

Oviedo, 2023

Biomarcadores de metilación del ADN para el diagnóstico precoz del cáncer de tiroides

Tesis doctoral

Paula Morales Sánchez

Directores

Elías Delgado Álvarez

Mario Fernández Fraga



Universidad de Oviedo
Programa Oficial de Doctorado
de Ciencias de la Salud

UNIVERSIDAD DE OVIEDO

**PROGRAMA OFICIAL DE DOCTORADO EN CIENCIAS
DE LA SALUD**



Universidad de Oviedo

**Biomarcadores de metilación del ADN para el diagnóstico precoz del
cáncer de tiroides**

DNA methylation biomarkers for early diagnosis of thyroid cancer

TESIS DOCTORAL

Paula Morales Sánchez

2023



Universidad de Oviedo

AUTORIZACIÓN PARA LA PRESENTACIÓN DE TESIS DOCTORAL

Año Académico: 2022/2023

1.- Datos personales del autor de la Tesis		
Apellidos: Morales Sánchez		Nombre: Paula
DNI/Pasaporte/NIE: 71728177D	Teléfono: 684629387	Correo electrónico: paulamsov76@gmail.com
2.- Datos académicos		
Programa de Doctorado cursado: PROGRAMA OFICIAL DE DOCTORADO EN CIENCIAS DE LA SALUD		
Órgano responsable: Centro Internacional de Postgrado		
Título definitivo de la Tesis		
Español/Otro Idioma: Biomarcadores de metilación del ADN para el diagnóstico precoz del cáncer de tiroides		Inglés: DNA methylation biomarkers for early diagnosis of thyroid cancer
Rama de conocimiento		<input type="checkbox"/> Arte y Humanidades <input type="checkbox"/> Ciencias <input checked="" type="checkbox"/> Ciencias de la Salud <input type="checkbox"/> Ciencias Sociales y Jurídicas <input type="checkbox"/> Ingeniería y Arquitectura
3.- Autorización del Directores y Tutor de la tesis		
D/D ^a : ELIAS DELGADO ALVAREZ		DNI/Pasaporte/NIE: 09751167H
Departamento/Instituto: DEPARTAMENTO DE MEDICINA / FACULTAD DE MEDICINA Y CIENCIAS DE LA SALUD / UNIVERSIDAD DE OVIEDO		
D/D ^a : MARIO FERNANDEZ FRAGA		DNI/Pasaporte/NIE: 10875819Q
Departamento/Instituto/Institución: DEPARTAMENTO DE BIOLOGÍA DE ORGANISMOS Y SISTEMAS / FACULTAD DE BIOLOGIA / UNIVERSIDAD DE OVIEDO		
Autorización del Tutor de la tesis		
D/D ^a : <i>Jose Cuartero Audin</i>		DNI/Pasaporte/NIE: <i>09371005T</i>
Departamento/Instituto: <i>Medicine</i>		

Autoriza la presentación de la tesis doctoral en cumplimiento de lo establecido en el Art. 32 del Reglamento de los Estudios de Doctorado, aprobado por el Consejo de Gobierno, en su sesión del día 20 de julio de 2018 (BOPA del 9 de agosto de 2018)

Directores de la Tesis

Fdo.: ELIAS DELGADO Y MARIO F. FRAGA

Tutor de la Tesis

Fdo.: *Jose Cuartero Audin*

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO
EN _____



Universidad de Oviedo

RESOLUCIÓN DE PRESENTACIÓN DE TESIS DOCTORAL

Año Académico:2022/2023

1.- Datos personales del autor de la Tesis		
Apellidos: MORALES SANCHEZ	Nombre: PAULA	
DNI/Pasaporte/NIE: 71728177D	Teléfono: 684629387	Correo electrónico: paulamsov76@gmail.com

2.- Datos académicos	
Programa de Doctorado cursado: PROGRAMA DE CIENCIAS DE LA SALUD	
Órgano responsable: Centro Internacional de Postgrado	
Departamento/Instituto en el que presenta la Tesis Doctoral: INSTITUTO DE INVESTIGACIÓN DEL PRINCIPADO DE ASTURIAS	
Título definitivo de la Tesis	
Español/Otro Idioma: Biomarcadores de metilación del ADN para el diagnóstico precoz del cáncer de tiroides	Inglés: DNA methylation biomarkers for early detection of thyroid cancer
Rama de conocimiento: <input type="checkbox"/> Arte y Humanidades <input type="checkbox"/> Ciencias <input checked="" type="checkbox"/> Ciencias de la Salud <input type="checkbox"/> Ciencias Sociales y Jurídicas <input type="checkbox"/> Ingeniería y Arquitectura	
Señale si procede: <input type="checkbox"/> Mención Internacional <input type="checkbox"/> Mención Industrial <input type="checkbox"/> Idioma de presentación de la Tesis distinto al español <input type="checkbox"/> Presentación como compendio de publicaciones	

3.- Autorización del Presidente de la Comisión Académica	
D/D ^a : ANA MARIA NAVARRO INCIO	DNI/Pasaporte/NIE: 093785875
Departamento/Instituto: MORFOLOGIA Y BIOLOGIA CELULAR	

Resolución: La Comisión Académica del Programa de Doctorado /El Departamento en su reunión de fecha , acordó la presentación de la tesis doctoral a la Comisión de Doctorado, previa comprobación de que la tesis presentada y la documentación que la acompaña cumplen con la normativa vigente, según lo establecido en el Art.32.8 del Reglamento de los Estudios de Doctorado, aprobado por el Consejo de Gobierno, en su sesión del día 20 de julio de 2018 (BOPA del 9 de agosto de 2018)

Además, informa:

	Favorable	Desfavorable
• Mención Internacional	<input type="checkbox"/>	<input type="checkbox"/>
• Mención Industrial	<input type="checkbox"/>	<input type="checkbox"/>
• Idioma	<input type="checkbox"/>	<input type="checkbox"/>
• Presentación como compendio de publicaciones	<input type="checkbox"/>	<input type="checkbox"/>



Universidad de Oviedo

Justificación

SE ESTIMA QUE EL PRESENTE TRABAJO DE INVESTIGACIÓN PRESENTADO COMO TESIS DOCTORAL CUMPLE TODOS LOS REQUISITOS PARA SU DEFENSA Y PARA OPTAR AL GRADO DE DOCTOR POR LA UNIVERSIDAD DE OVIEDO

<lugar> <fecha> Oviedo, 12 de diciembre 2022

Presidente de la Comisión Académica del Programa de Doctorado

Fdo.:



Contra la presente Resolución, podrá interponer recurso de alzada ante el Rectorado, en el plazo de un mes, a partir del día siguiente al de la presente notificación, de conformidad con el art. 122 de la Ley 39/2015, de 1 de octubre, de Procedimiento Administrativo Común de las Administraciones Públicas

SR. _____

(Doctorando, Director de la Tesis Doctoral y Sr. Director del Centro Internacional de Postgrado)



RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: Biomarcadores de metilación del ADN para el diagnóstico precoz del cáncer de tiroides	Inglés: DNA methylation biomarkers for early diagnosis of thyroid cancer
2.- Autor	
Nombre: Paula Morales Sánchez	DNI/Pasaporte/NIE: 71728177D
Programa de Doctorado: Programa Oficial de Doctorado de Ciencias de Salud	
Órgano responsable: Centro Internacional de Postgrado	

RESUMEN (en español)

Los nódulos tiroideos son la patología endocrina más frecuente, diagnosticándose malignidad en una pequeña proporción (~15 %). La incidencia del cáncer de tiroides se ha triplicado en España en los últimos 25 años. Así mismo, afecta tres veces más a mujeres que a hombres y se incrementa su frecuencia con la edad, sobre todo a partir de los 40 años. En general, son tumores de gran pronóstico, con tasas de supervivencia cercanas al 100 %. Sin embargo, debido a la alta prevalencia de los nódulos en la población es necesario un triaje exhaustivo.

La punción por aspiración con aguja fina es la prueba no quirúrgica de referencia para el diagnóstico inicial de los nódulos tiroideos. Sin embargo, el 25-30 % de todas las punciones son designadas como indeterminadas y dada que la posibilidad de malignidad, muchos pacientes suelen ser derivados a cirugía. Tras la operación, solo el 25 % de las lesiones resultan malignas, por lo que el riesgo de cáncer no es lo suficientemente alto como para respaldar definitivamente la cirugía. En retrospectiva, estas tiroidectomías son innecesarias y generan complicaciones que los pacientes padecerán durante toda su vida. Por lo tanto, existe una necesidad urgente de desarrollar métodos efectivos para discernir benignidad de malignidad en las lesiones foliculares de tiroides.

El desarrollo de biomarcadores seguros y rentables podría ayudar en la toma de decisiones del médico tratante para manejar con mayor certeza los nódulos tiroideos. Si bien durante mucho tiempo se ha tratado de establecer el valor diagnóstico de las diferentes alteraciones genéticas ninguna parece ser totalmente resolutive. La metilación del ADN es una de las principales modificaciones epigenéticas y su papel en el inicio, desarrollo y metástasis de cánceres humanos ha sido ampliamente descrito, incluido en cáncer de tiroides. Dado que las firmas de metilación del ADN pueden ser indicativas de la plasticidad molecular que surge en la tumorigénesis temprana, en este estudio se planteó que esta marca epigenética podría funcionar como un marcador para la discriminación en casos difíciles de diagnosticar.



Inicialmente se realizó un análisis mutacional mediante secuenciación del exoma completo en una cohorte de descubrimiento multicéntrica retrospectiva que agrupa 54 muestras de tiroides con patrón folicular. Aunque se encontraron alteraciones propias de los nódulos tiroideos, no pudo establecerse una firma distintiva. Por ello, en la misma cohorte se llevó a cabo un análisis de metilación del ADN empleando la plataforma EPIC de Illumina. Utilizando herramientas de aprendizaje automatizado se desarrolló un modelo de diagnóstico basado en tres sitios CpG (cg17154646, cg21915100 y cg10189462) que posteriormente se validó mediante diferentes aproximaciones ortogonales (bases de datos externas y pirosecuenciación con bisulfito). Cuando se evaluó el desempeño del clasificador en una cohorte de validación independiente (n = 125) empleando la técnica de pirosecuenciación, se obtuvo una sensibilidad del 96 %, una especificidad de 82 % y unos valores predictivos positivos y negativos del 77 % y 97 %, respectivamente. Curiosamente, usando este modelo, cuando se asignaron puntuaciones de malignidad a los distintos tumores tiroideos, se pudo observar que algunos tumores benignos ya albergaban alteraciones moleculares propias de los carcinomas. Asimismo, se probó el desempeño en punciones de tiroides (n = 33), clasificando correctamente el 75 % de las muestras que indeterminadas que normalmente son derivadas a cirugía.

Dados los resultados, el sistema de clasificación basado con un conjunto mínimo de biomarcadores epigenéticos puede complementar el potencial de las técnicas diagnósticas actualmente disponibles. Así mismo, la clasificación pronóstica de las muestras en función de las puntuaciones de malignidad priorizaría un número considerable de intervenciones quirúrgicas que a menudo se realizan debido a una citología incierta.

RESUMEN (en Inglés)

Thyroid nodules are the most common encountered thyroid endocrine pathology, being unfrequently diagnosed as malignant (~15%). The incidence of thyroid cancer has tripled in Spain in the last 25 years. Likewise, it affects three times more women than men and its frequency increases with age, especially after the age of 40. In general, they are tumors with a great prognosis, with survival rates close to 100%. However, due to the high prevalence of nodules in the population, an exhaustive triage is necessary.

Fine needle aspiration puncture is the gold standard non-surgical test for the initial diagnosis of thyroid nodules. However, 25-30% of all cytologies are designated as indeterminate and given the possibility of malignancy, many patients are often referred for surgery. After the operation, only 25% of lesions result as cancer, so the risk of malignancy is not high enough to definitively support thyroidectomy. In hindsight, these thyroidectomies are necessary and lead to complications that patients



will endure for their entire lives. Therefore, there is an urgent need to develop effective methods to discern benign from malignant follicular-patterned thyroid lesions.

