

Prioritization of predisposition genes for familial non-medullary thyroid cancer by whole-genome sequencing

Aayushi Srivastava,^{1,2,†} Diamanto Skopelitou,^{1,2} Beiping Miao,¹ Sara Giagiobbe,^{1,3}
Nagarajan Paramasivam,⁴ Abhishek Kumar,^{1,5} Chiara Diquigiovanni,⁶ Elena Bonora,^{6,7}
Obul Reddy Bandapalli,^{1,2} Asta Försti,^{1,8,9,*‡} and Kari Hemminki^{1,10,11,*‡}

¹Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany

²Heidelberg University, Medical Faculty, Heidelberg 69120, Germany

³Department of Medicine I and Clinical Chemistry, University Hospital of Heidelberg, Heidelberg 69120, Germany

⁴Computational Oncology, Molecular Diagnostics Program, National Center for Tumor Diseases (NCT), Heidelberg 69120, Germany

⁵Clinical Cooperation Unit Molecular Hematology-Oncology, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany

⁶Unit of Medical Genetics, Department of Medical and Surgical Sciences, University of Bologna, Bologna 40126, Italy

⁷IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna 40126, Italy

⁸Hopp Children's Cancer Center (KiTZ), Heidelberg 69120, Germany

⁹Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), Heidelberg 69120, Germany

¹⁰Biomedical Center, Faculty of Medicine in Pilsen, Charles University in Prague, Pilsen 32300, Czech Republic

¹¹Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany

*Corresponding author: Asta Försti, Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, Heidelberg 69120, Germany. Email: a.foersti@kitz-heidelberg.de; Kari Hemminki, Biomedical Center, Faculty of Medicine, Charles University, Alej Svobody 1655/76, Pilsen 32300, Czech Republic. Email: k.hemminki@dkfz.de

Abstract

Objective: Thyroid cancer (TC) is the most common endocrine malignancy, with 90%–95% of the cases representing non–medullary thyroid cancer (NMTC). Familial cases account only for a few of all cases and the underlying genetic causes are still poorly understood.

Methods: We whole-genome sequenced affected and unaffected members of an Italian NMTC family and applied our in-house developed Familial Cancer Variant Prioritization Pipeline (FCVPPv2) which prioritized 12 coding variants. We refined this selection using the VarSome American College of Medical Genetics and Genomics (ACMG) implementation, SNAP2 predictions and further *in silico* scores.

Results: We prioritized 4 possibly pathogenic variants in 4 genes including *Ret proto-oncogene (RET)*, polypeptide *N*-acetylgalactosaminyltransferase 10 (*GALNT10*), *ubiquitin-1 (UBN1)*, and *prostaglandin I2 receptor (PTGIR)*. The role of *RET* point mutations in medullary thyroid carcinoma is well established. Similarly, somatic rearrangements of *RET* are known in papillary TC, a specific histotype of NMTC. In contrast to *RET*, no germline variants in *PTGIR*, *GALNT10*, or *UBN1* have been linked to the development of TC to date. However, alterations in these genes have been shown to affect pathways related to cell proliferation, apoptosis, growth, and differentiation, as well as posttranslational modification and gene regulation. A thorough review of the available literature together with computational evidence supported the interpretation of the 4 shortlisted variants as possibly disease-causing in this family.

Conclusions: Our results implicate the first germline variant in *RET* in a family with NMTC as well as the first germline variants in *PTGIR*, *GALNT10*, and *UBN1* in TC.

Keywords: RET, GALNT10, PTGIR, UBN1, familial non–medullary thyroid cancer, germline variant, whole-genome sequencing

Significance

Germline genetics of familial non–medullary thyroid cancer (NMTC) is poorly known. We applied whole-genome sequencing in an Italian NMTC family followed by stringent bioinformatics filtering and variant prioritization pipeline to identify novel predisposition genes for NMTC. Our results implicated the first germline variant in the proto-oncogene *Ret proto-oncogene (RET)* in a family with NMTC as well as the first germline variants in *prostaglandin I2 receptor (PTGIR)*, *polypeptide N-acetylgalactosaminyltransferase 10 (GALNT10)*, and *ubiquitin-1 (UBN1)* in thyroid cancer (TC). Although these variants and genes were identified in only one family, their involvement in cancer-related pathways and their clustering in a family suggest that these variants may jointly contribute on NMTC development in this family.

† Present address: Goethe University Frankfurt, University Hospital, Medical Clinic 1, Frankfurt Am Main, Germany

‡ A.F. and K.H. contributed equally.

Introduction

Thyroid cancer (TC) is the most common endocrine malignancy with an estimated 586 000 new cases and 43 600 deaths in 2020, making it the ninth most frequent cancer in that year.¹ Non-medullary thyroid carcinoma (NMTC) is the most common type of TC accounting for over 95% of all TC cases. Familial NMTC (FNMTC) accounts for a few percent of NMTC and can be categorized into syndromic and non-syndromic forms.^{2,3} Familial risk of NMTC is one of the highest of all cancers and there is a difference in risk between the parent and offspring (3.8), and between siblings (5.0).⁴ Familial risk is highest for the concordant type of NMTC.⁵ Germline susceptibility genes for syndromic forms are known, although these entities are very rare.⁶ A few potential susceptibility genes have been identified in families and low-risk variants have been identified in genome-wide association studies.⁶⁻⁹ In general, the genetic basis of non-syndromic FNMTC is still not well understood and is the subject of our research.

In recent years different approaches have been implemented to understand the genetic basis of FNMTC. These include genome-wide association studies, linkage analyses, targeted sequencing, whole-exome, and whole-genome sequencing (WGS). Several genes and loci, including low-penetrance variants near or in *FOXE1*, *DIRC3*, *SRGAP1*, *TITF-1/NKX2-1*, *NF1*, and *CHEK2*, have been suggested to affect non-syndromic FNMTC susceptibility.^{6,8,9} Our research group has performed WGS on 5 Italian NMTC families to identify rare, high-to-moderate-penetrance variants. In a previous study, we identified and functionally validated a pathogenic variant in the *POT1* gene (p.V29L).¹⁰ More recently, we identified potentially disease-causing variants in *CHEK2*, *EWSR1*, and *TIAM1*.¹¹

In this study, we analyzed WGS data of a family, in which multiple individuals were diagnosed with FNMTC, and short-listed variants using our in-house developed Familial Cancer Variant Prioritization Pipeline (FCVPPv2)¹² in conjunction with additional *in silico* tools. We identified 4 novel potentially disease-causing germline variants in *prostaglandin I2 receptor (PTGIR)*, *Ret proto-oncogene (RET)*, *polypeptide N-acetylgalactosaminyltransferase 10 (GALNT10)*, and *ubiquitin-1 (UBN1)*. Although available literature suggests the involvement of the identified genes in cancer pathways, no germline cancer predisposition variants have yet been reported in *PTGIR*, *GALNT10*, or *UBN1*. To our knowledge, we report the first of such germline variants in these 3 genes.

Methods

Patients

A family of Italian origin with NMTC aggregation was recruited through the International Consortium for the Genetics of Non-Medullary Thyroid Carcinoma. The epidemiological and clinical characteristics of this family collection have been reported elsewhere.¹³ Briefly, the criterion for eligibility of the families was to have at least 2 first- or second-degree relatives affected with NMTC. Family members having benign thyroid diseases, such as goiters, adenomas, or thyroiditis, were also sampled. The study was conducted in accordance with the Declaration of Helsinki. It was approved by the Committee for Protection of Persons in Biomedical Research of Lyon (CCPRB A-96.18), by the IARC Ethical Review Board (Project 101 95-050,

amendment 01-013) and by the Comitato Etico Indipendente dell'Azienda Ospedaliero-Universitaria di Bologna, Policlinico S. Orsola-Malpighi (prot n. 736/2016). Written consent was obtained from each patient or subject after full explanation of the purpose and nature of all procedures used and a blood sample (10 mL) was drawn from family members who agreed to participate to the study. A clinical questionnaire was also proposed to the patients, and their pedigrees were constructed. For the Italian family presented in this study, patients III-1 and III-3 developed papillary thyroid cancer (PTC) at an early age, as indicated in Figure 1A, and were subjected to partial/total thyroidectomy following diagnosis. Patients IV-3 and IV-4 had also a total thyroidectomy after diagnosis at an early age. Patient III-4 developed PTC-positive nodules at an older age (54 years) compared to her spouse and offspring and was subsequently operated for breast cancer. She was included in the study to exclude the possibility of recessive inheritance of the disease in her children. WGS was conducted on genomic DNA derived from peripheral blood of 3 affected members (III-1, III-3, and IV-3), one unaffected member (no thyroid disease[s] or nodules at ultrasound IV-2), one potential carrier with benign nodules and no evidence of cancer at time of analysis (IV-1), and one family member by marriage (III-4, see Figure 1A). Genomic DNA was extracted from blood using the QiAMP DNA Blood Mini kit following the manufacturer's instructions.

