Elucidating the genetic pathways of RET driven Medullary Thyroid Carcinogenesis

By

Vasudha Mishra

[LIFE09201304018]

Tata Memorial Centre - ACTREC
Mumbai

A thesis submitted to the

Board of Studies in Life Sciences

In partial fulfillment of requirements

For the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE

February 2020

Homi Bhabha National Institute

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by Ms. Vasudha Mishra entitled **"*Elucidating the genetic pathways of RET driven Medullary Thyroid Carcinogenesis*"** and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chairperson–Dr. Neelam Shirsat — Date: 7/02/2020

Guide/Convener –Dr. Rajiv Sarin — Date: 07/02/2020

External Examiner–Dr. Nihal Thomas — Date: 07/02/2020

Member –Dr. Abhijit De — Date: 10/02/2020

Member -Dr. Sudhir Nair — Date: 07/02/2020

Technical Advisor -Dr. Rita Mulherkar — Date: 07.02.20

Technical Advisor -Dr. Pradnya Kowtal — Date: 07/02/2020

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

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

Date: 07/02/2020
Place: Navi Mumbai

Dr.Rajiv Sarin
Guide

STATEMENT BY AUTHOR

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

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

Vasudha Mishra

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.

Vasudha Mishra

List of Publications arising from the thesis

Journal

Published

1. “Genetic risk association of CDKN1A and RET gene SNPs with Medullary Thyroid Carcinoma: Results from the largest MTC cohort and meta-analysis”, **Vasudha Mishra**; Pradnya Kowtal; Pallavi Rane; Rajiv Sarin. *Cancer Med*. 2019; 8: 6151–6161. DOI: 10.1002/cam4.2443.

Manuscripts under peer review

1. Modulatory role of Single Nucleotide Polymorphisms of distinct genetic pathways on clinical behavior of Medullary Thyroid Carcinoma.

Authors: Vasudha Mishra, Pradnya Kowtal, Pallavi Rane and *Rajiv Sarin.

Under peer review in **"Asian Pacific Journal of Cancer Prevention"**

2. False negative *RET* genetic testing in MEN2B cases due to Allele Dropout detected on exome sequencing: A clinical alert.

Authors: Vasudha Mishra, Pradnya Kowtal and *Rajiv Sarin.

Under peer review in **"Familial Cancer" journal.**

Peer reviewed conference proceedings

1. RET Genotype & MEN2 Phenotype correlation in a large Indian Medullary Thyroid Carcinoma (MTC) cohort & influence of SNPs in 3 genetic pathways on

MTC behaviour" **European Journal of Human Genetics, 2018**. ('Abstracts from the 51st European Society of Human Genetics Conference: Posters', 2019).

2. Influence of SNPs in distinct genetic pathways on the tumor burden and metastasis in Medullary Thyroid Carcinoma patients" **Journal of Carcinogenesis, 2016**. ('INVITED SPEAKERS' ABSTRACTS', 2016).

Chapters in books and lectures notes: Nil

Conferences

1. **Poster Presentation**: RET Genotype & MEN2 Phenotype correlation in a large Indian Medullary Thyroid Carcinoma (MTC) cohort & influence of SNPs in 3 genetic pathways on MTC behavior. **European Human Genetics Conference, 2018, *Milan Italy*.**

2. **Poster Presentation:** "Penetrance of distinct germline RET mutations in the largest Indian MEN2 syndrome cohort". **4th Indian Cancer Genetics conference & 3rd Indo-UK Cancer Genetic Conference 2018, ACTREC, Tata Memorial Centre, Navi Mumbai.**

3. **Poster Presentation**: Increased Sensitivity of Next Generation Sequencing over Sanger Sequencing in Mutation Detection: A call for reanalysis of False-Negative Sanger Results**. 9th HOPE Meeting with Nobel Laureates organized by Japan Society for Promotion of Science (JSPS), 2017, *Tokyo Japan***.

4. **Poster Presentation:** Utility of Next Generation Sequencing (NGS) in detecting germline allele dropouts in patients with hereditary cancer syndromes. **2nd International conference on Founder Population with the theme "The Landscape of Genetic Variants in Asian Founder Populations", 2017** ***Kochi, Kerela, India.***

5. **Poster Presentation**: Sequence Independent Allele Dropout event on Sanger Sequencing in classical MEN2B cases detected by Next Generation Sequencing. **3rd Indian Cancer Genetics Conference & Workshop 2016, ACTREC, Tata Memorial Centre,** ***Navi Mumbai, India.***

6. **Poster Presentation**: Influence of SNPs in distinct genetic pathways on tumor burden & metastasis in Medullary Thyroid Carcinoma patients. **12th International Conference of the Asian Clinical Oncology Society (ACOS) & 35th Annual Convention of Indian Association for Cancer research & Mid-Term Conference of IASO, Theme: CANCER IN ASIA: BRIDGING THE GAPS, 2016** ***New Delhi, India.***

7. **Poster Presentation**: Is CDKN1B V109G polymorphism a prognostic marker for Medullary Thyroid Carcinoma**. 2nd Indian Cancer Genetics Conference & Workshop, 2014, ACTREC , Tata Memorial centre,** ***Navi Mumbai, India***.

Vasudha Mishra

Vasudha Mishra

Dedicated to the hands that held me when I felt lost

& the support that stood by me at any cost.

'My Family'

ACKNOWLEDGEMENTS

These six years of my PhD tenure were the finest and most valued years of my life as they have turned me into a stronger and more confident person than ever. It gives me immense pleasure to acknowledge all those who stood beside me through thick and thin.

I would first like to acknowledge the source of my strength and my biggest inspiration – My Mother. She has been the strongest and the most inspiring person in my life. Words would fall short in expressing the love and respect I have for her and how much I owe her. Mom, I am who I am because of you. Thank you for always being there for me, blessing me, protecting me and guiding me. It was my mother's dream to see her daughters thrive and my Father's conviction to leave no stone unturned to make her dream come true. Thank you Dad for your unconditional love, for all the sacrifices you made, for believing in me and for letting me be me. Blessed are the children who have such parents and I am proud to be one.

This work would not have been possible without the guidance & support of my guide – Professor Rajiv Sarin. I would like to express my sincere gratitude to him for his continuous support and constant belief in me. He gave me the opportunity to work with him and provided a healthy working environment. During my PhD tenure I have learned a lot from him, right from the skills of making presentations to performing proper literature search, from designing experiments to thinking scientifically but even more important, he taught me not to run away from my fears but to confront and conquer them. I cannot forget to mention how he took the fear of statistics out of me which almost all the students without mathematics or statistics background have, by giving examples on how statistics is used in day to day life. Thank you Sir for instilling in me the qualities that a young researcher should acquire to flourish in future. You have been a true example of 'GURU' who recognized my capabilities and brought the best out of me.

I would also like to extend my gratitude to Dr. Pradnya Kowtal, Co-Investigator of Sarin lab. She is extremely proficient in molecular biology techniques and has always offered a helping

hand to the students whenever required. It has been a great experience working under her guidance. She has been a mother figure to me with whom I always speak my heart out.

I would like to thank the Director of ACTREC, Professor Sudeep Gupta, former Director ACTREC, Dr. SV Chiplunkar and the Deputy Director, Dr. Prasanna Venkatraman for providing full-fledged on-campus infrastructure to the students for their hassle free work. It is worth mentioning that ACTREC has got a beautiful campus with breathtaking views during monsoon. I am also thankful to Department of Science and Technology (DST), Government of India for providing me fellowship to pursue my PhD in ACTREC (DST-Inspire Fellowship) and ICMR-CAR for funding the project.

I would further like to acknowledge my Doctoral Committee members – Dr. Neelam Shirsat, Dr. Abhijit De, Dr. Sudhir Nair, Dr. Rita Mulherkar & Dr. Pradnya Kowtal. I am extremely thankful to them for critically evaluating my thesis work during all my doctoral committee meetings and for providing valuable suggestions towards completion of this thesis.

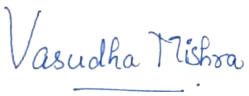
This work would have not been possible without the hard work of genetic counselors of the Cancer Genetics Clinic, Tata Memorial Hospital in patient's recruitment and sample collection. I would specially like to mention and thank Mr. Ravindra Reddy in recruiting and counseling medullary thyroid cancer patients on which this thesis work was carried out.

I would like to acknowledge all the central facilities of ACTREC including Genomics Facility, Common Instrument Facility, Library, SCOPE cell, Steno pool, Photography, IT department, Administration and Accounts department, Dispatch & Purchase department and last but not the least, ACTREC-Security staff. In some way or the other they all have helped me in completion of my thesis work smoothly. Here I would especially like to thank Mr. Naresh Mahida and Mrs. Sharada Haralkar from the Genomics facility for patiently and happily helping me whenever I rushed with hundreds of my samples tagging them as urgent. I would further like to thank Mr. Nilesh & Mr. Sunil for their help in bioinformatics based NGS data analysis and Mrs. Pallavi Rane for helping me with the statistical analysis.

This section would be incomplete without acknowledging the members of Sarin Lab – Nikhat, Moquit, Amogh, Pallavi Pandharipande, Dr. Preeti, Dr. Bhanu, Mr. Parag Madankar, Dr. Pallavi Shukla, Mrs. Prerana Dange, Mrs. Jyoti Patel, Sujath, Roshan, Imran as well as the trainees for their help, how big or small it might be. Here I would especially like to mention my sweet little girl – Snehal. She always made sure that the smile on my face doesn't fade a bit. I cannot thank enough my PhD seniors Nikhat & Moquit for all the guidance and support. I thank them for those much needed laughter and giggles, fond memories and the conviction I carried even on my darkest days. You both have been one of my biggest supports during this whole PhD journey.

I would also like to thank my batchmates for being there through my ups and downs. I remember all our batch trips, birthday parties and other get together. They used to be so refreshing and every time turned out to be a stress buster.

Lastly but importantly, I am grateful to God for blessing me with such a loving and caring sisters. They are my strength, motivation and source of my happiness. They are the ones who taught me not to quit, not breaking down but to rise every time with more vigor and perseverance.



Vasudha Mishra

TABLE OF CONTENTS

SYNOPSIS

Homi Bhabha National Institute

SYNOPSIS OF Ph.D. THESIS

1. **Name of the Student:** Vasudha Mishra
2. **Name of the Constituent Institution:** Tata Memorial Centre, ACTREC
3. **Enrolment No. & Date of Enrolment :** LIFE09201304018, 1st September 2013
4. **Title of the Thesis:** Elucidating the genetic pathways of RET driven Medullary Thyroid Carcinogenesis.
5. **Board of Studies:** Life Sciences

SYNOPSIS

Introduction

Medullary Thyroid Carcinoma (MTC) is a neuroendocrine tumor arising from neural crest-derived, calcitonin-producing parafollicular C-cells of the thyroid. It accounts for 3-4% of all thyroid malignancies[1]. Various studies have reported that around 75% of all MTC cases are sporadic, that is, without any family history or syndromic features whereas the remaining 25% are hereditary [2]. Hereditary MTC occurs as a component of an inherited cancer syndrome called Multiple Endocrine Neoplasia Type 2 or MEN2 with an autosomal dominant pattern of inheritance. This syndrome was first described by John Sipple in 1961 and has three clinical sub-

types: MEN 2A, MEN 2B and Familial MTC (FMTC). MEN 2A, which is the most common sub-type (55%) is characterized by presence of MTC or its precursor C cell hyperplasia (CCH), pheochromocytoma (rare tumor of adrenal chromaffin cells) and hyperparathyroidism [3] [35]. Two rare variants of MEN 2A have been identified, one where MEN2A is accompanied with Hirshsprung's disease which is marked by congenital absence of enteric ganglia and the other where MEN2A is accompanied with cutaneous lichen amyloidosis which manifests as a pruritic and pigmented lesion of skin on upper back [4]. The second subtype MEN 2B accounts for 5% of hereditary MTC cases. MEN2B is characterized by early onset and aggressive form of MTC, pheochromocytoma and is associated with distinct phenotypic features such as mucosal neuromas, thickened corneal nerves and marfanoid features. The third subtype that accounts for 35-40% of MEN2 is Familial Medullary Thyroid Carcinoma (FMTC) where the only clinical manifestation is familial MTC[35].

In 1993, RET proto-oncogene had been identified as the predisposing gene for the hereditary form of MTC [5]. Germline missense mutation in RET proto-oncogene is found in about 98% of hereditary MTC cases and even 6-8% of sporadic MTC cases. Somatic mutations in RET gene is reported in 40-60% of sporadic MTC tumors [1][28].

RET (REarranged during Transfection) proto-oncogene is located on chromosome 10q11.2 and has 21 coding exons. It was first identified in 1985 as a proto-oncogene which was able to undergo oncogenic activation by genetic rearrangement [6]. RET encodes a trans-membrane receptor tyrosine kinase which is expressed in neural-crest derived cell lineages including thyroid C cells and adrenal medulla where it plays a major role in cell proliferation, differentiation and survival. It also plays a role in the development of enteric nervous system and the kidney [2]. RET is a single pass receptor trans-membrane protein and has an extracellular ligand binding domain, a

trans-membrane and an intracellular tyrosine kinase domain. The extracellular domain includes four cadherin-like repeats (to induce and stabilize conformational changes required for receptor interaction with ligand and co-receptor) and a highly conserved cysteine rich region (maintains the tertiary structure of the protein and is also involved in receptor dimerization) [3]. The trans-membrane domain is involved in maintaining the close proximity of RET monomers through non-covalent receptor-receptor interaction [1]. The intracellular domain includes two tyrosine kinase subdomain (TK1 and TK2) which are involved in activation of downstream intracellular signaling pathways [3]. Alternative splicing generates three isoforms of RET: RET 9, 43 and 51, of which RET 9 and RET 51 are the functional isoforms in vivo [1].

Activation of RET is mediated by complex of ligand and co-receptors [3]. The ligands for RET activation are growth factors of glial cell line derived neurotropic factors (GDNF) family which includes GDNF, artemin (ARTN), persephin (PSPN) and neurturin (NRTN). These ligands in association with GDNF family of alpha receptors (GFRαs) form GFL/GFRα complex which in turn leads to receptor dimerization[3]. RET receptor dimerization further leads to phosphorylation of intracellular tyrosine kinase domain and subsequently activates the downstream signal transduction pathway [2]. RET signals through multiple downstream signaling pathways including RAS/ERK pathway, PI3K/AKT/NF-κB pathway, MAP Kinase and JAK/STAT pathways [2].

The role of RET proto-oncogene in MTC development and progression has been well studied and characterized [2][3][5][13]. Activating germline point mutation in RET has been identified in 98% MEN 2A, 95% MEN2B and 88% FMTC. Also, 40-60% of all sporadic MTC cases bear somatic point mutations in RET [2]. There is a strong and established correlation between site specific codon mutation within the RET gene and the phenotypic heterogeneity observed among patients with hereditary MTC [2][3]. For instance, majority of patients with MEN 2A harbour mutations in

exon 11 (codon 634) in cysteine rich extracellular domain of RET [13]. A few patients also harbour mutations in exon 10 (codon 609,610,611,618,620) of the cysteine rich extracellular domain of RET. Mutation in these cysteine residues leads to receptor homodimerization via the formation of intrarmolecular disulfide bonds leading to constitutive activated RET receptor even in the absence of ligand [2]. In MEN 2B, more than 95% cases have mutation in exon 16 (codon 918) and a few cases have mutation in exon 15 (codon 883) in the tyrosine kinase domain of RET. These mutations leads to RET activation in its monomeric state and alters its substrate specificity by structural alterations in the binding pocket of tyrosine kinase domain [3].Patients with FMTC harbor mutation in exon 10, 11, 13, (codon 768) and 14 (codon 804 and 806). Some germline RET mutations have weak transformation potential and will present the disease phenotype only in homozygous state. For example, mutation at codon 804 of exon 14 of RET is weakly transforming and causes low disease penetrance with late age at onset. Individuals heterozygous for such weakly transforming mutation require either a second mutation in RET or any other downstream signaling gene to present the clinical phenotype of MEN 2 [3].

Based on this strong genotype – phenotype correlation, the American Thyroid Association (ATA) has categorized germline RET mutations in 4 different risk categories based on the age of MTC onset and its aggressiveness. ATA level D mutations (M918T and A883F) are associated with highest risk of MTC which occurs in early childhood and is highly aggressive. ATA risk category C mutations are at codon 634 with substitution of cysteine by arginine or few other amino acids. Risk category C mutations are associated with a high lifetime risk of aggressive MTC. ATA risk category B mutations are at codon 609, 611, 618, 620 and 630 and present a moderate risk of MTC. ATA level A mutations occur at codon 768, 790, 791, 804 and 891 and have the lowest lifetime risk of developing MTC [1,14]. Although this strong genotype – phenotype correlation

based on the ATA system is widely accepted, it cannot explain the phenotypic heterogeneity seen in individuals carrying identical mutations within a family or in different families. The most common RET mutation is at codon 634 and it shows considerable inter and intra-familial phenotypic variability in terms of age of disease onset and its aggressiveness. This suggests the role of additional genetic alterations in other genes at either germline or somatic level which are not well understood.

In 98% of individuals with MEN2syndrome, the pathogenic mutation is specific to RET gene. However in the remaining 2% of MEN2 syndrome cases, there is a possibility of germline mutation in some other gene which is not known till date to be associated with MTC.

Hypothesis:

Role of RET proto-oncogene as a driver mutation in MTC pathogenesis is well established. Despite clear correlation between the site specific codon mutation in RET gene and the associated MEN2 phenotype, the molecular mechanism correlating the mutated RET receptor to different disease phenotype is not completely understood [15].**"There might be other germline or somatic genetic alterations that act in association with RET mutation to govern the inter-and intra-familial phenotypic variations seen in MTC cases."**

Aim of the study

To identify and characterize germline and somatic genetic variants and pathways associated with MTC and its phenotype (age at onset, aggressiveness) in a cohort of MTC patients withor without a germline RET mutation.

Objectives

1. To Identify novel and recurrent germline RET mutations, their penetrance and Genotype-Phenotype correlation in a cohort of Indian MTC cases.

2. To identify additional molecular signatures in RET driven MTC tumor samples of using Whole Exome Sequencing

3. To study the modulatory role of SNPs of distinct genetic pathways on risk of MTC development and its clinical behavior.

4. To identify driver mutation that could produce MEN2B phenotype in the absence of germline RET gene mutations.

Results

Objective 1

To Identify novel and recurrent germline RET mutations, their penetrance and Genotype-Phenotype correlation in a cohort of Indian MTC cases.

In this part of the study, 245 MTC cases which were enrolled in Cancer Genetics Clinic (CGC), Tata Memorial Centre, between 2014-2019 were included. Detailed family history & clinico-pathological details recorded using standard case record form. Blood sample was collected with informed consent & DNA was extracted. Samples were tested for hotspot RET mutations (6 exons) by PCR-Sanger Sequencing. Based on identified mutation & syndromic/phenotypic features, cases were classified as sporadic & hereditary (MEN2) MTC cases.

Deleterious germline RET mutations were identified in 39/245 (16%) MTC families tested. Of the 39 mutation carrier families, 50 first degree relatives were screened for family specific mutations and 16/50 (32%) were found to be mutation carrier. Therefore, a total of 55 mutation carrier

individuals (39+16) have been identified in our cohort. We have also identified one family with a double RET mutation (C634R + I852M). Most recurrent mutation was at codon 634 & was identified in 17/55 (31%) mutation carrier individuals (Figure 1).

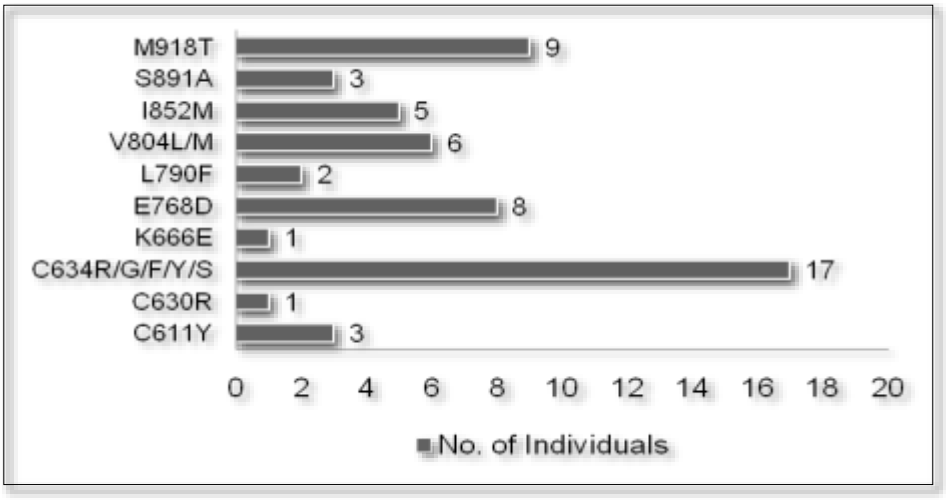


Figure 1: Recurrence of distinct germline RET mutations.

We have performed genotype-phenotype correlation between codon specific mutations in RET and MTC phenotype. Based on the genotype-phenotype correlation, in our cohort of 55 mutation carrier individuals, we have identified 17/55 (31%) MEN2A cases, 9/55 (16%) MEN2B cases and 29/55 (53%) FMTC cases. We have also grouped each RET mutation into different American Thyroid Association (ATA) risk category and calculated the age related penetrance of RET mutations of each risk category (Figure 2). We observed that risk category D mutation carriers had complete penetrance (100%) for MTC with an early age of disease onset & 65-70% lifetime risk of developing Pheochromocytoma. Risk Category C mutation carriers had 85% lifetime risk of MTC & 25% risk of developing Pheochromocytoma. Risk Category A&B mutation carriers had 70% lifetime risk of MTC with a late age of disease onset & these individuals were not at risk of developing pheochromocytoma (Figure 2).

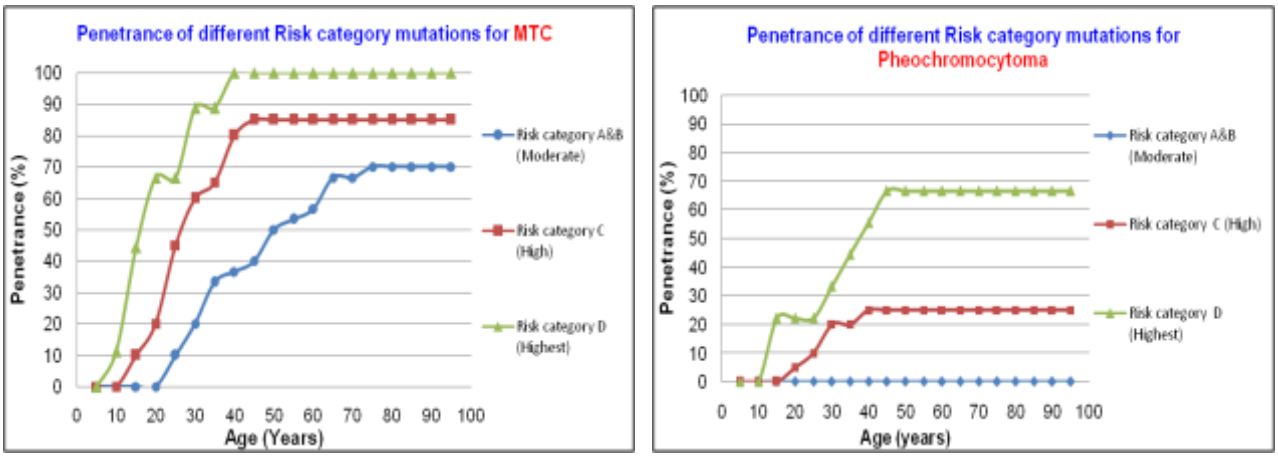


Figure 2: Age related penetrance of different ATA risk category mutations for MTC and Pheochromocytoma.

Summary (Objective 1):

In a cohort of 245 Indian MTC cases, we identified 39 deleterious germline RET mutation carrier families and 55 mutation carrier individuals including one family with double RET mutation. The high prevalence of codon 634 mutation could be explained by early onset of the disease and occurrence of pheochromocytoma/hyperparathyroidism that makes the referral and diagnosis earlier. Determining the Genotype-Phenotype correlation and the penetrance of different risk category mutations aids in clinical management of the patients, guiding prophylactic/therapeutic intervention and predictive genetic testing for family members.

Objective 2

To identify additional molecular signatures in RET driven MTC tumor samples of using Whole Exome Sequencing

In the only study of comprehensive genomic characterization of MTC done so far, Agarwal et al in 2013 [28] reported their findings of whole exome sequencing in 17 sporadic MTC tumors. They identified 305 high confidence somatic mutations in 283 genes. In this cohort of MTC cases without a germline RET mutation, the most common somatic mutation was in RET and observed in 12/17 (71%) sporadic MTC tumors followed by HRAS (18%) & KRAS (5%). Somatic mutations identified in RET, HRAS & KRAS were mutually exclusive. This study though small, confirmed the role of RET as the driver mutation not only in hereditary but also as a somatic mutation in sporadic MTC and that of RAS in RET negative tumors. However these studies could not explain the phenotypic heterogeneity observed in sporadic and hereditary MTC and could not identify any common downstream mutations. It is noteworthy that the mutational landscape of hereditary MTC with a known germline driver RET mutation has not been studied so far except for validation of somatic genetic alterations identified in sporadic MTC.

In the present study, Whole Exome Sequencing of 15 RET driven hereditary MTC fresh frozen tumor tissue and paired blood sample was done. These samples represent tumors arising from all the major classes of RET mutations – Risk Category B (C609R); Category C (C634R, C634W, C634Y) and Category D (M918T).

Sample Library was prepared using Agilent SureSelect v6 exome capture kit which enriches 50MB protein coding region from human genome corresponding to 21,000 genes. Sequencing was performed on HiSeq 2500 Illumina sequencer and data was analyzed using Varscan 2.0 pipeline. In the data coverage statistics, 99% reads were uniquely mapped with the reference genome and more than 95% target region was covered with a minimum of 20 reads (20X coverage).

Average mutation burden per Mb across all samples was 0.124. In a study published by Alexandrov et al, Nature; in 2013, they studied the tumor mutation burden (TMB) across different cancer types through whole genome sequencing and demonstrated that pilocytic astrocytoma had the lowest TMB of 0.04/Mb [29]. This cancer was shown to be driven by single abnormalities of MAPK pathway in all the cases. Similarly Lawrence et al; Nature; 2014 reported rhabdoid sarcoma to possess a low TMB of 0.1/Mb and is driven by germline or somatic genetic alteration in SMARCB1 gene [30]. These cancers with low tumor mutation burden (TMB) are driven by mutations in single gene or single genetic pathway that could also explain the low TMB observed in our cohort of MTC cases which are driven by mutations in RET gene.

Overall 94 high confidence somatic mutations were identified in 15 samples. The number of high confidence somatic mutation per tumor sample averaged 6.2 (range from 2 to 21 and SD $\pm$ 5.29). Mutations include: 66 missense mutations, 6 nonsense mutations, 3 frameshift mutations, 8 In-frame mutations and 9 splice-site mutations.19/66 missense variants were predicted to be pathogenic by 5 or more In-Silico pathogenicity prediction tools. None of the identified high confidence somatic mutation was found to be recurrent in any of the 15 samples studied. However mutations in a few a few important gene have been identified such as NF1, PTEN, BRCA2, PRSS3, ATXN3 and PRDM9.

Summary (Objective 2)

Our study on 15 paired MTC tumor tissues suggested that germline RET driven MTC have very few somatic mutations and the identified mutations are not recurrent. Extremely low somatic mutation burden is similar to rare brain tumors driven by single genetic pathways e.g. Pilocytic

astrocytoma & Rhabdoid tumors. Large cohort studies are therefore required to further understand the genomic landscape of RET driven MTC.

Objective 3

To study the modulatory role of SNPs of distinct genetic pathways on risk of MTC development and its clinical behavior.

Role of several SNPs as genetic modifiers of MTC phenotype have been studied, most common being the RET polymorphisms - G691S, L769L, S836S & S904S. Several studies have demonstrated that RET variants may be associated with the risk of MTC as well as its aggressive behavior [23,25]. However, other studies failed to demonstrate such associations [24,26]. Although the reason for the conflicting results remains unsolved, differences in the genetic background of distinct population have been suggested as a potential explanation.

Similarly, polymorphisms in genes of detoxification [21,22] are studied for their modifier role in MTC pathogenesis such as CYP1A2*F, CYP1A1m1, GSTP1, NAT2 & TP53. The variants were suggested to increase the risk of MTC and were not found to be associated with biological behavior of MTC. Recently, there have been reports associating polymorphisms of cell cycle regulatory genes with both the susceptibility & progression of MTC. One such study demonstrated that CDKN1B V109G polymorphism correlates with a more favourable disease progression than the wild type allele [19]. In another study, it was shown that CDKN1B & CDKN2A genes are associated with susceptibility to MTC, whereas, polymorphism of CDKN1A, CDKN2B & CDKN1C is associated with an aggressive behavior of MTC [27]. Further studies on a larger cohort will allow a better understanding on the role of these polymorphisms on MTC behavior.

With the largest cohort of 438 Indian MTC cases registered at our Cancer Genetics Clinic (CGC), Tata Memorial Hospital, between 2006 to 2018, the present study aimed to study the genetic risk association of 13 SNPs (Table 1) of genes belonging to 3 distinct genetic pathways with MTC and the association of these SNPs with clinical behavior of MTC in terms of age at disease diagnosis as well as its aggressiveness.

Table 1: List of SNPs selected for the study based on detailed literature review:

Polymorphisms of genes of Detoxification	**Polymorphisms of RET**	**Polymorphisms of Cell cycle regulatory genes**
Cyp1A1m1 (rs4646903) T/C	G691S (rs1799939) G/A	CDKN1A (rs1801270) C/A
Cyp1A2 (rs762551) A/C	L769L (rs1800861) T/G	CDKN1B (rs2066827) T/G
GSTP1 (rs1695) A/G	S836S (rs1800862) C/T	CDKN2A (rs11515) C/G
NAT2 (rs1041983) C/T	S904S (rs1800863) C/G	CDKN2B (rs1063192) T/C
		CDKN2C (rs12885) G/T

Objective 3a)

To study the genetic risk association of SNPs of distinct genetic pathways with Medullary Thyroid Carcinoma.

In this largest reported cohort of 438 Indian MTC cases (361 sporadic and 77 hereditary) and with 489 gender and ethnicity matched healthy controls from the South Asian cohort of the 1000 Genome Project, a comprehensive risk association study of all the above mentioned SNPs from the three genetic pathways was undertaken (Table 1). Further, based on detailed literature search, a

meta-analysis of all the reported case-control studies, including the present study was conducted on the risk association of the 4 *RET* gene SNPs with MTC susceptibility.

SNP genotyping was performed by RFLP or TaqMan method and the statistical analysis was performed using SPSS version 21.0. The SNP genotypes were tested for Hardy-Weinberg equilibrium (HWE) before performing the statistical analysis. As the homozygous genotype of several SNPs was either absent or very low in either the cases or controls, the analysis was performed using only the dominant model of inheritance (AA vs Aa+aa). Univariate and Multivariate logistic regression analysis was used to study the association of these SNPs with MTC risk. Odds ratio was calculated with 95% Confidence Interval (CI). Variables were considered eligible for multivariate logistic regression analysis when univariate analysis shows a trend of association with $p < 0.1$. As multiple comparisons were made in a single cohort for these 13 SNPs, a p value of <0.01 was used to consider any association to be statistically significant.

On univariate logistic regression analysis, the *CDKN1A* SNP presented a significant association with reduced risk of MTC in both the hereditary (OR = 0.52; 95% CI = 0.27-0.99; p = 0.048) and sporadic MTC groups (OR = 0.63; 95% CI = 0.45-0.88; p = 0.007). The variant allele A was overrepresented in the control population (26.2%) compared to both the hereditary cases (15.6%) and the sporadic cases (18.3%). Multivariate logistic regression analysis for CDKN1A SNP further confirmed that this SNP is strongly associated with a reduced risk of MTC in both the hereditary (OR = 0.27; 95% CI = 0.13-0.55; p = <0.001) and sporadic MTC groups (OR = 0.53; 95% CI = 0.36-0.78; p = 0.001). Multivariate logistic regression analysis also showed a significant risk association of the *RET* S904S SNP with hereditary MTC (OR = 2.82; 1.64-4.86; p = <0.001) and *CDKN2A* (OR = 1.89; 95% CI = 1.20-2.98; p = 0.006) and *NAT2* SNP (OR = 1.62; 95% CI = 1.17-2.25; p = 0.004) with sporadic MTC. Taken together, the logistic regression

analysis data indicated that the inheritance of *CDKN1A* SNP confers a protective effect in MTC development in both hereditary and sporadic MTC group whereas the inheritance of *CDKN2A, NAT2* and *RET* S904S variants increases the susceptibility to Medullary Thyroid Carcinoma.

In meta-analysis of hereditary MTC group, *RET* S836S and *GSTP1* SNPs were found to be associated with increased risk of MTC. This effect was observed under the random and fixed effect model for *RET* S836S variant with significant heterogeneity (I^2 = 72%, p = 0.01) whereas for *GSTP1* this effect was observed only with fixed-effect model with significant heterogeneity between studies (I^2 = 90%, p <0.01). We also observed *RET* L769L and *Cyp1A1m1* variant allele to be associated with reduced risk of hereditary MTC under both random and fixed effect models with no significant heterogeneity between studies. Similarly in meta-analysis of sporadic MTC group, we observed *RET* G691S, *RET* S904S and *CDKN1A* SNPs to be significantly associated with the risk of MTC. The G691S and S904S SNPs of *RET* were associated with increased risk of sporadic MTC whereas *CDKN1A* variant allele was showing a significantly protective effect on MTC development under both fixed effect and random effect model with no significant heterogeneity between studies (I^2 = 0%, p = 0.43).

Objective 3b)

To study the modulatory effect of SNPs of distinct genetic pathways on clinical behavior of Medullary Thyroid Carcinoma

A comprehensive gene dose-response relationship analysis between 13 SNPs of genes of three distinct genetic pathways was undertaken (Table 1). This gene dose-response relationship aimed to investigate the clinic-pathological differences between the wild-type versus heterozygous versus homozygous variant for each SNP separately in hereditary and sporadic MTC cases. The clinico-

pathological parameters with which the correlation was made included: Age at MTC diagnosis, Tumor Marker – Serum Calcitonin levels at diagnosis (indicative of tumor burden), Tumor Volume (cubic centimeter), regional nodal Metastasis and distant metastasis.

In gene dose-response association, the only clinic-pathological parameter with which any SNP showed significant association was nodal metastasis at diagnosis. In hereditary MTC group, *Cyp1A1m1*, *CDKN2A* and *CDKN2C* SNP showed this association. Patients with wildtype *Cyp1A1m1* showed higher rate of regional nodal metastasis compared to the heterozygous and homozygous variants (95.6% vs 86.2% vs 33.3%; p=0.01). Similarly patients with wildtype *CDKN2A* showed a higher incidence of nodal metastasis compared to their variant counterparts (91.3% vs 83.3% vs 33.3%; p=0.01). For CDKN2C, the comparison could only be made between wildtype and heterozygote as the homozygous variant for this SNP was absent. We observed that cases heterozygous for *CDKN2C* SNP showed higher rate of lymph node metastasis compared to wildtype *CDKN2C* (100% vs 86.3%; p=0.01).

In sporadic MTC group, this association was observed between the wildtype and the homozygous variant for *CDKN2C* SNP. The cases wildtype for *CDKN2C* SNP showed higher rate of nodal metastasis compared to homozygous variant genotype (83.8% vs 40%; p = 0.03). The rate of nodal metastasis for heterozygous variant was similar to the wildtype *CDKN2C* (84.4% vs 83.8%). No other significant association was observed between any of the SNPs and the patient's clinico-pathological behavior.

Summary (Objective 3):

This study is the first single cohort study to perform a comprehensive risk association analysis of 13 most frequently studied SNPs with MTC development as well as analysis of modulatory role of

these SNPs with clinical outcome of MTC patients. In this largest reported risk-association study of 13 different SNPs with MTC, a significant protective risk-association of *CDKN1A* Ser31Arg SNP with MTC was reported for the first time. In our expanded meta-analysis, we have identified significant risk-association of all the 4 *RET* gene SNPs with MTC, not seen in the previously reported meta-analysis. In our extended analysis on both hereditary and sporadic MTC cohort we reported no significant association of these SNPs with clinical behavior of MTC.

Objective 4

To identify driver mutation that could produce MEN2B phenotype in the absence of germline RET gene mutations.

In MEN2B cases, the most frequently found mutation is at codon 918 (M918T) in exon 16 of RET gene which is found in almost 95% MEN2B cases[3]. The remaining 5% cases either harbor mutation at codon 883(A883F) or a double mutation in RET gene. In our cohort of MEN2B cases, we could not identify M918T or any other mutation in RET gene on Sanger Sequencing in 3 classical MEN2B cases. These cases were taken for Whole Exome sequencing to identify driver mutation in other gene producing MEN2B phenotype in the absence of germline mutation in RET gene.

Exome Sequencing libraries were prepared using Nextera Rapid Capture Exome Kit v2 & 150x2bp paired end sequencing was performed on Hiseq 2500 NGS Sequencer. Exome sequencing data identified M918T mutation at 100X coverage in all the 3 MEN2B cases which were initially missed upon Sanger Sequencing. This observed phenomenon is called Allele dropout, a genotyping error, with preferential amplification of one of the two alleles where the other allele is under-represented or completely absent. Allele drop out may result from sequence

dependent factors or sequence independent factors. Sequence dependent allele dropout occurs when there is a single nucleotide variant present at primer binding site on any of the two allele which results in failure of primer binding to the template & causes allele specific allele dropout [31] or there could be polymerase hindering secondary structures within the sequence of an allele leading to failure of its amplification [32]. Sequence independent allele dropout could be caused by variety of sampling or molecular events such as DNA Quantity or Quality, thermocycler temperature imprecisions, variation in pipetting volumes of reagents, polymerase activity/quality etc [33,34]. Unfortunately, most of these causes are difficult to identify precisely because of their unpredictable & non-reproducible behavior [34].

In our cases, we observed allele dropout due to sequence independent factors. While we could confirm allele dropout, its cause could not be identified despite changing several PCR conditions and varying the concentration of different reagents and DNA. The only variable which could not be matched in replicative analysis was the batch of reagents as several years had lapsed between the initial Sanger Sequencing with false negative report and the replicative Sanger Sequencing which identified the allele dropout. It is therefore quite likely that allele dropouts observed in our study was due to the quality of reagents used for initial rounds of sequencing or due to equipment artifacts. Our failure to identify the exact cause of the Allele dropout despite a very systematic examination is not surprising as others have also reported that sequence independent Allele Dropouts are usually unpredictable, non reproducible and their exact cause remains unknown.

Summary (Objective 4)

This is the first report of series of clinically relevant sequence independent allele dropout in any cancer predisposing gene. This study highlights the importance of syndromic diagnosis based on phenotypic characterization, the strongest indication of genotyping errors such as allele dropout in cases showing Mendelian inheritance discrepancy.

Conclusion: Significance of the Study

The present study is the first Indian study to establish genotype-phenotype correlation and penetrance of different risk category RET mutations. This is useful for counseling and risk management of mutation carriers

This is also the largest and most comprehensive risk-association study of 13 SNPs of three distinct genetic pathways with MTC, which has shown for the first time a significant protective risk-association of *CDKN1A* Ser31Arg SNP with MTC. In the largest and most recent meta-analysis including our cases, we have observed a significant risk association of the 4 *RET gene* SNPs with MTC, not reported in previous meta-analysis.

Further, this is the first Exome analysis of germline RET driven MTC cases which reveals extremely low somatic mutation burden, similar to some rare forms of single genetic pathway driven brain tumors.

This is also the first report of series of clinically relevant sequence independent allele dropout in any cancer predisposing gene. Considering the major clinical implication of false negative genetic test result, our study could serve as an alert to systematically address the issue of sequence independent allele dropout in cases with Mendelian inheritance discrepancy.

Future Directions

The molecular mechanism through which the CDKN1A SNP or RET gene SNPs modifies the risk of MTC development needs to be elucidated through functional studies. Also, the genes and mutations identified in our exome sequencing studies on RET driven MTC tumor tissues needs further validation in a larger cohort studies to draw definitive conclusions.

References:

1. Medullary thyroid carcinoma (MTC) and RET proto-oncogene: Mutation spectrum in the familial cases and a meta-analysis of studies on the sporadic form. Figlioli G. *et al*: Mutation Research 752 (2013) 36–44

2. RET oncogene in MEN2, MEN2B, MTC, and other forms of thyroid cancer: molecular genetics and therapeutic advances.Lodish M.B. *et al*:Expert Rev Anticancer Ther. 2008 April ; 8(4): 625–632

3. RET as a Diagnostic and Therapeutic Target in Sporadicand Hereditary Endocrine Tumors. de Groot *et al:* Endocrine Reviews, August 2006, 27(5):535–560

4.RET proto-oncogene, a review and update of genotype-phenotypecorrelations in hereditary medullary thyroid cancer and associated endocrine tumors. Kouvaraki MA *et al:* Thyroid 2005;15(6):531–544.

5. Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. Mulligan LM *et al*: Nature 363, 458-460 (3 June 1993)

6. Activation of a novel human transforming gene, RET, by DNA rearrangement. Takahashi M *et al*: Cell 1985;42(2):581–588.

7.RET-Mediated Gene Expression Pattern is Affected by isoform but Not Oncogenic Mutation. Hickey JG *et al:* GENES, CHROMOSOMES & CANCER 48:429–440 (2009)

8.Molecular genetics of medullary thyroid carcinoma: the quest for novel therapeutic targets. Cerrato A. *et al:* Journal of Molecular Endocrinology (2009) 43, 143–155

9. Multiple endocrine neoplasia. Gagel FR & Marx SJ 2003: In Williams Textbook of Endocrinology, 10 edn, pp 1717–1762.

10. Amplification and Overexpression of Mutant RET in Multiple Endocrine Neoplasia Type 2-associated Medullary Thyroid Carcinoma.Huang SC *et al:* The Journal of Clinical Endocrinology & Metabolism 88(1):459–463

11.Chromosome 10 and RET gene copy number alterations in hereditary and sporadic Medullary thyroid carcinoma.RaffaeleCiampi*et al*: Molecular and Cellular Endocrinology 348 (2012) 176–182

12.Specific mutations of the RET proto-oncogene are related to disease phenotype in MEN 2A and FMTC. Mulligan LM *et al*, Nat. Genet. 6 (1994) 70–74.

13.The relationship between specific RET proto-oncogene mutations and disease phenotype in multiple endocrine neoplasia type 2.Eng C.*et al* : International RET mutation consortium analysis, JAMA 276 (1996) 1575–1579

14. Medullary Thyroid Cancer: Management Guidelines of the American Thyroid Association.The American Thyroid Association Guidelines Task Force.Kloos RT et al: THYROID :Volume 19, Number 6, 2009

15.RET-Familial Medullary Thyroid Carcinoma Mutants Y791F and S891A Activate a Src/JAK/STAT3 Pathway, Independent of Glial Cell Line–Derived Neurotrophic Factor. Ivan Plaza Menacho*et al:* Cancer Res 2005;65:1729-1737

16. Differential Gene Expression of Medullary Thyroid Carcinoma reveals specific markers associated with Genetic Conditions. Maliszewska A *et al*: The American Journal of Pathology, Vol. 182, No. 2, February 2013

17. Mutational screening of RET, *HRAS*, *KRAS*, *NRAS*, *BRAF*, *AKT1*, and *CTNNB1* in medullary thyroid carcinoma. Schulten *et al:* Anticancer Research. 2011.

18. High prevalence of RAS mutations in RET-negative sporadic medullary thyroid carcinomas. Moura et al.J Clin Endocrinol Metab. 2011.

19. CDKN1B V109G polymorphism a new prognostic factor in sporadic medullary thyroid carcinoma. Pasquali D et al. European Journal of Endocrinology, 2011.

20. P18 is a tumor suppressor gene involved in human medullary thyroid carcinoma and pheochromocytoma development. Veelen et al, Int.J.Cancer, 2009.

21. Evidence that polymorphisms in detoxification genes modulate the susceptibility for sporadic medullary thyroid carcinoma. Barbieri et al, European J. Endocrinology, 2012.

22. Genes of detoxification are important modulators of hereditary medullary thyroid carcinoma risk. Barbieri et al, Clinical Endocrinology, 2013.

23. Additive effect of RET polymorphisms on sporadic medullary thyroid carcinoma susceptibility and tumor aggressiveness. Ceolin et al, European J. Endocrinology, 2012.

24. Evaluation of the role of *RET* polymorphisms/haplotypes as modifier loci for MEN 2, and analysis of the correlation with the type of *RET* mutation in a series of Spanish patients. Fernandez et al, Int.J.Mol.Med, 2006.

25. Polymorphisms G691S/S904S of *RET* as Genetic Modifiers of MEN 2A. Robledo et al, Cancer Research, 2003.

26. Quantitative assessment of the association between L769L and S836S polymorphisms at RET gene and medullary thyroid carcinoma risk. Zhang et al, 2014.

27. Polymorphisms of cell cycle control genes influence the development of sporadic Medullary thyroid carcinoma. Barbieri RB et al, Eur J Endocrinol, 2014.

28. Exomic Sequencing of Medullary Thyroid Cancer reveals Dominant & Mutually Exclusive Oncogenic Mutations in RET & RAS. Agrawal N et al,J Clin Endocrinol Metab 2013.

29. Signatures of mutational processes in human cancer. Alexandrov et al.; Nature; 2013.

30. Discovery and saturation analysis of cancer genes across 21 tumor types. Lawrence M et al.; Nature 2014.

31. Allelic drop-out may occur with a primer binding site polymorphism for the commonly used RFLP assay for the -1131T>C polymorphism of the Apolipoprotein AV gene. Ward, K. J. *et al.Lipids Health Dis.* **5,** 1–6 (2006).

32. Allele dropout caused by a non-primer-site SNV affecting PCR amplification - A call for next-generation primer design algorithm. Lam, C. W. & Mak, C. M. *Clin. Chim. Acta* **421,** 208–212 (2013).

33. Genotyping errors: Causes, consequences and solutions. Pompanon, F., Bonin, A., Bellemain, E. & Taberlet, P. *Nat. Rev. Genet.* **6,** 847–859 (2005).

34.Risk of misdiagnosis due to allele dropout and false-positive PCR artifacts in molecular diagnostics analysis of 30,769 genotypes. Blais, J. *et al. J. Mol. Diagnostics* **17,** 505–514 (2015).

35. Change in the spectrum of RET mutations diagnosed between 1994 and 2006. Frank-Raue K, Rondot S, Schulze E, Raue F. Clin Lab. 2007;53(5–6):273—282.

Publications in Refereed Journal:

a. Published

Mishra V, Kowtal P, Rane P, Sarin R. Genetic risk association of CDKN1A and RET gene SNPs with medullary thyroid carcinoma: Results from the largest MTC cohort and meta-analysis. *Cancer Med*. 2019; 8: 6151–6161. DOI: 10.1002/cam4.2443.

b. Accepted: Nil

c. Communicated:

1. Modulatory role of Single Nucleotide Polymorphisms of distinct genetic pathways on clinical behavior of Medullary Thyroid Carcinoma.

Authors: Vasudha Mishra, Pradnya Kowtal, Pallavi Rane and *Rajiv Sarin.

Communicated in **"Asian Pacific Journal of Cancer Prevention".**

2. False negative *RET* genetic testing in MEN2B cases due to Allele Dropout detected on exome sequencing: A clinical alert.

Authors: Vasudha Mishra, Pradnya Kowtal and *Rajiv Sarin.

Under Peer Review in **"Familial Cancer" Journal.**

Other Publications:

a. Book/Book Chapter: Nil

b. Conference/Symposium

1. **Poster Presentation**: RET Genotype & MEN2 Phenotype correlation in a large Indian Medullary Thyroid Carcinoma (MTC) cohort & influence of SNPs in 3 genetic pathways on MTC behavior. **European Human Genetics Conference, 2018, *Milan Italy*.**

2. **Poster Presentation:** "Penetrance of distinct germline RET mutations in the largest Indian MEN2 syndrome cohort". **4th Indian Cancer Genetics conference & 3rd Indo-UK Cancer Genetic Conference 2018, ACTREC, Tata Memorial Centre, Navi Mumbai.**

3. **Poster Presentation**: Increased Sensitivity of Next Generation Sequencing over Sanger Sequencing in Mutation Detection: A call for reanalysis of False-Negative Sanger Results**. 9th HOPE Meeting with Nobel Laureates organized by Japan Society for Promotion of Science (JSPS), 2017, *Tokyo Japan***.

4. **Poster Presentation:** Utility of Next Generation Sequencing (NGS) in detecting germline allele dropouts in patients with hereditary cancer syndromes. **2nd International conference on Founder Population with the theme "The Landscape of Genetic Variants in Asian Founder Populations", 2017 *Kochi, Kerela, India.***

5. **Poster Presentation**: Sequence Independent Allele Dropout event on Sanger Sequencing in classical MEN2B cases detected by Next Generation Sequencing. **3rd Indian Cancer Genetics Conference & Workshop 2016, ACTREC, Tata Memorial Centre, *Navi Mumbai, India.***

6. **Poster Presentation**: Influence of SNPs in distinct genetic pathways on tumor burden & metastasis in Medullary Thyroid Carcinoma patients. **12th International Conference of the Asian Clinical Oncology Society (ACOS) & 35th Annual Convention of Indian Association for Cancer research &**

Mid-Term Conference of IASO, 2016 ***New Delhi, India.*** **Theme: CANCER IN ASIA: BRIDGING THE GAPS.**

7. **Poster Presentation**: Is CDKN1B V109G polymorphism a prognostic marker for Medullary Thyroid Carcinoma. **2nd Indian Cancer Genetics Conference & Workshop, 2014, ACTREC , Tata Memorial centre,** ***Navi Mumbai, India.***

Vasudha Mishra
Signature of Student:

Date: 8/8/19

Doctoral Committee:

S. No.	Name	Designation	Signature	Date
1.	Dr. Neelam Shirsat	Chairman		5.8.19
2.	Dr. Rajiv Sarin	Guide/ Convener		7.8.19
3.	Dr. Abhijit De	Member		8.8.19
4.	Dr. Sudhir Nair	Member		8.8.19
5.	Dr. Rita Mulherkar	Technical Advisor		8.8.19
6.	Dr. Pradnya Kowtal	Technical Advisor		7/8/19

Forwarded Through:

S.V. Chiplunkar 19/8/19

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

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

20/8/19

Prof. S. D. Banavali,
Dean (Academics)
T.M.C.

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

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

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Abbreviations	
ADO	Allele DropOut
ATA	American Thyroid Association
BAM	Binary Alignment/Map
CCH	C-Cell Hyperplasia
CDK	Cyclin Dependent Kinase
CGC	Cancer Genetic Clinic
CI	Confidence Interval
CLA	Cutaneous lichen amyloidosis
CT	Computed Tomography
dNTPs	Deoxyribonucleotide triphosphate
EBV	Epstein-Barr Virus
EDTA	EthyleneDiamineTetraAcetic Acid
EGFR	Epidermal Growth Factor Receptor
EtBr	Ethidium Bromide
FAP	Familial Adenomatous Polyposis
FMTC	Familial Medullary Thyroid Carcinoma
FNAC	Fine Needle Aspiration
FP	False Positive
FRET	Fluorescence Resonance Energy Transfer
GATK	Genome Analysis Tool kit
gDNA	Genomic Deoxyribonucleic acid
GDNF	Glial cell line-Derived Neurotrophic Factor
GFLs	GDNF family of Ligands
GFRαs	GDNF Family of α Receptors
HBOC	Hereditary Breast and Ovarian Cancer
HBV	Hepatitis B Virus

HIV	Human Immunodeficiency Virus
HPV	Human Papilloma Virus
HSCR	Hirschprung Disaese
HTLV-1	Human T-cell Leukemia Virus
HWE	Hardy-Weinberg Equilibrium
IARC	International Agency for Research on Cancer
IGV	Integrated Genome Viewer
LFS	Li-Fraumini Syndrome
LOH	Loss of Heterozygosity
LS	Lynch Syndrome
MEN2	Multiple Endocrine Neoplasia type2
MTC	Medullary Thyroid Carcinoma
NCBI	National Centre for Biotechnology Information
NCI	National Cancer Institute
NCRP	National Cancer Registry Program
NF	Neurofibromatosis
NGS	Next Generation Sequencing
PBCRs	Population Based Cancer Registries
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
PDGF	Platelet Derived Growth Factor
PJS	Peutz-Jeghers syndrome
PTH	Parathyroid Hormone
RBC	Red Blood Cell
RET	Rearranged during Transfection
RFLP	Restriction Fragment Length Polymorphism
RSV	Rous Sarcoma Virus
SAM	Sequence Alignment/Map
SEER9	Surveillance, Epidemiology, and End Results 9

SD	Standard Deviation
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
TBE	Tris Borate EDTA
TGFα	Transforming Growth Factor α
TMH	Tata Memorial Hospital
TSGs	Tumor Suppressor Genes
T3	Triiodothyronine
T4	Tetraiodothyronine (Thyroxine)
UCSC	University of California Santa Cruz
UV	Ultra Violet
VAF	Variant Allele Frequency
VCF	Variant Call Format
VHL	Von Hippel Lindau syndrome
WBCs	White Blood Cells
XP	Xeroderma Pigmentosa

Chapter 1

Introduction

1.1 Introduction to Cancer

Cancer can be defined as an enormous spectrum of diseases that all originates from uncontrolled cellular proliferation where a group of abnormal cells grow uncontrollably by disregardingthe normal rules of cell division. Under normal conditions, the cells of our body are under strict regulation of molecular signals which decides whether the cell should divide, differentiate or die. Cancer cells escapes these regulations and develop a degree of autonomy from these signals, resulting in uncontrolled growth and proliferation(1). If this proliferation continues it results in spreading of cancer cells to other sites in the body through the multistep process of metastasis. In fact, almost 90% of cancer-related deaths are due to metastasis of the disease (1).

After cardiovascular diseases, cancer is the second leading cause of death worldwide. As per the GLOBOCAN 2018 database released by the International Agency for Research on Cancer (IARC) in September 2018, the global cancer burden in 2018 has risen to 18.1 million new cases with 9.6 million deaths. One in 5 men and one in 6 women worldwide is developing cancer during their lifetime, and one in 8 men and one in 11 women die from the disease. Worldwide, the 5-year prevalence is estimated to be 43.8 million (2). Several factors are responsible for this increase in the cancer burden including population growth and ageing as well as the changing prevalence of certain causes of cancer linked to social and economic development. Global patterns show that for men and women combined, nearly half of the new cases and more than half of the cancer related deaths worldwide in 2018 are estimated to occur in Asia, in part because the region has nearly 60% of the global population (2).

In India, more than 1 million new cases of cancers are diagnosed every year in a population of 1.3 billion (3,4) and this number is expected to double by 2035 (4).The most common cancers in

Indian men include oral cavity, lung, stomach, colorectal and esophagus whereas in women, the most prevalent cancers include cancers of the breast, cervix, oral cavity and colorectum. In developing countries like India, patients with cancer generally have a poorer prognosis because of relatively low cancer awareness, late diagnosis, and the lack of affordable curative services compared with patients in high-income countries(3,4).

1.2 Etiology of Cancer

The question as to what causes cancer has intrigued researchers for generations. The aim was to identify the tumor initiators and promoters as the most effective and desirable way in fight against cancer. There are several environmental, occupational and lifestyle related risk factors contributing to cancer development along with the genetic risk factors.

- **Occupational factors**

The earliest carcinogens to be identified were associated with specific occupations. Sir Percivel Pott in 1775 reported that chimney sweeper boys frequently developed cancer of the scrotum which could be caused by heavy exposure to soot(5). Joseph Bell described two cases of scrotal cancer among shale oil workers in Scotland in 1876, and deduced that the cancer was very common among shale oil workers(6). Afterwards several reports emerged that a variety of occupational exposures such as chimney soot, coal tar, dye chemicals and X-rays predisposes individuals to specific types of cancers.

- **Environmental and Lifestyle related factors**

Among the environmental factors ultraviolet radiations are the major susceptibility factor to cancer development. Melanomas (skin cancers) are caused mainly due to high exposures of

ultraviolet radiation in the sun's rays and tanning salons (1). Paul Unna, a general physician who took a strong interest in skin diseases was the first to establish the association of ultraviolet rays and skin cancer(7).

John Hill in 1761(8) reported that immoderate use of tobacco snuff was associated with the occurrence of nasal cancers. Doll and Hill in the year 1950 were the first to show that cigarette smoking was correlated with the incidence of lung cancer(9). Among the other lifestyle related factors consumption of alcohol, red meat, processed meats, salted fish, low fibre diets along with no breast feeding and obesity are associated with susceptibility to cancer.

- **Biological agents**

In the 1980s ground breaking evidence began to emerge that a variety of viruses cause cancer in humans. Robert Gallo in 1980, identified for the first timea retrovirus which causes leukemia in humans and termed it as human T-cell leukemia retrovirus (HTLV-1) (10). Further in 1981, Robert Beasley found that Hepatitis B virus (HBV) was strongly associated with hepatocellular carcinoma through a prospective study of 22,707 men in Taiwan(11). These findings lead to the discovery of several viruses to be associated with human malignancies including human immunodeficiency virus (HIV), hepatitis C virus, human papillomavirus (HPV), Epstein-Barr virus (EBV), and human herpes virus.

Other than viruses, other biological agents such as parasites and bacteria are also known to cause cancers. Schistosoma haematobium (*S. Haematobium*), a parasitic fluke has been reported to be associated with bladder cancer(12). Similarly, *Helicobacter pylori (H. pylori*), a bacteria residing in gastric mucosa is known to be associated with gastric cancer (13).

- **Genetic risk-factors**

Specific genetic alterations can confer genetic susceptibility for development of certain cancer. Genetic alteration in certain genes confer a very high lifetime risk i.e. they are highly penetrant whereas the common variants such as SNPs have low penetrance and through their interaction with the environmental factors they may result in progression of cancer or act as disease modifiers.

1.3 The Hallmarks of Cancer

The fundamental theory of cancer development relies on the fact that cancer cells have defects in the regulatory circuits that govern normal cell proliferation. There are more than 100 different types of malignancies originating in different organs of the body (14). This complexity provokes a number of questions such as how many distinct regulatory circuits must be disrupted in order for a normal cell to become cancerous and whether same set of regulatory circuits undergo disruption across all different types of malignancies(14). To answer these questions Hanahan and Weinberg in the year 2000 proposed the six hallmarks or alterations of cancer development that collectively transform a normal cell into a malignant cell (Figure 1.1). These six hallmarks of cancer development include Self sufficiency in growth signals, insensitivity to anti-growth signals, tissue invasion and metastasis, evading apoptosis, limitless replicative potential, and sustained angiogenesis (14).

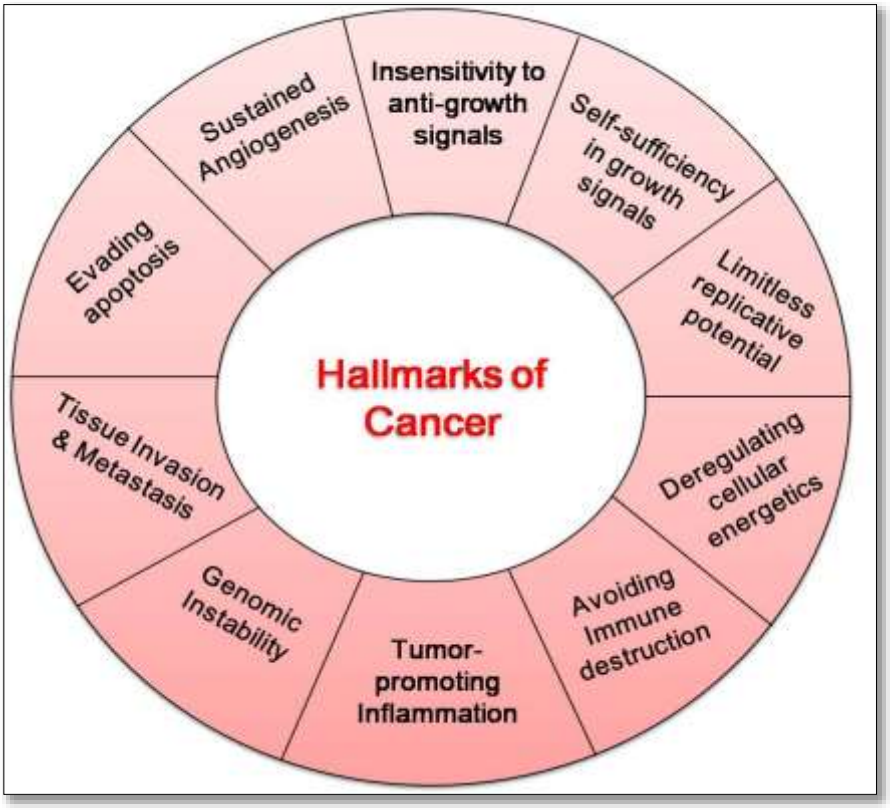


Figure 1.1 : Hallmarks of Cancer: Adapted from Hanahan & Weinberg, 2000 & 2011 (14,15)

The hallmarks are the functional capabilities acquired by the cancer cells to proliferate, survive and metastasize. These properties are acquired by different cancer types via different mechanism during the course of multistep carcinogenesis. Later, a few other distinct characteristics of cancer cells have been proposed to be functionally important for the development of cancer and were therefore added to the list of the core hallmarks (Figure 1.1). These include genomic instability, avoiding immune destruction, tumor-promoting inflammation and deregulating cellular energetics (15).These hallmarks of cancer provide a strong foundation in understanding the biology of cancer.