Safe and cost-effective biomarkers could help in the decision-making of the treating physician to manage thyroid nodules with greater certainty. Although many attempts have been made to establish the diagnostic value of the different genetic alterations, none of them seems to be fully resolvable. DNA methylation is one of the main epigenetic modifications and its role in the initiation, development and metastasis of human cancers has been widely described, including thyroid cancer. Since DNA methylation signatures may be indicative of molecular plasticity that arises in early tumorigenesis, this study postulated that this epigenetic mark could be used as a marker for discrimination in difficult-to-diagnose cases.

Initially, mutation analysis was performed by whole-exome sequencing in a retrospective multicenter discovery cohort of 54 follicular-patterned thyroid samples. Although common mutations of thyroid nodules were detected, a distinctive signature could not be established. Therefore, in the same cohort, a DNA methylation analysis was carried out using the Illumina EPIC platform. Using a machine learning pipeline, a diagnostic model based on three CpG sites (cg17154646, cg21915100 and cg10189462) was developed. Subsequently it was validated by different orthogonal approaches (external databases and bisulfite pyrosequencing). Subsequently, the performance of the classifier was evaluated in an independent validation cohort (n = 125) using the pyrosequencing technique, obtaining a sensitivity of 96%, a specificity of 82%, and positive and negative predictive values of 77% and 97%, respectively. Interestingly, when the same model was applied to assign malignancy scores to the tumours in the validation cohort, it was observed that some thyroid nodules already harbour molecular alterations which were typically seen in carcinomas. Finally, the diagnostic performance in thyroid punctures (n = 33) was tested, our model was capable of correctly classifying 75% of the indeterminate samples that are normally referred to surgery.

Our classification system based on a minimal set of epigenetic biomarkers could complement the potential of the diagnostic techniques currently available. Likewise, the prognostic classification of the samples based on the malignancy scores would prioritize a considerable number of surgical interventions that are often performed due to an uncertain cytology.

Declaración de intereses

Se ha presentado una solicitud de patente (EP22382165) sobre el uso de este sistema de clasificación epigenética para el diagnóstico del cáncer de tiroides. Los autores implicados en el desarrollo de este proyecto no revelan ningún otro conflicto de interés potencial.

Realización de la tesis y financiación

Esta tesis para la consecución del título de doctora, incluido en el Programa de Investigación en Medicina Clínica dentro del Doctorado en Ciencias de la salud de la Universidad de Oviedo, ha sido llevada a cabo en los laboratorios de Endocrinología, Nutrición, Diabetes y Obesidad (ENDO) (Dr. Elías Delgado Álvarez -ISPA/HUCA/CIBERER/Universidad de Oviedo-), y Epigenética del Cáncer y Nanomedicina (Dr. Mario F. Fraga -ISPA/CINN-CSIC/IUOPA/CIBERER/Universidad de Oviedo-; M.F.F) pertenecientes al Instituto de Investigación Sanitaria del Principado de Asturias (ISPA). La investigación reportada en esta tesis ha sido apoyada por la Asociación Española Contra el Cáncer (AECC, PROYE18061FERN a M.F.F.), el Gobierno de Asturias (PCTI) cofinanciación 2018-2022/FEDER (IDI/2018/146 a M.F.F.), la Fundación General CSIC (0348_CIE_6_E a M.F.F.) y el Instituto de Salud Carlos III (Plan Nacional de I+D+I) cofinanciando FEDER (PI18/01527 y PI21/01067 a M.F.F). Asimismo, esta tesis ha sido posible gracias a las Ayudas Predoctorales en Oncología (APRO) de la AECC (PRDAS18003FERN). También agradecemos el apoyo de la IUOPA-ISPA-FINBA (IUOPA cuenta con el apoyo de la Obra Social Cajastur-Liberbank, España).

"Y tú en verdad la atravesarás, claro está. La violenta tormenta de arena. La tormenta de arena metafísica y simbólica. Pero por más metafísica y simbólica que sea, te rasgará cruelmente la carne como si de mil cuchillas se tratase. Muchas personas han derramado allí su sangre y tú, asimismo, derramarás allí la tuya. Sangre caliente y roja. Y esa sangre se verterá en tus manos. Tu sangre y, también, la sangre de los demás. Y cuando la tormenta de arena haya pasado, tu no comprenderás cómo has logrado cruzarla con vida. ¡No! Ni siquiera estarás seguro de que la tormenta haya cesado de verdad. Pero una cosa sí quedara clara. Y es que la persona que surja de la tormenta no será la misma persona que penetró en ella."

Haruki Murakami, Kafka en la Orilla. 2002.

Agradecimientos

Si estoy escribiendo este apartado de la tesis, significa que casi he llegado al final de esta etapa... Aunque todavía no se si me puedo creer que este pasando. Es difícil expresar la sensación que ahora baila en mi cabeza (o en mi pecho) y ser capaz de transmitir lo que han significado todos estos años en mi desarrollo no solo como profesional, sino también como persona.

No me avergüenza reconocer que el Síndrome del Impostor me ha acompañado durante gran parte del trayecto, y creo que no soy la única en pensar que el verse estúpido y hacer un doctorado son amigos inseparables. Digo creo porque justo cuando más fuerte era este sentimiento (y aquí viene la primera de las muchas referencias del manuscrito) me tope con la misma reflexión en un ensayo publicado en Journal of Cell Science (*). Sin embargo, aunque en ese momento no podía verlo así, ahora cuando lo rememoro con más perspectiva, coincido con el autor de este artículo: "La ciencia implica enfrentarnos a nuestra «estupidez más absoluta». Ese tipo de estupidez es un hecho existencial, inherente a nuestros esfuerzos para penetrar en lo desconocido... Una de las cosas más hermosas de la ciencia es que nos permite tambalearnos, fallar una vez tras otra, y sentirnos perfectamente bien mientras aprendemos algo cada una de las veces... Sin duda, los niveles de confianza y la resiliencia emocional ayudan, pero creo que la educación científica debería hacer más fácil lo que es una gran transición: pasar de aprender de lo que otra gente descubrió una vez a aprender de tus propios descubrimientos. Cuanto más cómodos nos sintamos con nuestra estupidez, más profundo será el nivel de lo desconocido, y más probable es que vayamos a hacer un gran descubrimiento". No podemos estar seguros de si estamos haciéndonos las preguntas apropiadas o los experimentos correctos hasta que conseguimos (o no) la respuesta o el resultado. Lo que si tengo claro es que llegar hasta aquí no habría sido posible sin el trabajo, apoyo y ayuda de muchas personas, por lo que mis agradecimientos no pueden dejar de ser numerosos y sinceros. Este trabajo no hubiese sido posible sin la participación desinteresada de todas y cada una de las personas que aún están o han estado presentes durante este periodo de tiempo.

A mi director de tesis, el Dr. Mario Fernández Fraga. Fuiste el primero en darme la oportunidad de entrar en el mundo de la ciencia y la epigenética. Siempre recordaré aquella primera reunión en el antiguo laboratorio del EpiLab en el antiguo HUCA, cuando todavía estaba empezando cuarto de carrera y me abriste las puertas de tu laboratorio. Estaba convencida de que quería aprender ahí y llegara a hacer mi tesis. Gracias a ti por la confianza depositada para guiarme (aunque no haya sido fácil en muchas ocasiones) y a todas las personas que conforman el laboratorio de Epigenética del Cáncer y Nanomedicina por estar dispuestos a ofrecer un consejo o ayudar con un experimento en cualquier momento. Especialmente a Ramón, ya que sin él esta tesis definitivamente no habría sido posible.

A mi director de tesis, el Dr. Elías Delgado Álvarez. Por confiar en mí, no solo para ser la primera estudiante predoctoral de la pequeña gran rama básica del grupo ENDO, sino por permitirme ser la base junto con Sandra en esos inicios oficiales del laboratorio en 2017. Siempre confiando en mi juico, aun cuando yo misma no lo hacía, y alentándome a ir tomando

cada vez mayor responsabilidad. Por abrirme las puertas de tu consulta siempre que necesitaba cualquier cosa o tenía cualquier inquietud y, sobre todo, por apoyarme cuando tras la pandemia la situación en el laboratorio fue complicada. Aunque eso significara que tus mañanas fueran más locas y ajetreadas. Gracias por todo lo que me has enseñado y por siempre estar dispuesto a buscar financiación hasta debajo de las piedras. Por como concibes al grupo, separar la parte clínica del Servicio de Endocrinología y Nutrición del HUCA del laboratorio básico se hace complicado en nuestro caso.

Esta, a mi parecer, es una de nuestras virtudes como grupo: una fuerte relación entre profesionales clínicos y básicos para generar conocimiento por y para los pacientes. Por ello, hay muchas personas a las que agradecer y me disculpo de antemano por no mencionar expresamente a cada una de ellas. Al jefe de Servicio Edelmiro Menéndez Torre. Gracias por ser una fuente interminable de cualquier tipo de conocimiento y compartirlo con los demás de forma tan cercana y amable. A Pedro y Alicia, por ser amigos y psicólogos en ese cuarto que comenzó como una sala de reuniones, pero que se convirtió en una guarida segura en las peores épocas. Gracias por preocuparos en todo momento por mi bienestar debido al estrés derivada tanto de la tesis como de cualquier percance que pude llegar a tener sobre cualquier otra cosa (aunque eso significara que Alicia me persiguiera con una aguja en la mano para hacerme otra analítica). A todos los médicos y personal de enfermería que componen el Servicio de Endocrino, por tener siempre las consultas abiertas y estar dispuestos a ayudar en cualquier momento. A los residentes que están y a los que ya fueron. En especial a Carlos, Lucia, Raúl, Silvia y Gema por hacerme sentir como una más y tener incluso más ganas que yo de que acabe con éxito la tesis. Por los días de congreso (o mejor dicho las noches), las quedadas, los regalos de cumpleaños hechos siempre fuera de tiempo o por presentarme siempre como "la R mayor" de laboratorio. Gracias, de verdad. A Carmen, llegas en el momento con mayor cúmulo de trabajo (por decirlo de manera fina) y gracias a ti pude volver a respirar en el laboratorio. Por ofrecerte desinteresadamente a ayudar con la tesis, por dejarme mi espacio cuando lo necesitaba y por tratarme como si tuviéramos el mismo rango en el laboratorio. A Elsa y Vicky, los pollitos del laboratorio, por llenar el laboratorio de alegría y esas ganas tan grandes que teneis de aprender y de ayudar. A Miguel, quién diría que tras conocernos en los cursos sobre inferencia Bayesiana o "hablar de nuestro libro" en el de Comunicación de la Ciencia, volveríamos a encontrarnos como compañeros de laboratorio. Con vosotros cuatro el laboratorio básico seguro que llegará mucho más lejos, mucha suerte y ánimo.

A todas las personas que conforman el "Laboratorio Clínico" del ISPA. A Yamina por todos los descansos para tomar café que lograron, durante un rato, despejar las preocupaciones del laboratorio. Tantos que en la cafetería del hospital ya podíamos decir: "ponme lo de siempre, por fa". A Pilar, desde las PLs de la carrera hasta volver a encontrarnos para hacer el doctorado. Si la tesis no ha acabado con nosotras, mucho menos lo harán las comas al programar los scripts. A Gabi, porque sabemos que en el momento en el que defendamos la tesis, no solo firmaremos la aceptación del título de doctor... sino que también firmaremos nuestra nueva tara. Espero que eso no nos quite las cervezas curativas de los jueves. Porque, ¿A quién no le va a gustar una tesis romana del siglo I? ¿A quién no le va a gustar?

