



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS MÉDICAS: ENDOCRINOLOGIA

Vicente Rodrigues Marczyk

**UTILIZAÇÃO DE FERRAMENTAS DE BIOINFORMÁTICA E DE APRENDIZADO
DE MÁQUINA NA CLASSIFICAÇÃO DAS NEOPLASIAS DE TIREOIDE**

Porto Alegre

2023

Vicente Rodrigues Marczyk

**UTILIZAÇÃO DE FERRAMENTAS DE BIOINFORMÁTICA E DE
APRENDIZADO DE MÁQUINA NA CLASSIFICAÇÃO DAS NEOPLASIAS DE
TIROIDE**

Tese de Doutorado apresentada ao Programa
de Pós-Graduação em Ciências Médicas:
Endocrinologia da Faculdade de Medicina
da Universidade Federal do Rio Grande do
Sul como requisito parcial para obtenção do
título de Doutor em Endocrinologia

Orientadora: Profa. Dra. Ana Luiza Maia

Porto Alegre

2023

CIP - Catalogação na Publicação

Marczyk, Vicente Rodrigues
UTILIZAÇÃO DE FERRAMENTAS DE BIOINFORMÁTICA E DE
APRENDIZADO DE MÁQUINA NA CLASSIFICAÇÃO DAS NEOPLASIAS
DE TIREOIDE / Vicente Rodrigues Marczyk. -- 2023.
65 f.
Orientadora: Ana Luiza Maia.

Tese (Doutorado) -- Universidade Federal do Rio
Grande do Sul, Faculdade de Medicina, Programa de
Pós-Graduação em Ciências Médicas: Endocrinologia,
Porto Alegre, BR-RS, 2023.

1. Câncer de tireoide. 2. Neoplasias de tireoide.
3. Oncologia. 4. Biologia computacional. 5.
Aprendizado de máquina. I. Maia, Ana Luiza, orient.
II. Título.

Vicente Rodrigues Marczyk

**UTILIZAÇÃO DE FERRAMENTAS DE BIOINFORMÁTICA E DE
APRENDIZADO DE MÁQUINA NA CLASSIFICAÇÃO DAS NEOPLASIAS DE
TIREOIDE**

Tese de Doutorado apresentada ao Programa
de Pós-Graduação em Ciências Médicas:
Endocrinologia da Faculdade de Medicina
da Universidade Federal do Rio Grande do
Sul como requisito parcial para obtenção do
título de Doutor em Endocrinologia

Orientadora: Profa. Dra. Ana Luiza Maia

Aprovada em: Porto Alegre, 12 de dezembro de 2023

BANCA EXAMINADORA:

Profa. Dra. Fernanda Vaisman Balieiro
Instituto Nacional do Câncer (INCA)

Profa. Dra. Vanessa Suñé Mattevi
Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA)

Profa. Dra. Marcia Silveira Graudenz
Universidade Federal do Rio Grande do Sul (UFRGS)

AGRADECIMENTOS

Agradeço aos meus pais, Carlos e Selma, e à minha irmã, Joana, por todo o amor e carinho que recebi. Nada disso seria possível sem vocês, que sempre acreditaram em mim e me apoiaram integralmente.

Agradeço à minha orientadora, Profa. Dra. Ana Luiza Maia, por ter me acolhido como seu orientando desde o início da iniciação científica até o final do doutorado, numa caminhada de oito anos. Tenho a ela a agradecer pela confiança, pelas oportunidades, pelos ensinamentos e pela amizade.

Agradeço ao amigo Iuri Goemann, que abraçou todos os projetos de pesquisa comigo até hoje, sempre me estimulando a sonhar. Foi ele quem inicialmente instigou minha curiosidade sobre o universo de possibilidades contido em bioinformática, *machine learning* e *big data*.

Agradeço ainda a todos os professores e alunos da Unidade de Tireoide do HCPA, que estiveram sempre presentes, contribuindo e dando suporte aos projetos.

Agradeço também à Profa. Dra. Mariana Recamonde-Mendoza, ao Núcleo de Bioinformática e ao Instituto de Informática, que contribuíram com essas pesquisas.

Por fim, agradeço a todas as instituições que me deram treinamento e suporte ao longo da minha formação (UFRGS, HCPA, CAPES, CNPq, FAPERGS).