1.4 Classification of Cancer

1.4.1 Classification by Tissue of Origin

Based on their primary site of origin, there are more than 100 different types of cancer which are further classified on the basis of the tissue type from which these cancers originates (histological type)[Cancer Classification, 2004]. From the histological standpoint, different cancers are grouped into the following major categories:

- **Carcinoma:** These are the malignant neoplasm of epithelial origin which are the cells that form the external and internal lining of the body. Carcinomas are the most common type of cancer and accounts for 80-90% of all cancer. Carcinomas are divided into four major subtypes: Adenocarcinoma, cancer that originates from the epithelial cells that produce mucous (cancer of mucous membrane); Squamous cell carcinoma, cancer that originates from the squamous cells, which are epithelial cells that lie just beneath the outer surface of the skin; Transitional cell carcinoma, that originates from the transitional epithelium; and Basal cell carcinoma, that begins from the basal layer of the epidermis.

- **Sarcoma:** Sarcoma refers to cancers that originate from the supportive and connective tissues of the body such as bones, tendons, cartilage, muscle, and fat. Sarcomas are grouped into two categories: Osteosarcoma and Soft tissue sarcoma. Osteosarcoma is the most common cancer of bone. Examples of soft tissue sarcoma include leiomyosarcoma (smooth muscle), Kaposi sarcoma, liposarcoma (adipose tissue) etc.

- **Leukemia:** Leukemias, also called as "liquid cancers" or "blood cancers" are cancers of the bone marrow. The disease is associated with overproduction of abnormal white blood cells crowding out the normal blood cells. Leukemia is the most common childhood cancer.

- **Lymphoma:** Lymphoma is the cancer that originates from lymphocytes (T cells or B cells). These cells are a type of WBCs that function as part of the immune system in eliminating foreign invaders from the body called antigens. In lymphoma, abnormal lymphocytes accumulate in lymph nodes, lymph vessels as well as in other organs of the body. There are two types of lymphoma: Hodgkin lymphoma, in which the abnormal B-lymphocytes are produced and Non-Hodgkin lymphoma, which can be formed from both the abnormal B-lymphocytes or T-lymphocytes

- **Myeloma:** Myeloma is the cancer type that originates from another type of immune cells called plasma cells. These abnormal plasma cells, also called as myeloma cells, accumulates in the bone marrow and form tumors in bones all through the body.

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

- **Other types:** These include germ cell tumors which are a type of tumor that begins in the pleuripotent cells that give rise to sperm or eggs and neuroendocrine tumors arising from cells derived from neural crest.

1.4.2 Classification by Grade

Cancers can also be classified according to their grade. Grading involves examining tumor cells under a microscope. The grade of the cancer is determined by the extent of abnormality of these cells with respect to the surrounding normal tissues. Increasing abnormality increases the grade from 1–4(18).

- **Grade 1**: Cells with slight abnormality but are well differentiated.
- **Grade 2**: Cells more abnormal and are moderately differentiated.
- **Grade 3**: Cells highly abnormal and are poorly differentiated.
- **Grade 4**: Immature, primitive and undifferentiated.

Cells that are well differentiated and closely resemble the normal specialized cells belong to less aggressive low grade tumors whereas cells that are immature and undifferentiated belong to highly aggressive high grade tumors.

1.4.3 Classification by Stage

Staging is the classification of cancer based on the the extent of the disease spread. There are several types of staging methods. The two most common are the TNM staging and the numerical staging (Stage 0 –Stage IV). In the TNM staging, 'T' determines the tumor size, 'N' suggests the regional nodal spread of the disease and 'M' denotes the degree of distant metastasis. T0 signifies no evidence of tumor, T 1 to 4 signifies increasing tumor, N0 signifies no nodal involvement and N 1 to 4 signifies increasing degrees of lymph node involvement, M0 signifies no evidence of distant spread while M1 signifies distant spread of the disease.

Similarly in numerical staging, Stage 0 indicates cancer being in situ or limited to surface cells, Stage I indicates cancer being limited to the tissue of origin, Stage II indicates limited local spread, Stage III indicates extensive local and regional spread while stage IV is advanced cancer with distant metastasis(16,18).

1.4.4 Classification by Population at risk

Majority of cancers occur as a result of lifestyle choices or by environment conditions. However a few proportion of cancers develop due to inherited gene mutations and cancer risk in such families are much higher than in general population. Based on the population at risk, cancer has been classified as:

- **Hereditary Cancer:** These cancers accounts for 2-5% of all cancers and are caused by mutations in germ cells which are then passed on from parents to their children. Such cancersare monogenic and follow Mendelian inheritance pattern (autosomal dominant or recessive transmission). Individuals with hereditary cancer are identified if they have characterstic phenotypic features of an inherited cancer syndrome or they have a strong family history. In the absence of a family history or characteristic syndromic features, cancer is sometimes suspected to be hereditary if it has an early age of onset, presence of multiple primary cancers or bilateral cancers in an individual.

- **Sporadic Cancer:** These are the most common form of cancers and contribute to almost 80% of all cancers(19).Sporadic cancers do not show familial aggregation or Mendelian pattern of inheritance and arise from genetic alterations acquired by an individual during

their lifetime from environmental exposures, dietary factors, hormones, normal aging, and other influences. These acquired genetic alterations are present in somatic cells of the body and therefore are not inherited.

- **Familial Cancer:** Cancers that occur in a family more often than would be expected by chance but not in a classical Mendelian inheritance pattern, are called familial cancers(20). In familial cancers, several individuals across wide range of ages have similar cancer type(s) but without any specific pattern of inheritance. They account for 5-10% of all cancers and may result from common genetic background, similar environment and or lifestyle factors(21).

1.5. Cancer: A Genetic Disease

Cancer in essence, is a genetic disease resulting from congenital or acquired genetic alterations in certain cells of the body that accumulate over lifetime(22). In the last few decades, many important cancer predisposing genes have been discovered, their mutations have been precisely identified and the pathways through which they act have been characterized(23).Majority of these genetic alterations are somatic and occurs as a result of exposure to environmental carcinogens such as chemical carcinogens, radiation exposure or tumor viruses(24). However, germline genetic alterations although rare, also contribute to cancer initiation and progression. These alterations may either result from a single nucleotide change or may involve larger stretch of DNA resulting from deletions, duplications or large genomic rearrangements. Sometimes these genetic changes do not alter the actual DNA sequence and are called as epigenetic modifications. Such modifications regulate gene expression without actually disturbing the normal DNA sequence. The majority of genetic changes in the process of carcinogenesis fall into

two categories: gain-of-function mutations in proto-oncogenes, which stimulate cell growth, division, and survival; and loss-of-function mutations in tumor suppressor genes that normally help prevent unrestrained cellular growth and promote DNA repair and cell cycle checkpoint activation(25).

1.5.1 Proto-Oncogenes

Proto-oncogenes are a group of genes that encodes for proteins that regulates cell division, differentiation and cell death all of which are required for normal development and maintenance of tissues and organs (26). Mutations in proto-oncogenes convert them to oncogene which lacks these regulations leading to uncontrolled cell proliferation resulting inthe neoplastic phenotype. Conversion of a proto-oncogeneinto an oncogene generally involves gain-of-function mutations all of which confer growth and survival advantage to cells.These mutations include:

- Point mutations in proto-oncogene that result in a constitutively activated protein product
- Gene duplication or gene amplification leading to over-expression of the encoded protein.
- Chromosomal translocation that brings a growth-regulatory gene under the regulation of a different promoter causing altered gene expression (27).

Oncogenes were first identified in cancer-causing retroviruses from a series of studies by Peyton Rous in 1911. He excised fibro sarcoma from chicken, generated a filtrate and injected them into chicks. Majority of the chicks developed sarcoma and the transforming agent in the filtrate was shown to be a virus called *Rous Sarcoma Virus (RSV).* It was later shown that RSV is a retro-virus whose RNA was reverse-transcribed into DNA and incorporated into the host genome. It was also shown that along with the normal retroviral genes, the virus contained *v-src* gene which

was the cancer inducing gene. Further, Michael Bishop and Harold Varmus in 1977 showed that normal cells from chickens and other species contain a gene similar to the RSV v-*src* gene. This normal cellular gene was the proto-oncogene, distinguished from the viral counterpart by the prefix "c" (c-*src*) (27)

Based on the functional properties, oncogenes can be classified as(26):

- **Growth Factors**: Constitutive activation of growth factor genes can lead to malignant transformation,e.g. Platelet Derived Growth Factor (*PDGF)*, Fibroblast growth factors, Transforming growth factor α (*TGFα*).
- **Growth Factor Receptors**: Alterations in growth factor receptor gene leads to their constitutive activation in the absence of ligand binding, e.g. *EGF* receptor family, *PDGF* receptor.
- **Signal Transducers**: Binding of ligand to its receptor leads to auto-phosphorylation of the kinase domain of the receptor which in turn activates several downstream signaling pathway, e.g. *RAS, ABL, BRAF, ß-catenin.*
- **Nuclear regulatory proteins/Transcription factors**: They are members of multigene families which regulate expression of several genes that control cell division, e.g.*Myc, Fos, Jun.*
- **Cell cycle regulators**: Regulates cell proliferation, e.g. Cyclins, Cyclin dependent kinases (CDKs)
- **Apoptosis regulators:** The product of these genes have pro-apoptotic properties, e.g**.** c-Myc, c-Fos/c-Jun and bcl2.

1.5.2 Tumor Suppressor Genes

Activation of oncogenes through gain of function mutations represents only one of two distinct genetic alterations involved in cancer development; the other is inactivation of tumor suppressor genes (TSGs) through loss of function mutations. Unlike oncogene which drives abnormal cell proliferation, tumor suppressor genes unless mutated, restrain cell proliferation and stimulate cell death. In many cancers, these TSGs are either lost or inactivated, thereby removing negative regulators of cell proliferation and contributing to the abnormal proliferation of tumor cells(28). Proteins that are generally encoded by tumor suppressor genes include (24):

- Receptor proteins for secreted hormones that function in inhibiting cell proliferation
- Negative regulators of cell cycle progression
- Negative regulators of growth signaling pathways (e.g. APC or PTEN)
- Checkpoint-control proteins that arrest cell cycle upon DNA damage.
- Proteins that promote apoptosis
- DNA damage-repair enzymes.

Unlike oncogenes, the mutation spectrum of tumor suppressor genes is broad and includes truncating mutations such as nonsense or frameshift small insertion or deletions; splice site mutations; missense mutations or large genomic rearrangements. Since these loss of function mutations in tumor suppresser genes are recessive in nature, both the alleles of the gene must be mutated in order for a particular cell to become cancerous. This is known as the "two-hit" hypothesis, and it was first proposed by geneticist Alfred Knudson in 1971in order to understand the genetic mechanisms underlying retinoblastoma, a childhood form of retinal cancer. (29). The tumor suppressor genefound to be involved in retinoblastoma development was *Rb1* gene which

was also the first tumor suppressor gene to be identified. This discovery served as a prototype for identification of several other TSGs including *TP53*(28).

1.5.3 The burden of Hereditary Cancers

Mutations in oncogenes and tumor suppressor genes affect the overall growth and behavior of the mutant cell and also confer a selective growth advantage to thesecells. Although unlike oncogenes, both the alleles of the tumor suppressor genes must be inactivated for a physiologic effect to occur (23). When mutations in these genes occur at germline level, they result in hereditary predisposition to cancer. Although the burden of hereditary cancers is less and accounts for only 5% of all malignancies, it is important to identify such patients and family for risk assessment and clinical management (30). Germline mutations in oncogenes or tumor suppressor genes cause cancer predisposition and individuals with these mutations have a 'head start' on the neoplastic process as a mutation that can initiate the carcinogenic process isalready present in every cell of the body. Such individuals therefore often develop multiple primary tumors that occur at an earlier age than in individuals whose cancer-gene mutations have occurred at somatic level (23). Majority of hereditary cancer syndromes follow autosomal dominant pattern of inheritance in which the patient's first degree relatives have a 50% life-time risk of inheriting the proband's mutation (30). Each hereditary cancer syndrome has a certain characteristic spectrum of tumors with specific mode of inheritance. Table 1.1 summarizes a few common hereditary cancer syndromes with their predisposing genes, characteristic tumor spectrum and their mode of inheritance.

Table 1.1 : Mendelian Cancer Predisposition Syndromes (23)

Syndrome	Predisposing gene(s)	Mode of Inheritance	Tumor Spectrum
Hereditary Breast and Ovarian Cancer (HBOC)	BRCA1 & BRCA2	Dominant	Breast, Ovary, Prostate, Pancreas
Li-Fraumini Syndrome (LFS)	Tp53	Dominant	Sarcoma (mainly osteosarcoma), Breast, Brain, Blood, Adrenal
Lynch Syndrome (LS)	MLH1, MSH2, MSH6 & PMS2	Dominant	Colorectum, Endometrium, Stomach, small intestine, ovary, liver, pancreas, glioblastoma, breast, prostate
Familial Adenomatous Polyposis (FAP)	APC	Dominant	Colorectum, Upper GI, Medulloblastoma, Papillary Thyroid, Hepatoblastoma
Multiple Endocrine Neoplasia type 2 (MEN2)	RET	Dominant	Thyroid (Medullary), Parathyroid, Adrenal
Hereditary Retinoblastoma (RB)	RB1	Dominant	Eye (Retina)
Xeroderma Pigmentosa (XP)	XPA, C; ERCC2-5, DDB2	Recessive	Skin
Von Hippel-Lindau Syndrome (VHL)	VHL	Dominant	Kidney
Familial Paraganglioma	SDHB, C, D	Dominant	Paraganglioma, Pheochromocytoma
Peutz-Jeghers Syndrome (PJS)	STK11	Dominant	Intestine, Ovary, Pancreas
Cowden Syndrome	PTEN	Dominant	Hamartoma, Glioma, Uterus, Thyroid, Breast
Neurofibromatosis type I &II (NFI &II)	NF1 & NF2	Dominant	Neurofibroma, Meningioma
Hereditary leiomyomatosis	FH	Dominant	Leiomyomas, Renal Cancer, Uterus, Skin

1.6. Thyroid Cancer

Thyroid cancer originates from the cells of thyroid - a large, ductless, butterfly shapedendocrine gland located on the anterior portion of the neck in front of trachea(31). Thyroid gland consists of two encapsulated lateral lobes which are connected by a structure called isthmus (Figure 1.2A). Each lobe of thyroid is composed of thyroid follicles which are the structural and functional unit of the thyroid(32). The cavity of thyroid follicles is filled with colloidal fluid (Figure 1.2B) which is the site for synthesis of thyroid hormones-Triiodothyronine (T3) and Thyroxine/Tetraiodothyronine (T4) that regulates the rate of metabolism in the body(33). Thyroid follicles are surrounded by thin connective tissue stroma, which contains parafollicular cells (C cells). The parafollicular cells secrete a hormone called calcitonin which regulates calcium metabolism.

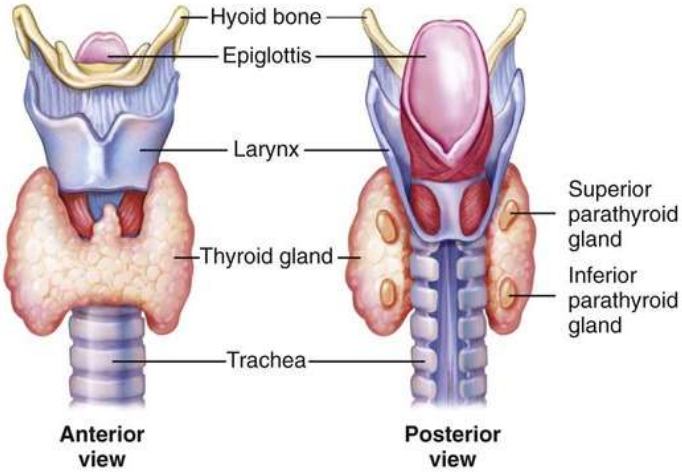


Figure 1.2A: Anatomy of Thyroid and Parathyroid Gland (Taken from Ref. No.31)

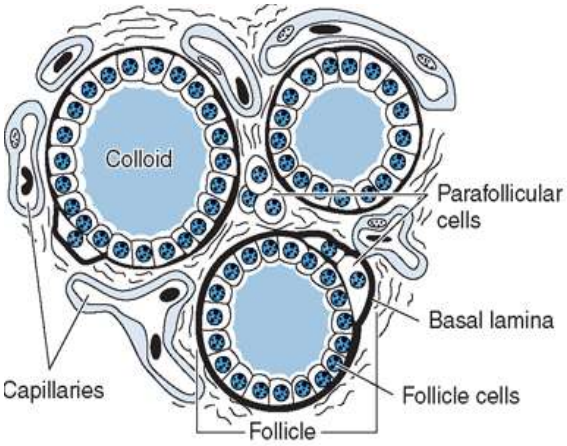


Figure 1.2B: Histology of Thyroid Gland (Taken from Ref No - 32)

On the posterior side of the thyroid gland are embedded two pair of parathyroid glands, one pair on each lobe of thyroid. (Figure 1.2A). The parathyroid gland secrete parathyroid hormone (PTH) called parathormone whose function is to regulate blood calcium levels.

1.6.1 Types of Thyroid Cancer

There are four types of thyroid cancer based on their cells of origin and aggressiveness (Figure 1.3):

- **Papillary Thyroid Cancer**: This is the most common type of thyroid cancer originating from follicular cells of the thyroid. It accounts for 70-80% of all thyroid cancers (Figure 1.3). Papillary Thyroid cancer tends to grow slow and has excellent prognosis.
- **Follicular Thyroid Cancer:** Follicular thyroid cancer also arises from the follicular cells of thyroid and makes up to 10-12% of all thyroid cancers. Follicular cancer can spread to lymph nodes and is also likely to metastasize to distant organs.
- **Medullary Thyroid Cancer**: Medullary Thyroid cancer is an aggressive subtype of thyroid cancer that originates from parafollicular cells (C-cells) of thyroid. The C-cells secretes a hormone called calcitonin which is a biochemical marker of MTC diagnosis and its clinical follow-up. Medullary Thyroid Cancer makes up to 5% of all thyroid cancer (Figure 1.3) and is associated with a poor prognosis if metastasized.
- **Anaplastic Thyroid Cancer:** This is the rarest and the most aggressive subtype of thyroid cancer that accounts for 2% of all thyroid cancers (Figure 1.3).Anaplastic thyroid cancer occurs in adults of 60 years or above and are least likely to respond to treatments.

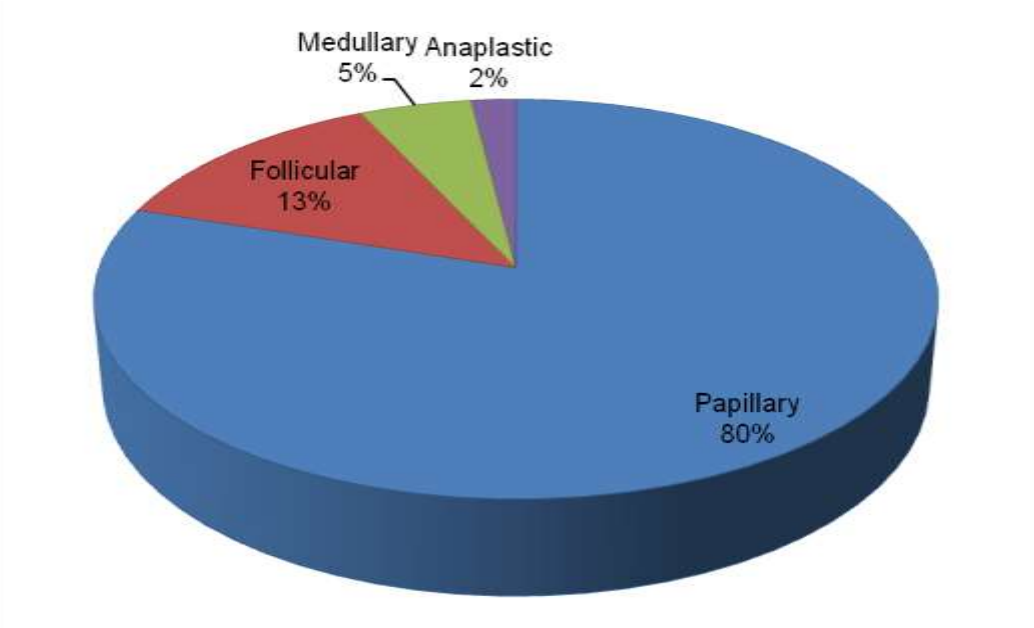


Figure 1.3: Types of Thyroid Cancer & their incidence

1.6.2 Epidemiology of Thyroid Cancer

Thyroid cancer is relatively uncommon compared to other cancers. In the United States, by 2016 approximately 64,000 new patients were diagnosed with thyroid cancer, compared to 240,000 patients with breast cancer and 135,000 patients with colon cancer(34). In a study conducted by Surveillance, Epidemiology, and End Results 9 (SEER 9) of the National Cancer Institute reported in *JAMA*, trends in thyroid cancer incidence from 1974 to 2013 in the United States were analyzed and reported(35). Among the total of 77,276 patients registered, mean age of cancer diagnosis was 48 years and 75% were women. The most common type of thyroid cancer by histology was papillary thyroid cancer (84%) followed by follicular thyroid cancer (11%). The incidence of thyroid cancer increased from 4.56/100,000 person-year in 1974-1977 to

14.42/100,000 person-years in 2010-2013, representing an average annual increase of 3.6% (95% CI: 3.2%–3.9%, P < .001)(35).

In India,the epidemiological trends of thyroid cancer was studied through the Population Based Cancer Registries (PBCRs) compiled by the National Cancer Registry Program (NCRP), Government of India. Data from 14 regions of India which reported data from 2004/05 to 2013/14 was analyzed. They reported that over a decade, the incidence rate of thyroid cancer in Indian women increased from 2.4% (95% CI: 2.2-2.7) to 3.9% (95%CI: 3.6-4.2) and in men from 0.9% (95%CI 0.8-1.1) to 1.3% (95%CI 1.2-1.5), (36). In conclusion, the incidence of thyroid cancers in India is rapidly increasing particularly among the younger population.

1.7 Medullary Thyroid Carcinoma

Medullary Thyroid Carcinoma (MTC) was first described by Jaquet (1906) in the German literature as ‘‘malignant goiter with amyloid’’ (37). Later in 1959, Hazard et al., provided a definitive histological description of MTC as a solid tumor of nonpapillary configuration yet with a high potential of metastasizing to the lymph nodes(38). Williams further suggested that MTC originated from the calcitonin-secreting parafollicular C cells of thyroid (39). Since then, several group of investigators demonstrated MTC to be derived from the thyroid parafollicular cells which under neoplastic condition produces excessive calcitonin which would help with the biochemical diagnosis of MTC and follow-up of the patients(40,41).

MTC is unresponsive to conventional chemotherapy & radiotherapy if metastasized. Once MTC metastasizes, it has a tendency to spread to local and regional lymph nodes, andmore distantly to lung, liver, and bone(42). Therefore the current standard treatment is surgical resection of thyroid at an early stage when MTC is confined to thyroid itself. The overall survival rate of MTC

depends on the stage and aggressiveness of the disease. The 10 year survival rate of patients with MTC restricted to thyroid is around 95%, decreasing to less than 40% for metastatic MTC(43,44). Although majority of MTCs are sporadic, that is, without any family history or syndromic features, a hereditary form of the disease also exists(44).

1.7.1 Hereditary MTC/MEN2 Syndrome

Hereditary MTC accounts for 25% of MTC cases and occurs as a main feature of an autosomal dominant inherited cancer syndrome called Multiple Endocrine Neoplasia type 2 (MEN2) which implies that the offspring is at 50% risk to inherit the disorder(45) . MEN2 syndrome, first identified by John Sipple in 1969(46), has three distinct clinical subtypes, MEN2A, MEN2B & Familial MTC (Figure 1.4 & Table 1.2) with MTC being the common clinical manifestation. Affected individuals initially develop C-cell hyperplasia (CCH) that later progresses to invasive medullary thyroid carcinoma (47). The three subtypes of MEN 2 differ in terms of age of disease onset and its aggressiveness, tissue affected, and prognosis(45).

- **MEN2A:** MEN2A represents 55% MEN2 cases (48) and is characterized by the presence of bilateral & multifocal MTC, adrenal Pheochromocytoma and/or hyperparathyroidism. The lifetime risk for MTC is over 90% in patients with MEN 2A, while that of pheochromocytoma and parathyroid hyperplasia are 40–50% and 10–20% respectively(47). MTC is generally the first manifestation in MEN 2Apatients and presents between the age ranging from 5years to 25 years(47).Two rare variants of MEN 2A have been described, one in which MTC is accompanied with cutaneous lichen amyloidosis(pigmented lesion of skin on upper back) (49) and the other in

which MTC occurs with Hirschsprung's disease which is marked by congenital absence of enteric ganglia(50).

- **MEN 2B:** MEN2B is the rarest but the most aggressive subtype and accounts for 5% of MEN2 cases (47). It is characterized by early onset aggressive MTC with syndrome associated phenotypic features such as mucosal neuromas, blubbery lips, intestinal ganglioneuromatosis & skeletal abnormalities (Marfanoid habitus). Pheochromocytoma is associated with MTC in about 50% MEN2B cases(45).

- **FMTC:** The clinical diagnosis of Familial MTC is made when four or more family members across a wide range of ages have isolated MTC and no family history of pheochromocytoma or hyperparathyroidism (51). FMTC is the mildest form of MEN2 and accounts for 35-40% of all MEN2 cases (48,52). Recently, FMTC is classified as a phenotypic variant of MEN 2A with decreased penetrance of pheochromocytoma and primary hyperparathyroidism(47,53).

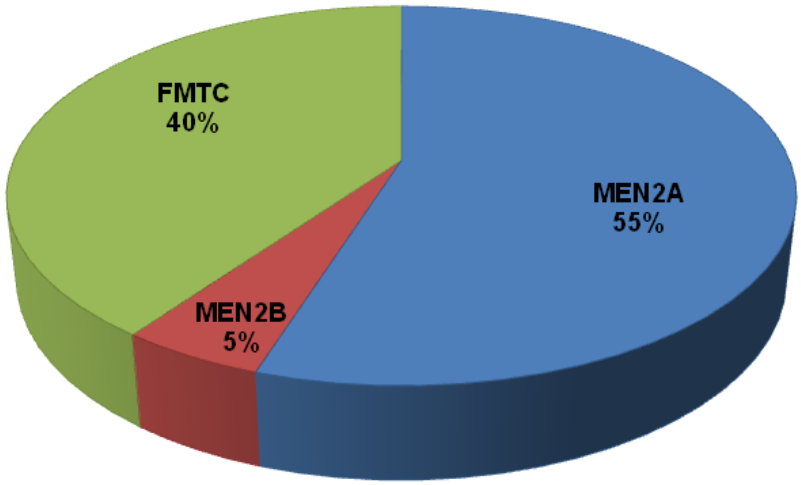


Figure 1.4: MEN2 subtypes & their incidence rate

Table 1.2: Phenotypic features common to more than one subtype and others specific to a particular subtype of MEN2 syndrome.

	MTC	Pheochromocytoma	HPT	CLA	Age at presentation	Mucosal neuromas
MEN2A	+	+	+	*	<20yrs	-
MEN2B	+	+	-	-	<10yrs	*
FMTC	+	-	-	-	20-50yrs	-

HPT: Hyperparathyroidism; CLA: Cutaneous Lichen Amyloidosis.

+ Features seen in more than one subtype, * Features specific to one subtype

1.7.2 Sporadic MTC

Sporadic MTC is the most common form of MTC and accounts for 75% of all MTC cases (45,53). It is characterized by presence of unilateral MTC without any family history or syndromic features associated with MEN2 syndrome. Most patients suspected with sporadic MTC present with a palpable thyroid nodule and the diagnosis is made by fine needle aspiration of the thyroid nodule confirmed by immune staining for calcitonin(54). Sporadic MTC presents with a slow growth which spreads quickly to regional cervical lymph nodes. At the time of diagnosis, up to 70% of cases may present with local cervical metastasis. Distant metastasis may occur mainly in the liver, lung, bone and brain(55).

Sporadic MTC differs from its hereditary form (MEN2) in several ways: (i) It is associated with later disease onset; (ii) Usually less aggressive compared to MEN2 associated MTC; (iii) The thyroid nodule is unilateral and unifocal; (iv) No other MEN2-associated Neoplasia exists; (v) Family members not affected with the disease(55).

1.8 RET Proto-Oncogene

The predisposing gene for Medullary Thyroid Carcinoma is RET, mutations in which is found in virtually all MEN2 syndrome cases and 40-60% of sporadic MTC cases (53). RET is a proto-oncogene located on chromosome 10 (10q11.2) and has 21 exons. The gene encodes a receptor tyrosine kinase expressed in neural crest derived cell lineages such as noradrenergic and dopamanergic neurons and neuroendocrine tissues including thyroid C-cells and adrenal chromaffin cells(44,56).

RET (REarranged during Transfection) was first identified by Takahashi et al., in 1985 as a proto-oncogene which was able to undergo oncogenic activation by genetic rearrangement (57). In their experiment, they transfected NIH3T3 mouse embryonic fibroblast cells with human lymphoma DNA and observed that DNA of the transformed NIH3T3 cells induced transformation with high efficiency. The transformedNIH3T3 cells was shown to harbor a fusion gene, whichwas absent in the original tumor(57). This fusion gene contained part of a gene that encoded a tyrosine kinase domain, andthat gene of which the tyrosine kinase domain was a partwas thereafter called "*RE*arranged during *T*ransfection or RET" gene(56,57).

RET is essential for early development of the sympathetic, parasympathetic, and enteric nervous systems, the kidney, and spermatogenesis(58). Inactivating germline *RET* mutations leads to the development of Hirschsprung disease(HSCR), a congenital absence of enteric neurons in the gastrointestinal tract. On the other hand, activating germline *RET* mutations can cause Multiple Endocrine Neoplasia type 2 (MEN2)and genetic rearrangements leads to papillary thyroid cancer(56).

1.8.1 RET structure

RET is a single pass trans-membrane receptor tyrosine kinase protein (Figure 1.5) with three functional domains-an extracellular ligand binding domain with four cadherin-like repeats and a cysteine rich region, a transmembrane and an intracellular tyrosine kinase domain (44). The Ca^{+2}dependent cadherin like repeats in the extracellular domain induces and stabilizes the conformational changes required for interaction of the receptor with the ligand and the co-receptor. The cysteine rich region is involved in receptor dimerization upon ligand binding by forming inter-molecular disulfide between two cysteine residues of the adjacent RET monomers. The intracellular tyrosine kinase domain is involved in intracellular signaling upon activation of RET(56). The extracellular domain also contains a number of glycosylation sites. The fully glycosylated RET is 170KDa (mature protein) and is expressed on cell membrane while the immature 150KDa protein lacks glycosylation and is present only in the endoplasmic reticulum and the cytoplasm.

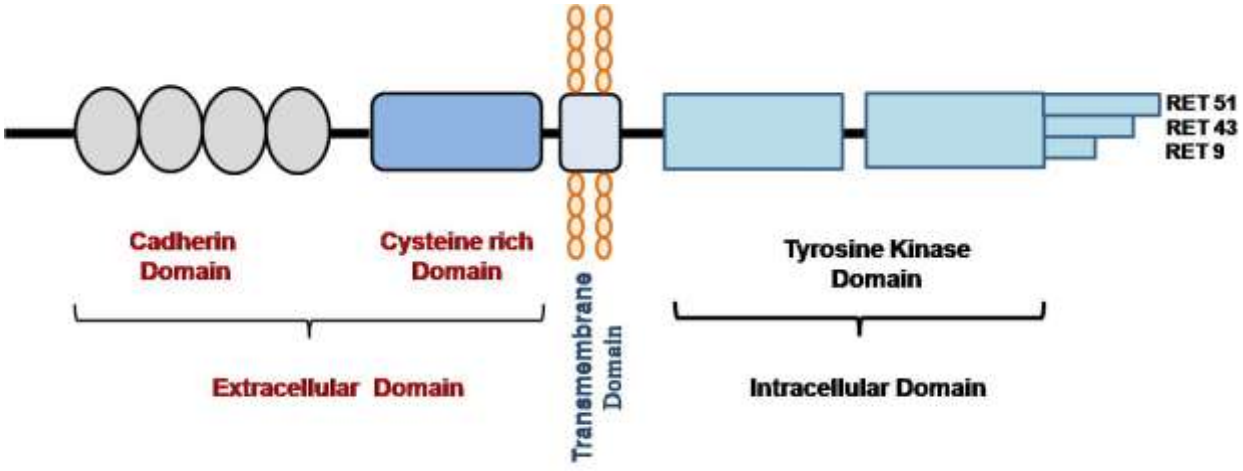


Figure 1.5: Structure of RET receptor tyrosine kinase.

RET undergoes alternative splicing at the carboxy-terminal end generating three protein isoforms that differ in 9 (RET9), 43 (RET43) and 51 (RET51) amino acids in the carboxy-terminal tail downstream from glycine 1063 (Figure 1.5). RET9 and RET51 are the functional isoforms in *vivo.*

1.8.2 RET activation and signaling

Activation of RET is mediated by complex of ligands and the co-receptors (Figure 1.6). The ligands belong to Glial cell line-derived neurotrophic factor (GDNF) family of ligands (GFLs) including neurturin, artemin, and persephin and the co-receptors belong to GDNF family of α receptors (GFRαs)(56). On the cell surface, upon external stimulus, the ligand binds to the co-receptor forming ligand-co-receptor complex. This ligand-co-receptor complex brings about dimerization of two RET monomers leading to the auto-phosphorylation of the tyrosine residues in the tyrosine kinase domain and activates RET.

Activated RET serves as a docking site for a number of adaptor molecules that links the activated RET with downstream signaling pathways. RET signals through multiple downstream pathways including RAS/MEK/ERK pathway that promotes cell cycle progression, PI3K/AKT/NF-κB pathway that leads to increased cell motility, survival, and progression through the cell cycle, JAK-STAT pathway and the MAP kinase pathway (44).

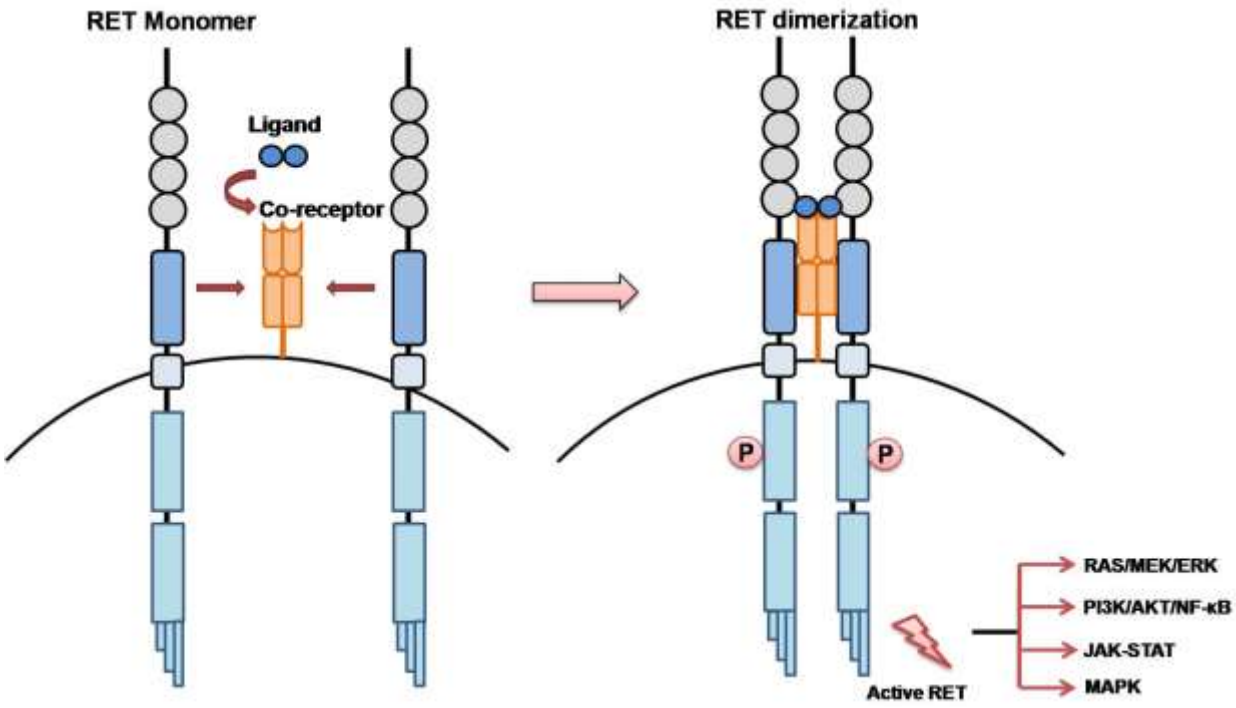


Figure 1.6: RET activation and Signaling

1.8.3 RET & MEN2 Syndrome: Genotype Phenotype correlation

Unlike most other cancer syndromes, which are associated within activation of tumor suppressor genes, MEN 2 arises as a result of activating mutations of the RET proto-oncogene (59).Germline mutations in RET proto-oncogene has been identified in 95-98% MEN2 cases(45,53,60,61). Also, 5-8% apparent sporadic MTC cases harbor germline mutations in RET gene(62).Similarly somatic mutations in RET is found in 40-60% sporadic MTC cases and the most common mutation being the M918T mutation of the tyrosine kinase domain of the gene (63) Mutations within RET are clustered in specific regions of the gene called as "Mutation Hotspots" and include exon 10 and 11 of extracellular cysteine rich region and exon 13-16 of the intracellular tyrosine kinase domain (59).

Clinical implication of RET proto oncogene in MEN 2 syndrome was first identified in 1993 (61,64)and it was reported that majority of MEN2A and Familial Medullary Thyroid Carcinoma (FMTC) results from missense mutations at codon 609, 611, 618, 620 and 634 in the cysteine rich region of the extracellular domain of RET. Later, mutations at codon 768 and codon 804 within the intra- cellular tyrosine kinase domain of RET have only been identified in FMTC patients (65).Further Eng et al., in 1994 reported point mutation in the tyrosine kinase catalytic domain of RET resulting in substitution of Threonine for Methionine at codon 918 to be associated with MEN2B phenotype (66).

The largest cohort study to establish the relationship between specific RET mutations and MEN2 phenotype was performed in 1996 by Eng et al., in International RET mutation Consortium analysis (60). This was a correlatives urvey study of 477 MEN 2 families from eighteen tertiary referral centers worldwide with an aim to establish an association between specific RET mutations and phenotypic features associated with MEN2 which could help in clinical management of the patients and their family in terms of screening, surveillance and prophylaxis. They identified that 95% MEN2B families had M918T mutation in the intracellular tyrosine kinase domain. The remaining 5% cases have mutation at either codon 883 or by rare occurrence of double *RET* mutation in codons such as 804, 904 or 768,which alone have a low penetrance for MTC(45,56). The most commonly associated mutation with MEN2A phenotype was at codon 634, specifically C634R mutation where MTC was accompanied with pheochromocytoma and/or hyperparathyroidism. FMTC was associated with mutations at codon 609, 611, 618, 620 and 634 of the extracellular cysteine rich domain and at codon 768 and 804 of the tyrosine kinase domain (60). Following this, several studies in different cohorts supported these findings and

RET was established as the driver or predisposing gene for MEN2 syndrome and genetic testing of RET became an integral part of clinical management of MTC patients.

American Thyroid Association (ATA) guidelines task force further stratified different RET codon mutations into different risk categories A-D based on the aggressiveness of the disease and the phenotype they produce (45). Risk category A mutations presents the least aggressive phenotype of FMTC in which MTC is the only clinical manifestation whereas risk category D mutations presented the most aggressive phenotype of MEN2B (Figure 1.7). Risk category C mutations were associated with MEN2A phenotype where MTC is accompanied with pheochromocytoma or HPT whereas risk category B mutations give phenotype of either FMTC or MEN2A (Figure 1.7).

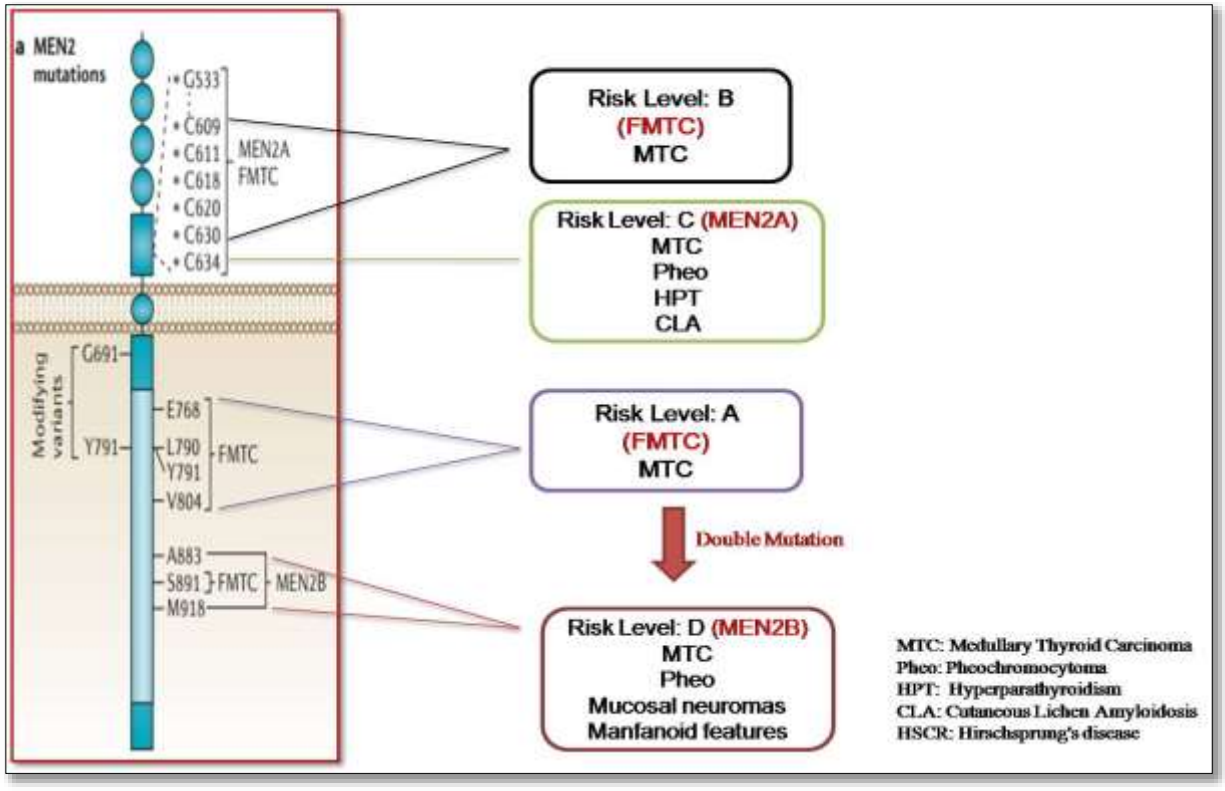


Figure 1.7: Risk stratification of different RET mutations by American Thyroid Association guidelines (45).

Further, in their revised guidelines proposed in 2015 (53), they modified the risk categories for hereditary MTC into 'Highest', 'High' and 'Moderate' based on the disease aggressiveness. The current risk level D category was changed to "Highest risk" and included patients with M918T mutations that gives phenotype of MEN2B. Further, risk level C category was changed to 'High risk" category and included MEN2A patients with codon 634 mutations. The risk category A & B was combined and made a new category "Moderate risk" which included patients with FMTC phenotype and all other RET mutations except M918T C634 (53).

American Thyroid association also proposed guidelines (Table 1.3) for prophylactic thyroidectomy as part of clinical management of hereditary MTC (45). As per ATA guidelines, individuals with a family history consistent with FMTC or MEN2 are at risk of autosomal dominant inheritance of the syndrome and should be offered RET genetic testing. For MEN2B this should be done shortly after birth whereas for MEN2A and FMTC it should be done by the age of 5 years or may be postponed beyond 5 years under strict surveillance for annual basal and stimulated serum calcitonin which is a biochemical marker of MTC diagnosis and its clinical follow-up (Table 1.3).

Table 1.3: ATA guidelines for MTC management and prophylaxis.

ATA risk level/Mutations	Age of genetic testing	Age of prophylactic surgery
D(Highest) (M918T/A883F/V804M+E805K, V804M+Y806C, V804M+S904C)	ASAP and within 1 year of life	ASAP and within 1 year of life
C (High) (C634R/G/F/W/Y)	<3-5 years	Before age of 5 years
B(Moderate) (C609, C611, C618, C620, C630)	<3-5 years	Consider Before age of 5 years. May delay surgery beyond age 5 years if stringent criteria are met such as normal annual basal and stimulated serum calcitonin
A(Low) (E768D, V804M, V804L, K666E, S891A	<3-5 years	May delay beyond 5years. May delay surgery beyond age 5 years if stringent criteria are met such as normal annual basal and stimulated serum calcitonin

1.9 Rationale of the study

The Genotype-Phenotype correlation between RET mutations and MEN2 phenotypeis important for the clinical management of MTC patients and advising screening or preventive surgery for mutation carriers. In the Caucasian population a strong genotype-phenotype correlationhas been established between location and type of RET mutation and MEN2 phenotype but this has not been studied in the Indian population.Even for specific and common mutations such as C634R, there is significant variability in terms of age of onset and aggressiveness within and between families(67). Also, there are a few individuals with classical MEN2B phenotype but without MEN2B specific M918T or A883F mutation identified. Such cases therefore need to be studied in order to identify other genetic loci that could produce MEN2B phenotype.

1.10 Hypothesis

The Genotype-phenotype correlation could be different between populations because of the difference in their geo-ethnic background and needs to be established for clinical management of patients and the family of that population. Also, while the presence and location of a germline mutation in RET is the major determinant of the MEN2 phenotype, the phenotypic differences within and between families with a specific RET mutation can be due to germline or somatic genetic alterations in other genes.

1.11 Aim of the study

To establish the Genotype-Phenotype correlation in the Indian MTC cohort and to identify additional germline and somatic genetic variants and pathways associated with MTC and its phenotype (age at onset, aggressiveness) in a cohort of MTC patients with or without a germline RET mutation.

1.12 Objectives

1. To identify novel and recurrent germline RET mutations, their penetrance and Genotype-Phenotype correlation in a cohort of Indian MTC cases.
2. To identify additional molecular signatures in RET driven MTC tumor samples using Whole Exome Sequencing.
3. To study the modulatory role of SNPs of distinct genetic pathways on risk of MTC development and its clinical behavior.
4. To identify driver mutation that could produce MEN2B phenotype in the absence of germline RET gene mutations.

Chapter 2

Materials and Methods

2.1 Patients

Patients diagnosed with Medullary Thyroid Carcinoma (MTC) either by fine needle aspiration biopsy (FNAC), histopathological examination of surgically resected tumor tissues or by investigation for biochemical markers for MTC such as serum calcitonin levels at diagnosis were enrolled in the study as part of Institutional ethics committee approved study. The patients were enrolled directly or referred from Head & Neck, Disease Management Group, Tata Memorial Hospital (TMH) between 2006 to 2019 at our Cancer Genetics Clinic, TMH. Blood sample was collected fromthe patients with written informed consent for biobanking and germline *RET*genetic testing. For minors, the written informed consent was provided by the parents. Personal and family history with clinico-pathological details was recorded using a standard case record form. All experiments were carried out in accordance with the approved guidelines and regulations.

2.2. Overview of patient recruitment process

Pre-test counseling session: Before patient recruitment or blood sample collection for *RET* genetic testing, a pre-test counseling session was conducted between the patients, their accompanying family members and the genetic counselor at Cancer Genetics Clinic. During this session, the genetic counselors explained the patient and their family members about the hereditary aspect of the disease and how it is inherited. They also explained them the advantages of performing the genetic testing, its expected outcomes and possible implications. Further, a pedigree chart was drawn based on the detailed family history and all the clinic-pathological details were recorded using a standard case-record form. Following this, 3-6 ml of blood was collected in EDTA vacutainers after taking written informed consent.

Post-test counseling session

In the post-test counseling session the genetic test results were communicated to the patients in the form of a printed genetic test report. During this session, the patients were explained about the test results and if they were found to be mutation carriers, appropriate surveillance and prophylactic measures were advised. Also, screening of the family specific mutation in first and second degree relatives of mutation carriers were also offered during this session.

2.3. Study Schema

- For Genotype-Phenotype correlation study, 245 Indian MTC cases enrolled between 2014-2019 were included. Germline mutation analysis was performed for the six hotspot exons of RET – exon 10, 11, 13, 14, 15 and 16 by PCR and Sanger Sequencing. For this, blood sample was collected from the patients and DNA was extracted followed by exon specific PCR and Sanger Sequencing.
- For MTC cases with classical MEN2 phenotype and negative for the hotspot RET mutations (3 cases), full RET gene sequencing was performed and if no mutation was identified, those cases were taken up for whole exome sequencing. Exome Library was prepared by Nextera rapid capture exome kit and data was analyzed using GATK pipeline.
- Further, whole exome sequencing was performed on 15 tumor and paired blood samples of RET driven MTC cases to identify additional genetic alterations which might be responsible for the clinical heterogeneity observed in patients harboring same RET mutations. For this DNA was extracted by PAXgene Tissue DNA kit from tumor tissues

and by Qiagen QIAmp blood DNA mini kit from blood samples followed by their exome sequencing and data analysis.

- Also, the modifier role of 13 SNPs of genes of 3 distinct genetic pathways were studied in a cohort of 438 MTC cases enrolled between 2006-2018 at CGC clinic, TMH using RFLP/TaqMan SNP genotyping approach followed by statistical analysis.

The methodology employed for the completion of each of these objectives is described in detail below:

2.4 Genomic DNA Extraction

2.4.1 DNA extraction from Blood (Phenol-Chloroform Method)

For genomic DNA extraction from blood samples, conventional Phenol-Chloroform method was used initially which was later replaced with Qiagen QIAmp Blood DNA extraction kit which is a rapid method of DNA extraction and also provides a better quality DNA compared to the conventional method.

For genomic DNA extraction by phenol-chloroform method, fresh blood sample was collected in 3 ml EDTA vacutainer and kept at room temperature until the plasma was separated.Lymphocytes were separated from the 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 is as follows:

RBC Lysis Buffer:

1. Ammonium Chloride - 82.9g

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. Working 1X concentration can be made by diluting it with distilled water.

Phosphate Buffer Saline (PBS):

1. NaCL : 8 g

2. KCL : 0.2g

3. Na2 HPO4 : 1.44g

4. KH2PO4: 0.24g

Dissolved in 800ml of distilled water and pH was adjusted to 7.4. The volume was make up to 1 litre with distilled water, autoclaved at 15 lb for 15 mins and stored at room temperature.

Steps involved:

1. The fresh blood collected in the EDTA vacutainer was transferred to 15ml falcon tube and three volumes of RBC lysis buffer was added and mixed by rotating for 10 mins.

2. Tubes were centrifuged at 1500rpm for 15 mins and the supernatant was discarded leaving about 1ml behind to prevent loss of lymphocyte cells.

3. Step 1 and 2 was repeated until a clear supernatant and a clean white pellet is obtained (usually obtained in 3 washes).

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

5. The solution was transferred to two 1.5 ml fresh eppendorf, tubes distributing half the volume in each tube followed by centrifugation of the tubes at 12,000rpm for 10 mins.

6. The supernatant was discarded and the pellets were re-suspended in 500µl of fresh PBS and the eppendorf tubes were centrifuged at 12000 rpm for 10 mins and the supernatant was discarded completely. This pellet obtained at this step was used for DNA extraction.

7. To the pellet containing 1.5 ml eppendorf tube, cell lysis buffer (400ul) was added and mixed gently by rotating. The tubes were incubated at 56° C in a water bath for an hour.

8. In the next step 10 µl of Proteinase K was mixed (10mg/ml stock) gently directly onto the pellet and mixed gently by pipetting up and down 8-10 times (do not vortex).

9. The tubes were incubated overnight at 56 °C for proper lysis.

10. The next day, 500 µl of phenol (equilibrated with TRIS to get a pH above 7) was added to the tube, mixed well by inverting/rotating ten mins and centrifuged at 12000 rpm for 10 mins to separate the aqueous and organic phase.

11. The aqueous upper layer was transferred to a fresh 1.5 ml eppendorf tube containing 250µl phenol and 250µl chloroform: Iso-amyl alcohol (24:1), mixed well by inverting/rotating, for 10 mins and then centrifuge for 10 mins at 12,000 rpm.

12. The aqueous upper layer was again transferred to a fresh 1.5 ml eppendorf tube containing 500µl of chloroform: iso-amyl alcohol (24:1), mixed well by inverting, for 10 mins and then centrifuged at 12,000 rpm for 10 mins.

13. The aqueous phase was transferred to a fresh 1.5 ml eppendorf tube and 1ml of absolute ethanol (96-100%) was added and inverted gently a few times. The DNA tends to be seen as thread like structures.

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

15. In the next step, the tube was brought at room temperature and centrifuged at 12,000rpm for 20 mins. The supernatant was discarded and two washes with 70% alcohol was given (1 ml of 70% ethanol was added and centrifuged at 12,000rpm for 10 mins). The pellet was the air dried to get rid of ethanol completely.

16. To the pellet containing tube, 200 µl of TE buffer was added and incubated at 56° C overnight to dissolve the DNA completely which can be then stored at -20° C.

2.4.2 DNA extraction from Blood (Qiagen Column Method)

Genomic DNA was extracted from 200µl of fresh blood sample using the QIAmp Blood DNA Mini kit (Cat#51304) as per the manufacturer's protocol.

Steps involved:

1. In a 1.5ml micro-centrifuge tube 20 µl proteinase K (QIAGEN) was added to the bottom of the tube followed by addition of 200 µl of blood sample and 200 µl of buffer AL. This was mixed gently by pulse-vortexing for 15 sec or pipetting the entire solution up and down 6-8 times.

2. The tube was incubated at 56°C for 30 mins followed by brief centrifugation to settle down the droplets formed inside the lid of the microcentrifuge tube.

3. In the next step, 200 µl absolute ethanol (96-100%)was added to the sample tube and mixed by pulse-vortexing or pipetting.

4. The micro-centrifuge tube was then briefly centrifuged and the entire solution was transferred to the Qiagen QIAmp mini spin column provided in a 2 ml collection tube without wetting the rim. The column containing 2ml micro-centrifuge tube was then centrifuged at 8000 rpm (6000 x g) for 1 min.

5. The QIAmp Mini spin column was then transferred in a fresh 2 ml collection tube (provided with kit), and the previous tube containing the supernatant was discarded.

6. In the next step, 500 µl Buffer AW1 was added in the column placed in 2 ml collection tube without wetting the rim. The column containing tube was then centrifuged at 8000 rpm (6000 x g) for 1 min.

7. The column was then transferred in another fresh 2 ml collection tube and 500 µl Buffer AW2 was added without wetting the rim and centrifuged at full speed (20,000 x g; 14,000 rpm) for 3 min.

8. The QIAamp Mini spin column was again transferred into a fresh 2 ml collection tube and centrifuged at full speed for 1 min (dry spin).

9. The QIAamp Mini spin column was then placed in a fresh 1.5 ml micro-centrifuge tube and 100-200 µl Buffer AE (elution buffer) was added. The tube was incubated at room temperature for 1-5 min and then centrifuged at8000 rpm (6000 x g) for 1 min to elute the DNA from the column.

10. The eluted DNA was then quantified using Nanodrop spectrophotometer and/or by loading the DNA on 0.8% agarose gel (semi-quantitative approach).

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

2.4.3 DNA Extraction from Fresh Frozen Tumor tissues (PAXgene Tissue DNA kit)

DNA was extracted from 20mg MTC tumor tissue using PAXgene Tissue DNA kit (PreAnalytiX, Cat#767134) as per the manufacturer's protocol.

Steps Involved:

1**.** The tumor tissue (20mg) was minced properly in 50-100µl PBS on a petri-plate with a scalpel. The entire minced tissue was then pipetted/transferred into a 1.5ml microcentrifuge tube and brief spin is given.

2. In the next step, 180µl TD1 buffer (lysis buffer) and 20 µl proteinaseK was added to the tissue containing micro-centrifuge tube and mixed well by pipetting. The micro-centrifuge tube was then incubated at 56°C waterbath overnight.

3. The next day, the micro-centrifuge tube was centrifuged briefly to settle down the droplets formed inside the lid of the tube. If the tissue was not lysed completely upon overnight incubation, 10µl proteinaseK was again added and incubated at 56°C for another 1 hour for complete lysis of the tissue.

4. The microcentrifuge tube was again centrifuged briefly and 200 µl TD2 buffer was added and mixed by pulse vortexing for 15 s. The tube was incubated at 80°C waterbath for 30 mins.

5. In the next step, 200µl absolute ethanol (96-100%) was added to the sample and mixed by vortexing (Note: a white precipitate may form upon addition of ethanol which does not interfere with the downstream steps of DNA extraction). The sample tube was the centrifuged briefly to remove the droplets from the inside of the lid.

6. The entire sample was transferred into the PAXgene DNA spin column placed in a 2 ml collection tube, and centrifuged for 1 min at 6000 x g. The spin column was then placed in a new 2 ml collection tube and the old collection tube containing the flow-through was discarded.

7. Next, 500 µl TD3 buffer was added into the PAXgene DNA spin column and centrifuged for 1 min at 6000 x g.

8. The spin column was again placed into a new 2ml collection tube discarding the previous one containing the flow-through and 500µl TD4 buffer was added followed by centrifugation at 6000 x g for 1 min.

9. The spin column was transferred into a new 2ml collection tube and centrifuged at maximum speed (20,000 x g) for 3 min to dry the membrane completely.

10. The collection tube containing any flow-through was discarded and the spin column was placed in a 1.5 ml micro-centrifuge tube. 100-200 µl buffer TD5 (elution buffer) was added directly onto the PAXgene DNA spin column membrane and incubated at room temperature for 5 min followed by centrifugation for 1 min at maximum speed to elute the DNA.

11. After estimating the purity and concentration, the DNA samples was stored at -20°C till further use.

2.4.4 DNA Extraction from FFPE samples (Qiagen DNA FFPE Tissue Kit)

For genomic DNA extraction from formalin-fixed, paraffin-embedded tissues, Qiagen QIAamp DNA FFPE Tissue Kit (Cat#56404) was used and DNA was extracted as per manufacturer's protocol.

Steps Involved:

1. Using a scalpel, excess of paraffin was trimmed off the sample block and 7-8 sections, 5-10 µm thick were cut and immediately placed in a 1.5 ml microcentrifuge tube. To the sample tube, 1 ml xylene was added, lid was closed and the microcentrifuge tube was vortexed vigorously for 10 s.

2. The sample tube was then centrifuged at full speed for 2 min at room temperature and the supernatant was removed by pipetting without disturbing the pellet.

3. To the pellet, 1 ml ethanol (96–100%) was added and mixed by vortexing (the ethanol extracts residual xylene from the sample).

4. The tube was centrifuged at full speed for 2 min and the supernatant was removed by pipetting without disturbing the pellet. All the residual ethanol was further removed using a fine pipette tip. The tube was kept open at room temperature until all ethanol gets evaporated.

5. The pellet was resuspended in 180 µl Buffer ATL (lysis buffer) and 20 µl proteinase K. The microcentrifuge tube containing the resuspended pellet was vortexed and centrifuged briefly and incubated at 56°C for 1 h (or until the sample has been completely lysed).

6. The 1.5 ml microcentrifuge tube was centrifuged briefly to remove drops from the inside of the lid and 200 µl AL buffer was added and mix thoroughly by vortexing.

7. Then 200 µl ethanol (96–100%) was added and mixed by vortexing (The buffer AL and ethanol are mixed immediately and thoroughly by vortexing or pipetting to yield a homogeneous solution). Buffer AL and ethanol can be premixed and added together in one step to save time when processing multiple samples.

8. The entire lysate was then transferred to the QIAamp MinElute column placed in a 2 ml collection tube without wetting the rim and and centrifuged at 6000 x g (8000 rpm) for 1 min.

9. The QIAamp MinElute column was then transferred to a fresh 2 ml collection tube, and the previous collection tube containing the flow-through was discarded.

10. To the column 500 µl Buffer AW1 was added without wetting the rim and centrifuged at 6000 x g (8000 rpm) for 1 min. The QIAamp MinElute column was again transferred in a new 2 ml collection tube discarding the previous collection tube containing the flow-through.

11. In the next step, 500 µl Buffer AW2 was added to the column without wetting the rim and centrifuged at 6000 x g (8000 rpm) for 1 min. The column was further transferred to a new collection tube and centrifuged at full speed (20,000 x g; 14,000 rpm) for 3 min to dry the membrane completely.

12. In the final step, the QIAamp Mini Elute column was placed in a clean 1.5 ml microcentrifuge tube and 20–50 µl Buffer ATE (elution buffer) was added to the center of the membrane.

13. After estimating the purity and concentration, the DNA samples was stored at -20°C till further use.

2.5 Polymerase Chain Reaction

Germline *RET* mutation analysis was performed initially for the 6 hotspot exons of *RET* (exon 10, 11, 13, 14, 15 16) by PCR and Sanger Sequencing using specific primers designed for respective exons. In patients with classical syndrome associated features for MEN2 and are found to be negative for hotspot RET mutations, full gene sequencing of RET was performed (20 exons). The primer sequences and annealing temperatures of all exons of RET gene designed as part of the thesis project are given in Table 2.1. For PCR, 5µl (20ng/µl) gDNA was amplified in a 25µl PCR reaction volume containing 0.5µl of each Forward and Reverse primers (10pmol), 1µl dNTPs (2.5 mmol), 0.5µl Taq Polymerase (2U/µl - Thermo Scientific), 2.5µl Taq Buffer (10X) and the total volume was adjusted to 25µl with molecular biology grade water. Primers for PCR were designed using Oligo Explorer version 1.5. The composition of the PCR mastermix and the thermal cycling conditions are summarized in Table 2.2 & 2.3.

