Non-medullary Thyroid Cancer Susceptibility Genes: Evidence and Disease Spectrum

Cite this article as: Jingan Zhou<Degrees>MD</Degrees>, Preeti Singh<Degrees>MD</Degrees>, Kanhua Yin<Degrees>MD, MPH</Degrees>, Jin Wang<Degrees>MD</Degrees>, Yujia Bao<Degrees>MS</Degrees>, Menghua Wu<Degrees>BS</Degrees>, Kush Pathak<Degrees>MD</Degrees>, Sophia K. McKinley<Degrees>MD, EdM</Degrees>, Danielle Braun<Degrees>PhD</Degrees>, Carrie C. Lubitz<Degrees>MD, MPH</Degrees>, Kevin S. Hughes<Degrees>MD, FACS</Degrees>, Non-medullary Thyroid Cancer Susceptibility Genes: Evidence and Disease Spectrum, Annals of Surgical Oncology <https://doi.org/10.1245/s10434-021-09745-x>

This Author Accepted Manuscript is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication but has not been copyedited or corrected. The official version of record that is published in the journal is kept up to date and so may therefore differ from this version.

Terms of use and reuse: academic research for non-commercial purposes, see here for full terms. <https://www.springer.com/aam-terms-v1>

Non-medullary Thyroid Cancer Susceptibility Genes: Evidence and Disease Spectrum

Running title: Thyroid cancer susceptibility genes

Jingan Zhou, MD^{1,2}; Preeti Singh, MD²; Kanhua Yin, MD, MPH^{2,3}; Jin Wang, MD^{2,4};
Yujia Bao, MS⁵; Menghua Wu, BS⁵; Kush Pathak, MD⁶; Sophia K. McKinley, MD, EdM⁷;
Danielle Braun, PhD^{3,8}; Carrie C. Lubitz, MD, MPH^{2,9}; and Kevin S. Hughes, MD²

¹Department of General Surgery, Beijing Anzhen Hospital, Capital Medical University, Beijing, China

²Division of Surgical Oncology, Massachusetts General Hospital, Boston, MA, USA

³Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA, USA

⁴Department of Breast Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center of Cancer Medicine, Guangzhou, China

⁵Computer Science & Artificial Intelligence, Massachusetts Institute of Technology, Boston, MA, USA

⁶Department of Surgical Oncology, P. D Hinduja Hospital, Mumbai, India

⁷Department of Surgery, Massachusetts General Hospital, Boston, MA, USA

⁸Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

⁹Institute for Technology Assessment, Massachusetts General Hospital, Boston, MA, USA

Correspondence

Kevin S. Hughes, MD, FACS, Division of Surgical Oncology, Massachusetts General Hospital, 55 Fruit Street, Yawkey 7, Boston, MA 02114.

Email: kshughes@partners.org

Conflict of interest

Kevin S. Hughes receives Honoraria from Hologic (Surgical implant for radiation planning with breast conservation and wire-free breast biopsy) and Myriad Genetics and has a financial interest in CRA Health (Formerly Hughes RiskApps). CRA Health develops risk assessment models/software with a particular focus on breast cancer and colorectal cancer. Dr. Hughes is a founder and owns equity in the company. Dr. Hughes is the Co-Creator of Ask2Me.Org which is freely available for clinical use and is licensed for commercial use by the Dana Farber Cancer Institute and the MGH. Dr. Hughes's interests in CRA Health and Ask2Me.Org were reviewed and are managed by Massachusetts General Hospital and Partners Health Care in accordance with their conflict of interest policies. Dr. Braun co-leads the BayesMendel laboratory, which licenses software for the computation of risk prediction models. She does not derive any personal income from these licenses. All revenues are assigned to the lab for software maintenance and upgrades. The other authors declare that they have no conflict of interest.

Abstract

Background: The prevalence of non-medullary thyroid cancer (NMTC) is increasing worldwide. Although most NMTCs grow slowly, conventional therapies are less effective in advanced tumors. Approximately 5-15% of NMTC have a significant germline genetic component. Awareness of the NMTC susceptibility genes may lead to earlier diagnosis and better cancer prevention. The aim of this study is to provide the current panorama of susceptibility genes associated with NMTC and the spectrum of diseases associated with these genes.

Methods: Twenty-five candidate genes were identified by searching for relevant studies in PubMed. Each candidate gene was carefully checked using six authoritative genetic resources: ClinGen, National Comprehensive Cancer Network guidelines, Online Mendelian Inheritance in Man, Genetics Home Reference, GeneCards, and Gene-NCBI, and a validated natural language processing (NLP)-based literature review protocol was used to further assess gene-disease associations where there was ambiguity.

Results: Among 25 candidate genes, ten (*APC*, *DICER1*, *FOXE1*, *HABP2*, *NKX2-1*, *PRKAR1A*, *PTEN*, *SDHB*, *SDHD*, and *SRGAP1*) were verified among the six genetic resources. Two additional genes, *CHEK2* and *SEC23B*, were verified using the NLP protocol. Seventy-nine diseases were found to be associated with these 12 NMTC susceptibility genes. The following diseases were associated with more than one NMTC susceptibility gene: colorectal cancer, breast cancer, gastric cancer, kidney cancer, gastrointestinal stromal tumor, paraganglioma, pheochromocytoma, and benign skin conditions.