A todas las personas de la planta 3 de la UFIEC del ISCIII. A Antonio por permitirme realizar mi estancia doctoral en el laboratorio de Tumores Endocrinos, porque, aunque los anticuerpos se hagan los difíciles sabemos que NIS está ahí. Gracias al laboratorio de Proteómica Funcional por la paciencia que mostraron y el hueco que sacaron para poder enseñarme todas esas nuevas técnicas. Porque las promesas se cumplen, a Edu por esa inesperada coincidencia que te llevó a reevaluar los últimos tumores de una de las cohortes de esta tesis. A Mari Paz (sin olvidar a Ruper y a Levi), llegar al laboratorio en tu último día de trabajo pudo ser el más extraño de los comienzos, sin embargo, gracias por la alegría con la que me recibiste e incluirme en la vida de un laboratorio que no era el mío.

A mis ratones biólogos: Andre, Amanda, Carmen, Sara, Olaya y Nere. Por vuestra inquebrantable amistad desde que empezamos la carrera, sé que será eterna como la vida del ficus de Biología. Por tantos momentos que pasamos tanto en los eternos días encerradas en la facultad, como los que disfrutamos fuera de ella. Por preocuparos y enorgulleceros de cada una de nosotras como se fuerais vosotras mismas. Siempre tratando de mantener el contacto, aunque estemos desperdigadas por el mundo.

A mis amigos y compañeros de farándula: Sofi, Gus, Nere, Carlota, Javi, Carlos, Pablo, Carla y Dube, gracias por simplemente estar ahí, siempre es un alivio poder salir del ambiente científico y aunque os preocupéis por lo que hago, no tener que hablar de ello. Aunque no os lo diga: os quiero. Especialmente agradecer a Sofi, se dice pronto más de media vida siendo las mejores amigas. Darte las gracias por ser como eres y por dejarme ser yo misma. Por ser terriblemente sincera y compartir no solo los momentos felices, sino también los más desastrosos. Gracias a ti y a tú gran familia por tratarme siempre como una Pérez Sotomayor más.

A mi familia, por el orgullo que sienten y su inquebrantable apoyo. A mis abuelos, los que están para enorgullecerse y los que no (pero que sé, a ciencia cierta, que estarían en primera fila para verme cumplir mis sueños). Pero, sobre todo, a mis padres y hermana. Gracias por haberme permitido hacer de mi vida lo que yo elegí y haberme apoyado siempre y sin fisuras. Gracias por empujarme a volar cuando alguna vez he tenido miedo y por ser mi red cuando las alas que me aseguráis que tengo han fallado. A mi hermana Celia, mi máximo apoyo en el mundo. Somos como dos caras de una misma moneda: tan diferentes, pero a la vez tan iguales. Un "os quiero" se queda pequeño. Esta tesis es para vosotros.

Dar las gracias a todos los pacientes que desinteresadamente dieron su consentimiento para formar parte de estudios como el que aquí se presenta. Finalmente, también agradecer a la Asociación Española Contra el Cáncer, ya que, debido a su compromiso con la sociedad y la investigación, la realización de proyectos como los que a continuación se exponen es posible.

RESUMEN

Los nódulos tiroideos son la patología endocrina más frecuente, diagnosticándose malignidad en una pequeña proporción (~15 %). La incidencia del cáncer de tiroides se ha triplicado en España en los últimos 25 años. Asimismo, afecta tres veces más a mujeres que a hombres y se incrementa su frecuencia con la edad, sobre todo a partir de los 40 años. En general, son tumores de gran pronóstico, con tasas de supervivencia cercanas al 100 %. Sin embargo, debido a la alta prevalencia de los nódulos en la población es necesario un triaje exhaustivo.

La punción por aspiración con aguja fina es la prueba no quirúrgica de referencia para el diagnóstico inicial de los nódulos tiroideos. Sin embargo, el 25-30 % de todas las punciones son designadas como indeterminadas y dada que la posibilidad de malignidad, muchos pacientes suelen ser derivados a cirugía. Tras la operación, solo el 25 % de las lesiones resultan malignas, por lo que el riesgo de cáncer no es lo suficientemente alto como para respaldar definitivamente la cirugía. En retrospectiva, estas tiroidectomías son innecesarias y generan complicaciones que los pacientes padecerán durante toda su vida. Por lo tanto, existe una necesidad urgente de desarrollar métodos efectivos para discernir benignidad de malignidad en las lesiones foliculares de tiroides.

El desarrollo de biomarcadores seguros y rentables podría ayudar en la toma de decisiones del médico tratante para manejar con mayor certeza los nódulos tiroideos. Si bien durante mucho tiempo se ha tratado de establecer el valor diagnóstico de las diferentes alteraciones genéticas ninguna parece ser totalmente resolutoria. La metilación del ADN es una de las principales modificaciones epigenéticas y su papel en el inicio, desarrollo y metástasis de cánceres humanos ha sido ampliamente descrito, incluido en cáncer de tiroides. Dado que las firmas de metilación del ADN pueden ser indicativas de la plasticidad molecular que surge en la tumorigénesis temprana, en este estudio se planteó que esta marca epigenética podría funcionar como un marcador para la discriminación en casos difíciles de diagnosticar.

Inicialmente se realizó un análisis mutacional mediante secuenciación del exoma completo en una cohorte de descubrimiento multicéntrica retrospectiva que agrupa 54 muestras de tiroides con patrón folicular. Aunque se encontraron alteraciones propias de los nódulos tiroideos, no pudo establecerse una firma distintiva. Por ello, en la misma cohorte se llevó a cabo un análisis de metilación del ADN empleando la plataforma EPIC de Illumina. Utilizando herramientas de aprendizaje automatizado se desarrolló un modelo de diagnóstico basado en tres sitios CpG (cg17154646, cg21915100 y cg10189462) que posteriormente se validó mediante diferentes aproximaciones ortogonales (bases de datos externas y pirosecuenciación con bisulfito). Cuando se evaluó el rendimiento del clasificador en una cohorte de validación independiente (n = 125) empleando la técnica de pirosecuenciación, se obtuvo una sensibilidad del 96 %, una especificidad de 82 % y unos valores predictivos positivos y negativos del 77 % y 97 %, respectivamente. Curiosamente, usando este modelo, cuando se asignaron puntuaciones de malignidad a los distintos tumores tiroideos, se pudo observar que algunos tumores benignos ya albergaban alteraciones moleculares propias de los carcinomas. Asimismo, se probó el rendimiento en punciones de tiroides (n = 33), clasificando correctamente el 75 % de las muestras que indeterminadas que normalmente son derivadas a cirugía.

Dados los resultados, el sistema de clasificación basado con un conjunto mínimo de biomarcadores epigenéticos puede complementar el potencial de las técnicas diagnósticas actualmente disponibles. Asimismo, la clasificación pronóstica de las muestras en función de las puntuaciones de malignidad priorizaría un número considerable de intervenciones quirúrgicas que a menudo se realizan debido a una citología incierta.

ABSTRACT

Thyroid nodules are the most common encountered thyroid endocrine pathology, being unfrequently diagnosed as malignant (~15 %). The incidence of thyroid cancer has tripled in Spain in the last 25 years. Likewise, it affects three times more women than men and its frequency increases with age, especially after the age of 40. In general, they are tumors with a great prognosis, with survival rates close to 100 %. However, due to the high prevalence of nodules in the population, an exhaustive triage is necessary.

Fine needle aspiration puncture is the *gold standard* non-surgical test for the initial diagnosis of thyroid nodules. However, 25-30 % of all cytologies are designated as indeterminate and given the possibility of malignancy, many patients are often referred for surgery. After the operation, only 25 % of lesions result as cancer, so the risk of malignancy is not high enough to definitively support thyroidectomy. In hindsight, these thyroidectomies are necessary and lead to complications that patients will endure for their entire lives. Therefore, there is an urgent need to develop effective methods to discern benign from malignant follicular-patterned thyroid lesions.

Safe and cost-effective biomarkers could help in the decision-making of the treating physician to manage thyroid nodules with greater certainty. Although many attempts have been made to establish the diagnostic value of the different genetic alterations, none of them seems to be fully resolvable. DNA methylation is one of the main epigenetic modifications and its role in the initiation, development and metastasis of human cancers has been widely described, including thyroid cancer. Since DNA methylation signatures may be indicative of molecular plasticity that arises in early tumorigenesis, this study postulated that this epigenetic mark could be used as a marker for discrimination in difficult-to-diagnose cases.

Initially, mutation analysis was performed by whole-exome sequencing in a retrospective multicenter discovery cohort of 54 follicular-patterned thyroid samples. Although common mutations of thyroid nodules were detected, a distinctive signature could not be established. Therefore, in the same cohort, a DNA methylation analysis was carried out using the Illumina EPIC platform. Using a machine learning pipeline, a diagnostic model based on three CpG sites (cg17154646, cg21915100 and cg10189462) was developed. Subsequently it was validated by different orthogonal approaches (external databases and bisulfite pyrosequencing). Subsequently, the performance of the classifier was evaluated in an independent validation cohort (n = 125) using the pyrosequencing technique, obtaining a sensitivity of 96 %, a specificity of 82 %, and positive and negative predictive values of 77 % and 97 %, respectively. Interestingly, when the same model was applied to assign malignancy scores to the tumours in the validation cohort, it was observed that some thyroid nodules already harbour molecular alterations which were typically seen in carcinomas. Finally, the diagnostic performance in thyroid punctures (n = 33) was tested, our model was capable of correctly classifying 75 % of the indeterminate samples that are normally referred to surgery.

Our classification system based on a minimal set of epigenetic biomarkers could complement the potential of the diagnostic techniques currently available. Likewise, the prognostic classification of the samples based on the malignancy scores would prioritize a considerable number of surgical interventions that are often performed due to an uncertain cytology.

ÍNDICE

ÍNDICE

LISTA DE ABREVIATURAS	I
GLOSARIO DE ANGLICISMOS	V
LISTA DE TABLAS	IX
LISTA DE FIGURAS	XI
INTRODUCCIÓN	1
1 La encrucijada entre el genoma y el fenotipo	3
1.1 Epigenética	4
1.1.1 Modificaciones postraduccionales de histonas	5
1.1.2 Metilación del ADN	8
1.1.3 Impulsores epigenéticos del cáncer	11
2 Glándula tiroides	15
2.1 Anatomía macro- y microscópica general	15
2.2 Biogénesis, regulación y función de las hormonas tiroideas	16
3 Nódulos tiroideos	21
3.1 Epidemiología	21
3.2 Factores de riesgo demográficos y ambientales en cáncer de tiroides	24
3.3 Clasificación histopatológica	27
3.3.1 Hiperplasia nodular	29
3.3.2 Adenoma folicular	30
3.3.3 Carcinomas de tiroides bien diferenciados	30
3.3.4 Tumores tiroideos de potencial maligno incierto	33
3.3.5 Carcinomas desdiferenciados	33
3.4 Alteraciones moleculares en las lesiones foliculares de tiroides	34
3.4.1 Perfil de alteraciones genéticas	35
3.4.2 Perfil de modificaciones epigenéticas	40
4 Evaluación diagnóstica inicial de los nódulos tiroideos	46
4.1 Limitación de las punciones en el diagnóstico exploratorio de los nódulos tiroideos	53
6 Marcadores para el diagnóstico de cáncer de tiroides	55
6.1 Marcadores moleculares de diagnóstico en cáncer de tiroides	57
6.2 Marcadores epigenéticos de diagnóstico en cáncer de tiroides	61
HIPÓTESIS Y OBJETIVOS	65
MATERIAL Y MÉTODOS	69
1 Cohortes de pacientes	71
2 Procesado de las muestras embebidas en el compuesto preservador frente la congelación	73
3 Procesado de las punciones por aspiración con aguja fina	74
4 Procesado de las muestras de las extensiones citológicas	74
5 Extracción del ADN genómico	75
6 Cuantificación y calidad del ADN	76
7 Caracterización del perfil mutacional	77