Whole-genome sequencing

WGS was performed using Illumina-based short-read sequencing. BWA mem (version 0.7.8) was used to map the results to the human reference genome (assembly GRCh37 version hs37d5), and duplicates were removed with biobambam (version 0.0.148).¹⁴ Small nucleotide variants and InDels were called with Platypus through joint calling on all samples from the family.¹⁵ All variants were annotated using data from ANNOVAR, 1000 Genomes, dbSNP, and ExAC.¹⁶⁻¹⁹ We retained variants based on QUAL scores (>20) and coverage (>5×), as well as those that passed all of the internal Platypus filters. We selected rare variants with minor allele frequency (MAF) < 0.1% in the 1000 Genomes Phase 3 and non-TCGA ExAC data. To search for sample swaps and family relatedness, a pairwise comparison of shared rare variants was performed.

Analysis and prioritization of coding variants

We filtered and prioritized variants according to the criteria of the FCVPPv2.^{12,20} For the present study, we updated the MAF data using the gnomAD_genomes_AF, gnomAD_genomes_NFE_AF and gnomAD_genomes_POPMAX_AF < 0.1% (gnomAD v.4.0.0; <https://gnomad.broadinstitute.org>). Based on variant segregation with the disease in the family, the variant should be present in cancer patients, it may be present in individuals with benign nodules and it should not be present in unaffected individuals. We selected the variants belonging to the top 10% of probable deleterious variants in the human genome according to the web-based CADD tool v1.3.²¹ Variants predicted to be deleterious by at least 60% of 12 tools accessed using dbNSFP, namely SIFT, PolyPhen V2-HDV, PolyPhen V2-HVAR, LRT, MutationTaster, Mutation Assessor, FATHMM, MetaSVM, MetLR, VEST3, PROVEAN, and RI, were selected.²² Evolutionary conservation of the genomic position of the variants was evaluated using 3 scores, Genomic Evolutionary Rate Profiling (GERP, >2), PhastCons (>3), and PhyloP (≥0.3).²³⁻²⁵

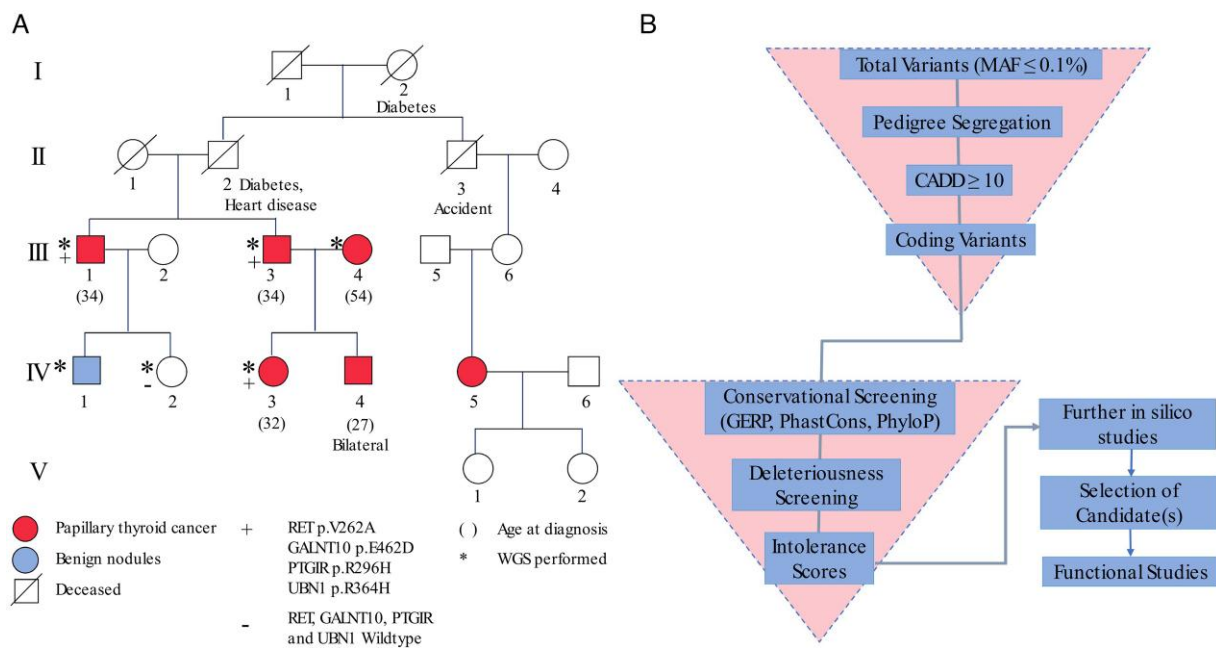


Figure 1. (A) Pedigree of the NMTC family with *RET*, *GALNT10*, *PTGIR*, and *UBN1* variants. (B) An overview of the analysis of coding variants following the FCVPPv2.

We used scores derived from NHLBI-ESP6500, ExAC and a local dataset, all of which were developed using allele frequency data, to evaluate the intolerance of genes to functional variants.²⁶ Two other scoring systems developed by the ExAC consortium with the help of large-scale exome sequencing data were regarded for loss-of-function (LOF) variants (pLI) and missense variants (Z-scores).

In silico studies, protein alignment, and structural modeling

We used additional *in silico* tools for final selection of the candidate predisposition variants. We obtained the American College of Medical Genetics and Genomics (ACMG) classification of the selected variants via Varsome.²⁷ DANN, ClinPred, and REVEL scores were generated using the web-based tool OpenCRAVAT.²⁸ DANN is a deep learning-based approach to annotate pathogenicity of genetic variants and uses the same feature set and training data as CADD to train a deep neural network.²⁹ REVEL is an ensemble method for predicting the pathogenicity of missense variants based on a combination of scores from 13 individual tools, that are also part of the FCVPPv2.³⁰ Both REVEL and DANN scores can range from 0 to 1, with higher scores showing greater probability that the variant is disease-causing. ClinPred is a machine learning-based tool for identifying disease-relevant nonsynonymous variants. ClinPred scores >0.5 are defined as damaging.

We studied the tolerance of the proteins to amino acid (AA) substitution of the shortlisted variants using SNAP2, a neural network-based classifier.³¹ A heat map representation was generated and the area surrounding the variant as well as the variant itself was evaluated.

The conservation of the variants across species was assessed further. Orthologue gene groups for *PTGIR*, *RET*, *UBN1*, and *GALNT10* were obtained from the National Center for Biotechnology Information (NCBI). Multiple homologous sequences were selected, aligned, visualized

and formatted by COBALT, a constraint-based multiple alignment tool.³²

The structural stability of the protein in response to the respective variant was evaluated using the mutation Cutoff Scanning Matrix (mCSM).³³ We accessed the PDB entry and 3D representation of human RET (4ux8) and human GALNT10 (2d7i) from the Protein Data Bank in Europe.³⁴ PDB files of the complete human PTGIR and UBN1 proteins were not found and not evaluated for structural stability.

Validation of candidate variants

The sequencing data for the shortlisted variants were rechecked manually using the Integrative Genomics Viewer.³⁵ We screened WGS data from the other 4 NMTC families that were part of our previous studies for rare variants in the selected genes,²⁰ and found none that passed the criteria of FCVPPv2.