Esta Tese de Doutorado segue o formato proposto pelo Programa de Pós-Graduação em Ciências Médicas: Endocrinologia, Faculdade de Medicina, Universidade Federal do Rio Grande do Sul, sendo apresentada na forma de uma introdução e dois artigos originais sobre o tema da tese:

- **Artigo original 1:** Marczyk VR, Recamonde-Mendoza M, Maia AL, Goemann IM. Classification of Thyroid Tumors Based on DNA Methylation Patterns. *Thyroid*. 2023;33(9):1090-1099. doi:10.1089/thy.2023.0074
Fator de Impacto: 6.6
- **Artigo original 2:** TERT promoter mutations C228T and C250T have distinct impacts on the transcriptional program and clinical behavior of papillary thyroid carcinomas
(Em submissão)

Dados preliminares do artigo original da presente tese foram apresentados nos seguintes eventos científicos:

- Congresso Brasileiro de Atualização em Endocrinologia e Metabologia, 2023, João Pessoa, PB
TERT Promoter Mutations C228T and C250T Have Distinct Impacts on the Transcriptional Program and Clinical Behavior of Papillary Thyroid Carcinomas
3º Lugar – Melhores Temas Livres Orais
- Congresso Brasileiro de Atualização em Endocrinologia e Metabologia, 2023, João Pessoa, PB
DNA Methylation Patterns Define Subtypes of Differentiated Follicular Cell-Derived Thyroid Neoplasms
.
- XIX Latin American Thyroid Society Congress, 2023, Curitiba, PR
TERT Promoter Mutation C228T is Associated with Major Transcriptional Alterations and Aggressive Clinical Course in Papillary Thyroid Carcinomas.
- XIX Latin American Thyroid Society Congress, 2023, Curitiba, PR
A DNA Methylation-Based Classification for Thyroid Neoplasms Using an Unsupervised Machine Learning Approach.

Além dos artigos que fazem parte da presente tese, ao longo do período do doutorado foram desenvolvidos os seguintes trabalhos:

Artigos Completos:

- Goemann IM, Marczyk VR, Recamonde-Mendoza M, Wajner SM, Graudenz MS, Maia AL. Decreased expression of the thyroid hormone-inactivating enzyme type 3 deiodinase is associated with lower survival rates in breast cancer. *Sci Rep* (2020) 10:13914. doi: 10.1038/s41598-020-70892-4
- Marczyk VR, Rosa DD, Maia AL, Goemann IM. Overall Survival for HER2-Positive Breast Cancer Patients in the HER2-Targeted Era: Evidence From a Population-Based Study. *Clin Breast Cancer* (2022) 22:418–423. doi: 10.1016/j.clbc.2022.03.004

Resumos publicados em anais de congressos:

- Marczyk, V.R., Goemann, I.M., & Maia, A.L.. (2020, Nov). Evidence of Estrogen Signaling and RET Proto-Oncogene Interplay in Breast Cancer. *Archives of Endocrinology and Metabolism*, 64(S2), S13.
- Marczyk, V.R., Goemann, I.M., & Maia, A.L.. (2020, Nov). mRNA and miRNA Expression Patterns Predict BRAF Mutational Status with High Accuracy in Thyroid Cancer: a Supervised Machine Learning Approach. *Archives of Endocrinology and Metabolism*, 64(S2), S158.
- Ning, M., Dudzinski, S., Busaidy, N., Dadu, R., Welsh, J., Hu, M., Hamidi, S., Marczyk, V.R., Gunn, B., Wang, C., Waguespack, S., Phan, J., Beckham, T., Chang, J., Sherman, S., Reddy, J., Ying, A., O'Reilly, M., Chen, A., Lee, A., Gandhi, S., Liao, Z., Lu. (2023, Sep). Definitive Radiotherapy for Oligometastatic and Oligoprogressive Differentiated Thyroid Carcinoma: a Potential Strategy to Defer Escalation of Systemic Therapy. *Thyroid*, 33(S1), A-151.