7.1	Secuenciación del exoma	77
7.2	Determinación de mutaciones somáticas: BRAF, KRAS, NRAS y HRAS	79
8	Caracterización del perfil de metilación	79
8.1	Procesado de los datos brutos	79
8.2	Detección de sondas CpG diferencialmente metiladas	81
9	Desarrollo del clasificador de tumores tiroideos mediante aprendizaje automático	82
10	Validación del rendimiento del clasificador de tumores tiroideos	84
10.1	Validación mediante pirosecuenciación con bisulfito	84
10.1.1	Genes de estudio	84
10.1.2	Diseño de cebadores	84
10.1.3	Modificación del ADN mediante bisulfito sódico	85
10.1.4	Amplificación por Reacción en Cadena de la Polimerasa	86
10.1.5	Pirosecuenciación	87
10.2	Análisis del rendimiento del clasificador mediante bases de datos externas	89
11	Disponibilidad y accesibilidad de los datos	89
RESULTADOS		91
1	Perfil mutacional de las lesiones tiroideas con patrón folicular	93
2	Alteraciones epigenómicas en lesiones tiroideas con patrón folicular	95
3	Desarrollo de un sistema de clasificación de diagnóstico de lesiones tiroideas malignas basado en biomarcadores epigenéticos	99
4	Rendimiento del sistema de clasificación de diagnóstico de lesiones tiroideas malignas mediante bases de datos externas	101
5	Traslación del sistema de clasificación de diagnóstico de lesiones tiroideas malignas a un modelo desarrollado mediante pirosecuenciación con bisulfito	104
6	Puntuaciones de malignidad en los nódulos tiroideos según el sistema mínimo de clasificación epigenético	106
7	Validación del sistema de clasificación epigenética en punciones de nódulos tiroideos	110
DISCUSIÓN		115
1	Las limitaciones de la punción por aspiración con aguja fina en el diagnóstico preoperatorio de los nódulos tiroideos ponen de manifiesto la necesidad de nuevos biomarcadores	117
1.1	Análisis el perfil mutacional de las lesiones nodulares originadas a partir de células foliculares	118
1.2	Evaluación de las alteraciones epigenómicas mediante <i>arrays</i> de metilación del ADN	121
1.3	Identificación nuevos biomarcadores que permitan el desarrollo de un clasificador de tipo diagnóstico mediante técnicas de <i>aprendizaje automático</i>	123
1.4	Discriminación pronóstica de las lesiones foliculares indolentes según los biomarcadores epigenéticos	127
CONCLUSIONES		133
REFERENCIAS		137
ANEXO 1: TABLAS SUPLEMENTARIAS		155
ANEXO 2: FIGURAS SUPLEMENTARIAS		167

LISTA DE ABREVIATURAS

Abreviatura	En español	En inglés
5caC	5-carboxilcitosina	5-carboxylcytosine
5fC	5-formilcitosina	5-formylcytosine
5hmC	5-hidroximetilcitosina	5-hydroxymethylcytosine
5mC	5-metilcitosina	5-methylcytosine
ADN	Ácido desoxiribonucleico	Deoxyribonucleic acid
AECC	Asociación Española Contra el Cáncer	Spanish Association Against Cancer
ARN	Ácido ribonucleico	Ribonucleic acid
ATC	Carcinoma anaplásico de tiroides	Anaplastic thyroid carcinoma
AUC	Área bajo la curva	Area Under the Curve
AUS/FLUS	Atíпия de significado incierto/Lesión folicular de significado incierto	Atypia of Undetermined Significance/Follicular Lesion of Undetermined Significance
BAM	Binary Alignment Map	Binary Alignment Map
BMIQ	Beta-Mixture Quantile	Beta-Mixture Quantile
Bs	Bisulfito Sódico	Sodium bisulphite
CA	Carcinoma maligno	Malignant carcinoma
CIA	Fenol-cloroformo-alcohol isoamílico	Phenol-chloroform-isoamyl alcohol
DIT	Diiodotirosina	Diiodotyrosine
dmCpG	CpG diferencialmente metiladas	Differentially methylated CpG
dNTPs	Deoxinucleótido trifosfato	Deoxynucleotide triphosphate
EGA		European Genome-phenome Archive
EIVFPTC	Variante folicular del carcinoma papilar de tiroides encapsulado invasivo	Encapsulated-infiltrative follicular variant of papillary thyroid carcinoma
ENCODE		Encyclopedia of DNA Elements
FDR	Tasa de descubrimiento falso	False Discovery Rate
FN/FSN	Neoplasia folicular/Sospechoso de neoplasia folicular	Follicular neoplasm/Suspicious for follicular neoplasm
FTA	Adenoma folicular de tiroides	Follicular thyroid adenoma
FTC	Carcinoma folicular de tiroides	Follicular thyroid carcinoma
FT-UMP	Tumor folicular de potencial maligno incierto	Follicular tumor of undetermined malignant potential
fvPTC	Variante folicular del carcinoma papilar de tiroides	Follicular variant of papillary thyroid carcinoma
GEC	Clasificador de expresión de genes de Afirma	Afirma Gene Expression Classifier
GEO	Gene Expression Omnibus	Gene Expression Omnibus
GSC	Clasificador de secuenciación de genes de Afirma	Afirma Gene Sequencing Classifier
HCS	Células madre hematopoyéticas	Hematopoietic stem cells
HTA	Adenoma de células de Hürthle	Hürthle cell adenoma
HTC	Carcinoma de células de Hürthle	Hürthle cell carcinoma
HUCA	Hospital Universitario Central de Asturias	Central University Hospital of Asturias
IARC	Centro Internacional de Investigaciones sobre el Cáncer	The International Agency for Research on Cancer
ICG	Isla CpG	CpG Island

IDAT	Intensity Data Files	Intensity Data Files
IFVPTC	Variante folicular del carcinoma papilar de tiroides invasivo	Invasive follicular variant of papillary thyroid carcinoma
IGV	Integrative Genomics Viewer	Integrative Genomics Viewer
InDel	Inserción y delección	Insertion and deletion
INE	Instituto Nacional de Estadística español	Spanish National Statistics Institute
LB	Células B	B cells
LOD	Limite de detección	Limit of detection
LTC	Células T CD8	CD8 T cells
LTH	Células T CD4	CD4 T cells
M	Monocitos	Monocytes
miFTC	Carcinoma folicular de tiroides minimamente invasivo	Minimally invasive follicular thyroid cancer
MIT	Monoyodotirosina	Monoiodotyrosine
MS	Puntuación de malignidad	Malignancy score
N	Neutrófilo	Neutrophil
NEMII	Neoplasia endocrina múltiple tipo 2	Multiple endocrine neoplasia type 2
NH	Hiperplasia nodular	Nodular hyperplasia
NIFTP	Neoplasia tiroidea folicular no invasiva con características nucleares de tipo papilar	Non-invasive follicular thyroid neoplasm with papillarylike nuclear features
NK	Célula natural asesina	Natural killer cell
OCT	Compuesto preservador frente a la congelación	Optimal Cutting Temperature compound
OMS	Organización mundial de la salud	World Health Organization
PAAF	Punción por aspiración con aguja fina	Fine-needle aspiration cytology
PCR	Reacción en cadena de la polimerasa	Polymerase chain reaction
PDTC	Carcinoma de tiroides pobremente diferenciado	Poorly differentiated thyroid carcinoma
PON	Panel of Normals	Panel of Normals
PPi	Pirofosfato	Pyrophosphate
PTC	Carcinoma papilar de tiroides	Papillary thyroid cancer
REDECAN	Red Española de Registros de Cáncer	Spanish Network of Cancer Registries
RT-qPCR	Reacción en cadena de la polimerasa en tiempo real	Real-time polymerase chain reaction
SAM	Sequence Alignment Map	Sequence Alignment Map
SN	Sospechoso de malignidad	Suspicious for malignancy
SNP	Polimorfismo de un solo nucleotido	Single-Nucleotide Variants
T3	Triyodotironina	Triiodothyronine
T4	Tiroxina	Thyroxine
TBSRTC	Sistema Bethesda para informar la citopatología tiroidea	The Bethesda System for Reporting Thyroid Cytopathology
TCGA	El Atlas del Cáncer	The Cancer Genome Atlas
THCA	Atlas del cáncer de tiroides	Thyroid Cancer Atlas
Tm °C	Temperatura de hibridación	Hybridization temperature
TPTC	Variante de células altas del carcinoma papilar de tiroides	Tall cell variant of papillary thyroid carcinoma
TSS	Sitio de inicio de la transcripción	Transcription start site

UMP	Tumores de potencial maligno incierto o limitrofes	Tumors of undetermined malignant potential or borderline
UTR	Región no traducida	Untranslated Trailer
VCF	Variant Call Format	Variant Call Format
VPN	Valor predictivo negativo	Negative predictive value
VPP	Valor predictivo positivo	Positive predictive value
WDT-UMP	Tumores bien diferenciados de potencial maligno incierto	Well differentiated tumours of undetermined malignant potential
wiFTC	Carcinoma folicular de tiroides ampliamente invasivo	Widely invasive follicular thyroid cancer

GLOSARIO DE ANGLICISMOS

- ❖ **Beta-Mixture Quantile (BMIQ):** un método de normalización empleado para corregir el sesgo del diseño de las sondas en los datos de metilación del ADN de la plataforma Illumina Infinium.
- ❖ **Binary Alignment Map (BAM):** es un formato de datos utilizado para mapear secuencias del genoma. Es la representación binaria de SAM (vease a continuación en esta sección) y mantiene exactamente la misma información que este fichero de texto.
- ❖ **Enhancer:** en castellano se traduce como amplificador o potenciador. Es una región del ADN que puede unirse a factores de transcripción para aumentar los niveles de expresión génica. Esta activación no tiene por qué ocurrir cerca de los genes, ni siquiera en el mismo cromosoma. Esto es así, porque en el proceso de plegamiento del ADN dentro del núcleo, dos regiones distantes pueden llegar a estar próximas geoméricamente.
- ❖ **Frameshift deletion:** la mutación por corrimiento o desfase es una alteración génica causada por la deleción de un número de nucleótidos que no es múltiplo de tres en la secuencia de ADN.
- ❖ **Frameshift insertion:** la mutación por corrimiento o desfase es una alteración génica causada por la inserción de un número de nucleótidos que no es múltiplo de tres en la secuencia de ADN.
- ❖ **Gold standard** (adaptación al español estándar): este término se puede traducir como estándar de oro, patrón de oro o patrón de referencia. En este texto científico hace referencia aquellas pruebas de diagnóstico que tienen máxima fiabilidad a la hora de diagnosticar una patología. Debido a la falta de consenso en la forma de traducir este concepto, utilizaremos su denominación en inglés.
- ❖ **Integrative Genomics Viewer (IGV):** es una herramienta interactiva de alto rendimiento y fácil de usar para la exploración visual de datos genómicos. Admite la integración flexible de todos los tipos comunes de datos y metadatos genómicos, generados por investigadores o disponibles públicamente, cargados desde fuentes locales o en la nube.
- ❖ **Intensity Data Files (IDAT):** los archivos de datos sin procesar de Illumina suelen estar en texto sin formato o en formato binario. En este caso, los archivos binarios "IDAT" son generados por el escáner y se pueden analizar con paquetes R/BioConductor que contiene toda la información necesaria para analizar los datos.
- ❖ **Learn Orientation Model:** hace referencia a un paso específico en el procesamiento de los datos de secuenciación del exoma completo. En este caso es siguiendo el protocolo de buenas prácticas de GATK. La guía paso por paso describe que si sospecha que alguna de sus muestras tiene errores de sustitución que ocurren en una sola hebra antes de la secuenciación, se debe usar el filtro de sesgo de orientación de Mutect2.
- ❖ **Mean Decrease Accuracy:** es un resultado imprescindible del algoritmo *Random Forest*. Expresa cuánta precisión pierde el modelo al excluir cada variable. Cuanto más comprometida se ve la

precisión, más importante es la variable para la clasificación exitosa. Las variables se presentan de importancia decreciente.