Results

Whole-genome sequencing and variant prioritization

This study is based on an Italian family, with 5 members diagnosed with PTC across 2 generations and 1 member with benign nodules (Figure 1A). Three family members diagnosed with PTC were subjected to WGS (III-1, III-3, and IV-3), together with the family member with benign nodules (IV-1) and one healthy family member (IV-2). Additionally, we sequenced III-4, wife of the affected family member III-3, and mother of two affected family members IV-3 and IV-4. She was also affected by TC, however, at an older age (54 years) compared to her husband (34 years) and his first degree relatives (27, 32, and 34 years). The variants were filtered based on the pipeline shown in Figure 1B.

WGS of the family identified a total of 207 873 variants with a MAF < 0.1%. This number was reduced by pedigree-based filtering to 9373. Application of FCVPPv2 resulted in

Table 1. Top 7 shortlisted variants with SNAP2, DANN, REVEL, and ClinPred scores as well as ACMG classifications.

Gene	Variant position	HGVS ^a	SNAP2 predicted effect	SNAP2 score	SNAP2 expected accuracy (%)	DANN score ^b	REVEL score ^b	ClinPred ^c	Varsome ACMG classification
<i>PTGIR</i>	19_47124811_C_T	Arg296His	Effect	82	91	0.9996	0.299	0.9810	VUS
<i>GALNT10</i>	5_153789322_G_C	Glu462Asp	Effect	17	59	0.9964	0.279	0.9700	Pathogenic
<i>UBN1</i>	16_4911084_G_A	Arg364His	Effect	30	66	0.9993	0.353	0.5095	VUS with minor pathogenic evidence
<i>RET</i>	10_43600559_T_C	Val262Ala	Effect	1	53	0.9908	0.748	0.1109	VUS with some pathogenic evidence
<i>KLHL18</i>	3_47385160_A_G	Tyr485Cys	Effect	36	66	0.9973	0.809	0.9820	VUS with minor pathogenic evidence
<i>OSGIN2</i>	8_90921899_A_T	Glu50Val	Effect	50	28	0.9883	0.205	0.1219	VUS
<i>GSR</i>	8_30585111_C_T	Arg81His	Neutral	-85	93	0.9994	0.313	0.9161	VUS

^aAmino acid change based on the human genome variant society (HGVS) nomenclature.

^bScores can range from 0 to 1, with higher scores showing greater probability that the variant is disease-causing.

^cScores >0.5 are defined damaging.

the prioritization of 12 coding variants that segregated with the disease²⁰; the summary of the variants is shown in Table S1.

After inclusion only the variants with CADD >20 (representing the top 1% of probable deleterious variants according to CADD) and exclusion of the variants with low intolerance scores (<60% of scores indicating intolerance), 7 variants remained (Table 1). These variants were further analyzed using SNAP2 prediction on impact of AA substitutions on proteins as well as DANN, REVEL, and ClinPred tools that have been shown to be very strong ensemble methods, and classified using ACMG criteria (Table 1). However, since DANN and REVEL are already included in the ACMG Varsome analysis, we did not weight these points as heavily to avoid double counting evidence. For these variants and their related genes, we looked at the available literature thoroughly and weighted this information strongly. We evaluated whether the loss-of-function is a known mechanism in genes with variants predicted to be null variants. We evaluated if the gene is known to play a role in cancer and which domains have been implicated so far. Another factor was whether the gene is involved in a pathway that has been implicated in the carcinogenesis of thyroid or other cancers. Based on these data, we prioritized the variants in the genes *PTGIR*, *RET*, *GALNT10*, and *UBN1*. All the 4 variants were present in all TC cases, but not in the healthy individual or the family member by marriage (Figure 1A).

PTGIR and RET are involved in MAPK/ERK and PI3K/AKT pathways

Two variants located in the genes *PTGIR* and *RET*, respectively, may have an effect on the mitogen-activated protein kinase/extracellular signal-regulated kinase (MAPK/ERK) and the phosphatidylinositol 3-kinase/AKT serine/threonine kinase (PI3K/AKT) pathways. The tolerance of AA substitutions at position 296 of the *PTGIR* protein was predicted by SNAP2, which showed that the arginine to histidine substitution has a very strong effect with a score of 82 and expected accuracy of 91% (Table 1). The region immediately preceding this position is also shown to be highly intolerant to AA substitutions as shown in the heat map (Figure 2A). Comparative

sequence analysis of *PTGIR* showed the position (R296H) and the region preceding it to be highly conserved among selected representative species within the phylogeny (Figure 2B). As expected, the intolerant region in the SNAP2 analysis (Figure 2A, red) and the highly conserved region in Figure 2B coincide with each other. *PTGIR* has an extracellular N-terminus, followed by 7 transmembrane α -helices that are connected by 3 intracellular and 3 extracellular loops (Figure 2C). The variant is located immediately after the seventh helix of *PTGIR* in the intracellular space; a 3D model of *PTGIR* is shown in Figure 2D.

The second gene we analyzed further was *RET*. Although the newly created scores were not very strong (Table 1), we included *RET* in our further evaluation, as it is a proto-oncogene with a well-known role in medullary TC (MTC).³⁷ Furthermore, somatic rearrangements of *RET* can cause PTC, a specific histotype of NMTC.³⁷ The original FCVPPv2 analysis prioritized the V262A variant with a CADD score of 26.3, with supporting information from intolerance and deleteriousness scores (Table S1). The SNAP2 analysis showed a weak effect (score = 1) for the valine to alanine change (Table 1). Most AA changes of this position and the area immediately after it are predicted to be intolerant (Figure 3A; shades of red), in contrast with the region prior to position 262, which is shown to be tolerant to substitutions (blue). Multiple sequence alignment of amino acids 257-285 of *RET* from selected vertebrate species indicated the position of our variant to be highly conserved (Figure 3B). The 3D crystal structure of *RET* (PDB entry 4ux8) is shown in Figure 3C. We utilized the PDB entry for the mCSM analysis, which showed the V262A variant to cause destabilization of the *RET* protein, by predicting a negative thermodynamic change in Gibb's free energy ($\Delta\Delta G = -2.388$ kcal/mol, Figure 3D).

Since *RET* is an established proto-oncogene, many germline variants in *RET* have already been implicated and functionally validated in MTC and PTC.^{37,38} Also, germline LOF variants in *RET* are found in Hirschsprung disease.³⁹ These variants and their positions can be seen in Figure 4. Known variants in multiple endocrine neoplasia, type 2 (MEN2) were checked with the MEN2 *RET* database developed by the ARUP Institute for Clinical and Experimental Pathology and last updated in 2020.³⁸ The majority of germline variants,