Apresentações Orais Premiadas em Congressos Científicos:

- Melhor Trabalho Científico Acadêmico – Congresso Internacional Oncologia D'OR, Rio de Janeiro, RJ, 2019
Trabalho: Increased RET Expression in Breast Cancer: a Robust Model within the TCGA Dataset
- 3º Lugar – Melhores Trabalho Área Básica/Translacional – 34º Congresso Brasileiro de Endocrinologia e Metabologia, 2020, Online
Trabalho: Evidence of Estrogen Signaling and RET Proto-Oncogene Interplay in Breast Cancer.
- 3º Lugar – Melhores Temas Livres Orais - Congresso Brasileiro de Atualização em Endocrinologia e Metabologia, 2023, João Pessoa, PB
Trabalho: TERT Promoter Mutations C228T and C250T Have Distinct Impacts on the Transcriptional Program and Clinical Behavior of Papillary Thyroid Carcinomas

RESUMO

A compreensão dos mecanismos genéticos envolvidos na oncogênese das neoplasias de tireoide tem modificado nosso entendimento desse grupo de doenças. A análise molecular vem ganhando espaço, já sendo utilizada clinicamente na avaliação diagnóstica, prognóstica e terapêutica. Dessa forma, o manejo clínico das neoplasias de tireoide cada vez mais é guiado pelas alterações genéticas. Os estudos publicados até hoje tiveram como enfoque a busca por alterações na sequência do DNA, mas pouco se compreende sobre as alterações transcricionais e epigenéticas. O presente estudo objetiva avançar o conhecimento sobre as bases moleculares das neoplasias de tireoide a partir do estudo das alterações da metilação do DNA e da expressão gênica. Essa tese é composta por dois estudos: o primeiro dedicado à análise dos padrões de metilação do DNA encontrados nesse grupo de doenças; o segundo dedicado a investigar o impacto das mutações na região promotora do gene *TERT* sobre a expressão gênica. No primeiro estudo, foram analisados dados de metilação reunidos de seis estudos prévios, totalizando 810 amostras. Esses dados foram empregados para o treinamento de um modelo de aprendizado de máquina não supervisionado, sem qualquer dado clínico ou histopatológico. O modelo identificou que as amostras podiam ser divididas em três subtipos (*clusters*) com base exclusivamente no perfil de metilação. Esses subtipos estavam fortemente associados ao diagnóstico histopatológico ($P < 0.001$) e foram, portanto, denominados *normal-like*, *folicular-like* e *papilar-like*. Adenomas e carcinomas foliculares ou oncocíticos compunham o subtipo *folicular-like*. Carcinomas papilíferos clássicos ou de células altas compunham o subtipo *papilar-like*. Os subtipos identificados estavam também associados aos *drivers* oncogênicos ($P < 0.001$): 99% dos tumores com a mutação *BRAF* V600E eram *papilar-like*, enquanto 96% daqueles com mutações em *RAS* eram *folicular-like*. Por outro lado, amostras diagnosticadas como variante folicular do carcinoma papilífero acabaram dispersas em ambos os subtipos, demonstrando que esse tipo histológico não representa uma única doença, mas sim duas doenças distintas que compartilham características histopatológicas. Na segunda parte, foram empregados dados do *The Cancer Genome Atlas* para análise das alterações transcricionais associadas às mutações da região promotora do gene *TERT* (C228T e C250T). Sabe-se que essas mutações aumentam a expressão do gene *TERT*, conferindo maior agressividade e piores desfechos. Não se sabe, entretanto, se ambas as mutações são igualmente oncogênicas. Em nossa análise, observou-se que a mutação C228T estava associada a um aumento de 15 vezes ($P < 0.001$) nos níveis de RNAm do gene *TERT*, enquanto a mutação C250T estava associada a um aumento de apenas 2 vezes ($P = 0.034$). Clinicamente, a mutação C228T foi associada a idade avançada, invasão linfonodal e metástases. A sobrevida livre de metástases em 5 anos para pacientes com a mutação C228T foi de 79% em comparação a 98% para aqueles sem a mutação. Dessa forma, nossos dados sugerem que a mutação C228T possa conferir aos tumores maior agressividade, servindo como marcador prognóstico. Os dados apresentados aqui demonstram o potencial das ferramentas computacionais em alavancar nosso entendimento das bases moleculares das neoplasias de tireoide, oferecendo novos *insights* sobre a classificação e o prognóstico dessas doenças.

