblastoid cell lines of varying p53 status. Cell proliferation. 1999;32(1):15-24.

109. Levy-Lahad E, Catane R, Eisenberg S, Kaufman B, Hornreich G, Lishinsky 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. American journal of human genetics. 1997;60(5):1059-67.

110. Rasmussen-Torvik LJ, Almoguera B, Doheny KF, Freimuth RR, Gordon AS, Hakonarson H, et al. Concordance between Research Sequencing and Clinical Pharmacogenetic Genotyping in the eMERGE-PGx Study. The Journal of molecular diagnostics : JMD. 2017;19(4):561-6.

111. Pompanon F, Bonin A, Bellemain E, Taberlet P. Genotyping errors: causes, consequences and solutions. Nature reviews Genetics. 2005;6(11):847-59.

112. Gonzalez KD, Noltner KA, Buzin CH, Gu D, Wen-Fong CY, Nguyen VQ, et al. Beyond Li Fraumeni Syndrome: clinical characteristics of families with p53 germline mutations. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(8):1250-6.

113. Malkin D. p53 and the Li-Fraumeni syndrome. Cancer genetics and cytogenetics. 1993;66(2):83-92.

114. Formiga M, de Andrade KC, Kowalski LP, Achatz MI. Frequency of Thyroid Carcinoma in Brazilian TP53 p.R337H Carriers With Li Fraumeni Syndrome. JAMA oncology. 2017;3(10):1400-2.

115. Liu Y, Bodmer WF. Analysis of P53 mutations and their expression in 56 colorectal cancer cell lines. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(4):976-81.

116. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic acids research. 2006;34(Web Server issue):W676-82.

117. Stothard P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. BioTechniques. 2000;28(6):1102, 4.

118. Olivier M, Goldgar DE, Sodha N, Ohgaki H, Kleihues P, Hainaut P, et al. Li-Fraumeni and related syndromes: correlation between tumor type, family structure, and TP53 genotype. Cancer research. 2003;63(20):6643-50.

119. Li FP, Fraumeni JF, Jr. Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? Annals of internal medicine. 1969;71(4):747-52.

120. Li FP, Fraumeni JF, Jr. Prospective study of a family cancer syndrome. Jama. 1982;247(19):2692-4.

121. Pearson AD, Craft AW, Ratcliffe JM, Birch JM, Morris-Jones P, Roberts DF. Two families with the Li-Fraumeni cancer family syndrome. Journal of medical genetics. 1982;19(5):362-5.

122. Lynch HT, Mulcahy GM, Harris RE, Guirgis HA, Lynch JF. Genetic and pathologic findings in a kindred with hereditary sarcoma, breast cancer, brain tumors, leukemia, lung, laryngeal, and adrenal cortical carcinoma. Cancer. 1978;41(5):2055-64.

123. Hartley AL, Birch JM, Marsden HB, Harris M. Malignant melanoma in families of children with osteosarcoma, chondrosarcoma, and adrenal cortical carcinoma. Journal of medical genetics. 1987;24(11):664-8.

124. Li FP, Fraumeni JF, Jr., Mulvihill JJ, Blattner WA, Dreyfus MG, Tucker MA, et al. A cancer family syndrome in twenty-four kindreds. Cancer research. 1988;48(18):5358-62.

125. Birch JM, Hartley AL, Tricker KJ, Prosser J, Condie A, Kelsey AM, et al. Prevalence and Diversity of Constitutional Mutations in the p53 Gene among 21 Li-Fraumeni Families. Cancer research. 1994;54(5):1298-304.

126. Chompret A, Abel A, Stoppa-Lyonnet D, Brugieres L, Pages S, Feunteun J, et al. Sensitivity and predictive value of criteria for p53 germline mutation screening. Journal of medical genetics. 2001;38(1):43-7.

127. Bougeard G, Sesboue R, Baert-Desurmont S, Vasseur S, Martin C, Tinat J, et al. Molecular basis of the Li-Fraumeni syndrome: an update from the French LFS families. Journal of medical genetics. 2008;45(8):535-8.

128. Tinat J, Bougeard G, Baert-Desurmont S, Vasseur S, Martin C, Bouvignies E, et al. 2009 version of the Chompret criteria for Li Fraumeni syndrome. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(26):e108-9; author reply e10.

129. Eeles RA. Germline mutations in the TP53 gene1995. 101-24 p.

130. Ognjanovic S, Olivier M, Bergemann TL, Hainaut P. Sarcomas in TP53 germline mutation carriers: a review of the IARC TP53 database. Cancer. 2012;118(5):1387-96.

131. Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, et al. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype:

lessons from recent developments in the IARC TP53 database. Human mutation. 2007;28(6):622-9.

132. Tabori U, Shlien A, Baskin B, Levitt S, Ray P, Alon N, et al. TP53 alterations determine clinical subgroups and survival of patients with choroid plexus tumors. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(12):1995-2001.