Table 2.1: **Primer Sequence of RET gene used for exons amplification**

Exon	Sequence (F: Forward, R: Reverse)	Length	Annealing Temperature
Exon 1	F: CCTAGCCGCAGTCCCTCCAG	20	67.5
	R: TGCCCACGGCAAACAGAAAG	20	
Exon 2	F: CAGTTCTTTTCTAGCCCGTG	20	65
	R: GAAGCCATGCCAAGCTTCAG	20	
Exon 3	F: CTGAGCCTTGTGGACCTTGG	20	65
	R: GGCAGGCATGCAAGAGGTAC	20	
Exon 4	F: CTTCCCGAGGAAAGCGGCTG	20	65
	R: ACGGAGGCGAAGGACAAATG	20	
Exon 5	F: GACGTGCAGCATTCTAAGG	19	62.5
	R: CCAGGTGTGCATGTGTGTAG	20	
Exon 6	F: TGTTTGCACCAGTGTGAGTG	20	62.5
	R: TTCTGTGCACACAGCTGTAC	20	
Exon 7	F: CTACCCTCAGGCCATTACAG	20	62.5
	R: GGAGCAACCATTTACTGCTG	20	
Exon 8	F: GTGCTGTTCCCTGTCCTTGG	20	65
	R: CACTGCTGCCCGCTATGCTG	20	
Exon 9	F: TCCAGTTGCTCCTCCCTAGAG	21	62.5
	R: GCATAGAACTGACAGCCCTG	20	
Exon 10	F: TATGCTTGCGACACCAGTTG	20	59
	R: ACTCGCCTCCCAGCAATTTC	20	
Exon 11	F: TGAGGCAGAGCATACGCAGC	20	67.5
	R: GCGTTGGCAGCCCCTCACAG	20	
Exon 12	F: GGGTCTTGCCTTCTTCCTCC	20	62.5
	R: AACTGCAGCCGCTCTAGAAC	20	
Exon 13	F: CCAGGAGCGATCGTTTGCAAC	21	67.5
	R: AAAGAGGGAGAACAGGGCTG	20	

Exon 14	F: GTCCACCCCCTTACTCATTG	20	62.5
	R: GTGGGCTAGAGTGTGGCATG	20	
Exon 15	F: TGCTGGTCACACCAGGCTGA	20	59
	R: CGGTATCTTTCCTAGGCTTC	20	
Exon 16	F: CTCCTCTGGTTACTGAAAGC	20	59
	R: GGCCCCACTACATGTATAAG	20	
Exon 17	F: CATCTGTGCTTGAAGCCGAC	20	62.5
	R: GGATGGCTCTCTAACCTCTG	20	
Exon 18	F: TGGTGGGCTGTCCTTCTGAG	20	65
	R: GGGTTCAATCTGCTGTCTGC	20	
Exon 19	F: CAGGAGATTGGTGCGAGGTG	20	62.5
	R: AACTGGATGCAAAGTCAGAC	20	
Exon 20	F: GAGTTTTGCCAAGGCCTTAC	20	62.5
	R: TCACAAAGAAAGGGCCGGTA	20	

Table 2.2: Mastermix composition for PCR

Components	**Volume /reaction**
MilliQ	15.0 µl
10 X PCR buffer	2.5µl
2.5 mM dNTP	1.0µl
10 pmoles/µl Primer (F)	0.5 µl
10 pmoles/µl Primer (R)	0.5 µl
20 ng/µl DNA template	5.0µl
2U/µl Taq polymerase	0.5 µl
Total volume	**25 µl**

Table 2.3: Thermal 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
Go to Step 2-4: 35 cycles		
5. Final Extension	72ºC	5 mins

* annealing temperature for specific primer pairs are given in Table 4.

2.6 Agarose Gel Electrophoresis

For visualization of PCR products, 1% agarose gel was prepared whereas for visualization of DNA, 0.8% agarose gel was prepared.

Steps involved:

1.The required amount of agarose powder was weighed andadded in 1X TBE (Tris, Boric acid, EDTA buffer) to make a solution.

2. The agarose solution was boiled in a microwave to completely dissolve the agarose in the buffer. The solution was further cooled down to less than 50ºC.

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

4. The solution was poured on the gel casting tray and allowed to solidify.

5. The samples to be loaded were mixed with 1X loading dye containing xylene cynol, bromophenol blue and 40% glycerol.

6. The samples were loaded and the apparatus was run at the required voltage ($5V/cm^2$).

7. It was allowed to run until the dye traversed $3/4^{th}$ of the gel.

8. The DNA bands were visualized using a UV Transilluminator.

9. Pictures were taken and stored using Gel Documenter.

5X TBE Buffer Composition (1 litre)

Tris-HCL: 54 gms

Boric Acid: 27.5 gms

EDTA: 4.68 gms or 20ml 0.5M EDTA solution (pH 8.0)

Weigh and dissolve in 1 litre of distilled water for 5X concentration and can be stored at room temperature. Working 1X concentration can be made by diluting it with distilled water.

6X gel loading dye composition (50 ml)

Xylene cynol: 0.125 gms

Bromophenol blue: 0.125

MilliQ (distilled water): 30 ml

Glycerol: 20 ml

Weigh and dissolve in 30ml of distilled water for 6X concentration and add 20ml glycerol to provide the density to the dye.

2.7. Purification of PCR products

The PCR products were diluted according to the intensity of the band on agarose gel such that 30-50ng of template is available for Sanger sequencing. Two microliter of EXO-SAP IT [USB-Affimetrix] was added to 7µl of diluted PCR product. This mix was incubated at 37ºC for 15 mins for enzyme activity followed by incubation at 80ºC for 15 mins to deactivate the enzyme. The EXO-SAP enzyme contains Exonuclease and Shrimp Alkaline Phosphatase to remove the

unused dNTPs and primers which would otherwise interfere with downstream steps of sequencing of the PCR amplified products.

2.8. Sanger sequencing

2ul of cleaned PCR products with 1ul of 1.5 pmol amplicon specific 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 (8 capillary) or 3730 (48 capillary)(Applied Biosystems). Electropherograms were analyzed by Chromas Lite version 2.6.4 using reference sequence of *RET* gene extracted from National Centre for Biotechnology Information (NCBI) NG_007489.1.The threshold for mixed base for detection of heterozygous mutations was set at a ratio of 0.3 of the wild type.

Step 1: Cycle sequencing

Add 2µl of cleaned-up template DNA (PCR amplified products) and 1µl of 1.5 pmol primer in the 96-well sequencing plate.

Prepare the cycle sequencing reaction master mix as:

- MilliQ water - 4.75 µl
- 5X Sequencing Buffer - 1.75 µl
- Ready Reaction mix - 0.50 µl

Add 7µl of mastermix to each sample

Setup the cycle sequencing reaction in the thermal cycler as:

- 96ºC for 2 minutes
- 96ºC for 10 seconds
- 1ºC/second to 55ºC

- 55ºC for 0.05 minutes (5 seconds)
- 1ºC/second to 60ºC
- 60ºC for 4 minutes
- Go to steps 2-5 for 24 times
- 4ºCforever

Step 2: Post cycle sequencing cleanup

A mixture of the two BigDye X Terminator reagents (Premix) is prepared as:

- X Terminator Solution - 10µl
- SAM solution - 45µl

Add 55µl of the premix to each sample and tap spin the reaction plate

Vortex the reaction plate for 30 minutes

Centrifuge the reaction plate at 2500 rpm for 2 minutes

Tap spin and load the reaction plate on the sequencer.

2.9.Mutation analysis

The chromatograms were analyzed using Chromas Lite software, version 2.6.4 [Technelysium Pty Ltd].The data was compared with the reference sequence of the *RET* gene extracted from National Centre for Biotechnology Information (NCBI) NG_007489.1 to identify mismatches from the reference sequence. The mismatches were then checked for their pathogenecity using ARUP MEN2 database(http://www.arup.utah.edu/database/MEN2/MEN2_welcome.php). This database serves as a repository for MEN2-associated *RET* sequence variation and a reference for *RET* genotype-MEN2 phenotype correlations(68).

2.10. SNP Genotyping

SNP genotyping was done using Restriction Fragment Length Polymorphism (RFLP) or by TaqMan approach when restriction enzyme for a particular SNP could not be identified. Single site cutter restriction enzymes were identified using online tool NEBcutter version 2.0. This tool takes DNA sequences as an input and identifies the restriction sites/enzymes which will cleave that sequence at specific sites. For TaqMan, fluorescently labeled probes (Applied Biosystems) specific for a particular SNP was used. The primer sequences used to amplify the amplicons harboring the SNPs are provided in Table 2.4. Also, the restriction enzymes and the TaqMan probes used for SNP genotyping are given in Table 2.5 & 2.7.

Table 2.4: Sequence of Primers used for SNP genotyping

SNP	Sequence (F: Forward; R: Reverse)	Length	Annealing Temperature
Cyp1A1m1	ACTGTAGGGAGGAAGAAGAG	20	59.0
	GAGGGCGTAAGTCAGCACAG	20	
Cyp1A2*F	ACTGAGATGATGTGTGGAGG	20	62.0
	GCCGAGAAGGGAACAGACTG	20	
NAT2	CATGGAGTTGGGCTTAGAGG	20	62.0
	TGCTCTCTCCTGATTTGGTC	20	
GSTP1	TGTGTTGATCAGGCGCCCAG	20	65.0
	ATAAGGGTGCAGGTTGTGTC	20	
CDKN1A	GACATTAGCTTGCCCTTCAG	20	62.0
	CATGGTCTTCCTCTGCTGTC	20	
CDKN1B	TAGAGGGCAAGTACGAGTGG	20	55.0
	TGGTTGGGAAAGGGTCAT	18	
CDKN2A	CGGTAGGGACGGCAAGAGAG	20	59.0
	TGCTCACTCCAGAAAACTCC	20	

CDKN2B	ACTCAATCATTAGAGGCTAC	20	55.0
	CAGGTGGCTTCGAAAATGGA	20	
CDKN2C	GCGGGGTAACTACTCATCTC	20	55.0
	AATCCACTTACACAGACCAC	20	
RET G691S	TGAGGCAGAGCATACGCAGC	20	67.5
	GCGTTGGCAGCCCCTCACAG	20	
RET L769L	CCAGGAGCGATCGTTTGCAAC	21	67.5
	AAAGAGGGAGAACAGGGCTG	20	
RET S836S	GTCCACCCCCTTACTCATTG	20	62.5
	GTGGGCTAGAGTGTGGCATG	20	
RET S904S	TGCTGGTCACACCAGGCTGA	20	59.0
	CGGTATCTTTCCTAGGCTTC	20	

2.10.1 Restriction Fragment Length Polymorphism (RFLP)

RFLP is a molecular biology technique used to identify the variations in homologous DNA sequence such as Single Nucleotide Polymorphisms (SNPs) by detecting DNA fragments of different length after digestion with locus specific restriction endonuclease. For RFLP, 5µlgDNA (20ng/µl) was first PCR amplified followed by restriction digestion using reaction conditions as per manufacturer's protocol (Table 2.6). The digested products were visualized on 2 % agarose gel and the genotypes were inferred from band sizes in the gel. The incubation temperature of each restriction enzyme, amplicon product size post restriction digestion and specific allele recognized by each restriction enzyme is summarized in Table 2.5. The incubation time varies for each restriction enzyme but the general incubation time ranges between 1-16 hrs with no star activity. Around 5% of all the genotyping results were validated by Sanger Sequencing.

Table 2.5: Restriction enzymes used for SNP Genotyping

SNP	Restriction enzyme	R.E. digested product size	Incubation Temp.	Allele recognized by enzyme
Cyp1A1m1	SmaI	394, 232, 162bp	30	Variant allele
Cyp1A2*F	ApaI	300, 200, 100bp	30	Variant allele
NAT2	BtsCI	363, 222, 141bp	55	Wild-type allele
GSTP1	BsmAI	340, 201, 139bp	37	Variant allele
CDKN1A	HgaI	575, 345, 230bp	37	Variant allele
CDKN1B	BglI	288, 173, 115bp	37	Variant allele
CDKN2A	BspEI	385, 255, 130bp	55	Wild-type allele
RET L769L	TaqI	228, 163, 67bp	65	Wild-type allele
RET S836S	BsrI	453, 232, 221bp	65	Variant allele
RET S904S	BsiWI	293, 182, 111bp	55	Variant allele

Table 2.6: RFLP reaction condition:

Components	Volume/Reation
Nuclease free water	8-11 µl
10X Buffer	2.0 µl
Restriction Enzyme	0.1 µl
PCR amplified product	7-10 µl (150-200ng)
Total reaction volume	**20 µl**

2.10.2 TaqMan SNP Genotyping

The TaqMan SNP genotyping technique utilizes the 5' nuclease activity of *Taq* polymerase to generate a fluorescent signal during PCR. The TaqMan assay requires a DNA template, Taq polymerase, forward and reverse primers specific to the region of interest (region harboring the polymorphic site) and two probes with different fluorescent reporter dyes. One probe is labeled with VIC fluorescent dye and is specific for wildtype sequence whereas the other probe is labeled with FAM fluorescent dye specific for the polymorphic sequence. If the sample is homozygous for wildtype allele, the fluorescence readout will show VIC fluorescence, while if the sample is homozygous for variant sequence, the readout will be the FAM dye. If the sample is heterozygous, there would be almost equal signal for both the dyes.

Working Principle:

The technique is based on the FRET technology where a 5' reporter dye and a 3' quencher dye are covalently linked to the wild-type and variant allele probes. When the probes are intact, fluorescence is suppressed because of the quencher dyes. During the PCR, the TaqMan probes hybridize to the targeted polymorphic site. During PCR extension, the fluorescent reporter and quencher are cleaved due to the 5' exonuclease activity of the Taq polymerase, resulting in fluorescence of the reporter dye. The exonuclease activity only happens on the perfectly hybridized probes, since a probe containing a mismatched base will not be recognized by the Taq polymerase (Figure 2.1).

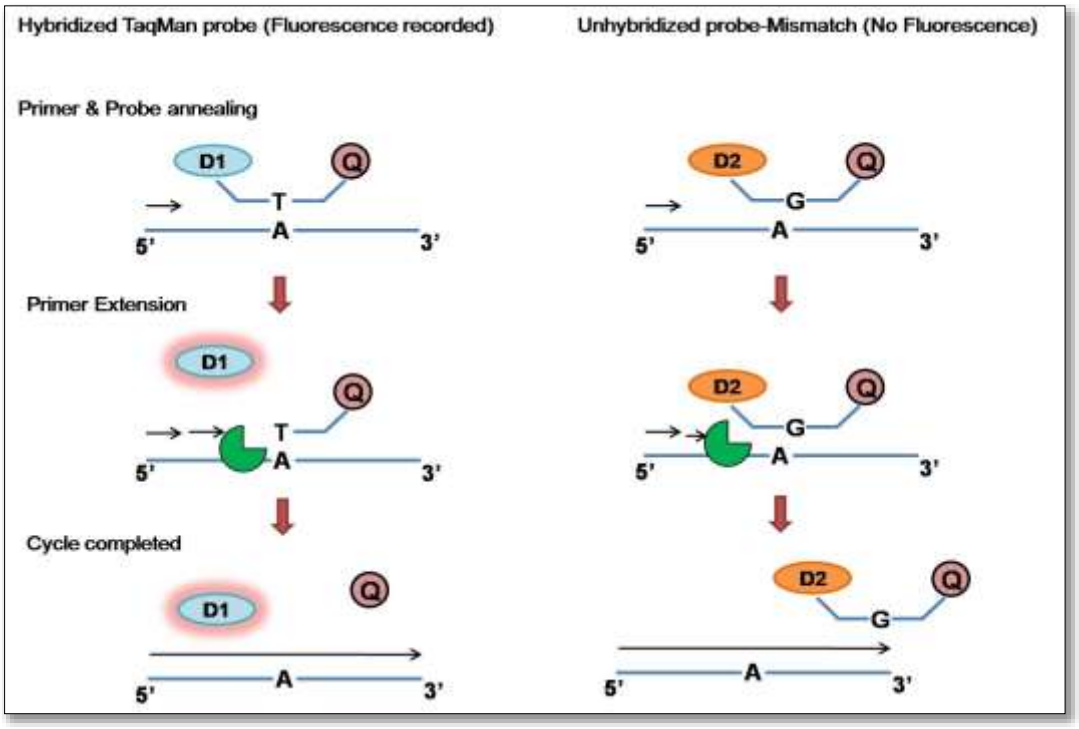


Figure 2.1: TaqMan working principle

Steps involved:

For TaqMan SNP genotyping, 1µl of gDNA (10ng/µl) was mixed with 2.5µl of TaqMan universal master mix II with UNG (Applied Biosystems, cat#4440038) and 0.1µl probe mix (Applied Biosystems) designed for each SNP (Table 2.8). TaqMan realtime PCR was performed on QuantStudio 5.0 and genotypes were inferred from amplification plot and allelic discrimination plots. About 5% of all the genotyping results were validated by Sanger Sequencing. The IDs of the taqMan probes used for SNP genotyping are given in Table 2.7.

Table 2.7: TaqMan probes used for SNP Genotyping

SNP	TaqMan Probe
CDKN2B	C___2618046_10
CDKN2C	C___1452499_10
RET G691S	C___3204350_10

Table 2.8: TaqMan reaction condition:

Components	Volume/Reation
Nuclease free water	1.4 µl
2X Master Mix	2.5 µl
40X Assay Mis	0.1 µl
gDNA	1.0 µl (5-10ng)
Total reaction volume	**5 µl**

2.10.3 Statistical Analysis

All Statistical analysis was performed on SPSS v21.0. SNP genotypes were tested for Hardy-Weinberg equilibrium (HWE) using Chi-square Hardy-Weinberg equilibrium test calculator for biallelic markers (http://www.oege.org/software/hwe-mr-calc.shtml). Genotypic frequency and minor allele frequencies were calculated for all the studied SNPs in both cases and controls. Risk-association study as well as association with MTC progression was done using Dominant Model of inheritance (AA vs Aa+aa) as the homozygous status of several SNPs was either absent or very low in either cases or controls. Logistic regression analysis was used to analyze the association between the polymorphisms and MTC risk and odds ratio was calculated with 95% Confidence

Interval (CI). Odds ratio is a measure of association between an exposure and outcome. All SNPs showing a trend for association on univariate analysis with $p<0.1$ were included in the multivariate logistic regression analysis. When multiple comparisons were made, a p value of <0.01 was used to consider an association as statistically significant.

Odds Ratio	Possible outcomes
OR = 1	Exposure does not affect odds of outcome.
OR > 1	Exposure is associated with higher odds of outcome.
OR < 1	Exposure is associated with lower odds of outcome.

2.11. Whole Exome Sequencing

Exome library was prepared using Nextera Rapid Capture Exome kit (Cat# FC-140-1002) designed to sequence the entire coding region (exome) of human genome. The probes in the kit were designed to enrich 214,405 exons which covers 98.3% of exome providing a uniform and specific coverage of 37Mb. The starting material for sample library preparation using nextera rapid capture exome kit is 50ng high quality gDNA.

2.11.1.Exome library preparation:

1.First step of exome library preparation is tagmentation of gDNA where Nextera transposomes (enzymes) fragments the gDNA and adds adapter sequence to their ends in a single step called tagmentation (fragmentation + adapter tagging). After this step, the tagmented DNA is purified

from transposomes which would otherwise bind to DNA ends & interfere with the downstream process. For this, sample purification beads (magnetic beads) are used.

2. In this step, index adapters & sequencing adapters are ligated to the ends of tagmented DNA & are then amplified. Index adapters provides unique identity to each sample & sequencing adapters provides binding sites for sequencing primers which binds & sequence the template DNA.

3. The next step is 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.

4. Each library is pooled at a concentration of 500 ng. So if we are pooling 10 samples, then the final library pool concentration should be 5000 ng. Once the pooling is done, the probes are added to the library to enrich the target region.

5. In this step the probes that are hybridized to the target region are captured using streptavidin magnetic beads & non specific binding are washed from the beads resulting in exome enriched library.

6. The step of probe hybridization & capture is done twice to ensure high specificity of captured regions in enriched libraries & removal of nonspecific bindings.

7. In the final step, clean up of enriched library is done using sample purification magnetic beads and is amplified to generate exome library to be used for the subsequent steps of cluster generation and sequencing.

2.11.2. Exome Sequencing Data Analysis

The raw data output of exome sequencing run is in the form of image files(tiff files) which are processed to generate base calling files (.bcl files) and then fastq files. There are several steps involved in processing and analysis of the data starting from processing of files for quality check,alignment and analysis for exome, transcriptome or whole genome analysis. First step of exome data analysis is the quality checking of fastq files. If the base sequence quality deteriorates towards the end of the read then those reads are trimmed from the ends. Also, before mapping / alignment of the reads to the reference genome, the genome file needs to be indexed.

Steps of Exome data analysis (GATK Pipeline)

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:** Processed FastQ files are aligned to human reference genome sequence, extracted from UCSC.Alignment is done using BWA software/Mapping tool. After alignment .sai files are generated which is converted into readable text files-SAM files.Large SAM files are then compressed into BAM files using SAMtools command.

3. **Post-Alignment processing:**After alignment with reference genome, post alignment processing of BAM files is done which includes marking of PCR duplicates using PICARD tools, local realignment around InDels 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 or VarScan which Calls SNPs & short InDels at the same time & gives well annotated VCF file as an output.

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

Steps of Exome data analysis (VARSCAN Pipeline)

1.Pre-Alignment QC: Quality Check (QC) of the Fastq files with FastQC software and poor quality reads are removed.

2.Alignment: The QC passed reads are aligned with the indexed reference genome sequence (hg19) using BWA mapping tool and generates BAM files.

3.Post-Alignment processing: The aligned BAM files are processed before variant calling. These processing steps include:

- **Samtools Fixmate:** For paired end sequencing data, the BAM files are fixed using samtools fixmate command. Here, the reads whose mates does not map as expected with the reference genome are removed (mate 1 and mate 2 are read 1 and read 2 in paired end sequencing). This fixmate command generates fixmate.bam files.

- **Picard Samsort:** Sorted BAM has the aligned data sorted by chromosomes/contigs/scaffolds in order to efficiently display/access the BAM files. The Samsort command generates fixmate_sorted.bam files.
- **Picard Markduplicates:** This command removes the reads which are the exact copy of other reads and are called as duplicate reads. They might lead to false positive calls during variant calling step hence needs to be removed.
- **Samtools mpileup:**The normal and tumor BAM files are first converted into SAMtools pileup file format for variant calling using VarScan.

4. Variant Calling: VarScan 2.0 was used for variant calling which calls somatic variants (SNPs and Indels) based on the number of aligned reads supporting each allele. At every position where both the normal and the tumor have sufficient coverage (minimum coverage for normal: 8X, for tumor: 6X), a comparison is made using the following algorithm:

- **<u>If Tumor matches normal:</u>**

  If the tumor and the normal matches the reference sequence

  $\Longrightarrow$ Call Reference

  If the tumor and the normal do not match the reference

  $\Longrightarrow$ Call Germline

- **<u>If Tumor does not match the normal:</u>**

  Calculate the significance of allele frequency difference by Fisher exact test. If the difference is significant ($p<0.05$)

⟹ Call Somatic

If difference is not significant (p>0.05)

⟹ Call Germline

If normal and tumor are variant but different

⟹ Call Indel filter or Unknown

Output: Varscan variant call generates two files:

- Indel.vcf
- SNP.vcf

[Criteria for somatic variant call: Minimum variant allele frequency in tumor should be 0.02 (2 reads per 100 reads should have the variant), and the variant should either be absent in normal or with the maximum variant allele frequency of 0.05]

5. Isolating variants by type and confidence: The command'ProcessSomatic' will separate the somatic output file by somatic status (germline, somatic and LOH). Somatic variants will further be classified as high confidance (.hc) or low-confidance (.lc). The above command will produce the following output files:

- SNP germline.vcf Indel Germline.vcf
- SNP germline hc.vcf Indel germline hc.vcf
- SNP somatic.vcf Indel somatic.vcf
- SNP somatic hc.vcf Indel somatic hc.vcf

[For high confidence somatic variant call: Minimum VAF in tumor should be 0.02, Maximum VAF in normal should be 0.05 and $p<0.05$].

6. **Somatic Filter**: Variants adjacent to Indels are filtered out considering them to be false-positive.

7. FP Filter (False-Positive): Base Quality score should be more than Q15 (Q15 and above), $p<0.05$, depth required to test somatic p value should be 8X (at least 8 reads should cover the region of the genome harboring the variant and the variant should be present in at least 4 reads).

8. Further, Intronic variants are removed and only exonic and splice-site variants are retained.

9. Synonymous exonic variants are removed and non-synonymous (missense) and stopgain (nonsense) variants are retained.

10. Variants are annotated to different databases to sort them based on known/reported and novel variants.

Chapter 3

To identify novel and recurrent germline RET mutations, their penetrance and Genotype-Phenotype correlation in a cohort of Indian MTC cases.

3.1 Background

3.1.1 RET genetic testing in MTC and predisposition to MEN2 syndrome

The genes responsible for predisposition to MEN2 syndrome was identified in 1985 during transfection of NIH3T3 cells with human lymphoma DNA (discussed in chapter 1 in detail) and was later found to be localized on chromosome 10 (10q11.2) by genetic linkage analysis in 1987(57,69). Further in 1993, Mulligan et al and Donnis-Keller et al., identified the first missense mutations in the extracellular cysteine rich domain of *RET* in patients with MTC(61,64). Subsequently, point mutations of the RET (REarranged during Transfection) proto-oncogene were identified in MEN2A, MEN2B, and FMTC patients. The mutations in the RET gene predisposing to MEN2 syndrome result in single amino acid substitutions affecting key residues in the extra-cellular and kinase domains, called as 'mutation hotspots' (exon 10, 11, 13, 14, 15 & 16). This leads to ligand independent inappropriate constitutive activation of RET receptor.

Genetic testing of RET gene in families with MEN2 syndrome revealed that nearly all of these families have germline mutations in RET gene(53). This discovery lead to major advances in the understanding the genetic basis of this rare syndrome and made a significant contribution in the clinical management of the patient. It also enabled the pre-symptomatic interventions to at-risk family members. Since a strong genotype-phenotype correlation exists for MEN2, the detection of RET mutations has produced a major impact in early recognition and treatment of MTC and MEN2(70). Genetic testing detects about 98% of mutation carriers and is considered to be standard care for all affected individuals and their first-degree relatives(71). Therefore all individuals presenting with MTC or primary CCH should be offered germline RET testing

regardless of other syndrome associated features or family history as 4-10% patients with apparently sporadic MTC are found to harbor germline RET mutations(52).

3.1.2 Mechanism of pathogenic RET activation

RET is activated by the complex of ligands and co-receptor (described in Chapter 1). The ligands form a complex with the co-receptors which then brings about the dimerization of the two RET monomers leading to auto-phosphorylation and activation of the tyrosine kinase domain of RET. The activated RET provides a docking site for a number of adaptor molecules including STAT3, GRB7/10, SRC, DOK2, DOK4, DOK5, PKA, SHANK3, GRB2, PLC etc. that links the activated RET to the downstream signaling pathways involved in cell proliferation, differentiation and survival. The major pathways that gets activated upon RET activation include JAK-STAT pathway, RAS-MAPK pathway and PI3K-AKT pathway(56).

The activation of oncogenic RET depends upon the location of mutation in the RET protein. Under normal physiological conditions, the cysteine residues of the cysteine rich extracellular domain of RET are involved in intra-molecular disulfide bond formation. A mutation in the cysteine rich domain substitutes a cysteine with a non-cysteine residue which leaves an unpaired cysteine residue in the RET monomer. This unpaired cysteine then forms an inter-molecular disulfide bond with unpaired cysteine of another RET monomer leading to dimerization of RET in the absence of ligand(72). This receptor dimerization leads to phosphorylation of the tyrosine kinase domain and constitutive activation of the RET protein(56).

For the mutations in the intracellular tyrosine kinase domain, which gives phenotype of MEN2B or FMTC, RET gets activated in the monomeric state and does not require receptor dimerization. The specific mutation M918T producing MEN2B phenotype lies in the catalytic domain of RET

and alters the substrate specificity of the RET protein due to structural changes of the binding pocket. This leads to aberrant phosphorylation of the substrates preferred by cytoplasmic tyrosine kinase such as src and abl rather than those preferred by receptor tyrosine kinase (73). Thus the mutated RET no longer needs dimerization to become active.

Although mutated RET signals independent of ligand, in certain mutation types RET gets further activated by GDNF(74). This was shown for MEN2B associated intracellular mutations as opposed to intracellular FMTC mutations. The same phenomenon was observed for extracellular codon 634 mutations which was not seen for other extracellular codon mutations(74,75).

3.1.3 Study Objective

RET genetic testing is widely used to detect the inherited pattern of the disease and identification of mutation carriers for clinical management & preclinical diagnosis of MTC. Since the spectrum of mutations may vary between populations, it is important to characterize the mutations in each population. The present study is the first large and comprehensive study aimed to identify novel and recurrent germline RET mutations in Indian population, determine their penetrance and establish their Genotype-Phenotype correlation in a cohort of a large Indian MTC cases.

3.2 Methodology

A total of 245 MTC cases enrolled in the Cancer Genetics Clinic, Tata Memorial Hospital, between 2014-2019 were included in this part of the study. The clinical diagnosis of MTC was made by histopathological examination of surgically resected tumor tissues along with biochemical confirmation of elevated serum Calcitonin, a biochemical marker of MTC diagnosis. Detailed family history & clinico-pathological details were recorded for each patient. Blood sample was collected after written informed consent. Genomic DNA was extracted as

described in Materials & Methods chapter. All samples were tested by PCR-Sanger Sequencing for RET gene exons 10, 11, 13, 14, 15 & 16, which covers >95% of all known RET mutations (59). Detailed steps of sequencing are provided in Materials & Methods chapter. Based on identified mutation & syndromic/phenotypic features, cases were classified as sporadic & hereditary (MEN2) MTC cases.

3.3 Results

3.3.1 RET mutation spectrum in Indian MTC cohort

Of the 245 probands with MTC, 139 were males and 106 were females. The mean age of MTC diagnosis in females was 38 years (SD $\pm$ 13.9; Range:12y – 76y) and in males it was 43 years (SD $\pm$ 14.7; Range: 9y – 72y). Based on phenotypic characterization, the 245 MTC families included 19 MEN2A/FMTC families, 8 MEN2B families and 218 sporadic non-syndromic MTC families. Germline RET mutation was identified in all 19/19 MEN2A/FMTC and 8/8 MEN2B families as well as 12/218 sporadic non-syndromic MTC families. Fifty first degree relatives from these mutation carrier families were screened for family specific mutations and 16/50 (32%) were found to be mutation carriers. Thus, a total of 55 RET mutation carrier individuals were identified which included 39 probands and 16 first degree relatives.

Table 3.1: RET mutations identified in MEN2 and sporadic MTC families

Phenotype	MTC family tested (n = 245)	Mutation carrier families (n=39)	First degree relatives tested (n=50)	First degree mutation carriers (n=16)
MEN2A/FMTC	19	19	22	10
MEN2B	8	8	12	1
Sporadic Non-syndromic	218	12 (6%)	16	5

We identified 15 distinct RET mutations spanning across all the six hotspot RET exons (Table 3.1). Most frequently found mutation was at codon 634 and was found in 17/55 (31%) mutation carrier individuals (Figure 3.1).The exon wise prevalence of the mutations is shown in Figure 3.2. Mutations in exon 10 was identified in 3/55 cases (5%), exon 11 in 18/55 (33%) cases, exon 13 in 10/55 (18%) cases, exon 14 in 12/55 (22%) cases, exon 15 in 3/55 (5%) cases & exon 16 in 9/55 (16%) cases. We observed a high prevalence of mutations in exon 11 of RET (33%) which included mutations of codon 630, 634 and 666. Representative electropherograms of a few different types of RET mutations identified in our cohort are shown in Figure 3.5.

Table 3.2: Distinct germline RET mutations identified in ACTREC cohort

Sr. No	Distinct RET mutations	Exon	No. of families	No. of individuals with mutation
1	C611Y	Exon10	1	3
2	C630R	Exon 11	1	1
3	C634F	Exon 11	1	1
4	C634G	Exon 11	1	1
5	C634R	Exon 11	9	11
6	C634S	Exon 11	1	2
7	C634Y	Exon 11	1	1
8	E768D	Exon 13	6	8
9	I852M	Exon14	3	6
10	K666E	Exon 11	1	1
11	L790F	Exon 13	1	2
12	M918T	Exon 16	8	9
13	S891A	Exon 15	1	3
14	V804L	Exon 14	1	3
15	V804M	Exon 14	3	3
Total			**39**	**55**

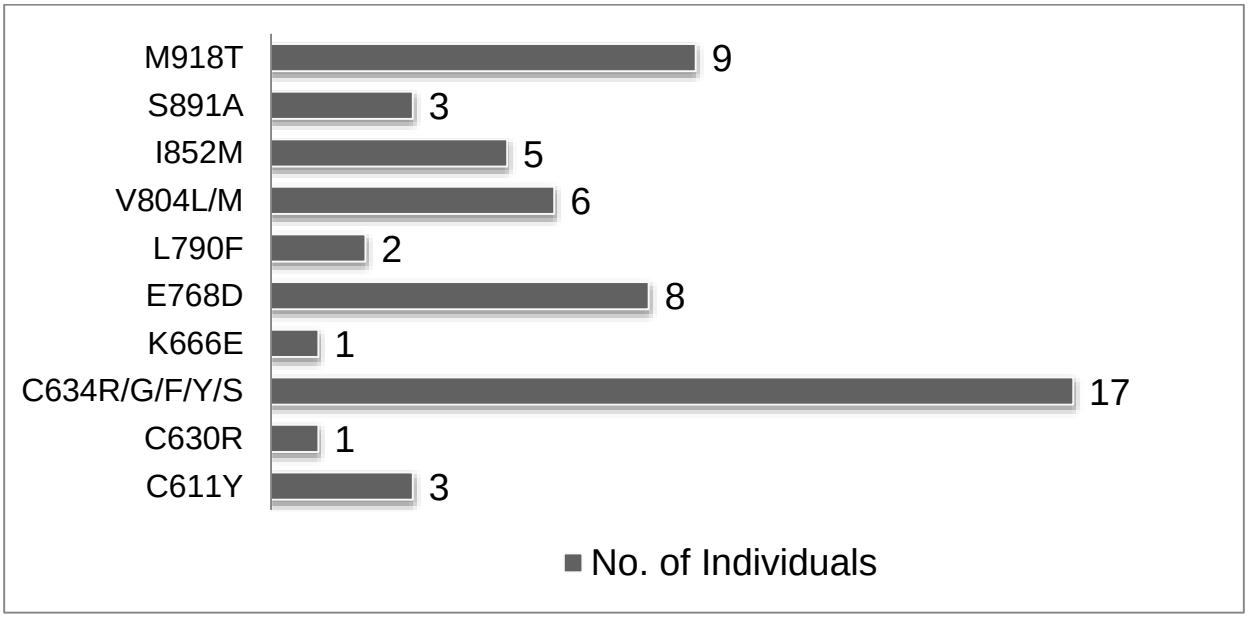


Figure 3.1: Frequency of distinct germline RET mutations in mutation carrier individuals

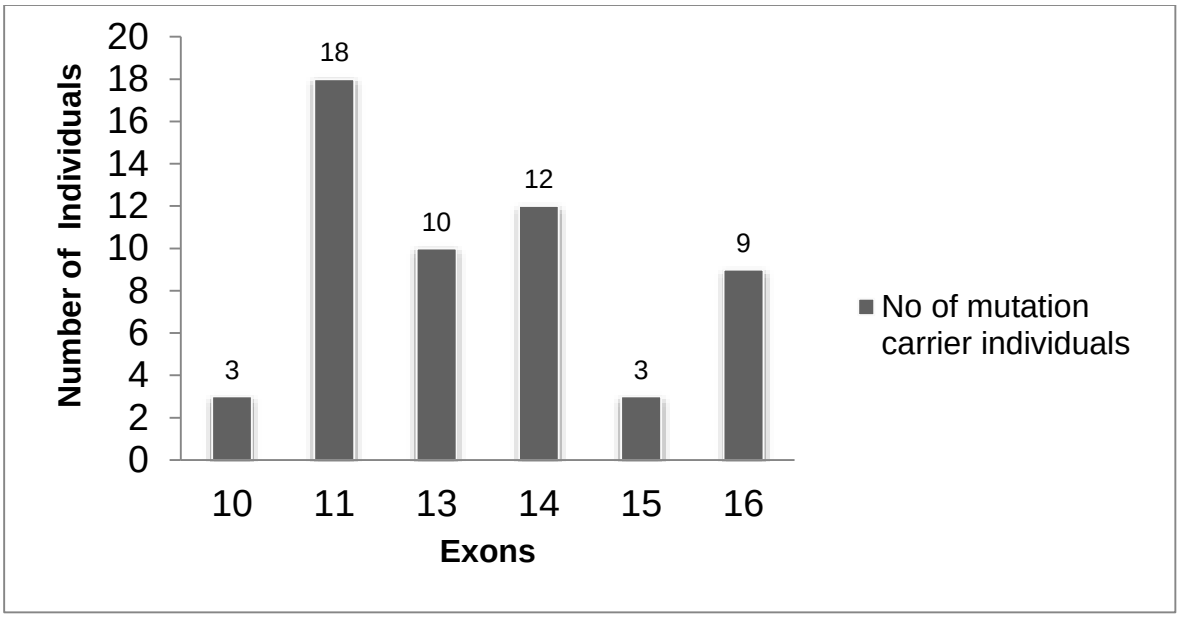


Figure 3.2: Exon wise distribution of RET mutations (n=55)

A rare combination of double germline RET mutation was identified in one family in our cohort (Figure 3.3 & 3.4). The proband was diagnosed with highly aggressive, metastatic MTC at the age of 31 years and genetic testing identified germline double mutation – C634R and I852M in exon 11 and 14 of the RET gene. On testing his first degree relatives, it was found that he had inherited the C634R mutation from his father whereas the I852M mutation was inherited from his mother. His brother and daughter were found to be carriers of I852M mutation. The C634R mutation belong to ATA risk category 'C' with a high level of disease aggressiveness whereas the I852M mutation belong to ATA risk level 'A' with moderate level of disease aggressiveness. All family members with a single I852M mutation were asymptomatic and on biochemical evaluation, serum calcitonin was in the normal range (<2pg/ml). The father of the proband who was the carrier of C634R mutation was diagnosed with a thyroid nodule at the age of 65 years with raised serum calcitonin level (800pg/ml).

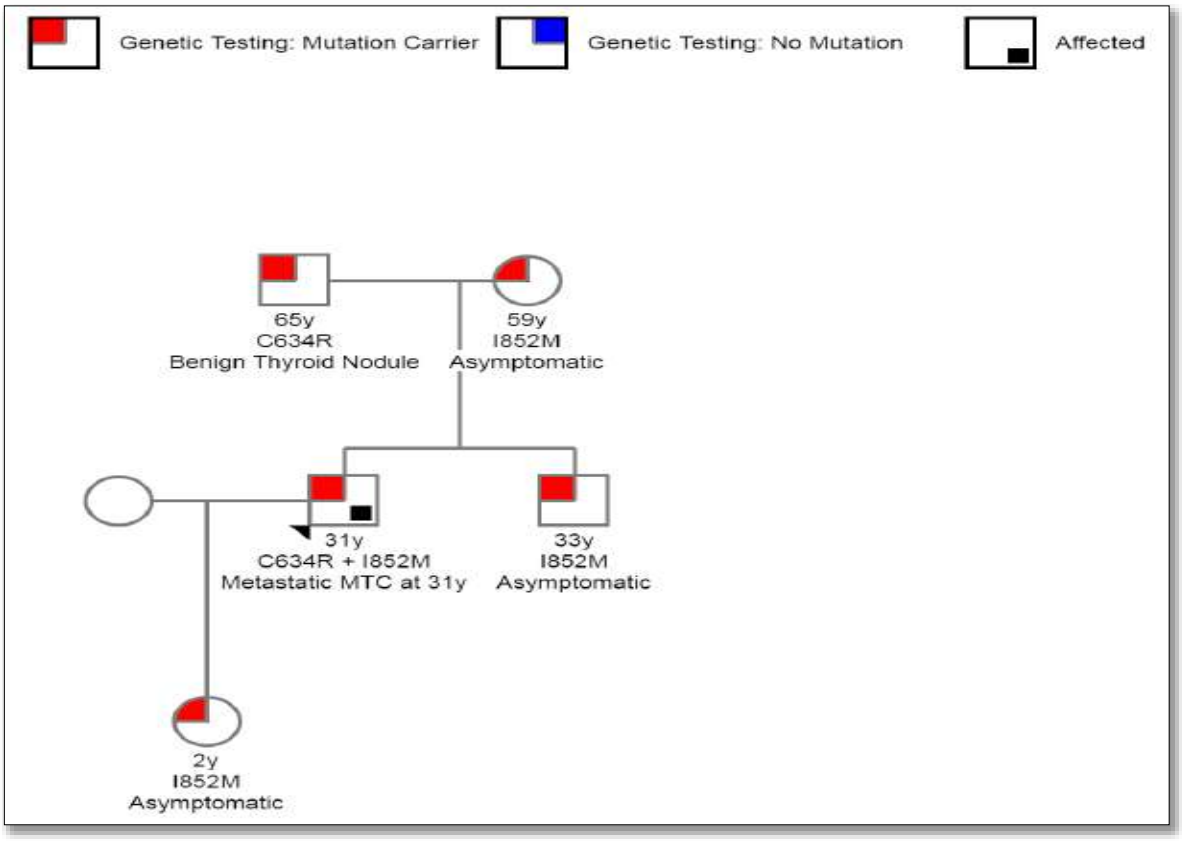


Figure 3.3: The pedigree of double RET mutation carrier family

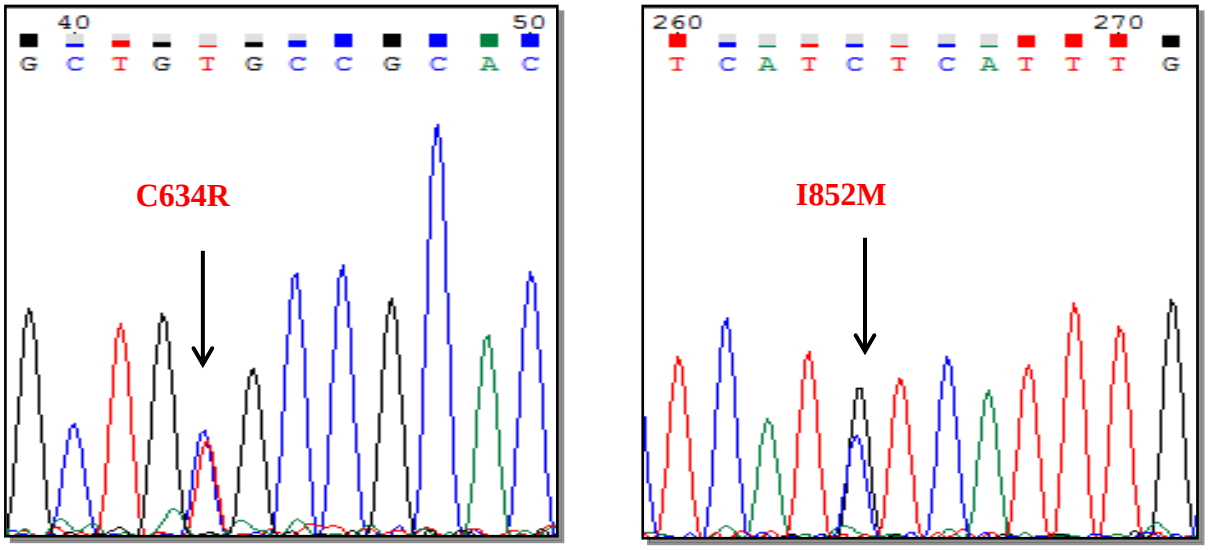
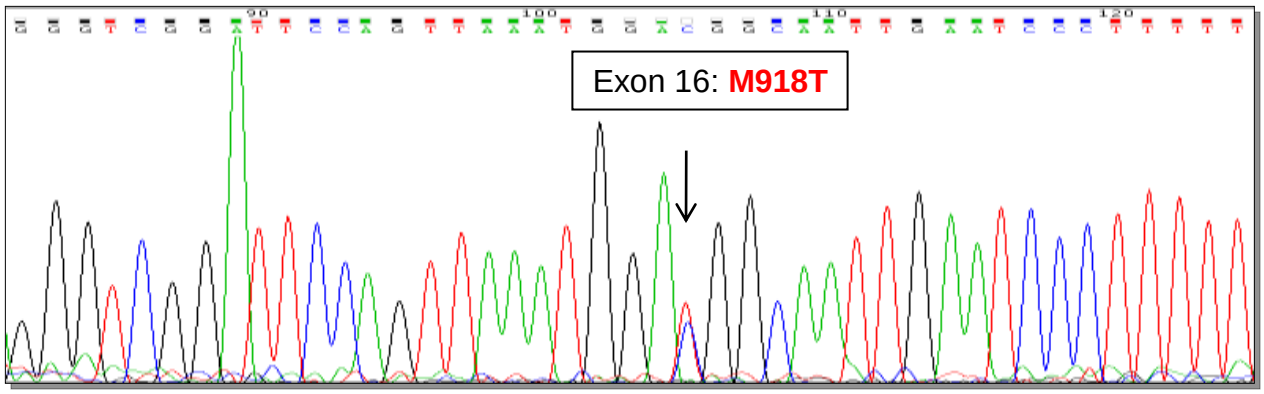
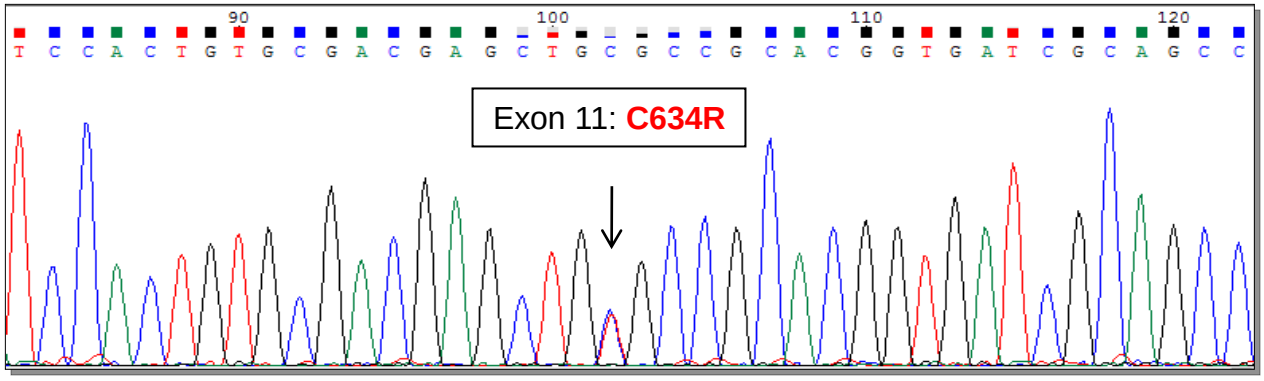
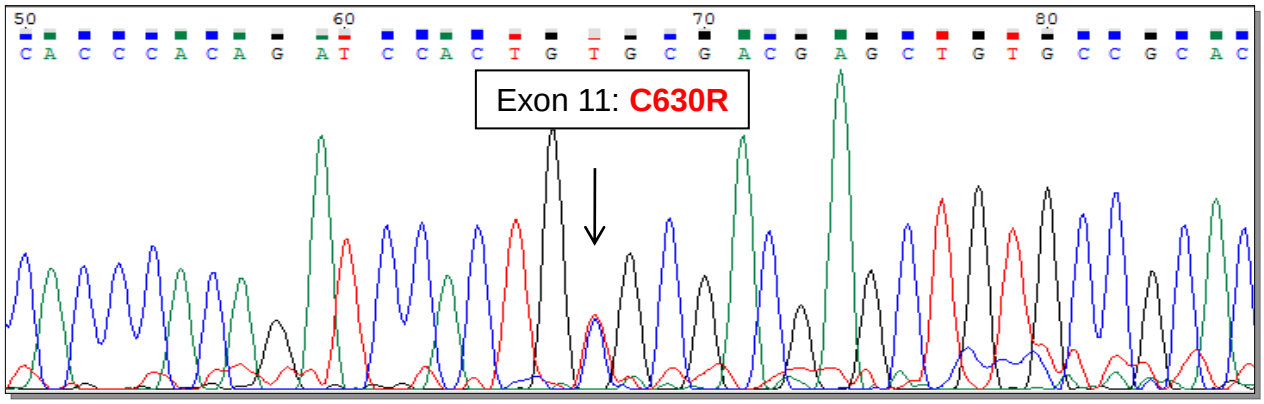
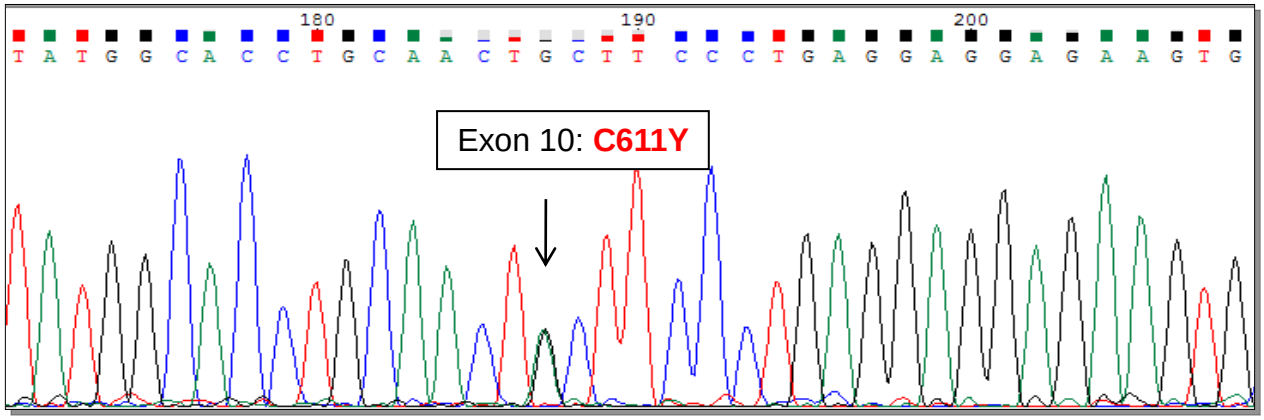


Figure 3.4: Electropherogram showing C634R (exon 11) & I852M (exon 14) mutation


Exon 16: M918T


90
100
110
120
T C C A C T G T G C G A C G A G C T G C G C C G C A C G G T G A T C G C A G C C
Exon 11: C634R


50
60
70
80
C A C C C A C A G A T C C A C T G T G C G A C G A G C T G T G C C G C A C
Exon 11: C630R


180
190
200
T A T G G C A C C T G C A A C T G C T T C C C T G A G G A G G G A G A A A G T G
Exon 10: C611Y

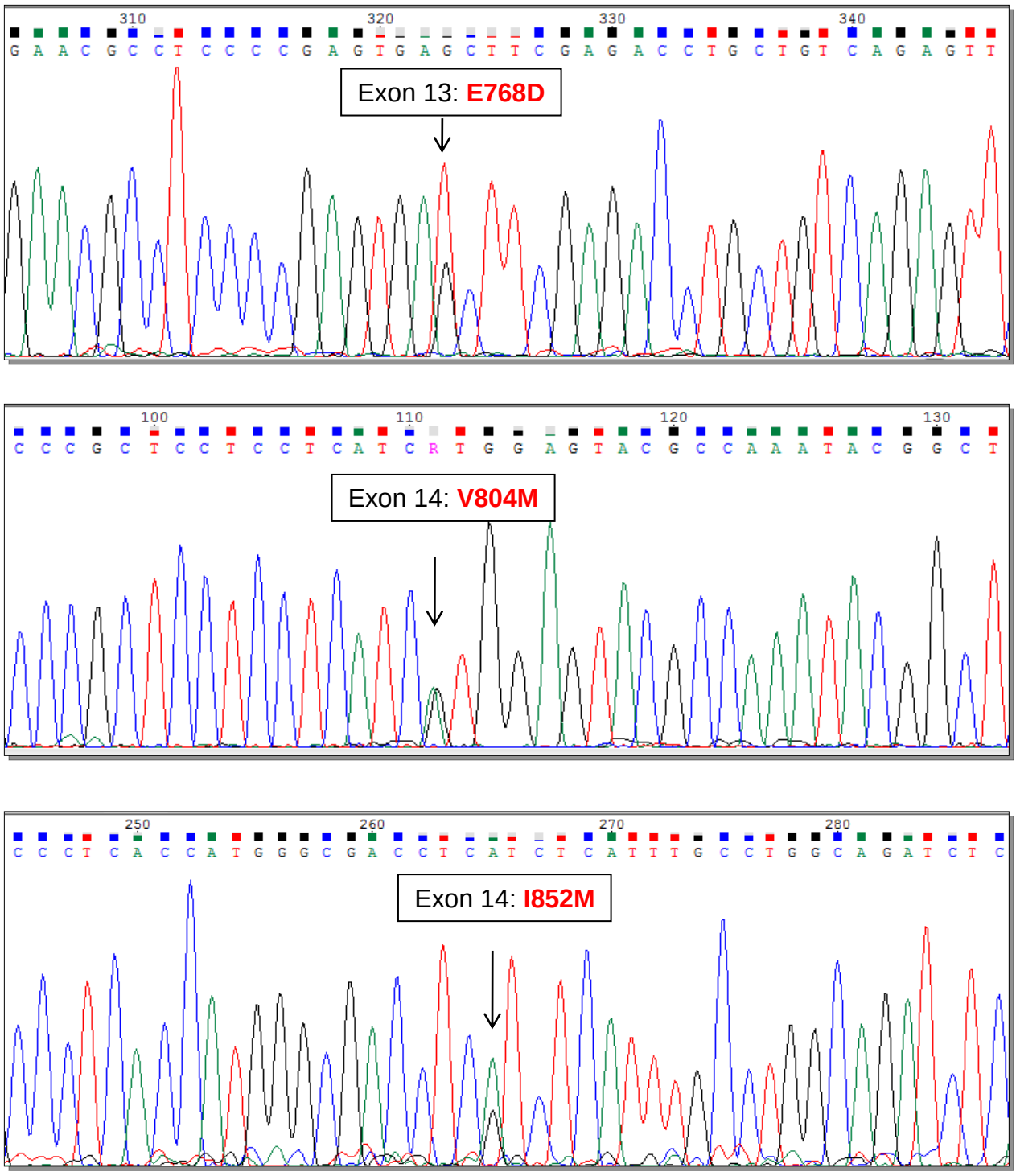


Figure 3.5: Electropherograms of different RET mutations identified in our cohort.

3.3.2 Genotype-Phenotype Correlation

A strong genotype-phenotype has been established between location of mutation in RET gene and the phenotype of MEN2 syndrome in terms of age of disease onset, its aggressiveness and the presence of other syndrome associated phenotypic features such as pheochromocytoma, hyperparathyroidism, mucosal neuromas, Marfanoid features etc. The genotype-phenotype correlation explained in detail in Chapter 1, helps in risk stratification of MEN2 by its genotype.

The genotype-phenotype correlation in our cohort is summarized in Table 3.2 & Figure 3.6. All the 9/55 (16%) individuals with M918T codon mutation in the tyrosine kinase domain of the RET protein developed MEN2B phenotype with presence of early onset aggressive MTC and MEN2B specific phenotypic features such as mucosal neuromas, blubbery lips, Marfanoid features etc. The median age of MTC diagnosis in MEN2B patients was 17y (Table 3.3). The 16/55 (29%) individuals with codon 634 (C634R/G/F/Y/S) mutation in exon 11 of RET gene that falls in the extracellular cysteine rich domain of RET receptor presented phenotype of MEN2A.The median age at MTC diagnosis for patients with MEN2A phenotype was 24y (Table 3.3). The 5/16 (31%) MEN2A patients also developed pheochromocytoma. For other remaining 30/55 (55%) individuals with mutations at codon 611, 630, 666 of the extracellular cysteine rich domain and codon 768, 790, 804, 852 and 891 of the tyrosine kinase domain, the patients presented the phenotype of FMTC, where MTC was the only clinical manifestation. The median age at MTC diagnosis in FMTC patients was 40y (Table 3.3). Based on the observed genotype-phenotype correlation, recommendations for prophylactic thyroidectomy has been proposed (Table 3.3).

Table 3.3: Genotype-Phenotype correlation in our cohort of MTC cases

Exon	Codon	Phenotype	ATA risk category	No of Individuals
10	611	FMTC	A (Moderate risk)	3
11	630	FMTC	B (Moderate risk)	1
	634	MEN2A	C (High risk)	16
	666	FMTC	A (Moderate risk)	1
13	768	FMTC	A (Moderate risk)	8
	790	FMTC	A (Moderate risk)	2
14	852	FMTC	A (Moderate risk)	6
	804	FMTC	A (Moderate risk)	6
15	891	FMTC	A (Moderate risk)	3
16	918	MEN2B	D (Highest risk)	9

Table 3.4: Recommendation for prophylactic thyroidectomy based on Genotype-Phenotype correlation

Phenotype	No of Individuals (%)	ATA risk category	Median age at MTC diagnosis	Recommended age of prophylactic thyroidectomy
FMTC	30 (55%)	Moderate	40y	May be delayed beyond 5 year with surveillance for annual serum calcitonin levels
MEN2A	16 (29%)	High	24y	Before the age of 5year
MEN2B	9 (16%)	Highest	17y	Before the age of 5year

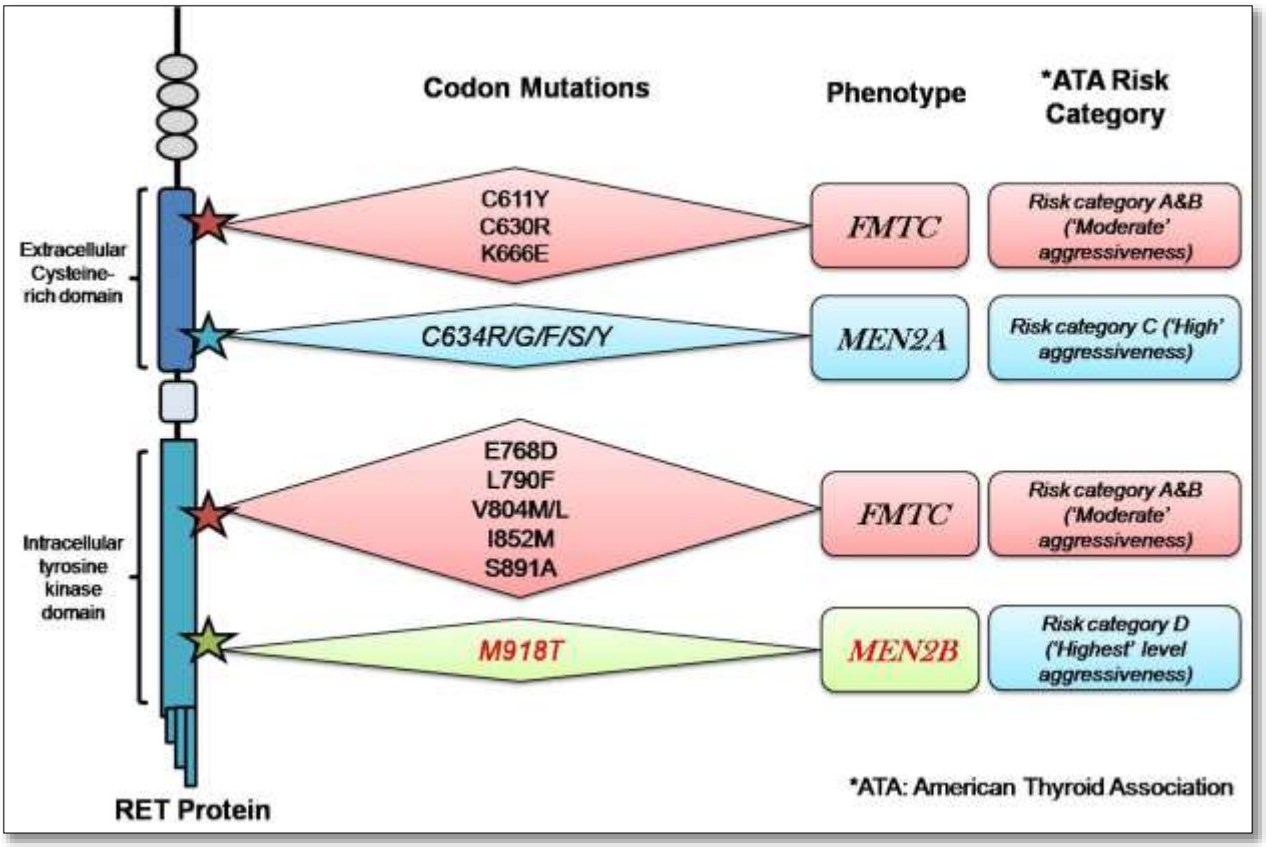


Figure 3.6: Correlation of MEN2 pheotype with site specific codon mutation in RET.

3.3.3 Penetrance of distinct germline RET mutations in Indian MEN2 syndrome cohort

We have also grouped each RET mutation into different American Thyroid Association (ATA) risk category and calculated the age related penetrance of RET mutations of each risk category identified in our cohort. Penetrance is the probability of developing the disease in the mutation carriers. Here, the penetrance was calculated separately for MTC and pheochromocytoma, that is, the probability of developing MTC or pheochromocytoma with different risk category mutations in terms of cumulative incidence.

For penetrance calculation, 39 mutation carrier probands, 16 mutation carrier family members and 4 obligate carriers were included (n=59). Among these 59 mutation carrier individuals, 9 belonged to 'highest risk' category mutations and presented MEN2B phenotype, 20 were 'high risk' category mutation carriers with MEN2A phenotype & 30 belonged to 'moderate risk' category mutation and produced FMTC phenotype (Figure 3.7).