Palavras-chave: Câncer de tireoide. Neoplasias de tireoide. Oncologia de precisão. Aprendizado de máquina. Machine learning. Big data. Bioinformática.

ABSTRACT

Understanding the genetic mechanisms involved in the oncogenesis of thyroid neoplasms has changed our comprehension of this group of diseases. Molecular analysis has been gaining ground and is already used clinically in diagnostic, prognostic, and therapeutic evaluation. Therefore, the clinical management of thyroid neoplasms is increasingly guided by genetic alterations. The studies published to date have focused on the search for changes in the DNA sequence, but little is known about transcriptional and epigenetic changes. The present study aims to advance knowledge about the molecular bases of thyroid neoplasms by studying changes in DNA methylation and gene expression. This thesis is composed of two studies: the first is dedicated to the analysis of DNA methylation patterns found in this group of diseases; the second is dedicated to investigating the impact of mutations in the promoter region of the *TERT* gene on gene expression. In the first study, methylation data gathered from six previous studies were analyzed, totaling 810 samples. These data were used to train an unsupervised machine learning model without any clinical or histopathological data. The model identified that samples could be divided into three subtypes (clusters) based solely on the methylation profile. These subtypes were strongly associated with histopathological diagnosis ($P < 0.001$) and were therefore named normal-like, follicular-like, and papillary-like. Follicular or oncocytic adenomas and carcinomas comprised the follicular-like subtype. Classical papillary or tall cell carcinomas comprised the papillary-like subtype. The identified subtypes were also associated with oncogenic drivers ($P < 0.001$): 99% of tumors with the *BRAF* V600E mutation were papillary-like, while 96% of those with *RAS* mutations were follicular-like. On the other hand, samples diagnosed as a follicular variant of papillary carcinoma ended up dispersed across both subtypes, demonstrating that this histological type does not represent a single disease but rather two distinct diseases that share histopathological characteristics. In the second part, data from The Cancer Genome Atlas were used to analyze transcriptional changes associated with mutations in the promoter region of the *TERT* gene (C228T and C250T). It is known that these mutations increase the expression of the *TERT* gene, resulting in greater aggressiveness and worse outcomes. It is not known, however, whether both mutations are equally oncogenic. In our analysis, it was observed that the C228T mutation was associated with a 15-fold increase ($P < 0.001$) in *TERT* gene mRNA levels, while the C250T mutation was associated with an increase of only 2-fold ($P = 0.034$). Clinically, the C228T mutation was associated with advanced age, lymph node invasion, and metastases. The 5-year metastasis-free survival for patients with the C228T mutation was 79% compared to 98% for those without the mutation. Therefore, our data suggest that the C228T mutation may confer tumors greater aggressiveness, serving as a prognostic marker. The data presented herein demonstrate the potential of computational tools in leveraging our understanding of the molecular basis of thyroid neoplasms, offering new insights into the classification and prognosis of these diseases.

Keywords: Thyroid cancer. Thyroid neoplasms. Precision oncology. Machine learning. Big data. Bioinformatics.

SUMÁRIO

INTRODUÇÃO	13
PARTE 1 - Classification of Thyroid Tumors Based on DNA Methylation Patterns....	18
PARTE 2 - <i>TERT</i> promoter mutations C228T and C250T have distinct impacts on the transcriptional program and clinical behavior of papillary thyroid carcinomas.....	43
CONCLUSÃO.....	63
REFERÊNCIAS	64

INTRODUÇÃO

Nódulos de tireoide são encontrados em aproximadamente 5% das mulheres e 1% dos homens através do exame físico do pescoço (1). A prevalência desses nódulos pode chegar a 68% com o uso de exames ultrassonográficos de alta resolução (1,2). Entretanto, estima-se que somente 5 a 13% dos nódulos identificados sejam malignos (1,2).

As neoplasias de tireoide derivadas da célula folicular são o tipo mais comum de tumor de tireoide e são o foco do presente estudo. Recentemente, a nomenclatura dessas doenças foi revisada e atualizada pela Organização Mundial da Saúde (OMS), com a publicação de um novo sistema de classificação (5ª edição, 2022)(3,4). Esse novo sistema tem como objetivo conciliar a tradicional classificação histopatológica com o conhecimento adquirido ao longo dos últimos anos sobre as alterações genéticas responsáveis pelo desenvolvimento e progressão dessas neoplasias.