- ❖ **Mean Decrease Gini:** la disminución media del coeficiente de Gini es una medida resultante del algoritmo *Random Forest*. Hace referencia a cómo cada variable contribuye a la homogeneidad de los nodos y las hojas en el bosque resultante. Cuanto mayor sea la puntuación, mayor será la importancia de la variable en el modelo.
- ❖ **Open sea:** se refiere a regiones aisladas que no tienen una designación específica.
- ❖ **Ómica:** este neologismo se utiliza como sufijo para referirse al estudio de la totalidad o del conjunto de algo. En este caso, se utiliza para hablar del estudio completo, mediante diferentes técnicas de nueva generación, de las modificaciones epigenéticas que ocurren sobre genoma, transcriptoma y proteoma.
- ❖ **Panel of Normals (PON):** es un tipo de recurso utilizado en el análisis de variantes somáticas. Lo que todos los PON tienen en común es que (1) están generados a partir de muestras normales (en este contexto, "normal" significa derivado de tejido sano que se cree que no tiene alteraciones somáticas) y (2) su propósito principal es capturar artefactos técnicos recurrentes para mejorar los resultados del análisis de variantes somáticas.
- ❖ **Punch:** biopsia punch o biopsia en sacabocados se realiza para extraer un pequeño crecimiento extirpando un cilindro de tejido con un sacabocados.
- ❖ **Random Forest:** la traducción al castellano sería Árboles aleatorios y se puede definir como una técnica de machine learning o aprendizaje automático que se encarga de solucionar problemas relacionados con la regresión y la clasificación. Es un algoritmo en el que los datos de entrenamiento las categorías son conocidas para el investigador.
- ❖ **Rule-in test:** en una prueba molecular significa incluir la posibilidad de una condición médica en el diagnóstico o diagnóstico diferencial.
- ❖ **Rule-out test:** en una prueba molecular significa excluir la posibilidad de una condición médica en el diagnóstico o diagnósticos diferenciales.
- ❖ **Sequence Alignment Map (SAM):** es un formato de texto de alineación genérico para almacenar alineaciones de lectura contra secuencias de referencia, que admite lecturas cortas y largas (hasta 128 Mbp) producidas por diferentes plataformas de secuenciación.
- ❖ **Shelf:** son sitios CpG en el genoma que están hasta 4kb de la isla CpG más cercana.
- ❖ **Shore:** son sitios CpG en el genoma que están hasta 2kb de la isla CpG más cercana.
- ❖ **Splicing:** El empalme o splicing alternativo permite obtener a partir de un transcrito primario de ARNm o pre-ARNm distintas isoformas de ARNm y proteínas, las cuales pueden tener funciones diferentes.

- ❖ **Startloss:** mutación no sinónima es una mutación puntual en el codón de inicio AUG de un transcrito, que provoca la reducción o eliminación de la producción de proteínas.
- ❖ **Stopgain:** mutación no sinónima que da lugar a un codón de terminación prematuro, que señala el final de la traducción y provoca un acortamiento anormal de la proteína.
- ❖ **Variant Call Format (VCF):** es un formato de texto que se usa en Bioinformática para almacenar variantes de una o varias secuencias de genes respecto a un genoma de referencia.

LISTA DE TABLAS

INTRODUCCIÓN

Tabla 1 Resumen de las características clínicas y epigenéticas en las lesiones foliculares de tiroides	34
Tabla 2 Resumen de estudios en cáncer de tiroides utilizando métodos para el análisis de la metilación del ADN en todo el genoma	44
Tabla 3 El Sistema Bethesda 2017 para informar la citopatología tiroidea: riesgo de malignidad (%) y recomendaciones clínicas	48

MATERIAL Y MÉTODOS

Tabla 4 Biobancos pertenecientes a la Red Nacional de Biobancos de España que han participado en el estudio	71
Tabla 5 Datos de pacientes para las cohortes de entrenamiento, validación y punciones de tiroides	72
Tabla 6 Resumen de las características de los genes seleccionados para la validación del clasificador mediante pirosecuenciación con bisulfito	84
Tabla 7 Secuencias de los cebadores directo, inverso y secuenciaciones correspondientes a cada gen	85

ANEXO 1 TABLAS SUPLEMENTARIAS

Tabla S1 Clasificación de tumores tiroideos de la Organización Mundial de la Salud de 2017	157
Tabla S2 Resumen de los resultados del experimento de secuenciación del exoma completo correspondientes a las lesiones tiroideas humanas analizadas en la cohorte de descubrimiento	158
Tabla S3 Información clínico-patológica y puntuaciones moleculares de malignidad calculados usando los valores de metilación del ADN obtenidos del array de metilación y el algoritmo de predicción en la cohorte de descubrimiento	159
Tabla S4 Información clínico-patológica y puntuaciones moleculares de malignidad calculados usando los valores de metilación del ADN obtenidos del array de metilación y el algoritmo de predicción en la cohorte de validación externa	160
Tabla S5 Información clínico-patológica y puntuaciones moleculares de malignidad calculados usando los valores de metilación del ADN obtenidos mediante pirosecuenciación con bisulfito y el algoritmo de predicción en la cohorte de descubrimiento	161
Tabla S6 Información clínico-patológica y puntuaciones moleculares de malignidad calculados usando los valores de metilación del ADN obtenidos mediante pirosecuenciación con bisulfito y el algoritmo de predicción en la cohorte de validación	163
Tabla S7 Información clínico-patológica y puntuaciones moleculares de malignidad calculados usando los valores de metilación del ADN obtenidos mediante pirosecuenciación con bisulfito y el algoritmo de predicción en la cohorte de PAAF	166

LISTA DE FIGURAS

INTRODUCCIÓN

Figura 1 La creciente complejidad del dogma central de la biología molecular. Organización, compartimentalización y modificaciones que ocurren sobre el ADN en el núcleo	5
Figura 2 Representación esquemática del nucleosoma	6
Figura 3 Metilación de citosinas	8
Figura 4 Proceso de metilación y desmetilación del ADN	10
Figura 5 Representación de las distintas características comunes a los distintos tipos de cáncer	12
Figura 6 Esquema de los patrones de metilación del ADN en células normales y tumorales	13
Figura 7 Localización (a), anatomía (b) e histología (b) de la glándula tiroidea normofuncionante en el cuerpo humano	16
Figura 8 Descripción general del eje hipotálamo-hipófisis-tiroides, las influencias directas e interacciones de retroalimentación entre los distintos componentes	17
Figura 9 Representación esquemática de los aspectos clave del transporte de yodo tiroideo y la síntesis de hormona tiroidea en los tirocitos altamente polarizados	18
Figura 10 Fisiología y estructura de las hormonas tiroideas	19
Figura 11 Incidencia de cáncer de tiroides en el mundo. La tasa se presenta como nuevos casos por cada 100.000 habitantes	22
Figura 12 Representación demográfica de las tasas de incidencia (a), prevalencia (b) y mortalidad (c) del cáncer de tiroides en España por cada 100.000 habitantes	23
Figura 13 Factores de riesgo que podrían influenciar el desarrollo de cáncer de tiroides en la población	24
Figura 14 Esquema representativo de la clasificación de los tumores tiroideos referenciados a lo largo de este trabajo	29
Figura 15 Vías MAPK y PI3K/AKT/mTOR en el cáncer de tiroides versus en condiciones normales	35
Figura 16 Resumen de las mutaciones más frecuentes de neoplasias derivadas de células foliculares	36
Figura 17 Ejemplos de punciones realizadas por Martin y Ellis en 1930	47
Figura 18 Imágenes representativas de PAAF con clasificación Bethesda II	49
Figura 19 Imágenes representativas de PAAF con clasificación Bethesda III	50
Figura 20 Panel superior: Imágenes representativas de PAAF con clasificación Bethesda IV. Panel inferior: Imágenes representativas de PAAF con clasificación Bethesda V	51
Figura 21 Imágenes representativa de PAAF con clasificación Bethesda VI	52
Figura 22 Imagen representativa de la transición desde nódulo hiperplásico, a adenoma y finalmente carcinoma de tiroides	54
Figura 23 Resumen del rendimiento diagnóstico para los clasificadores moleculares de nódulos tiroideos con citología indeterminada disponibles en el mercado	57

MATERIAL Y MÉTODOS

Figura 24 Resumen del proceso de extracción de ADN genómico según los distintos tipos de muestras biológicas	75
Figura 25 Información adicional relacionada con el protocolo seguido para la caracterización mutacional mediante secuenciación del exoma en las muestras de la cohorte de descubrimiento	77
Figura 26 Esquema representativo del procedimiento llevado a cabo para realizar la caracterización del perfil de metilación mediante el experimento con Infinium HumanMethylationEPIC 850K Beadchip y la generación del clasificador empleando técnicas de aprendizaje automático	80

Figura 27 Evaluación del rendimiento del sistema de clasificación Random Forest utilizando diferentes tamaños de muestra	83
Figura 28 Esquema del proceso de modificación del ADN genómico mediante bisulfito sódico	85
Figura 29 Esquema del proceso llevado a cabo para realizar la amplificación de los genes seleccionados mediante PCR	87
Figura 30 Esquema del funcionamiento de la pirosecuenciación de ADN genómico modificado con bisulfito	88

RESULTADOS

Figura 31 La secuenciación del exoma identifica alteraciones genéticas en lesiones tiroideas foliculares benignas y malignas	93
Figura 32 Representación detallada de los genes mutados identificados en este estudio	95
Figura 33 Identificación de patrones aberrantes de metilación del ADN en los tumores tiroideos de origen folicular en la cohorte de descubrimiento	96
Figura 34 Metilación global del ADN, sitios CpG diferencialmente metiladas y ubicación en el genoma en las muestras analizadas en la cohorte de descubrimiento mediante arrays de metilación	97
Figura 35 Identificación de patrones aberrantes de metilación del ADN en el cáncer folicular de tiroides con respecto al tejido normal	99
Figura 36 Generación del modelo de clasificación de lesiones tiroideas mediante 12845 dmCpG utilizando el algoritmo Random Forest	100
Figura 37 Clasificador epigenético de tres CpG discrimina entre nódulos tiroideos benignos y malignos en la cohorte de descubrimiento	101
Figura 38 Clasificador epigenético de tres CpG (cg17154646, cg21915100 y cg10189462) discrimina entre nódulos tiroideos benignos y malignos en la cohorte de validación GSE121377	102
Figura 39 Generación del modelo de clasificación de lesiones tiroideas mediante las sondas cg10705422, cg17707274 y cg26849382 utilizando el algoritmo Random Forest	103
Figura 40 Validación del clasificador epigenético de tres CpG (cg17154646, cg21915100 y cg10189462) mediante la técnica de pirosecuenciación con bisulfito	105
Figura 41 Puntuaciones de malignidad inferidas para cada nódulo tiroideo a través del clasificador desarrollado con 3 CpG	108
Figura 42 Puntuaciones de malignidad inferidas para la cohorte de validación de tumores papilares y oncocíticos a través del clasificador desarrollado con 3 CpG	109
Figura 43 Puntuaciones de malignidad inferidas para cada PAAF a través del clasificador desarrollado con 3 CpG	111
Figura 44 Ejemplos del rendimiento del clasificador en PAAFs de la cohorte analizada	112

DISCUSIÓN

Figura 45 Resumen del rendimiento de los clasificadores diagnósticos encontrados en la bibliografía en comparación con los desarrollados por nuestro laboratorio	124
Figura 46 Cambio de paradigma en el diagnóstico de los nódulos tiroideos	130

ANEXO 2: FIGURAS SUPLEMENTARIAS

Figura S1 Información adicional relacionada con el experimento de secuenciación del exoma	169
Figura S2 Niveles de metilación de las 3 CpG (cg17154646, cg21915100 y cg10189462) en NIFTP con respecto a los nódulos no tumorales	170