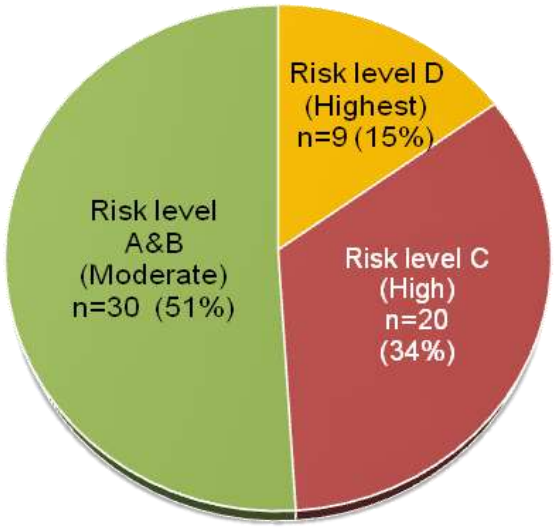


Figure 3.7: Frequency of different risk level mutations. Risk levels indicates MTC penetrance and aggressiveness.

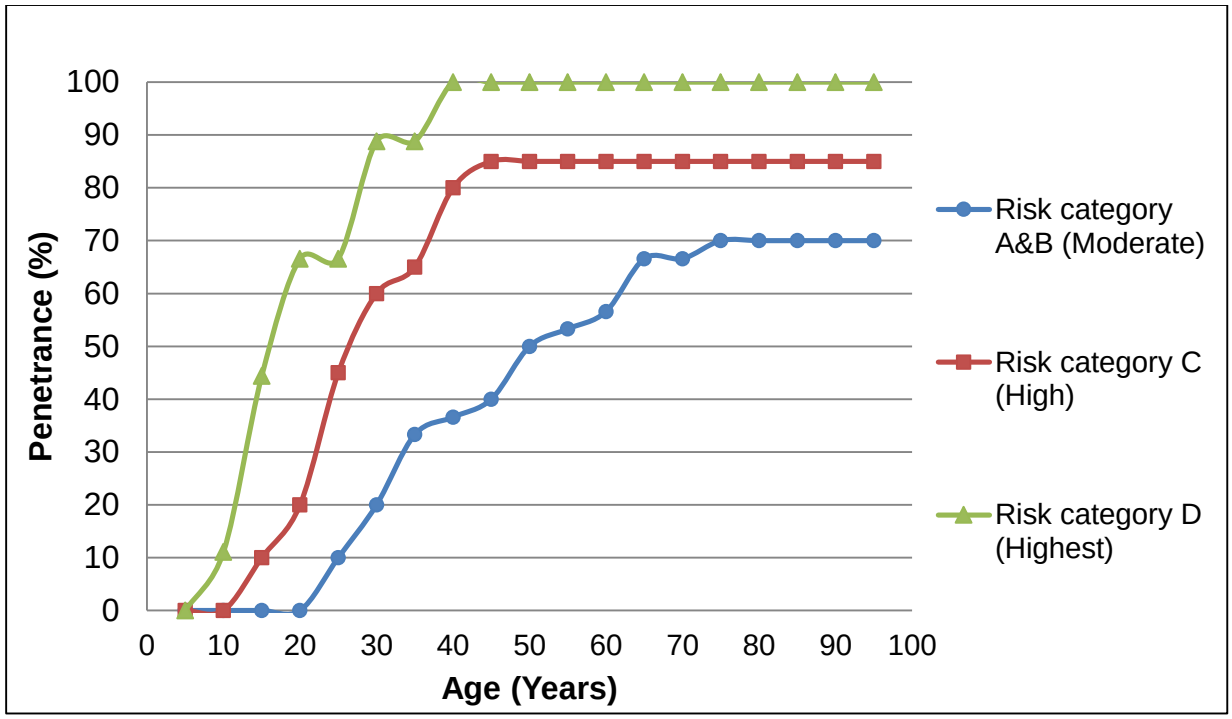


Figure 3.8: Cumulative incidence of MTC with different Risk category mutations.

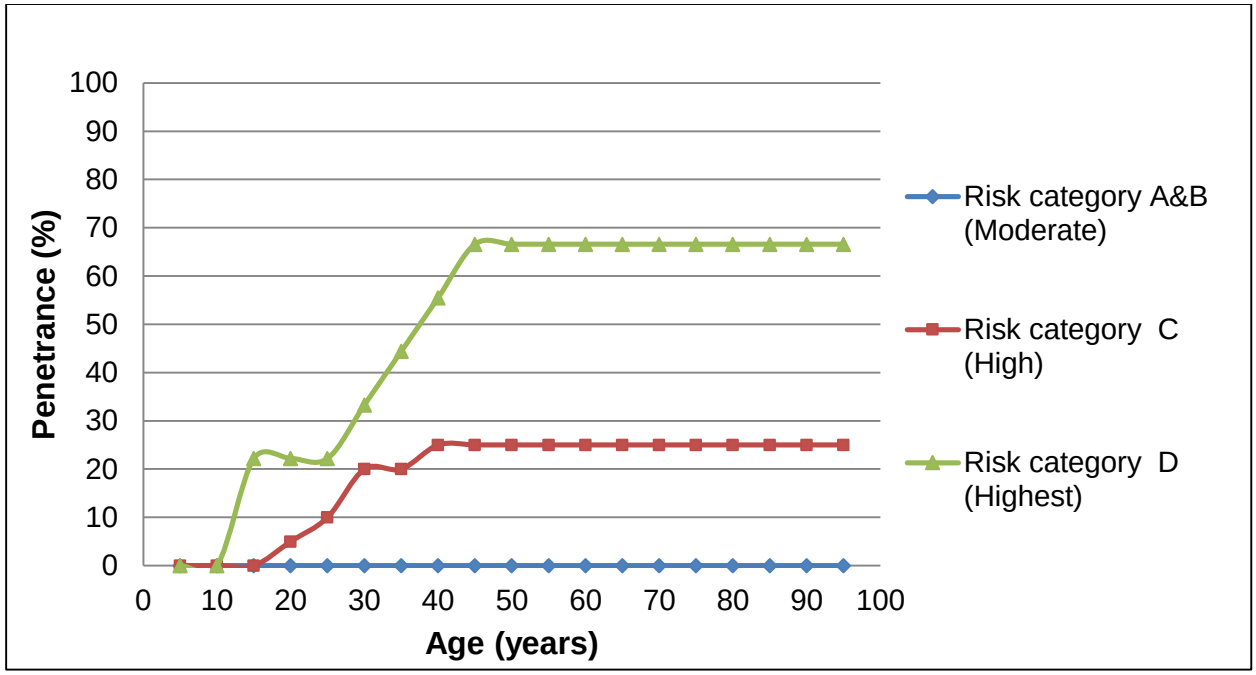


Figure 3.9: Cumulative incidence of Pheochromocytoma withdifferent risk category mutations

We observed that risk category D mutation carriers had complete penetrance (100%) for MTC by the age of 40 years and an early age of disease onset. These patients also had 65-70% lifetime risk of developing Pheochromocytoma. Risk Category C mutation carriers had 85% lifetime risk of MTC & 25% risk of developing Pheochromocytoma. Risk Category A&B mutation carriers had 70% lifetime risk of MTC with a late age of disease onset & these individuals were not at risk of developing pheochromocytoma (Figure 3.8 & Figure 3.9).

3.4 Discussion

MTC is an aggressive malignancy of thyroid and is curable only if it is diagnosed early and treated surgically when the disease is confined to the thyroid with or without limited regional nodal spread. Current systemic treatment including Receptor Tyrosine Kinase Inhibitors (RTKIs) such as sorafenib or cytotoxic chemotherapy does not produce long lasting disease control or cure. In the US Surveillance, Epidemiology and End Results registry (SEER) database, of the 793 MTC cases diagnosed between 1993-2002, the 10 year Disease specific survival was 96% for patients with MTC localized to the thyroid, 71% for patients with regional nodal spread and 26% in patients with distant spread(76).

Management of MEN2 syndrome patients using RET genetic testing represents an excellent exampleof how molecular DNA diagnosis have improved the clinical management of patients (70).The genotype-phenotype correlations between mutation, age of disease onset, its aggressiveness, and presence of pheochromocytoma or hyperparathyroidism are widely used to predict the phenotype of MTC and to make recommendations regarding the agesat which prophylactic thyroidectomy should be performed. This is particularly important in presymptomatic RET mutation carriers because a prophylactic thyroidectomy should be

performed either before the development of MTC, or when the disease is confined to the thyroid (71). In addition, tumor marker such as serum calcitonin levels for MTC and plasma metanephrine for pheochromocytoma helps guide the extent of surgery and detection of recurrence of the disease (76).

In the present study, in the cohort of 245 MTC cases, we identified 39 RET mutation carrier families and 55 mutation carrier individuals. The most frequently found mutation identified in our cohort was at codon 634, identified in 17/55 (31%) cases which is similar to other population(77–80). Early onset of the disease and occurrence of pheochromocytoma along with MTC makes the referral and diagnosis earlier explaining a high prevalence of this codon mutation across different population (79). We have also correlated the RET genotype with MEN2 phenotype and calculated age related penetrance for each risk category RET mutation. We observed that risk category D mutation carriers which present the most aggressive phenotype of MEN2B had complete penetrance for MTC and 65-70% lifetime risk of developing pheochromocytoma. Risk Category C mutation carriers that results in MEN2A phenotype and presents a high aggressiveness of the disease had 85% lifetime risk of developing MTC & 25% risk of developing Pheochromocytoma. Risk Category A and B mutation carriers with moderate level of disease aggressiveness had 70% lifetime risk of MTC with a late age of disease onset. Also, these patients were not at risk of developing pheochromocytoma and MTC was the only clinical manifestation in them.

Our study is the first Indian study to establish genotype-phenotype correlation in patients with MEN2 syndrome and also determined the age related penetrance of different risk category RET mutations which will be useful for genetic counseling and risk management of mutation carriers.

The only study on RET mutations in MTC cases from India described the mutation spectrum in 51 clinically diagnosed MTC cases but the genotype - phenotype correlation was not described. Specific RET mutations in patients with MEN 2 syndrome have been shown to predict the phenotype of the disease and its aggressiveness in different population(77–80). This provides prognostic information to guide prophylactic thyroidectomy and screen for pheochromocytoma. This is especially valuable for pre-symptomatic mutation carriers. The current approach for prevention of hereditary cancers to pass on to next generation involves prenatal genetic testing or pre-implantation genetic diagnosis and assisted reproductive techniques to prevent inheritance of the mutation to the future generations. However, a strong genotype-phenotype correlation for RET mutations and MEN2 phenotype and established guidelines for prophylactic interventions provides opportunity for early detection of the disease and its prevention.

Chapter 4

To identify additional molecular signatures in RET driven MTC tumor samples of using Whole Exome Sequencing

4.1 Background

As discussed in previous chapters, a strong genotype – phenotype correlation exists between location of RET mutations and MEN2 phenotype, based on which the American Thyroid Association (ATA) has categorized germline RET mutations in 4 risk levels based on the age of onset of MTC and its aggressiveness. ATA level D mutations (M918T and A883F) are associated with highest risk of MTC which occurs in early childhood and is very aggressive. ATA level C mutations are at codon 634 with substitution of cysteine by arginine or few other amino acids. Risk Level C mutations are associated with a high lifetime risk of aggressive MTC. ATA risk level B and A mutations are at codon 609, 611, 618, 620 and 630 of cysteine rich domain of RET or at codon 768, 790, 791, 804 and 891of the tyrosine kinase domain of RET. These mutations are associated with moderate to lowest lifetime risk of developing MTC(45,53). These strong correlations provide the basis for clinical management of the patients along with prophylactic interventions in healthy carriers.

However, the widely accepted genotype – phenotype correlation based on the ATA system cannot explain the phenotypic heterogeneity seen in individuals carrying identical mutations within a family or in different families. The most common RET mutation is at codon 634 and it shows considerable inter and intra-familial phenotypic variability in terms of age of disease onset and its aggressiveness. This suggests the modulatory role of additional somatic genetic alterations in other genes which are not well understood.

In the only study of comprehensive genomic characterization of MTC done so far, Agarwal et al in 2013(63) reported their findings of whole exome sequencing in 17 sporadic MTC tumors. They identified 305 high confidence somatic mutations in 283 genes. In this cohort of MTC

cases without a germline RET mutation, the most common somatic mutation was in RET and observed in 12/17 (71%) sporadic MTC tumors. Other than RET, the genes that were mutated in at least 2 of the 17 MTCs were MDC1, SF3B1, MGAM, DOCK9, SEMA6A, TDG & DISP2. To evaluate the frequency of these mutations as well as some other genes that were previously known to be implicated in MTC tumorigenesis such as HRAS, KRAS (81,82) they analyzed these gene mutations in an independent set of 40 MTC cases (19 sporadic & 21 hereditary). Somatic mutation in RET, HRAS, KRAS were identified in 78%, 18%, 5% of cases and these mutations were mutually exclusive. Overall, in the discoveryand validation sets, 91% of MTC tumors harbored mutations in *RET*, *HRAS*, or *KRAS*. This study though small, confirmed the role of RET as the driver mutation not only in hereditary but also as a somatic mutation in sporadic MTC and that of RAS in RET negative tumors. However,inclusion of only sporadic tumors in the discovery set may have precluded identification of other important genes and pathways involved in germline RET driven MTC development & how they influence phenotypic heterogeneity.

It is noteworthy that the mutational landscape of hereditary MTC with a known germline driver RET mutation has not been studied so far except for validation of somatic genetic alterations identified in sporadic MTC by Agrawal et al (2013) (63). Therefore, to study the genomic landscape of germline RET driven hereditary MTC cases, whole exome sequencing was performed in a cohort of 15 MTC tumor tissues with paired blood sample. The aim was to identify additional somatic genetic alterations that might govern the phenotypic heterogeneity observed in MTC patients with the same RET mutations.

4.2 Methodology

The study was conducted on 15 fresh frozen germline RET driven MTC tumor tissues and paired blood samples collected under institutional ethics committee approved study with written informed consent. Germline RET mutation was identified in the DNA extracted from the blood sample by PCR-Sanger Sequencing in each of these 15 individuals. The samples represented tumors arising from all the major classes of RET mutations – Risk Category B (C609R); Category C (C634R, C634W, C634Y) and Category D (M918T) (Table 4.1).

Table 4.1: Details of MTC tumor tissues included for Whole Exome Sequencing

Sr no.	Mutation	*ATA Risk category	No. of paired samples (Families)
1	M918T	D (Highest)	3 (3 families)
2	C634W	C (High)	2 (1 family)
3	C634R	C (High)	8 (3 families)
4	C634Y	C (High)	1 (1 family)
5	C609R	B (Moderate)	1 (1 family)

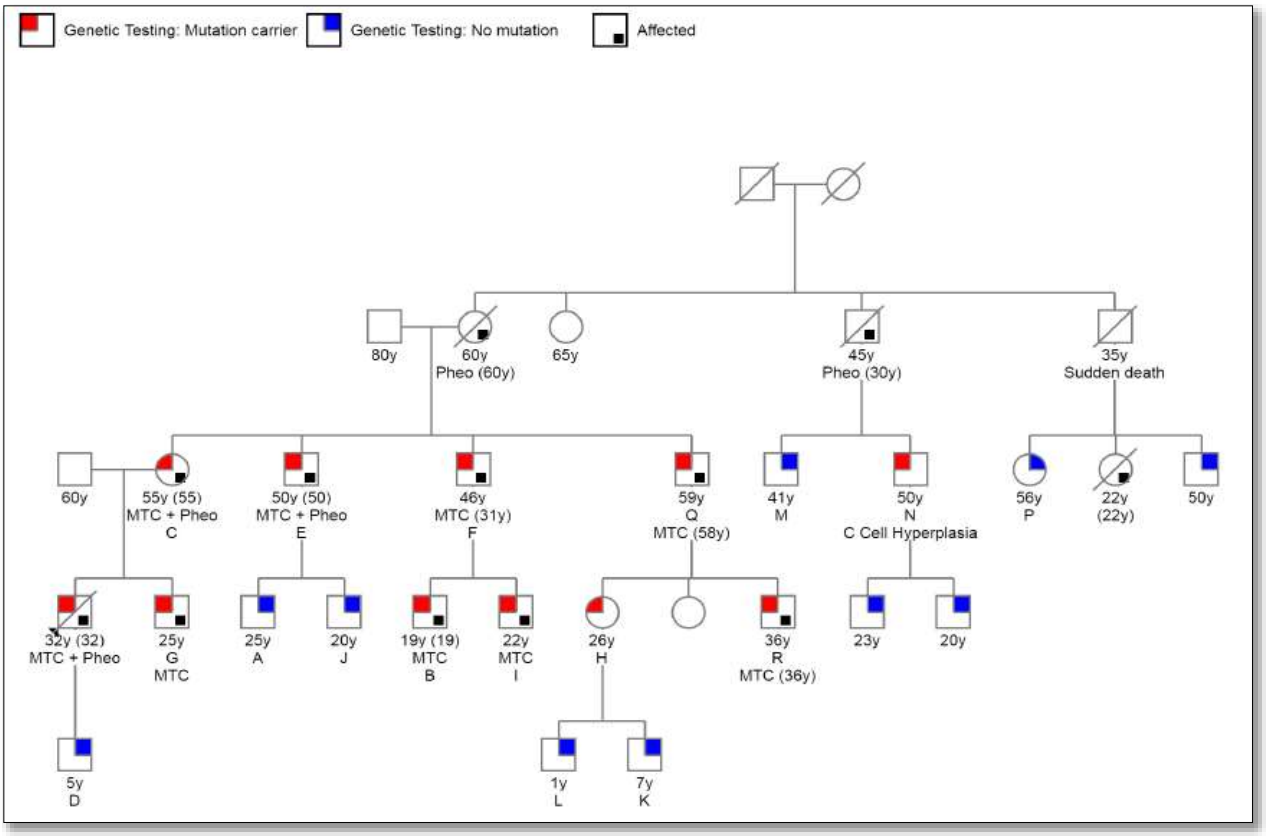
Whole exome sequencing was performed at Genotypic pvt. Ltd.where exome sequencing library was prepared using Agilent SureSelect v6 exome capture kit which enriches 50MB protein coding region from human genome corresponding to 21,000 genes. 150x2 bp paired end sequencing was performed on Illumina HiSeq 2500 genome sequencer and data was analyzed using Varscan 2.0 pipeline as described inMaterials & Methods chapter.

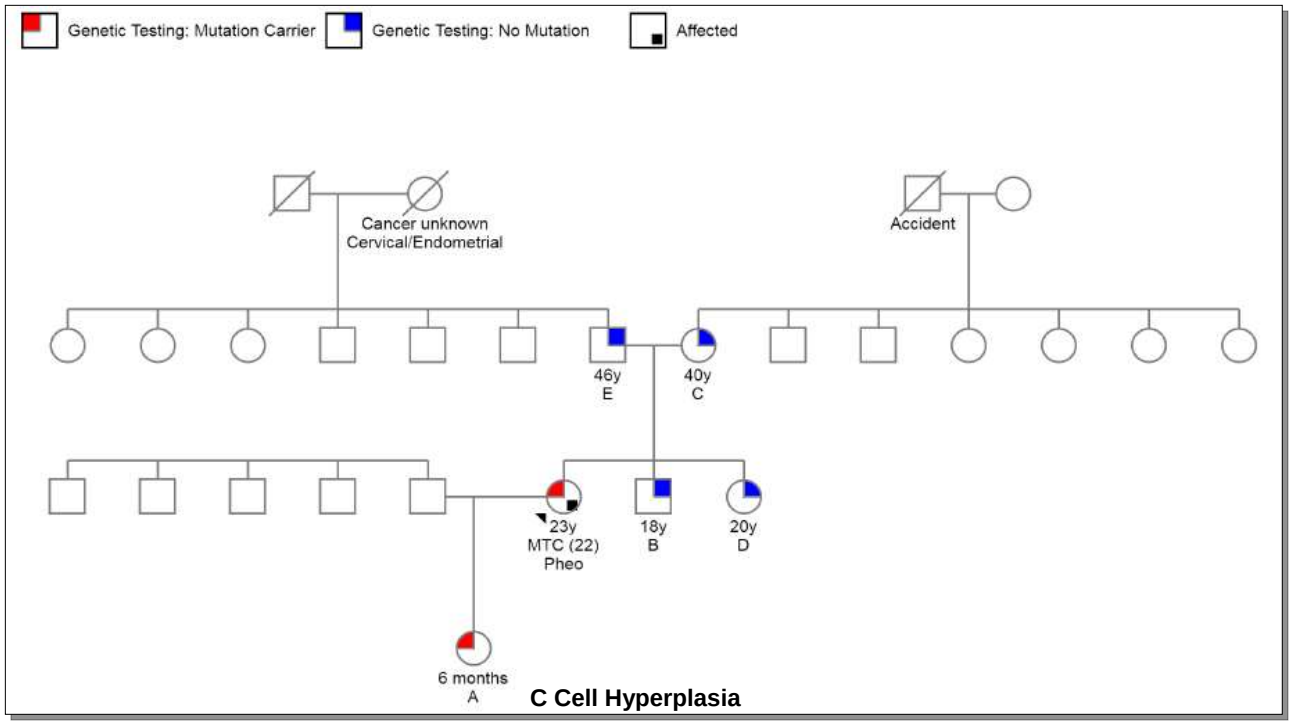
4.3 Results

Clinical details of the 15 MTC cases taken for whole exome sequencing are given in Table 4.2. and a few representative pedigree of the MTC cases included in the study are also shown below (Figure 4.1). MTC progression starts from C-Cell Hyperplasia (CCH), followed by microscopic foci of MTC, followed by localized MTC which is confined to thyroid which then metastasizes to distant organs of the body (metastatic MTC).

Table 4.2: Clinical details of MTC cases taken for whole exome sequencing (n=15)

Sr. No. (Cases)	RET Mutation	ATA Risk Category	Age at MTC diagnosis	Syndrome associated features
1	C634W	C	22y	MTC + Pheo
2	C634W	C	6 months	CCH, No Pheo
3	M918T	D	21y	Metastatic MTC, Pheo
4	M918T	D	15y	Metastatic MTC, No Pheo
5	M918T	D	11y	MTC, No Mets, No Pheo
6	C634Y	C	51y	MTC, No Pheo
7	C634R	C	27y	Metastatic MTC, b/l pheo
8	C634R	C	10y	CCH, No Pheo
9	C609R	B	9y	CCH, No Pheo
10	C634R	C	55y	MTC + Pheo
11	C634R	C	33y	Multiple foci of MTC, No Pheo
12	C634R	C	22y	Microscopic foci of MTC, No Pheo
13	C634R	C	50y	CCH, No Pheo
14	C634R	C	58y	MTC, No Pheo
15	C634R	C	36y	MTC, No Pheo





C Cell Hyperplasia

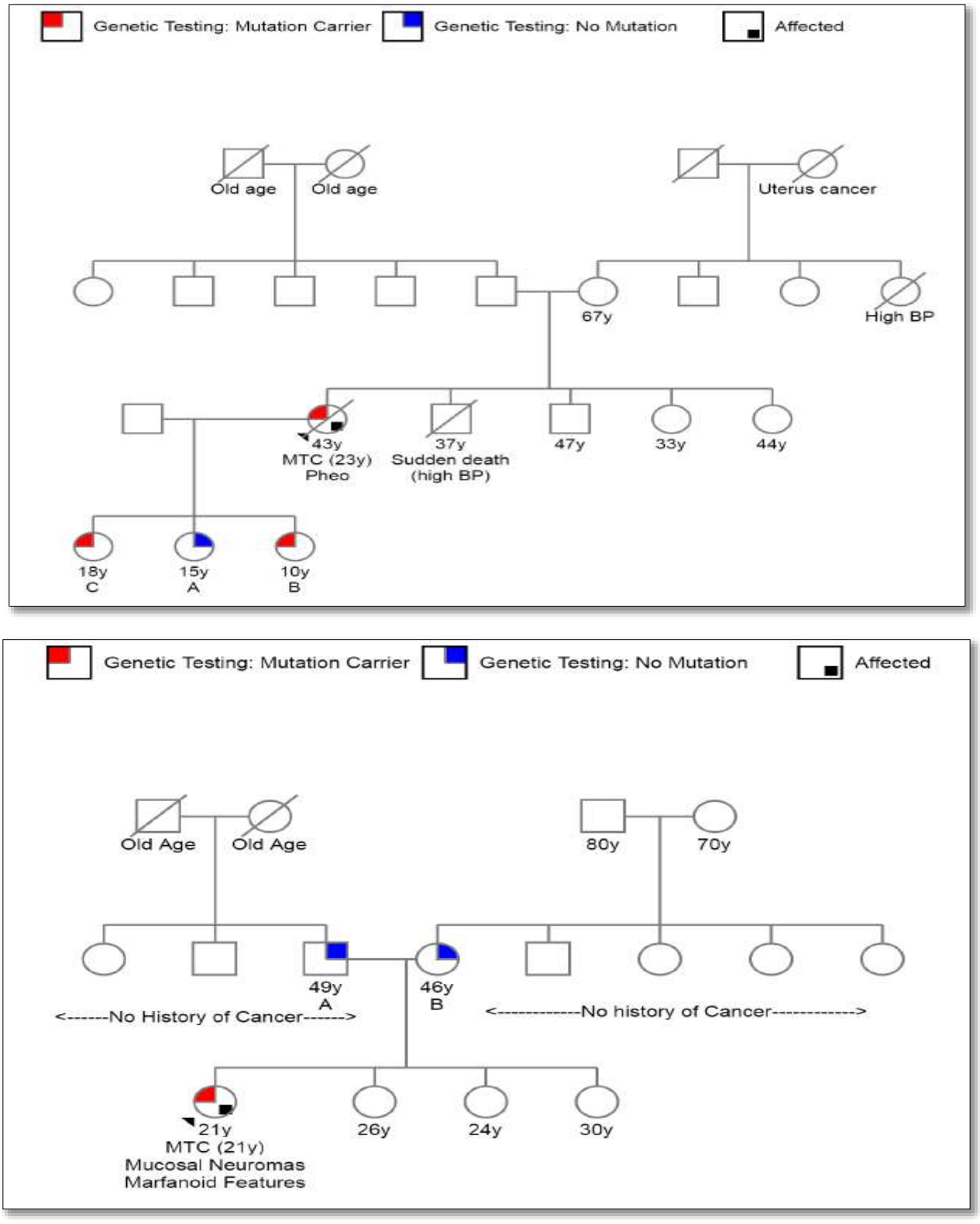


Figure 4.1: Representative pedigree of RET mutation carrier families

4.3.1 Exome Sequencing data Quality Check

The exome sequencing data (150x2 bp paired end sequencing) had first undergone initial quality check before proceeding for further steps of data analysis. For this the de-multiplexed Fastq files were uploaded in Fastqc software tool and the sequencing quality was assessed (Figure 4.2). FastQC software assess the quality of each base based on Phred score which predicts the probability of an error in base calling.The below figure 4.2 shows the quality values across all bases at each position in the FastQ file.

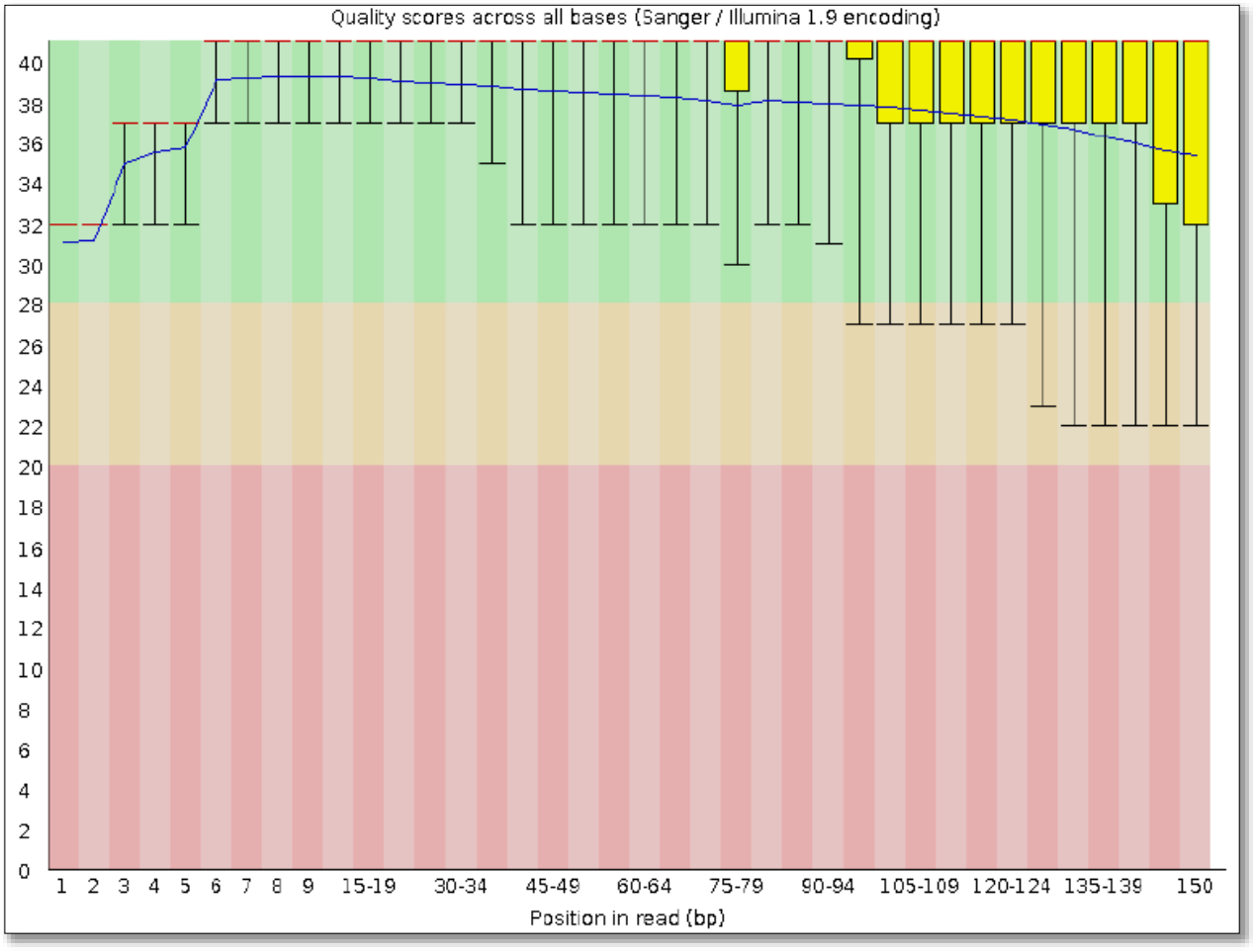


Figure 4.2: FASTQC analysis of the raw data (Fastq) showing the quality scores across all the 150 bases sequenced.

In the ‘per base sequencing quality graph’ shown above, for each position a Box & Whisker plot is drawn. The elements of the plot are as follows: The central red line is the median value. The yellow box represents the inter-quartile range (25-75%). The upper and lower whiskers represent the 10% and 90% points and the blue line represents the mean sequence quality. The y-axis on the graph shows the quality scores. The higher the score the better the quality of the base call at that particular position. The background of the graph divides the y-axis into very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red) region. The mean quality score should preferably be in green region to be called as good quality reads to be used for further steps of data analysis. The detailed steps of the further steps of data analysis using VarScan 2.0 pipeline is discussed in Chapter 2: Materials and Methods.

4.3.2 Coverage Statistics

The initial quality check passed reads were aligned with the reference human genome hg19. In tumor samples, 99.7% reads were uniquely mapped with the reference genomewith the mean mapping quality of 57.6. The average target coverage was 80X. For paired normal samples, 99.6% reads were uniquely mapped with the reference genome with the mean mapping quality of 57.6. The average target coverage was 78X. Detailed description of the exome coverage statistics are given in Table 4.3 for all the tumor and in Table 4.4 for the paired normal samples. More than 89% of the bases were covered with a minimum of 20 reads (20X coverage) in tumor samples and more than 88% of the bases were covered with a minimum of 20 reads (20X coverage) in paired normal samples.

Table 4.3: Coverage statistics for Exome sequencing of Tumor Samples

Tumor Sample	Mapped reads %	Mean Mapping Quality	Regions with 10X coverage	Regions with 20X coverage	Mean coverage (X)
1T	99.7	57.4	98.9	96.7	76.4
2T	99.6	57.4	99.1	97.4	78.3
3T	99.7	57.5	99.1	97.5	83.0
4T	99.7	57.4	99.1	97.4	84.3
5T	99.7	57.4	99.2	97.3	81.7
6T	99.6	57.3	99.3	97.8	92.0
7T	99.8	57.4	99.1	97.5	89.5
8T	99.7	57.4	99.0	97.3	95.1
9T	99.7	57.3	99.2	97.4	89.2
10T	99.7	57.4	99.3	97.7	88.5
11T	99.8	57.5	99.1	97.3	80.6
12T	99.7	57.4	99.2	97.4	85.2
13T	99.9	58.6	57.4	53.2	55.9
14T	99.9	58.6	59.0	55.7	68.4
15T	99.9	58.5	56.5	51.5	47.2

Table 4.4: Coverage statistics for Exome sequencing of paired Normal Samples

Sample	Mapped reads %	Mean Mapping Quality	Regions with 10X coverage	Regions with 20X coverage	Mean coverage (X)
1N	99.7	57.4	99	97.4	85.5
2N	99.6	57.4	98.9	96.8	72.0
3N	99.6	57.4	99.1	97.5	82.2
4N	99.5	57.4	99.0	97.4	81.5
5N	99.6	57.3	99.2	97.4	80.0
6N	99.7	57.3	99.3	97.8	92.3
7N	99.7	57.4	99.1	97.4	83.0
8N	99.7	57.4	99.2	97.7	91.0
9N	99.7	57.3	99.3	97.7	86.0
10N	99.6	57.3	99.3	97.8	92.6
11N	99.6	57.3	99.2	97.6	87.0
12N	99.6	57.4	99.2	97.6	88.4
13N	99.9	58.6	57.6	53.4	53.0
14N	99.7	58.6	53.5	45.3	34.4
15N	99.9	58.5	57.2	53.0	55.7

4.3.3 Variant Statistics

A detailed summary of variant statistics is given in (Table 4.5). Overall we identified 351691 Indels and 2310904 single nucleotide variants (SNVs) using Varscan 2.0 variant calling tool. Of these, 5528 variants were called as high confidence somatic variants which included both Indels & Single Nucleotide Variants. For a variant to be called as high confidence somatic variant, it should have a minimum variant allele frequency of 0.02 in tumor and the maximum allele frequency of 0.05 in normal (explained in detail in Materials & Methods chapter). Further, a false-positive filter (FP filter) was applied to further filter the variants. In this command, a base quality of more than Q15 and above was required for each variant, significance should be less than 0.05 ($p<0.05$) and depth required to test somatic p value should be 8X (at least 8 reads should cover the region of the genome harboring the variant and the variant should be present in at least 4 reads). The FP filter also removes the variants that were present in vicinity of Indels as they could be false-positive variants. After applying false-positive (FP) filter , the number of high confidence somatic variants reduced to 390. Further, the insertions and deletions that were present in homo-polymer region were also removed as they could be a result of sequencing artifact. This was done by checking the filtered Indels individually in dbSNP database using their rsID or by visualizing them in Integrated Genome viewer using their genomic coordinates. In the next step, the intronic variants were removed and then synonymous exonic variants were also filtered out. This resulted in a total of 93 significant somatic exonic variants across all the 15 MTC samples (Table 4.5 & Table 4.6) and included missense, nonsense, splice-site, In-frame & Frameshift variants (Table 4.6)

Table 4.5: Whole Exome Sequencing variant statistics across all 15 samples

Sample	Total Indels	Total SNVs	High Confidence filtered somatic variants (Indels &SNVs)	Somatic variants after removing False Positive	Somatic Variants after removing Intronic variants	Somatic Variants (Synonymous removed)
1	24059	152949	320	16	6	4
2	24522	155450	329	24	3	2
3	25091	150883	270	12	6	3
4	23619	148628	317	30	14	9
5	23983	153299	354	27	8	5
6	21755	156646	535	44	19	15
7	22153	147232	317	26	5	4
8	23067	150203	351	13	11	8
9	22363	149147	321	15	5	3
10	23207	168399	550	65	30	21
11	24061	154860	342	18	6	5
12	25542	157650	371	19	5	3
13	23293	148371	291	20	6	4
14	21937	167970	527	43	7	5
15	23039	149217	333	18	5	2
Total	351691	2310904	5528	390	136	**93**

Table 4.6: Significant somatic variants identified in our cohort

Samples	Somatic Variants (Synonymous removed)	Missense	Nonsense	Splicesite	InFrame	Frameshift
1	4	2	0	1	0	1
2	2	1	0	1	0	0
3	3	2	1	0	0	0
4	9	9	0	0	0	0
5	5	4	0	0	1	0
6	15	10	2	3	0	0
7	4	4	0	0	0	0
8	8	2	1	0	4	1
9	3	2	0	1	0	0
10	21	17	2	0	2	0
11	5	4	0	1	0	0
12	3	3	0	0	0	0
13	4	1	1	1	1	0
14	5	5	0	0	0	0
15	2	0	1	1	0	0
Total	**93**	66	8	9	8	2

4.3.4 Tumor Mutation Burden/Mb

Tumor mutation burden per Mb (TMB/Mb) was calculated from the 93 high confidence somatic variants identified across the 15 samples and is shown in Figure 4.3. The average TMB/Mb across all samples was **0.124/Mb**. The TMB in hereditary MTC cases in our cohort was compared with the TMB of sporadic MTC reported by Agarwal et al. in 2013 and papillary thyroid carcinoma by TCGA (63,83). While TMB in thyroid cancers is in general very low with 0.35 in sporadic MTC and 0.41 in papillary thyroid carcinoma, hereditary MTC has the lowest TMB (Figure 4.4).

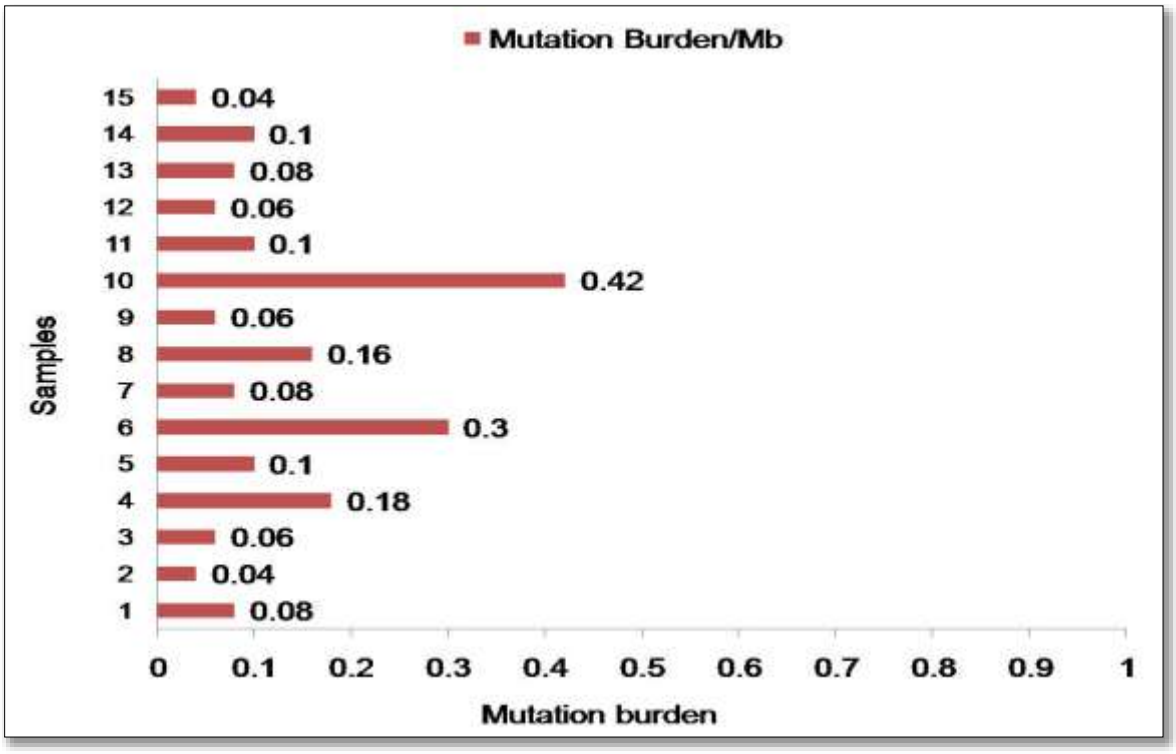
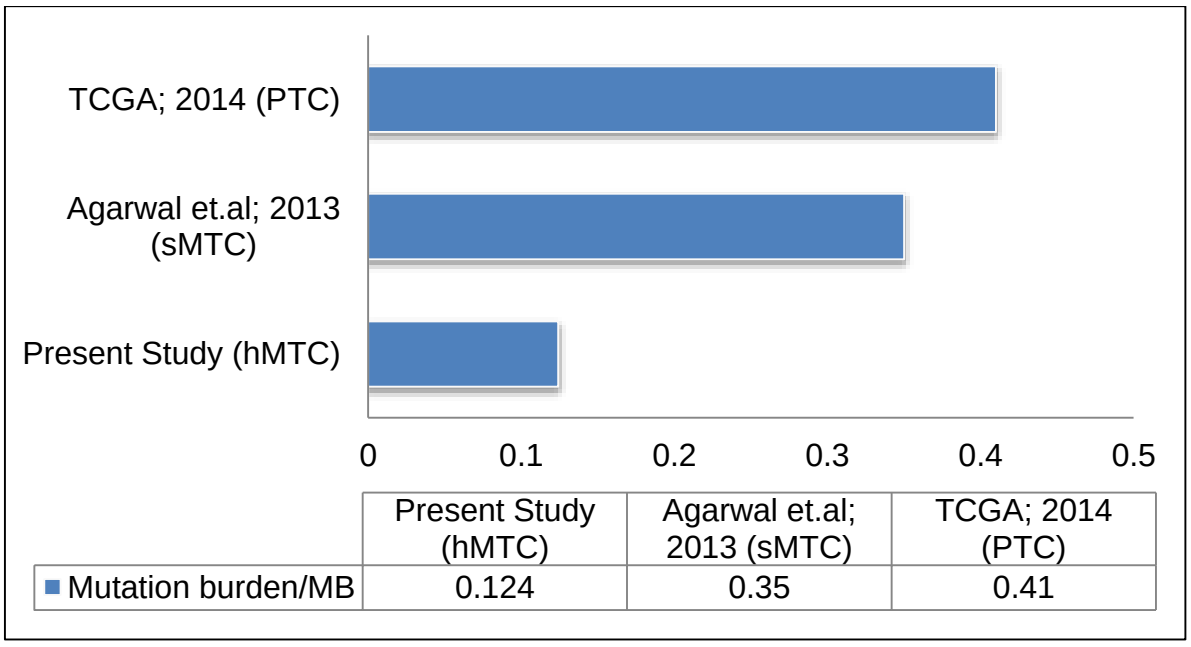


Figure 4.3: Tumor Mutation burden/Mb in all 15 samples (high confidence variants include - Missense, Nonsense, Splice-site, Frameshift & Inframe mutations)



	Present Study (hMTC)	Agarwal et.al; 2013 (sMTC)	TCGA; 2014 (PTC)
Mutation burden/MB	0.124	0.35	0.41

Figure 4.4: Tumor Mutation burden/Mb across different thyroid cancer

4.3.5 Somatic mutation spectrum by nucleotide change (Transition/Transversion)

Transitions (Ts's) are point mutations where a purine is substituted for a purine or a pyrimidine for a pyrimidine whereas transversions (Tv's) are the point mutations where a purine is substituted for a pyrimidine and vice versa.The generation of transversions requires much greater distortion of the double helix than transition mutations. Since Tv's cause larger changes in the shape of a DNA backbone, they have larger impacts on binding of Transcription Factors (TFs) or other DNA binding proteins affecting gene expression(84). The somatic mutation spectrum in all the15 samples in terms of nucleotide change (transitions/transversions) is shown in Figure 4.5. In a few samples (sample 5, 6, 7 & 10) we observed a high frequency of transversions compared to other samples.

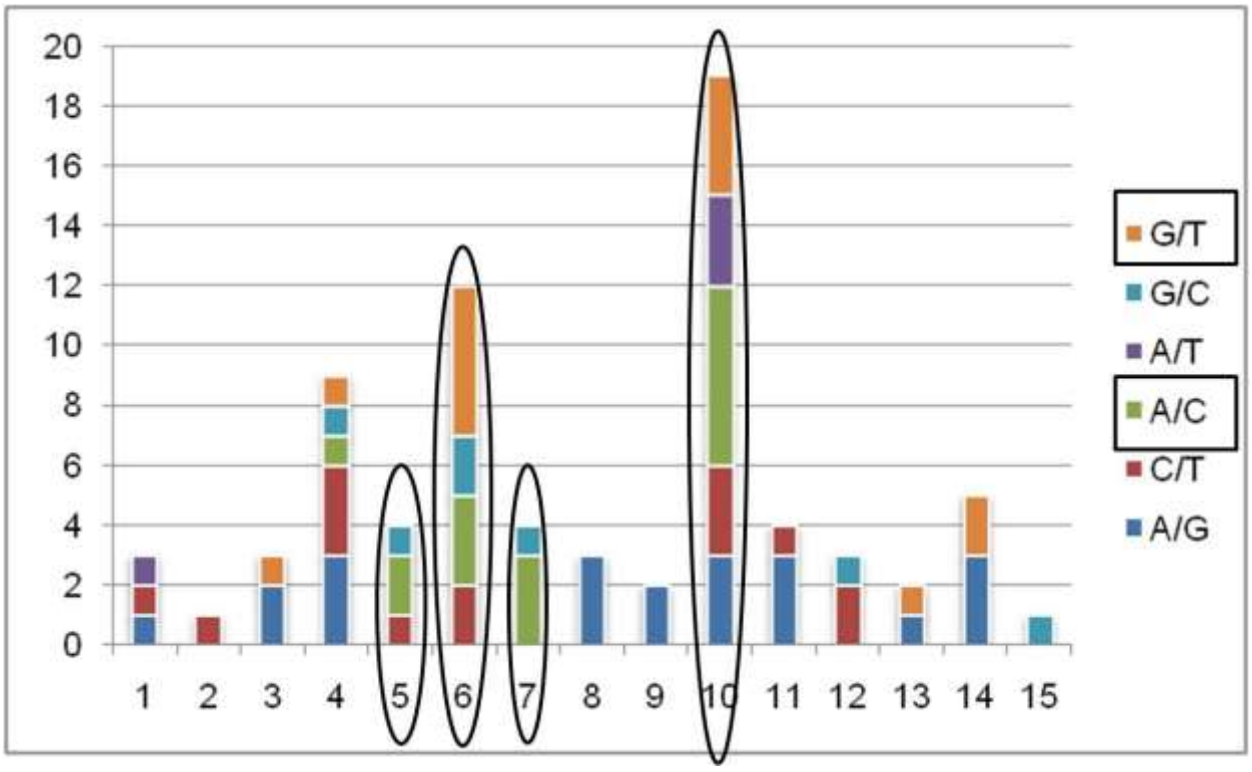


Figure 4.5: Somatic mutation spectrum by nucleotide change: Transition/Transversion in all 15 samples [A/G & C/T are transition mutations & A/C, A/T, G/C & G/T are transversions]

We further compared the Transitions/Transversions frequency in Hereditary MTC cases (our Cohort) & Sporadic MTC cases studied by Agarwal et al., group in 2013 (discussed in the 'Background' section of this chapter)(63). We could not identify any significant difference in the transition/transversion ratio between hereditary and sporadic MTC groups (p=0.57) (Figure 4.6, 4.7 & 4.8)

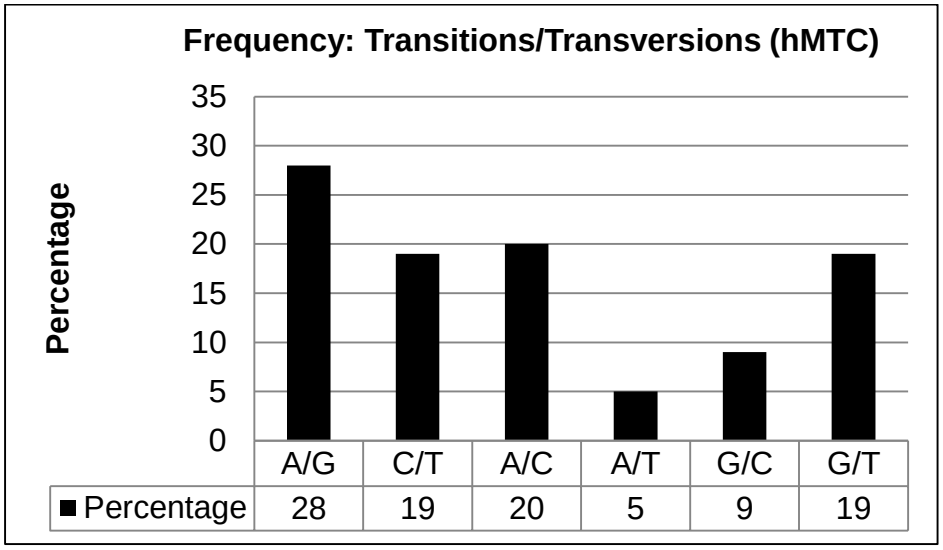


Figure 4.6: Overall Transition/Transversion frequency in hereditary MTC cohort (n=15)

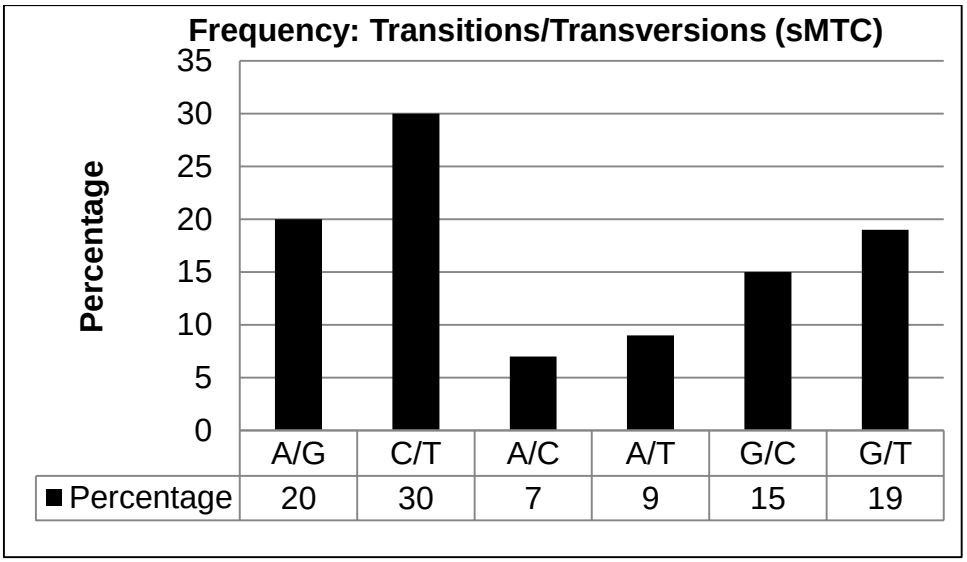


Figure 4.7: Overall Transition/Transverion frequency in sporadic MTC cohort (Agarwal et al 2013; n=17)

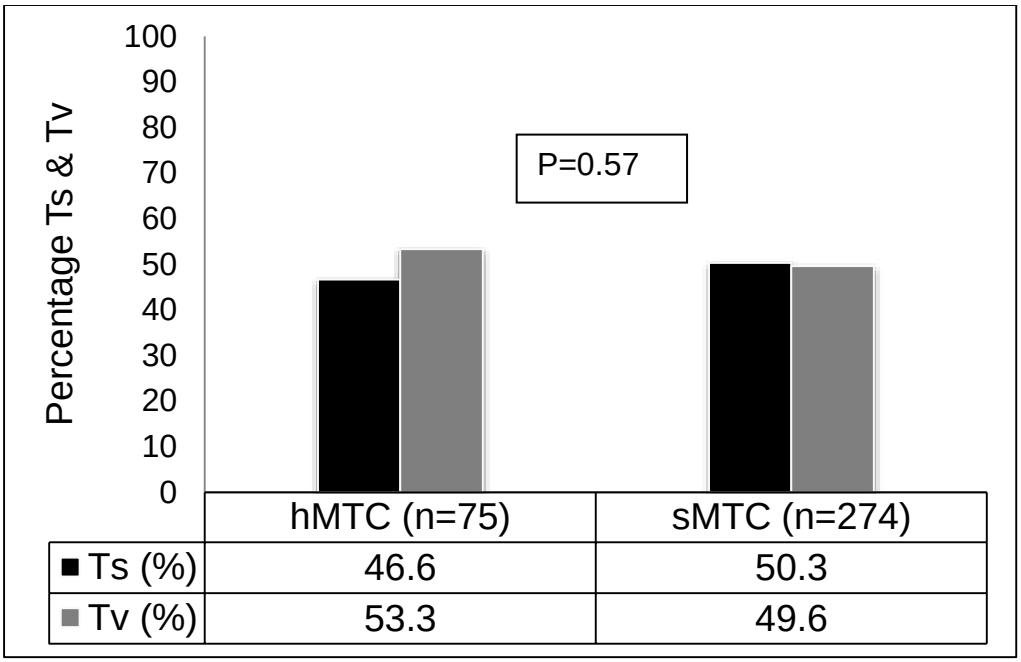


Figure 4.8: Comparison of Transition/Transversion frequency in hereditary & sporadic MTC cases

4.3.6 Number of significant somatic mutations identified in our cohort

Overall 93 significant somatic mutations in 15 samples were identified. The number of significant somatic mutation per tumor sample is shown in Figure 4.9. The average number of significant mutations per sample was 6.2 (range from 2 to 21 and SD $\pm$5.29). The mutations include: 66 missense mutations, 8 nonsense mutations, 2 frameshift mutations, 8 In-frame mutations and 9 splice-site mutations. Out of the 66 missense mutations, 18 mutations were predicted to be pathogenic by 5 or more In-Silico pathogenicity prediction tools (Table 4.7). All of the identified somatic mutations were private events and none of them was found to be recurrent in any of the 15 samples studied. The list of the potential pathogenic variants identified in each sample is given in Table 4.7. Further, the physiological function of each gene harboring

these pathogenic variants is given in Table 4.8. Although none of the somatic pathogenic variant was found to be recurrent among samples, we have identified mutations in a few genes that are known to have a role in different cancer types (Table 4.7 & Table 4.8).

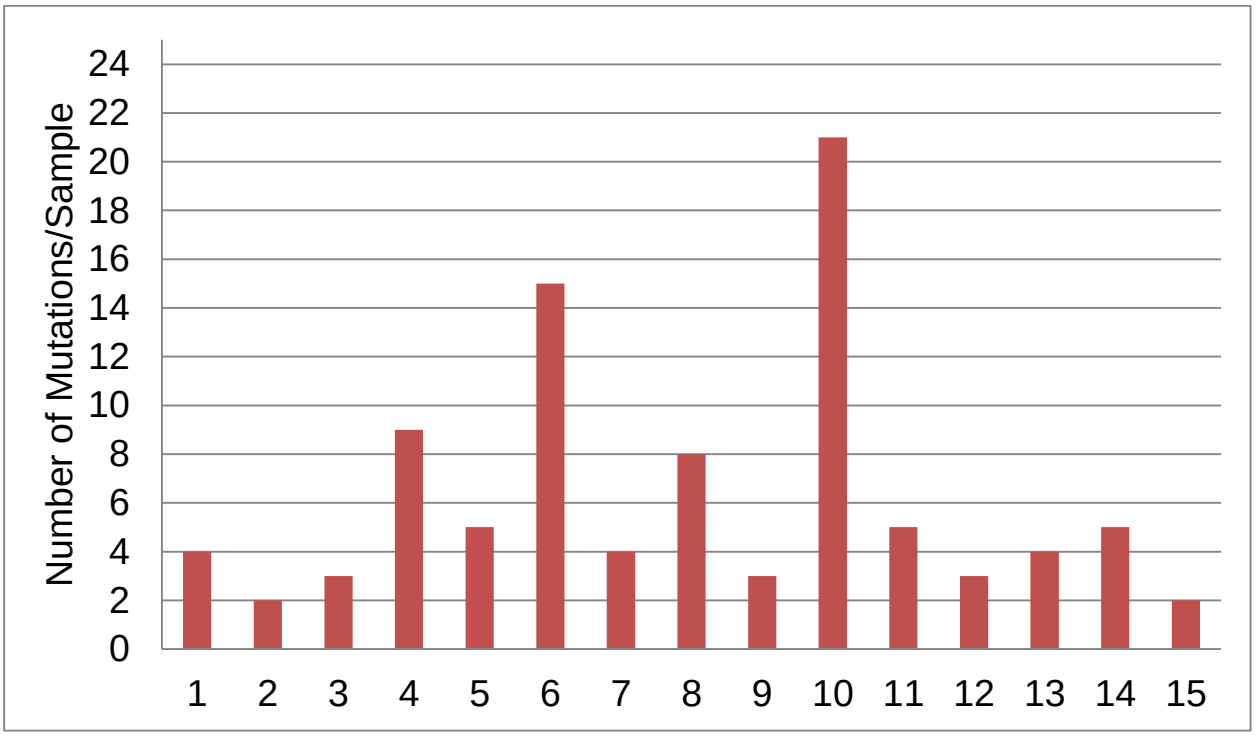


Figure 4.9: Number of significant somatic variants identified per sample in our cohort

Table 4.7: Significantly mutated genes & variants identified in all samples

Sr. No.	Gene	cDNA change	Amino acid change	Type of mutation
1	NUBPL	c.184G>T	G62C	Missense
2	IGDCC4	c.764T>A	V255E	Missense
3	RHOBTB3	c.1643T>C	L548P	Missense
4	NF1	c.814A>T	I272F	Missense
5	ADSL	c.1267T>A	L423I	Missense
6	PTGER3	c.400G>T	G134W	Missense
7	NNMT	641T>C	L214P	Missense
8	NPFFR1	c.353T>G	M118R	Missense
9	GCNT2	c.1148G>A	R383H	Missense
10	ADRA2A	c.479G>A	R160H	Missense
11	LAMB1	c.4643G>A	R1548H	Missense
12	GALNT12	c.G918G>T	R306S	Missense
13	PCED1B	c.1223C>T	P408L	Missense
14	ANO5	c.1240G>C	V414L	Missense
15	SPEG	c.4366C>G	R1456G	Missense
16	PLA2G4E	c.1865G>T	C622F	Missense
17	GPD2	c.1081G>T	V361F	Missense
18	PTEN	c.256G>T	A86S	Missense
19	PRDM9	c.1032C>A	C344X	Nonsense
20	USP47	c.1060G>T	E354X	Nonsense
21	BRCA2	c.8494G>T	E2832X	Nonsense
22	ESCO1	c.16G>T	E6X	Nonsense
23	TMEM41A	c.175C>T	R59X	Nonsense
24	USP17L20	c.984G>A	W328X	Nonsense
25	PRSS3	c.638DelT	C213fs	Frameshift
27	COL4A2-AS2	c.757DelC	D253fs	Frameshift
28	ATXN1	c.633DelTGA	H211Del	Inframe
29	HOXA1	c.216DelCGA	H74Del	Inframe
30	MICALCL	c.1408DelCTA	T471Del	Inframe

32	RP1L1	c.3925DelTCC	G1309Del	Inframe
33	RP1L1	c.4027DelTTC	E1343Del	Inframe
34	SUPT20HL1	c.1564DelTAG	L548Del	Inframe
35	TRIOBP	c.1190DelTCA	Q398Del	Inframe
36	ATXN3	..	…	Splice-site
37	CDO1	..	…	Splice-site
38	CEP104	..	…	Splice-site
39	FAM120A	..	…	Splice-site
40	HFM1	..	…	Splice-site
41	PRPF40A	..	…	Splice-site
42	TMED10	..	…	Splice-site
43	UBE2Q1	..	…	Splice-site
44	UVRAG	..	…	Splice-site

Table 4.8: Biological function of the significantly mutated genes identified in our cohort

Gene	Function
NUBPL	This gene encodes a member of the Mrp/NBP35 ATP-binding proteins family. Required for the assembly of the mitochondrial membrane respiratory chain NADH dehydrogenase (Complex I). Clinical disorders have variable severity, ranging from lethal neonatal disease to adult-onset neurodegenerative disorders
IGDCC4	Proteins of the immunoglobulin (Ig) superfamily, play essential roles in biologic functions of the cell surface, including cell adhesion, cell migration, and signal transduction. Identified as an oncofetal marker of murine and human hepatocellular carcinoma
RHOBTB3	Members of the RHOBTB subfamily of Rho GTPases and are evolutionarily conserved. Involved in transport vesicle docking at the Golgi complex, possibly by participating in release M6PRBP1/TIP47 from vesicles to permit their efficient docking and fusion at the Golgi.
NF1	NF1 gene encodes neurofibromin protein, a tumor suppressor protein. This protein prevents cell overgrowth by turning off another protein (ras) that stimulates cell growth and division
ADSL	The ADSL gene provides instructions for making an enzyme called adenylosuccinate lyase involved in purine synthesis. Mutations in this gene causes brain dysfunction (encephalopathy) that leads to delayed development of mental and motor skills (psychomotor delay), autistic behaviors and seizures
PTGER3	The protein encoded by this gene is a member of the G-protein coupled receptor

	family. This protein is one of four receptors identified for prostaglandin E2 (PGE2). This receptor may have many biological functions, which involve digestion, nervous system, kidney reabsorption, and uterine contraction activities
NNMT	N-methylation is one method by which drug and other xenobiotic compounds are metabolized by the liver. This gene (Nicotinamide N-Methyltransferase) encodes the protein responsible for this enzymatic activity which uses S-adenosyl methionine as the methyl donor.
NPFFR1	Neuropeptide FF receptor 1 (NPFFR1) is the receptor for NPAF (A-18-F-amide) and NPFF (F-8-F-amide) neuropeptides, also known as morphine-modulating peptides. This receptor mediates its action by association with G proteins that activate a phosphatidylinositol-calcium second messenger system.
ADRA2A	Alpha-2-adrenergic receptors, members of G protein-coupled receptor superfamily. These receptors have a critical role in regulating neurotransmitter release from sympathetic nerves and from adrenergic neurons in the central nervous system
LAMB1	This gene encodes the beta chain isoform laminin, beta 1. Laminins, a family of extracellular matrix glycoproteins, are the major noncollagenous constituent of basement membranes. They have been implicated in a wide variety of biological processes including cell adhesion, differentiation, migration, signaling, neurite outgrowth and metastasis.
GALNT12	This gene encodes a member of a family of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases, which catalyze the transfer of N-acetylgalactosamine (GalNAc) from UDP-GalNAc to a serine or threonine residue on a polypeptide acceptor in the initial step of O-linked protein glycosylation.