Conclusion: Twelve genes predisposing to NMTC and their associated disease spectrums were identified and verified. Clinicians should be aware that patients with certain pathogenic variants may require more aggressive surveillance beyond their thyroid cancer risk.

Keywords: Non-medullary thyroid cancer; Cancer susceptibility gene; Germline pathogenic variant; Natural language processing; Genetic testing.

Synopsis

Using a semi-structured approach combined with natural language processing, we identified 12 non-medullary thyroid cancer susceptibility genes and 79 associated diseases.

Author accepted manuscript

Introduction

Currently, the incidence rate of thyroid cancer in the United States is 15.7 per 100,000 people per year. In 2020, an estimated 52,890 new cases of thyroid cancer will be diagnosed in the U.S., and 2,180 patients will die from this disease [1]. Globally, there were 567,233 new cases of thyroid cancer and an estimated 41,071 deaths per year [2]. Genetic predisposition is one factor responsible for the development of thyroid cancer. Compared to sporadic thyroid cancer, the hereditary form is not only more likely to recur but is more likely to be aggressive and present at a younger age [3-7]. There are two types of thyroid cancer: medullary thyroid cancer, arising from parafollicular cells, and non-medullary thyroid cancer (NMTC), arising from follicular cells. NMTC accounts for 95% to 97% of all thyroid cancer cases [8]. Hereditary NMTC, a subtype with a germline genetic component, accounts for approximately 5-15% of NMTCs and makes up 3-9% of all thyroid cancers [3, 9-11]. Hereditary NMTC susceptibility genes can lead to a spectrum of diseases beyond thyroid cancer as has been observed in pathogenic variants responsible for cancer syndromes, including *APC* (familial adenomatous polyposis, FAP) [12, 13], *PTEN* (Cowden syndrome) [14], and *PRKAR1A* (Carney complex) [15,16].

Patients at high risk of cancer can benefit from close surveillance and timely surgical intervention. Genetic testing helps identify at-risk individuals, making preventive treatment possible for appropriate populations while avoiding overtreatment in others. Yet, even though genetic testing has become a routine part of clinical care, the NMTC susceptibility genes are still under-recognized [17], and germline testing is not routinely clinically available [18].

When pathogenic variants are found, clinicians who are unaware of the full spectrum of cancers associated with a particular gene may turn to genetic resources for information, such as ClinGen, National Comprehensive Cancer Network (NCCN) guidelines, Online Mendelian Inheritance in Man (OMIM), Genetics Home Reference, GeneCards, and Gene-NCBI. Unfortunately, many of these resources are incomplete, ambiguous, or cumbersome to navigate. This study aims to provide an easy to use, accessible resource to improve clinician awareness of hereditary NMTC susceptibility genes and the disease spectrum associated with these genes.

Methods

Identifying NMTC susceptibility genes

The NMTC susceptibility genes were curated using the approach detailed below. We began with an initial search of the English-language literature in the PubMed database using the following query: (familial [TIAB] OR hereditary [TIAB]) AND (non-medullary thyroid cancer). Following the literature review, abstracts relevant to germline mutations were included, followed by reviewing the corresponding full-texts and identifying the candidate genes. Abstracts/papers focused on somatic mutation or polymorphism were excluded. Twenty-five candidate genes were eventually curated. The final list of NMTC susceptibility genes was identified and validated from these initial candidate genes using a two-step approach, described in detail in our previous paper by Wang et al. [19] and are summarized in Figure 1. The first step involved checking six commonly used genetic resources. These resources included ClinGen, NCCN guidelines, OMIM, Genetics Home Reference, GeneCards, and Gene-NCBI. ClinGen and NCCN guidelines were deemed the most authoritative resources as they were curated by expert panels. As a second step, uncertain associations were validated using our natural language processing (NLP)-based literature review protocol [20, 21].

Identification and evaluation of the gene-disease associations

For each of the identified NMTC susceptibility genes, we used a similar two-step approach to identify other gene-disease associations besides thyroid cancer. Once again, the first step involved the six genetic resources, followed by a second step for uncertain gene-disease associations using our NLP-based literature review protocol.

The date of last access for all resources was January 8, 2020.

Each gene-disease association received an overall classification, as described previously and outlined in Figure 1 [19]. Two individual researchers independently coded each gene-disease association using this process, and their results were brought to our group meeting for consensus review and verification. A final classification code of '1' was assigned under two circumstances: (i) if the gene-disease association was considered accurate in either ClinGen or NCCN (for the current study, there was no conflict found between ClinGen and NCCN); or (ii) if at least three out of four other genetic resources (i.e., OMIM, Genetics Home Reference, GeneCards, and Gene-NCBI) considered the gene-disease association correct. Gene-disease associations fitting neither of these circumstances were classified as 'uncertain.' Finally, any gene-disease association that was coded as not associated in ClinGen was classified as 'no association.'