Figura S3 Correlación entre los valores de metilación mediante pirosecuenciación con bisulfito y los arrays de metilación para cada una de las 3 CpGs (cg17154646, cg21915100 y cg10189462) que componen el clasificador de lesiones tiroideas de la cohorte de validación	170
Figura S4 Puntuaciones de malignidad del conjunto de datos de descubrimiento utilizando arrays de metilación EPIC	171
Figura S5 Diferencias estadísticas entre los niveles de metilación en las muestras NH y FTA de las cohortes de descubrimiento y validación	171
Figura S6 Comparación de la calidad de las muestras tras la extracción de material genético a partir de PAAF (n = 24), citologías de archivo (n = 5) o la combinación de ambos (n = 23)	172
Figura S7 Discriminación de PAAF mediante el clasificador epigenético	172

INTRODUCCIÓN

1 La encrucijada entre el genoma y el fenotipo

El ciclo de vida de los mamíferos comienza con la fertilización del ovocito que prolifera rápidamente para dar lugar a todo el organismo conformado por millones de células. Durante el desarrollo embrionario las células se diversifican en más de 200 clases distintas [1]. Esta diferenciación desde una célula totipotente a una somática es el resultado de la expresión de un subconjunto de genes que es específico para cada tipo celular, y determinará su función en el organismo [2]. Todas ellas son unidades indispensables para la vida y contienen en su interior el mismo material genético, con la información necesaria para dar lugar a toda la complejidad biológica del organismo. El llamado “libro de la vida” o genoma, se encuentra dentro de la membrana nuclear de las células eucariotas en forma de moléculas conocidas como ácido desoxirribonucleico (ADN). La estabilidad de los tejidos, su crecimiento controlado y su eventual renovación dependerán de que cada célula diferenciada sea capaz de recordar su identidad y de transmitirla correctamente a sus células hijas, todo ello sin cambios en la secuencia de ADN.

El genoma contiene 3.200 millones de pares de bases de ADN específicos para cada individuo (genotipo), y albergan unos 20.000 genes que codifican las proteínas que conferirán al organismo unas características morfológicas, fisiológicas y de comportamiento determinadas (fenotipo) [3]. Si una célula transcribe 10.000 genes a ácido ribonucleico (ARN), dará como resultado la producción de, al menos, la misma cantidad de proteínas (que tienen, aproximadamente, un número de copias acumulado de 10^9 - 10^{11} moléculas) [3]. Aunque impresionantes, estos números no logran demostrar la verdadera complejidad de un organismo.

Entonces, ¿cómo es posible que cada célula tenga un fenotipo diferente, si todas las células comparten la misma información genética? El dogma central de la biología representa la transferencia directa de información genética del ADN al ARN y a las proteínas, y sigue siendo la base para comprender la función genómica. Sin embargo, la célula representa un entorno dinámico en el que la transcripción y traducción de estas macromoléculas biológicas pueden modificarse a través de innumerables eventos de procesamiento que cambian la funcionalidad [4, 5]. Se trata de una coreografía compleja en la que intervienen múltiples elementos regulatorios en todos los puntos del proceso

[5]. Todos ellos contribuyen en gran medida a la diversidad que encontramos en el organismo. En particular, en esta tesis se prestará especial atención a aquellos mecanismos que se necesitan para la expresión selectiva de los genes en función del contexto y que en general se conocen como regulación epigenética.

1.1 Epigenética

El concepto de epigenética fue originalmente definido en 1942 por Conrad Hal Waddington como el estudio de los mecanismos causales que intervienen entre el genotipo y el fenotipo [6]. Semánticamente, el término significa "sobre la genética" y se refiere a todos los mecanismos moleculares por los cuales el ADN, el ARN y las proteínas, se modifican químicamente, lo que permite la transformación de un genoma en cientos de fenotipos y funcionalidades diferentes, sin que se cambie la secuencia primaria del genoma. Un cambio que es específico para cada tipo de célula y etapa de desarrollo [7, 8].

Si bien las marcas epigenéticas pueden permanecer estables y transmitirse a las generaciones posteriores, también pueden modificarse en respuesta a estímulos ambientales endógenos y exógenos [8, 9]. La importancia de estas marcas en el establecimiento de la identidad tisular durante el desarrollo embrionario, la reparación del daño genómico y el envejecimiento ha sido ampliamente descrita [10–12]. Por ello, es lógico pensar que las alteraciones en estas modificaciones a través de la maquinaria epigenética conllevan a la activación o inhibición inapropiada de distintos genes, que en muchos casos tiene como consecuencia el desarrollo y progreso de diferentes enfermedades.

Así como el doma central de la biología se centra en el paso consecutivo del genoma, al transcriptoma y finalmente al proteoma, la regulación epigenética puede dividirse en tres grandes bloques ómicos¹: epigenómica; epitranscriptómica y epiproteómica [4]. De ellos, los mecanismos más estudiados son los que ocurren sobre el genoma e incluyen la modificación química covalente del ADN (metilación e

¹ El neologismo ómica se utiliza como sufijo para referirse al estudio de la totalidad o del conjunto de algo. En este caso, se utiliza para hablar del estudio completo, mediante diferentes técnicas de nueva generación, de las modificaciones epigenéticas que ocurren sobre genoma, transcriptoma y proteoma.

hidroximetilación) y la cromatina (modificaciones de histonas covalentes y la disposición espacial en el núcleo de la cromatina), que en última instancia están relacionados con la regulación de la expresión génica (Figura 1).

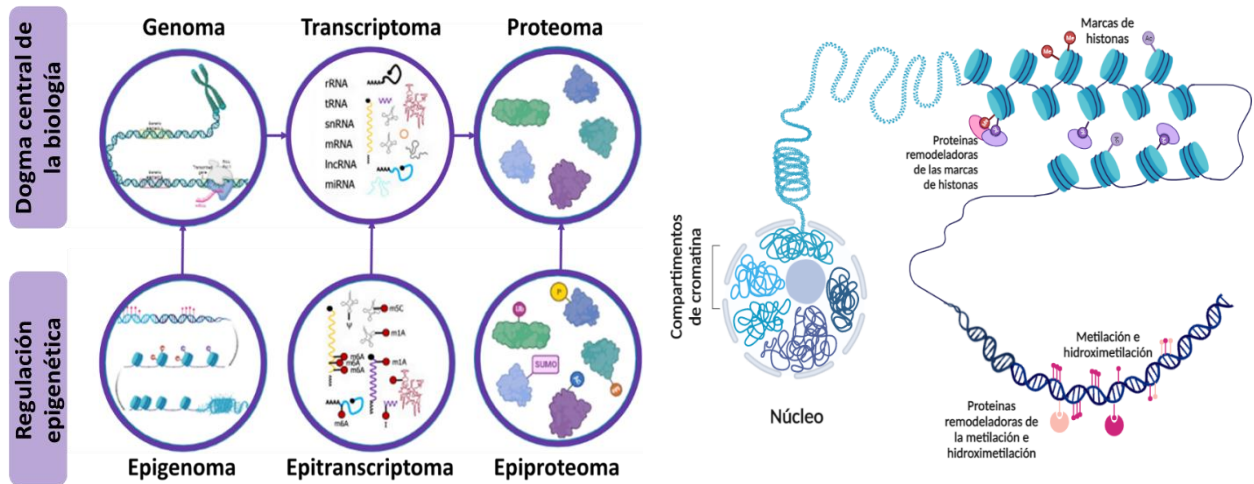


Figura 1 | La creciente complejidad del dogma central de la biología molecular. El flujo de información genómica desde el ADN al ARN y las proteínas sigue siendo la base para comprender la función genómica. Sin embargo, las células son sistemas dinámicos en los que tanto genoma, como transcriptoma como proteoma están sujetos a modificaciones, entre ellas la regulación epigenética. **Organización, compartimentalización y modificaciones que ocurren sobre el ADN en el núcleo.** El elemento estructural básico del ADN es el nucleosoma, compuesto por un octámero de histonas cuya función es la de empaquetar el genoma, organizarlo de forma que este contenido en su totalidad dentro del núcleo. Asimismo, tanto las colas de las histonas como el propio ADN son susceptibles de ser modificados por distintas marcas epigenéticas, y esto supondrá un impacto sobre la expresión de los genes. Imagen creada con BioRender.

1.1.1 Modificaciones postraduccionales de histonas

El genoma está compartimentado espacialmente dentro del núcleo, no es aleatorio y se sabe que esto juega un papel importante en el control transcripcional de los genes [13]. El plegamiento permite la interacción entre los promotores y los elementos reguladores en cis [14, 15] y, del mismo modo, la posición de los genes en el espacio tridimensional del núcleo se correlaciona con los patrones de expresión y la función de los mismos [16]. Las histonas proteínas conservadas evolutivamente juegan un papel importante en el empaquetamiento, organización y regulación del ADN dentro del núcleo de todos los eucariotas [17].

Así, el elemento estructural básico de la compactación del ADN, el nucleosoma, está compuesto por una secuencia de 146 pares de bases de nucleótidos que se pliegan alrededor de un octámero de proteínas que comprende dos copias de cada una de las

histonas centrales: H2A; H2B; H3 y H4 (**Figura 2**). Además, la histona H1 no forma parte de la estructura en "cuenta" de los nucleosomas, sino que une el nucleosoma con el ADN extranucleosómico para establecer la estructura de la cromatina.

Una característica estructural esencial de las histonas son las colas amino-terminales que se extienden más allá del octámero de proteína de ADN, lo que hace que sus residuos de aminoácidos sean accesibles para la modificación (**Figura 2**) [18]. Se han detectado una serie de variaciones en los residuos de histonas y se han descrito al menos ocho tipos diferentes de cambios, entre ellos destacan acetilación, metilación, fosforilación y ubiquitinación [19]. Además, cabe destacar que las modificaciones que ocurren sobre estas proteínas no se producen de forma aislada, sino que existe una gran variedad de marcas postraduccionales y muchas combinaciones posibles. El equilibrio en la interacción entre las histonas modificadas y el ADN puede regular muchos eventos biológicos, incluida la expresión génica, la reparación del ADN, la compactación de la cromatina y la estabilidad genómica, así como importantes procesos genéticos como la inactivación del cromosoma X [7, 20, 21].

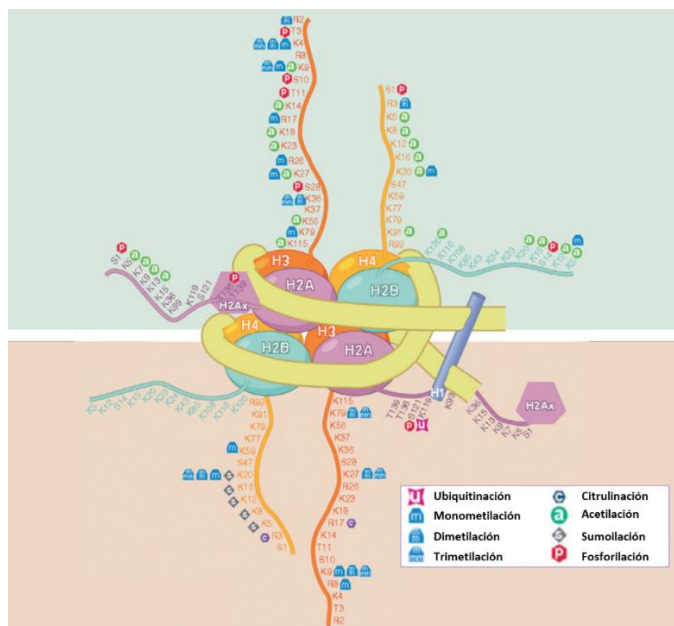


Figura 2| Representación esquemática del nucleosoma. El ADN se envuelve alrededor del octámero de histonas de las cuatro histonas centrales H2A, H2B, H3 y H4. La histona H1, la proteína enlazadora, se une al ADN entre los nucleosomas. Los diferentes aminoácidos que constituyen las colas de las histonas se representan junto con las diferentes modificaciones covalentes específicas de cada residuo. Las marcas activas se representan en la parte superior de la figura y las marcas represivas se representan en la parte inferior de la figura. Residuos: lisina, K; arginina, R; serina, S; y treonina, T; Marcas: acetilación, a; citrulinación, c; dimetilación, mm; fosforilación, p; monometilación, m; sumoilación, s; trimetilación, mmm; y ubiquitinación, u. Imagen tomada y adaptada de Sawan y Herceg, 2010 [22].