133. Villani A, Tabori U, Schiffman J, Shlien A, Beyene J, Druker H, et al. Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: a prospective observational study. The Lancet Oncology. 2011;12(6):559-67.

134. Wong P, Verselis SJ, Garber JE, Schneider K, DiGianni L, Stockwell DH, et al. Prevalence of early onset colorectal cancer in 397 patients with classic Li-Fraumeni syndrome. Gastroenterology. 2006;130(1):73-9.

135. Masciari S, Dewanwala A, Stoffel EM, Lauwers GY, Zheng H, Achatz MI, et al. Gastric cancer in individuals with Li-Fraumeni syndrome. Genetics in medicine : official journal of the American College of Medical Genetics. 2011;13(7):651-7.

136. Birch JM, Alston RD, McNally RJ, Evans DG, Kelsey AM, Harris M, et al. Relative frequency and morphology of cancers in carriers of germline TP53 mutations. Oncogene. 2001;20(34):4621-8.

137. Hisada M, Garber JE, Fung CY, Fraumeni JF, Jr., Li FP. Multiple primary cancers in families with Li-Fraumeni syndrome. Journal of the National Cancer Institute. 1998;90(8):606-11.

138. Gonzalez K, Fong C, Buzin C, Sommer SS, Saldivar JS. p53 Testing for Li-Fraumeni and Li-Fraumeni-like syndromes. Current protocols in human genetics. 2008;Chapter 10:Unit 10

139. Giacomazzi CR, Giacomazzi J, Netto CB, Santos-Silva P, Selistre SG, Maia AL, et al. Pediatric cancer and Li-Fraumeni/Li-Fraumeni-like syndromes: a review for the pediatrician. Revista da Associacao Medica Brasileira. 2015;61(3):282-9.

140. Chompret A, Brugieres L, Ronsin M, Gardes M, Dessarps-Freichey F, Abel A, et al. P53 germline mutations in childhood cancers and cancer risk for carrier individuals. British journal of cancer. 2000;82(12):1932-7.

141. McBride KA, Ballinger ML, Killick E, Kirk J, Tattersall MH, Eeles RA, et al. Li-Fraumeni syndrome: cancer risk assessment and clinical management. Nature reviews Clinical oncology. 2014;11(5):260-71.

142. Amadou A, Waddington Achatz MI, Hainaut P. Revisiting tumor patterns and penetrance in germline TP53 mutation carriers: temporal phases of Li-Fraumeni syndrome. Current opinion in oncology. 2018;30(1):23-9.

143. Renaux-Petel M, Charbonnier F, Thery JC, Fermey P, Lienard G, Bou J, et al. Contribution of de novo and mosaic TP53 mutations to Li-Fraumeni syndrome. Journal of medical genetics. 2018;55(3):173-80.

144. Wasserman JD, Novokmet A, Eichler-Jonsson C, Ribeiro RC, Rodriguez-Galindo C, Zambetti GP, et al. Prevalence and functional consequence of TP53 mutations in pediatric adrenocortical carcinoma: a children's oncology group study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2015;33(6):602-9.

145. Petitjean A, Achatz MI, Borresen-Dale AL, Hainaut P, Olivier M. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. Oncogene. 2007;26(15):2157-65.

146. Bouaoun L, Sonkin D, Ardin M, Hollstein M, Byrnes G, Zavadil J, et al. TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data. Human mutation. 2016;37(9):865-76.

147. Birch JM, Blair V, Kelsey AM, Evans DG, Harris M, Tricker KJ, et al. Cancer phenotype correlates with constitutional TP53 genotype in families with the Li-Fraumeni syndrome. Oncogene. 1998;17(9):1061-8.

148. Burt EC, McGown G, Thorncroft M, James LA, Birch JM, Varley JM. Exclusion of the genes CDKN2 and PTEN as causative gene defects in Li-Fraumeni syndrome. British journal of cancer. 1999;80(1-2):9-10.

149. Stone JG, Eeles RA, Sodha N, Murday V, Sheriden E, Houlston RS. Analysis of Li-Fraumeni syndrome and Li-Fraumeni-like families for germline mutations in Bcl10. Cancer letters. 1999;147(1-2):181-5.

150. Brown LT, Sexsmith E, Malkin D. Identification of a novel PTEN intronic deletion in Li-Fraumeni syndrome and its effect on RNA processing. Cancer genetics and cytogenetics. 2000;123(1):65-8.

151. Bougeard G, Limacher JM, Martin C, Charbonnier F, Killian A, Delattre O, et al. Detection of 11 germline inactivating TP53 mutations and absence of TP63 and HCHK2 mutations in 17 French families with Li-Fraumeni or Li-Fraumeni-like syndrome. Journal of medical genetics. 2001;38(4):253-7.

152. Vahteristo P, Tamminen A, Karvinen P, Eerola H, Eklund C, Aaltonen LA, et al. p53, CHK2, and CHK1 genes in Finnish families with Li-Fraumeni syndrome: further evidence of CHK2 in inherited cancer predisposition. Cancer research. 2001;61(15):5718-22.

153. Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R, et al. Low-penetrance susceptibility to breast cancer due to

CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet. 2002;31(1):55-9.

154. Vahteristo P, Bartkova J, Eerola H, Syrjakoski K, Ojala S, Kilpivaara O, et al. A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. American journal of human genetics. 2002;71(2):432-8.

155. Bell DW, Varley JM, Szydlo TE, Kang DH, Wahrer DC, Shannon KE, et al. Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. Science. 1999;286(5449):2528-31.

156. Sodha N, Williams R, Mangion J, Bullock SL, Yuille MR, Eeles RA. Screening hCHK2 for mutations. Science. 2000;289(5478):359.

157. Bachinski LL, Olufemi SE, Zhou X, Wu CC, Yip L, Shete S, et al. Genetic mapping of a third Li-Fraumeni syndrome predisposition locus to human chromosome 1q23. Cancer research. 2005;65(2):427-31.

158. Ramachandra C, Challa VR, Shetty R. Constitutional mismatch repair deficiency syndrome: Do we know it? Indian journal of human genetics. 2014;20(2):192-4.

159. Bond GL, Hu W, Bond EE, Robins H, Lutzker SG, Arva NC, et al. A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. Cell. 2004;119(5):591-602.

160. Ruijs MW, Schmidt MK, Nevanlinna H, Tommiska J, Aittomaki K, Pruntel R, et al. The single-nucleotide polymorphism 309 in the MDM2 gene contributes to the Li-Fraumeni syndrome and related phenotypes. European journal of human genetics : EJHG. 2007;15(1):110-4.

161. Bond GL, Hirshfield KM, Kirchhoff T, Alexe G, Bond EE, Robins H, et al. MDM2 SNP309 accelerates tumor formation in a gender-specific and hormone-dependent manner. Cancer research. 2006;66(10):5104-10.

162. Renaux-Petel M, Sesboue R, Baert-Desurmont S, Vasseur S, Fourneaux S, Bessenay E, et al. The MDM2 285G-309G haplotype is associated with an earlier age of tumour onset in patients with Li-Fraumeni syndrome. Familial cancer. 2014;13(1):127-30.

163. Marcel V, Palmero EI, Falagan-Lotsch P, Martel-Planche G, Ashton-Prolla P, Olivier M, et al. TP53 PIN3 and MDM2 SNP309 polymorphisms as genetic modifiers in the Li-Fraumeni syndrome: impact on age at first diagnosis. Journal of medical genetics. 2009;46(11):766-72.

164. Id Said B, Kim H, Tran J, Novokmet A, Malkin D. Super-Transactivation TP53 Variant in the Germline of a Family with Li-Fraumeni Syndrome. Human mutation. 2016;37(9):889-92.

165. Evans DG, Birch JM, Thorneycroft M, McGown G, Lalloo F, Varley JM. Low rate of TP53 germline mutations in breast cancer/sarcoma families not fulfilling classical criteria for Li-Fraumeni syndrome. Journal of medical genetics. 2002;39(12):941-4.

166. Diller L, Sexsmith E, Gottlieb A, Li FP, Malkin D. Germline p53 mutations are frequently detected in young children with rhabdomyosarcoma. The Journal of clinical investigation. 1995;95(4):1606-11.

167. Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R, et al. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(14):8424-9.

168. Pickens A, Vickers SM, Brown KL, Reddy VV, Thompson JA. An unusual etiology of biliary hilar obstruction and the potential role of acidic fibroblast growth factor in the development of a biliary neuroma. The American surgeon. 1999;65(1):47-51.

169. FitzGerald LM, Karlins E, Karyadi DM, Kwon EM, Koopmeiners JS, Stanford JL, et al. Association of FGFR4 genetic polymorphisms with prostate cancer risk and prognosis. Prostate cancer and prostatic diseases. 2009;12(2):192-7.

170. Taylor JGt, Cheuk AT, Tsang PS, Chung JY, Song YK, Desai K, et al. Identification of FGFR4-activating mutations in human rhabdomyosarcomas that promote metastasis in xenotransplanted models. The Journal of clinical investigation. 2009;119(11):3395-407.

171. Giacomazzi J, Correia RL, Palmero EI, Gaspar JF, Almeida M, Portela C, et al. The Brazilian founder mutation TP53 p.R337H is uncommon in Portuguese women diagnosed with breast cancer. The breast journal. 2014;20(5):534-6.

172. Chen X, Bahrami A, Pappo A, Easton J, Dalton J, Hedlund E, et al. Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. Cell reports. 2014;7(1):104-12.

173. Ribi S, Baumhoer D, Lee K, Edison, Teo AS, Madan B, et al. TP53 intron 1 hotspot rearrangements are specific to sporadic osteosarcoma and can cause Li-Fraumeni syndrome. Oncotarget. 2015;6(10):7727-40.