	Mutations in this gene are associated with an increased susceptibility to colorectal cancer
PCED1B	This gene encodes a protein that belongs to the GDSL/SGNH-like acyl-esterase family. Members of this family are hydrolases that function in modification of biopolymers on the cell surface.
ANO5	The ANO5 gene provides instructions for making a protein called anoctamin-5. While the specific function of this protein is not well understood, it belongs to a family of proteins, called anoctamins, that act as chloride channels
SPEG	This gene encodes a protein with similarity to members of the myosin light chain kinase family. This protein family is required for myocyte cytoskeletal development. Along with the desmin gene, expression of this gene may be controlled by the desmin locus control region. Mutations in this gene are associated with centronuclear myopathy 5
GPD2	The protein encoded by this gene localizes to the inner mitochondrial membrane and catalyzes the conversion of glycerol-3-phosphate to dihydroxyacetone phosphate, using FAD as a cofactor.
DOCK10	This gene encodes a member of the dedicator of cytokinesis protein family. Members of this family are guanosine nucleotide exchange factors for Rho GTPases and defined by the presence of conserved DOCK-homology regions.
USP47	Ubiquitin carboxyl-terminal hydrolase 47. Ubiquitin-specific protease that specifically deubiquitinates mono ubiquitinated DNA polymerase beta (POLB), stabilizing POLB thereby playing a role in base-excision repair (BER). Acts as a regulator of cell growth and genome integrity. May also indirectly regulate

	CDC25A expression at a transcriptional level
BRCA2	Tumor suppressor gene involved in DNA damage repair pathway
PTEN	Tumor Suppressor gene. The PTEN enzyme is part of a chemical pathway that signals cells to stop dividing and triggers cells to self-destruct through a process called apoptosis. Evidence suggests that this enzyme also helps control cell movement (migration), the sticking (adhesion) of cells to surrounding tissues, and the formation of new blood vessels (angiogenesis)
PRSS3	Serine protease 3 (PRSS3) is an isoform of trypsinogen, and plays an important role in the development of many malignancies. PRSS3 acts as an oncogene in invasive ductal carcinoma of the breast development and progression
ATXN3	The ATXN3 gene provides instructions for making an enzyme called ataxin-3. Acts as a deubiquitylase and is involved in protein degradation.

4.3.7 Variant Visualization: Integrative Genome Viewer (IGV)

The high confidence variants were visualized in Integrative Genome Viewer (IGV) which is an open source genome visualization tool that is used to view the alignments of sequencing reads with reference genome. For a variant to be called as a true variant, the region of the genome where the variant is located should be covered with at least 8 reads and the variant should be present in at least 4 reads, else it will be called as a false-positive variant call. IGV helps in removing false-positive variants. Figure 4.10 shows variant visualization on IGV for missense variant, Inframe Deletion and Frameshift Insertion identified in our cohort.

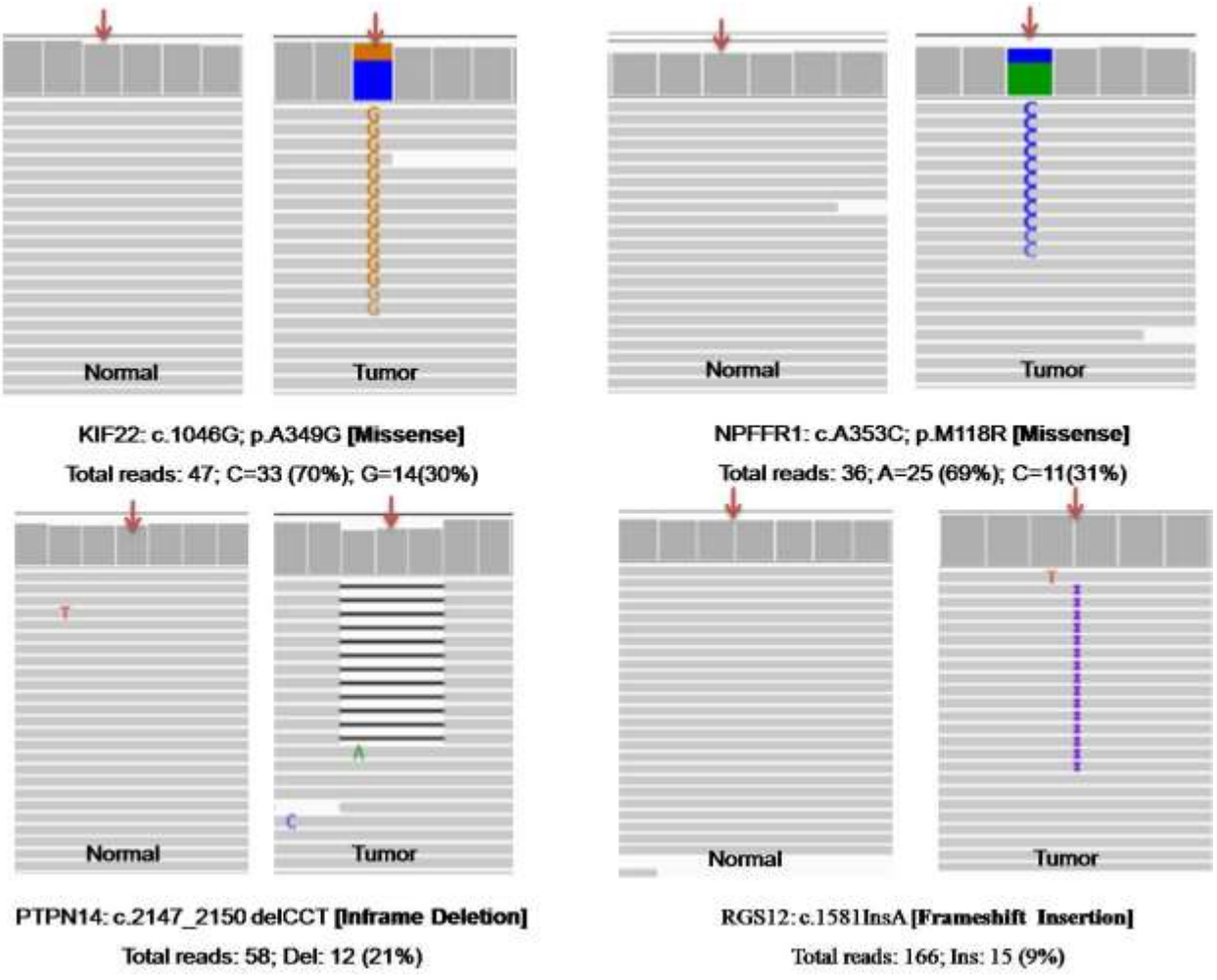


Figure 4.10: Variant visualization on Integrative Genome Viewer (IGV)

4.4 Discussion

The present study shows that in germline RET driven hereditary MTC tumors, other somatic mutations are rare, resulting in an extremely low somatic tumor mutation burden of 0.124 per Mb. Although the somatic tumor mutation burden in sporadic MTC and papillary thyroid cancer are also low (63,83), in hereditary MTC it is even lower (Figure 4.4). In a study published in Nature by Alexandrov et al in 2013 (85), the mutational signatures across different cancer types have been studied. They have shown that the prevalence of somatic mutations was highly variable between different cancer types ranging from 0.001mutations/Mb to more than 400 mutations/Mb (Figure 4.11). The lowest somatic mutation burden was observed in pilocytic astrocytoma which is a benign, slow growing brain tumor that most commonly occurs in children or young adults. The highest mutation burden was seen in lung squamous cell carcinoma and melanoma (Figure 4.11), both arising from chronic mutagenic exposures such as tobacco smoking for lung carcinoma and exposure to ultraviolet radiation for malignant melanoma (85). In another study published in Nature in 2014 by Lawrence et al; somatic mutations in 4,742 tumor-normal pairs across 21 cancer types was studied by exome sequencing(86).They have also shown that the highest somatic mutation burden was found in lung squamous cell carcinoma (9.9/Mb) and Melanoma (12.9/Mb) (Figure 4.12). Although they have not included pilocytic astrocytoma in their study group, they have observed the lowest somatic mutation burden of 0.1/Mb in another rare childhood brain tumour– Rhabdoid tumour (Figure 4.12).

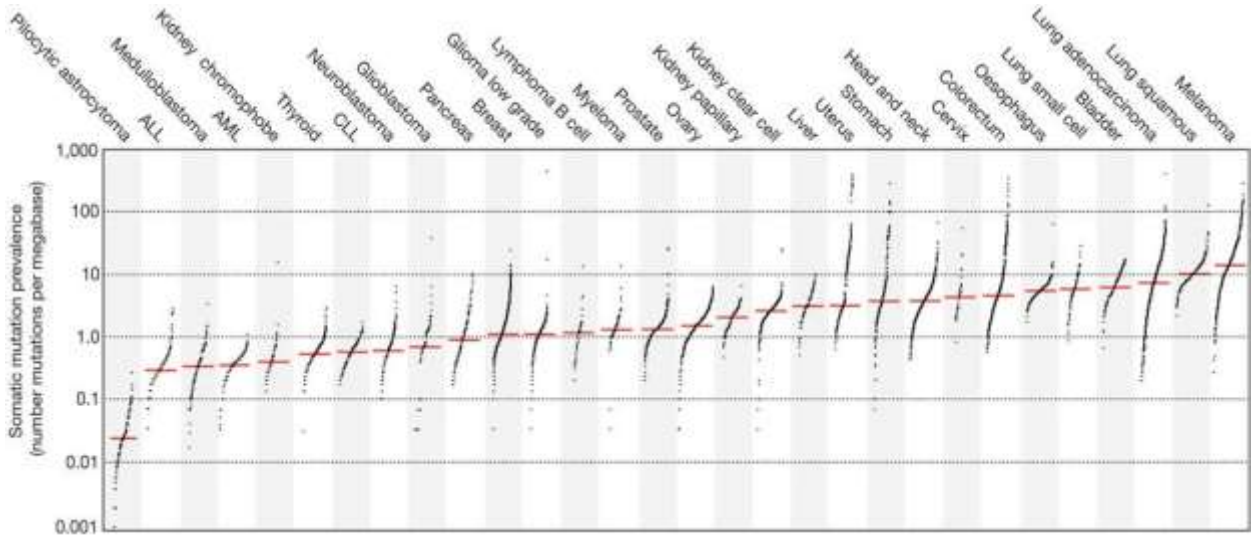


Figure 4.11: The prevalence of somatic mutations across human cancer types (Figure taken from Ref. 85; Alexandrov et al; 2013; Nature).

It is noteworthy that both these rare childhood brain tumors- pilocytic astrocytoma and rhabdoid tumor are single gene or single pathway driven cancers similar to hereditary medullary thyroid carcinoma which is driven by germline mutations in RET proto-gene. Almost all malignant rhabdoid tumors present with a loss of function in the SMARCB1 gene(87). SMARCB1, a tumor suppressor gene is a member of the SWI/SNF chromatin-remodeling complex(88). Germline mutations of the SMARCB1 gene predisposes individuals torhabdoid tumor predisposition syndrome(89). Similarly for pilocytic astrocytoma, it was shown that single abnormalities of the mitogen-activating protein kinase (MAPK) pathway including mutations in BRAF, N-RAS, RAF-1 and NF1 are exclusively found in almost all the cases of pilocytic astrocytoma, indicating that it represents a one-pathway disease(90).

Table 4.9: Median somatic mutation burden (per Mb) across 21 different cancer types (Taken from Ref. 86)

Tumor type	No. of tumor–normal pairs	Median somatic mutation frequency (per Mb)	No. of significantly mutated genes	No. of additional significant genes found under RHT
Acute myeloid leukaemia	196	0.4	26	1
Bladder	99	7.1	24	10
Breast	892	1.2	32	5
Carcinoid	54	0.5	1	0
Chronic lymphocytic leukaemia	159	0.6	7	8
Colorectal	233	3.1	23	12
Diffuse large B-cell lymphoma	58	3.3	16	7
Endometrial	248	2.5	58	15
Oesophageal adenocarcinoma	141	4	8	7
Glioblastoma multiforme	291	2.2	22	4
Head and neck	384	3.9	25	9
Kidney clear cell	417	1.9	15	6
Lung adenocarcinoma	405	8.1	22	10
Lung squamous cell carcinoma	178	9.9	11	13
Medulloblastoma	92	0.3	2	1
Melanoma	118	12.9	19	9
Multiple myeloma	207	1.6	11	3
Neuroblastoma	81	0.5	1	0
Ovarian	316	1.7	5	5
Prostate	138	0.7	4	2
Rhabdoid tumour	35	0.1	1	0

Further, although we have not identified any recurrent somatic mutation in our cohort, mutations in a few important cancer predisposing genes have been identified including PTEN, BRCA2 and NF1 (Table 4.7 & Table 4.8). Among other mutated genes, PRSS3 (Serine protease 3) is shown to play important role in the development of several malignancy including invasive ducal carcinoma of the breast (91). PRSS3 mRNA and protein levels were shown to be significantly unregulated in IDC of the breast tissue compared to the normal tissue. The findings suggested that PRSS3 expression levels could be used as a prognosis marker for IDC and targeting PRSS3 can potentially represent a new treatment strategy for IDC patients (91).Further, we have identified a nonsense mutation in PRDM9 gene. PRDM9 is known as meiosis specific gene and is involved in induction of double stranded breaks during meiotic recombination. Recently, this gene was found to be over-expressed in different cancer types and this aberrant expression was found to be associated with genomic instability in cancer (92).

Although we have found a few important genes which have been shown to have a role in cancer development and progression, none of them was found to be recurrent in our samples which may be because of small cohort size (n=15). In the exome study performed by Agrawal et al. group on sporadic MTC (n=17), except for mutually exclusive mutation in RET or RAS gene, no other recurrent gene mutation was found. Even in a combined analysis of somatic mutations in our hereditary MTC cohort along with the sporadic MTC cohort from USA, no gene was found to be mutated in more than one tumour sample. Therefore, a larger cohort study is needed for this rare yet aggressive thyroid malignancy to identify the underlying genetic signatures that governs phenotypic heterogeneity observed in patients.

Chapter 5

To Study of Modulatory role of SNPs of distinct genetic pathways on risk of MTC development and its clinical behavior

5.1 Background

Gain-of-function point mutations in RET proto-oncogene is well established as the primary susceptibility factor for MTC development and modifier of its clinical behavior. However certain aspects of the disease such as clinical heterogeneity despite having same mutation in RET gene is still unclear (93,94). Similarly for sporadic MTC cases, somatic *RET* mutations are not found in all the cases and also shows heterogeneity in different subpopulation of cells within the tumor (95–97).

Other than the high penetrance germline or somatic RET mutations, no other genetic, environmental or lifestyle related risk factors have been identified or established for their risk association with MTC. A few small studies have examined association of certain lifestyle related risk factors to be associated with MTC but with either inconclusive results or have identified a paradoxical protective risk association of tobacco smoking and alcohol with MTC(98–100). Several case control studies across different population have examined the association of SNPs of different genetic pathways with the risk of MTC. These include SNPs of RET gene, xenobiotic metabolism and cell cycle regulation pathways (101–123). A number of studies have also performed meta-analysis on SNPs of RET gene in order to investigate whether the presence of these SNPs could be associated with susceptibility to MTC development. However, most of these studies and their meta-analysis were either inconclusive or showed contradictory results. The possible reasons for not finding significant and consistent risk association could be the small cohort size of this rare cancer, geo-ethnic differences or poorly matched controls.

In addition to the studies on risk association of different SNPs with MTC development, several studies have also investigated the influence of SNPs of low penetrance genes on the clinical

behavior of MTC in terms of age of disease onset and its aggressiveness. However, it is still a matter of debate to what extent these neutral polymorphisms could have modifying effect in the pathogenesis of hereditary or sporadic MTC. Also, none of the risk association study or study on SNPs as modifier of clinical behavior of the disease has examined the association of all the above mentioned SNPs together in a single cohort.

Using the largest cohort of 438 MTC cases (361 sporadic and 77 hereditary) and 489 gender and ethnicity matched healthy controls from the 1000 Genome Project(124), South Asian population, we undertook a comprehensive risk-association analysis of 13 SNPs of 3 distinct genetic pathways. These include SNPs of genes of detoxification (*Cyp1A1m1, Cyp1A2*F, NAT2, GSTP1),* cell cycle regulation (*CDKN1A, CDKN1B, CDKN2A, CDKN2B, CDKN2C*) and the *RET* gene (G691S, L769L, S836S, S904S) (Table 5.1). We have also performed a meta-analysis of all the case control studies examining risk association of the 4 *RET* gene SNPs with MTC, including the present study, in order to derive definitive conclusions on the associations.

Further in the same cohort of 438 MTC cases, we have performed a comprehensive gene dose-response relationship analysis between these 13 SNPs and the clinical behavior of MTC. This gene dose-response relationship aimed to investigate the clinic-pathological differences between the wild-type versus heterozygous versus homozygous variant for each SNP separately in hereditary and sporadic MTC cases.

Table 5.1: List of SNPs of genes of distinct pathways selected for the study based on detailed literature search

Gene/SNP	Reference ID	Variant	Nucleotide change	Ancestral Allele	Amino acid Change
Genes of Detoxification					
Cyp1A1m1	rs4646903	3′UTR	T/C	T	-
Cyp1A2	rs762551	Intronic	A/C	A	-
NAT2	rs1041983	Synonymous	C/T	C	Y94Y
GSTP1	rs1695	Missense	A/G	A	I105V
Cell Cycle regulatory genes					
CDKN1A	rs1801270	Missense	C/A	C	S31R
CDKN1B	rs2066827	Missense	T/G	T	V109G
CDKN2A	rs11515	3′ UTR	G/C	G	-
CDKN2B	rs1063192	3′UTR	T/C	T	-
CDKN2C	rs12885	3′ UTR	G/T	G	-
RET Proto oncogene					
G691S RET	rs1799939	Missense	G/A	G	G691S
L769L RET	rs1800861	Synonymous	T/G	T	L769L
S836S RET	rs1800862	Synonymous	C/T	C	S836S
S904S RET	rs1800863	Synonymous	C/G	C	S904S

5.2 Methodology

5.2.1 Patient Cohort

The study was conducted on a cohort of 438 Indian MTC cases registered at Cancer Genetics Clinic, Tata Memorial Hospital as part of institutional ethics committee approved study over a period of 13 years (2006 to 2018). The inclusion criteria was histologically confirmed diagnosis of MTC with raised serum calcitonin in patients of any age or gender. Exclusion criteria included a previous history of another cancer except pheochromocytoma which is a part of MEN2 syndrome. The438 MTC cases include218 cases enrolled as part of this thesis and the remaining 220 case that were registered and were tested for germline RET mutation previously in our laboratory. For each patient, personal and family history with clinico-pathological details was recorded. Blood sample was collected with written informed consent and detailed genetic counseling. The cases were then tested for hotspot RET mutations and based on identified RET mutation and phenotyping, the cases were classified as hereditary and sporadic MTC cases. Of the 438 MTC cases, 77 cases were hereditary cases with a germline RET mutation and the remaining 361 were sporadic MTC cases without a RET mutation. Data for allele frequency of 13 SNPs in healthy controls was extracted from the 1000 genome project, South Asian population (http://www.ensembl.org/Homo_sapiens/Info/Index).This South Asian cohort included 489 healthy individuals representing major ethnicities of Indian origin - Punjabis from Lahore, Gujarati from Houston, Telugu from UK, Bengali from Bangladesh and Sri Lankan Tamil from UK.

5.2.2 Clinico-Pathological examination

The clinical diagnosis of MTC was made on biochemical evaluation of serum Calcitonin levels at diagnosis (reference range: 2-10mg/ml) and histopathological examination of surgically resected tumor tissues. The tumor volume was calculated using the ellipsoid volume formula 0.5 x l × w × h cm^3 (Tomayko and Reynolds, 1989)(125). The diagnosis of nodal metastasis was made on the histopathological examination of lymph nodes resected at the time of total thyroidectomy. Distant metastasis was confirmed by imaging in a background of very high serum calcitonin levels.

5.2.3 Molecular Genetic Testing

DNA was extracted from the blood sample using Qiagen QIAmp DNA mini kit (Cat#51304) and tested for hotspot RET mutation by PCR and Sanger Sequencing (details on steps of PCR and Sanger Sequencing are discussed in Chapter 2; Materials and Methods). Sequencing data was analyzed using Chromas Lite software version 2.6.4 and reference sequence of RET was extracted from National Centre for Biotechnology Information (NCBI) NG_007489.1. SNP genotyping was performed by Restriction Fragment Length Polymorphism (RFLP) for *Cyp1A1m1, Cyp1A2*F, GSTP1, NAT2, CDKN1A, CDKN1B, CDKN2A, RET* L769L, S836S and S904S polymorphisms and by TaqMan for *CDKN2B, CDKN2C* and *RET* G691S polymorphisms (Details discussed in Chapter 2; Materials and Methods). TaqMan SNP genotyping was employed for these 3 SNPs because no restriction site for a single cutter restriction enzyme was identified either for the wildtype or variant allele of the SNP. About 5% of all the SNP genotyping results were validated by Sanger Sequencing.

5.2.4 Statistical Analysis

All Statistical analysis was performed on SPSS v21.0. SNPs were tested for Hardy-Weinberg equilibrium (HWE) using Chi-square Hardy-Weinberg equilibrium test calculator for biallelic markers (http://www.oege.org/software/hwe-mr-calc.shtml) in both cases and controls (Table 5.2). Genotypic frequency was calculated for all the 13 SNPs and compared between cases and controls using chi-square test (Table 5.3 & Table 5.4).As the homozygous status of several SNPs was either absent or very low in either cases or controls, analysis was performed only for the dominant model which compares the variant allele either as heterozygous or homozygousform (Aa+aa) with the homozygous wild type allele (AA). Logistic regressions were used to analyze the association between these polymorphisms and MTC risk and odds ratio was calculated with 95% Confidence Interval (CI).All SNPs showing a trend for association on univariate analysis with $p<0.1$ were included in the multivariate logistic regression analysis.As multiple comparisons were made for 13 SNPs in a single cohort, a p value of <0.01 was used to consider an association as statistically significant.

Table 5.2: Hardy-Weinberg Equilibrium calculations for Cases & Controls

Gene/ SNP	Controls (n=489)		MTC cases(n=438)	
	MAF	***HWE p-value***	***MAF***	***HWE p-value***
Cyp1A1m1	0.34	0.35	0.31	**0.006**
Cyp1A2	0.47	0.35	0.43	0.65
NAT2	0.43	0.18	0.45	1.00
GSTP1	0.29	0.10	0.26	0.65
CDKN1A	0.15	0.18	0.09	0.14
CDKN1B	0.32	0.57	0.29	0.57
CDKN2A	0.05	0.76	0.07	**0.003**
CDKN2B	0.27	0.09	0.25	0.11
CDKN2C	0.11	0.69	0.10	0.31
G691S RET	0.24	0.84	0.27	0.17
L769L RET	0.40	0.57	0.35	0.60
S836S RET	0.08	0.74	0.08	0.33
S904S RET	0.24	0.75	0.29	0.05

For association study with patient's clinical outcome, the clinico-pathological details were collected from the patients during the time of their enrollment. The clinico-pathological parameters used for correlation studies included age at MTC diagnosis, serum Calcitonin levels at diagnosis (Pre-Op), tumor volume (cm^3), nodal and distant metastasis. To examine the gene dose-response relationship, the clinic-pathological variables were stratified between wild-type, heterozygous and homozygous variants and compared. Categorical data was tested with Pearson Chi-Square test and two-tailed Fisher exact test whereas for continuous data, the means were compared using one-way analysis of variance (ANOVA) and medians was compared using Kruskal-Wallis test. The level of significance was set at <0.05.

Table 5.3: Genotype frequency comparison between hereditary MTC (hMTC) cases and controls

Gene/ SNP	Genotype Frequency – hMTC cases (n=77)		Genotype Frequency – Controls (n=489)		
	Wt	***Hz + Hm***	***Wt***	***Hz + Hm***	***p-value***
Cyp1A1m1	37 (48.1%)	40 (51.9%)	218 (44.6%)	271 (55.4%)	0.57
Cyp1A2	28 (36.4%)	49 (63.6%)	145 (29.7%)	344 (70.3%)	0.23
NAT2	31 (42.2%)	46 (59.7%)	167 (34.2%)	322 (65.8)	0.29
GSTP1	35 (45.5%)	42 (54.5%)	251 (51.3%)	238 (48.7%	0.34
CDKN1A	**65 (84.4%)**	**12 (15.6%)**	**361 (73.8%)**	**128 (26.2%)**	**0.045**
CDKN1B	35 (45.5%)	42 (54.5%)	224 (45.8%)	265 (54.2%)	0.96
CDKN2A	67 (87%)	10 (13%)	439 (89.7%)	50 (10.2%)	0.46
CDKN2B	46 (59.8%)	31 (40.2%)	266 (54.4%)	223 (45.6%)	0.38
CDKN2C	62 (80.5%)	15 (19.5%)	387 (79.1%)	102 (20.8%)	0.78
G691S RET	37 (48.1%)	40 (51.9%)	283 (57.8%)	206 (42.1%)	0.11
L769L RET	35 (45.5%)	42 (54.5%)	178 (36.4%)	311 (36.6%)	0.13
S836S RET	64 (83.1%)	13 (16.9%)	411 (84%)	78 (15.9%)	0.84
S904S RET	36 (46.8%)	41 (53.2)	285 (58.3%)	204 (41.7%)	0.06

Table 5.4: Genotype frequency comparison between sporadic MTC (sMTC) cases and controls

Gene/ SNP	Genotype Frequency – sMTC cases (n=361)		Genotype Frequency – Controls (n=489)		
	Wt	***Hz + Hm***	***Wt***	***Hz + Hm***	***p-value***
Cyp1A1m1	161 (44.6%)	200 (553.4%)	218 (44.6%)	271 (55.4%)	0.996
Cyp1A2	117 (32.4%)	244 (67.6%)	145 (29.7%)	344 (70.3%)	0.390
NAT2	103 (28.5%)	258 (71.5%)	167 (34.2%)	322 (65.8)	0.082
GSTP1	208 (57.6%)	153 (42.4%)	251 (51.3%)	238 (48.7%	0.069
CDKN1A	**295 (81.7%)**	**66 (18.3%)**	**361 (73.8%)**	**128 (26.2%)**	**0.007**
CDKN1B	182 (50.4%)	179 (49.6%)	224 (45.8%)	265 (54.2%)	0.184
CDKN2A	310 (85.9%)	51 (14.1%)	439 (89.7%)	50 (10.2%)	0.083
CDKN2B	206 (57%)	155 (42.9%)	266 (54.4%)	223 (45.6%)	0.439
CDKN2C	297 (82.2%)	64 (17.7%)	387 (79.1%)	102 (20.8%)	0.256
G691S RET	201 (55.7%)	160 (44.3%)	283 (57.8%)	206 (42.1%)	0.523
L769L RET	146 (40.4%)	215 (59.6%)	178 (36.4%)	311 (36.6%)	0.231
S836S RET	311 (86%)	50 (13.9%)	411 (84%)	78 (15.9%)	0.398
S904S RET	194 (53.7%)	167 (46.3%)	285 (58.3%)	204 (41.7%)	0.187

5.2.5 Literature search and Meta-analysis

PUBMED search was conducted in order to identify the eligible studies for meta-analysis using the following search words: 'Polymorphism AND MTC', 'SNPs AND MTC', '*RET* Polymorphisms AND MTC'. The Inclusion criteria of the studies include the case-control studies examining association of the desired SNPs with MTC risk.

Table 5.5: Details of sporadic MTC studies included in the meta-analysis

Author	Year	Population	Variant	Cases	Controls
Gimm	1999	German+USA	G691S, L769L, S836S, S904S	50	70
Ruiz	2001	Spain	S836S	32	250
Berard	2004	France	L769L, S836S	92	86
Elisei	2004	Italy	G691S, L769L, S836S, S904S	106	106
Wiench	2004	Poland	L769L, S836S	135	90
Baumgartner	2005	Austria	L769L, S836S	45	79
Cebrian	2005	British	G691S, L769L, S836S, S904S	124	517
Costa	2005	Portugal	G691S, L769L, S836S, S904S	50	50
Severskava	2006	Russian	G691S, L769L, S836S, S904S	45	178
Fernandez	2006	Spanish	G691S, S904S	58	100
Fugazzola	2008	Italy	G691S, L769L, S836S, S904S	140	190
Sromek	2010	Poland	L769L, S836S, S904S	217	420
Ceolin	2012	Brazil	G691S, L769L, S836S, S904S	107	308
Lantieri	2013	Italy	G691S	78	85
Present study	2019	Indian	G691S, L769L, S836S, S904S, Cyp1A1m1, Cyp1A2, NAT2, GSTP1, CDKN1A, CDKN1B, CDKN2A, CDKN2B, CDKN2C	361	489
Total				1640	2968

Genotyping studies with only MTC cases and no controls were excluded from meta-analysis as comparison of MTC cases with healthy controls was a prerequisite for the association study. Details of all the studies included in the meta-analysis is provided in Table 5.5 & Table 5.6. Meta-Analysis was performed with R-Software package using minor allele frequency data as the genotype frequencies were not available for many studies. W applied both the fixed-effect (126) and the random effect (127) model for meta-analysis. The significance of overall OR was calculated by Z-test. Heterogeneity between studies was investigated using I^2 and τ^2 statistics. The results of meta-analysis were reported as conventional forest plots.

Table 5.6: Details of hereditary MTC studies included in the meta-analysis

Author	Year	Population	Variant	Cases	Controls
Robledo	2003	Spanish	G691S	104	653
Severskava	2006	Russian	G691S, L769L, S836S, S904S	22	178
Tamanaha	2009	Brazilian	G691S, L769L, S836S, S904S	77	100
Sharma	2011	Indian	G691S, L769L, S836S, S904S	51	50
Lantieri	2013	Italian	G691S	15	85
Present Study	2019	Indian	G691S, L769L, S836S, S904S, Cyp1A1m1, Cyp1A2, NAT2, GSTP1, CDKN1A, CDKN1B, CDKN2A, CDKN2B, CDKN2C	77	489
Total				346	1555

5.3 Results

5.3.1) To study the genetic risk association of SNPs of distinct genetic pathways with MTC.

The details of all the 13 SNPs included in the study are summarized in Table 5.1. The 438 MTC cases in our cohort included 239 males (54.5%) and 199 (45.4%) females. The mean age at MTC diagnosis was 40.64 ±14.24, Median: 40years with the range of 8-80years. The 489 controls used for the risk association study included 260 males (53.2%) and 229 females (46.8%). Both the cases and controls were matched for both gender (p = 0.67) and ethnicity. The genotype frequencies of all the SNPs included in the study are summarized in Table 5.3 & Table 5.4. Hardy-Weinberg Equilibrium was maintained for all 13 SNPs in the controls and for 11/13 SNPs in the MTC cases (Table 5.2).

- **Risk Association: Present Study**

The genotype of each SNP as homozygous wild-type, heterozygous and homozygous variant was compared between cases and controls using dominant model of inheritance. The selection of this model was to evaluate the role of variant allele either as homozygous or heterozygous form in conferring the risk of MTC development.

On univariate logistic regression analysis, *CDKN1A* SNP showed consistent and significant association with reduced risk of MTC in both hereditary (OR = 0.52; 95% CI = 0.27-0.99; p = 0.048) and sporadic MTC groups (OR = 0.63; 95% CI = 0.45-0.88; p = 0.007). The variant allele A was overrepresented in the control population (26.2%) as compared to both the hereditary cases (15.6%) and the sporadic cases (18.3%) (Table 5.7 & Table 5.8)

Table 5.7: Univariate logistic regression analysis for association between SNPs and risk of hereditary MTC development.

Gene/ SNP	Genotype Frequency – hMTC cases (n=77)		Genotype Frequency – Controls (n=489)				
	Wt	*Hz + Hm*	*Wt*	*Hz + Hm*	*OR*	*95% CI*	*p-value*
Cyp1A1m1	37 (48.1%)	40 (51.9%)	218 (44.6%)	271 (55.4%)	0.870	0.537-1.407	0.570
Cyp1A2	28 (36.4%)	49 (63.6%)	145 (29.7%)	344 (70.3%)	0.738	0.446-1.220	0.236
NAT2	31 (42.2%)	46 (59.7%)	167 (34.2%)	322 (65.8)	0.770	0.470-1.259	0.297
GSTP1	35 (45.5%)	42 (54.5%)	251 (51.3%)	238 (48.7%	1.266	0.781-2.050	0.339
CDKN1A	**65 (84.4%)**	**12 (15.6%)**	**361 (73.8%)**	**128 (26.2%)**	**0.521**	**0.272-0.995**	**0.048**
CDKN1B	35 (45.5%)	42 (54.5%)	224 (45.8%)	265 (54.2%)	1.014	0.626-1.643	0.954
CDKN2A	67 (87%)	10 (13%)	439 (89.7%)	50 (10.2%)	1.310	0.634-2.708	0.465
CDKN2B	46 (59.8%)	31 (40.2%)	266 (54.4%)	223 (45.6%)	0.804	0.493-1.311	0.382
CDKN2C	62 (80.5%)	15 (19.5%)	387 (79.1%)	102 (20.8%)	0.918	0.501-1.680	0.781
G691S RET	37 (48.1%)	40 (51.9%)	283 (57.8%)	206 (42.1%)	1.485	0.917-2.404	0.108
L769L RET	35 (45.5%)	42 (54.5%)	178 (36.4%)	311 (36.6%)	0.687	0.423-1.115	0.129
S836S RET	64 (83.1%)	13 (16.9%)	411 (84%)	78 (15.9%)	1.070	0.562-2.037	0.836
S904S RET	36 (46.8%)	41 (53.2)	285 (58.3%)	204 (41.7%)	1.591	0.982-2.578	0.06

Table 5.8: Univariate logistic regression analysis for association between SNPs and risk of sporadic MTC development.

Gene/ SNP	Genotype Frequency – sMTC cases (n=361)		Genotype Frequency – Controls (n=489)				
	Wt	*Hz + Hm*	*Wt*	*Hz + Hm*	*OR*	*95% CI*	*p-value*
Cyp1A1m1	161 (44.6%)	200 (55.4%)	218 (44.6%)	271 (55.4%)	0.999	0.760-1.314	0.996
Cyp1A2	117 (32.4%)	244 (67.6%)	145 (29.7%)	344 (70.3%)	0.879	0.655-1.179	0.390
NAT2	103 (28.5%)	258 (71.5%)	167 (34.2%)	322 (65.8)	1.299	0.967-1.745	0.082
GSTP1	208 (57.6%)	153 (42.4%)	251 (51.3%)	238 (48.7%	0.776	0.590-1.020	0.069
CDKN1A	**295 (81.7%)**	**66 (18.3%)**	**361 (73.8%)**	**128 (26.2%)**	**0.631**	**0.452-0.882**	**0.007**
CDKN1B	182 (50.4%)	179 (49.6%)	224 (45.8%)	265 (54.2%)	0.831	0.633-1.092	0.184
CDKN2A	310 (85.9%)	51 (14.1%)	439 (89.7%)	50 (10.2%)	1.444	0.953-2.190	0.083
CDKN2B	206 (57%)	155 (42.9%)	266 (54.4%)	223 (45.6%)	0.898	0.682-1.180	0.439
CDKN2C	297 (82.2%)	64 (17.7%)	387 (79.1%)	102 (20.8%)	0.818	0.578-1.157	0.256
G691S RET	201 (55.7%)	160 (44.3%)	283 (57.8%)	206 (42.1%)	1.094	0.831-1.439	0.523
L769L RET	146 (40.4%)	215 (59.6%)	178 (36.4%)	311 (36.6%)	0.843	0.637-1.115	0.231
S836S RET	311 (86%)	50 (13.9%)	411 (84%)	78 (15.9%)	0.847	0.577-1.244	0.398
S904S RET	194 (53.7%)	167 (46.3%)	285 (58.3%)	204 (41.7%)	1.203	0.914-1.582	0.187

Multivariate logistic regression analysis further confirmed the strong association of CDKN1A SNP with reduced MTC risk in both the hereditary (OR = 0.27; 95% CI = 0.13-0.55; p = <0.001) and sporadic MTC groups (OR = 0.53; 95% CI = 0.36-0.78; p = 0.001) (Table 5.9 & Table 5.10). Multivariate logistic regression analysis also identified significant risk association for the *RET* S904S SNP in the hereditary MTC group (OR = 2.82; 1.64-4.86; p = <0.001) (Table 5.9) whereas for *CDKN2A* (OR = 1.89; 95% CI = 1.20-2.98; p = 0.006) and *NAT2* SNP (OR = 1.62; 95% CI = 1.17-2.25; p = 0.004) in the sporadic MTC group (Table 5.10).Taken together, the results of univariate and multivariate logistic regression analysis indicated that the inheritance of *CDKN1A* SNP confers a protective effect in MTC development in both hereditary and sporadic MTC group whereas the inheritance of *CDKN2A*, *NAT2* and *RET* S904S variants increases the susceptibility to Medullary Thyroid Carcinoma.

Table 5.9: Multivariate logistic regression analysis for association between SNPs and risk of hereditary MTC development **(SNPs with significance <0.1 from univariate analysis were included in multivariate analysis).**

Gene/ SNP	Genotype Frequency – hMTC cases (n=77)		Genotype Frequency – Controls (n=489)				
	Wt	***Hz + Hm***	***Wt***	***Hz + Hm***	***OR***	***95% CI***	***p-value***
CDKN1A	65 (84.4%)	12 (15.6%)	361 (73.8%)	128 (26.2%)	0.266	0.129-0.549	*<0.001*
S904S RET	36 (46.8%)	41 (53.2)	285 (58.3%)	204 (41.7%)	2.821	1.636-4.862	*<0.001*

Table 5.10: Multivariate logistic regression analysis for association between SNPs and risk of sporadic MTC development **(SNPs with significance <0.1 from univariate analysis were included in multivariate analysis).**

Gene/ SNP	Genotype Frequency – sMTC cases (n=361)		Genotype Frequency – Controls (n=489)				
	Wt	Hz + Hm	Wt	Hz + Hm	OR	95% CI	p-value
NAT2	103 (28.5%)	258 (71.5%)	167 (34.2%)	322 (65.8)	1.622	1.168-2.251	0.004
GSTP1	208 (57.6%)	153 (42.4%)	251 (51.3%)	238 (48.7%	0.741	0.540-1.018	0.065
CDKN1A	295 (81.7%)	66 (18.3%)	361 (73.8%)	128 (26.2%)	0.526	0.357-0.776	0.001
CDKN2A	310 (85.9%)	51 (14.1%)	439 (89.7%)	50 (10.2%)	1.888	1.197-2.978	0.006

- **Meta-analysis including present study**

To compare the findings of the present study with the published studies in literature in different population, a pooled meta-analysis was conducted. We identified 23 case control studies examining risk associations of one or more of these 13 SNPs with MTC. However, for 9 SNPs in the cell cycle regulation (*CDKN1A, CDKN1B, CDKN2A, CDKN2B, CDKN2C*) and detoxification pathway (*CYP1A1m1, CYP1A2*F, NAT2, GSTP1*), only single small cohort studies had examined their risk association with MTC(108,110,111,128). Hence the meta-analysis was performed only for the 4 *RET* gene SNPs (G691S, L769L, S836S, S904S) one or more of which are reported in 19 case control studies. This included a total of 346 cases and 1,555 controls in hereditary MTC group and 1,640 cases and 2,968 controls in sporadic MTC

group. The Odds Ratio (ORs) with 95% CIs calculated for the allelic distribution of SNPs for each study is shown in their respective Forest plots (Figure 5.1 – 5.4).

The meta-analysis identified a significant association between *RET* L769L and S836S SNPs with risk of hereditary MTC (Figure 5.2B & Figure 5.3B). The *RET* S836S variant allele was found to be associated with increased susceptibility to MTC. The effect was observed under both the fixed effect model (OR=1.69; 95% CI=1.17-2.43; p<0.01) and random effect model (OR=2.19; 95% CI=1.02-4.71; p=0.04).For *RET* L769L variant, a significant protective risk association with MTC was observed under both fixed effect model (OR=0.77; 95% CI=0.60-0.98; p=0.04) and random effect model (OR=0.77; 95% CI=0.60-0.98; p=0.04). Further in the sporadic MTC group, meta-analysis showed significantly increased risk of MTC with the *RET* G691S and S904S (Figure 5.1A and Figure 5.4A). For G691S, the association was observed under both fixed effect model (OR=1.21 95% CI=1.07-1.37; p<0.01) and random effect model (OR=1.24; 95% CI=1.04-1.47; p=0.02). For S904S, this effect was observed under only fixed effect model (OR=1.16; 95% CI=1.03-1.30; p=0.02).

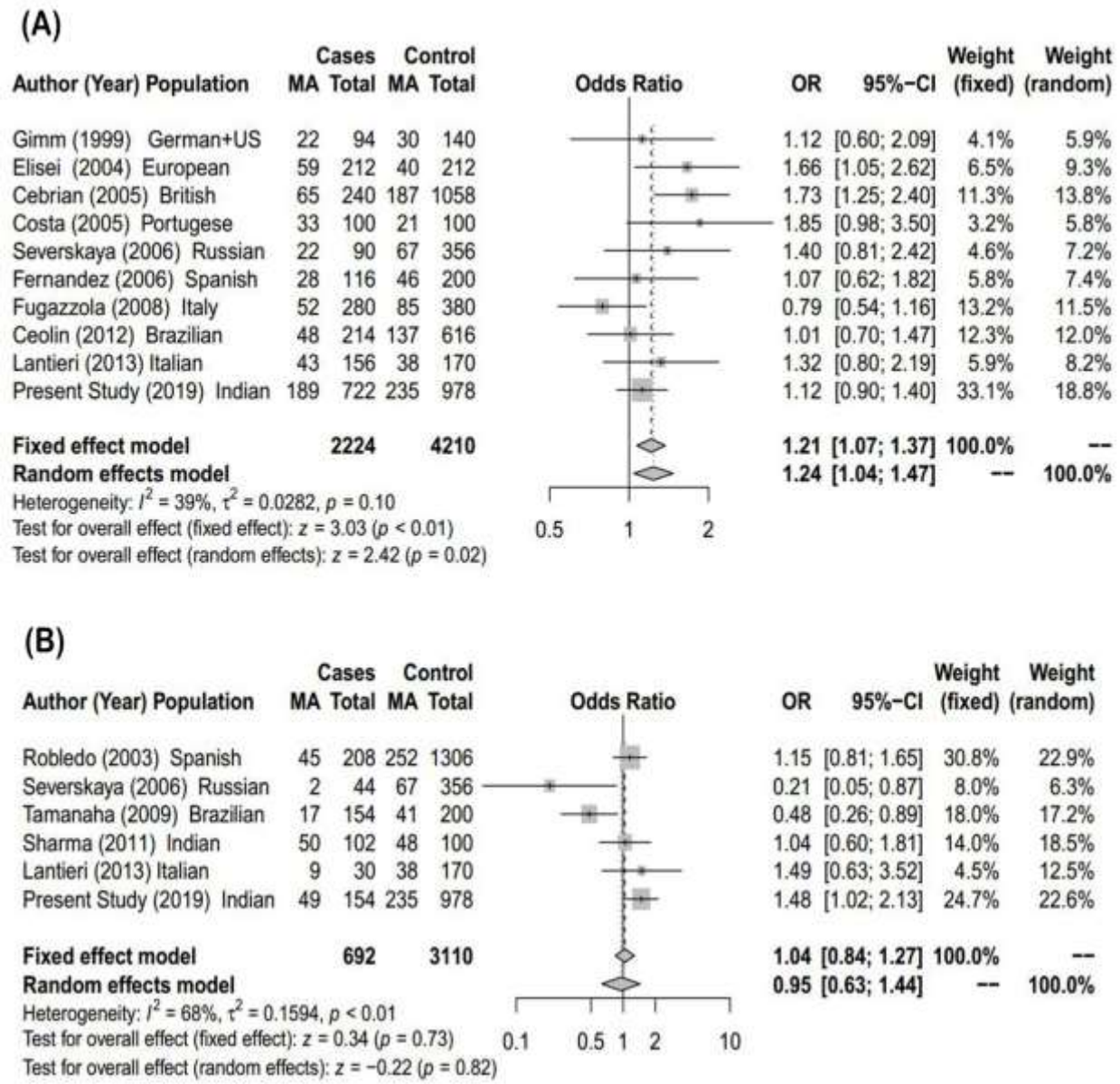


Figure 5.1: Forest Plot for meta-analysis on allelic association of *RET* G691S with **(A)** Sporadic MTC; **(B)** Hereditary MTC **[The total for cases and controls are allelic count (2n)]**

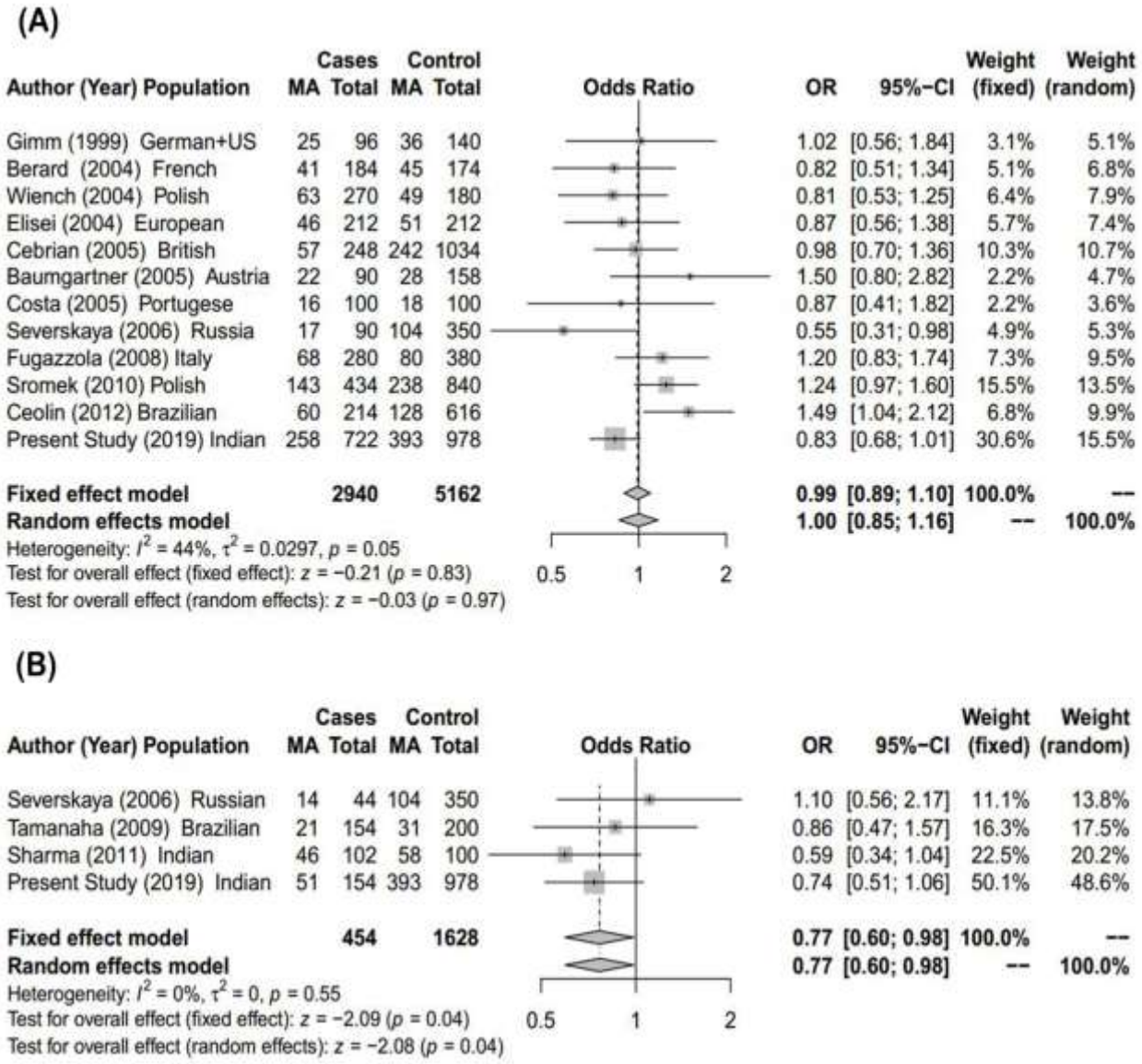


Figure 5.2: Forest Plot for meta-analysis on allelic association of *RET* L769L with **(A)** Sporadic MTC; **(B)** Hereditary MTC **[The total for cases and controls are allelic count (2n)]**

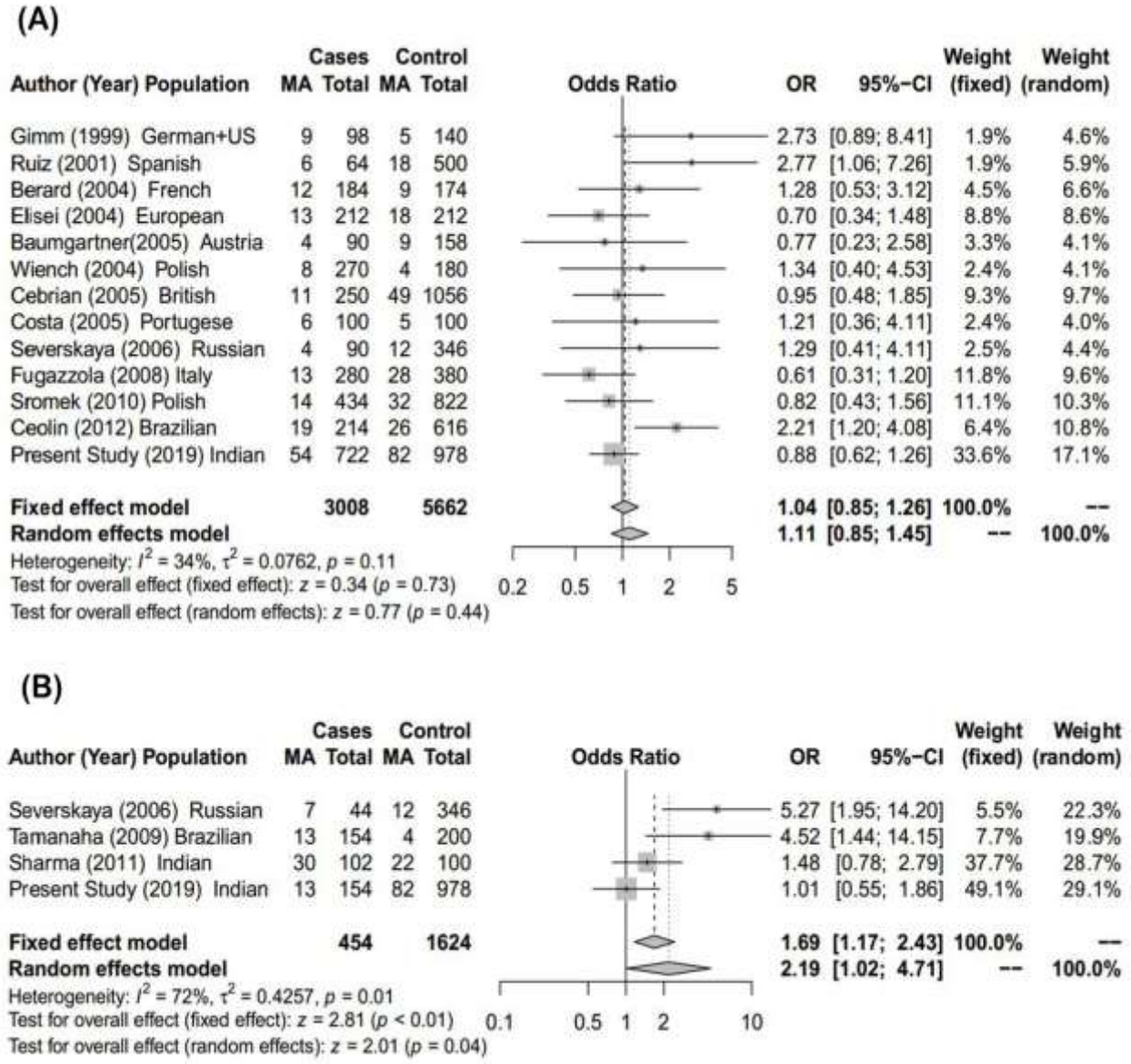


Figure 5.3: Forest Plot for meta-analysis on allelic association of *RET* S836S with **(A)** Sporadic MTC; **(B)** Hereditary MTC **[The total for cases and controls are allelic count (2n)]**

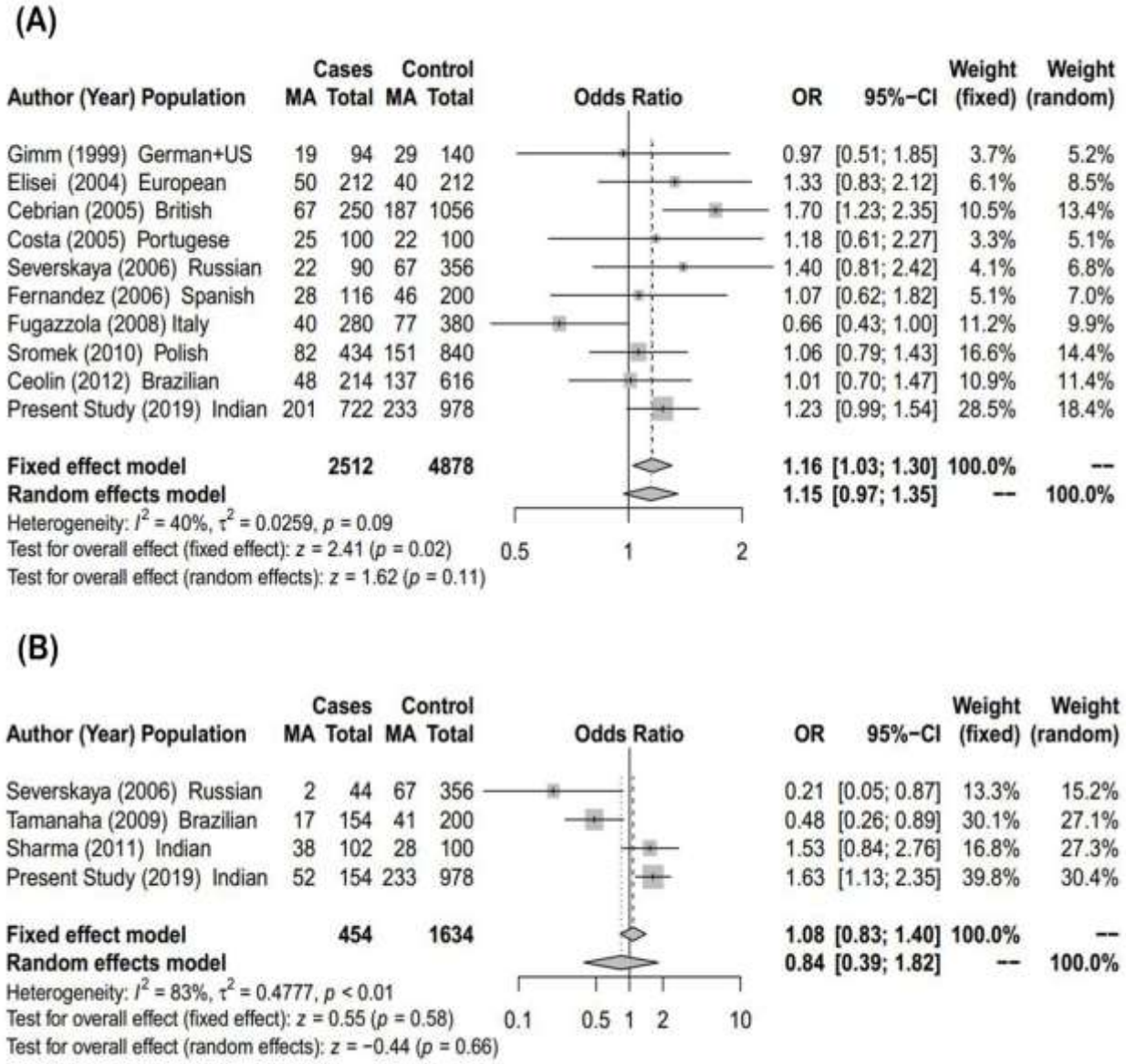


Figure 5.4: Forest Plot for meta-analysis on allelic association of *RET* S904S with **(A)** Sporadic MTC; **(B)** Hereditary MTC **[The total for cases and controls are allelic count (2n)]**

5.3.2 To study the Modulatory effect of SNPs of distinct genetic pathways with clinical behavior of MTC.

Several studies have investigated whether the presence of Single Nucleotide Polymorphisms (SNPs) in low penetrance genes could modulate the clinical behavior of the disease. If the presence of an SNP correlates with a change in clinical presentation of the disease then they may be considered as disease modifiers(114). In our gene dose-response association study, the only clinic-pathological parameter with which any SNP showed significant association was nodal metastasis at diagnosis. In hereditary MTC group, *Cyp1A1m1*, *CDKN2A* and *CDKN2C* SNP showed this association (Table 5.11). Patients with wildtype *Cyp1A1m1*showed higher rate of regional nodal metastasis compared to the heterozygous and homozygous variants (95.6% vs 86.2% vs 33.3%; p=0.01). Similarly patients with wildtype *CDKN2A* showed a higher incidence of nodal metastasis compared to their variant counterparts (91.3% vs 83.3% vs 33.3%; p=0.01). For CDKN2C, the comparison could only be made between wildtype and heterozygote as the homozygous variant for this SNP was absent. We observed that cases heterozygous for *CDKN2C* SNP showed higher rate of lymph node metastasis compared to wildtype *CDKN2C* (100% vs 86.3%; p=0.01) (Table 5.11).

In sporadic MTC group, this association was observed between the wildtype and the homozygous variant for *CDKN2C* SNP (Table 5.12). The cases wildtype for *CDKN2C* SNP showed higher rate of nodal metastasis compared to homozygous variant genotype (83.8% vs 40%; p = 0.03). The rate of nodal metastasis for heterozygous variant was similar to the wildtype *CDKN2C*(84.4% vs 83.8%) (Table 5.12). No other significant association was observed between any of the SNPs and the patient's clinico-pathological behavior.