Como ya se ha mencionado una de las principales funciones de las modificaciones que ocurren sobre las colas de las histonas que componen la cromatina es la organización del genoma. El código de marcas postraduccionales puede establecer diferentes “ambientes”, es decir, cromatina con un bajo estado de condensación (eucromatina), que es más “accesible” a los factores de transcripción, o cromatina que presenta un alto grado de compactación (heterocromatina), que, a diferencia de la eucromatina, impide la transcripción de genes. Dependiendo del tipo de modificación postraduccional, el efecto sobre la conformación de la cromatina y, por tanto, sobre la regulación de la expresión génica, es diferente. Las modificaciones de la cola de histonas se establecen o borran por la acción catalítica de los sistemas enzimáticos asociados con la cromatina. En general, las enzimas que realizan estas modificaciones forman parte de complejos multiproteicos implicados en la regulación de la transcripción u otros procesos genómicos. Como tales, afectan directamente a la configuración de la cromatina a través de la interacción del ADN y las histonas, o bien constituyen señales que son reconocidas por otros complejos [23].

En general, la adición de grupos acetilo a las lisinas corresponde a una conformación más abierta de la cromatina y, por tanto, a un aumento de la transcripción. Esto se debe a que la acetilación cancela la carga positiva de la lisina. La metilación de la lisina, por otro lado, no altera la carga, por lo que cualquier efecto directo sobre el plegamiento de la cromatina tendría que ocurrir a través de mecanismos no electrostáticos [24]. La acetilación es la modificación de histonas más estudiada y generalmente está asociada con la transcripción activa del gen y es realizada por enzimas histona acetiltransferasa (HAT) que acetilan residuos de lisina específicos en el sustrato de histona y son revertidas por la acción de las histonas desacetilasas (HDAC) [25]. Los activadores unidos al ADN reclutan HAT para acetilar histonas nucleosómicas, mientras que los represores reclutan HDAC para desacetilar histonas. Se ha demostrado que otros coactivadores y correpresores poseen actividad HAT o HDAC o se asocian con dichas enzimas [26, 27]. Se han descrito dos clases generales de enzimas remodeladoras: las histonas metiltransferasas (HMT) actúan para agregar grupos metilo a los residuos de lisina y/o arginina en las histonas, mientras que otro grupo de enzimas, las histonas desmetilasas (HDM), eliminan la metilación [28]. Un nivel adicional de complejidad reside en la posible existencia de múltiples estados metilados en cada residuo. Las lisinas

pueden estar mono, di o trimetiladas, mientras que las argininas solo pueden estar mono o dimetiladas [24].

1.1.2 Metilación del ADN

La metilación del ADN es una modificación covalente post-replicación que no interfiere estéricamente con el apareamiento de bases, aunque puede tener consecuencias importantes en la regulación de la expresión génica. En mamíferos consiste en la adición de un grupo metilo al anillo aromático de una sola base de ADN, principalmente citosinas que preceden a las guaninas donde genera 5-metilcitosina (5mC) (**Figura 3**).

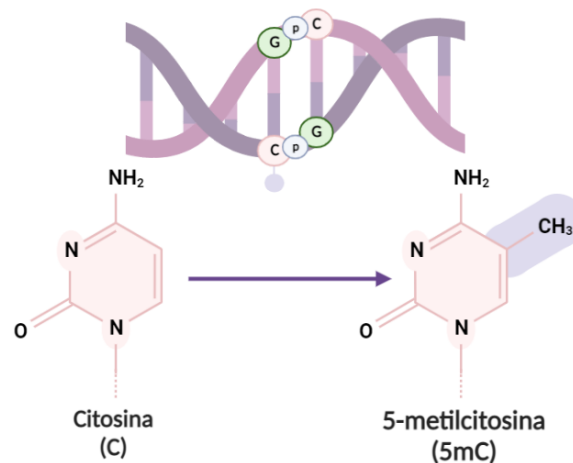


Figura 3 | Metilación de citosinas. La metilación de citosinas que preceden a las guaninas mediante la adición de un grupo metilo (-CH₃) al carbono 5 del anillo pirimidínico da lugar a la 5-metilcitosina.

Estos sitios de dinucleótidos suelen denominarse sitios CpG [29]. A pesar de que la estimación de sitios CpG en el genoma es de 4,4 % (dado el 42 % son GpCs) en realidad suponen el 1 % [30]. Posiblemente esto sea así debido a la que los dinucleótidos CpG metilados son propensos a las desaminaciones espontáneas mutando a timina y, por lo tanto, pasan a ser dinucleótidos TpG [31]. Si bien es conocido que la distribución del dinucleótido CpG es desigual a lo largo de la secuencia del genoma, gran parte de estos sitios se encuentran metilado [32]. Por un lado, el 45 % se encuentra metilado en los transposones y elementos virales [31]. Por otro, existen áreas ricas en CpGs, entre las que encontramos islas CpG (ICG), que se cree que normalmente no están metiladas y están asociadas con el 70 % de los promotores génicos en humanos [31]. El papel de la 5mC en las regiones promotoras está bien establecido, asociándose la hipermetilación

con la represión génica. Las regiones llamadas orillas CpG (del inglés, *CpG shores*), ubicadas a una distancia de hasta 2 kb de las islas, tienen patrones altamente conservados de metilación específica de tejido y su metilación se asocia con una expresión génica reducida [31]. Asimismo, la metilación del ADN aguas abajo del inicio de la transcripción, en la región del primer exón, está más estrechamente relacionada con el silenciamiento transcripcional que la que ocurre en la región promotora [33]. La metilación de primer intrón también muestra una asociación inversa con la expresión génica independientemente del tejido y la especie [34]. Por último, el papel que desempeña en las regiones intragénicas sigue siendo poco conocido. Trabajos recientes indican que la metilación en regiones intragénicas también podría estar involucrada en la regulación de múltiples procesos, incluida la elongación de la transcripción, la expresión de regiones intragénicas tanto codificantes como no codificantes, procesamiento o *splicing* alternativo² y activación de *enhancers* o amplificadores³ [12, 35].

La metilación del genoma es catalizada por la familia de enzimas conocidas como ADN metiltransferasas (DNMT), que transfieren un grupo metilo de la S-adenosil-metionina al carbono 5 de las citosinas (**Figura 4**). Se han identificado cinco isoformas de estas proteínas en mamíferos: DNMT1, DNMT2, DNMT3a, DNMT3b y DNMT3L. La metiltransferasa de mantenimiento, DNMT1, reconoce específicamente el ADN hemimetilado, copia el patrón de metilación y lo conserva después de cada replicación. DNMT3a y 3b son metiltransferasas *de novo*, capaces de establecer nuevos patrones de metilación en hebras de ADN que previamente no estaban metiladas, siendo incapaces de distinguir entre hemimetiladas y no metiladas. Se acumulan en gran medida en las células madre embrionarias, pero va disminuyendo a medida que las células se diferencian. DNMT3L, aunque no es catalíticamente activa, juega un papel importante en el establecimiento de la impronta genómica durante la gametogénesis. DNMT2 no

² El empalme o *splicing* alternativo permite obtener a partir de un transcrito primario de ARNm o pre-ARNm distintas isoformas de ARNm y proteínas, las cuales pueden tener funciones diferentes.

³ Un *enhancer* o amplificador, es una región del ADN que puede unirse a factores de transcripción para aumentar los niveles de expresión génica. Esta activación no tiene por qué ocurrir cerca de los genes, ni siquiera en el mismo cromosoma. Esto es así, porque en el proceso de plegamiento del ADN dentro del núcleo, dos regiones distantes pueden llegar a estar próximas geoméricamente.

exhibe actividad de ADN metiltransferasa, aunque hay evidencia de que puede participar en la adición covalente de grupos metilo para transferir ARN [36, 37].

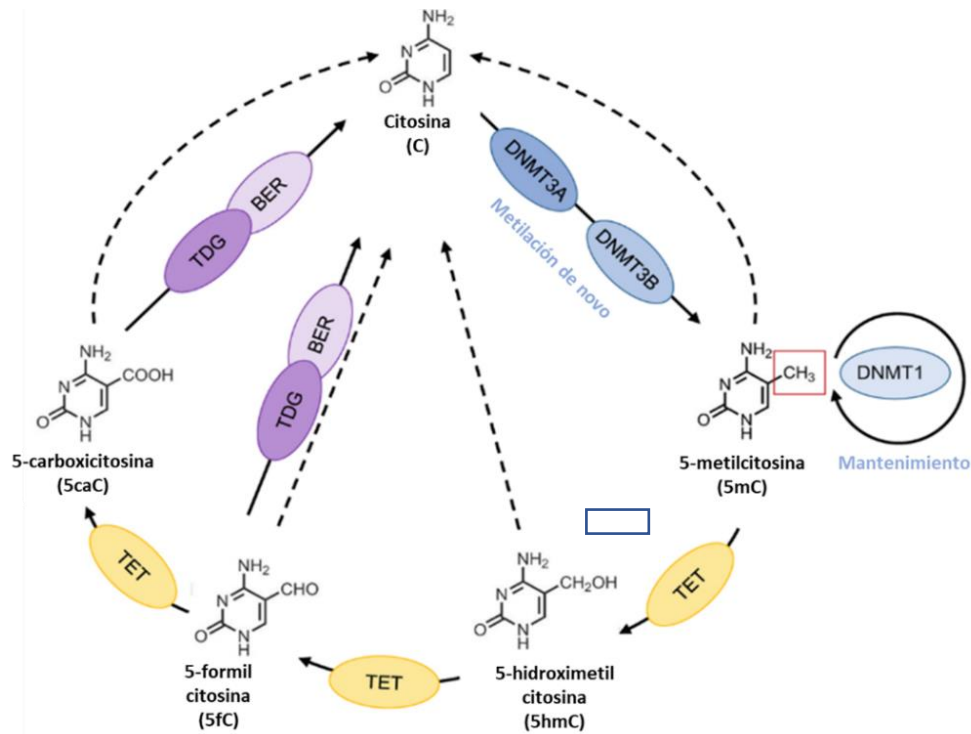


Figura 4 | Proceso de metilación y desmetilación del ADN. Las ADN metiltransferasas (DNMT) catalizan la metilación de la citosina al agregar un grupo metilo en el carbono 5. DNMT3A y DNMT3B son responsables de la metilación de novo, mientras que DNMT1 mantiene los patrones de metilación del ADN después de la replicación del ADN. La desmetilación del ADN puede ocurrir a través de diferentes mecanismos: (1) desmetilación pasiva debido al deterioro de la maquinaria de mantenimiento de la metilación del ADN que da como resultado la pérdida progresiva de la metilación del ADN después de múltiples rondas de replicación; y (2) desmetilación activa que ocurre a través de la oxidación iterativa de 5mC a 5-hidroximetilcitosina (5hmC), 5-formilcitosina (5fC) y 5-carboxilcitosina (5caC) mediada por proteínas de translocación diez-once (TET). Luego, estas formas oxidadas se pueden diluir posteriormente durante la replicación del ADN, o 5fC y 5caC se pueden escindir mediante glicosilasas de ADN de timina (TDG) junto con el mecanismo de reparación por escisión de bases (BER). Aunque se han propuesto otros mecanismos de desmetilación activa (rayas discontinuas), el mecanismo TDG-BER ha obtenido el mayor apoyo. Imagen tomada y adaptada al castellano de Zafón y colaboradores (2019) [38].