174. Macedo GS, Araujo Vieira I, Brandalize AP, Giacomazzi J, Inez Palmero E, Volc S, et al. Rare germline variant (rs78378222) in the TP53 3' UTR: Evidence for a new mechanism of cancer predisposition in Li-Fraumeni syndrome. Cancer genetics. 2016;209(3):97-106.

175. Ravine D, Suthers G. Quality standards and samples in genetic testing. Journal of clinical pathology. 2012;65(5):389-93.

176. Mullins FM, Dietz L, Lay M, Zehnder JL, Ford J, Chun N, et al. Identification of an intronic single nucleotide polymorphism leading to allele dropout during validation of a CDH1 sequencing assay: implications for designing polymerase chain reaction-

based assays. Genetics in medicine : official journal of the American College of Medical Genetics. 2007;9(11):752-60.

177. De Sarkar N, Majumder M, Roy B. Differential haplotype amplification leads to misgenotyping of heterozygote as homozygote when using single nucleotide mismatch primer. Electrophoresis. 2012;33(23):3564-73.

178. Silva FC, Torrezan GT, Brianese RC, Stabellini R, Carraro DM. Pitfalls in genetic testing: a case of a SNP in primer-annealing region leading to allele dropout in BRCA1. Molecular genetics & genomic medicine. 2017;5(4):443-7.

179. Saunders CJ, Friez MJ, Patterson M, Nzabi M, Zhao W, Bi C. Allele drop-out in the MECP2 gene due to G-quadruplex and i-motif sequences when using polymerase chain reaction-based diagnosis for Rett syndrome. Genetic testing and molecular biomarkers. 2010;14(2):241-7.

180. Stevens AJ, Taylor MG, Pearce FG, Kennedy MA. Allelic Dropout During Polymerase Chain Reaction due to G-Quadruplex Structures and DNA Methylation Is Widespread at Imprinted Human Loci. G3. 2017;7(3):1019-25.

181. Soulsbury CD, Iossa G, Edwards KJ, Baker PJ, Harris S. Allelic dropout from a high-quality DNA source. Conservation Genetics. 2007;8 (3):733-8.

182. Dolan SM, Goldwaser TH, Jindal SK. Preimplantation Genetic Diagnosis for Mendelian Conditions. Jama. 2017;318(9):859-60.

183. Blais J, Lavoie SB, Giroux S, Bussieres J, Lindsay C, Dionne J, et al. Risk of Misdiagnosis Due to Allele Dropout and False-Positive PCR Artifacts in Molecular Diagnostics: Analysis of 30,769 Genotypes. The Journal of molecular diagnostics : JMD. 2015;17(5):505-14.

184. Zucca S, Villaraggia M, Gagliardi S, Grieco GS, Valente M, Cereda C, et al. Analysis of amplicon-based NGS data from neurological disease gene panels: a new method for allele drop-out management. BMC bioinformatics. 2016;17(Suppl 12):339.

185. Costa HA, Neal JW, Bustamante CD, Zehnder JL. Identification of a Novel Somatic Mutation Leading to Allele Dropout for EGFR L858R Genotyping in Non-Small Cell Lung Cancer. Molecular diagnosis & therapy. 2017;21(4):431-6.

186. India Project Team of the International Cancer Genome C. Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. Nature communications. 2013;4:2873.

187. Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, et al. Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. The Journal of molecular diagnostics : JMD. 2017;19(3):341-65.

188. Gill P, Buckleton J. A universal strategy to interpret DNA profiles that does not require a definition of low-copy-number. Forensic science international Genetics. 2010;4(4):221-7.

189. Budowle B, van Daal A. Comment on "A universal strategy to interpret DNA profiles that does not require a definition of low copy number" by Peter Gill and John Buckleton, 2010, Forensic Sci. Int. Genetics 4, 221-227. Forensic science international Genetics. 2011;5(1):15.

190. Ewen KR, Bahlo M, Treloar SA, Levinson DF, Mowry B, Barlow JW, et al. Identification and analysis of error types in high-throughput genotyping. American journal of human genetics. 2000;67(3):727-36.

191. Hasmats J, Green H, Solnestam BW, Zajac P, Huss M, Orear C, et al. Validation of whole genome amplification for analysis of the p53 tumor suppressor gene in limited amounts of tumor samples. Biochemical and biophysical research communications. 2012;425(2):379-83.

192. Austin MC, Smith C, Pritchard CC, Tait JF. DNA Yield From Tissue Samples in Surgical Pathology and Minimum Tissue Requirements for Molecular Testing. Archives of pathology & laboratory medicine. 2016;140(2):130-3.

193. Forest F, Stachowicz ML, Casteillo F, Karpathiou G, Gouzy-Grosjean F, Guilaubey C, et al. EGFR, KRAS, BRAF and HER2 testing in metastatic lung adenocarcinoma: Value of testing on samples with poor specimen adequacy and analysis of discrepancies. Experimental and molecular pathology. 2017;103(3):306-10.