Figura 1: Classificação das neoplasias de tireoide derivadas da célula folicular conforme a 5ª classificação da Organização Mundial da Saúde

Lesões benignas

- Doença nodular folicular
- Adenoma folicular
- Adenoma oncocítico
- Adenoma folicular com arquitetura papilífera

Neoplasias de baixo risco

- Neoplasia folicular não invasiva com características nucleares tipo papilíferas
- Tumor folicular de potencial maligno incerto
- Tumor bem diferenciado de potencial maligno incerto
- Tumor trabecular hialinizante

Neoplasias malignas

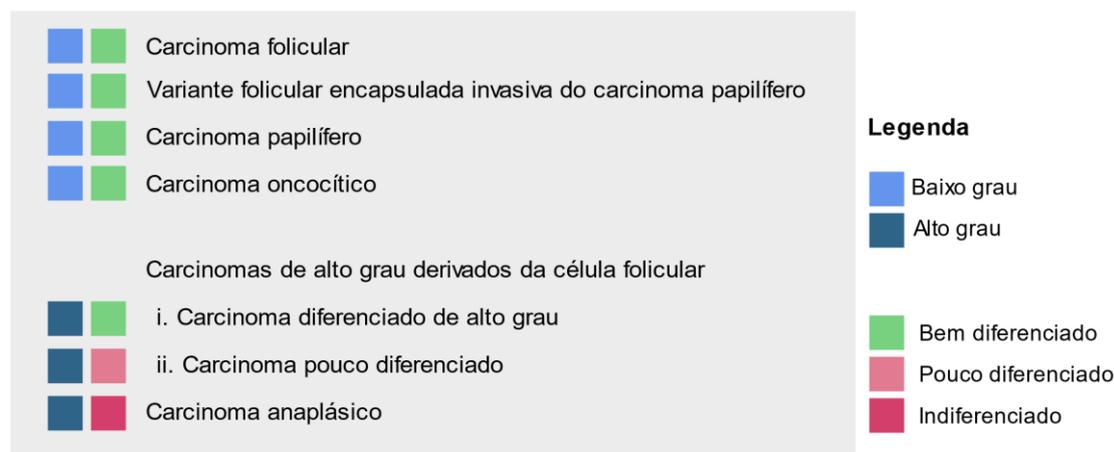
- Carcinoma folicular
- Variante folicular encapsulada invasiva do carcinoma papilífero
- Carcinoma papilífero
- Carcinoma oncocítico
- Carcinomas de alto grau derivados da célula folicular
 - Carcinoma diferenciado de alto grau
 - Carcinoma pouco diferenciado
- Carcinoma anaplásico

Fonte: Elaborado pelo autor.

O novo sistema de classificação desenvolvido pela OMS classifica os tumores de tireoide derivados da célula folicular de acordo com seu comportamento biológico em três grandes categorias: tumores benignos, neoplasias de baixo risco e neoplasias malignas (**Figura 1**)(3,4). Dentro das neoplasias malignas (*i.e.*, carcinomas de tireoide), encontram-se os cânceres bem diferenciados e de baixo grau (carcinoma papilífero, carcinoma folicular, carcinoma oncocítico e variante folicular encapsulada invasiva do carcinoma papilífero), assim como os cânceres de alto grau (carcinoma diferenciado de alto grau, carcinoma pouco diferenciado e carcinoma anaplásico) (**Figura 2**). Com esse novo modelo de nomenclatura, reconhece-se que as neoplasias de tireoide derivadas da célula folicular devem ser compreendidas como um espectro contínuo, que vai desde os tumores benignos até as neoplasias malignas de alto grau (3,5,6).

Figura 2: Características das neoplasias malignas de tireoide derivadas da célula folicular conforme a 5ª classificação da Organização Mundial da Saúde (2022)

Neoplasias malignas derivadas da célula folicular



Fonte: Elaborado pelo autor.