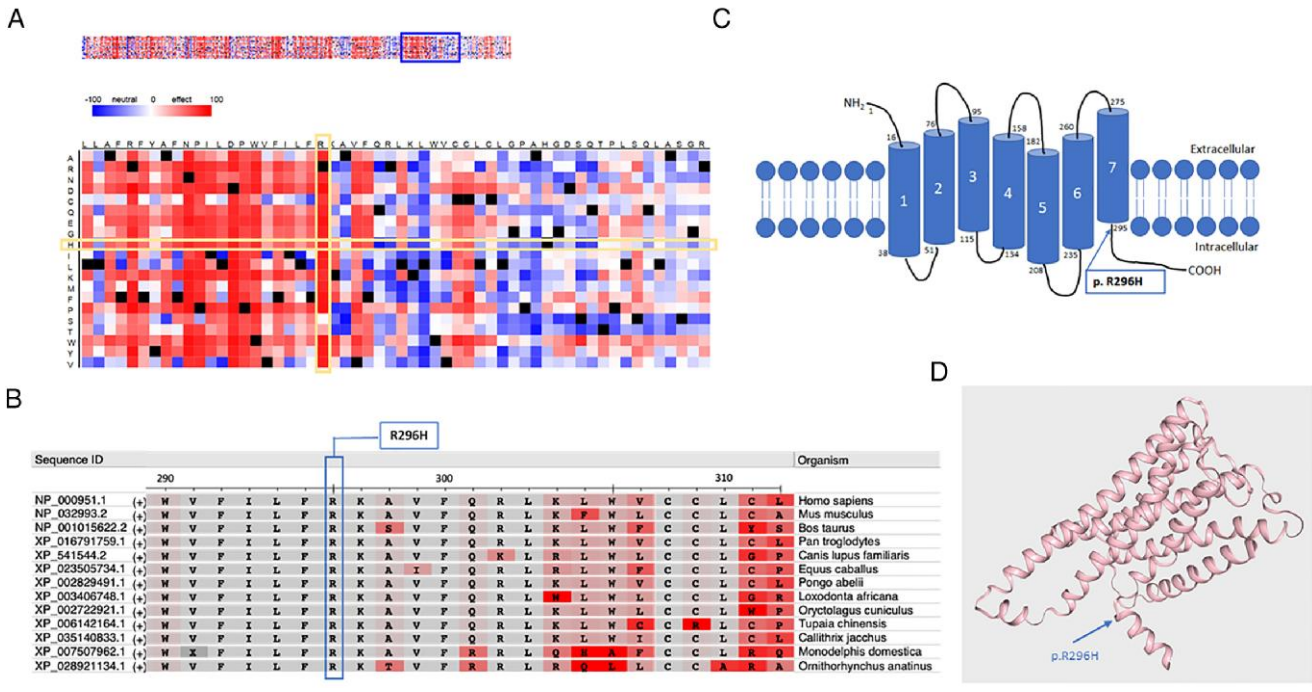


Figure 2. *In silico* studies of PTGIR p. R296H. (A) Heat map representation of SNAP2 results depicting the likely impact of individual AA substitutions (y-axis) for each position (x-axis) on protein function. Substitutions with strong effects are depicted with dark red (score = 100), whereas weak substitutions are represented with blue (score = -100). White represents neutral substitutions. Black represents the corresponding wild-type residue. The R296H variant is shown with yellow rectangles. (B) Comparative sequence analysis across representative phylogeny. The variant site is shown with a blue box. The organisms and their sequence IDs are also included. Differences are highlighted in red. (C) Schematic diagram of PTGIR with all its domains. The variant identified in this study is marked in blue box. (D) 3D structure of PTGIR obtained via SwissModel (SMTL ID: 7d7m.1).³⁶ The position of p. R296H is marked with a blue arrow.

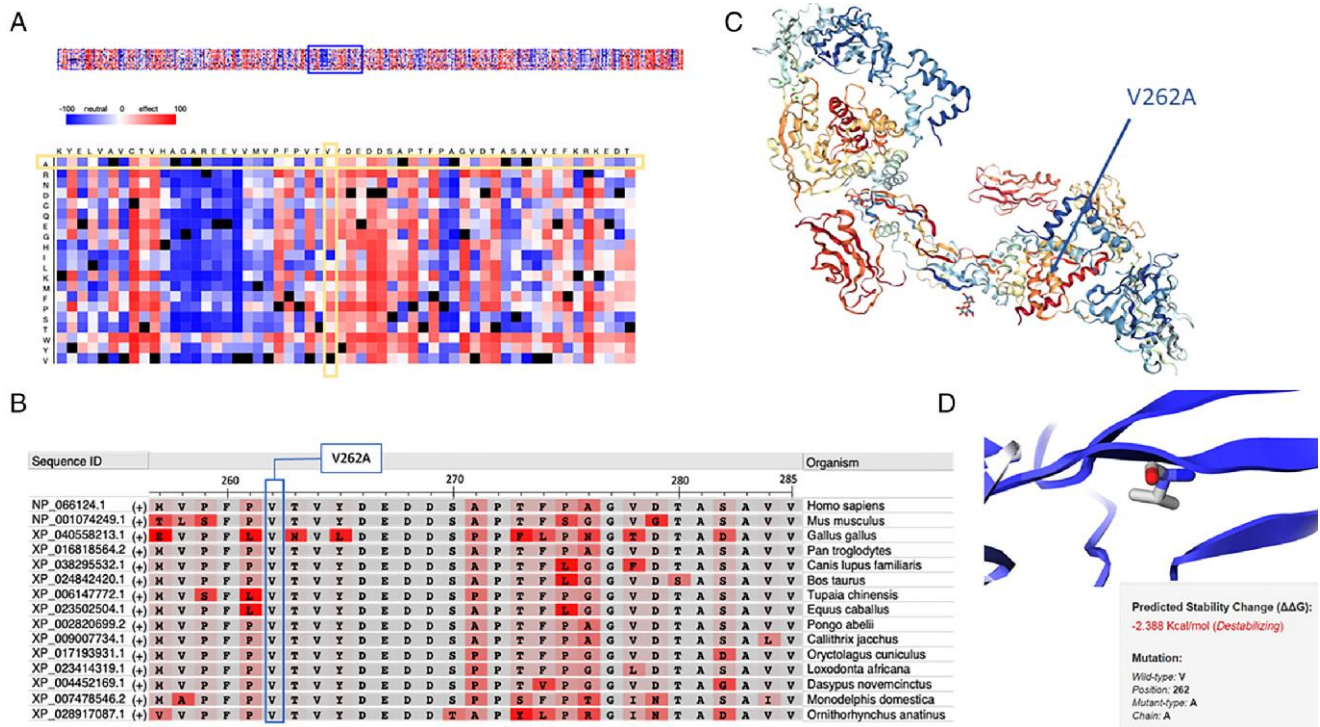


Figure 3. *In silico* studies of RET p.V262A. (A) Heat map representation of SNAP2 results depicting the likely impact of individual AA substitutions (y-axis) for each position (x-axis) on protein function. Substitutions with strong effects are depicted with dark red (score = 100), whereas weak substitutions are represented with blue (score = -100). White represents neutral substitutions. Black represents the corresponding wild-type residue. The V262 position is shown with yellow rectangles. (B) Comparative sequence analysis across representative phylogeny. The position of the variant is shown in blue box. Differences are highlighted in red. (C) 3D crystal structure of PDB 4ux8; p.V262A is marked with an arrow. (D) Thermodynamic change in Gibb's free energy caused by the p.V262A substitution as predicted by the mCSM.