Table 5.11 : Gene-dose response relationship comparison (wildtype vs heterozygous vs homozygous) in hereditary MTC group

SNPs	Age at MTC diagnosis; Median (n)				Calcitonin levels at diagnosis; Median (n)				Tumor Volume; cm^3 ;Median (n)				Lymph Node metastasis (n)				Distant Metastasis (n)			
	Wt (n)	Hz (n)	Hm (n)	p	Wt (n)	Hz (n)	Hm (n)	p	Wt (n)	Hz (n)	Hm (n)	p	Wt (n)	Hz (n)	Hm (n)	p	Wt (n)	Hz (n)	Hm (n)	p
RET G691S	31 (37)	38 (29)	21 (9)	0.06	1807 (23)	3397 (23)	1076 (5)	0.43	5.3 (10)	2.1 (16)	4.6 (5)	0.5	26 (28)	16 (20)	6 (7)	0.42	11 (27)	7 (18)	1 (7)	0.41
RET L769L	34 (33)	31 (33)	31 (9)	0.89	2243 (25)	2937 (23)	1101 (3)	0.92	2.1 (17)	6.0 (13)	12 (1)	0.27	24 (26)	19 (23)	5 (6)	0.56	8 (24)	8 (22)	3 (6)	0.75
RET S836S	31 (62)	34 (13)	0 (0)	0.59	2590 (44)	1670 (7)	0 (0)	0.96	4.6 (28)	1.4 (3)	0 (0)	0.3	45 (51)	3 (4)	0 (0)	0.44	17 (48)	2 (4)	0 (0)	0.56
RET S904S	31 (36)	36 (28)	21 (11)	0.07	2937 (23)	2243 (23)	1076 (5)	0.32	6.1 (10)	2.0 (14)	3.0 (7)	0.36	24 (26)	16 (20)	8 (9)	0.45	11 (26)	7 (18)	1 (8)	0.3
Cyp1A1m1	32 (35)	31 (36)	25 (4)	0.7	5859 (24)	1522 (26)	2937 (1)	0.1	3.8 (16)	2.1 (13)	4.7 (2)	0.83	22 (23)	25 (29)	1 (3)	**0.01**	9 (22)	10 (28)	0 (2)	0.51
Cyp1A2*F	31 (28)	30 (33)	39 (14)	0.22	3403 (17)	1753 (24)	6546 (10)	0.66	6.2 (10)	1.4 (17)	5.6 (4)	0.23	19 (21)	21 (25)	8 (9)	0.79	8 (18)	5 (24)	6 (10)	0.07
NAT2	32 (31)	31 (32)	29 (12)	0.77	2702 (18)	2437 (24)	1699 (9)	0.99	2.1 (13)	4.5 (15)	2.0 (3)	0.98	21 (23)	18 (22)	9 (10)	0.61	9 (22)	7 (20)	3 (10)	0.82
GSTP1	32 (34)	30 (34)	39 (7)	0.75	1716 (24)	3403 (23)	770 (4)	0.32	1.3 (11)	7.5 (16)	3.3 (4)	0.15	21 (24)	23 (27)	4 (4)	0.71	12 (25)	6 (24)	1 (3)	0.25
CDKN1A	31 (63)	31 (11)	35 (1)	0.74	1753 (42)	4443 (8)	3403 (1)	0.48	2.0 (26)	7.5 (4)	2.1 (1)	0.31	36 (43)	11 (11)	1 (1)	0.32	13 (41)	6 (10)	0 (1)	0.18
CDKN1B	31 (34)	31 (32)	32 (9)	0.76	3403 (23)	1753 (22)	15175 (6)	0.59	6.2 (14)	2.0 (13)	1.0 (4)	0.21	21 (23)	22 (26)	5 (6)	0.74	7 (21)	9 (25)	3 (6)	0.75
CDKN2A	31 (65)	30 (7)	23 (3)	0.6	2090 (42)	2484 (6)	2937 (3)	0.84	5.3 (24)	1.5 (5)	0.3 (2)	0.16	42 (46)	5 (6)	1 (3)	**0.01**	18 (43)	1 (6)	0 (3)	0.19
CDKN2B	31 (46)	31 (25)	28 (4)	0.66	3049 (30)	1237 (18)	3403 (3)	0.38	5.3 (16)	2.0 (11)	3.3 (4)	0.76	27 (33)	17 (18)	4 (4)	0.34	11 (31)	8 (18)	0 (3)	0.32
CDKN2C	31 (62)	31 (12)	23 (1)	0.56	2702 (40)	1672 (10)	2937 (1)	0.95	4.5 (25)	2.0 (5)	0.4 (1)	0.57	38 (44)	10 (10)	0 (1)	**0.01**	13 (40)	6 (11)	0 (1)	0.3

Table 5.12: Gene-dose response relationship comparison (wildtype vs heterozygous vs homozygous) in sporadic MTC group (n=361)

SNPs	Age at MTC diagnosis; Median (n)				Calcitonin levels at diagnosis; Median (n)				Tumor Volume; cm^3;Median (n)				Lymph Node metastasis (n)				Distant Metastasis (n)			
	Wt (n)	Hz (n)	Hm (n)	p	Wt (n)	Hz (n)	Hm (n)	p	Wt (n)	Hz (n)	Hm (n)	p	Wt (n)	Hz (n)	Hm (n)	p	Wt (n)	Hz (n)	Hm (n)	p
RET G691S	41 (199)	42 (129)	44 (29)	0.42	3334 (123)	2727 (68)	1673 (16)	0.65	4.9 (120)	7.8 (73)	2.5 (8)	0.17	143 (170)	75 (95)	20 (21)	0.17	50 (145)	33 (80)	6 (15)	0.58
RET L769L	42 (145)	40 (170)	40 (42)	0.41	2300 (85)	3008 (101)	5000 (21)	0.24	6.9 (73)	5.9 (100)	3.2 (28)	0.12	99 (113)	108 (137)	31 (36)	0.16	34 (94)	42 (112)	13 (34)	0.97
RET S836S	42 (307)	38 (46)	33 (4)	0.32	3008 (181)	3631 (25)	9367 (1)	0.67	5.8 (173)	5.0 (25)	2.8 (3)	0.41	206 (247)	30 (35)	2 (4)	0.19	78 (205)	11 (31)	0 (4)	0.29
RET S904S	40 (192)	42 (131)	44 (34)	0.29	3295 (122)	2727 (66)	2805 (19)	0.46	5.1 (118)	7.6 (70)	2.8 (13)	0.31	138 (165)	76 (96)	24 (25)	0.13	49 (140)	31 (82)	9 (18)	0.45
Cyp1A1m1	42 (159)	41 (173)	39 (25)	0.75	3008 (101)	2668 (94)	12094 (12)	0.32	6.0 (93)	5.4 (94)	2.3 (14)	0.38	105 (128)	118 (140)	15 (18)	0.88	43 (111)	40 (112)	6 (17)	0.89
Cyp1A2*F	41 (116)	42 (174)	40 (67)	0.89	2532 (69)	3415 (96)	4403 (42)	0.76	5.0 (67)	5.6 (92)	6.3 (42)	0.81	78 (91)	115 (136)	45 (59)	0.27	23 (79)	46 (114)	20 (47)	0.19
NAT2	41 (102)	41 (181)	42 (74)	0.54	2512 (62)	3298 (97)	2692 (48)	0.93	5.9 (54)	4.5 (101)	7.0 (46)	0.21	64 (75)	117 (145)	57 (66)	0.5	29 (66)	39 (117)	21 (57)	0.36
GSTP1	40 (206)	43 (127)	42 (24)	0.74	2692 (116)	4500 (78)	2922 (13)	0.53	4.7 (120)	8.8 (71)	4.5 (10)	0.22	132 (161)	88 (104)	18 (21)	0.81	49 (133)	31 (86)	9 (21)	0.84
CDKN1A	42 (291)	35 (66)	0 (0)	0.06	3402 (176)	1571 (31)	0 (0)	0.16	5.9 (157)	4.9 (44)	0 (0)	0.86	198 (234)	40 (52)	0 (0)	0.18	75 (196)	14 (44)	0 (0)	0.42
CDKN1B	43 (179)	39 (152)	39 (26)	0.14	2580 (111)	3511 (84)	3276 (12)	0.87	5.0 (97)	6.3 (87)	4.0 (17)	0.45	123 (144)	98 (121)	17 (21)	0.61	51 (120)	30 (101)	8 (19)	0.13
CDKN2A	42 (306)	40 (47)	42 (4)	0.89	3277 (170)	1970 (33)	1745 (4)	0.88	6.0 (169)	3.0 (29)	8.0 (3)	0.21	200 (244)	34 (38)	4 (4)	0.34	72 (199)	15 (37)	2 (4)	0.76
CDKN2B	42 (204)	40 (124)	43 (29)	0.46	2805 (119)	4670 (71)	1230 (17)	0.29	6.5 (112)	5.1 (72)	4.5 (17)	0.26	141 (167)	83 (99)	14 (20)	0.26	54 (137)	30 (85)	5 (18)	0.57
CDKN2C	42 (293)	39 (59)	40 (5)	0.49	3373 (167)	1593 (36)	6839 (4)	0.27	5.2 (167)	5.6 (32)	22.9 (2)	0.88	198 (236)	38 (45)	2 (5)	**0.03**	74 (200)	15 (36)	0 (4)	0.26

5.4 Discussion

In hereditary cancer syndromes, highly penetrant germline mutations in proto-oncogene or tumor suppressor genes confers a very high lifetime risk of cancer development (101). However in sporadic cancers, the risk of cancer development is mediated by both environmental or lifestyle related factors and genetic variants of low penetrance genes. While there are no clearly established lifestyle or environmental risk factors for susceptibility to MTC, several SNPs in *RET* as well as other genes have been reported to slightly increase or decrease the risk of MTC development (103,111,115,121,129). However, the findings of these studies are inconclusive.

The two large studies on association of *RET* G691S with susceptibility to sporadic MTC by Elisei *et al*. (2004) and Cebrian *et al*. (2005) demonstrated that the frequency of this SNP was significantly higher in patients with sporadic MTC compared to controls (130,131). The influence of this polymorphism on *RET* transcription and splicing was also studied but the results were inconclusive (130,131). Although several studies have found an association between *RET* G691S and S904S SNP with susceptibility to MTC, others could not observe the difference in frequency of these variants between patients and controls (109,116,132,133). In the present study, we observed a significant association between *RET* S904S SNP and susceptibility to hereditary MTC under multivariate logistic regression analysis (Table 5.9). The SNP is overrepresented in MTC cases compared to controls (53.2% vs 41.7%) thereby increasing the susceptibility to hereditary MTC (OR = 2.82, 95% CI = 1.64-4.86, p <0.001). Further, in our meta-analysis, we observed that both G691S and S904S SNPs are associated with increased susceptibility to sporadic MTC (Figure 5.1A & Figure 5.4A). For G691S, we observed this association under both random effect (OR = 1.24, 95% CI = 1.04-1.47, p = 0.02) and fixed effect model (OR = 1.21, 95% CI = 1.07-1.37, p = <0.01). For S904S, the association was only observed under the fixed effect model (OR

= 1.16, 95% CI = 1.03-1.30, p = 0.02). In meta-analysis study performed by Lantieri *et al*. (2013) for role of G691S in MTCs, an overall significant allelic association was observed under the fixed effect model (p=0.0049) and a borderline significance under the random effect model (p = 0.057) (134). They postulated through *in- silico* analysis, that change in phosphorylation pattern might account for enhanced signaling for G691S which in turn explains its overrepresentation in MTC cases (134).

For *RET* L769L and S836S SNP, several studies in different population studied the presence of these SNPs in both hereditary and sporadic MTC cases and controls including meta-analysis (117) (93,120,123,131,135). In the present study, we did not observe any association between L769L or S836S SNP and risk of MTC on univariate or multivariate logistic regression analysis in either the hereditary or sporadic MTC group. However, in our meta-analysis, we observed a significant protective effect of L769L polymorphism on MTC development and an increased risk of MTC with S836S polymorphism in hereditary MTC group under both fixed effect and random effect model (Figure 5.2B & Figure 5.3B). The only study demonstrating the significant protective effect of L769L polymorphism with MTC risk was performed by Severskaya *et al*. (2006) in Russian population. They observed that in sporadic MTC group, the allele frequency of RET L769L variant was significantly lower than the control group (p <0.04)(104). They have also showed a significant over-representation of RET S836S polymorphism suggesting it to be associated with increased susceptibility to hereditary MTC(104). Similar findings of association of S836S SNP with increased risk of MTC was observed by Gimm *et al*. (1999), Siqueira *et al*. (2010) and Ceolin *et al*. (2012) (109,136,137). Although the exact molecular mechanism by which these *RET* polymorphisms exert their effect is still poorly understood, it is postulated that these SNPs could influence the stability of mRNA or create a new alternative splicing site(109).

However the S836S variant failed to affect mRNA stability or alternative splicing (138), but it is speculated to create an unstable sequence upstream or downstream at germline or somatic *RET* mutations instead of directly participating in the tumorigenic process(136).

Over the past few years, studies on SNPs of genes of detoxification and cell cycle regulatory genes have also been studied as modulators of both hereditary and sporadic MTC (111,128,139,140). Most of the carcinogens that we are exposed to are metabolized by phase I & II xenobiotic metabolizing enzymes such as cytochrome P450 family of phase I enzymes *Cyp1A1m1* and *Cyp1A2*F* as well as phase II enzymes N-Acetyltransferase 2 (*NAT2*), Glutathione S-transferase (*GSTP1*) (139). Different genetic makeup due to SNPs in the detoxifying genes may modulate the risk of cancer development(139). Similarly alterations in genes regulating cell cycle progression such as cyclins, cyclin dependent kinase (CDKs) and CDK inhibitors (CDKIs) play a significant role in pathogenesis of different human Neoplasia (141).

Barbieri *et al*. in 2012 and 2013 reported that genes of detoxification are important modulators of both hereditary as well as sporadic MTC risk (111,139). They have demonstrated that the *Cyp1A2*F, GSTP1 and NAT2* SNPs increases the risk for hereditary MTC (111) whereas in sporadic MTC, *NAT2 SNP* reduces the risk of disease development(139). Barbieri *et al.*(2014) also reported that cell cycle regulation SNPs *CDKN1B* and *CDKN2A* to be associated with susceptibility to sporadic MTC (128). The *CDKN1B* SNP was more frequent in patients (62.2%) than controls (40.2%) increasing the risk of sporadic MTC whereas the *CDKN2A* SNP was more frequent in controls (32.65%) than the patients (15.56%) reducing the risk of sporadic MTC.

In our study, there was an increased risk association of *CDKN2A* SNP with sporadic MTC under multivariate logistic regression analysis (OR = 1.88, 95% CI = 1.19-2.97, p=0.006) (Table 5.10). Further, univariate and multivariate logistic regression analysis in our cohort also demonstrated a strong protective effect of *CDKN1A* SNP and the risk of MTC development in both hereditary and sporadic MTC group (Table 5.7 – 5.10). *CDKN1A* gene, also known as $p21^{CIP1/WAF1}$, encodes a cyclin-dependent kinase inhibitor which binds to and inhibits the activity of Cyclin-CDK2 or CDK4 complexes regulating cell cycle progression at G1 stage (142,143). *CDKN1A* activity is regulated by p53 which binds to its promoter and induces cell cycle arrest in response to various stimuli (143). This gene is often deregulated in human cancers with altered expression reported in several cancers including cervical, breast, ovarian, liver, uterine, and head and neck cancers (144). The *CDKN1A* SNP (rs1801270) at codon 31 (Ser31Arg) reported in the present study falls in a highly conserved N-terminal region of the protein, which is demonstrated to contain the tumor suppressor function(142). Functional studiessuggested that while the *CDKN1A*-Ser and Arg variant possess similar kinase inhibitory and growth suppression abilities (145), their transcriptional efficiency is significantly different (146). The allelic frequency of this SNP varies significantly among different populations with minor allele frequency of 15% in the South Asian Population (1000 Genome Project). Several molecular epidemiological studies of *CDKN1A* Ser31Arg SNP show conflicting result with some studies reporting increased risk association with tobacco related upper aerodigestive tract cancers (147), while a protective effect in HPV related cervical cancers (148,149). The only study of this SNP in MTC has been reported by Barbieri et al (128) in a small cohort of 45 sporadic MTC cases. Even though no significant risk association for MTC development was identified, perhaps due to small sample size, extrathyroidal tumor extension was significantly less in patients with the *CDKN1A* SNP as

compared to those with wild type *CDKN1A* (50% versus 92% , p=0.037). In our study of much larger cohort of this rare cancer, univariate and multivariate logistic regression analysis showed a highly significant protective effect of *CDKN1A* SNP on risk of MTC development in sporadic as well as hereditary MTC. Although the molecular mechanism of the possible protective effect of *CDKN1A* is not clear, it is suggested that the transcription of *CDKN1A*/*p21* is affected by the polymorphism which in turn may affect its function (128,148,150).

Along with studies of SNPs with the risk of MTC development, the role of several SNPs as modulators of clinical behavior of MTC has also been studied in different populations in both hereditary and sporadic MTC patients. A few studies have reported a significant dose-response relationship between the *RET* SNPs and the age at MTC diagnosis(107,114,151,152) whereas many others have failed to establish these associations (106,115,153,154).

Barbieri et al in 2012 and 2013 have studied the modulatory role of SNPs of genes of detoxification (*Cyp1A1m1, Cyp1A2*F, NAT2* and *GSTP1*) in both hereditary and sporadic MTC patients in Brazilian population (110,111). Although in their sporadic MTC cohort they failed to establish any significant relationship between the genetic profiles of these SNPs and patient's clinical outcome, they attributed it to their small cohort size (47 sporadic MTC cases) (110). Further in their study on hereditary MTC cohort of 132 patients, they demonstrated a significant association between these SNPs and the MTC clinical behavior (111). They showed that *Cyp1A2*F*SNP was associated with lesser tumor burden in patients and *GSTP1* SNP was associated with later age at MTC diagnosis (111). The same group also studied the role of cell cycle regulation gene SNPs (*CDKN1A, CDKN1B, CDKN2A, CDKN2B* and *CDKN2C*) in a small cohort of 45 sporadic MTC cases(128). They have reported that patients with wildtype *CDKN1A* presented extrathyroidal tumor extension more frequently (p=0.037) and patients with wildtype

CDKN2C presented larger tumors (p=0.032) whereas patients with polymorphic *CDKN2B* presented higher rate of distant metastasis (p=0.026). Pasquali *et al.*(2011) in an Italian cohort of 84 sporadic MTC patients and 90 matched controls, demonstrated for the first time *CDKN1B* V109G polymorphism to be associated with more favorable disease progression than the wild-type allele. Based on the protective role of this SNP in their study, they proposed this SNP can be considered a new promising prognostic marker in MTC (108). The Bradford-Hill criteria (155) established in 1965 for association studies comprehensively illustrates what aspects of the association should be considered before concluding upon the most likely interpretation in association studies. These include: (i) Strength of the association study; (ii) Consistency of the observed association with other studies; (iii) Specificity of the cause-effect relationships and; (iv) Biological credibility and plausibility of the study. Although statistically significant associations were observed in some of these studies, their cohort size was comparatively small. A small cohort study is likely to produce false-positive results than the conclusions drawn from a larger cohort study. Also, majority of the association studies reported so far have failed to fulfill these criteria in order to establish their effect on MTC causation.

In a large cohort study of 152 sporadic MTC cases, Machens et al in 2012 have investigated the clinical relevance of SNPs of *RET* gene (G691S, L769L, S836S & S904S) (153). They investigated the differences in clinic-pathological characteristics of sporadic MTC patients with or without these *RET* SNPs. They failed to identify any significant dose-response relationship between the *RET* SNPs and MTC clinical behavior and called for another comprehensive and larger cohort study for definitive conclusions.

The present study of 438 MTC represents the largest MTC cohort reported from any centre in the world. . In this cohort, a comprehensive investigation of 13 SNPs of three major genetic

pathways was undertaken Although a few SNPs show a statistically significant association ($p<0.01$) with regional nodal metastasis (Table 5.11 & Table 5.12), they may not represent a true gene dose-response relationship as they were not associated with any other clinico-pathological parameters. Our findings on *RET* SNPs were consistent with Machens et al group as we too have failed to establish any gene dose-response relationship between the carriers and non-carriers of *RET* gene SNPs in our large MTC cohort. Further, we could not replicate the findings of the Italian study which had reported significant prognostic association of *CDKN1B* SNP in MTC cases. This may be possibly explained by the lack of follow-up data in our cohort unlike the Italian study.

Taken together, in this largest SNP risk association study for MTC, a significant protective risk association of *CDKN1A*Ser31Arg SNP with MTC was shown for the first time. Expanded meta-analysis including our cohort identified significant risk association of 4 *RET*SNPs with MTC development, not observed in previous meta-analysis. The present study underscores the need for replicative risk-association studies for *CDKN1A* SNP as well as functional studies to understand molecular mechanisms through which these low penetrance alleles modulate MTC risk.

Also, this study is the first single cohort study to perform a comprehensive analysis of modulatory role of 13 most frequently studied SNPs with clinical outcome of MTC patients. We have extended our analysis on both the hereditary and sporadic MTC cohort and reported no significant association of these SNPs with clinical behavior of MTC. There is a need for additional genetic association studies on large cohort of MTC cases with clinic-pathological details and long-term follow-up data in order to draw consistent and definitive conclusions.

One of the limitations of our study is that unlike classical case control studies, instead of recruiting and genotyping matched controls, we used healthy gender and ethnicity matched south asian controls from the 1000 genome database. Matching for age was not possible as MTC, especially the hereditary MTC, is known to occur in childhood and recruiting minor subjects as healthy controls for genotyping study raises ethical issues. Of all the MTC case control studies, some have not reported whether controls were matched (107,129) whereas many have failed to obtain controls matched for age or gender (109,111,113). Moreover, in absence of clearly established lifestyle or environmental factors for MTC risk, none of the MTC SNP case control studies have described or matched for these factors, as is the case in our study.

Chapter 6

Identification of driver mutation in RET negative MEN2B cases

6.1. Background

Multiple Endocrine Neoplasia type 2B (MEN2B) is the most aggressive subtype of MEN2 syndrome and accounts for 5% of all MEN2 syndrome cases (53). It is characterized by Medullary Thyroid Carcinoma (MTC), pheochromocytoma and syndrome associated phenotypic features such as mucosal neuromas, blubbery lips, thickened eyelids and prominent corneal nerves, intestinal ganglioneuromatosis & skeletal abnormalities (Marfanoid habitus, narrow long facies) (44). MTC in MEN2B syndrome patients have an early age of onset and are more aggressive, metastasizing early to regional lymph nodes and other organs compared to MEN2A and FMTC. It is therefore important to establish the diagnosis of MEN2B at an early age when there is a possibility that thyroidectomy can be curative.

In contrast to MEN2A which has over 100 reported *RET* variants, the most frequently found mutation in MEN2B patients is at codon 918 of exon 16 of RET (M918T) & is found in 95% cases (51,102).The mutation produces a methionine to threonine substitution in the intracellular tyrosine kinase domain of the RET gene(156). This amino acid substitution occurs in the highly conserved substrate recognition pocket of the catalytic core of the tyrosine kinase domain and is postulated to alter substrate specificity. (157).Alteration in substrate specificity modifies the downstream intra-cellular signaling and when this occurs during embryonic and fetal development, the resulting phenotype is MEN2B (156).

Around 2-3% MEN2B cases have mutation at codon 883 of exon 15 of RET(158,159) in the absence of M918T mutation. Little is known about the clinical behavior of patients with the *RET* codon A883F mutation. Recent reports, however, suggest that patients with the A883F codon

mutation have a less aggressive MTC compared to patients with the M918Tcodon mutation. In addition to this, compound heterozygous mutations of codon 804, 768, 790, 791 also gives MEN2B phenotype which alone has a very low disease penetrance (102). Traditional Molecular Genetic Testing for MEN2 syndrome includes PCR amplification of 6 hotspot exons of RET (exon 10, 11, 13-16) followed by Sanger Sequencing for variant identification. American Thyroid Association guidelines task force recommends sequencing of the entire coding region of RET gene when no RET mutationis identified is identified in the hotspot exons, or there is a discrepancy between the MEN2 phenotype and the expected genotype(53).

In this part of the study, we examined 3 classical MEN2B syndrome cases which were associated with syndrome specific phenotypic features such as mucosal neuromas, marfanoid features etc. and the initial genetic testing by Sanger Sequencing failed to identify most common MEN2B associated mutation M918T or any other mutation on full gene testing for RET (20 exons). However, since these were classical MEN2B cases we presumed that the negative genetic test result could either be because of genotyping error on initial genetic testing or the involvement of another yet to be reported mutation in other gene which could give rise to MEN2B phenotype in the absence of Mutation in RET gene. Since genotyping errors in monogenic hereditary diseases like MEN2 can have major clinical implications, a systematic literature search was undertaken and an algorithm was proposed for suspecting and investigating genotyping errors. The algorithm was then tested in these 3*RET* negative MEN2B cases to confirm the possibility of genotyping error during the initial genetic testing and its possible causes.

6.2. Methodology

6.2.1 Clinical Data

The 4 RET negative patients reported in this part of the study were enrolled in the Cancer Genetics Clinic, Tata Memorial Hospital, Mumbai, India, at different time point between 2009-2014. Syndromic diagnosis was made by recording detailed personal and family history as well as the classical phenotypic features associated with MEN2B syndrome in the form of photographs. After detailed genetic counseling, they provided written informed consent for collection of photographs and for providing blood sample for genetic testing and biobanking as part of an Institutional Ethics Committee approved study. For minors, the consent was provided by the parents.

- **Case 1**

A 13 year old male with MEN2B associated phenotypic features (mucosal Neuromas and Marfanoid features) was diagnosed with synchronous MTC and Pheochromocytoma at the age of 13 years (Figure 6.1 & Figure 6.2). Total Thyroidectomy revealed multifocal MTC with metastasis to bilateral neck nodes. CT scan revealed a nodule in left adrenal gland with elevated serum normetanephrine of 459 ng/ml suggestive of Pheochromocytoma. A family history of cancer of Head & Neck was present without complete information of the nature of cancer was recorded.

Figure 6.1: Patient showing classical MEN2B phenotype: Mucosal Neuromas, blubbery lips & Marfanoid features

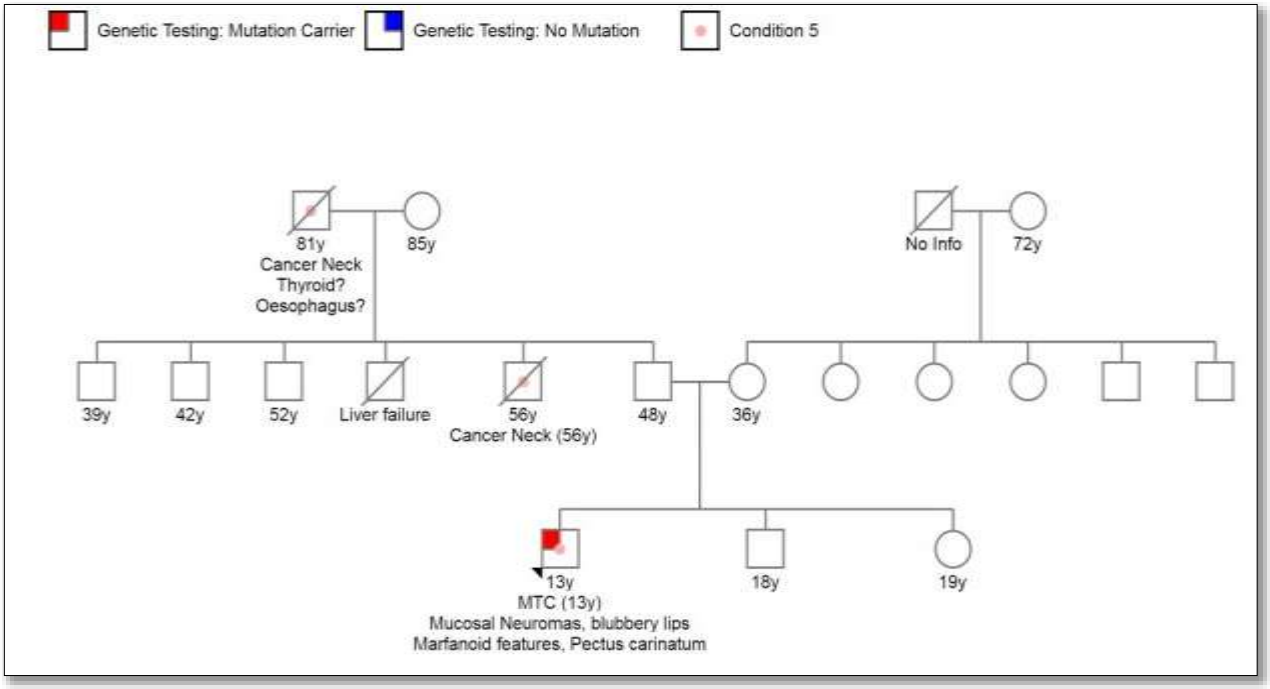


Figure 6.2: Pedigree of the classical MEN2B syndrome family with germline *RET* M918T mutation.

- **Case 2**

A 40 year old male with characteristic phenotypic features of MEN2B syndrome.was diagnosed with synchronous MTC and pheochromocytoma at the age of 40 year (Figure 6.3 & Figure 6.4). Total Thyroidectomy revealed multifocal MTC with metastasis to bilateral neck nodes.Bilateral adrenalectomy confirmed pheochromocytomain both adrenal glands.

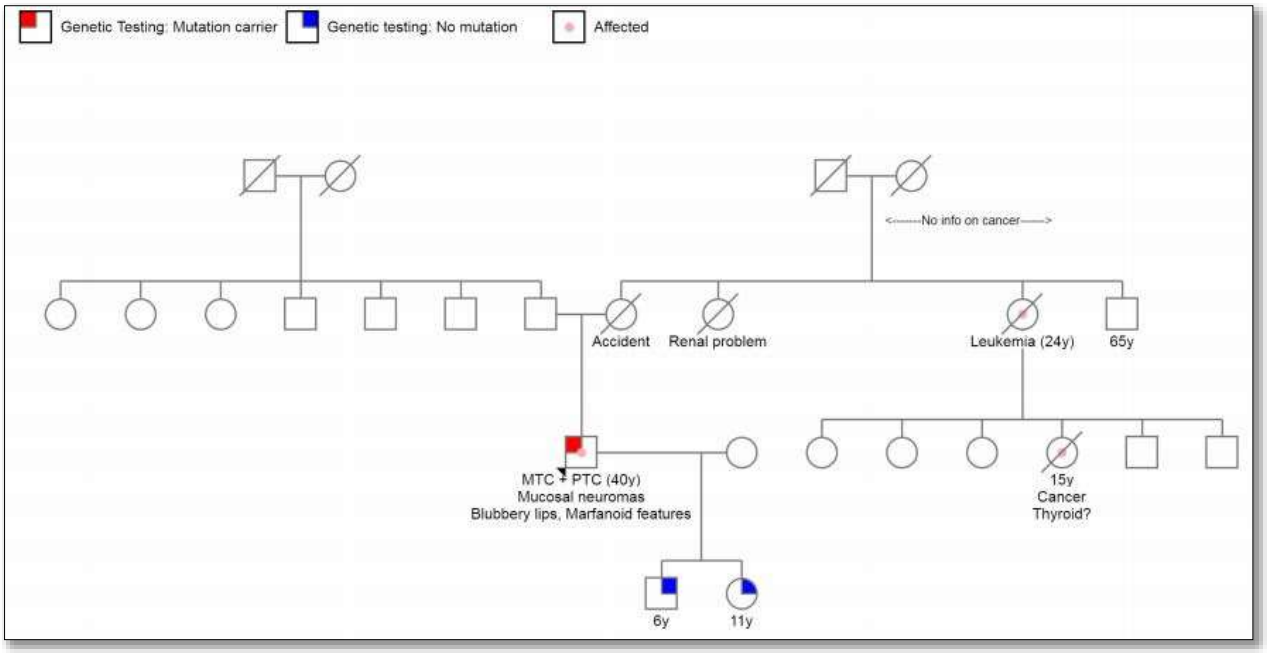


Figure 6.3: Pedigree of the classical MEN2B syndrome family with germline *RET* M918T mutation.

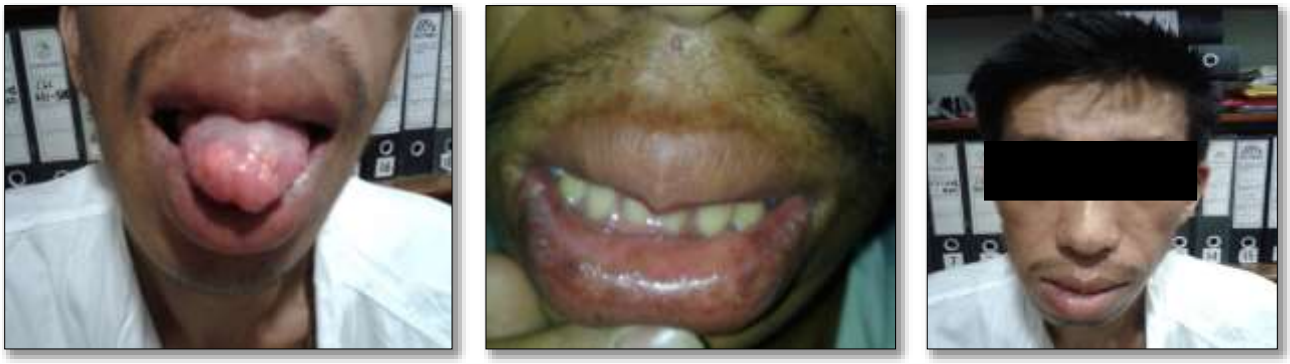


Figure 6.4: Patient showing classical MEN2B phenotype: Mucosal Neuromas, blubbery lips and marfanoid features.

- **Case 3**

A 17 year old male with characteristic phenotypic features of MEN2B (mucosal neuromas on lips, tongue, eyelids, Marfanoid features, blubbery lips) was diagnosed with MTC at the age of 15 years (Figure 6.5 & Figure 6.6). Total Thyroidectomy confirmed MTC. A part of intestine was also removed because of intestinal obstruction and history of constipation indicating ganglioneuromatosis, a clinical feature of MEN2B syndrome.

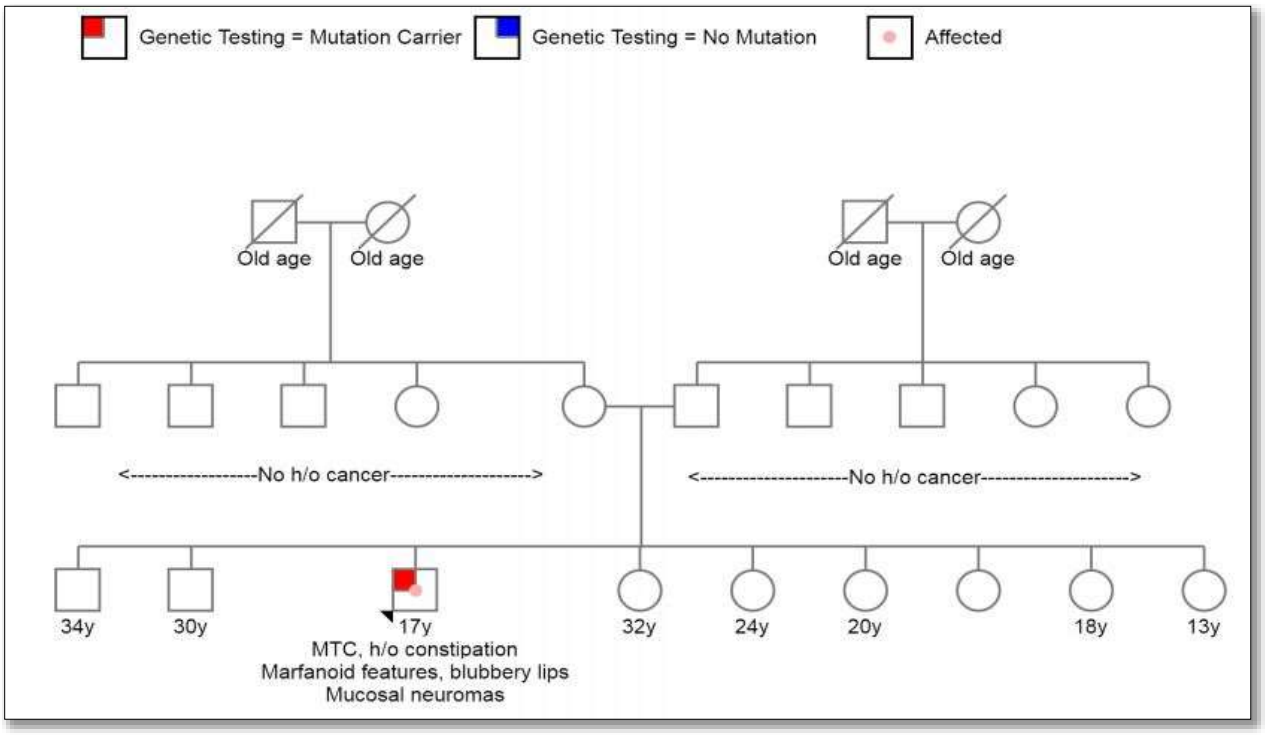


Figure 6.5: Pedigree of the classical MEN2B syndrome family with germline RET M918T mutation identified through whole exome sequencing.

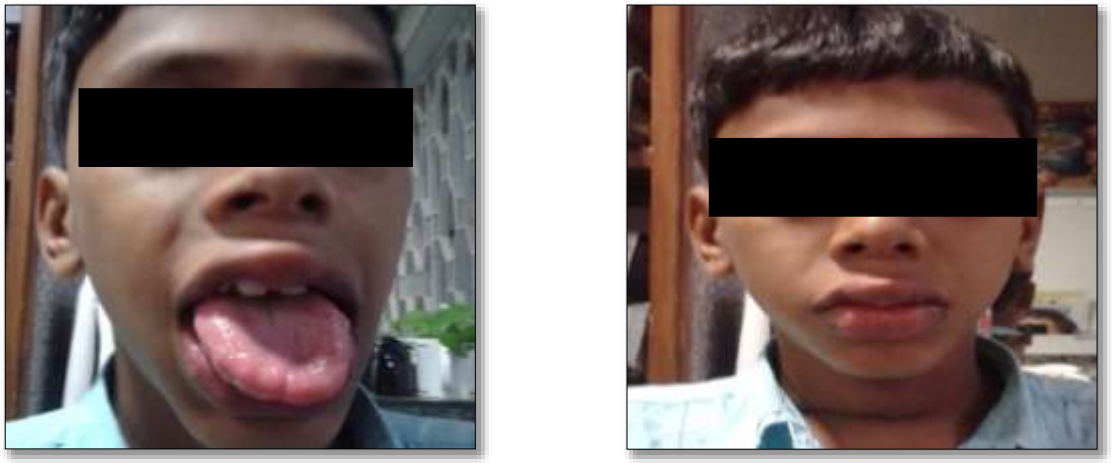


Figure 6.6: Patient showing classical MEN2B phenotype: Mucosal Neuromas, blubbery lips and Marfanoid features.

6.2.2 Molecular Genetic Testing

Blood sample was collected & DNA was extracted using QiagenQIAamp DNA Mini Kit (Cat#51304). Germline mutation analysis was carried out in three phases. Initially germline mutation analysis was carried out by bidirectional sequencing of the six hot spot exons of RET gene. When the cases were found to be negative for hotspot RET mutations, bidirectional sequencing of the remaining 14 exons of RET was performed. In the absence of mutation in the entire coding region of RET, germline whole exome sequencing was performed. Detailed steps of PCR reaction composition and thermal cycler conditions are described in chapter 2 (Materials and Methods).Primers for PCR were designed using Oligo Explorer version 1.5. Primer sequence for each exon of RET gene along with their annealing temperature is given in chapter 2 of Materials and Methods. Purification of PCR products was done using ExoSAP IT (USB Products,

Affimetrix). Sanger Sequencing was performed using BigDyes Terminator Cycle Sequencing kit v3.1 (Applied Biosystems)on ABI 3500 & 3730 DNA Sequencer (Applied Biosystems) and electropherograms were analyzed by chromas lite version 2.6.4.

Paired end whole exome sequencing was performed for all the 3 cases with a read length of 150 base pairs. Exome library was prepared from 50ng/µl gDNA using Nextera Rapid Capture Exome library preparation Kit following Illumina's standard protocol. Sequencing was done on HiSeq 2500 Genome Sequencer. Data Analysis was done using linux based standardized pipeline for exome data analysis which includes data quality check by FastQC, alignment withGRCh37/Hg19 reference genome using BWA alignment tools, post alignment processing using PICARD Tools, Variant Calling by GATK/Samtools and variant annotation using Annovar. Integrated Genome Viewer (IGV) version 2.3 was used for visualization of the variants.

6.2.3 Developing the algorithm for investigating genotyping errors

A systematic review of literature was undertaken on Pubmed to identify human studies on genotyping errors and its causes. Key words used for search were "genotyping errors", "false negative AND genetic testing", "Allele Dropout", "MEN2B AND *RET* negative". The search was not restricted to language of publication or year of publication. Based on our analysis of different types of false negative genotyping errors and its underlying causes that are described in the literature, we developed an algorithm for suspecting and systematically investigating false-negative genotyping errors in Mendelian diseases.

6.3 Results

The 3 MEN2B cases were enrolled in the Cancer Genetics Clinic at different time points between 2009 and 2014. Initial Sanger Sequencing was performed for exon 16 of RET gene which harbors the most frequently found mutation at codon 918 (M918T) associated with MEN2B. We failed to identify the M918T mutation in any of the 3 cases. We further tested all the hotspot exons of RET (exon 10, 11, 13, 14, 15 and 16) as there are reports of mutation at codon 883 (exon 15) or compound heterozygous mutations in the hotspot exons of RET which could also give classical MEN2B phenotypein the absence of M918T mutation. As we failed to identify any mutation in the hotspot exons, the entire coding region of RET (all 20 exons) was screened but had not revealed any mutation. Since these patients had an unambiguous clinical diagnosis of monogenetic MEN2B syndrome, we had a high index of suspicion for a false negative genetic test report for *RET*. Pre-analytic errors like sample mix-up were ruled out by performing *RET* sequencing on a second independent blood sample collected carefully and with full traceability.

Considering the characteristic MEN2B phenotype in absence of an identified *RET* mutation on multiple rounds of Sanger sequencing, and also since no mutation in any other gene is reported in literature which could give MEN2B phenotype in the absence of mutation in RET, germline whole exome sequencing was performed to identify any novel genotype phenotype association. However, exome sequencing unexpectedly revealed MEN2B specific *RET* M918T mutation in all these three patients Figure 6.7. *RET* codon 918 was sequenced at 92X, 102X and 158X depth in these patients as visualized on Integrated Genome Viewer (IGV) (Figure 6.8 – 6.10)

Func	Gene	ExonicFunc	AAChange	Conserved	SegDup	ESP5400_ALL	1000g2010nov_ALL	dbSNP130	AVSIFT	LJB_PhyloP	LJB_PhyloP_Pred	LJB_SIFT	LJB_SIFT_Pred	LJB_Polyphen2	LJB_Polyphen2_Pred
exonic	BMS1	synonymous SNV	NM_014753:c.T336C:p.T112T	480;Name=lod=119	0.92	0.465142	0.5	rs7074877							
exonic	BMS1	synonymous SNV	NM_014753:c.C2445T:p.P815P		0.96	0.135764	0.15	rs11239786							
exonic	BMS1	nonsynonymous SNV	NM_014753:c.G3421A:p.V1141I	522;Name=lod=177	0.96	0.132859	0.18	rs12764004	1	0.1726	N	0	T	0	B
exonic	RET	synonymous SNV	NM_020630:c.A135G:p.A45A			0.809072	0.74	rs1800858							
exonic	RET	synonymous SNV	NM_020630:c.A1296G:p.A432A	483;Name=lod=123		0.751348	0.81	rs1800860							
exonic	RET	synonymous SNV	NM_020630:c.G2307T:p.L769L	621;Name=lod=444		0.800335	0.72	rs1800861							
exonic	RET	nonsynonymous SNV	NM_020630:c.T2753C:p.M918T	506;Name=lod=152				rs74799832	0	0.9977	C	0.99	D	1	D
exonic	ZNF239	nonsynonymous SNV	NM_001099283:c.T625G:p.C209G	284;Name=lod=19		0.536494	0.5	rs2230661	0.03	0.9227	N	1	D	1	D
exonic	ZNF239	nonsynonymous SNV	NM_001099283:c.C515G:p.A172G			0.534767	0.52	rs2230660	0.08	0.7986	N	0.93	T	0.481	P
exonic	ZNF485	nonsynonymous SNV	NM_145312:c.C64T:p.R22W	278;Name=lod=18		0.123377	0.07	rs45545532	0.02						
exonic	ZNF485	synonymous SNV	NM_145312:c.G315A:p.L105L			0.374791	0.26	rs10899839							

Func	Gene	ExonicFunc	AAChange	Conserved	SegDup	ESP5400_ALL	1000g2010nov_ALL	dbSNP130	AVSIFT	LJB_PhyloP	LJB_PhyloP_Pred	LJB_SIFT	LJB_SIFT_Pred	LJB_Polyphen2	LJB_Polyphen2_Pred
exonic	ZNF33B	nonsynonymous SNV	NM_006955:c.C433T:p.R145C		0.93	0.766927	0.76	rs210280	0.07	0.06137	N	0.82	T	0	B
exonic	BMS1	nonsynonymous SNV	NM_014753:c.A1955G:p.K652R			0.158154	0.16	rs787795	0.26	0.97233	C	0.72	T	0	B
exonic	RET	synonymous SNV	NM_020630:c.A135G:p.A45A			0.809072	0.74	rs1800858							
exonic	RET	nonsynonymous SNV	NM_020630:c.G2071A:p.G691S			0.156256	0.16	rs1799939	0.62	0.23387	N	0.86	T	0.373	P
exonic	RET	synonymous SNV	NM_020630:c.G2307T:p.L769L	621;Name=lod=444		0.800335	0.72	rs1800861							
exonic	RET	synonymous SNV	NM_020630:c.C2712G:p.S904S	595;Name=lod=348		0.160376	0.14	rs1800863							
exonic	RET	nonsynonymous SNV	NM_020630:c.T2753C:p.M918T	506;Name=lod=152				rs74799832	0	0.99773	C	0.99	D	1	D
exonic	RET	nonsynonymous SNV	NM_020630:c.C2944T:p.R982C	580;Name=lod=304		0.016174	0.01	rs17158558	0.03	0.11117	N	0.98	D	0.977	D
exonic	ZNF485	synonymous SNV	NM_145312:c.G315A:p.L105L			0.374791	0.26	rs10899839							
exonic	ZNF485	nonsynonymous SNV	NM_145312:c.G754A:p.A252T			0.242703	0.17	rs12354886	0.63	0.80991	N	0.32	T	0	B

Func	Gene	ExonicFunc	AAChange	Conserved	SegDup	ESP5400_ALL	1000g2010nov_ALL	dbSNP130	AVSIFT	LJB_PhyloP	LJB_PhyloP_Pred	LJB_SIFT	LJB_SIFT_Pred	LJB_Polyphen2	LJB_Polyphen2_Pred
exonic	ZNF37A	synonymous SNV	NM_001178101:c.A1440G:p.S480S		0.92	0.50093	0.49	rs176889							
exonic	ZNF33B	nonsynonymous SNV	NM_006955:c.C433T:p.R145C		0.93	0.766927	0.76	rs210280	0.07	0.061372	N	0.82	T	0	B
exonic	BMS1	nonsynonymous SNV	NM_014753:c.A1955G:p.K652R			0.158154	0.16	rs787795	0.26	0.972332	C	0.72	T	0	B
exonic	RET	synonymous SNV	NM_020630:c.A135G:p.A45A			0.809072	0.74	rs1800858							
exonic	RET	synonymous SNV	NM_020630:c.A1296G:p.A432A	483;Name=lod=123		0.751348	0.81	rs1800860							
exonic	RET	synonymous SNV	NM_020630:c.G2307T:p.L769L	621;Name=lod=444		0.800335	0.72	rs1800861							
exonic	RET	nonsynonymous SNV	NM_020630:c.T2753C:p.M918T	506;Name=lod=152				rs74799832	0	0.99773	C	0.99	D	1	D
exonic	ZNF239	nonsynonymous SNV	NM_001099283:c.T625G:p.C209G	284;Name=lod=19		0.536494	0.5	rs2230661	0.03	0.922737	N	1	D	1	D
exonic	ZNF239	nonsynonymous SNV	NM_001099283:c.C515G:p.A172G			0.534767	0.52	rs2230660	0.08	0.798641	N	0.93	T	0.481	P
exonic	ZNF485	synonymous SNV	NM_145312:c.G315A:p.L105L			0.374791	0.26	rs10899839							
exonic	ZNF485	nonsynonymous SNV	NM_145312:c.G754A:p.A252T			0.242703	0.17	rs12354886	0.63	0.809905	N	0.32	T	0	B

Figure 6.7: Exome Sequencing analysis data showing RET M918T mutation in all the 3 samples

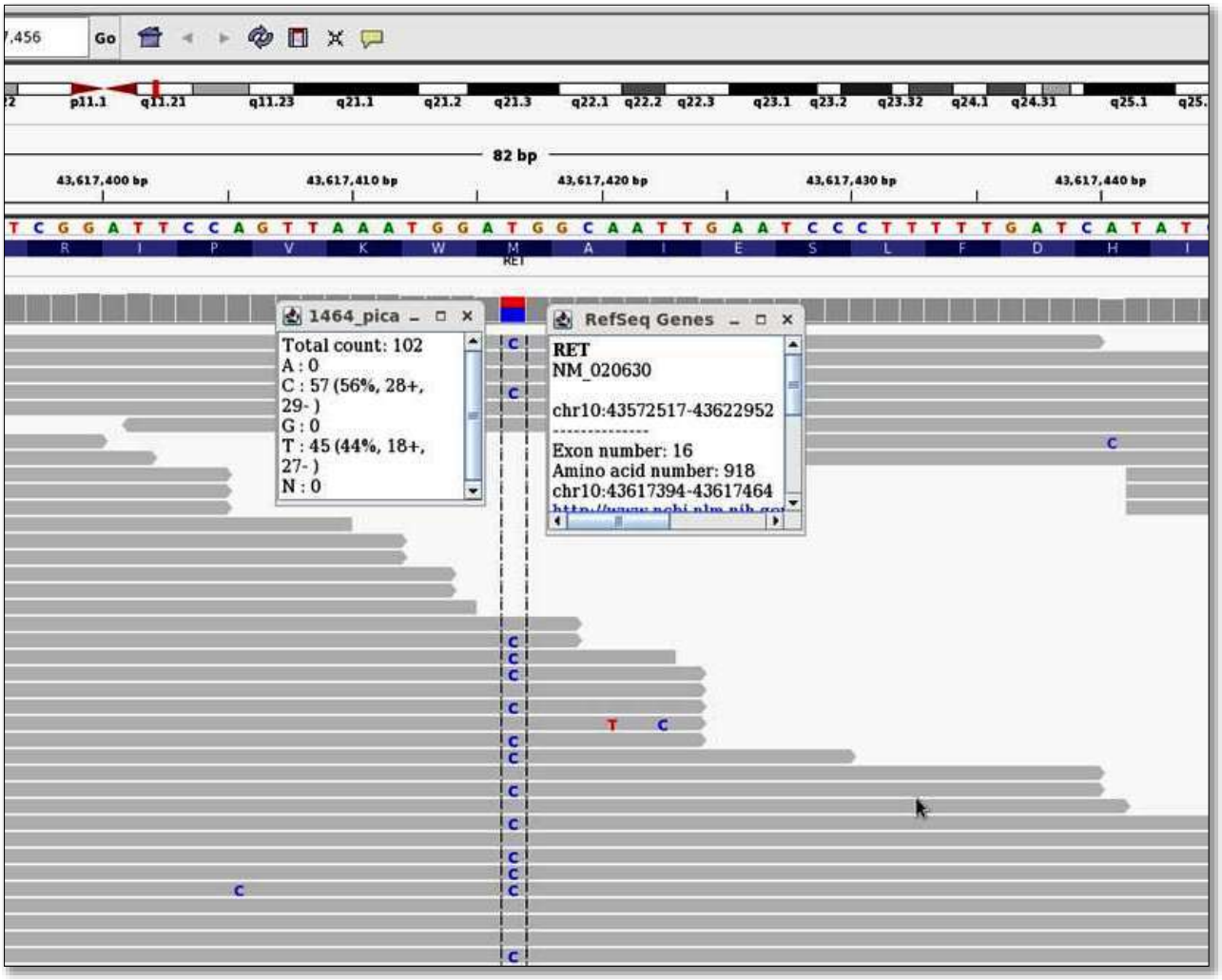


Figure 6.8: Exome data (.BAM file) visualization on Integrated Genome Viewer (IGV) showing read count of Wildtype (T) & Variant (C) allele for **p.M918T** mutation identified in exon 16 of RET gene.

Total no. of aligned reads: 102

Wildtype allele (T) frequency: 44% (45 reads)

Variant allele (C) frequency: 56% (57 reads)

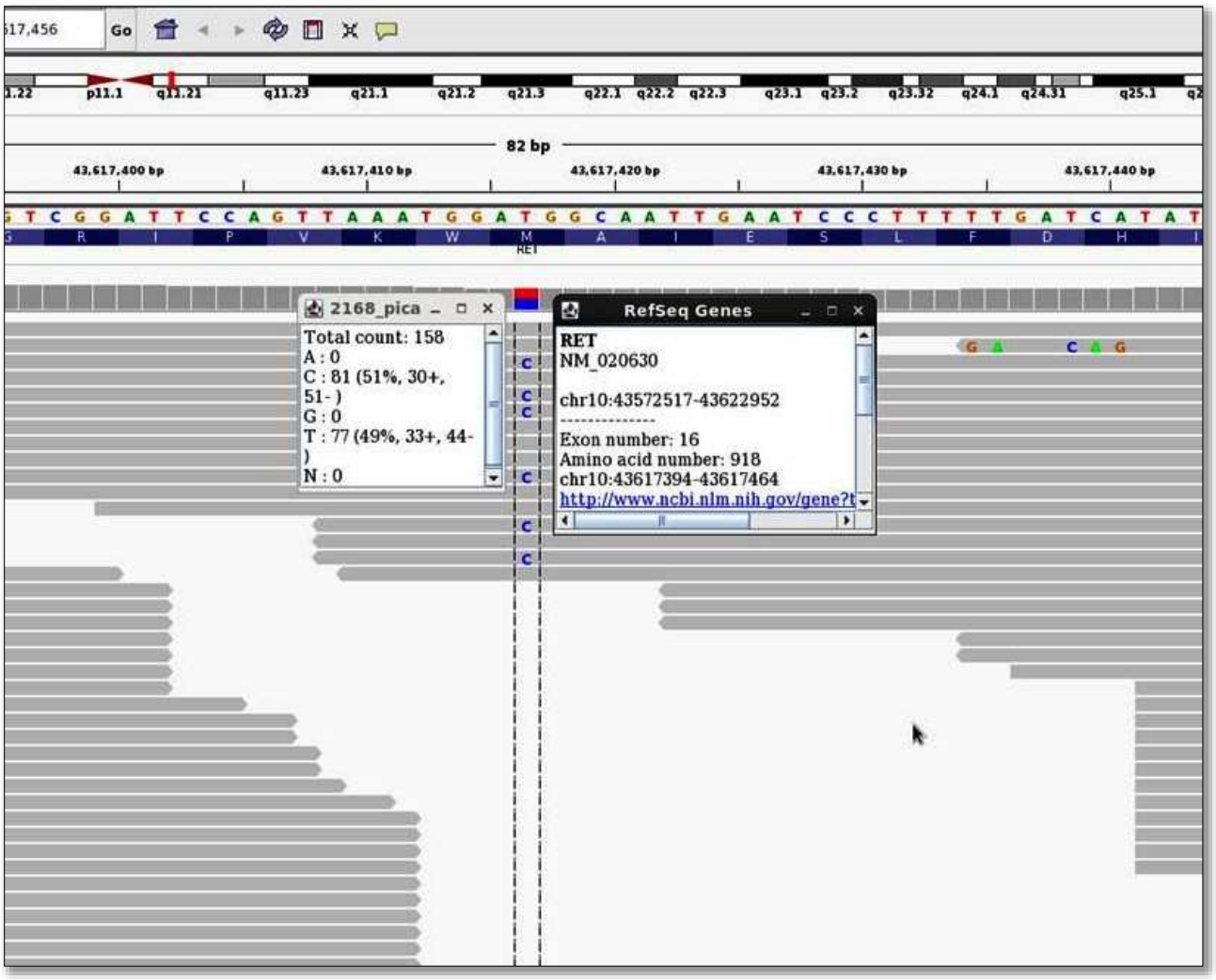


Figure 6.9: Exome data (.BAM file) visualization on Integrated Genome Viewer (IGV) showing read count of Wildtype (T) & Variant (C) allele for **p.M918T** mutation identified in exon 16 of RET gene.

Total no. of aligned reads: *158*

Wildtype allele (T) frequency: 49% (77 reads)

Variant allele (C) frequency: 51% (81 reads)

Figure 6.10: Exome data (.BAM file) visualization on Integrated Genome Viewer (IGV) showing read count of Wild-type (T) & Variant (C) allele for **p.M918T** mutation identified in exon 16 of RET gene.

Total no. of aligned reads: 93

Wildtype allele (T) frequency: 51% (47 reads)

Variant allele (C) frequency: 48% (45 reads)

Exome sequencing results confirmed our suspicion of the initial false negative genotyping result. The reason identified was allele dropout, a PCR based genotyping error where one of the two alleles of the gene is either underrepresented or completely absent. To confirm the results of exome sequencing, exon 16 primers were redesigned (Figure 6.11) and replicative Sanger sequencing was performed which identified the M918T mutation in all the three patients (Figure 6.12).

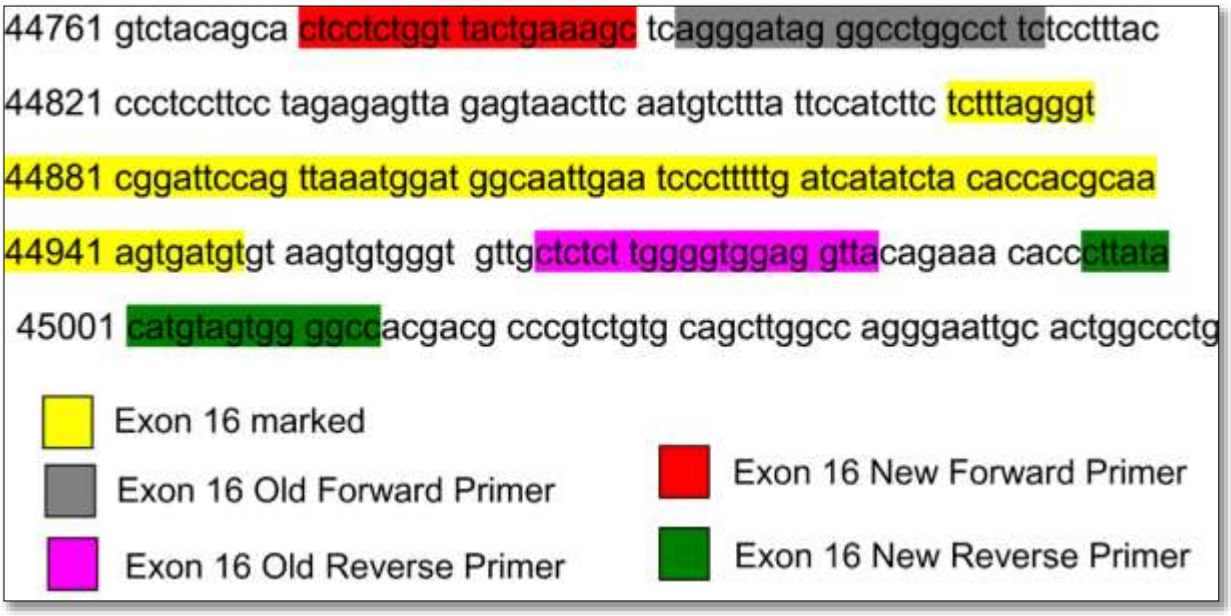


Figure 6.11: Sequence of Exon 16 of RET gene highlighted with region of old and new primers sites

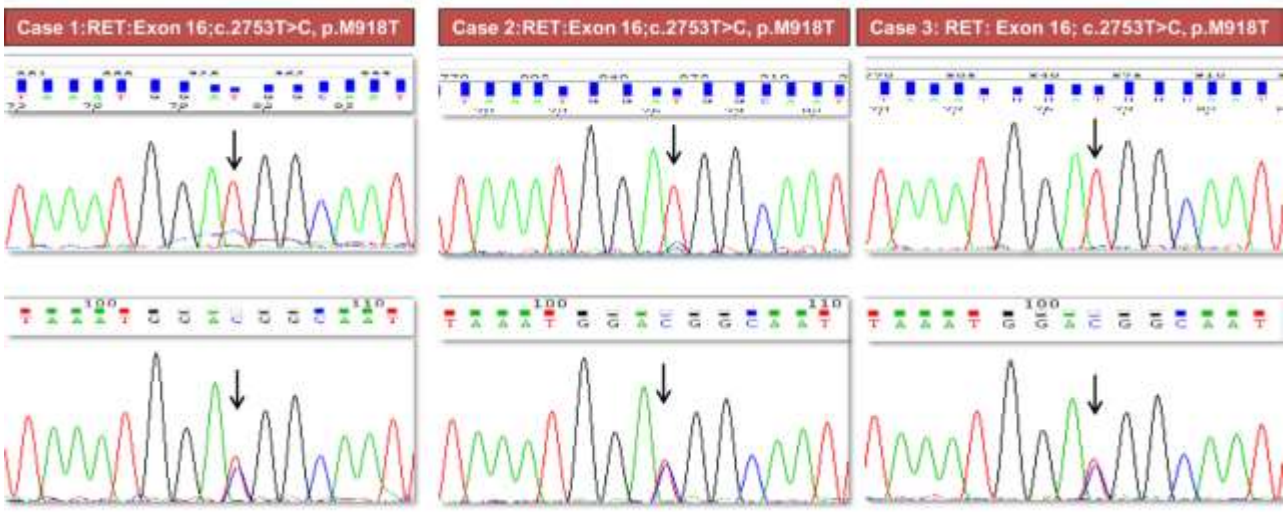


Figure 6.12: RET M918T allele dropout. Upper panels: M918T mutation not seen on Initial Sanger sequencing. Lower panels: M918T mutation identified on replicative Sanger Sequencing.

Various known causes of allele dropout were systematically evaluated. Any variant in the annealing region of exon 16 primers which may have hampered primer annealing was ruled out. For this, we sequenced the 80bp region of exon 16 and flanking regions that covered the annealing region of the primers used initially. We ruled out G-Quadruplexes by the QGRS mapper tool (160), hairpin structures by Oligocalc: Oligonucleotide properties calculator (http://biotools.nubic.northwestern.edu/OligoCalc.html) and CpG methylation sites by Sequence Manipulation suite: CpG islands(http://www.bioinformatics.org/sms2/cpg_islands.html). To identify the G-Quadruplexes, 300bp sequence of the exon 16 of RET and its flanking region covering both the old and redesigned primer sequences was submitted to QGRS mapper tool. We identified 17 QGRS sequences (including overlaps) and 3 QGRS sequences (without overlaps) but with very low QGRS score suggesting them to be neutral sequences and not involved in allele dropout. (Table 6.1 and Table 6.2)

Table 6.1: QGRS sequence in exon 16 of RET without overlaps.

Position	Length	QGRS	G-Score
18	30	**GGTTACTGAAAGCTCAGGGATAGGGCCTGG**	11
117	26	**GGGTCGGATTCCAGTTAAATGGATGG**	10
212	10	**GGGGTGGAGG**	20

Table 6.2: QGRS sequence in exon 16 of RET including overlaps.

Position	Length	QGRS	G-Score
18	30	**GGTTACTGAAAGCTCAGGGATAGGGCCTGG**	11
18	30	GGTTACTGAAAGCTCAGGGATAGGGCCTGG	10
18	30	GGTTACTGAAAGCTCAGGGATAGGGCCTGG	9
18	30	GGTTACTGAAAGCTCAGGGATAGGGCCTGG	9
117	26	**GGGTCGGATTCCAGTTAAATGGATGG**	10
118	25	GGTCGGATTCCAGTTAAATGGATGG	10
197	22	GGGTGTTGCTCTCTTGGGGTGG	8
197	25	GGGTGTTGCTCTCTTGGGGTGGAGG	8
197	25	GGGTGTTGCTCTCTTGGGGTGGAGG	9
197	25	GGGTGTTGCTCTCTTGGGGTGGAGG	8
197	25	GGGTGTTGCTCTCTTGGGGTGGAGG	7
198	21	GGTGTTGCTCTCTTGGGGTGG	9
198	24	GGTGTTGCTCTCTTGGGGTGGAGG	9
198	24	GGTGTTGCTCTCTTGGGGTGGAGG	10
198	24	GGTGTTGCTCTCTTGGGGTGGAGG	9
198	24	GGTGTTGCTCTCTTGGGGTGGAGG	8
212	10	**GGGGTGGAGG**	20

Since a specific mutant allele was dropped out in all four patients, possibility of this M918T mutation itself causing a hairpin or a secondary structure that may cause allele dropout was examined. For secondary structure prediction, 'Predict a secondary structure web server' was used (https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html). No secondary structure or any significant differences in hairpins between two alleles was observed.