All 'uncertain' gene-disease associations underwent further review using our NLP abstract classification algorithm [20, 21]. This algorithm extracts all abstracts from PubMed based on a specific PubMed search query and classifies them as penetrance and/or incidence abstracts or neither. All queries we used are listed in Supplement 1. The sensitivity (99%) of this procedure in identifying cancer penetrance studies has been validated [20, 21]. Two investigators independently reviewed the NLP classification algorithm results for all papers classified as penetrance and/or prevalence. The full text of penetrance studies was selected and then analyzed to validate the gene-disease association based on study design, sample size, carrier numbers, ascertainment criteria, and statistical significance of the results.

The gene-disease association in question was changed from 'uncertain' to 'no association' if relevant penetrance studies could not validate the association. If relevant penetrance studies were identified, all of them were presented at the group meeting attended by the principal investigator, surgeons, and research fellows (MDs). The participants selected high quality penetrance studies based on the design of the study, patient population, number of pathogenic variant carriers, and ascertainment mechanism and reached a final consensus on the assessment of these high-quality studies. If a statistically significant association was found in at least one high-quality penetrance study with at least a two-fold increase in the disease risk, the gene-disease association was upgraded from 'uncertain' to 'confirmed.' The association remained 'uncertain' if the literature failed to demonstrate a statistically significant increase in risk for those gene-disease associations. The group meetings reviewed all gene-disease associations and looked at the evidence for each gene-disease association reported in the study to ensure accuracy.

Results

NMTC susceptibility genes identified in six genetic resources and NLP-based literature review

We identified 12 NMTC susceptibility genes from the 25 candidate genes (Table 1, Supplement 2). Among these 12 genes, the association of NMTC with *APC* and *PTEN* was identified in ClinGen, NCCN, and OMIM. *DICER1*, *SDHB*, and *SDHD* were identified in ClinGen and Genetics Home Reference, while OMIM also identified *DICER1*. *PRKAR1A* was identified in ClinGen and OMIM. *FOXE1*, *HABP2*, and *SRGAP1* were identified in all the genetic resources except for ClinGen and NCCN, whereas *NKX2-1* was identified in only three of the six genetic resources OMIM, GeneCards, and Gene-NCBI. The association of NMTC with *SEC23B* was identified in only OMIM and GeneCards, while *CHEK2* was not identified in any of the six genetic resources. The NLP literature review protocol verified the associations between NMTC and both *SEC23B* and *CHEK2*.

Disease spectrum associated with NMTC susceptibility genes

A total of 79 gene-disease associations were verified for the 12 NMTC susceptibility genes (Table 2, Table 3, and Supplement 3). In addition to NMTC, all genes except for *FOXE1*, *SEC23B*, and *SRGAP1* showed an increased risk of other diseases as well. Diseases associated with individual genes varied and are listed in Table 3.

Notably, the most common disease associations were colorectal cancer and kidney cancer. These two cancers were associated with three NMTC susceptibility

genes each. Both cancers were associated with *CHEK2* and *PTEN*, while *APC* was associated with colorectal cancer but not kidney cancer, and *SDHB* was associated with kidney cancer but not colorectal cancer. The NMTC susceptibility genes with the highest number of disease associations were *APC* and *PTEN*. These genes are also associated with distinctive syndromes: *APC* with FAP and *PTEN* with Cowden syndrome. Other NMTC susceptibility genes associated with syndromes are *DICER1* with *DICER1* Syndrome and *PRKAR1A* with Carney complex (Table 2 and 3).

Gene-disease spectrum and corresponding resources

Among the 79 gene-disease associations ascertained by our methodology, approximately 71% (56 out of 79) were found in either ClinGen or NCCN or both. Twelve gene-disease associations were not found in either ClinGen or NCCN but were found in at least three out of the other four remaining genetic resources. Eleven gene-disease associations were identified in fewer than three of the six genetic resources. Among these, six gene-disease associations were not mentioned in any of the six genetic resources. For instance, *PRKAR1A* and large cell calcifying Sertoli cell tumors (LCCSCT), breast myxoma, pancreatic neoplasm, and *CHEK2*'s association with gastric and kidney cancer were negative in all six genetic resources and could be verified only by literature review using NLP. Supplement 4 summarizes the gene-disease associations that were verified by NLP.

Discussion

We curated multiple genetic resources to produce a single clinical resource for clinicians treating or counseling patients with NMTCs, many of whom harbor hidden vulnerabilities to other malignant and nonmalignant diseases. Using a novel quantitative approach combined with NLP, we found 12 NMTC susceptibility genes and 79 associated diseases. Our resource allows practitioners to review both the disease spectrum for each gene and the data in these six genetic resources and literature which support these gene-disease associations. Our overall intention is to provide an accessible, reliable, and easy-to-use clinical resource to help clinicians be more aware of the NMTC susceptibility genes and their spectrum. This work may help clinicians provide appropriate germline testing to patients with strong family history of other cancers besides NMTC (e.g., if a family has both NMTC and colorectal cancer, genetic testing might be indicated) or clinical characteristics of the related syndromes (e.g., Cowden syndrome). This resource will also help clinicians be more aware of other diseases in the gene spectrums that their patient or their family might be at risk for.