194. Coulet F, Pires F, Rouleau E, Lefol C, Martin S, Colas C, et al. A one-step prescreening for point mutations and large rearrangement in BRCA1 and BRCA2 genes using quantitative polymerase chain reaction and high-resolution melting curve analysis. Genetic testing and molecular biomarkers. 2010;14(5):677-90.

195. Chong HK, Wang T, Lu HM, Seidler S, Lu H, Keiles S, et al. The validation and clinical implementation of BRCAplus: a comprehensive high-risk breast cancer diagnostic assay. PloS one. 2014;9(5):e97408.

196. Dahl F, Stenberg J, Fredriksson S, Welch K, Zhang M, Nilsson M, et al. Multigene amplification and massively parallel sequencing for cancer mutation discovery. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(22):9387-92.

197. Lerman DM, Monument MJ, McIlvaine E, Liu XQ, Huang D, Monovich L, et al. Tumoral TP53 and/or CDKN2A alterations are not reliable prognostic biomarkers in

patients with localized Ewing sarcoma: a report from the Children's Oncology Group. Pediatric blood & cancer. 2015;62(5):759-65.

198. Broniscer A, Baker SJ, West AN, Fraser MM, Proko E, Kocak M, et al. Clinical and molecular characteristics of malignant transformation of low-grade glioma in children. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2007;25(6):682-9.

199. Hoe SL, Lee ES, Khoo AS, Peh SC. p53 and nasopharyngeal carcinoma: a Malaysian study. Pathology. 2009;41(6):561-5.

200. Hao XD, Yang Y, Song X, Zhao XK, Wang LD, He JD, et al. Correlation of telomere length shortening with TP53 somatic mutations, polymorphisms and allelic loss in breast tumors and esophageal cancer. Oncology reports. 2013;29(1):226-36.

201. Sibin MK, Lavanya CH, Bhat DI, Rao N, Geethashree N, Vibhuti W, et al. CD133 and BMI1 expressions and its prognostic role in primary glioblastoma. Journal of genetics. 2015;94(4):689-96.

202. Harajly M, Zalzali H, Nawaz Z, Ghayad SE, Ghamloush F, Basma H, et al. p53 Restoration in Induction and Maintenance of Senescence: Differential Effects in Premalignant and Malignant Tumor Cells. Molecular and cellular biology. 2016;36(3):438-51.

203. Sato S, Itamochi H, Oumi N, Chiba Y, Oishi T, Shimada M, et al. Establishment and characterization of a novel ovarian clear cell carcinoma cell line, TU-OC-2, with loss of ARID1A expression. Human cell. 2016;29(4):181-7.

204. Abedi-Ardekani B, Kamangar F, Sotoudeh M, Villar S, Islami F, Aghcheli K, et al. Extremely high Tp53 mutation load in esophageal squamous cell carcinoma in Golestan Province, Iran. PloS one. 2011;6(12):e29488.

205. Ognjanovic S, Martel G, Manivel C, Olivier M, Langer E, Hainaut P. Low Prevalence of TP53 Mutations and MDM2 Amplifications in Pediatric Rhabdomyosarcoma. Sarcoma. 2012;2012:492086.

206. Tornesello ML, Buonaguro L, Tatangelo F, Botti G, Izzo F, Buonaguro FM. Mutations in TP53, CTNNB1 and PIK3CA genes in hepatocellular carcinoma associated with hepatitis B and hepatitis C virus infections. Genomics. 2013;102(2):74-83.

207. Sagne C, Marcel V, Bota M, Martel-Planche G, Nobrega A, Palmero EI, et al. Age at cancer onset in germline TP53 mutation carriers: association with polymorphisms in predicted G-quadruplex structures. Carcinogenesis. 2014;35(4):807-15.

208. Liu Q, Thorland EC, Sommer SS. Inhibition of PCR amplification by a point mutation downstream of a primer. BioTechniques. 1997;22(2):292-4, 6, 8, passim.

209. Lam CW, Mak CM. Allele dropout caused by a non-primer-site SNV affecting PCR amplification--a call for next-generation primer design algorithm. Clinica chimica acta; international journal of clinical chemistry. 2013;421:208-12.

210. Ramaswamy A, Rekhi B, Bakhshi S, Hingmire S, Agarwal M. Indian data on bone and soft tissue sarcomas: A summary of published study results. South Asian journal of cancer. 2016;5(3):138-45.

211. Peltier LF. Historical note on bone and soft tissue sarcoma. Journal of surgical oncology. 1985;30(4):201-5.

212. Ham SJ, Schraffordt Koops H, van der Graaf WT, van Horn JR, Postma L, Hoekstra HJ. Historical, current and future aspects of osteosarcoma treatment. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 1998;24(6):584-600.