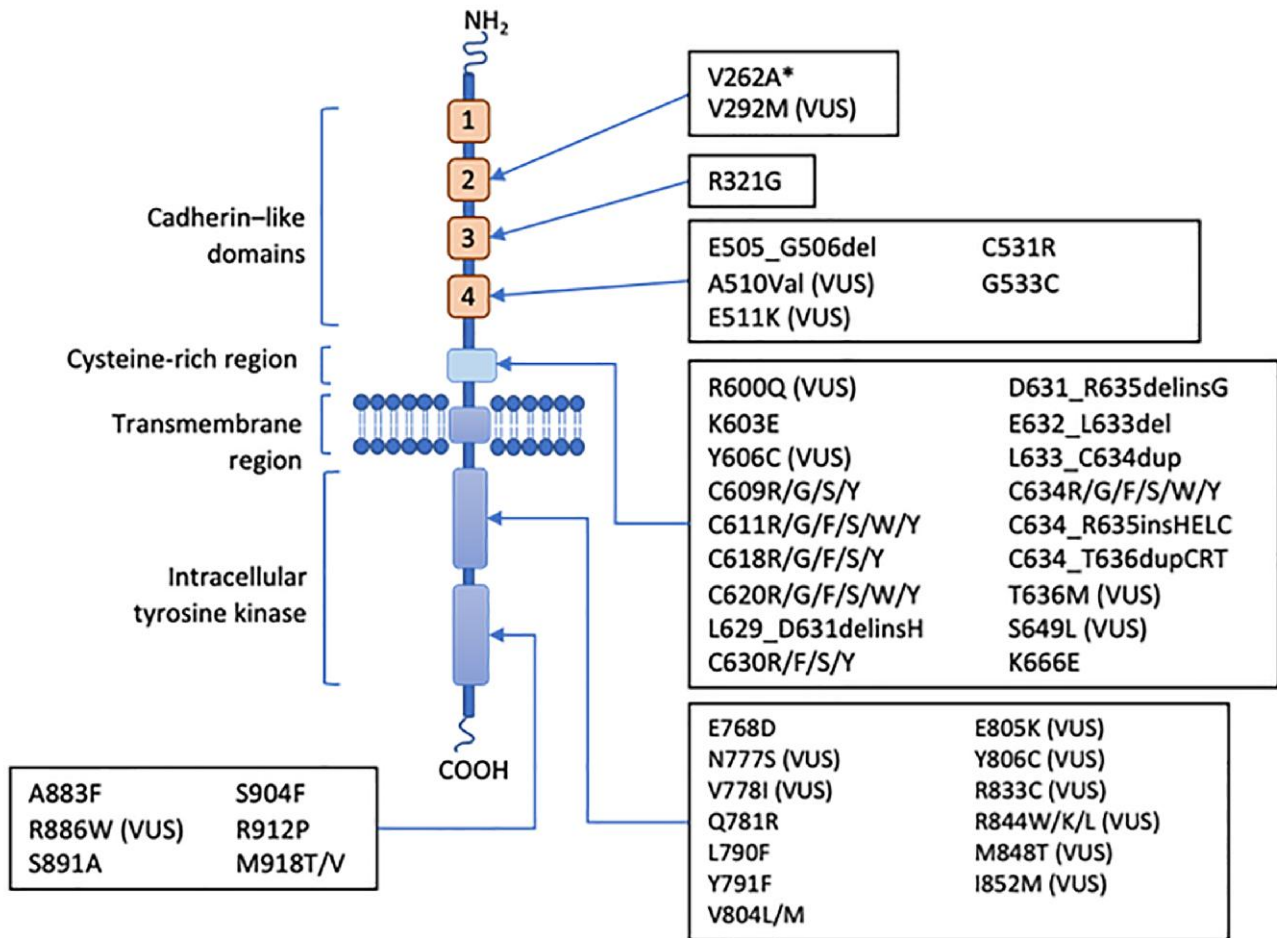


Figure 4. Schematic primary structure of the RET protein with its domains. Germline variants in RET identified thus far are shown next to their domains. The variant identified in the present study, p.V262A, is marked with an asterisk (*). Variants of unknown significance (VUS) are denoted as such.

especially those implicated in MTC and MEN2, are located in the extracellular cysteine-rich region or the intracellular tyrosine kinase domain (Figure 4). Only a few variants have been found in the cadherin-like domains of RET, where the V262A variant is located. So far, these have only been implicated in Hirschsprung disease.

Variants in *GALNT10* and *UBN1* are potentially predisposing to NMTC

Two other variants were located in the genes *GALNT10* and *UBN1*, respectively, that have functions relevant to cancer development. *GALNT10* is involved in posttranslational modification of proteins and *UBN1* has been implicated in chromatin regulation. For *GALNT10*, we attained relatively strong results from the *in silico* tools (Table 1). The ACMG classification of this variant was pathogenic. The SNAP2 results show that the substitution E462D has a slightly deleterious effect (score = 17). The region immediately surrounding this position is also shown to be intolerant to substitutions (Figure 5A). Comparative sequence analysis of amino acids 448-480 of *GALNT10* from selected representative species within the phylogeny shows the variant site and the residues surrounding it to be highly conserved (Figure 5B). The protein consists of an N-terminal glycosyltransferase domain involved in manganese coordination, substrate binding, catalytic

reaction, and UDP-Gal binding (Figure 5C). The C-terminal ricin B-type lectin domain, in which the variant identified in our study is located, binds to GalNAc and contributes to the glycopeptide specificity.⁴⁰ The 3D crystal structure of *GALNT10* along with the mCSM analysis results of this PDB entry are shown in Figure 5D and E, respectively. E462D was shown to cause destabilization of the *GALNT10* protein ($\Delta\Delta G = -1.609$ Kcal/mol).

For *UBN1*, SNAP2 predicted the arginine to histidine substitution at position 364 to have a moderately strong effect (score = 30; Table 1, Figure 6A). Comparative sequence analysis of *UBN1* showed the position (R364H) to be highly conserved among selected representative species within the phylogeny (Figure 6B). The variant identified in this study is located in the middle domain of *UBN1*. A schematic diagram of the structure is shown in Figure 6C, in which the domains and their interactions with histones and histone cell cycle regulator (HIRA) are shown.

Discussion

In recent years, WGS has emerged as a “state-of-the-art” tool in the identification of novel cancer-predisposing genes in Mendelian diseases. Unlike MTC, the genetic basis of FNMTTC is not yet well understood. A consensus as to the mode of inheritance, whether monogenic or polygenic or

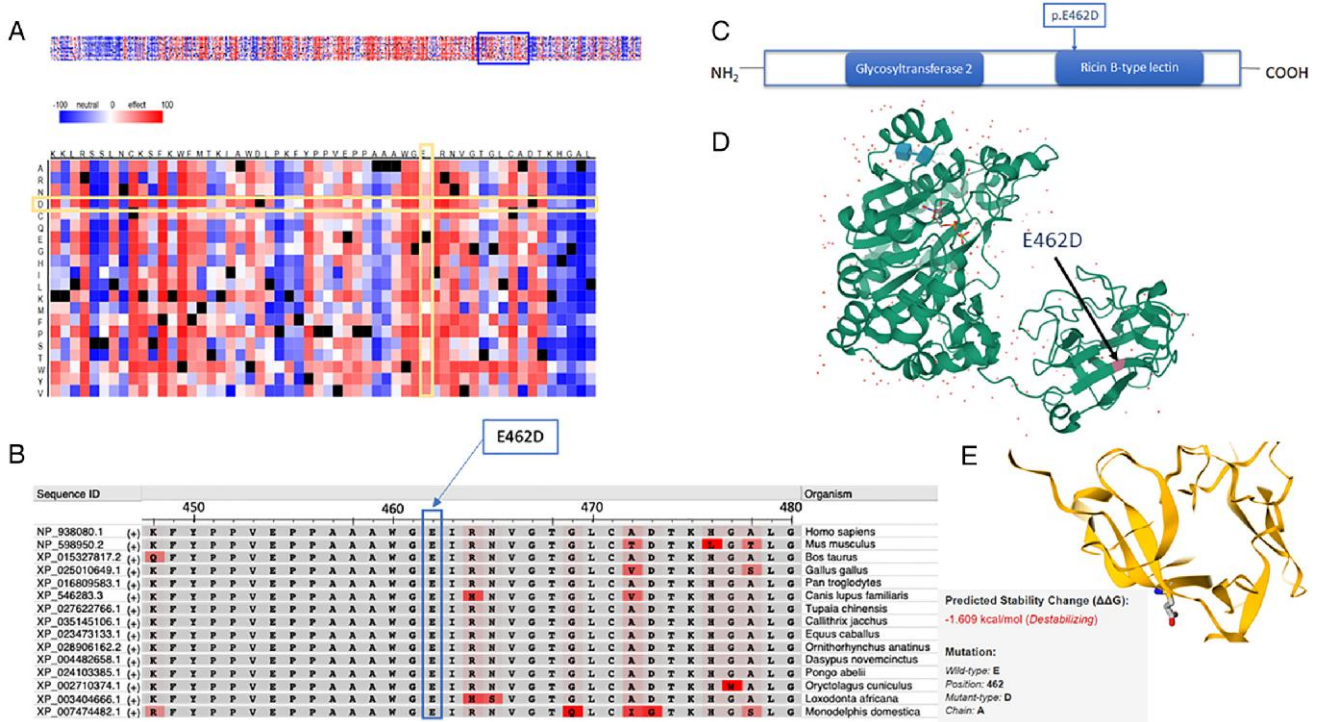


Figure 5. *In silico* studies of GALNT10 p.E462D. (A) Heat map representation of SNAP2 results depicting the predicted impact of individual AA substitutions (y-axis) for each position (x-axis) on protein function. Substitutions with strong effects are depicted with dark red (score = 100), whereas weak substitutions are represented with blue (score = -100). White represents neutral substitutions. Black represents the corresponding wild-type residue. The E462D variant is shown with yellow rectangles. (B) Comparative sequence analysis across representative phylogeny. The variant site is shown with a blue box. The organisms and their sequence IDs are also included. Differences are highlighted in red. (C) Schematic diagram of the GALNT10 protein with its domains. The variant identified in this study is marked with an arrow. (D) 3D crystal structure of PDB 2d7i. AA position 462 is marked with an arrow. (E) Thermodynamic change in Gibb’s free energy caused by the p.E462D substitution as predicted by the mCSM.