The NMTC susceptibility genes remain unfamiliar to most practicing clinicians. Only two NMTC susceptibility genes, *APC* and *PTEN*, have thyroid-specific clinical management recommendations in the current NCCN guidelines [22, 23]. Recent studies have revealed a far more complex genetic background for NMTC than we initially thought, indicating that even more genes may be associated with this cancer. For example, studies have shown direct evidence that *DICER1* pathogenic variants are linked with NMTC [24, 25]. A family-based cohort study found a 16-fold increased risk of NMTC in *DICER1* pathogenic variant carriers (95% confidence interval [CI]: 4.3-41; $p <$

0.05) compared with the general population [26]. Germline alterations in *SDHB* and *SDHD* genes have also been linked with NMTC [27-29], and recent studies have revealed some newly identified genes, including *FOXE1* [30], *HABP2* [31, 32], and *SRGAP1* [33], to be associated with increased risk of NMTC.

Although the 25 candidate genes initially selected for this study have been frequently reported in the literature [17], there remain significant inconsistencies among these reports. For example, one study described *HRAS* and *MUTYH* as the most commonly mutated somatic genes in the thyroid [34]; yet review articles fail to mention the association in the germline [35, 36]. Conflicting conclusions have also appeared regarding the association between *NKX2-1* and NMTC [37, 38]. A more comprehensive, systematic approach was therefore needed to examine the gene-cancer associations implicated in NMTC.

To ensure a more focused evaluation and assessment, we designed an integrated approach that relies exclusively on germline pathogenic variants and autosomal dominant inheritance. Two authoritative genetic resources (ClinGen and NCCN guidelines) combined with four other reputable but less highly curated genetic resources (OMIM, Genetics Home Reference, GeneCards, and Gene-NCBI) were selected to serve as the foundation for evaluating gene-disease associations. These represent the most common, publicly available genetic information resources utilized by practicing physicians and genetic counselors. Additionally, we adopted an NLP protocol as an auxiliary approach to ensure that all 'uncertain associations' were subject to rigorous review based on a contemporaneous analysis of the published literature [20, 21, 39].

Twelve specific NMTC susceptibility genes were verified among the 25 candidate genes. Six known susceptibility genes (*APC*, *PTEN*, *DICER1*, *PRKAR1A*, *SDHB*, and *SDHD*) were identified in the two most authoritative resources (ClinGen and NCCN). Four other genes (*FOXE1*, *HABP2*, *NKX2-1*, and *SRGAP1*) were identified among the four genetic resources. Although the association between NMTC and *SEC23B* was identified in only two of the six genetic resources (OMIM and GeneCards), the finding was verified by our NLP protocol, which retrieved literature showing clear evidence of the association. Specifically, Yehia et al. reported a significantly increased age-adjusted standardized incidence rate (SIR) of thyroid cancer in *SEC23B* pathogenic variant carriers (SIR = 242.6; 95% CI: 150.4-371.8, $p < 0.001$) when compared with the general population [40]. The association between *CHEK2* and NMTC was not found in any of the six resources but was verified by our NLP protocol. In 2004, Cybulski et al. reported that *CHEK2*-I157T was associated with an increased risk of thyroid carcinoma (OR = 1.9, $p = 0.04$) [41]. In 2015, Siolek et al. compared the frequencies of four *CHEK2* variants, 1100delC, IVS211G>A, del5395, and I157T, between unselected papillary thyroid carcinoma patients and their age- and sex-matched cancer-free controls. They showed that the *CHEK2* truncating pathogenic variant carriers with IVS211G>A, 1100delC, or del5395 had a higher risk of papillary thyroid carcinoma (OR = 5.7, $p = 0.006$). In addition, the missense variant I157T was associated with a 2.8-fold increased risk ($p < 0.001$) [42]. These studies strongly suggest that *CHEK2* is also an NMTC susceptibility gene.

We notice that controversies still exist regarding the association between NMTC and these 12 genes, especially for the genes that were newly identified. For example,

the association between *HAPB2* and NMTC was first reported in 2015 [43]. This association was identified in four genetic resources (Genetics Home Reference, OMIM, Gene Cards, and Gene-NCBI), making *HAPB2* qualified as an NMTC susceptibility gene in our study. However, several sizeable studies reported negative associations in different populations, such as in the U.K., China, and Australia [44-47]. This conflicting evidence suggests that further studies are required to elucidate the links between *HAPB2* and NMTC.

As critical as verifying the identity of the 12 NMTC susceptibility genes is, our analysis of each gene's disease spectrum is just as important. This information allowed us to categorize these genes into two groups: those that are known syndromic versus non-syndromic (not having a recognized or named syndrome) associations. The former includes *APC*, *PTEN*, *PRKAR1A*, and *DICER1*. Individuals with syndromic pathogenic variants are at high risk of developing various types of malignancies and benign diseases. Thus, they may require additional surveillance, especially those who have disease associations other than NMTC [16, 22, 23, 25]. *PTEN* pathogenic variant carriers, for example, should undergo more active surveillance for breast and endometrial cancer [23].

The non-syndromic susceptibility genes include *CHEK2*, *HABP2*, *NKX2-1*, *SDHB*, and *SDHD*, each with a specific disease spectrum. For instance, *CHEK2* pathogenic variant carriers may have a higher risk of breast cancer, colorectal cancer, gastric cancer, prostate cancer, kidney cancer, and osteosarcoma (Table 3). In particular, individuals who carry the *SDHB* pathogenic variant have a significantly higher risk of malignant pheochromocytoma or paraganglioma [48]. However, several of the genes,

FOXE1, *SEC23B*, and *SRGAP1*, are only associated with NMTC.