213. Mirabello L, Troisi RJ, Savage SA. International osteosarcoma incidence patterns in children and adolescents, middle ages and elderly persons. International journal of cancer. 2009;125(1):229-34.

214. Zambo I, Vesely K. [WHO classification of tumours of soft tissue and bone 2013: the main changes compared to the 3rd edition]. Ceskoslovenska patologie. 2014;50(2):64-70.

215. Yeole BB, Jussawalla DJ. Descriptive epidemiology of bone cancer in greater Bombay. Indian journal of cancer. 1998;35(3):101-6.

216. Rao VS, Pai MR, Rao RC, Adhikary MM. Incidence of primary bone tumours and tumour like lesions in and around Dakshina Kannada district of Karnataka. Journal of the Indian Medical Association. 1996;94(3):103-4, 21.

217. VG A, Banerjee R, Bandopadhyay D. EPIDEMIOLOGY, PATHOLOGY, TYPES AND DIAGNOSIS OF SOFT TISSUE SARCOMA: A RESEARCH REVIEW. 2013. 2013:8.

218. Jo VY, Fletcher CDM. WHO classification of soft tissue tumours: an update based on the 2013 (4th) edition. PATHOL Pathology. 2014;46(2):95-104.

219. Brennan MF, Antonescu CR, Moraco N, Singer S. Lessons learned from the study of 10,000 patients with soft tissue sarcoma. Annals of surgery. 2014;260(3):416-21; discussion 21-2.

220. Campanacci M. Bone and soft tissue tumors : clinical features, imaging, pathology and treatment. Wien: Springer; 1999.

221. Clark JC, Dass CR, Choong PF. A review of clinical and molecular prognostic factors in osteosarcoma. Journal of cancer research and clinical oncology. 2008;134(3):281-97.

222. Peltier LF. Orthopedics : a history and iconography. San Francisco: Norman Publishing; 1993.

223. Simon MA, Springfield DS. Surgery for bone and soft-tissue tumors. Philadelphia: Lippincott-Raven Publishers; 1998.

224. Eriksson M, Hardell L, Adami HO. Exposure to dioxins as a risk factor for soft tissue sarcoma: a population-based case-control study. Journal of the National Cancer Institute. 1990;82(6):486-90.

225. Eriksson M, Hardell L, Berg NO, Moller T, Axelson O. Soft-tissue sarcomas and exposure to chemical substances: a case-referent study. British journal of industrial medicine. 1981;38(1):27-33.

226. Karlsson P, Holmberg E, Samuelsson A, Johansson KA, Wallgren A. Soft tissue sarcoma after treatment for breast cancer--a Swedish population-based study. European journal of cancer. 1998;34(13):2068-75.

227. Weiss RA, Whitby D, Talbot S, Kellam P, Boshoff C. Human Herpesvirus Type 8 and Kaposi's Sarcoma. JNCI Monographs. 1998;1998(23):51-4.

228. Farid M, Ngeow J. Sarcomas Associated With Genetic Cancer Predisposition Syndromes: A Review. The oncologist. 2016;21(8):1002-13.

229. Biermann JS, Chow W, Reed DR, Lucas D, Adkins DR, Agulnik M, et al. NCCN Guidelines Insights: Bone Cancer, Version 2.2017. Journal of the National Comprehensive Cancer Network : JNCCN. 2017;15(2):155-67.

230. von Mehren M, Randall RL, Benjamin RS, Boles S, Bui MM, Ganjoo KN, et al. Soft Tissue Sarcoma, Version 2.2018, NCCN Clinical Practice Guidelines in Oncology. Journal of the National Comprehensive Cancer Network : JNCCN. 2018;16(5):536-63.

231. Group ESESNW. Bone sarcomas: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of oncology : official journal of the European Society for Medical Oncology. 2014;25 Suppl 3:iii113-23.

232. Trojani M, Contesso G, Coindre JM, Rouesse J, Bui NB, de Mascarel A, et al. Soft-tissue sarcomas of adults; study of pathological prognostic variables and definition of a histopathological grading system. International journal of cancer. 1984;33(1):37-42.

233. Costa J, Wesley RA, Glatstein E, Rosenberg SA. The grading of soft tissue sarcomas. Results of a clinicohistopathologic correlation in a series of 163 cases. Cancer. 1984;53(3):530-41.

234. American Joint Committee on C, Amin MB. AJCC cancer staging manual. New York: Springer; 2017.

235. Enneking WF. A system of staging musculoskeletal neoplasms. Clinical orthopaedics and related research. 1986(204):9-24.

236. Ballinger ML, Goode DL, Ray-Coquard I, James PA, Mitchell G, Niedermayr E, et al. Monogenic and polygenic determinants of sarcoma risk: an international genetic study. The Lancet Oncology. 2016;17(9):1261-71.

237. Hu Z, Li X, Qu X, He Y, Ring BZ, Song E, et al. Intron 3 16 bp duplication polymorphism of TP53 contributes to cancer susceptibility: a meta-analysis. Carcinogenesis. 2010;31(4):643-7.