Atualmente, reconhece-se que a maioria dos tumores de tireoide se inicia a partir de um único *driver* oncogênico (evento genético inicial) e que a aquisição de alterações genéticas adicionais (eventos genéticos tardios) é responsável pela progressão dessas neoplasias (3,5–9)(5,6,8,9). Os *drivers* oncogênicos mais frequentes são as mutações nos genes *BRAF* e *RAS* (*N/H/KRAS*), assim como as fusões nos genes *RET*, *NTRK1/3*, *ALK*, *BRAF*, *PPARG* e *THADA*. Os eventos tardios mais frequentes são as mutações nos genes *TERT*, *TP53*, *PTEN*, *EIF1AX*, *AKT1* e *PIK3CA* (3,5–9).

Como recentemente revisado por *Fagin e Nikiforov (5)*, esse espectro das neoplasias derivadas da célula folicular difere conforme o evento genético inicial. Certos eventos iniciais, como as mutações *BRAF V600E*, originam universalmente lesões malignas, sem que haja uma lesão benigna ou de baixo risco precursora. Nesses casos, a única progressão que pode ser observada é a perda de diferenciação, ocorrendo a transformação de um carcinoma bem diferenciado em um carcinoma pouco diferenciado ou anaplásico. Contrariamente, outros eventos genéticos iniciais, como as mutações *RAS*, podem originar tanto neoplasias benignas e de baixo risco, quanto neoplasias malignas (5). Dessa forma, podemos encontrar o mesmo *driver* oncogênico em ambos os extremos desse espectro, desde o adenoma folicular *RAS*-mutado até o carcinoma anaplásico *RAS*-mutado, sendo os eventos genéticos adicionais os determinantes da apresentação e do comportamento clínico do tumor (3,5,6).

A elucidação dos mecanismos moleculares envolvidos na patogênese dessas doenças já impacta diretamente a prática clínica. A testagem molecular já é bem reconhecida, por exemplo, como ferramenta diagnóstica em certos cenários clínicos (10). Usualmente, o diagnóstico dos nódulos de tireoide é feito através de uma punção aspirativa com agulha fina (PAAF) e subsequente análise citológica. A análise citológica do material da PAAF é feita conforme os critérios do Bethesda System for Reporting Thyroid Cytopathology, daqui em diante apenas “Bethesda”, que classifica os nódulos em categorias de I a VI.(11) A categoria Bethesda I (“não diagnóstico”) indica uma punção com material insatisfatório para análise, devendo ser realizada nova PAAF. As categorias II a VI indicam amostras satisfatórias, sendo o risco de malignidade progressivamente crescente. Nódulos classificados como Bethesda II (“benignos”) apresentam um risco de malignidade ao redor de 4%. Enquanto isso, nódulos classificados como Bethesda V (“suspeito de malignidade”) e VI (“maligno”) apresentam um risco de malignidade de 74% e 97%, respectivamente. Tanto em um extremo (Bethesda II) quanto no outro (Bethesda IV e VI), a classificação citológica é suficiente para o tomada de decisões clínicas. Por outro lado, nódulos classificados como Bethesda III ou IV apresentam um risco de malignidade intermediário, variando entre 13-30% e 23-34%, respectivamente, dificultando a tomada de decisões. Nesses casos, é frequente a indicação de cirurgia, mesmo sabendo-se que aproximadamente 70% desses nódulos são benignos. Nesse cenário, isto é, diante de nódulos com citologia indeterminada, os painéis moleculares podem auxiliar a diferenciação de lesões malignas e benignas (12–15). Certas alterações genéticas, como as mutações *BRAF V600E* e as mutações da região promotora do gene

TERT, são somente encontradas em lesões malignas, de forma que sua identificação é suficiente para estabelecer o diagnóstico (5). Por outro lado, várias alterações genéticas, como é o caso das mutações *RAS*, são encontradas tanto em lesões benignas e de baixo risco, quanto em lesões malignas (5). Dessa forma, os painéis moleculares atuais analisam simultaneamente centenas de genes e empregam modelos probabilísticos complexos para refinar a estratificação do risco de malignidade (10).