Finally, NLP played an important role in evaluating the disease spectra for *CHEK2* and *PRKAR1A*, respectively. The literature retrieved through NLP verified the associations between *CHEK2* and gastric and kidney cancer [49, 50], and the related *PRKAR1A*-disease associations were supported by recent review articles [15, 16].

The ultimate goal of identifying these 12 NMTC susceptibility genes and their associated disease spectrums is to inform clinicians and prompt better cancer/disease prevention and management. One direct implication of our work is possibly adding more of these 12 genes to the commercially available germline genetic testing panels, ensuring these genes are tested and that the pathogenic variant carriers are informed about the thyroid and other cancer risks. Clinicians could also refer to the disease spectrum information we identified when evaluating patients' family history and include these genes into their differential diagnoses. Genetic testing should be offered to identify the pathogenic variant carriers for those whose family histories include thyroid cancer and other associated cancers in these spectrums. Cascade testing may also be considered to benefit patients' family members. In the most recent NCCN guidelines, thyroid-specific surveillances and management recommendations have been mentioned regarding only two (*APC* and *PTEN*) out of these 12 NMTC susceptibility gene [22, 23]. For these two genes, thyroid examination and ultrasound are suggested as possibly useful additions to management. These recommendations could potentially be expanded to the other ten genes after carefully evaluating the cost-effectiveness.

Limitations

This study is subject to certain limitations. First, the accuracy of our conclusions largely relies upon the accuracy of the six genetic resources. Thus, we used a consensus review of all six genetic resources supplemented with the NLP protocol. Two individual researchers independently coded each gene-disease association, and their individual coding results were brought to our group meeting for consensus review and verification. Second, we understand that our findings are a snapshot of the current understanding of these gene-disease associations and that our knowledge could change with additional studies. These six genetic resources are updated periodically. With genetic research increasing, future research efforts may clarify NMTC susceptibility gene and disease associations that we could not find or verify. This study tries to provide a comprehensive view of the current knowledge and understanding of NMTC susceptibility genes, rather than a definitive conclusion. Third, the data regarding the penetrance of these NMTC susceptibility genes are still sparse. Most currently available NMTC risk for these genes are estimated from pedigree analysis, and large-scale population-based studies are still uncommon. We are well aware of these limitations and are actively trying to keep both the gene-cancer association data and penetrance data up to date on our Ask2Me website (<https://ask2me.org/index.php>) [51].

Conclusion

Twelve NMTC susceptibility genes and their associated disease spectrums were identified and verified. Clinicians who treat thyroid cancer should be aware that patients with certain pathogenic variants may develop other types of cancers and require more aggressive surveillance beyond their thyroid cancer risk.

Author accepted manuscript

Acknowledgements

The authors thank Ann S. Adams (Department of Surgery, Massachusetts General Hospital) for editorial and writing assistance.

Author accepted manuscript

References

1. National Cancer Institute: Surveillance, Epidemiology, and End Results Program. Cancer Stat Facts: Thyroid Cancer [Available from <https://seer.cancer.gov/statfacts/html/thyro.html>] Accessed November 20, 2020
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. *C.A. Cancer J Clin.* 2018; 68(6), 394-424.
3. Moses W, Weng J, Kebebew E. Prevalence, clinicopathologic features, and somatic genetic mutation profile in familial versus sporadic non-medullary thyroid cancer. *Thyroid.* 2011; 21(4):367-71.
4. Bano G, Hodgson S. Diagnosis and management of hereditary thyroid cancer. *Recent Results Cancer Res.* 2016; 205:29-44.
5. Rowland KJ, Moley JF. Hereditary thyroid cancer syndromes and genetic testing. *J Surg Oncol.* 2015; 111(1):51-60.
6. Navas-Carrillo D, Ríos A, Rodríguez JM, Parrilla P, Orenes-Piñero E. Familial nonmedullary thyroid cancer: screening, clinical, molecular and genetic findings. *Biochim Biophys Acta.* 2014; 1846(2):468-76.
7. Guilmette J, Nosé V. Hereditary and familial thyroid tumours. *Histopathology.* 2018; 72(1):70-81.
8. Fagin JA, Wells SA Jr. Biologic and clinical perspectives on thyroid cancer. *N Engl J Med.* 2016;375(11):1054-67.
9. Alsanea O, Clark OH. Familial thyroid cancer. *Curr Opin Oncol.* 2001; 13(1), 44-51.
10. Nagy R, Ringel M. Genetic predisposition for non-medullary thyroid cancer. *Horm Cancer.* 2015; 6(1): 13-20.
11. Lamartina L, Grani G, Durante C, Filetti S, Cooper DS. Screening for differentiated thyroid cancer in selected populations. *Lancet Diabetes Endocrinol.* 2020; 8(1):81-8.
12. Waller A, Findeis S, Lee M. Familial Adenomatous Polyposis. *J Pediatr Genet.* 2016; 5(2): 78-83.
13. Dinarvand P, Davaro EP, Doan JV, Ising ME, Evans NR, Phillips NJ, et al. Familial adenomatous polyposis syndrome: An update and review of extraintestinal manifestations. *Arch Pathol Lab Med.* 2019;143(11):1382-98.
14. Mester J, Eng C. Cowden syndrome: Recognizing and managing a not - so - rare hereditary cancer syndrome. *Journal of Surgical Oncology.* 2015; 111(1):125-130.
15. Correa R, Salpea P, Stratakis CA. Carney complex: an update. *Eur J Endocrinol.* 2015; 173(4):M85-97.
16. Kamilaris C, Faucz F, Voutetakis A, Stratakis CA. Carney Complex. *Exp Clin Endocrinol Diabetes.* 2019; 127(2-03):156-64.
17. Kamihara J, Laduca H, Dalton E, Speare V, Garber JE, Black MH. Germline mutations in cancer predisposition genes among patients with thyroid cancer. *J Clin Oncol.* 2017; 35(S15),1581.
18. Patel KN, Yip L, Lubitz CC, Grubbs EG, Miller BS, Shen W, et al. The American Association of Endocrine Surgeons Guidelines for the Definitive Surgical Management of Thyroid Disease in Adults. *Ann Surg.* 2020; 271(3):e21-e93.
19. Wang J, Singh P, Yin K, Zhou J, Bao Y, Wu M, et al. Disease Spectrum of Breast Cancer Susceptibility Genes. medRxiv doi: 10.1101/2020.08.11.20172007