238. Qin X, Peng Q, Tang W, Lao X, Chen Z, Lai H, et al. An updated meta-analysis on the association of MDM2 SNP309 polymorphism with colorectal cancer risk. PloS one. 2013;8(9):e76031.

239. Wu D, Zhang Z, Chu H, Xu M, Xue Y, Zhu H, et al. Intron 3 sixteen base pairs duplication polymorphism of p53 contributes to breast cancer susceptibility: evidence from meta-analysis. PloS one. 2013;8(4):e61662.

240. Brookes AJ. The essence of SNPs. Gene. 1999;234(2):177-86.

241. Toffoli G, Biason P, Russo A, De Mattia E, Cecchin E, Hattinger CM, et al. Effect of TP53 Arg72Pro and MDM2 SNP309 polymorphisms on the risk of high-grade osteosarcoma development and survival. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15(10):3550-6.

242. Paulin FE, O'Neill M, McGregor G, Cassidy A, Ashfield A, Ali CW, et al. MDM2 SNP309 is associated with high grade node positive breast tumours and is in linkage disequilibrium with a novel MDM2 intron 1 polymorphism. BMC cancer. 2008;8:281.

243. Lum SS, Chua HW, Li H, Li WF, Rao N, Wei J, et al. MDM2 SNP309 G allele increases risk but the T allele is associated with earlier onset age of sporadic breast cancers in the Chinese population. Carcinogenesis. 2008;29(4):754-61.

244. Gemignani F, Moreno V, Landi S, Moullan N, Chabrier A, Gutierrez-Enriquez S, et al. A TP53 polymorphism is associated with increased risk of colorectal cancer and with reduced levels of TP53 mRNA. Oncogene. 2004;23(10):1954-6.

245. Costa S, Pinto D, Pereira D, Rodrigues H, Cameselle-Teijeiro J, Medeiros R, et al. Importance of TP53 codon 72 and intron 3 duplication 16bp polymorphisms in prediction of susceptibility on breast cancer. BMC cancer. 2008;8:32.

246. Perriaud L, Marcel V, Sagne C, Favaudon V, Guedin A, De Rache A, et al. Impact of G-quadruplex structures and intronic polymorphisms rs17878362 and rs1642785 on basal and ionizing radiation-induced expression of alternative p53 transcripts. Carcinogenesis. 2014;35(12):2706-15.

247. Winiecka-Klimek M, Szybka M, Rieske P, Piaskowski S, Bienkowski M, Walczak M, et al. PIN3 duplication may be partially responsible for TP53 haploinsufficiency. BMC cancer. 2014;14:669.

248. Sagne C, Marcel V, Amadou A, Hainaut P, Olivier M, Hall J. A meta-analysis of cancer risk associated with the TP53 intron 3 duplication polymorphism (rs17878362): geographic and tumor-specific effects. Cell death & disease. 2013;4:e492.

249. Masters JR. HeLa cells 50 years on: the good, the bad and the ugly. Nature Reviews Cancer. 2002;2:315.

250. Epstein MA, Achong BG, Barr YM. Virus Particles in Cultured Lymphoblasts from Burkitt's Lymphoma. Lancet. 1964;1(7335):702-3.

251. Pulvertaft JV. Cytology of Burkitt's Tumour (African Lymphoma). Lancet. 1964;1(7327):238-40.

252. Stewart SE, Lovelace E, Whang JJ, Ngu VA. Burkitt Tumor: Tissue Culture, Cytogenetic and Virus Studies. Journal of the National Cancer Institute. 1965;34:319-27.

253. Rabson AS, O'Conor GT, Baron S, Whang JJ, Legallais FY. Morphologic, cytogenetic and virologic studies in vitro of a malignant lymphoma from an African child. International journal of cancer. 1966;1(1):89-106.

254. Henle W, Diehl V, Kohn G, Zur Hausen H, Henle G. Herpes-type virus and chromosome marker in normal leukocytes after growth with irradiated Burkitt cells. Science. 1967;157(3792):1064-5.

255. Pope JH, Horne MK, Scott W. Transformation of foetal human keukocytes in vitro by filtrates of a human leukaemic cell line containing herpes-like virus. International journal of cancer. 1968;3(6):857-66.

256. Gerper P, Whang-Peng J, Monroe JH. Transformation and chromosome changes induced by Epstein-Barr virus in normal human leukocyte cultures. Proceedings of the National Academy of Sciences of the United States of America. 1969;63(3):740-7.

257. Chesnokova LS, Hutt-Fletcher LM. Epstein-Barr virus infection mechanisms. Chinese journal of cancer. 2014;33(11):545-8.

258. Miller G, Shope T, Lisco H, Stitt D, Lipman M. Epstein-Barr virus: transformation, cytopathic changes, and viral antigens in squirrel monkey and marmoset leukocytes. Proceedings of the National Academy of Sciences of the United States of America. 1972;69(2):383-7.