After ruling out sequence dependent factors as a potential cause of allele dropout, we investigated various sequence independent factors sequentially. Replicative Sanger sequencing on the same DNA samplesas used for exome sequencing using old primer sets could identify the M918T mutation that was missed initially. This suggested that this allele dropout was a result of some sequence independent factors interfering with PCR amplification of the mutant allele during the initial rounds of Sanger sequencing performed several years ago. To investigate this, several variables were examined sequentially. First we checked the quality of DNA and then performed PCRs with different concentration of DNA (25, 37.5, 50, 75 and 100ng)as variations in DNA concentration & quality is also reported to be a cause of allele dropout. Further, different primer annealing temperature with a gradient PCR (65°C to 55°C); different concentration of dNTPs and PCR buffer; and spiking with different concentration of normal DNA was done. But this time the M918T mutation could be reliably identified with all different conditions described above.

6.4 Discussion

Identification of genetic variants in cancer have not only helped in molecular sub classification of different cancers but have also revealed the hereditary nature of certain cancers. Although the burden of hereditary cancer is less & accounts for 5% of all cancers(30), it is important to identify

such patients where appropriate genetic testing helps to identify a pathogenic or likely pathogenic mutation in a relevant gene for genetic counseling and risk management of the proband and for extended family testing.

For highly penetrant and potentially lethal hereditary diseases like MEN2, a false negative genetic test report may increase the risk of cancer morbidity and mortality for the proband as well as the healthy carriers in the family who do not get the benefit of reflex genetic testing(161). While failure to identify germline *RET* mutation in MEN2 syndrome cases has been reported in several studies(66,156,162–164), we report a systematic investigation for suspected false negative *RET* gene testing. The cause identified in our study was Allele Dropout, a genotyping error resulting from preferential amplification of one of the two alleles of a gene where the other allele is either under-represented or completely absent(165). Allele dropout can result from either allele specific sequence variations or sequence independent factors (Table 6.3).

Table 6.3: Types and causes of Allele Dropout

Type	Cause	Mechanism	Reference
Sequence Dependent Allele Dropout	Single Nucleotide variant (SNV) in the primer annealing region	Prevents annealing of the primer leading to amplification failure of that particular allele.	Haque 2018, Ward 2006, Laios 2008
	Non-primer annealing site SNV	Causes a strong secondary hairpin structure formation of the PCR products and leads to amplification failure	Lam 2013

	G-Quadruplexes in the GC-rich region	Causes arrest of DNA polymerase during PCR amplification	Stevans 2014, Stevans 2017
	Methylated cytosines at CpG Islands	Causes polymerase arrest on the methylated template during PCR amplification leading to allele dropout	Stevans 2017
Sequence Independent Allele Dropout	Sample Quality/Quantity	Insufficient or degraded DNA for PCR amplification	Nagy 2005, Hedell 2015
	Biochemical Artifacts	Variation in Taq polymerase quality/activity or low quality reagents	Pompanon 2005, Blais et al
	Equipment Artifacts	Imprecisions in thermal cycler temperature, evaporation or concentration of reaction mixture	Pompanon 2005, Blais et al
	Experimental Errors	Pipetting errors or sample contamination	Pompanon 2005, Blais et al

Our study highlights the importance of considering characteristic syndromic features in known monogenetic conditions while interpreting genetic test results. As these four cases had characteristic MEN2B phenotype, their *RET* negative report raised a strong suspicion of it being false negative. There could be several reasons for genotyping errors and a false negative genetic analysis report. These include (i) Pre-Analytical errors due to sample mix up (166), 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 samples, and rarely due to Allele Dropout; (iii) Post analytical errors of incorrect interpretation arising from human interpretative errors or use of inappropriate bio-informatic approaches or *in-silico* tools or databases; (iv) Mosaicism with low copy number or absence of mutation in tissues tested(167,168). Taking into consideration various underlying causes, we propose an algorithm for suspecting and evaluating genotyping errors, including allele dropout (Figure 6.13).

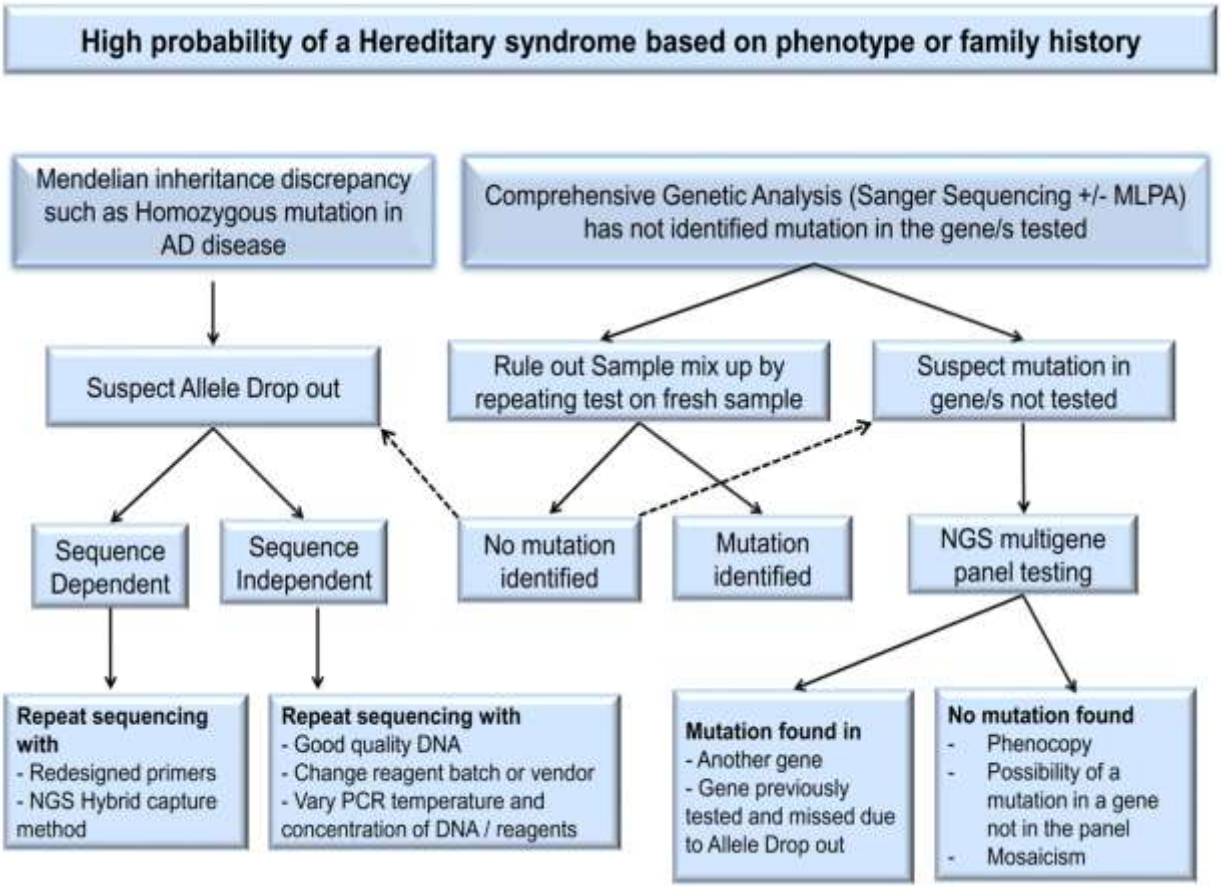


Figure 6.13: Algorithm for investigating suspected genotyping errors in Hereditary syndromes.AD: Autosomal Dominant, MLPA: Multiple Ligation dependent Probe Amplification.

In this algorithm we suggest that in cases with an unequivocal clinical diagnosis of a specific hereditary syndrome based on specific phenotypic features or strong family history, negative genetic test result could be false negative. Suspected false negative genetic test result may be resolved using the approach mentioned in the proposed algorithm. Using this algorithm in our cases, we first ruled out pre-analytical errors by replicative Sanger sequencing on an independent DNA sample collected with full traceability. Further we ruled out 2-5% probability of a single or double mutation in the remaining 14 exons of *RET* which could also explain the MEN2B phenotype (53). As this failed to identify any *RET* mutation in 3 cases, expanded genetic analysis to identify mutation in other gene was considered. As none of the genes included in the commonly used NGS multi-gene cancer panels have ever been associated with MEN2 syndrome, we directly performed exome sequencing. While in our study, whole exome sequencing led to the serendipitous finding of a *RET* Allele dropout, in few other studies on *RET* negative MTC cases which are mostly represented by familial MTC kindreds (102,162,163), exome sequencing identified germline mutations in other genes which may have a role in inherited predisposition to MTC. Recently, Smith et al., (162) reported a novel frameshift mutationc.948delT, p.Gly318Alafs*22 in *ESR2* gene in a family with two generations affected with MTC/CCH.The mutation was identified in affected individuals as well as the family members who were diagnosed with C-cell hyperplasia and was absent in individuals with normal thyrioid histology. In another Italian family with two siblings affected with MTC, exome sequencing revealed a possibly pathogenic germline *MET* proto-oncogene mutationp.Arg417Glnin its extracellular Sema domain(163). The authors have shown that ectopic expression of *MET* p.Arg417Gln in *MET*-negative T47D breast cancer cells restores mutant receptor's functionality and its ability to enhance cell migration and invasion(163). Pathogenic mutations in*ESR2* or *MET* gene were not

identified on exome analysis in our MEN2B cases, though a few Single Nucleotide Polymorphisms (SNPs) were identified in both the genes.

Several causes of sequence dependent Allele dropout have been reported in literature. These include single nucleotide variant (SNV) within primer binding region(169–171), SNV at non-primer binding site which results in formation of hairpin structure (172), G Quadruplex formation due to high GC content (173) or methylation at CpG islands which stabilizes the G quadruplexes (174). In contrast, causes of sequence independent allele dropout are difficult to predict or reproduce. These include low template or poor quality DNA(175); biochemical artifacts such as variation in Taq polymerase quality or activity (176); low quality reagents or equipment artifacts such as imprecisions in thermal cycler temperature; evaporation or concentration of reaction mixture; or experimental errors in pipetting or sample contamination (176).

Sequence independent Allele dropout are generally non reproducible and may be uncovered by replicative analysis with same or different PCR conditions (165). While we were able to confirm allele dropout in our cases, its cause could not be identified despite changing several PCR conditions and varying the concentration of different reagents and DNA. The only variable which could not be matched in replicative analysis was the batch of reagents as several years had lapsed between the initial Sanger Sequencing with false negative report and the replicative Sanger Sequencing which identified the allele dropout. It is therefore quite likely that allele dropouts observed in our study was due to the quality of reagents used for initial rounds of sequencing or due to equipment artifacts. Our failure to identify the exact cause of the Allele dropout despite a very systematic examination is not surprising as others have also reported that sequence independent Allele Dropouts are usually unpredictable, non reproducible and their exact cause remains unknown (165).

Of the 18 cases of germline allele dropouts reported so far in a cancer predisposing gene (169,171,177–181) (Table 6.4) only one was sequence independent (169) and was reported in the p53gene. Since this was a sequence independent allele dropout, the exact cause of it remained unknown. The only Allele dropout reported previously in *RET* gene was a sequence dependent dropout of the wildtype allele resulting in the *RET* Y791F mutation appearing incorrectly as homozygous germline mutation(171). This is the first series of four cases of clinically relevant sequence independent allele dropout in any cancer predisposing gene.

Table 6.4: Reports of germline Allele Dropout in cancer predisposing gene

Germline Allele Dropout in cancer predisposing gene	**No. of Cases with allele dropout**	**Gene/s involved**	**Type of Allele Dropout**	**Mechanism of Allele dropout**
Coulet et al 2010 (179)	3 cases	BRCA1, BRCA2	Sequence dependent Allele dropout	**Case 1**: SNP at primer binding region (BRCA2) **Case 2** : 104bp exonic deletion (BRCA1) **Case 3** : Insertion of Alu sequence (BRCA2)
Mullins et al. 2007 (178)	1 case	CDH1	Sequence dependent Allele dropout	SNP at primer binding region
Haque et al. 2018 (169)	9 cases	p53	Both Sequence dependent and Sequence Independent Allele dropout	**8 cases**: SNP at primer binding region leading to sequence dependent allele dropout. **1 case:** Sequence independent allele dropout due to unknown genotypic error.

Rondot et al. 2004 (171)	1 case	RET	Sequence dependent Allele dropout	SNP at primer binding region
Silva et al. 2017 (177)	1 case	BRCA1	Sequence dependent Allele dropout	SNP at primer binding region
Chong et al. 2014 (180)	2 cases	BRCA1	Sequence dependent Allele dropout	SNP at primer binding region
Ellison et al.2017 (181)	1 case	BRCA2	Sequence dependent Allele dropout	Deletion at the primer binding region

The only comprehensive study determining the frequency of allele dropout in clinical diagnostic setting was carried out by Blais et al in 2015 in which they assessed the risk of misdiagnosis in PCR based genotyping approaches by analyzing 30,769 patient's genotypes. Their study identified allele dropout in 135 (0.44%) of the 30769 genotypes and affected 1:227 genotype which shows that it is a rare event. They have also shown that majority of times (94%) these artifacts are associated with non reproducible PCR failures rather than sequence specific variants suggesting that careful design of PCR primers cannot always prevent such errors. Sequence dependent allele dropouts were also rare in their study as the genotyping was done with allele specific assays.Considering the extreme rarity of reported germline allele dropouts in cancer predisposing genes, it is very likely that these are not suspected or investigated as a cause of false negative genetic test report.

Based on this series of four cases with allele dropout we raise an important alert about their existence and the need for systematically preventing, suspecting and confirming allele dropouts in different clinical and technical contexts. If suspected false negative testing status remains unresolved on replicative Sanger Sequencing or mutation in other genes is a possibility, multi-gene panel or clinical exome analysis using hybrid capture with deep target coverage can be considered. This could help in detecting and minimizing genotyping errors which are observed in PCR based genotyping assays. In hybrid capture based assay sequence specific probes are used which are specific for target regions in the genome. Since these probes are longer in size than PCR primers they can tolerate single nucleotide mismatches in their target binding region preventing genotyping errors like allele dropout. Considering that MEN2 is a high penetrance monogenetic disorder and the known occurrence of genotyping errors, MEN2 families previously reported to be *RET* negative may be retested for *RET* mutations with Sanger sequencing or directly with NGS multi-gene panels including *RET* gene.

In conclusion, our study serves as an alert to systematically address the issue of sequence independent allele dropout in cases with Mendelian inheritance discrepancy. Through our proposed algorithm we explain when to suspect a false negative genetic test report which may require retesting of cases by appropriate methods.This study also highlights the importance of syndromic diagnosis based on phenotypic characterization, the strongest indication of genotyping errors such as allele dropout in cases showing Mendelian inheritance discrepancy.

Chapter 7

Conclusion & Future Directions

In the largest Indian cohort of MTC cases reported so far, through comprehensive genotype-phenotype correlation studies, we have established the mutation spectrum of RET and its phenotypic manifestations. Our findings of specific RET gene mutations being associated with more aggressive and early onset MTC in Indian families, as has been previously reported in the Caucasian population, will help in clinical management of MTC patients and risk management of pre-symptomatic RET mutation carriers.

We have studied the modulatory role of 13 SNPs of 3 distinct genetic pathways in a cohort of 436 Indian MTC cases with risk of MTC development as well as its clinical behavior. This is the largest MTC cohort ever reported and unlike previous studies, examines all three important genetic pathways. We show for the first time a significant protective association of *CDKN1A* Ser31Arg SNP with MTC development. Further, in our meta-analysis we have identified a significant risk association of all the 4 *RET* SNPs with MTC. It is noteworthy that these significant associations were not observed in previously reported meta-analysis.

We report the first comprehensive somatic genomic analysis on MTC tumors in patients with germline RET driven hereditary MTC. High quality exome analysis of paired normal and tumor tissue revealed an extremely low somatic mutation burden of 0.124/Mb similar to some rare forms of single gene or single genetic pathway driven brain tumors including pilocytic astrocytoma and rhabdoid tumor.

Our study has also for the first time reported a series of clinically relevant sequence independent allele dropout in any cancer predisposing gene which we identified in 3 classical hereditary MTC cases. This study highlights the importance of syndromic diagnosis based on phenotypic characterization, the strongest indication of genotyping errors such as allele dropout in cases

showing Mendelian inheritance discrepancy. Based on a systematic review of literature, we have also developed an algorithm for suspecting and investigating different types of genotyping errors including allele dropout. Through our proposed algorithm we explain when to suspect a false negative genetic test report which may require retesting of cases by appropriate methods.

The molecular mechanism through which the CDKN1A SNP or RET gene SNPs modifies the risk of MTC development needs to be further elucidated through functional studies. Although germline RET driven MTC is a rare tumor with very low tumor mutation burden, comprehensive somatic genomic analysis on a larger cohort of such tumors may identify recurrently mutated gene and the genomic landscape of germline RET driven MTC.

Chapter 8

Bibliography

1. Hejmadi M. Introduction to cancer biology [Internet]. Frederiksberg, Denmark: BoonBooks.com; 2010. Available from: http://bookboon.com/int/student/biology/introduction-to-cancer-biology

2. 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 Cancer J Clin. 2018 Nov;68(6):394–424.

3. Dhillon PK, Mathur P, Nandakumar A, Fitzmaurice C, Kumar GA, Mehrotra R, et al. The burden of cancers and their variations across the states of India: the Global Burden of Disease Study 1990–2016. Lancet Oncol. 2018 Oct;19(10):1289–306.

4. 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. Lancet Oncol. 2014 May;15(6):e205-12.

5. Pott P. Chirurgical observations : relative to the cataract, the polypus of the nose, the cancer of the scrotum, the different kinds of ruptures, and the mortification of the toes and feet [Internet]. London: Printed by T.J. Carnegy, for L. Hawes, W. Clarke, and R. Collins; 1775. Available from: http://galenet.galegroup.com/servlet/ECCO?c=1&stp=Author&ste=11&af=BN&ae=T007127&tiPG=1&dd=0&dc=flc&docNum=CW107455718&vrsn=1.0&srchtp=a&d4=0.33&n=10&SU=0LRM&locID=ureginalib

6. Bell J. Paraffin Epithelioma of the Scrotum. Edinb Med J. 1876 Aug;22(2):135–7.

7. Hollander AW. Development of dermatopathology and Paul Gerson Unna. J Am Acad

Dermatol. 1986 Oct;15(4 Pt 1):727–34.

8. Hill J. Cautions against the immoderate use of snuff. Founded on the known qualities of the tobacco plant;...and enforced by instances of persons who have perished miserably of diseases, occasioned...by its use. [Internet]. Medicine, Science and Technology. London : printed for R. Baldwin; 1761. Available from: http://lib.ugent.be/catalog/rug01:001097560

9. DOLL R, HILL AB. Smoking and carcinoma of the lung; preliminary report. Br Med J. 1950 Sep;2(4682):739–48.

10. Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc Natl Acad Sci U S A. 1980 Dec;77(12):7415–9.

11. Beasley RP, Hwang LY, Lin CC, Chien CS. Hepatocellular carcinoma and hepatitis B virus. A prospective study of 22 707 men in Taiwan. Lancet (London, England). 1981 Nov;2(8256):1129–33.

12. Ferguson AR. Associated bilharziosis and primary malignant disease of the urinary bladder, with observations on a series of forty cases. J Pathol Bacteriol [Internet]. 1911 Jan 1;16(1):76–94. Available from: https://doi.org/10.1002/path.1700160107

13. Huang JQ, Sridhar S, Chen Y, Hunt RH. Meta-analysis of the relationship between Helicobacter pylori seropositivity and gastric cancer. Gastroenterology. 1998 Jun;114(6):1169–79.

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

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

16. Cancer Classification [Internet]. Encyclopedic Dictionary of Genetics, Genomics and Proteomics. 2004. Available from: http://doi.wiley.com/10.1002/0471684228.egp01697

17. Pinter D, Sumowski J, DeLuca J, Fazekas F, Pichler A, Khalil M, et al. Higher education moderates the effect of T2 lesion load and third ventricle width on cognition in multiple sclerosis [Internet]. Vol. 9, PLoS ONE. 2014. p. 2–3. Available from: https://www.cancer.net/cancer-types

18. Dr. Ananya Mandal M. Cancer Classification [Internet]. Encyclopedic Dictionary of Genetics, Genomics and Proteomics. 2004. Available from: https://www.news-medical.net/health/Cancer-Classification.aspx

19. Essays U. "Hereditary Vs Sporadic Cancer: Analysis." Available from: https://www.ukessays.com/essays/biology/study-on-hereditary-vs-sporadic-cancer-biology-essay.php?vref=1

20. Venugopal R, Bavle RM, Konda P, Muniswamappa S, Makarla S. Familial Cancers of Head and Neck Region. J Clin Diagn Res. 2017 Jun;11(6):ZE01–6.

21. Amor D. Familial cancers. An overview. Aust Fam Physician. 2001 Oct;30(10):937–45.

22. Goldberg YP, Parker MI, Gevers W. The genetic basis of cancer. S Afr Med J. 1991 Jul;80(2):99–104.

23. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med. 2004 Aug;10(8):789–99.

24. Bommer U-A, Perrow KL. Cancer biology : Molecular and genetic basis. Cancer Counc Austrailia [Internet]. 2014;1–16. Available from: https://wiki.cancer.org.au/oncologyformedicalstudents/Cancer_biology:_Molecular_and_genetic_basis

25. Lee EYHP, Muller WJ. Oncogenes and tumor suppressor genes. Cold Spring Harb Perspect Biol. 2010 Oct;2(10):a003236.

26. Chail H. Proto-oncogenes to oncogenes to cancer. Nat Educ [Internet]. 2008; Available from: http://www.nature.com/scitable/topicpage/proto-oncogenes-to-oncogenes-to-cancer-883

27. Lodish H, Berk A, Zipursky SL et al. The Actin Cytoskeleton - Molecular Cell Biology - 4th edition. In: Freeman W.H., editor. Molecular Cell Biology 4th edition New York: W H Freeman [Internet]. Section 24. New York; 2000. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21662/

28. Cooper GM, Hausman RE. The Cell: A Molecular Approach 2nd Edition. In: Sinauer Associates [Internet]. 2nd editio. Sunderland (MA): Sinauer Associates; 2007. p. 1–820. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9894/

29. Chial BH. Tumor Suppressor (TS) Genes and the Two-Hit Hypothesis. Nat Educ [Internet]. 2008;1:1–5. Available from: https://www.nature.com/scitable/topicpage/tumor-suppressor-ts-genes-and-the-two-887

30. Rahner N, Steinke V. Hereditary cancer syndromes. Dtsch Arztebl Int. 2008 Oct;105(41):706–14.

31. Susan C. Landis. Nursing Assessment: Endocrine System. In. Available from: https://nursekey.com/nursing-assessment-endocrine-system/

32. Paulsen DF. Histology & Cell Biology: Examination & Board Review. In: 5th Editio. Available from: https://accessmedicine.mhmedical.com/book.aspx?bookid=563

33. Panawala L. Difference Between Thyroid and Parathyroid. 2017.

34. American Thyroid Association. Thyroid Cancer (Papillary and Follicular). Available from: https://www.thyroid.org/thyroid-cancer/

35. Lim H, Devesa SS, Sosa JA, Check D, Kitahara CM. Trends in Thyroid Cancer Incidence and Mortality in the United States, 1974-2013Thyroid Cancer Incidence and Mortality TrendsThyroid Cancer Incidence and Mortality Trends. JAMA [Internet]. 2017 Apr 4;317(13):1338–48. Available from: https://doi.org/10.1001/jama.2017.2719

36. Sekkath Veedu J, Wang K, Lei F, Chen Q, Huang B, Mathew A. Trends in thyroid cancer incidence in India. J Clin Oncol [Internet]. 2018 May 20;36(15_suppl):e18095–e18095. Available from: https://doi.org/10.1200/JCO.2018.36.15_suppl.e18095

37. Jaquet AJ. Ein Fall von metastasierenden Amyloidtumoren (Lymphosarcoma). Virchows Arch A Pathol Anat Histol 185251-267. 1906;

38. HAZARD JB, HAWK WA, CRILE GJ. Medullary (solid) carcinoma of the thyroid; a clinicopathologic entity. J Clin Endocrinol Metab. 1959 Jan;19(1):152–61.

39. Williams ED. Histogenesis of medullary carcinoma of the thyroid. J Clin Pathol. 1966 Mar;19(2):114–8.

40. Melvin KE, Miller HH, Tashjian AHJ. Early diagnosis of medullary carcinoma of the thyroid gland by means of calcitonin assay. N Engl J Med. 1971 Nov;285(20):1115–20.

41. Melvin KE, Tashjian AHJ. The syndrome of excessive thyrocalcitonin produced by medullary carcinoma of the thyroid. Proc Natl Acad Sci U S A. 1968 Apr;59(4):1216–22.

42. Kebebew E, Ituarte PH, Siperstein AE, Duh QY, Clark OH. Medullary thyroid carcinoma: clinical characteristics, treatment, prognostic factors, and a comparison of staging systems. Cancer. 2000 Mar;88(5):1139–48.

43. Roman S, Lin R, Sosa JA. Prognosis of medullary thyroid carcinoma: demographic, clinical, and pathologic predictors of survival in 1252 cases. Cancer. 2006 Nov;107(9):2134–42.

44. Lodish MB, Stratakis CA. RET oncogene in MEN2, MEN2B, MTC and other forms of thyroid cancer. Expert Rev Anticancer Ther. 2008 Apr;8(4):625–32.

45. Kloos RT, Eng C, Evans DB, Francis GL, Gagel RF, Gharib H, et al. Medullary Thyroid Cancer : Management Guidelines of the American Thyroid Association. 2009;19(6).

46. Sipple JH. The association of pheochromocytoma with carcinoma of the thyroid gland. Am J Med [Internet]. 1961 Jul 1;31(1):163–6. Available from: https://doi.org/10.1016/0002-9343(61)90234-0

47. Frank-Raue K, Rondot S, Raue F. Molecular genetics and phenomics of RET mutations: Impact on prognosis of MTC. Mol Cell Endocrinol. 2010;322(1–2):2–7.

48. Frank-Raue K, Rondot S, Schulze E, Raue F. Change in the spectrum of RET mutations diagnosed between 1994 and 2006. Clin Lab [Internet]. 2007;53(5–6):273—282.

Available from: http://europepmc.org/abstract/MED/17605401

49. Gagel RF, Levy ML, Donovan DT, Alford BR, Wheeler T, Tschen JA. Multiple endocrine neoplasia type 2a associated with cutaneous lichen amyloidosis. Ann Intern Med. 1989 Nov;111(10):802–6.

50. Borst MJ, VanCamp JM, Peacock ML, Decker RA. Mutational analysis of multiple endocrine neoplasia type 2A associated with Hirschsprung's disease. Surgery. 1995 Apr;117(4):386–91.

51. Eng C. The Relationship Between Specific RET Proto-oncogene Mutations and Disease Phenotype in Multiple Endocrine Neoplasia Type 2. Jama [Internet]. 1996;276(19):1575. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.1996.03540190047028

52. Elisei R, Romei C, Cosci B, Agate L, Bottici V, Molinaro E, et al. RET genetic screening in patients with medullary thyroid cancer and their relatives: experience with 807 individuals at one center. J Clin Endocrinol Metab. 2007 Dec;92(12):4725–9.

53. Wells SA, Asa SL, Dralle H, Elisei R, Evans DB, Gagel RF, et al. Revised American Thyroid Association Guidelines for the Management of Medullary Thyroid Carcinoma. Thyroid [Internet]. 2015;25(6):567–610. Available from: http://online.liebertpub.com/doi/10.1089/thy.2014.0335

54. Sippel RS, Kunnimalaiyaan M, Chen H. Current management of medullary thyroid cancer. Oncologist. 2008 May;13(5):539–47.

55. Correia-Deur JEM, Toledo RA, Imazawa AT, Lourenço DMJ, Ezabella MCL, Tavares

MR, et al. Sporadic medullary thyroid carcinoma: clinical data from a university hospital. Clinics (Sao Paulo). 2009 May;64(5):379–86.

56. De Groot JWB, Links TP, Plukker JTM, Lips CJM, Hofstra RMW. RET as a diagnostic and therapeutic target in sporadic and hereditary endocrine tumors. Endocr Rev. 2006;27(5):535–60.

57. Takahashi M, Ritz J, Cooper GM. Activation of a novel human transforming gene, ret, by DNA rearrangement. Cell. 1985 Sep;42(2):581–8.

58. Schuchardt A, D'Agati V, Larsson-Blomberg L, Costantini F, Pachnis V. Defects in the kidney and enteric nervous system of mice lacking the tyrosine kinase receptor Ret. Nature. 1994 Jan;367(6461):380–3.

59. Arighi E, Borrello MG, Sariola H. RET tyrosine kinase signaling in development and cancer. Cytokine Growth Factor Rev. 2005;16(4–5):441–67.

60. Eng C, Clayton D, Schuffenecker I, Al E. The relationship between specific ret proto-oncogene mutations and disease phenotype in multiple endocrine neoplasia type 2: International ret mutation consortium analysis. JAMA [Internet]. 1996 Nov 20;276(19):1575–9. Available from: http://dx.doi.org/10.1001/jama.1996.03540190047028

61. Donis-Keller H, Dou S, Chi D, M. Carlson K, Toshima K, C. Lairmore T, et al. Mutations in the RET proto-oncogene are associated with Men2A and FMTC. Vol. 2, Human molecular genetics. 1993. 851–856 p.

62. Eng C, Mulligan LM, Smith DP, Healey CS, Frilling A, Raue F, et al. Low frequency of

germline mutations in the RET proto-oncogene in patients with apparently sporadic medullary thyroid carcinoma. Clin Endocrinol (Oxf). 1995 Jul;43(1):123–7.

63. Agrawal N, Jiao Y, Sausen M, Leary R, Bettegowda C, Roberts NJ, et al. Exomic sequencing of medullary thyroid cancer reveals dominant and mutually exclusive oncogenic mutations in RET and RAS. J Clin Endocrinol Metab. 2013 Feb;98(2):E364-9.

64. Mulligan LM, Kwok JB, Healey CS, Elsdon MJ, Eng C, Gardner E, et al. Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. Nature. 1993 Jun;363(6428):458–60.

65. Bolino A, Schuffenecker I, Luo Y, Seri M, Silengo M, Tocco T, et al. RET mutations in exons 13 and 14 of FMTC patients. Oncogene. 1995 Jun;10(12):2415–9.

66. Eng C, Smith DP, Mulligan LM, Nagai MA, Healey CS, Ponder MA, et al. Point mutation within the tyrosine kinase domain of the RET proto-oncogene in multiple endocrine neoplasia type 2B and related sporadic tumours. Hum Mol Genet [Internet]. 1994;3(2):237–41. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7911697

67. Wells SAJ, Pacini F, Robinson BG, Santoro M. Multiple endocrine neoplasia type 2 and familial medullary thyroid carcinoma: an update. J Clin Endocrinol Metab. 2013 Aug;98(8):3149–64.

68. Margraf RL, Crockett DK, Krautscheid PMF, Seamons R, Calderon FRO, Wittwer CT, et al. Multiple endocrine neoplasia type 2 RET protooncogene database: repository of

MEN2-associated RET sequence variation and reference for genotype/phenotype correlations. Hum Mutat. 2009 Apr;30(4):548–56.

69. Takahashi M, Cooper GM. ret transforming gene encodes a fusion protein homologous to tyrosine kinases. Mol Cell Biol. 1987 Apr;7(4):1378–85.

70. Toledo SP de A, dos Santos MACG, Toledo R de A, Lourenco DMJ. Impact of RET proto-oncogene analysis on the clinical management of multiple endocrine neoplasia type 2. Clinics (Sao Paulo). 2006 Feb;61(1):59–70.

71. Raue F, Frank-Raue K. Genotype-phenotype correlation in multiple endocrine neoplasia type 2. Clinics (Sao Paulo). 2012;67 Suppl 1(Suppl 1):69–75.

72. Wagner SM, Zhu S, Nicolescu AC, Mulligan LM. Molecular mechanisms of RET receptor-mediated oncogenesis in multiple endocrine neoplasia 2. Clinics (Sao Paulo). 2012;67 Suppl 1(Suppl 1):77–84.

73. Borrello MG, Smith DP, Pasini B, Bongarzone I, Greco A, Lorenzo MJ, et al. RET activation by germline MEN2A and MEN2B mutations. Oncogene. 1995 Dec;11(11):2419–27.

74. Carlomagno F, Melillo RM, Visconti R, Salvatore G, De Vita G, Lupoli G, et al. Glial cell line-derived neurotrophic factor differentially stimulates ret mutants associated with the multiple endocrine neoplasia type 2 syndromes and Hirschsprung's disease. Endocrinology. 1998 Aug;139(8):3613–9.

75. Arighi E, Popsueva A, Degl'Innocenti D, Borrello MG, Carniti C, Perala NM, et al. Biological effects of the dual phenotypic Janus mutation of ret cosegregating with both

multiple endocrine neoplasia type 2 and Hirschsprung's disease. Mol Endocrinol. 2004 Apr;18(4):1004–17.

76. Randle RW, Balentine CJ, Leverson GE, Havlena JA, Sippel RS, Schneider DF, et al. Trends in the presentation, treatment, and survival of patients with medullary thyroid cancer over the past 30 years. Surgery [Internet]. 2016/11/11. 2017 Jan;161(1):137–46. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27842913

77. Chang C-F, Yang W-S, Su Y-N, Wu I-L, Chang T-C. Mutational spectrum of multiple endocrine neoplasia type 2 and sporadic medullary thyroid carcinoma in taiwan. J Formos Med Assoc. 2009 May;108(5):402–8.

78. Chung YJ, Kim H-H, Kim H-J, Min Y-K, Lee M-S, Lee M-K, et al. RET proto-oncogene mutations are restricted to codon 634 and 618 in Korean families with multiple endocrine neoplasia 2A. Thyroid. 2004 Oct;14(10):813–8.

79. Sarika H-L, Papathoma A, Garofalaki M, Saltiki K, Pappa T, Pazaitou-Panayiotou K, et al. Genetic screening of patients with medullary thyroid cancer in a referral center in Greece during the past two decades. Eur J Endocrinol. 2015 Apr;172(4):501–9.

80. Jindrichova S, Vcelak J, Vlcek P, Neradilova M, Nemec J, Bendlova B. Screening of six risk exons of the RET proto-oncogene in families with medullary thyroid carcinoma in the Czech Republic. J Endocrinol. 2004 Nov;183(2):257–65.

81. Moura MM, Cavaco BM, Pinto AE, Leite V. High prevalence of RAS mutations in RET-negative sporadic medullary thyroid carcinomas. J Clin Endocrinol Metab. 2011 May;96(5):E863-8.

82. Schulten H-J, Al-Maghrabi J, Al-Ghamdi K, Salama S, Al-Muhayawi S, Chaudhary A, et al. Mutational screening of RET, HRAS, KRAS, NRAS, BRAF, AKT1, and CTNNB1 in medullary thyroid carcinoma. Anticancer Res. 2011 Dec;31(12):4179–83.

83. Integrated genomic characterization of papillary thyroid carcinoma. Cell. 2014 Oct;159(3):676–90.

84. Guo C, McDowell IC, Nodzenski M, Scholtens DM, Allen AS, Lowe WL, et al. Transversions have larger regulatory effects than transitions. BMC Genomics. 2017 May;18(1):394.

85. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin A V, et al. Signatures of mutational processes in human cancer. Nature. 2013 Aug;500(7463):415–21.

86. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014 Jan;505(7484):495–501.

87. Kohashi K, Oda Y. Oncogenic roles of SMARCB1/INI1 and its deficient tumors. Cancer Sci. 2017 Apr;108(4):547–52.

88. Sredni ST, Tomita T. Rhabdoid tumor predisposition syndrome. Pediatr Dev Pathol. 2015;18(1):49–58.

89. Vitte J, Gao F, Coppola G, Judkins AR, Giovannini M. Timing of Smarcb1 and Nf2 inactivation determines schwannoma versus rhabdoid tumor development. Nat Commun. 2017 Aug;8(1):300.

90. Collins VP, Jones DTW, Giannini C. Pilocytic astrocytoma: pathology, molecular mechanisms and markers. Acta Neuropathol. 2015 Jun;129(6):775–88.

91. Qian L, Gao X, Huang H, Lu S, Cai Y, Hua Y, et al. PRSS3 is a prognostic marker in invasive ductal carcinoma of the breast. Oncotarget. 2017 Mar;8(13):21444–53.

92. Houle AA, Gibling H, Lamaze FC, Edgington HA, Soave D, Fave M-J, et al. Aberrant PRDM9 expression impacts the pan-cancer genomic landscape. Genome Res. 2018 Nov;28(11):1611–20.

93. Wiench M, Wygoda Z, Gubala E, Wloch J, Lisowska K, Krassowski J, et al. Estimation of risk of inherited medullary thyroid carcinoma in apparent sporadic patients. J Clin Oncol. 2001 Mar;19(5):1374–80.

94. Robledo M, Gil L, Polla M, Cebria A, Ruı S, Azan M, et al. Polymorphisms G691S / S904S of RET as Genetic Modifiers of MEN 2A 1. 2003;(8):1814–7.

95. Eng C, Mulligan LM, Healey CS, Houghton C, Frilling A, Raue F, et al. Heterogeneous mutation of the RET proto-oncogene in subpopulations of medullary thyroid carcinoma. Cancer Res. 1996 May;56(9):2167–70.

96. Marsh DJ, Learoyd DL, Andrew SD, Krishnan L, Pojer R, Richardson AL, et al. Somatic mutations in the RET proto-oncogene in sporadic medullary thyroid carcinoma. Clin Endocrinol (Oxf). 1996 Mar;44(3):249–57.

97. Ceolin L, Siqueira DR, Romitti M, Ferreira C V., Maia AL. Molecular basis of medullary thyroid carcinoma: The role of RET polymorphisms. Int J Mol Sci. 2012;13(1):221–39.

98. Stansifer KJ, Guynan JF, Wachal BM, Smith RB. Modifiable risk factors and thyroid

cancer. Otolaryngol Head Neck Surg. 2015 Mar;152(3):432–7.

99. Negri E, Ron E, Franceschi S, La Vecchia C, Preston-Martin S, Kolonel L, et al. Risk factors for medullary thyroid carcinoma: a pooled analysis. Cancer Causes Control. 2002 May;13(4):365–72.

100. Kalezic NK, Zivaljevic VR, Slijepcevic NA, Paunovic IR, Diklic AD, Sipetic SB. Risk factors for sporadic medullary thyroid carcinoma. Eur J Cancer Prev. 2013 May;22(3):262–7.

101. Lantieri F, Caroli F, Ceccherini I, Griseri P. The involvement of the RET variant G691S in medullary thyroid carcinoma enlightened by a meta-analysis study. Int J cancer. 2013 Jun;132(12):2808–19.

102. Figlioli G, Landi S, Romei C, Elisei R, Gemignani F. Medullary thyroid carcinoma (MTC) and RET proto-oncogene: Mutation spectrum in the familial cases and a meta-analysis of studies on the sporadic form. Mutat Res - Rev Mutat Res [Internet]. 2013;752(1):36–44. Available from: http://dx.doi.org/10.1016/j.mrrev.2012.09.002

103. Cebrian A, Lesueur F, Martin S, Leyland J, Ahmed S, Luccarini C, et al. Polymorphisms in the initiators of RET (rearranged during transfection) signaling pathway and susceptibility to sporadic medullary thyroid carcinoma. J Clin Endocrinol Metab. 2005 Nov;90(11):6268–74.

104. Severskaya N, Radiological M, Saenko V. Germline Polymorphisms of RET and GFRA1 Genes in Patients with Medullary Thyroid Carcinoma. 2006;(January).

105. Fernandez RM, Pecina A, Antinolo G, Navarro E, Borrego S. Analysis of RET

polymorphisms and haplotypes in the context of sporadic medullary thyroid carcinoma. Thyroid. 2006 Apr;16(4):411–7.

106. Fugazzola L, Muzza M, Mian C, Cordella D, Barollo S, Alberti L, et al. RET genotypes in sporadic medullary thyroid cancer: studies in a large Italian series. Clin Endocrinol (Oxf). 2008 Sep;69(3):418–25.

107. Sromek M, Czetwertynska M, Skasko E, Zielinska J, Czapczak D, Steffen J. The frequency of selected polymorphic variants of the RET gene in patients with medullary thyroid carcinoma and in the general population of central Poland. Endocr Pathol. 2010 Sep;21(3):178–85.

108. Pasquali D, Circelli L, Faggiano A, Pancione M, Renzullo A, Elisei R, et al. CDKN1B V109G polymorphism a new prognostic factor in sporadic medullary thyroid carcinoma. Eur J Endocrinol. 2011 Mar;164(3):397–404.

109. Ceolin L, Siqueira DR, Ferreira CV, Romitti M, Maia SC, Leiria L, et al. Additive effect of RET polymorphisms on sporadic medullary thyroid carcinoma susceptibility and tumor aggressiveness. Eur J Endocrinol. 2012 May;166(5):847–54.

110. Barbieri RB, Bufalo NE, Secolin R, Silva ACN, Assumpcao LVM, Maciel RMB, et al. Evidence that polymorphisms in detoxification genes modulate the susceptibility for sporadic medullary thyroid carcinoma. Eur J Endocrinol. 2012 Feb;166(2):241–5.

111. Barbieri RB, Bufalo NE, Cunha LL, Assumpcao LVM, Maciel RMB, Cerutti JM, et al. Genes of detoxification are important modulators of hereditary medullary thyroid carcinoma risk. Clin Endocrinol (Oxf). 2013 Aug;79(2):288–93.

112. Barbieri RB, Bufalo NE, Secolin R, Assumpcao LVM, Maciel RMB, Cerutti JM, et al. Polymorphisms of cell cycle control genes influence the development of sporadic medullary thyroid carcinoma. Eur J Endocrinol. 2014 Dec;171(6):761–7.

113. Costa P, Domingues R, Sobrinho LG, Bugalho MJ. RET polymorphisms and sporadic medullary thyroid carcinoma in a Portuguese population. Endocrine. 2005 Aug;27(3):239–43.

114. Robledo M, Gil L, Pollan M, Cebrian A, Ruiz S, Azanedo M, et al. Polymorphisms G691S/S904S of RET as genetic modifiers of MEN 2A. Cancer Res. 2003 Apr;63(8):1814–7.

115. Tamanaha R, Camacho CP, Pereira AC, da Silva AMA, Maciel RMB, Cerutti JM. Evaluation of RET polymorphisms in a six-generation family with G533C RET mutation: specific RET variants may modulate age at onset and clinical presentation. Clin Endocrinol (Oxf). 2009 Jul;71(1):56–64.

116. Sharma BP, Saranath D. RET gene mutations and polymorphisms in medullary thyroid carcinomas in Indian patients. J Biosci. 2011 Sep;36(4):603–11.

117. Zhang Y, Wang S, Chen X, Huang S, Li J. Quantitative assessment of the association between L769L and S836S polymorphisms at RET gene and medullary thyroid carcinoma risk. Tumour Biol. 2014 Jul;35(7):6641–7.

118. Gimm O, Neuberg DS, Marsh DJ, Dahia PL, Hoang-Vu C, Raue F, et al. Over-representation of a germline RET sequence variant in patients with sporadic medullary thyroid carcinoma and somatic RET codon 918 mutation. Oncogene. 1999

Feb;18(6):1369–73.

119. Ruiz A, Antinolo G, Fernandez RM, Eng C, Marcos I, Borrego S. Germline sequence variant S836S in the RET proto-oncogene is associated with low level predisposition to sporadic medullary thyroid carcinoma in the Spanish population. Clin Endocrinol (Oxf). 2001 Sep;55(3):399–402.

120. Berard I, Kraimps J-L, Savagner F, Murat A, Renaudin K, Nicolli-Sire P, et al. Germline-sequence variants S836S and L769L in the RE arranged during Transfection (RET) proto-oncogene are not associated with predisposition to sporadic medullary carcinoma in the French population. Vol. 65, Clinical genetics. Denmark; 2004. p. 150–2.

121. Elisei R, Cosci B, Romei C, Bottici V, Sculli M, Lari R, et al. RET exon 11 (G691S) polymorphism is significantly more frequent in sporadic medullary thyroid carcinoma than in the general population. J Clin Endocrinol Metab. 2004 Jul;89(7):3579–84.

122. Wiench M, Wloch J, Wygoda Z, Gubala E, Oczko M, Pawlaczek A, et al. RET polymorphisms in codons 769 and 836 are not associated with predisposition to medullary thyroid carcinoma. Cancer Detect Prev. 2004;28(4):231–6.

123. Baumgartner-Parzer SM, Lang R, Wagner L, Heinze G, Niederle B, Kaserer K, et al. Polymorphisms in exon 13 and intron 14 of the RET protooncogene: Genetic modifiers of medullary thyroid carcinoma? J Clin Endocrinol Metab. 2005;90(11):6232–6.

124. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. Nature. 2015 Oct;526(7571):68–74.

125. Tomayko MM, Reynolds CP. Determination of subcutaneous tumor size in athymic

(nude) mice. Cancer Chemother Pharmacol. 1989;24(3):148–54.

126. MANTEL N, HAENSZEL W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst. 1959 Apr;22(4):719–48.

127. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986 Sep;7(3):177–88.

128. Barbieri RB, Bufalo NE, Secolin R, Assumpcao LVM, Maciel RMB, Cerutti JM, et al. Polymorphisms of cell cycle control genes influence the development of sporadic medullary thyroid carcinoma. Eur J Endocrinol. 2014 Dec;171(6):761–7.

129. Severskaia N V, Saenko VA, Il'in AA, Chebotareva I V, Rumiantsev PO, Isaev PA, et al. [RET and GFRA1 germline polymorphisms in medullary thyroid cancer patients]. Mol Biol (Mosk). 2006;40(3):425–35.

130. Elisei R, Cosci B, Romei C, Bottici V, Sculli M, Lari R, et al. More Frequent in Sporadic Medullary Thyroid Carcinoma Than in the General Population. 2016;89(June):3579–84.

131. Cebrian A, Lesueur F, Martin S, Leyland J, Ahmed S, Luccarini C, et al. Polymorphisms in the Initiators of RET (Rearranged during Transfection) Signaling Pathway and Susceptibility to Sporadic Medullary Thyroid Carcinoma. 2005;90(11):6268–74.

132. Fernández RM, Peciña A, Antiñolo G, Navarro E, Borrego S. Case History Analysis of RET Polymorphisms and Haplotypes in the Context of Sporadic Medullary Thyroid Carcinoma. 2006;16(4).

133. Sromek M, Czetwerty M, Steffen J. The Frequency of Selected Polymorphic Variants of the RET Gene in Patients with Medullary Thyroid Carcinoma and in the General

Population of Central Poland. 2010;178–85.

134. Lantieri F, Caroli F, Ceccherini I, Griseri P. thyroid carcinoma enlightened by a meta-analysis study. 2013;11:2808–19.

135. Costa P, Domingues R, Sobrinho LG, Bugalho MJ. RET Polymorphisms and Sporadic Medullary Thyroid Carcinoma in a Portuguese Population. 2005;27(3):239–43.

136. Gimm O, Donna S, Marsh DJ, Dahia PLM, Hoang-vu C, Raue F, et al. Over-representation of a germline RET sequence variant in patients with sporadic medullary thyroid carcinoma and somatic RET codon 918 mutation. 1999;1369–73.

137. Siqueira R, Rocha AP, Ceolin L, Maia AL, Meotti C, Estivalet A, et al. The RET polymorphic allele S836S is associated with early metastatic disease in patients with hereditary or sporadic medullary thyroid carcinoma. 2010;953–63.

138. Griseri P, Sancandi M, Patrone G, Bocciardi R, Hofstra R, Ravazzolo R, et al. A single-nucleotide polymorphic variant of the RET proto-oncogene is underrepresented in sporadic Hirschsprung disease. Eur J Hum Genet. 2000 Sep;8(9):721–4.

139. Study C, Barbieri RB, Bufalo NE, Secolin R, Silva ACN, Assumpc LVM, et al. Evidence that polymorphisms in detoxification genes modulate the susceptibility for sporadic medullary thyroid carcinoma. 2012;241–5.

140. Pasquali D, Circelli L, Faggiano A, Pancione M, Renzullo A, Elisei R, et al. CDKN1B V109G polymorphism a new prognostic factor in sporadic medullary thyroid carcinoma. 2011;397–404.

141. Barbieri F, Cagnoli M, Ragni N, Foglia G, Bruzzo C, Pedulla F, et al. Increased cyclin D1

expression is associated with features of malignancy and disease recurrence in ovarian tumors. Clin Cancer Res. 1999 Jul;5(7):1837–42.

142. Chedid M, Michieli P, Lengel C, Huppi K, Givol D. A single nucleotide substitution at codon 31 (Ser/Arg) defines a polymorphism in a highly conserved region of the p53-inducible gene WAF1/CIP1. Oncogene. 1994 Oct;9(10):3021–4.

143. Balbin M, Hannon GJ, Pendas AM, Ferrando AA, Vizoso F, Fueyo A, et al. Functional analysis of a p21WAF1,CIP1,SDI1 mutant (Arg94 --> Trp) identified in a human breast carcinoma. Evidence that the mutation impairs the ability of p21 to inhibit cyclin-dependent kinases. J Biol Chem. 1996 Jun;271(26):15782–6.

144. Ma H, Zhou Z, Wei S, Wei Q. Association between p21 Ser31Arg polymorphism and cancer risk: a meta-analysis. Chin J Cancer. 2011 Apr;30(4):254–63.

145. May A, Sun Y, Hildesheim A, Cheng Y, Hayes RB, Cao Y, et al. No Point Mutation Carcinoma but a Codon Polymorphism of the Tumor Suppressor Gene in Nasopharyngeal (N PC): The Polymorphism Distinguishes Caucasians from Chinese. 1995;4(May):261–7.

146. Johnson GG, Sherrington PD, Carter A, Lin K, Liloglou T, Field JK, et al. A novel type of p53 pathway dysfunction in chronic lymphocytic leukemia resulting from two interacting single nucleotide polymorphisms within the p21 gene. Cancer Res. 2009 Jun;69(12):5210–7.

147. Wang Z, Sturgis EM, Zhang F, Lei D, Liu Z, Xu L, et al. Genetic variants of p27 and p21 as predictors for risk of second primary malignancy in patients with index squamous cell

carcinoma of head and neck. Mol Cancer. 2012 Mar;11:17.

148. Roh J-W, Kim BK, Lee CH, Kim J, Chung HH, Kim JW, et al. P53 codon 72 and p21 codon 31 polymorphisms and susceptibility to cervical adenocarcinoma in Korean women. Oncol Res. 2010;18(9):453–9.

149. Wang N, Wang S, Zhang Q, Lu Y, Wei H, Li W, et al. Association of p21 SNPs and risk of cervical cancer among Chinese women. BMC Cancer. 2012 Dec;12:589.

150. Facher EA, Becich MJ, Deka A, Law JC. Association between human cancer and two polymorphisms occurring together in the p21Waf1/Cip1 cyclin-dependent kinase inhibitor gene. Cancer. 1997 Jun;79(12):2424–9.

151. Siqueira DR, Romitti M, da Rocha AP, Ceolin L, Meotti C, Estivalet A, et al. The RET polymorphic allele S836S is associated with early metastatic disease in patients with hereditary or sporadic medullary thyroid carcinoma. Endocr Relat Cancer. 2010 Dec;17(4):953–63.

152. Cardot-Bauters C, Leteurtre E, Leclerc L, Vantyghem MC, Do Cao C, Wemeau JL, et al. Does the RET variant G691S influence the features of sporadic medullary thyroid carcinoma? Clin Endocrinol (Oxf). 2008 Sep;69(3):506–10.

153. Machens A, Frank-Raue K, Lorenz K, Rondot S, Raue F, Dralle H. Clinical relevance of RET variants G691S, L769L, S836S and S904S to sporadic medullary thyroid cancer. Clin Endocrinol (Oxf). 2012;76(5):691–7.

154. Lesueur F, Cebrian A, Robledo M, Niccoli-Sire P, Svensson KA, Pinson S, et al. Polymorphisms in RET and its coreceptors and ligands as genetic modifiers of multiple

endocrine neoplasia type 2A. Cancer Res. 2006;66(2):1177–80.

155. HILL AB. THE ENVIRONMENT AND DISEASE: ASSOCIATION OR CAUSATION? Proc R Soc Med. 1965 May;58(5):295–300.

156. Toogood AA, Eng C, Smith DP, Ponder BAJ, Shalet SM. No mutation at codon 918 of the RET gene in a family with multiple endocrine neoplasia type 2B. Clin Endocrinol (Oxf) [Internet]. 1995;43(6):759–62. Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L26015022%5Cnhttp://linksource.ebsco.com/linking.aspx?sid=EMBASE&issn=03000664&id=doi:&atitle=No+mutation+at+codon+918+of+the+RET+gene+in+a+family+with+multiple+endocrine+neoplasia+t

157. Hanks SK, Quinn AM, Hunter T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science. 1988 Jul;241(4861):42–52.

158. Gimm O, Marsh DJ, Andrew SD, Frilling A, Dahia PL, Mulligan LM, et al. Germline dinucleotide mutation in codon 883 of the RET proto-oncogene in multiple endocrine neoplasia type 2B without codon 918 mutation. J Clin Endocrinol Metab. 1997 Nov;82(11):3902–4.

159. Smith DP, Houghton C, Ponder BAJ. Germline mutation of RET codon 883 in two cases of de novo MEN 2B. 1997;1213–7.

160. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. 2006;34(WEB. SERV. ISS.):676–82.

161. Hampel H, Bennett RL, Buchanan A, Pearlman R, Wiesner GL. A practice guideline from the American College of Medical Genetics and Genomics and the National Society of Genetic Counselors: Referral indications for cancer predisposition assessment. Genet Med. 2015;17(1):70–87.

162. Smith J, Read ML, Hoffman J, Brown R, Bradshaw B, Campbell C, et al. Germline ESR2 mutation predisposes to medullary thyroid carcinoma and causes up-regulation of RET expression. Hum Mol Genet. 2016;25(9):1836–45.

163. Sponziello M, Benvenuti S, Gentile A, Pecce V, Rosignolo F, Virzì AR, et al. Whole exome sequencing identifies a germline MET mutation in two siblings with hereditary wild-type RET medullary thyroid cancer. Hum Mutat. 2018;39(3):371–7.

164. Lima J, Teixeira-Gomes J, Soares P, Máximo V, Honavar M, Williams D, et al. Germline Succinate Dehydrogenase Subunit D Mutation Segregating with Familial Non-RET C Cell Hyperplasia. J Clin Endocrinol Metab. 2003;88(10):4932–7.

165. Blais J, Lavoie SB, Giroux S, Bussières 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. J Mol Diagnostics. 2015;17(5):505–14.

166. Ravine D, Suthers G. Quality standards and samples in genetic testing. J Clin Pathol. 2012;65(5):389–93.

167. Dehainault C, Golmard L, Millot GA, Charpin A, Laugé A, Tarabeux J, et al. Mosaicism and prenatal diagnosis options: Insights from retinoblastoma. Eur J Hum Genet. 2017;25(3):381–3.

168. Prochazkova K, Pavlikova K, Minarik M, Sumerauer D, Kodet R, Sedlacek Z. Somatic TP53 mutation mosaicism in a patient with Li-Fraumeni syndrome. Am J Med Genet Part A. 2009;149(2):206–11.

169. Haque MM, Kowtal P, Sarin R. Identification and characterization of TP53 gene Allele Dropout in Li-Fraumeni syndrome and Oral cancer cohorts. Sci Rep [Internet]. 2018;8(1):11705. Available from: http://www.nature.com/articles/s41598-018-30238-7

170. Ward KJ, Ellard S, Yajnik CS, Frayling TM, Hattersley AT, Venigalla PNS, et al. Allelic drop-out may occur with a primer binding site polymorphism for the commonly used RFLP assay for the -1131T>C polymorphism of the Apolipoprotein AV gene. Lipids Health Dis. 2006;5:1–6.

171. Stratatis CA, Marx SJ. Abstracts for the Ninth International Workshop on Multiple Endocrine Neoplasia (MEN2004). 2004;696–730.

172. Lam CW, Mak CM. Allele dropout caused by a non-primer-site SNV affecting PCR amplification - A call for next-generation primer design algorithm. Clin Chim Acta. 2013;421:208–12.

173. Stevens AJ, Stuffrein-Roberts S, Cree SL, Gibb A, Miller AL, Doudney K, et al. G-Quadruplex structures and CpG methylation cause drop-out of the maternal allele in polymerase chain reaction amplification of the Imprinted MEST gene promoter. PLoS One. 2014;9(12):1–24.

174. Stevens AJ, Kennedy MA. Methylated Cytosine Maintains G-Quadruplex Structures during Polymerase Chain Reaction and Contributes to Allelic Dropout. Biochemistry.

2017;56(29):3691–8.

175. Hedell R, Dufva C, Ansell R, Mostad P, Hedman J. Enhanced low-template DNA analysis conditions and investigation of allele dropout patterns. Forensic Sci Int Genet. 2015;14:61–75.

176. Pompanon F, Bonin A, Bellemain E, Taberlet P. Genotyping errors: Causes, consequences and solutions. Nat Rev Genet. 2005;6(11):847–59.

177. 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. Mol Genet Genomic Med. 2017;5(4):443–7.

178. 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. Genet Med. 2007;9(11):752–60.

179. 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. Genet Test Mol Biomarkers [Internet]. 2010;14(5):677–90. Available from: http://www.liebertonline.com/doi/abs/10.1089/gtmb.2009.0183

180. Chong HK, Wang T, Lu H-M, 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 [Internet]. 2014;9(5):e97408. Available from:

http://dx.plos.org/10.1371/journal.pone.0097408

181. Ellison G, Wallace A, Kohlmann A, Patton S. A comparative study of germline BRCA1 and BRCA2 mutation screening methods in use in 20 European clinical diagnostic laboratories. Br J Cancer [Internet]. 2017;117(5):710–6. Available from: http://dx.doi.org/10.1038/bjc.2017.223

Publication

Received: 29 May 2019 | Revised: 11 July 2019 | Accepted: 15 July 2019

DOI: 10.1002/cam4.2443

ORIGINAL RESEARCH

Cancer Medicine Open Access WILEY

Genetic risk association of CDKN1A and RET gene SNPs with medullary thyroid carcinoma: Results from the largest MTC cohort and meta-analysis

Vasudha Mishra[1,2] | **Pradnya Kowtal**[1,2] | **Pallavi Rane**[3] | **Rajiv Sarin**[1,2,4]

[1]Sarin Lab, Advanced Centre for Treatment Research and Education in Cancer, Tata Memorial Centre, Navi Mumbai, India

[2]Homi Bhabha National Institute, Mumbai, India

[3]Clinical Research Centre (CRC), Advanced Centre for Treatment Research and Education in Cancer, Tata Memorial Centre, Navi Mumbai, India

[4]Cancer Genetics Clinic, Tata Memorial Hospital, Tata Memorial Centre, Mumbai, India

Correspondence
Rajiv Sarin, Advanced Centre for Treatment Research and Education in Cancer (ACTREC) & Tata Memorial Hospital, Tata Memorial Centre, Room No. KS#153, ACTREC-Tata Memorial Centre, Sector-22, Kharghar, Navi Mumbai, Maharashtra 410210, India.
Email: rsarin@actrec.gov.in

Funding information
The study was supported by the Indian Council of Medical Research (ICMR) and the Department of Science and Technology (DST), Government of India.

Abstract

Background: Medullary thyroid carcinoma (MTC) is a rare subtype of thyroid cancer. Other than gain-of-function *RET* mutations, no other genetic, lifestyle or environmental risk associations have been established for MTC. Several case-control studies and meta-analysis have examined the risk association of different SNPs with MTC in different populations but with contradictory or inconclusive results.

Methods: In a large cohort of 438 Indian MTC cases and 489 gender and ethnicity matched healthy controls from 1000 genome project, a comprehensive risk association of 13 SNPs of three pathways—detoxification, cell cycle regulation and *RET* was performed along with meta-analysis of RET SNPs.

Results: Multivariate logistic regression analysis identified a protective risk association of *CDKN1A*Ser31Arg SNP with both hereditary (OR 0.26; 95% confidence interval [CI] 0.13-0.55; $P < .001$) and sporadic MTC (OR 0.53; 95% CI 0.36-0.78; $P = .001$). An increased risk association was identified for *NAT2*Y94Y SNP (OR 1.62, 95% CI 1.17-2.25, $P = .004$) and *CDKN2A*3′UTR SNP (OR 1.89, 95% CI 1.19-2.98, $P = .006$) with sporadic MTC and *RET* S904S with hereditary MTC (OR 2.82, 95% CI 1.64-4.86, $P < .001$). Meta-analysis of *RET* SNPs including our cohort identified increased risk association of all four RET SNPs with MTC.

Conclusion: In this largest SNP risk association study for MTC and the only risk association study of the 13 most commonly studied MTC associated SNPs in a single cohort of this rare cancer, a significant protective risk association of *CDKN1A*Ser31Arg SNP with MTC was shown for the first time. Meta-analysis identified significant risk association of all four *RET* SNPs, not observed in previous meta-analysis.