20. Deng Z, Yin K, Bao Y, Armengol VD, Wang C, Tiwari A, et al. Validation of a semiautomated natural language processing-based procedure for meta-analysis of cancer susceptibility gene penetrance. *JCO Clin Cancer Inform.* 2019;3:1-9.
21. Bao Y, Deng Z, Wang Y, Kim H, Armengol VD, Acevedo F, et al. Using machine learning and natural language processing to review and classify the medical literature on cancer susceptibility genes. *JCO Clin Cancer Inform.* 2019;3:1-9.
22. National Comprehensive Cancer Network. Genetic/familial high-risk assessment: Colorectal. Version 2.2019. Retrieved from https://www.nccn.org/professionals/physician_gls/pdf/genetics_colon.pdf (accessed November 10, 2019)
23. National Comprehensive Cancer Network. Genetic/familial high-risk assessment: Breast and ovarian. Version 3.2019. Retrieved from https://www.nccn.org/professionals/physician_gls/pdf/genetics_bo.pdf (accessed November 10, 2019)
24. Rutter MM, Jha P, Schultz KA, Sheil A, Harris AK, Bauer AJ, et al. DICER1 mutations and differentiated thyroid carcinoma: Evidence of a direct association. *J Clin Endocrinol Metab.* 2016; 101(1):1-5.
25. Schultz KAP, Williams GM, Kamihara J, Stewart DR, Harris AK, Bauer AJ, et al. DICER1 and associated conditions: Identification of at-risk individuals and recommended surveillance strategies. *Clin Cancer Res.* 2018; 24(10):2251-61.
26. Khan NE, Bauer AJ, Schultz KAP, Doros L, Decastro RM, Ling A, et al. Quantification of thyroid cancer and multinodular goiter risk in the DICER1 syndrome: A family-based cohort study. *J Clin Endocrinol Metab.* 2017; 102(5):1614-22.
27. Ni Y, Seballos S, Ganapathi S, Gurin D, Fletcher B, Ngeow J, et al. Germline and somatic SDHx alterations in apparently sporadic differentiated thyroid cancer. *Endocr Relat Cancer.* 2015; 22(2):121-30.
28. Ngeow J, Mester J, Rybicki L, Ni Y, Milas M, Eng C. Incidence and clinical characteristics of thyroid cancer in prospective series of individuals with Cowden and Cowden-Like syndrome characterized by germline PTEN, SDH, or KLLN Alterations. *J Clin Endocrinol Metab* 2011;96(12):E2063-71.
29. Ni Y, He X, Chen J, Moline J, Mester J, Orloff MS, et al. Germline SDHx variants modify breast and thyroid cancer risks in Cowden and Cowden-like syndrome via FAD/NAD-dependant destabilization of p53. *Hum Mol Genet.* 2012; 21(2):300-10.
30. Pereira JS, da Silva JG, Tomaz RA, Pinto AE, Bugalho MJ, Leite V, et al. Identification of a novel germline FOXE1 variant in patients with familial non-medullary thyroid carcinoma (FNMTCT). *Endocrine.* 2015; 49(1):204-14.
31. Gara SK, Jia L, Merino MJ, Agarwal SK, Zhang L, Cam M, et al. Germline HABP2 mutation causing familial non-medullary thyroid cancer. *N Engl J Med.* 2015; 373(5):448-55.
32. Zhang T, Xing M. HABP2 G534E mutation in familial non-medullary thyroid cancer. *J Natl Cancer Inst.* 2016; 108(6):djv415.
33. He H, Bronisz A, Liyanarachchi S, Nagy R, Li W, Huang Y, et al. SRGAP1 is a candidate gene for papillary thyroid carcinoma susceptibility. *J Clin Endocrinol Metab.* 2013; 98(5):E973-80.