259. Baer R, Bankier AT, Biggin MD, Deininger PL, Farrell PJ, Gibson TJ, et al. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature. 1984;310(5974):207-11.

260. Sie L, Loong S, Tan EK. Utility of lymphoblastoid cell lines. Journal of neuroscience research. 2009;87(9):1953-9.

261. Hussain T, Mulherkar R. Lymphoblastoid Cell lines: a Continuous in Vitro Source of Cells to Study Carcinogen Sensitivity and DNA Repair. International journal of molecular and cellular medicine. 2012;1(2):75-87.

262. Kumar S, Curran JE, Glahn DC, Blangero J. Utility of Lymphoblastoid Cell Lines for Induced Pluripotent Stem Cell Generation. Stem cells international. 2016;2016:2349261.

263. Pierce BA. Genetics : a conceptual approach2017.

264. Kumar S, Subramanian S. Mutation rates in mammalian genomes. Proceedings of the National Academy of Sciences. 2002;99(2):803-8.

265. Darzynkiewicz Z, Traganos F, Zhao H, Halicka HD, Skommer J, Wlodkowic D. Analysis of individual molecular events of DNA damage response by flow- and image-assisted cytometry. Methods in cell biology. 2011;103:115-47.

266. Brandsma I, Gent DC. Pathway choice in DNA double strand break repair: observations of a balancing act. Genome integrity. 2012;3(1):9.

267. Ivashkevich A, Redon CE, Nakamura AJ, Martin RF, Martin OA. Use of the gamma-H2AX assay to monitor DNA damage and repair in translational cancer research. Cancer letters. 2012;327(1-2):123-33.

268. Bonner WM, Redon CE, Dickey JS, Nakamura AJ, Sedelnikova OA, Solier S, et al. GammaH2AX and cancer. Nature reviews Cancer. 2008;8(12):957-67.

269. Dickey JS, Redon CE, Nakamura AJ, Baird BJ, Sedelnikova OA, Bonner WM. H2AX: functional roles and potential applications. Chromosoma. 2009;118(6):683-92.

270. Yuan J, Adamski R, Chen J. Focus on histone variant H2AX: to be or not to be. FEBS letters. 2010;584(17):3717-24.

271. Van Oorschot B, Oei AL, Nuijens AC, Rodermond H, Hoeben R, Stap J, et al. Decay of gamma-H2AX foci correlates with potentially lethal damage repair and P53 status in human colorectal carcinoma cells. Cellular & molecular biology letters. 2014;19(1):37-51.

272. Schuler N, Palm J, Schmitz S, Lorat Y, Rube CE. Increasing genomic instability during cancer therapy in a patient with Li-Fraumeni syndrome. Clinical and translational radiation oncology. 2017;7:71-8.

273. Suchankova J, Legartova S, Ruckova E, Vojtesek B, Kozubek S, Bartova E. Mutations in the TP53 gene affected recruitment of 53BP1 protein to DNA lesions, but level of 53BP1 was stable after gamma-irradiation that depleted MDC1 protein in specific TP53 mutants. Histochemistry and cell biology. 2017;148(3):239-55.

274. Varley JM, Attwooll C, White G, McGown G, Thorncroft M, Kelsey AM, et al. Characterization of germline TP53 splicing mutations and their genetic and functional analysis. Oncogene. 2001;20(21):2647-54.

275. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic acids research. 2009;37(9):e67.

276. Noensie EN, Dietz HC. A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. Nature biotechnology. 2001;19(5):434-9.

277. Staib F, Robles AI, Varticovski L, Wang XW, Zeeberg BR, Sirotin M, et al. The p53 tumor suppressor network is a key responder to microenvironmental components of chronic inflammatory stress. Cancer research. 2005;65(22):10255-64.

278. Solozobova V, Rolletschek A, Blattner C. Nuclear accumulation and activation of p53 in embryonic stem cells after DNA damage. BMC cell biology. 2009;10:46.

279. Nilsson K. Human B-lymphoid cell lines. Human cell. 1992;5(1):25-41.

280. Lalle P, Moyret-Lalle C, Wang Q, Vialle JM, Navarro C, Bressac-de Paillerets B, et al. Genomic stability and wild-type p53 function of lymphoblastoid cells with germ-line p53 mutation. Oncogene. 1995;10(12):2447-54.

281. Okubo M, Tsurukubo Y, Higaki T, Kawabe T, Goto M, Murase T, et al. Clonal chromosomal aberrations accompanied by strong telomerase activity in immortalization of human B-lymphoblastoid cell lines transformed by Epstein-Barr virus. Cancer genetics and cytogenetics. 2001;129(1):30-4.

282. Zhang W, Ratain MJ, Dolan ME. The HapMap Resource is Providing New Insights into Ourselves and its Application to Pharmacogenomics. Bioinformatics and biology insights. 2008;2:15-23.