both, has yet to be reached.⁶ In the present study, we applied the FCVPPv2, followed by additional *in silico* tool evaluation, to prioritize WGS-identified variants in an Italian NMTC family. This led to the identification of 4 potentially deleterious coding variants in *PTGIR*, *RET*, *GALNT10*, and *UBN1*. All the 4 variants were present in all TC cases, but not in the healthy individual or the family member by marriage, who also was diagnosed with TC, although at an older age than the other TC-affected family members. Thus, a polygenic inheritance model might explain the TC clustering in the family. The variants were rare (MAF < 0.1% in the gnomAD genome data), and we did not find any variants in these genes fulfilling our prioritization pipeline criteria in 4 other families from our previous studies.²⁰ Unfortunately, we did not have access to other, sporadic, cohorts of TC, or benign thyroid diseases. However, the moderate to strong results of the *in silico* analyses were supported by the available literature that implicated the involvement of the corresponding proteins in cancer-related processes: *PTGIR* and *RET* as mediators of the MAPK/ERK and PI3K/AKT pathways, *UBN1* as a regulator of senescence and gene expression, and *GALNT10* as an initiator of posttranslational mucin-type-O-glycosylation. While *RET* is a known proto-oncogene in MTC and is also frequently mutated in Hirschsprung disease, or megacolon congenitus,^{37,39} we report here the first germline variant in predisposition to NMTC. Similarly, to our knowledge, no other germline variants in *PTGIR*, *GALNT10*, and *UBN1* have been implicated in cancer to date.

In our previous pathway analysis study, G-protein coupled receptor (GPCR) driven pathways like MAPK/ERK and PI3K/

AKT were enriched in families affected by TC.²⁰ The MAPK pathway is one of the most frequently mutated signaling pathways in human cancer and plays a central role in the stimulation of biological responses, including cell proliferation, differentiation, growth, migration, and apoptosis.⁴¹ The PI3K/AKT pathway is also involved in similar biological processes and has been extensively linked to carcinogenesis in the past.⁴² Akin to GPCRs, the overexpression or aberrant activation of receptor tyrosine kinases, such as *RET*, can result in the upregulation of the MAPK/ERK and PI3K/AKT pathways. Thus, the *RET* and *PTGIR* variants found in this study could jointly play an important role in the development of TCs in the studied family.

PTGIR is a GPCR with 7 transmembrane domains, from which the last one has been shown to play a vital role in the signal transduction of GPCRs.⁴³ The variant R296H is located in the intracellular space, immediately after the seventh transmembrane domain of *PTGIR*. As a GPCR, *PTGIR* activates MAPK/ERK and PI3K/AKT signaling and regulates various important biological processes, such as metabolism, growth and homeostasis and is involved in cancer initiation and progression. Furthermore, GPCRs are important targets for many currently prescribed drugs.⁴⁴

RET is a well-known oncogene in MTC and it is also rearranged leading to its ectopic activation in PTC.³⁷ Thus, the germline and somatic *RET* variants in these diseases are well-characterized.³⁷⁻³⁹ Many of the variants are located in the intracellular tyrosine kinase region. Interestingly, some variants, especially in *MEN2A*, are located in the extracellular domains, including the cadherin-like domains, in which the

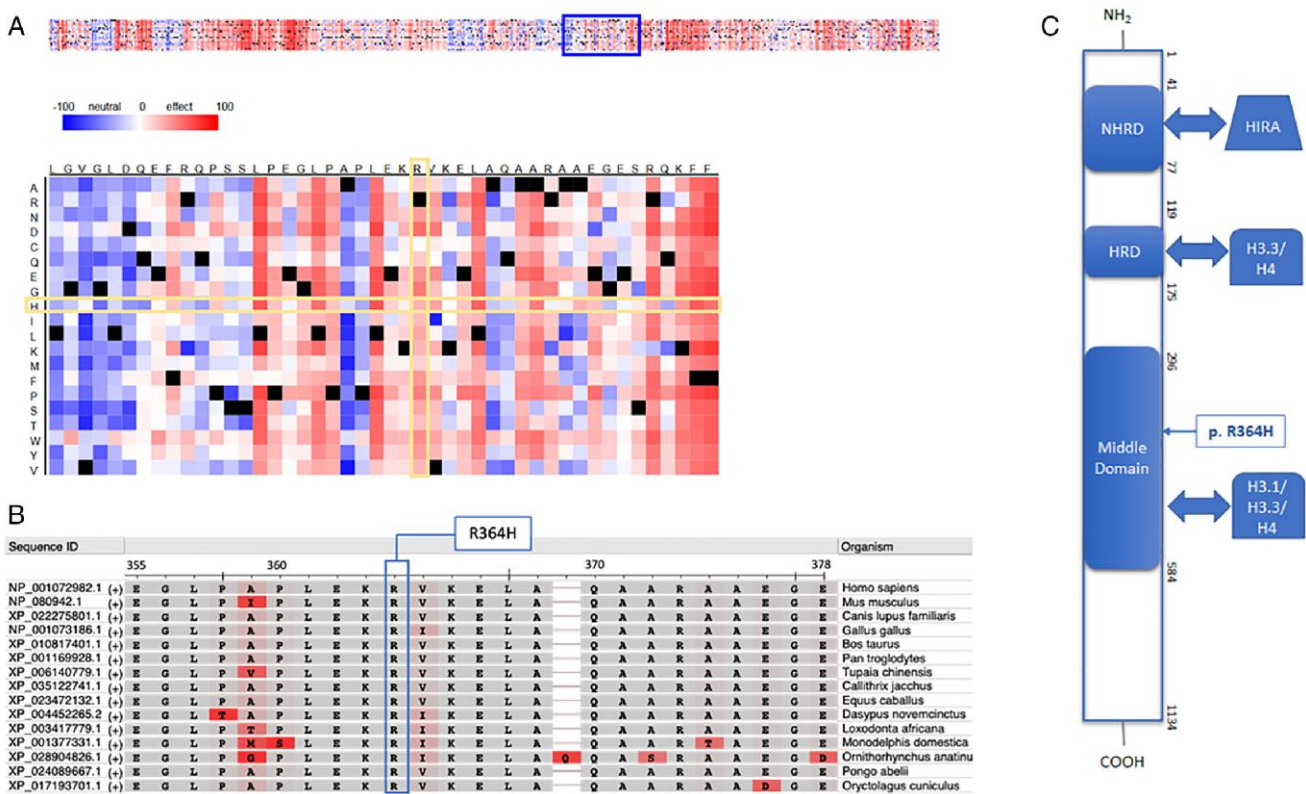


Figure 6. *In silico* studies of UBN1 p. R364H. (A) Heat map representation of SNAP2 results depicting the predicted impact of individual AA substitutions (y-axis) for each position (x-axis) on protein function. Substitutions with strong effects are depicted with dark red (score = 100), whereas weak substitutions are represented with blue (score = -100). White represents neutral substitutions. Black represents the corresponding wild-type residue. The R296H variant is shown with yellow rectangles. (B) Comparative sequence analysis across representative phylogeny. The organisms and their sequence IDs are also included. Differences are highlighted in red. The variant identified in this study is marked in blue box. (C) Schematic diagram of the UBN1 protein with all domains and their interactions, listed from the N-terminal to the C-terminal, these are N-terminal Hpc2-related domain (NHRD), HRD, and the middle domain. The variant identified in this study is marked in blue.

V262A variant from our study is located.⁴⁵ *In vitro* experiments conducted by Wang et al.³⁹ have confirmed LOF effects for nine Hirschsprung disease-associated RET mutations. The LOF effects were manifested either as disruption of RET phosphorylation or the production of a truncated protein with subcellular mis-localization or aberrant splicing.