34. Akras Z, Bungo B, Leach BH, Marquard J, Ahluwalia M, Carraway H, et al. Primer on hereditary cancer predisposition genes included within somatic next-generation sequencing panels. *JCO Precision Oncology*. 2019; 3(3), 1-11.
35. Yang SP, Ngeow J. Familial non-medullary thyroid cancer: unraveling the genetic maze. *Endocr Relat Cancer*. 2016; 23(12):R577-95.
36. Hińcza 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.
37. Ngan ES, Lang BH, Liu T, Shum CKY, So MT, Lau DKC, et al. A germline mutation (A339V) in thyroid transcription factor-1(TITF-1/NKX2.1) in patients with multinodular goiter and papillary thyroid carcinoma. *J Natl Cancer Inst*. 2009; 101(3):162-75.
38. Cantara S, Capuano S, Formichi C, Pisu M, Capezzone M, Pacini F. Lack of germline A339V mutation in thyroid transcription factor-1(TITF-1/NKX2.1) gene in familial papillary thyroid cancer. *Thyroid Res*. 2010; 3(1):4.
39. Hughes KS, Zhou J, Bao Y, Singh P, Wang J, Yin K. Natural language processing to facilitate breast cancer research and management. *Breast J*. 2020; 26:92-9.
40. Yehia L, Niazi F, Ni Y, Ngeow J, Sankunni M, Liu Z, et al. Germline heterozygous variants in SEC23B are associated with Cowden syndrome and enriched in apparently sporadic thyroid cancer. *Am J Hum Genet* 2015;97(5):661-76.
41. Cybulski C, Górski B, Huzarski T, Masojć B, Mierzejewski M, Debniak T, et al. CHEK2 is a multiorgan cancer susceptibility gene. *Am J Hum Genet*. 2004;75(6):1131-5.
42. Siólek M, Cybulski C, Gaśior - Perczak D, Kowalik A, Kozak-Klonowska B, Kowalska A, et al. CHEK2 mutations and the risk of papillary thyroid cancer. *Int J Cancer* 2015;137(3):548-52.
43. Gara SK, Jia L, Merino MJ, Agarwal SK, Zhang L, Cam M, et al. Germline HABP2 mutation causing familial nonmedullary thyroid cancer. Germline HABP2 mutation causing familial nonmedullary thyroid cancer. *N Engl J Med*. 2015;373(5):448-55.
44. Zhao X, Li X, Zhang X. HABP2 mutation and nonmedullary thyroid cancer. *N Engl J Med* 2015;373(21):2084.
45. Sahasrabudhe R, Stultz J, Williamson J, Lott P, Estrada A, Bohorquez M, et al. The HABP2 G534E variant is an unlikely cause of familial non-medullary thyroid cancer. *J Clin Endocrinol Metab*. 2016;10(3):1098-1103.
46. Alzahrani AS, Murugan AK, Qasem E, Al-Hindi H. Habp2 gene mutations do not cause familial or sporadic non-medullary thyroid cancer in a highly inbred middle eastern population. *Thyroid*. 2016;26(5):667-671.
47. Weeks AL, Wilson SG, Ward L, Goldblatt J, Hui J, Walsh JP. HABP2 germline variants are uncommon in familial nonmedullary thyroid cancer. *BMC Med Genet*. 2016;17(1):60.
48. Jochmanova I, Wolf KI, King KS, Nambuba J, Wesley R, Martucci V, et al. SDHB-related pheochromocytoma and paraganglioma penetrance and genotype-phenotype correlations. *J Cancer Res Clin Oncol*. 2017; 143(8):1421-35.
49. Teodorczyk U, Cybulski C, Wokołorczyk D, Jakubowska A, Starzyńska T, Lawniczak M, et al. The risk of gastric cancer in carriers of CHEK2 mutations. *Fam Cancer*. 2013; 12(3):473-8.

50. Näslund-Koch C, Nordestgaard BG, Bojesen SE. Increased risk for other cancers in addition to breast cancer for CHEK2*1100delC heterozygotes estimated from the Copenhagen General Population Study. *J Clin Oncol*. 2016; 34(11):1208-16.
51. Braun D, Yang J, Griffin M, Parmigiani G, Hughes KS. A clinical decision support tool to predict cancer risk for commonly tested cancer-related germline mutations. *J Genet Couns*. 2018; 27:1187-99.
52. Ricketts CJ, Forman JR, Rattenberry E, Bradshaw N, Laloo F, Izatt L, et al. Tumor risks and genotype-phenotype-proteotype analysis in 358 patients with germline mutations in SDHB and SDHD. *Hum Mutat*. 2010;31(1):41-51.

Author accepted manuscript

Figure legends

Figure 1. Workflow for classifying gene-NMTC associations. The number '1' indicates that the gene was associated with NMTC in the resource. The number '0' indicates that the gene's association with NMTC was refuted in the resource. The number '9' indicates that the gene's association with NMTC was unclear in the resource. * Uncertain association indicates that the gene's association with NMTC is unclear and it may or may not be associated with NMTC, further studies are required to refute or accept the association.

Abbreviations: NCBI, National Center for Biotechnology Information; NCCN, National Comprehensive Cancer Network; NLP, natural language processing; NMTC, Nonmedullary thyroid cancer; OMIM, Online Mendelian Inheritance in Man.