Apesar dos avanços recentes no entendimento desses mecanismos moleculares, permanecem ainda importantes lacunas no nosso conhecimento nessa área. A maioria dos estudos publicados até hoje teve como enfoque a busca por alterações na sequência do DNA (*e.g.* mutações, fusões, deleções, inserções), mas muito menos se compreende sobre as alterações transcricionais e epigenéticas envolvidas nessas doenças. Outra importante limitação do conhecimento atual advém do desenho dos estudos que caracterizaram as alterações moleculares nas neoplasias de tireoide. Tais estudos tomaram como ponto de partida o diagnóstico histopatológico das neoplasias e, geralmente, ficaram restritos a um único subtipo histológico. Dessa forma, existe um importante viés de seleção a ser considerado, sendo difícil a generalização dos achados para os demais subtipos. O pioneiro (e até hoje mais abrangente) desses trabalhos foi *The Cancer Genome Atlas* (TCGA), publicado em 2014 (7). Esse foi o primeiro estudo a realizar uma caracterização multi-ômica de um grande conjunto ($n=496$) de carcinomas papilíferos de tireoide, analisando alterações na sequência do DNA, alterações de expressão de RNAm, microRNA e proteínas, assim como alterações na metilação do DNA. Entretanto, esse estudo se limitou a investigar apenas os carcinomas papilíferos em estágio inicial, o que dificulta a generalização dos achados. Ademais, esse desenho de estudo focado em um único subtipo histológico traz uma limitação adicional. Assumimos que a classificação histopatológica seja a correta/ideal e olhamos apenas para um subgrupo dos tumores; assim, perdemos a oportunidade de analisar todo o conjunto de amostras e procurar por outras similaridades e diferenças que talvez não fossem evidentes à microscopia.

O presente estudo objetiva avançar o conhecimento sobre as bases moleculares das neoplasias de tireoide derivadas da célula folicular a partir do estudo das alterações da metilação do DNA (metiloma) e da expressão gênica (transcriptoma). Esse tipo de análise apresenta desafios que decorrem do grande volume de dados e, nesse contexto, as técnicas de bioinformática e de aprendizado de máquina possibilitam a análise desse vasto conjunto de informações de forma robusta.

Essa tese é composta por dois estudos distintos e complementares. O primeiro estudo é dedicado à análise dos padrões de metilação do DNA encontrados nesse grupo de doenças, tanto nas neoplasias benignas e de baixo risco, quanto nas neoplasias malignas. Nesse estudo, emprega-se um algoritmo de aprendizado de máquina não supervisionado com o objetivo de se agrupar e classificar essas doenças exclusivamente com base em seu perfil de metilação. Dessa forma, a análise conduzida não toma por base qualquer noção pré-concebida ou qualquer conhecimento prévio a respeito das neoplasias de tireoide, propondo-se a construir do zero uma nova visão sobre a oncogênese e a classificação dessas doenças.

O segundo estudo tem um objetivo mais delimitado, propondo-se a investigar o impacto das mutações na região promotora do gene *TERT* sobre a expressão gênica e sobre as vias de sinalização celular. Esse tipo de mutação tem uma característica muito diferente de todas as demais mutações patogênicas envolvidas nas neoplasias de tireoide: essa é a única mutação não codificante conhecida (16). Todas as demais mutações conhecidas da oncogênese das neoplasias de tireoide alteram a sequência de pares de base do DNA na região dos éxons (*i.e.* região codificante) e, assim, resultam em um RNAm alterado e em uma proteína alterada. Contrariamente, as mutações do gene *TERT* ocorrem na região promotora, anteriormente tanto ao sítio de início da tradução quanto de início da transcrição. Dessa forma, os mecanismos oncogênicos envolvidos são distintos e ainda pouco conhecidos (16).

CONCLUSÃO

Os resultados apresentados na primeira parte desse trabalho oferecem novos *insights* sobre as alterações epigenéticas encontradas nas neoplasias de tireoide derivadas da célula folicular. Esses resultados oferecem uma nova forma de compreendermos e classificarmos tais neoplasias com base no seu perfil de metilação. Por se tratar de um modelo de classificação não supervisionado construído com base apenas em dados de metilação, tal método oferece um sistema de classificação robusto e biologicamente significativo/coerente.

Os resultados apresentados na segunda parte demonstram que as duas mutações frequentemente encontradas na região promotora do gene *TERT* tem impacto distinto no aumento da expressão gênica, tendo a mutação C228T um maior potencial de ativação transcricional do que a mutação C250T. A mutação C228T foi consistentemente associada a agressividade tumoral e a piores desfechos clínicos, podendo ser empregada como um biomarcador prognóstico.