Posttranslational modification of proteins and its dysregulation has been described as a hallmark of many cancers.⁴⁶ Aberrant mucin-type-O-glycosylation, initiated by N-acetylgalactosaminyltransferases (GALNTs), is an important posttranslational modification that is frequently associated with uncontrolled proliferation, invasion and metastasis.⁴⁷ Human GALNT10 is ubiquitously expressed, with intermediate levels in the thyroid gland.⁴⁶ The ACMG annotation of this variant identified it as a possible null variant, as it is predicted to affect splicing. LOF is a known mechanism of disease in GALNT10, hence this prediction indicates a possible pathogenicity.

Cellular senescence is a hallmark of tumor suppression and depends on control of chromatin.⁴⁸ UBN1 is a protein involved in gene regulation through its effect on chromatin assembly. It is a member of the HIRA histone chaperone complex, composed of HIRA, UBN1, and CABIN1.⁴⁹ The HIRA complex specifically binds to and deposits H3.3 at selected genomic regions, including gene regulatory elements, gene bodies, and sites of DNA damage.^{50,51} The variant identified in this study is located in the UBN1 middle domain, which has dimer formation activity and binds to H3/H4.⁵²

Another histone chaperone complex consisting of ATRX and DAXX is essential for H3.3 deposition at heterochromatic loci. This pathway has previously been shown to suppress the development of pancreatic neuroendocrine tumors and pediatric glioblastomas.⁵³ In contrast, not much has been reported on the HIRA complex's role in cancer. It was initially identified through a study of DiGeorge syndrome patients and has since been linked to abnormalities in embryogenesis.⁵⁴ HIRA's role in establishing and maintaining senescence has only more recently been discovered.⁴⁸ In that study the complex was required to suppress oncogene-induced neoplasia in a mouse model.⁴⁸ Therefore, aberrant function of UBN1, as a member of the HIRA chaperone complex, may influence cell proliferation, DNA damage repair, and gene regulation.

The prioritization of the variants in our study was based on a set of *in silico* tools. We acknowledge this as a limitation of our study. *In silico* tools can only offer predictions, particularly because different algorithms are based on similar datasets. We recognize that a conclusive classification of variants is only possible with the help of functional studies. Therefore, further functional studies will be required to determine if the variants in genes identified in our study represent pathogenic variants for FNMTTC and to test if these variants indeed act in synergy.

In conclusion, WGS of an Italian family affected by NMTC combined with a thorough literature review allowed us to prioritize 4 potentially disease-causing variants in this family.

The affected genes were reported to be involved in important cancer-related pathways and processes, such as cell proliferation, differentiation, growth, DNA repair, migration, apoptosis, senescence, and gene regulation. One of the novel variants was located in an established oncogene, *RET*, known to cause MTC, MEN2, and PTC. In the present study, we propose a polygenic mode of inheritance in this family and aim to perform functional studies to further characterize these variants in the context of cancer predisposition. We hope to contribute to the growing database of known germline variants in NMTC to improve the current understanding of NMTC inheritance.

Acknowledgments

We thank the Genomics and Proteomics Core Facility (GPCF) and the Omics IT and Data Management Core Facility (ODCF) of the German Cancer Research Center (DKFZ) for the excellent technical support.

Supplementary material

[Supplementary material](#) is available at *European Journal of Endocrinology* online.

Funding

K.H. was supported by Czech Ministry grants NU21-03-00506, NU21-03-00145 and NW24-03-00521; Czech Science Agency grant 23-05609S; the Project National Institute for Cancer Research—NICR (Programme EXCELES, ID Project no. LX22NPO5102), funded by the European Union—Next Generation EU and the SALVAGE project, reg. no.: CZ.02.01.01/00/22_008/000464.

Authors' contributions

Aayushi Srivastava (Formal analysis [equal], Investigation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [supporting]), Diamanto Skopelitou (Investigation [equal], Writing—review & editing [supporting]), Beiping Miao (Investigation [equal], Writing—review & editing [supporting]), Sara Giagiobbe (Investigation [equal], Writing—review & editing [supporting]), Nagarajan Paramasivam (Formal analysis [equal], Investigation [equal], Methodology [equal], Software [lead], Writing—review & editing [supporting]), Abhishek Kumar (Investigation [supporting], Writing—review & editing [supporting]), Chiara Diquigiovanni (Investigation [supporting], Resources [supporting], Writing—review & editing [supporting]), Elena Bonora (Investigation [supporting], Resources [supporting], Writing—review & editing [supporting]), Obul Reddy Bandapalli (Investigation [equal], Methodology [equal], Supervision [lead], Writing—original draft [equal], Writing—review & editing [supporting]), Asta Försti (Conceptualization [lead], Project administration [equal], Supervision [equal], Writing—review & editing [lead]), and Kari Hemminki (Conceptualization [lead], Funding acquisition [lead], Project administration [supporting], Resources [lead], Supervision [supporting], Writing—review & editing [supporting])

Conflict of interest: The authors declare no conflicts of interest.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

- Sung H, Ferlay J, Siegel RL, *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209-249. <https://doi.org/10.3322/caac.21660>
- Bonora E, Tallini G, Romeo G. Genetic predisposition to familial nonmedullary thyroid cancer: an update of molecular findings and state-of-the-art studies. *J Oncol.* 2010;2010:385206. <https://doi.org/10.1155/2010/385206>
- Mazeh H, Sippel RS. Familial nonmedullary thyroid carcinoma. *Thyroid.* 2013;23(9):1049-1056. <https://doi.org/10.1089/thy.2013.0079>
- Hemminki K, Sundquist K, Sundquist J, Försti A, Hemminki A, Li X. Familial risks and proportions describing population landscape of familial cancer. *Cancers (Basel).* 2021;13(17):4385. <https://doi.org/10.3390/cancers13174385>
- Fallah M, Pukkala E, Tryggvadottir L, *et al.* Risk of thyroid cancer in first-degree relatives of patients with non-medullary thyroid cancer by histology type and age at diagnosis: a joint study from five nordic countries. *J Med Genet.* 2013;50(6):373-382. <https://doi.org/10.1136/jmedgenet-2012-101412>
- Hincza K, Kowalik A, Kowalska A. Current knowledge of germline genetic risk factors for the development of non-medullary thyroid cancer. *Genes (Basel).* 2019;10(7):482. <https://doi.org/10.3390/genes10070482>
- Brigante G, Lazzaretti C, Paradiso E, *et al.* Genetic signature of differentiated thyroid carcinoma susceptibility: a machine learning approach. *Eur Thyroid J.* 2022;11(5):e220058. <https://doi.org/10.1530/etj-22-0058>
- Köhler A, Chen B, Gemignani F, *et al.* Genome-wide association study on differentiated thyroid cancer. *J Clin Endocrinol Metab.* 2013;98(10):E1674-E1681. <https://doi.org/10.1210/jc.2013-1941>
- Truong T, Lesueur F, Sugier PE, *et al.* Multiethnic genome-wide association study of differentiated thyroid cancer in the EPITHYR consortium. *Int J Cancer.* 2021;148(12):2935-2946. <https://doi.org/10.1002/ijc.33488>
- Srivastava A, Miao B, Skopelitou D, *et al.* A germline mutation in the POT1 gene is a candidate for familial non-medullary thyroid cancer. *Cancers (Basel).* 2020;12(6):1441. <https://doi.org/10.3390/cancers12061441>
- Srivastava A, Giangioffe S, Skopelitou D, *et al.* Whole genome sequencing prioritizes CHEK2, EWSR1, and TIAM1 as possible predisposition genes for familial non-medullary thyroid cancer. *Front Endocrinol (Lausanne).* 2021;12:600682. <https://doi.org/10.3389/fendo.2021.600682>
- Kumar A, Bandapalli OR, Paramasivam N, *et al.* Familial cancer variant prioritization pipeline version 2 (FCVPPv2) applied to a papillary thyroid cancer family. *Sci Rep.* 2018;8(1):11635. <https://doi.org/10.1038/s41598-018-29952-z>
- Lesueur F, Stark M, Tocco T, *et al.* Genetic heterogeneity in familial non-medullary thyroid carcinoma: exclusion of linkage to RET, MNG1, and TCO in 56 families. NMTC Consortium. *J Clin Endocrinol Metab.* 1999;84(6):2157-2162. <https://doi.org/10.1210/jcem.84.6.5798>
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Rimmer A, Phan H, Mathieson I, *et al.* Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet.* 2014;46(8):912-918. <https://doi.org/10.1038/ng.3036>
- Auton A, Abecasis GR, Altshuler DM, *et al.* A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74. <https://doi.org/10.1038/nature15393>