Table 1. Associations between susceptibility genes and non-medullary thyroid cancer in six genetic resources

Gene	Genetic Resources					
	ClinGen	NCCN	Genetics Home Reference	OMIM	GeneCards	Gene-NCBI
<i>APC</i>	1	1		1		
<i>PTEN</i>	1	1		1		
<i>DICER1</i>	1		1	1		
<i>PRKAR1A</i>	1			1		
<i>SDHB</i>	1		1			
<i>SDHD</i>	1		1			
<i>FOXE1</i>			1	1	1	1
<i>HABP2</i>			1	1	1	1
<i>NKX2-1</i>				1	1	1
<i>SRGAP1</i>			1	1	1	1
<i>CHEK2</i>						
<i>SEC23B</i>				1	1	

Note: The number '1' indicates that the gene was associated with NMTC in the resource.

Abbreviations: NCBI, National Center for Biotechnology Information; NCCN, National Comprehensive Cancer Network; OMIM, Online Mendelian Inheritance in Man.

Table 2. Disease spectrum associated with NMTC susceptibility genes

Disease	NMTCSGs												
	Number												
	of genes	APC	CHEK2	DICER1	FOXE1	HABP2	NKX2-1	PRKAR1A	PTEN	SDHB	SDHD	SEC23B	SRGAP1
Non-medullary Thyroid Cancer	12	+	+	+	+	+	+	+	+	+	+	+	+
Colorectal Cancer	3	+	+						+				
Kidney cancer	3		+						+	+			
Breast cancer	2		+						+				
Gastric Cancer	2	+	+										
GIST	2									+	+		
Paraganglioma	2									+	+		
Pheochromocytoma	2									+	+		
Skin(Benign)	2							+	+				
Prostate Cancer	1		+										
Pancreatic Cancer	1	+											
Duodenal cancer	1	+											
Hepatoblastoma	1	+											
Medulloblastoma	1	+											
Small Intestine Cancer	1	+											
Adrenal Adenoma	1	+											
CHRPE	1	+											
Colorectal Polyps	1	+											
Desmoid	1	+											
Duodenal Polyps	1	+											
Epidermoid Cysts	1	+											
Gastric Polyps	1	+											
Odontoma	1	+											
Osteoma	1	+											
PMAH	1	+											
Osteosarcoma	1		+										
Thyroid Adenoma	1			+									
Pleuropulmonary blastoma	1			+									

[illegible]

Note: The '+' sign indicates that the gene was associated with the disease in the resources by our method.

Note: The + sign indicates that the gene was associated with the disease in the resources by our method.

Abbreviations: GIST, gastrointestinal stromal tumor; CHRPE, congenital hypertrophy of the retinal pigment epithelium; MNG, multinodular goiter; PPNAD, primary pigmented nodular adrenocortical disease; LCCSCT, large cell calcifying Sertoli cell tumors; PMAH, primary macronodular adrenal hyperplasia; PMS, psammomatous melanotic schwannoma; NMTCSGs, non-medullary thyroid cancer susceptibility genes

Table 3. NMTC susceptibility genes associated disease spectrum (malignant vs. benign vs. borderline)

Gene	Malignant	Benign	Borderline
APC	Colorectal, duodenal, gastric, hepatoblastoma, medulloblastoma, NMTC, pancreatic, small intestine	Adrenal adenoma, CHRPE, colorectal polyps, desmoid, duodenal polyps, epidermoid cysts, gastric polyps, odontoma, osteoma, PMAH	
CHEK2	Breast, colorectal, gastric, kidney, NMTC, osteosarcoma, prostate		
DICER1	Nephroblastoma, NMTC, pleuropulmonary blastoma	Cystic nephroma, multinodular goiter, thyroid adenoma	Sex cord tumor
FOXE1	NMTC		
HABP2	NMTC	Venous thrombosis	
NKX2-1	NMTC	Chorea, hypothyroid, neonatal respiratory distress	
PRKAR1A	NMTC	Breast myxoma, cardiac myxoma, cutaneous myxoma, eyelid myxoma, osteochondromyxoma, pituitary adenoma, PPNAD, skin(benign)	LCCSCT, pancreatic neoplasm, PMS
PTEN	Brain tumor, breast, colorectal, endometrial, kidney, melanoma, NMTC,	Autism, cerebrovascular malformations, facial papules, gastrointestinal hamartomatous polyps, lipoma, macrocephaly, oral mucosal papillomatosis, skin(benign), thyroid (benign), uterine fibroids	
SDHB	NMTC, Pheochromocytoma*, Paraganglioma, RCC		GIST
SDHD	NMTC, paraganglioma	Pheochromocytoma**	GIST
SEC23B	NMTC		
SRGAP1	NMTC		

Abbreviations: CHRPE, congenital hypertrophy of the retinal pigment epithelium; GIST, gastrointestinal stromal tumor; LCCSCT, large cell calcifying Sertoli cell tumors; NMTC, non-medullary thyroid cancer; PMAH, primary macronodular adrenal hyperplasia; PMS, psammomatous melanin schwannomas; PPNAD, primary pigmented nodular adrenocortical disease; RCC, renal cell carcinoma.

* Penetrance is not complete

** Risk is low (95% benign) [52]