KEYWORDS
CDKN1A, meta-analysis, MTC, RET, risk association, SNP

1 | INTRODUCTION

Thyroid cancers are broadly divided into less aggressive differentiated cancers—Papillary and Follicular thyroid cancer; and very aggressive poorly differentiated cancers—Medullary and Anaplastic Thyroid Cancer. Unlike the more common differentiated Thyroid Cancers, the risk factors for the less common but more aggressive thyroid cancers (medullary thyroid carcinoma [MTC] and ATC) are not well known. MTC originates from the parafollicular C cells of the thyroid. MTC is curable only if it is diagnosed and treated surgically when the disease is confined to the thyroid with or without limited regional nodal spread.[1] Current systemic treatment including Receptor Tyrosine Kinase Inhibitors such as sorafenib or cytotoxic chemotherapy does not produce long lasting disease control or cure. In the US SEER database, of the 793 MTC cases diagnosed between 1993 and 2002, the 10 year Disease specific survival was 96% for patients with MTC localized to the thyroid, 71% for patients with regional nodal spread and 26% in patients with distant spread.[2-4]

Around 75% MTC cases are sporadic while the remaining 25% cases are hereditary in nature and occur as part of an autosomal dominant inherited cancer syndrome called multiple endocrine neoplasia type 2 (MEN2).[5,6] MEN2 syndrome which affects multiple neuro-endocrine organs, has three clinical subtypes: MEN2A, MEN2B and Familial MTC.[7] MTC is the common clinical feature of all the three subtypes.

Mutations in *RET* gene have been identified as the primary susceptibility factor for MTC development. *RET* is a proto-oncogene that encodes a receptor tyrosine kinase expressed in neural crest derived cells.[8] In hereditary MTC cases germline point mutations in *RET* are identified in 95%-98% cases[5,9-11] whereas 40%-60% sporadic MTC cases have somatic *RET* mutations.[8,12,13] Other than the high penetrance gain-of-function germline or somatic *RET* mutations, no other genetic, lifestyle or environmental risk associations have been clearly established for MTC.

A few small studies which have examined certain lifestyle related risk associations with MTC have either failed to show any risk association or have paradoxically identified a protective role of tobacco smoking and alcohol.[14-16] Several case-control studies have examined the risk association of SNPs in *RET* and a few other genes involved in xenobiotic metabolism and cell cycle regulation with MTC in different populations.[6,17-38] However, most of these studies and their meta-analysis were either inconclusive or showed contradictory results. The possible reasons for not finding significant and consistent risk association could be the small cohort size of this rare cancer, geo-ethnic differences or poorly matched controls. Moreover, none of the studies have examined the risk association of SNPs in all these three pathways together in a single cohort. Hence, using the largest cohort of 438 MTC cases (361 sporadic and 77 hereditary) and gender and ethnicity matched 489 healthy controls from the 1000 Genome Project,[39] South Asian population, a comprehensive analysis of risk association of SNPs in all the three known MTC genetic modifier pathways was undertaken. These include a total of 13 SNPs from genes of detoxification (*Cyp1A1m1, Cyp1A2*F, NAT2, GSTP1)*, cell cycle regulation (*CDKN1A, CDKN1B, CDKN2A, CDKN2B, CDKN2C*) and the *RET* gene (G691S, L769L, S836S, S904S) (Table S1). Further, a meta-analysis of all the case-control studies examining risk association of the four *RET* gene SNPs with MTC, including the present study, was conducted to derive definitive conclusions.

2 | MATERIALS AND METHODS

2.1 | Study subjects

The study was conducted on 438 Indian MTC cases enrolled between 2006 and 2018 at the Cancer Genetics Clinic; Tata Memorial Hospital as part of Institutional Ethics Committee approved study. Personal and family history with clinico-pathological details was recorded. Blood sample was collected with written informed consent. The inclusion criteria were histologically confirmed diagnosis of MTC with raised serum calcitonin in patients of any age or gender. Exclusion criteria included a previous history of another cancer except pheochromocytoma which is a part of MEN2 syndrome. The hereditary MTC group consisted of those patients with germline RET proto-oncogene mutation, irrespective of family history or syndromic features. Those without a germline RET mutation were considered as sporadic MTC. In our cohort of 438 MTC cases, we have 77 hereditary and 361 sporadic MTC cases. Detailed lifestyle or exposure data were not systematically collected and analyzed as their risk association with MTC has not been established in earlier studies. A majority of the large studies on MTC risk association have not taken in to account the demographic or lifestyle factors of MTC patients.[27,28,40] Genotyping data for healthy controls were extracted from the South Asian population of the 1000 Genome Project (http://www.ensembl.org/Homo_sapiens/Info/Index). This South Asian cohort included all major ethnicities of Indian origin—Punjabis from Lahore, Gujarati from Houston, Telugu from UK, Bengali from Bangladesh and Sri Lankan Tamil from UK.

2.2 | Molecular genetic testing

2.2.1 | *RET* gene sequencing

From the peripheral blood sample, DNA was extracted using Qiagen QIAmp DNA Mini kit (Cat#51304). Germline *RET* mutation analysis was performed for six hotspot exons of *RET* (10, 11, 13, 14, 15 16) using polymerase chain reaction (PCR) and Sanger Sequencing. For PCR, 5 μL (20 ng/μL) gDNA was amplified in a 25 μL PCR reaction volume containing

0.5 μL of each Forward and Reverse primer (10 pmol), 1 μL dNTPs (2.5 mmol), 0.5 μL Taq Polymerase (2 U/μL—Thermo Scientific), 2.5 μL Taq Buffer (10X) and the total volume was adjusted to 25 μL with molecular biology grade water. Primers for PCR were designed using Oligo Explorer version 1.5. Purification of PCR products was done using ExoSAP IT (USB Products, Affimetrix). Sanger Sequencing was performed using BigDye Terminator Cycle Sequencing kit v3.1 (Applied Biosystems) on ABI 3500 and 3730 DNA Sequencer (Applied Biosystems) and electropherograms were analyzed using Chromas Lite version 2.6.4 using reference sequence of *RET* gene extracted from National Center for Biotechnology Information NG_007489.1.

2.2.2 | SNP genotyping

SNP genotyping was done using Restriction Fragment Length Polymorphism (RFLP) for 10/13 SNPs. For the remaining three SNPs, genotyping was done using TaqMan as no restriction site for a single cutter restriction enzyme was identified either for the wild type or variant allele. For both genotyping methods, 10% of the genotyping results were confirmed to be true using Sanger Sequencing. SNP genotyping using RFLP was done for *Cyp1A1m1, Cyp1A2*F, GSTP1, NAT2, CDKN1A, CDKN1B, CDKN2A, RET* L769L, S836S and S904S polymorphisms and using TaqMan for *CDKN2B, CDKN2C* and *RET* G691S polymorphisms. For RFLP, 100 ng gDNA was PCR amplified followed by restriction digestion using reaction conditions as per the manufacturer's protocol. The digested products were visualized on 2% agarose gel and the genotypes were inferred from band sizes in the gel. For TaqMan SNP genotyping, 1 μL gDNA (10 ng/μL) was mixed with 2.5 μL TaqMan universal master mix II with UNG (Applied Biosystems, cat#4440038) and 0.1 μL probe mix (Applied Biosystems) designed for each SNP. TaqMan realtime PCR was performed on QuantStudio 5.0 and genotypes were inferred from amplification plot and allelic discrimination plots. About 5% of all the genotyping results were validated using Sanger Sequencing.

2.3 | Statistical analysis

All Statistical analysis was performed on SPSS v21.0. SNP genotypes were tested for Hardy-Weinberg equilibrium (HWE) using Chi-square HWE test calculator for biallelic markers (http://www.oege.org/software/hwe-mr-calc.shtml) (Table S2). Genotypic frequency was calculated for all 13 SNPs and compared between cases and controls using chi-square test (Table S3). As the homozygous status of several SNPs was either absent or very low in either cases or controls, analysis was performed only for the dominant model which compares the variant allele either as heterozygous or homozygous form (Aa+aa) with the homozygous wild type allele (AA). Logistic regressions were used to analyze the association between these polymorphisms and MTC risk and odds ratio (ORs) was calculated with 95% confidence interval (CI). All SNPs showing a trend for association on univariate analysis with $P < .1$ were included in the multivariate logistic regression analysis. As multiple comparisons were made for 13 SNPs in a single cohort, a P-value of <.01 was used to consider an association as statistically significant.

2.4 | Literature search and meta-analysis

PUBMED search was conducted to identify eligible studies for meta-analysis using the following search words: "Polymorphism AND MTC", "SNPs AND MTC", "*RET* Polymorphisms AND MTC". All published case-control studies examining the risk association of these SNPs with sporadic or hereditary MTC were included in the meta-analysis, the details of which are provided in Table S4. Meta-Analysis was performed with R-Software package using minor allele frequency data as the genotype frequencies were not available for several studies. We applied both the fixed effect[41] and the random effect[42] model for meta-analysis. The significance of overall OR was calculated using Z test. Heterogeneity between studies was investigated using I^2 and τ^2 statistics. The results of meta-analysis were reported as conventional Forest plots.

3 | RESULTS

The 438 MTC cases in our cohort included 239 males (54.5%) and 199 (45.4%) females. The mean age at MTC diagnosis was 40.64 ± 14.24, Median: 40 years with the range of 8-80 years. The 489 controls used for the risk association study included 260 males (53.2%) and 229 females (46.8%). Both the cases and controls were matched for gender ($P = .67$) and ethnicity. The genotype frequencies of all the SNPs included in the study are summarized in Table S2. HWE was maintained for all 13 SNPs in the controls and for 11/13 SNPs in the MTC cases (Table S2).

3.1 | Risk associations

3.1.1 | Present study

On univariate logistic regression analysis, *CDKN1A* SNP showed consistent and significant association with reduced risk of MTC in both hereditary (OR = 0.52; 95% CI = 0.27-0.99; $P = .048$) and sporadic MTC groups (OR = 0.63; 95% CI = 0.45-0.88; $P = .007$). The variant allele A was overrepresented in the control population (26.2%) as compared to both the hereditary cases (15.6%) and the sporadic cases (18.3%) (Tables 1 and 3). The strong association of CDKN1A SNP with reduced MTC risk was further

confirmed on multivariate logistic regression analysis in both the hereditary (OR = 0.27; 95% CI = 0.13-0.55; $P < .001$) and sporadic MTC groups (OR = 0.53; 95% CI = 0.36-0.78; $P = .001$) (Table 2 and 4).

Multivariate logistic regression analysis also identified significant risk association for the *RET* S904S SNP in the hereditary MTC group (OR = 2.82; 1.64-4.86; $P < .001$) (Table 2) whereas for *CDKN2A* (OR = 1.89; 95% CI = 1.20-2.98; $P = .006$) and *NAT2* SNP (OR = 1.62; 95% CI = 1.17-2.25; $P = .004$) in the sporadic MTC group (Table 2).

3.1.2 | Meta-analysis including present study

We identified 23 case-control studies examining risk associations of one or more of these 13 SNPs with MTC. However, for nine SNPs in the cell cycle regulation (*CDKN1A, CDKN1B, CDKN2A, CDKN2B, CDKN2C*) and detoxification pathway (*CYP1A1m1, CYP1A2*F, NAT2, GSTP1*), only single small cohort studies had examined their risk association with MTC.[23,25-27] Hence the meta-analysis was performed only for the four *RET* gene SNPs (G691S, L769L, S836S, S904S) one or more of which are reported in 19 case-control studies. This included a total of 346 cases and 1555 controls in the hereditary MTC group and 1640 cases and 2968 controls in sporadic MTC group (Table S4). The ORs with 95% CIs calculated for the allelic distribution of SNPs for each study is shown in their respective Forest plots (Figures 1-4).

The meta-analysis identified a significant association between *RET* L769L and S836S SNPs with risk of hereditary MTC (Figures 2B and 3B).The *RET* S836S variant allele was found to be associated with increased susceptibility to MTC. The effect was observed under both the fixed effect model (OR = 1.69; 95% CI = 1.17-2.43; $P < .01$) and random effect model (OR = 2.19; 95% CI = 1.02-4.71; $P = .04$).For *RET* L769L variant, a significant protective risk association with

TABLE 1 Univariate logistic regression analysis for association between SNPs and risk of hereditary MTC development (hMTC: hereditary MTC; Wt: Wild type; Hz: Heterozygous; Hm: Homozygous)

	Genotype frequency—hMTC cases (n = 77)		Genotype frequency—Controls (n = 489)				
Gene/SNP	**Wt**	**Hz + Hm**	**Wt**	**Hz + Hm**	**OR**	**95% CI**	***P*-value**
Cyp1A1m1	37 (48.1%)	40 (51.9%)	218 (44.6%)	271 (55.4%)	0.870	0.537-1.407	.570
Cyp1A2	28 (36.4%)	49 (63.6%)	145 (29.7%)	344 (70.3%)	0.738	0.446-1.220	.236
NAT2	31 (42.2%)	46 (59.7%)	167 (34.2%)	322 (65.8)	0.770	0.470-1.259	.297
GSTP1	35 (45.5%)	42 (54.5%)	251 (51.3%)	238 (48.7%	1.266	0.781-2.050	.339
CDKN1A	**65 (84.4%)**	**12 (15.6%)**	**361 (73.8%)**	**128 (26.2%)**	**0.521**	**0.272-0.995**	**.048**
CDKN1B	35 (45.5%)	42 (54.5%)	224 (45.8%)	265 (54.2%)	1.014	0.626-1.643	.954
CDKN2A	67 (87%)	10 (13%)	439 (89.7%)	50 (10.2%)	1.310	0.634-2.708	.465
CDKN2B	46 (59.8%)	31 (40.2%)	266 (54.4%)	223 (45.6%)	0.804	0.493-1.311	.382
CDKN2C	62 (80.5%)	15 (19.5%)	387 (79.1%)	102 (20.8%)	0.918	0.501-1.680	.781
G691S RET	37 (48.1%)	40 (51.9%)	283 (57.8%)	206 (42.1%)	1.485	0.917-2.404	.108
L769L RET	35 (45.5%)	42 (54.5%)	178 (36.4%)	311 (36.6%)	0.687	0.423-1.115	.129
S836S RET	64 (83.1%)	13 (16.9%)	411 (84%)	78 (15.9%)	1.070	0.562-2.037	.836
S904S RET	36 (46.8%)	41 (53.2)	285 (58.3%)	204 (41.7%)	1.591	0.982-2.578	.06

Values in bold indicates significant associations.

Abbreviations: CI, confidence interval; MTC, medullary thyroid carcinoma; OR, odds ratio.

TABLE 2 Multivariate logistic regression analysis for association between SNPs and risk of hereditary MTC (hMTC) development (SNPs with significance <0.1 from univariate analysis were included in multivariate analysis)

	Genotype frequency—hMTC cases (n = 77)		Genotype frequency—Controls (n = 489)				
Gene/SNP	**Wt**	**Hz + Hm**	**Wt**	**Hz + Hm**	**OR**	**95% CI**	***P*-value**
CDKN1A	65 (84.4%)	12 (15.6%)	361 (73.8%)	128 (26.2%)	0.266	0.129-0.549	***<.001***
S904S RET	36 (46.8%)	41 (53.2)	285 (58.3%)	204 (41.7%)	2.821	1.636-4.862	***<.001***

P-value in bold indicates significant associations.

Abbreviations: CI, confidence interval; MTC, medullary thyroid carcinoma; OR, odds ratio.

TABLE 3 Univariate logistic regression analysis for association between SNPs and risk of sporadic MTC development (sMTC: sporadic MTC)

	Genotype frequency—sMTC cases (n = 361)		Genotype frequency—Controls (n = 489)				
Gene/SNP	**Wt**	**Hz + Hm**	**Wt**	**Hz + Hm**	**OR**	**95% CI**	***P*-value**
Cyp1A1m1	161 (44.6%)	200 (55.4%)	218 (44.6%)	271 (55.4%)	0.999	0.760-1.314	.996
Cyp1A2	117 (32.4%)	244 (67.6%)	145 (29.7%)	344 (70.3%)	0.879	0.655-1.179	.390
NAT2	103 (28.5%)	258 (71.5%)	167 (34.2%)	322 (65.8)	1.299	0.967-1.745	.082
GSTP1	208 (57.6%)	153 (42.4%)	251 (51.3%)	238 (48.7%	0.776	0.590-1.020	.069
CDKN1A	**295 (81.7%)**	**66 (18.3%)**	**361 (73.8%)**	**128 (26.2%)**	**0.631**	**0.452-0.882**	**.007**
CDKN1B	182 (50.4%)	179 (49.6%)	224 (45.8%)	265 (54.2%)	0.831	0.633-1.092	.184
CDKN2A	310 (85.9%)	51 (14.1%)	439 (89.7%)	50 (10.2%)	1.444	0.953-2.190	.083
CDKN2B	206 (57%)	155 (42.9%)	266 (54.4%)	223 (45.6%)	0.898	0.682-1.180	.439
CDKN2C	297 (82.2%)	64 (17.7%)	387 (79.1%)	102 (20.8%)	0.818	0.578-1.157	.256
G691S RET	201 (55.7%)	160 (44.3%)	283 (57.8%)	206 (42.1%)	1.094	0.831-1.439	.523
L769L RET	146 (40.4%)	215 (59.6%)	178 (36.4%)	311 (36.6%)	0.843	0.637-1.115	.231
S836S RET	311 (86%)	50 (13.9%)	411 (84%)	78 (15.9%)	0.847	0.577-1.244	.398
S904S RET	194 (53.7%)	167 (46.3%)	285 (58.3%)	204 (41.7%)	1.203	0.914-1.582	.187

Values in bold indicates significant associations.

Abbreviations: CI, confidence interval; MTC, medullary thyroid carcinoma; OR, odds ratio.

TABLE 4 Multivariate logistic regression analysis for association between SNPs and risk of sporadic MTC (sMTC) development (SNPs with significance <0.1 from univariate analysis were included in multivariate analysis)

	Genotype frequency—sMTC cases (n = 361)		Genotype frequency—Controls (n = 489)				
Gene/SNP	**Wt**	**Hz + Hm**	**Wt**	**Hz + Hm**	**OR**	**95% CI**	***P*-value**
NAT2	103 (28.5%)	258 (71.5%)	167 (34.2%)	322 (65.8)	1.622	1.168-2.251	**.004**
GSTP1	208 (57.6%)	153 (42.4%)	251 (51.3%)	238 (48.7%	0.741	0.540-1.018	.065
CDKN1A	295 (81.7%)	66 (18.3%)	361 (73.8%)	128 (26.2%)	0.526	0.357-0.776	**.001**
CDKN2A	310 (85.9%)	51 (14.1%)	439 (89.7%)	50 (10.2%)	1.888	1.197-2.978	**.006**

Values in bold indicates significant associations.

Abbreviations: CI, confidence interval; MTC, medullary thyroid carcinoma; OR, odds ratio.

MTC was observed under both fixed effect model (OR = 0.77; 95% CI = 0.60-0.98; $P = .04$) and random effect model (OR = 0.77; 95% CI = 0.60-0.98; $P = .04$). Further in the sporadic MTC group, meta-analysis showed significantly increased risk of MTC with the *RET* G691S and S904S (Figures 1A and 4A). For G691S, the association was observed under both fixed effect model (OR = 1.21 95% CI = 1.07-1.37; $P < .01$) and random effect model (OR = 1.24; 95% CI = 1.04-1.47; $P = .02$). For S904S, this effect was observed under only fixed effect model (OR = 1.16; 95% CI = 1.03-1.30; $P = .02$).

4 | DISCUSSION

In hereditary cancer syndromes, highly penetrant germline mutations in proto-oncogene or tumor suppressor genes confer a very high lifetime risk of cancer development.[17] However in several sporadic cancers, in addition to environmental or lifestyle factors there is a component of weak genetic susceptibility conferred by low penetrance genetic variants. While there are no clearly established lifestyle or environmental risk factors for susceptibility to MTC, several SNPs in *RET* as well as other genes have been reported to slightly increase or decrease the risk of MTC development.[18,26,30,36,43] However, the findings of these studies are inconsistent. Of the four previously reported meta-analysis of *RET* gene SNPs,[6,17,28,32] two demonstrated a significant risk association of *RET*G691S SNPs with MTC.[17,28] No other significant risk association has been observed in the other two studies.

Like previous case-control studies in MTC,[43] we have analyzed the risk association of SNPs independently in the hereditary and sporadic MTC groups for our cohort as well

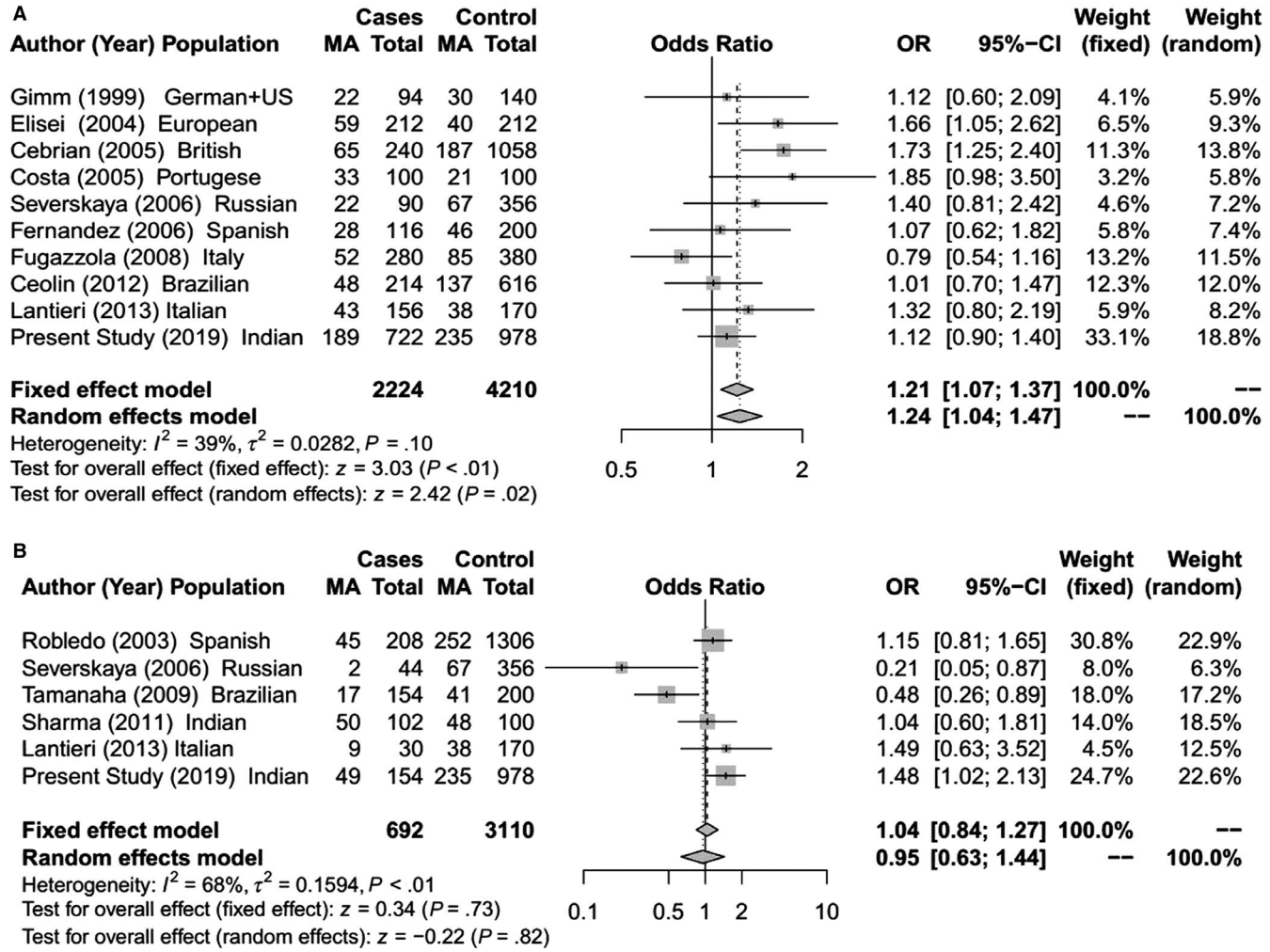


FIGURE 1 Forest Plot for meta-analysis on allelic association of *RET* G691S SNP with (A) Sporadic MTC; (B) Hereditary MTC [The total for cases and controls are allelic count (2n)]

as for the meta-analysis. In our cohort, multivariate logistic regression analysis identified a highly significant ($P < .01$) protective risk association of *CDKN1A* SNP for hereditary MTC as well as sporadic MTC (Tables 1-4). Two SNPs (*NAT2* and *CDKN2A)* had a significant increased risk association with sporadic MTC (Table 2) while another SNP (*RET* S904S) had a significant increased risk association with hereditary MTC (Table 2). With the inclusion of 346 hereditary MTC cases in the meta-analysis for 4 *RET* gene SNPs, a significant protective risk association was observed for *RET* L769L SNP while a significant increased risk association was seen with *RET* S836S SNP (Figure 2B and 3B). For the 1640 sporadic cases included in the meta-analysis, significant increased risk association was seen for the *RET* G691S and S904S SNPs (Figures 1A and 4A). A few functional and in-silico studies have postulated and examined how different *RET* SNPs modulate the risk of MTC development. These include their effect on RNA stability or its expression, creation of a new alternative splicing site[18,21,22,36] or changes in phosphorylation sites.[17] However, the findings of these studies have been inconclusive.

Univariate and multivariate logistic regression analysis in our cohort also demonstrated a strong protective association between *CDKN1A* SNP with hereditary and sporadic MTC. The *CDKN1A* gene, also known as $p21^{CIP1/WAF1}$, encodes a cyclin-dependent kinase inhibitor which binds to and inhibits the activity of Cyclin-CDK2 or CDK4 complexes regulating cell cycle progression at G1 stage.[44,45] *CDKN1A* activity is regulated by p53 which binds to its promoter and induces cell cycle arrest in response to various stimuli.[45] This gene is often deregulated in human cancers with altered expression reported in several cancers including cervical, breast, ovarian, liver, uterine, and head and neck cancers.[46] The *CDKN1A* SNP (rs1801270) at codon 31 (Ser31Arg) reported in the present study falls in a highly conserved N-terminal region of the protein, which is demonstrated to contain tumor suppressor function.[44] Functional studies suggested that while the *CDKN1A*-Ser and Arg variant possess similar kinase inhibitory and growth suppression abilities,[47]

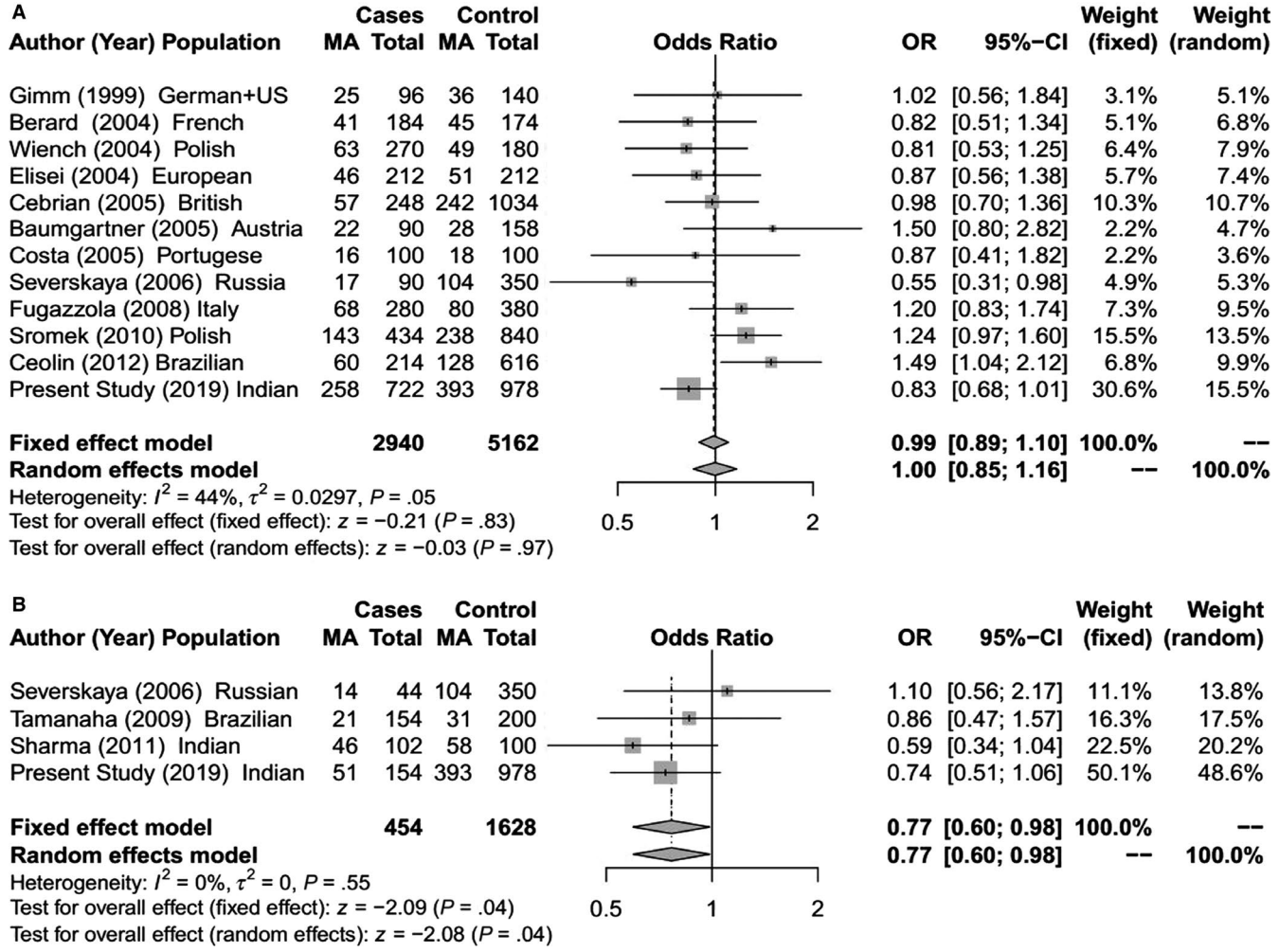


FIGURE 2 Forest Plot for meta-analysis on allelic association of *RET* L769L SNP with (A) Sporadic MTC; (B) Hereditary MTC [The total for cases and controls are allelic count (2n)]

their transcriptional efficiency is significantly different.[48] The allelic frequency of this SNP varies significantly among different populations with minor allele frequency of 15% in the South Asian Population (1000 Genome Project). Several molecular epidemiological studies of *CDKN1A* Ser31Arg SNP show conflicting results with some studies reporting increased risk association with tobacco related upper aerodigestive tract cancers,[49] while showing a protective effect in human papilloma virus related cervical cancers.[50,51] The only study of this SNP in MTC has been reported by Barbieri et al[27] in a small cohort of 45 sporadic MTC cases. Even though no significant risk association for MTC development was identified, perhaps due to the small sample size, extrathyroidal tumor extension was significantly less in patients with the *CDKN1A* SNP as compared to those with wild type *CDKN1A* (50% versus 92%, $P = .037$). In our study of much larger cohort of this rare cancer, univariate and multivariate logistic regression analysis shows the highly significant protective effect of *CDKN1A* SNP on risk of MTC development in sporadic as well as hereditary MTC.

The significant risk association of the variant allele C of *CDKN2A* 3'UTR SNP (rs11515), identified in our sporadic MTC cohort has also been reported as a risk allele in a Brazilian cohort of 45 sporadic MTC by Barbieri et al in 2014.[27] We have also identified a significantly increased risk association of the variant allele T of the *NAT2* Y94Y SNP (rs1041983) in our sporadic MTC cohort, as reported previously in a Brazilian cohort of 132 hereditary MTC cases.[26] However the same Brazilian group in their cohort of 47 sporadic MTC cases, found the variant allele T of this NAT2 SNP to be protective. This could be due to the small cohort size or difference in the frequency of alleles in the admixture population.[25]

The significant risk association of *CDKN2A*3'UTR SNP (rs11515) identified in our sporadic MTC cohort has also been reported in a Brazilian cohort of 45 sporadic MTC by Barberi et al in 2014.[27] For the *NAT2* Y94Y SNP (rs1041983) we identified a significantly increased risk association of the variant T allele in 361 sporadic MTC cases, as previously reported in a Brazilian cohort of 132

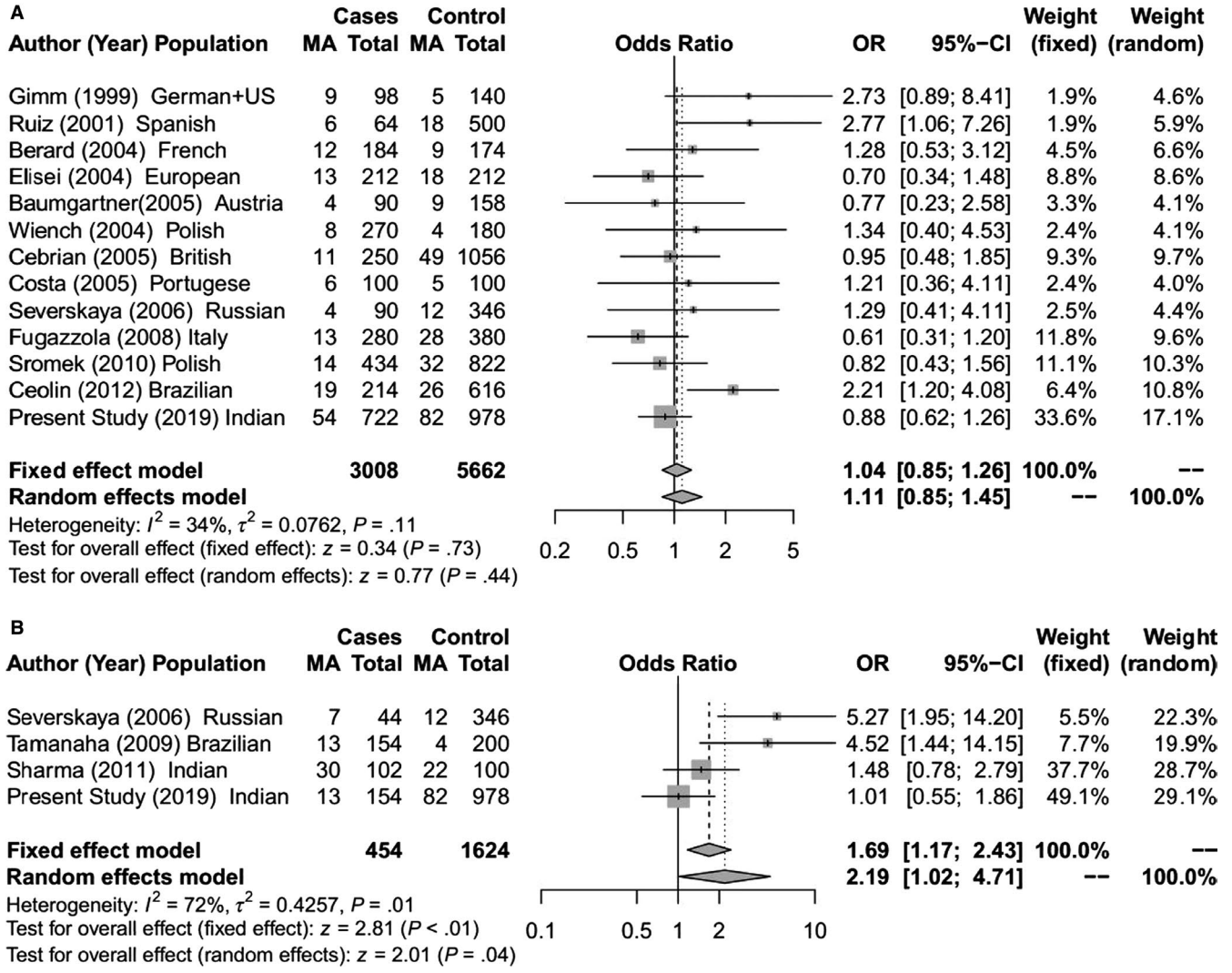


FIGURE 3 Forest Plot for meta-analysis on allelic association of *RET* S836S SNP with (A) Sporadic MTC; (B) Hereditary MTC [The total for cases and controls are allelic count (2n)]

hereditary MTC cases.[26] Paradoxically, in a study with 47 sporadic MTC cases, reported from the same Brazilian group,[25] the wild type C allele was associated with increased risk of MTC, the reasons for which have not been elaborated.

This is the first study to examine the MTC risk association of 13 different SNPs in genes of three distinct pathways in a single cohort, which is also the largest cohort of this rare cancer reported so far. The meta-analysis conducted by us, with the inclusion of MTC cases from our cohort, has increased the total sporadic MTC cases to 1640 and hereditary MTC cases to 346 (Table S4). While the previous meta-analysis by Figlioli et al in 2013 had failed to identify significant risk association with any of these four *RET* SNPs,[6] in our expanded meta-analysis cohort, we could identify significant risk association of *RET* L769L and S836S in hereditary MTC and of G691S and S904S in sporadic MTC.

One of the limitations of our study is that unlike classical case-control studies, instead of recruiting and genotyping matched controls, we used healthy gender and ethnicity matched South Asian controls from the 1000 genome database. Matching for age was not possible as MTC, especially the hereditary MTC, is known to occur in childhood and recruiting minor subjects as healthy controls for genotyping study raises ethical issues. Of all the MTC case-control studies, some have not reported whether controls were matched[22,43] whereas many have failed to obtain controls matched for age or gender.[24,26,28] Moreover, in the absence of a clearly established lifestyle or environmental factors for MTC risk, none of the MTC SNP case-control studies have described or matched for these factors, as is the case in our study.

Taken together, the findings from comprehensive genotyping of 13 SNPs in our large MTC cohort, we showed for the first time, a significant protective risk association of

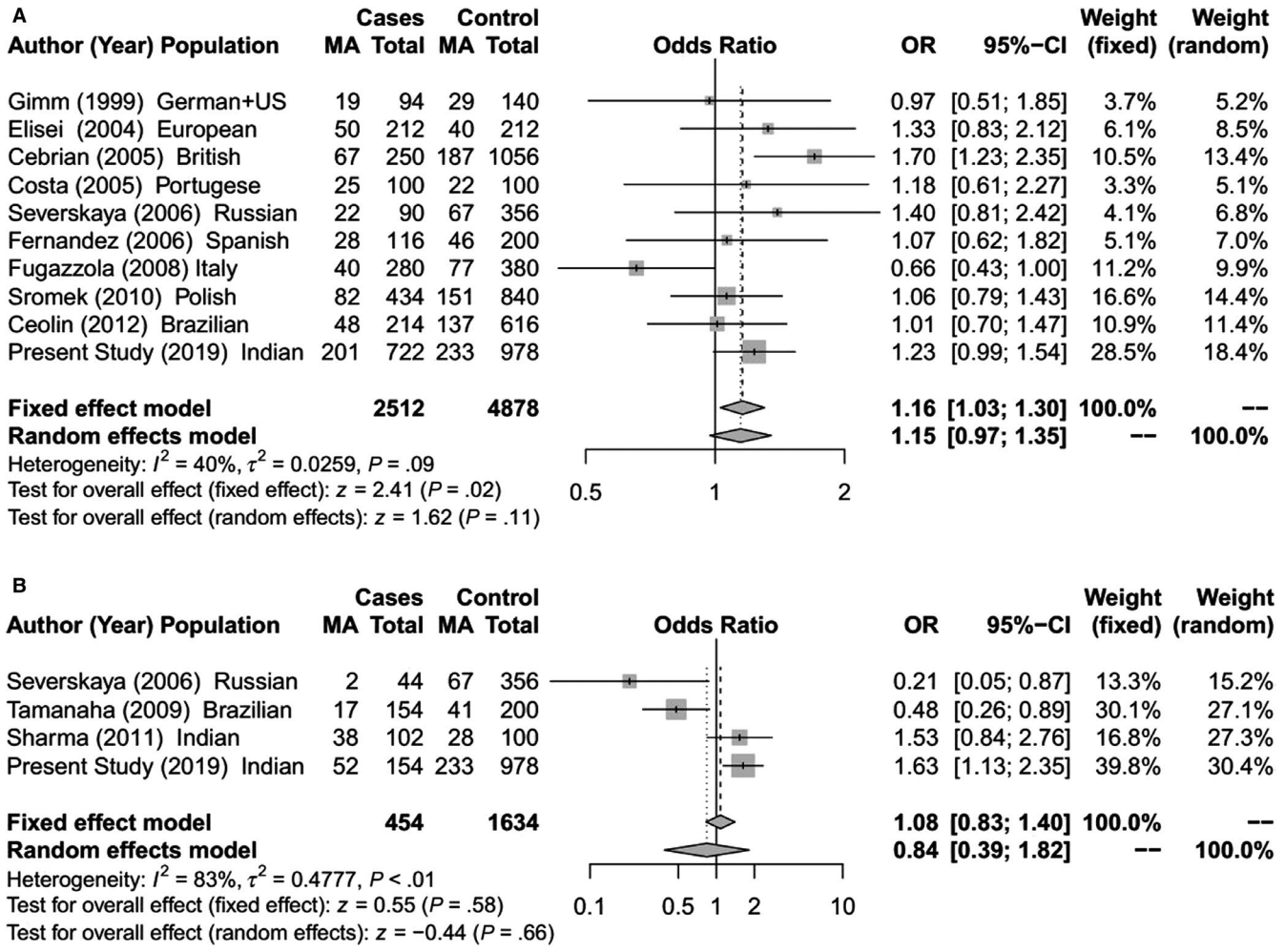


FIGURE 4 Forest Plot for meta-analysis on allelic association of *RET* S904S SNP with (A) Sporadic MTC; (B) Hereditary MTC [The total for cases and controls are allelic count (2n)]

CDKN1A SNP (rs1801270) with MTC and through meta-analysis of expanded cohort, we also showed a risk association of four *RET* SNPs with MTC. Identification of one or more low penetrance alleles in risk association studies in diverse cancers could provide some biological insight into cancer development but are not useful as biomarkers of prognosis or predisposition. However study of a large number of low penetrance alleles in large case-control studies could help in developing polygenic risk scores. The present study therefore underscores the need for large replicative risk association studies using a control group from the local population with well-defined characteristics to understand the molecular mechanisms through which these low penetrance alleles modulate MTC risk.

ACKNOWLEDGMENTS

We thank the Indian Council of Medical Research for funding the project and the Department of Science and Technology, Government of India for providing fellowship to Ms Vasudha Mishra. We acknowledge the cooperation of members of the Head and Neck Disease Management Group, Tata Memorial Hospital (TMH) for referring the patients. We thank Mr Ravindra Reddy and other genetic counselors at Cancer Genetics Clinic, TMH for providing counseling to the patients. We are thankful to all the patients and their family members for their participation in the study.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Rajiv Sarin https://orcid.org/0000-0002-6405-8282

REFERENCES

1. Kloos RT, Eng C, Evans DB, et al. Medullary thyroid cancer: management guidelines of the American Thyroid Association. *Thyroid*. 2009;19(6):565-612.
2. Leboulleux S, Baudin E, Travagli J, Schlumberger M. Medullary thyroid carcinoma. *Clin Endocrinol (Oxf)*. 2004;61(3):299-310.
3. Kebebew E, Ituarte PH, Siperstein AE, Duh QY, Clark OH. Medullary thyroid carcinoma: clinical characteristics, treatment, prognostic factors, and a comparison of staging systems. *Cancer*. 2000;88(5):1139-1148.
4. Randle RW, Balentine CJ, Leverson GE, et al. Trends in the presentation, treatment, and survival of patients with medullary thyroid cancer over the past 30 years. *Surgery*. 2017;161(1):137-146.
5. Eng C, Clayton D, Schuffenecker I, Al E. The relationship between specific RET proto-oncogene mutations and disease phenotype in multiple endocrine neoplasia type 2: international RET mutation consortium analysis. *JAMA*. 1996;276(19):1575-1579.
6. Figlioli G, Landi S, Romei C, Elisei R, Gemignani F. Medullary thyroid carcinoma (MTC) and RET proto-oncogene: mutation spectrum in the familial cases and a meta-analysis of studies on the sporadic form. *Mutat Res*. 2013;752(1):36-44.
7. Moline J, Eng C. Multiple endocrine neoplasia type 2: an overview. *Genet Med*. 2011;13(9):755-764.
8. De Groot J, Links TP, Plukker J, Lips C, Hofstra R. RET as a diagnostic and therapeutic target in sporadic and hereditary endocrine tumors. *Endocr Rev*. 2006;27(5):535-560.
9. Donis-Keller H, Dou S, Chi D, et al. Mutations in the RET proto-oncogene are associated with MEN 2A and FMTC. *Hum Mol Genet*. 1993;2(7):851-856.
10. Romei C, Mariotti S, Fugazzola L, et al. Multiple endocrine neoplasia type 2 syndromes (MEN 2): results from the ItaMEN network analysis on the prevalence of different genotypes and phenotypes. *Eur J Endocrinol*. 2010;163(2):301-308.
11. Eng C, Mulligan LM, Smith DP, et al. Low frequency of germline mutations in the RET proto-oncogene in patients with apparently sporadic medullary thyroid carcinoma. *Clin Endocrinol (Oxf)*. 1995;43(1):123-127.
12. Marsh DJ, Learoyd DL, Andrew SD, et al. Somatic mutations in the RET proto-oncogene in sporadic medullary thyroid carcinoma. *Clin Endocrinol (Oxf)*. 1996;44(3):249-257.
13. Agrawal N, Jiao Y, Sausen M, et al. Exomic sequencing of medullary thyroid cancer reveals dominant and mutually exclusive oncogenic mutations in RET and RAS. *J Clin Endocrinol Metab*. 2013;98(2):E364-E369.
14. Stansifer KJ, Guynan JF, Wachal BM, Smith RB. Modifiable risk factors and thyroid cancer. *Otolaryngol Head Neck Surg*. 2015;152(3):432-437.
15. Negri E, Ron E, Franceschi S, et al. Risk factors for medullary thyroid carcinoma: a pooled analysis. *Cancer Causes Control*. 2002;13(4):365-372.
16. Kalezic NK, Zivaljevic VR, Slijepcevic NA, Paunovic IR, Diklic AD, Sipetic SB. Risk factors for sporadic medullary thyroid carcinoma. *Eur J Cancer Prev*. 2013;22(3):262-267.
17. Lantieri F, Caroli F, Ceccherini I, Griseri P. The involvement of the RET variant G691S in medullary thyroid carcinoma enlightened by a meta-analysis study. *Int J cancer*. 2013;132(12):2808-2819.
18. Cebrian A, Lesueur F, Martin S, et al. Polymorphisms in the initiators of RET (rearranged during transfection) signaling pathway and susceptibility to sporadic medullary thyroid carcinoma. *J Clin Endocrinol Metab*. 2005;90(11):6268-6274.
19. Severskaya NV, Saenko VA, Ilyin AA, et al. Germline polymorphisms of RET and GFRA1 genes in patients with medullary thyroid carcinoma. *Mol Biol*. 2006;40(3):375-384.
20. Fernandez RM, Pecina A, Antinolo G, Navarro E, Borrego S. Analysis of RET polymorphisms and haplotypes in the context of sporadic medullary thyroid carcinoma. *Thyroid*. 2006;16(4):411-417.
21. Fugazzola L, Muzza M, Mian C, et al. RET genotypes in sporadic medullary thyroid cancer: studies in a large Italian series. *Clin Endocrinol (Oxf)*. 2008;69(3):418-425.
22. Sromek M, Czetwertynska M, Skasko E, Zielinska J, Czapczak D, Steffen J. The frequency of selected polymorphic variants of the RET gene in patients with medullary thyroid carcinoma and in the general population of central Poland. *Endocr Pathol*. 2010;21(3):178-185.
23. Pasquali D, Circelli L, Faggiano A, et al. CDKN1B V109G polymorphism a new prognostic factor in sporadic medullary thyroid carcinoma. *Eur J Endocrinol*. 2011;164(3):397-404.
24. Ceolin L, Siqueira DR, Ferreira CV, et al. Additive effect of RET polymorphisms on sporadic medullary thyroid carcinoma susceptibility and tumor aggressiveness. *Eur J Endocrinol*. 2012;166(5):847-854.
25. Barbieri RB, Bufalo NE, Secolin R, et al. Evidence that polymorphisms in detoxification genes modulate the susceptibility for sporadic medullary thyroid carcinoma. *Eur J Endocrinol*. 2012;166(2):241-245.
26. Barbieri RB, Bufalo NE, Cunha LL, et al. Genes of detoxification are important modulators of hereditary medullary thyroid carcinoma risk. *Clin Endocrinol (Oxf)*. 2013;79(2):288-293.
27. Barbieri RB, Bufalo NE, Secolin R, et al. Polymorphisms of cell cycle control genes influence the development of sporadic medullary thyroid carcinoma. *Eur J Endocrinol*. 2014;171(6):761-767.
28. Costa P, Domingues R, Sobrinho LG, Bugalho MJ. RET polymorphisms and sporadic medullary thyroid carcinoma in a Portuguese population. *Endocrine*. 2005;27(3):239-243.
29. Robledo M, Gil L, Pollan M, et al. Polymorphisms G691S/S904S of RET as genetic modifiers of MEN 2A. *Cancer Res*. 2003;63(8):1814-1817.
30. Tamanaha R, Camacho CP, Pereira AC, da Silva A, Maciel R, Cerutti JM. Evaluation of RET polymorphisms in a six-generation family with G533C RET mutation: specific RET variants may modulate age at onset and clinical presentation. *Clin Endocrinol (Oxf)*. 2009;71(1):56-64.
31. Sharma BP, Saranath D. RET gene mutations and polymorphisms in medullary thyroid carcinomas in Indian patients. *J Biosci*. 2011;36(4):603-611.
32. Zhang Y, Wang S, Chen X, Huang S, Li J. Quantitative assessment of the association between L769L and S836S polymorphisms at RET gene and medullary thyroid carcinoma risk. *Tumour Biol*. 2014;35(7):6641-6647.
33. Gimm O, Neuberg DS, Marsh DJ, et al. Over-representation of a germline RET sequence variant in patients with sporadic medullary thyroid carcinoma and somatic RET codon 918 mutation. *Oncogene*. 1999;18(6):1369-1373.
34. Ruiz A, Antinolo G, Fernandez RM, Eng C, Marcos I, Borrego S. Germline sequence variant S836S in the RET proto-oncogene is associated with low level predisposition to sporadic medullary

thyroid carcinoma in the Spanish population. *Clin Endocrinol (Oxf)*. 2001;55(3):399-402.
35. Berard I, Kraimps J-L, Savagner F, et al. Germline-sequence variants S836S and L769L in the RE arranged during Transfection (RET) proto-oncogene are not associated with predisposition to sporadic medullary carcinoma in the French population. *Clin Genet*. 2004;65(2):150-152.
36. Elisei R, Cosci B, Romei C, et al. RET exon 11 (G691S) polymorphism is significantly more frequent in sporadic medullary thyroid carcinoma than in the general population. *J Clin Endocrinol Metab*. 2004;89(7):3579-3584.
37. Wiench M, Włoch J, Wygoda Z, et al. RET polymorphisms in codons 769 and 836 are not associated with predisposition to medullary thyroid carcinoma. *Cancer Detect Prev*. 2004;28(4):231-236.
38. Baumgartner-Parzer SM, Lang R, Wagner L, et al. Polymorphisms in exon 13 and intron 14 of the RET protooncogene: genetic modifiers of medullary thyroid carcinoma? *J Clin Endocrinol Metab*. 2005;90(11):6232-6236.
39. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
40. Machens A, Frank-Raue K, Lorenz K, Rondot S, Raue F, Dralle H. Clinical relevance of RET variants G691S, L769L, S836S and S904S to sporadic medullary thyroid cancer. *Clin Endocrinol (Oxf)*. 2012;76(5):691-697.
41. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*. 1959;22(4):719-748.
42. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188.
43. Severskaia NV, Saenko VA, Il'in AA, et al. [RET and GFRA1 germline polymorphisms in medullary thyroid cancer patients]. *Mol Biol (Mosk)*. 2006;40(3):425-435.
44. Chedid M, Michieli P, Lengel C, Huppi K, Givol D. A single nucleotide substitution at codon 31 (Ser/Arg) defines a polymorphism in a highly conserved region of the p53-inducible gene WAF1/CIP1. *Oncogene*. 1994;9(10):3021-3024.
45. Balbin M, Hannon GJ, Pendas AM, et al. Functional analysis of a p21WAF1, CIP1, SDI1 mutant (Arg94 → Trp) identified in a human breast carcinoma. Evidence that the mutation impairs the ability of p21 to inhibit cyclin-dependent kinases. *J Biol Chem*. 1996;271(26):15782-15786.
46. Ma H, Zhou Z, Wei S, Wei Q. Association between p21 Ser31Arg polymorphism and cancer risk: a meta-analysis. *Chin J Cancer*. 2011;30(4):254-263.
47. May A, Sun Y, Hildesheim A, Point NO, et al.Mutation carcinoma but a codon polymorphism of the tumor suppressor gene in nasopharyngeal (N PC): The polymorphism distinguishes caucasians from Chinese. 1995;4:261-267.
48. Johnson GG, Sherrington PD, Carter A, et al. A novel type of p53 pathway dysfunction in chronic lymphocytic leukemia resulting from two interacting single nucleotide polymorphisms within the p21 gene. *Cancer Res*. 2009;69(12):5210-5217.
49. Wang Z, Sturgis EM, Zhang F, et al. Genetic variants of p27 and p21 as predictors for risk of second primary malignancy in patients with index squamous cell carcinoma of head and neck. *Mol Cancer*. 2012;11:17.
50. Roh J-W, Kim BK, Lee CH, et al. P53 codon 72 and p21 codon 31 polymorphisms and susceptibility to cervical adenocarcinoma in Korean women. *Oncol Res*. 2010;18(9):453-459.
51. Wang N, Wang S, Zhang Q, et al. Association of p21 SNPs and risk of cervical cancer among Chinese women. *BMC Cancer*. 2012;12:589.

SUPPORTING INFORMATION

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

How to cite this article: Mishra V, Kowtal P, Rane P, Sarin R. Genetic risk association of CDKN1A and RET gene SNPs with medullary thyroid carcinoma: Results from the largest MTC cohort and meta-analysis. *Cancer Med*. 2019;00:1–11. https://doi.org/10.1002/cam4.2443

Supplementary Table S1: List of SNPs of genes of distinct pathways selected for the study based on detailed literature search

Gene/SNP	Reference ID	Variant	Nucleotide change	Ancestral Allele	Amino acid Change
Genes of Detoxification					
Cyp1A1m1	rs4646903	3'UTR	T/C	T	-
Cyp1A2	rs762551	Intronic	A/C	A	-
NAT2	rs1041983	Synonymous	C/T	C	Y94Y
GSTP1	rs1695	Missense	A/G	A	I105V
Cell Cycle regulatory genes					
CDKN1A	rs1801270	Missense	C/A	C	S31R
CDKN1B	rs2066827	Missense	T/G	T	V109G
CDKN2A	rs11515	3' UTR	G/C	G	-
CDKN2B	rs1063192	3'UTR	T/C	T	-
CDKN2C	rs12885	3' UTR	G/T	G	-
RET Proto oncogene					
G691S RET	rs1799939	Missense	G/A	G	G691S
L769L RET	rs1800861	Synonymous	T/G	T	L769L
S836S RET	rs1800862	Synonymous	C/T	C	S836S
S904S RET	rs1800863	Synonymous	C/G	C	S904S

Supplementary Table S2: Hardy-Weinberg Equilibrium (HWE) Calculations for Cases (n=438) and Control (n=489) [MAF: Minor Allele Frequency]

Gene/ SNP	Controls (n=489)		MTC cases(n=438)	
	MAF	***HWE p-value***	***MAF***	***HWE p-value***
Cyp1A1m1	0.34	0.35	0.31	**0.006**
Cyp1A2	0.47	0.35	0.43	0.65
NAT2	0.43	0.18	0.45	1.00
GSTP1	0.29	0.10	0.26	0.65
CDKN1A	0.15	0.18	0.09	0.14
CDKN1B	0.32	0.57	0.29	0.57
CDKN2A	0.05	0.76	0.07	**0.003**
CDKN2B	0.27	0.09	0.25	0.11
CDKN2C	0.11	0.69	0.10	0.31
G691S RET	0.24	0.84	0.27	0.17
L769L RET	0.40	0.57	0.35	0.60
S836S RET	0.08	0.74	0.08	0.33
S904S RET	0.24	0.75	0.29	0.05

Supplementary Table S3: Genotype frequency comparison between: **(A):** hereditary MTC (hMTC) cases and controls; **(B):** sporadic MTC (sMTC) cases and controls

(A)

Gene/ SNP	Genotype Frequency – hMTC cases (n=77)		Genotype Frequency – Controls (n=489)		
	Wt	*Hz + Hm*	*Wt*	*Hz + Hm*	*p-value*
Cyp1A1m1	37 (48.1%)	40 (51.9%)	218 (44.6%)	271 (55.4%)	0.57
Cyp1A2	28 (36.4%)	49 (63.6%)	145 (29.7%)	344 (70.3%)	0.23
NAT2	31 (42.2%)	46 (59.7%)	167 (34.2%)	322 (65.8)	0.29
GSTP1	35 (45.5%)	42 (54.5%)	251 (51.3%)	238 (48.7%	0.34
CDKN1A	**65 (84.4%)**	**12 (15.6%)**	**361 (73.8%)**	**128 (26.2%)**	**0.045**
CDKN1B	35 (45.5%)	42 (54.5%)	224 (45.8%)	265 (54.2%)	0.96
CDKN2A	67 (87%)	10 (13%)	439 (89.7%)	50 (10.2%)	0.46
CDKN2B	46 (59.8%)	31 (40.2%)	266 (54.4%)	223 (45.6%)	0.38
CDKN2C	62 (80.5%)	15 (19.5%)	387 (79.1%)	102 (20.8%)	0.78
G691S RET	37 (48.1%)	40 (51.9%)	283 (57.8%)	206 (42.1%)	0.11
L769L RET	35 (45.5%)	42 (54.5%)	178 (36.4%)	311 (36.6%)	0.13
S836S RET	64 (83.1%)	13 (16.9%)	411 (84%)	78 (15.9%)	0.84
S904S RET	36 (46.8%)	41 (53.2)	285 (58.3%)	204 (41.7%)	0.06

(B)

Gene/ SNP	Genotype Frequency – sMTC cases (n=361)		Genotype Frequency – Controls (n=489)		
	Wt	*Hz + Hm*	*Wt*	*Hz + Hm*	*p-value*
Cyp1A1m1	161 (44.6%)	200 (553.4%)	218 (44.6%)	271 (55.4%)	0.996
Cyp1A2	117 (32.4%)	244 (67.6%)	145 (29.7%)	344 (70.3%)	0.390
NAT2	103 (28.5%)	258 (71.5%)	167 (34.2%)	322 (65.8)	0.082
GSTP1	208 (57.6%)	153 (42.4%)	251 (51.3%)	238 (48.7%	0.069
CDKN1A	**295 (81.7%)**	**66 (18.3%)**	**361 (73.8%)**	**128 (26.2%)**	**0.007**
CDKN1B	182 (50.4%)	179 (49.6%)	224 (45.8%)	265 (54.2%)	0.184
CDKN2A	310 (85.9%)	51 (14.1%)	439 (89.7%)	50 (10.2%)	0.083
CDKN2B	206 (57%)	155 (42.9%)	266 (54.4%)	223 (45.6%)	0.439
CDKN2C	297 (82.2%)	64 (17.7%)	387 (79.1%)	102 (20.8%)	0.256
G691S RET	201 (55.7%)	160 (44.3%)	283 (57.8%)	206 (42.1%)	0.523
L769L RET	146 (40.4%)	215 (59.6%)	178 (36.4%)	311 (36.6%)	0.231
S836S RET	311 (86%)	50 (13.9%)	411 (84%)	78 (15.9%)	0.398
S904S RET	194 (53.7%)	167 (46.3%)	285 (58.3%)	204 (41.7%)	0.187

Supplementary Table S4: (A): Details on sporadic MTC studies included in the meta-analysis. **(B):** Details on hereditary MTC studies included in the meta-analysis.

(A)

Author	Year	Population	Variant	Cases	Controls
Gimm	1999	German+USA	G691S, L769L, S836S, S904S	50	70
Ruiz	2001	Spain	S836S	32	250
Berard	2004	France	L769L, S836S	92	86
Elisei	2004	Italy	G691S, L769L, S836S, S904S	106	106
Wiench	2004	Poland	L769L, S836S	135	90
Baumgartner	2005	Austria	L769L, S836S	45	79
Cebrian	2005	British	G691S, L769L, S836S, S904S	124	517
Costa	2005	Portugal	G691S, L769L, S836S, S904S	50	50
Severskava	2006	Russian	G691S, L769L, S836S, S904S	45	178
Fernandez	2006	Spanish	G691S, S904S	58	100
Fugazzola	2008	Italy	G691S, L769L, S836S, S904S	140	190
Sromek	2010	Poland	L769L, S836S, S904S	217	420
Ceolin	2012	Brazil	G691S, L769L, S836S, S904S	107	308
Lantieri	2013	Italy	G691S	78	85
Present study	2019	Indian	G691S, L769L, S836S, S904S, Cyp1A1m1, Cyp1A2, NAT2, GSTP1, CDKN1A, CDKN1B, CDKN2A, CDKN2B, CDKN2C	361	489
Total				1640	2968

(B)

Author	Year	Population	Variant	Cases	Controls
Robledo	2003	Spanish	G691S	104	653
Severskava	2006	Russian	G691S, L769L, S836S, S904S	22	178
Tamanaha	2009	Brazilian	G691S, L769L, S836S, S904S	77	100
Sharma	2011	Indian	G691S, L769L, S836S, S904S	51	50
Lantieri	2013	Italian	G691S	15	85
Present Study	2019	Indian	G691S, L769L, S836S, S904S, Cyp1A1m1, Cyp1A2, NAT2, GSTP1, CDKN1A, CDKN1B, CDKN2A, CDKN2B, CDKN2C	77	489
Total				346	1555