17. Lek M, Karczewski KJ, Minikel EV, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-291. <https://doi.org/10.1038/nature19057>
18. Smigielski EM, Sirotkin K, Ward M, Sherry ST. dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res*. 2000;28(1):352-355. <https://doi.org/10.1093/nar/28.1.352>
19. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164. <https://doi.org/10.1093/nar/gkq603>
20. Srivastava A, Kumar A, Giangioffe S, *et al.* Whole genome sequencing of familial non-medullary thyroid cancer identifies germline alterations in MAPK/ERK and PI3K/AKT signaling pathways. *Biomolecules*. 2019;9(10):605. <https://doi.org/10.3390/biom9100605>
21. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310-315. <https://doi.org/10.1038/ng.2892>
22. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Hum Mutat*. 2016;37(3):235-241. <https://doi.org/10.1002/humu.22932>
23. Cooper GM, Stone EA, Asimenos G, *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res*. 2005;15(7):901-913. <https://doi.org/10.1101/gr.3577405>
24. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*. 2010;20(1):110-121. <https://doi.org/10.1101/gr.097857.109>
25. Siepel A, Bejerano G, Pedersen JS, *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005;15(8):1034-1050. <https://doi.org/10.1101/gr.3715005>
26. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet*. 2013;9(8):e1003709. <https://doi.org/10.1371/journal.pgen.1003709>
27. Kopanos C, Tsiolkas V, Kouris A, *et al.* VarSome: the human genomic variant search engine. *Bioinformatics*. 2019;35(11):1978-1980. <https://doi.org/10.1093/bioinformatics/bty897>
28. Pagel KA, Kim R, Moad K, *et al.* Integrated informatics analysis of cancer-related variants. *JCO Clin Cancer Inform*. 2020;4:310-317. <https://doi.org/10.1200/cci.19.00132>
29. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics (Oxford, England)*. 2015;31(5):761-763. <https://doi.org/10.1093/bioinformatics/btu703>
30. Ioannidis NM, Rothstein JH, Pejaver V, *et al.* REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet*. 2016;99(4):877-885. <https://doi.org/10.1016/j.ajhg.2016.08.016>
31. Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence variants. *BMC Genomics*. 2015;16(S8):S1. <https://doi.org/10.1186/1471-2164-16-S8-S1>
32. Papadopoulos JS, Agarwala R. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics*. 2007;23(9):1073-1079. <https://doi.org/10.1093/bioinformatics/btm076>
33. Pires DEV, Ascher DB, Blundell TL. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics (Oxford, England)*. 2014;30(3):335-342. <https://doi.org/10.1093/bioinformatics/btt691>
34. Velankar S, Alhroub Y, Alili A, *et al.* PDBE: Protein Data Bank in Europe. *Nucleic Acids Res*. 2011;39(Database):D402-D410. <https://doi.org/10.1093/nar/gkq985>
35. Robinson JT, Thorvaldsdottir H, Wenger AM, Zehir A, Mesirov JP. Variant review with the integrative genomics viewer. *Cancer Res*. 2017;77(21):e31-e34. <https://doi.org/10.1158/0008-5472.CAN-17-0337>
36. Nojima S, Fujita Y, Kimura KT, *et al.* Cryo-EM structure of the prostaglandin E receptor EP4 coupled to G protein. *Structure*. 2021;29(3):252-260.e6. <https://doi.org/10.1016/j.str.2020.11.007>
37. Romei C, Ciampi R, Elisei R. A comprehensive overview of the role of the RET proto-oncogene in thyroid carcinoma. *Nat Rev Endocrinol*. 2016;12(4):192-202. <https://doi.org/10.1038/nrendo.2016.11>
38. Margraf RL, Alexander RZ, Fulmer ML, Miller CE, Coupal E, Mao R. Multiple endocrine neoplasia type 2 (MEN2) and RET specific modifications of the ACMG/AMP variant classification guidelines and impact on the MEN2 RET database. *Hum Mutat*. 2022;43(12):1780-1794. <https://doi.org/10.1002/humu.24486>
39. Wang H, Li Q, Zhang Z, Xiao P, Li L, Jiang Q. Functional studies on novel RET mutations and their implications for genetic counseling for Hirschsprung disease. *Front Genet*. 2019;10:924. <https://doi.org/10.3389/fgene.2019.00924>
40. Consortium TU. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2020;49(D1):D480-D489. <https://doi.org/10.1093/nar/gkaa1100>
41. Morrison DK. MAP kinase pathways. *Cold Spring Harb Perspect Biol*. 2012;4(11):a011254. <https://doi.org/10.1101/cshperspect.a011254>
42. Hemmings BA, Restuccia DF. PI3K-PKB/Akt pathway. *Cold Spring Harb Perspect Biol*. 2012;4(9):a011189. <https://doi.org/10.1101/cshperspect.a011189>
43. Konvicka K, Guarnieri F, Ballesteros JA, Weinstein H. A proposed structure for transmembrane segment 7 of G protein-coupled receptors incorporating an asn-Pro/Asp-Pro motif. *Biophys J*. 1998;75(2):601-611. [https://doi.org/10.1016/s0006-3495\(98\)77551-4](https://doi.org/10.1016/s0006-3495(98)77551-4)
44. Liu Y, An S, Ward R, *et al.* G protein-coupled receptors as promising cancer targets. *Cancer Lett*. 2016;376(2):226-239. <https://doi.org/10.1016/j.canlet.2016.03.031>
45. Romeo G, Ceccherini I, Celli J, *et al.* Association of multiple endocrine neoplasia type 2 and Hirschsprung disease. *J Intern Med*. 1998;243(6):515-520. <https://doi.org/10.1046/j.1365-2796.1998.00332.x>
46. Hussain MRM, Hoessli DC, Fang M. N-acetylgalactosaminyltransferases in cancer. *Oncotarget*. 2016;7(33):54067-54081. <https://doi.org/10.18632/oncotarget.10042>
47. Hollingsworth MA, Swanson BJ. Mucins in cancer: protection and control of the cell surface. *Nat Rev Cancer*. 2004;4(1):45-60. <https://doi.org/10.1038/nrc1251>
48. Rai TS, Cole JJ, Nelson DM, *et al.* HIRA orchestrates a dynamic chromatin landscape in senescence and is required for suppression of neoplasia. *Genes Dev*. 2014;28(24):2712-2725. <https://doi.org/10.1101/gad.247528.114>
49. Daniel Ricketts M, Frederick B, Hoff H, *et al.* Ubinuclein-1 confers histone H3.3-specific-binding by the HIRA histone chaperone complex. *Nat Commun*. 2015;6(1):7711. <https://doi.org/10.1038/ncomms8711>
50. Goldberg AD, Banaszynski LA, Noh KM, *et al.* Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell*. 2010;140(5):678-691. <https://doi.org/10.1016/j.cell.2010.01.003>
51. Pchelintsev NA, McBryan T, Rai TS, *et al.* Placing the HIRA histone chaperone complex in the chromatin landscape. *Cell Rep*. 2013;3(4):1012-1019. <https://doi.org/10.1016/j.celrep.2013.03.026>
52. Ricketts MD, Dasgupta N, Fan J, *et al.* The HIRA histone chaperone complex subunit UBN1 harbors H3/H4- and DNA-binding activity. *J Biol Chem*. 2019;294(23):9239-9259. <https://doi.org/10.1074/jbc.RA119.007480>
53. Schwartztruber J, Korshunov A, Liu X-Y, *et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*. 2012;482(7384):226-231. <https://doi.org/10.1038/nature10833>
54. Roberts C, Daw SC, Halford S, Scambler PJ. Cloning and developmental expression analysis of chick Hira (Chira), a candidate gene for DiGeorge syndrome. *Hum Mol Genet*. 1997;6(2):237-245. <https://doi.org/10.1093/hmg/6.2.237>