Em conjunto, os dados apresentados aqui demonstram o potencial das ferramentas computacionais em alavancar nosso entendimento das bases moleculares das neoplasias de tireoide, oferecendo novos *insights* sobre a classificação e o prognóstico dessas doenças.

REFERÊNCIAS

1. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*®. 2015 Oct 14;26(1):1–133.
2. Durante C, Grani G, Lamartina L, Filetti S, Mandel SJ, Cooper DS. The Diagnosis and Management of Thyroid Nodules: A Review. *JAMA*. 2018 Mar 6;319(9):914–24.
3. Baloch ZW, Asa SL, Barletta JA, Ghossein RA, Juhlin CC, Jung CK, et al. Overview of the 2022 WHO Classification of Thyroid Neoplasms. Vol. 33, *Endocrine Pathology*. 2022. p. 27–63.
4. Christofer Juhlin C, Mete O, Baloch ZW. The 2022 WHO classification of thyroid tumors: novel concepts in nomenclature and grading. *Endocr Relat Cancer*. 2023;30(2):e220293.
5. Fagin JA, Nikiforov YE. Progress in Thyroid Cancer Genomics: A 40-Year Journey. *Thyroid*®. 2023 Sep 4;33(11):1271–86.
6. Fagin JA, Wells SA. Biologic and Clinical Perspectives on Thyroid Cancer. *N Engl J Med*. 2016 Sep 15;375(11):1054–67.
7. Agrawal N, Akbani R, Aksoy BA, Ally A, Arachchi H, Asa SL, et al. Integrated Genomic Characterization of Papillary Thyroid Carcinoma. *Cell*. 2014 Oct 23;159(3):676–90.
8. Landa I, Ibrahimasic T, Boucai L, Sinha R, Knauf JA, Shah RH, et al. Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. *J Clin Invest*. 2016 Mar 1;126(3):1052–66.
9. Pozdeyev N, Gay LM, Sokol ES, Hartmaier R, Deaver KE, Davis S, et al. Genetic Analysis of 779 Advanced Differentiated and Anaplastic Thyroid Cancers. *Clin Cancer Res*. 2018 Jul 1;24(13):3059–68.

10. Durante C, Hegedüs L, Czarniecka A, Paschke R, Russ G, Schmitt F, et al. 2023 European Thyroid Association Clinical Practice Guidelines for thyroid nodule management. *Eur Thyroid J*. 2023;12(5):e230067.
11. Ali SZ, Baloch ZW, Cochand-Priollet B, Schmitt FC, Vielh P, VanderLaan PA. The 2023 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid®*. 2023 Jul 8;33(9):1039–44.
12. Steward DL, Carty SE, Sippel RS, Yang SP, Sosa JA, Sipos JA, et al. Performance of a Multigene Genomic Classifier in Thyroid Nodules With Indeterminate Cytology: A Prospective Blinded Multicenter Study. *JAMA Oncol*. 2019 Feb 1;5(2):204–12.
13. Alexander EK, Kennedy GC, Baloch ZW, Cibas ES, Chudova D, Diggans J, et al. Preoperative Diagnosis of Benign Thyroid Nodules with Indeterminate Cytology. *N Engl J Med*. 2012 Jun 25;367(8):705–15.
14. Patel KN, Angell TE, Babiarz J, Barth NM, Blevins T, Duh Q-Y, et al. Performance of a Genomic Sequencing Classifier for the Preoperative Diagnosis of Cytologically Indeterminate Thyroid Nodules. *JAMA Surg*. 2018 Sep 1;153(9):817–24.
15. Livhits MJ, Zhu CY, Kuo EJ, Nguyen DT, Kim J, Tseng C-H, et al. Effectiveness of Molecular Testing Techniques for Diagnosis of Indeterminate Thyroid Nodules: A Randomized Clinical Trial. *JAMA Oncol*. 2021 Jan 1;7(1):70–7.
16. Vinagre J, Almeida A, Pópulo H, Batista R, Lyra J, Pinto V, et al. Frequency of TERT promoter mutations in human cancers. *Nat Commun*. 2013;4(1):2185.