Por otro lado, la desmetilación del ADN es un proceso que, en los últimos años, ha demostrado ser un importante regulador de la transcripción (**Figura 4**). Hay dos formas posibles de que esto suceda: (1) pasivamente debido a la pérdida gradual de la actividad de mantenimiento de la DNMT, un proceso que aumenta con replications adicionales (y el tiempo), o (2) la desmetilación activa del ADN, que es catalizada por las proteínas de translocación diez-once (del inglés *Ten-eleven-translocation proteins*; TET, y las isoformas 1, 2 y 3), que permite que las 5mC se transformen en 5-hidroximetilcitosina (5hmC), en un proceso que es independiente de la replicación [16,

17]. Asimismo, varios estudios han planteado la hipótesis de que la 5hmC podría desempeñar un papel activo en la regulación de la expresión génica y contribuir a la regulación epigenética específica de tejido [39–41]. Se sabe que 5hmC se encuentra en promotores con un nivel medio-bajo de islas CpG y que está ausente en los sitios de inicio de la transcripción (TSS) [42, 43]. También se encuentra en regiones con una alta densidad de ICG en cuerpos de genes, preferentemente en exones, donde se correlaciona con la expresión génica activa [44, 45].

La familia de enzimas TET no solo actúa sobre 5mC, sino que también puede oxidar 5hmC, transformándolo sucesivamente en 5-formilcitosina (5fC) y 5-carboxicitosina (5caC), siendo esta última convertida posteriormente en citosina por la tiamina-ADN-glicosilasa (TDG) [46]. Tanto la 5mC como la 5hmC actúan como sustratos en la reacción de desaminación llevada a cabo por la familia de deaminasas citidina desaminasa inducida por activación (AID) el complejo de edición de apolipoproteína B (APOBEC), que las transforman en timina o 5-hidroximetiluracilo, respectivamente [47]. Se ha postulado que ambos productos de esta reacción podrían eventualmente convertirse en citoquinas por la timina-ADN glicosilasa (TDG) o por el mecanismo de reparación por escisión de base (BER) [47].

1.1.3 Impulsores epigenéticos del cáncer

El cáncer es un conjunto de enfermedades heterogéneas y complejas. Un tumor se puede originar a partir de cualquier tejido y están compuestos por múltiples tipos celulares que interaccionan entre sí, además de ser capaces de reclutar células normales que participan activamente en el proceso de malignidad. Progresivamente las células, en un principio sanas, van evolucionando a un estado neoplásico, adquiriendo una serie de características y capacidades distintivas debido a distintos factores genéticos, ambientales, clínicos y de estilo de vida que contribuyen al inicio del tumor y la progresión maligna.

Hace 22 años Weinberg y Hanahan, establecieron las características distintivas del cáncer, influenciados por el reconocimiento de que los tumores malignos humanos se desarrollan como productos de un proceso progresivo en el tiempo, y sembraron los cimientos para comprender los procesos que ocurren a nivel del fenotipo celular [48–50]. A saber, la capacidad adquirida para mantener la señalización proliferativa; evasión

INTRODUCCIÓN

de las señales que suprimen la proliferación incontrolada; resistencia a la muerte celular; inmortalidad replicativa; inducción a la angiogénesis; invasión y la metástasis; reprogramación del metabolismo celular; y evasión del sistema inmunitario (Figura 5) [48, 49].

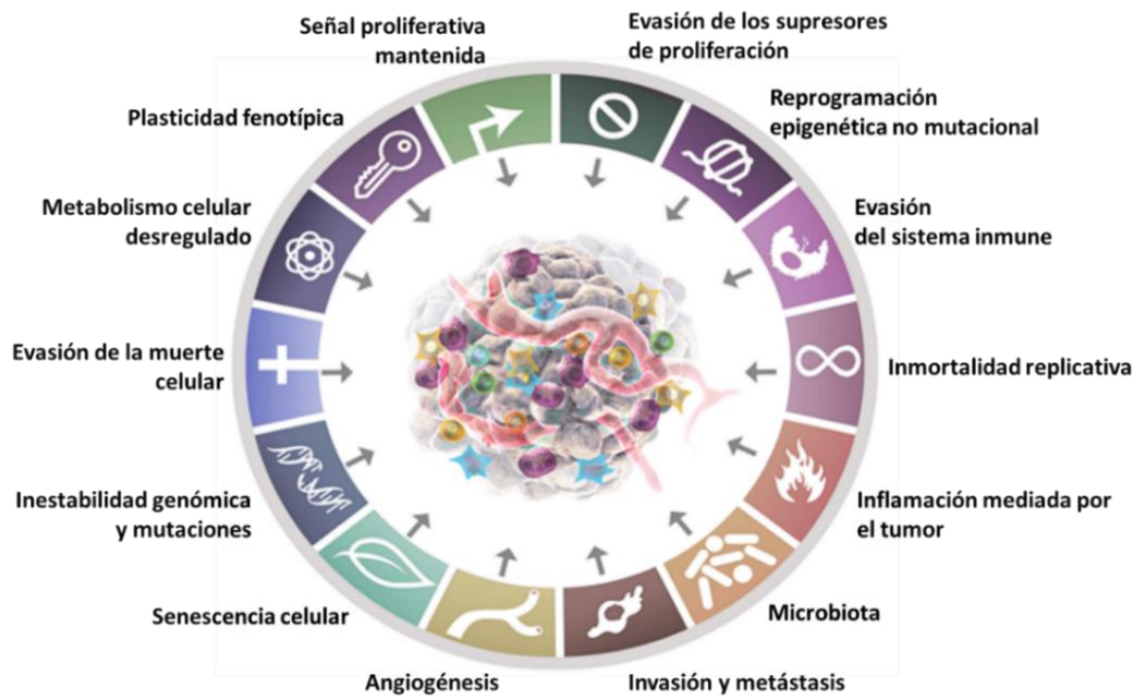


Figura 5 | Representación de las distintas características comunes a los distintos tipos de cáncer. Los sellos distintivos del cáncer actualmente incorporan ocho capacidades distintas y dos características habilitadoras. Capacidades distintas: la capacidad adquirida para mantener la señalización proliferativa; evasión de las señales que suprimen la proliferación incontrolada; resistencia a la muerte celular; inmortalidad replicativa; inducción a la angiogénesis; invasión y la metástasis; reprogramación del metabolismo celular; y evasión del sistema inmunitario. Capacidades habilitadoras: inestabilidad genómica e inflamación mediada por el tumor. Hanahan propone la adición de la plasticidad fenotípica como nueva capacidad distinta; y el microbioma, la reprogramación epigenética no mutacional y la senescencia celular como nuevas capacidades habilitadoras. Figura tomada y adaptada de Hanahan 2022 [50].

Sin embargo, los procesos tumorales son complicados, y estas características en sí mismas no pueden explicar el desarrollo de las neoplasias hacia tumores malignos. Por ello, se incorporó el concepto de características habilitadoras que son consecuencia inherente de la condición y reflejan los mecanismos moleculares y celulares mediante los cuales se adquieren estos rasgos funcionales [50]. A saber, las dos características ya establecidas son la inestabilidad genómica, y la inflamación mediada por el tumor, mientras que el microbioma, la senescencia celular y la reprogramación epigenética no mutacional se plantean como nuevas incorporaciones a la lista [50].

Si bien la investigación en torno a las causas moleculares subyacentes al desarrollo tumoral se ha centrado tradicionalmente en el análisis mutacional a nivel del genoma, las modificaciones epigenéticas también juegan un papel crucial [50], de forma que ambos procesos mantienen una estrecha relación.

Las marcas epigenéticas constituyen señales moleculares que se pueden heredar a través de la división celular. Sin embargo, una de las características que también las define es su carácter dinámico, ya que existe una amplia gama de modificadores epigenéticos que pueden agregar o eliminar marcas de forma activa [51]. Estas enzimas, si bien son dianas para la terapia farmacológica en diversos entornos, como el cáncer, también están sujetas a la influencia ambiental, lo que convierte a los mecanismos epigenéticos en un sistema crucial de comunicación entre el genoma y los estímulos externos [52].

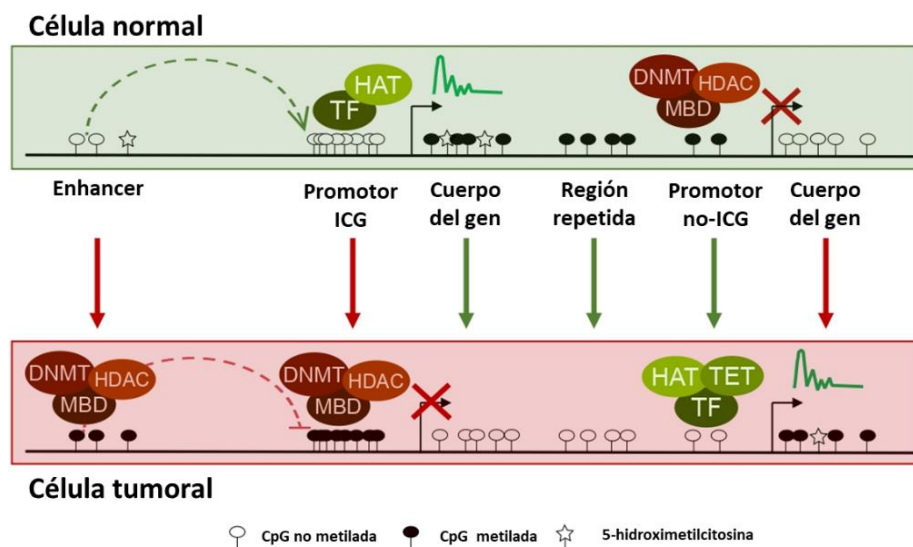


Figura 6 | Esquema de los patrones de metilación del ADN en células normales y tumorales. La mayoría de los CpG en el genoma humano están metilados, excepto los CpG ubicados en regiones ricas en CpG llamadas islas CpG (ICG) y en promotores y potenciadores pobres en CpG. Los CpG no metilados en promotores y enhancer permiten la unión de factores de transcripción (TF) apropiados y el reclutamiento de modificadores de ADN y de histonas asociados con la activación transcripcional (como histona acetilasas (HAT) y proteínas de translocación Ten-Eleven (TET)). Por el contrario, los sitios CpG metilados están unidos a proteínas de unión a metil-CpG (MBD) que reclutan modificadores de histonas y ADN asociados con la represión génica (como las histonas desacetilasas (HDAC) y las ADN metiltransferasas (DNMT)). Las repeticiones son silenciadas por metilación. Algunos CpG en elementos activos están hidroximetilados. Las alteraciones en estos patrones, incluidas las hipermetilaciones (generalmente en supresores tumorales) y las hipometilaciones (global y asociada a genes específicos como oncogenes), son un sello distintivo del cáncer. El escenario típico se representa en la figura superior y las alteraciones asociadas al cáncer se muestran en la figura inferior. Figura tomada y adaptada al castellano de Zafón y colaboradores (2019) [38].

La metilación del ADN ha sido, con diferencia, la marca epigenética estudiada en mayor profundidad en el cáncer (**Figura 6**). Curiosamente, tanto los eventos de hipometilación como los de hipermetilación se pueden observar en los procesos oncogénicos [53, 54]. La hipometilación global fue la primera alteración epigenética que se reconoció en los tumores humanos [53, 54]. Aunque supone un evento temprano, que se da incluso antes de que tenga lugar la transformación tumoral en las hiperplasias, también está asociado a la progresión tumoral [54]. Provoca inestabilidad cromosómica, derepresión de genes impresos, y reactivación de retrotransposones, así como expresión génica aberrante [55]. La hipometilación del ADN de genes individuales en células cancerosas es poco común y específica del tumor [53, 54], sin embargo, es destacable que puede estar involucrada en la sobreexpresión de impulsores oncogénicos [55]. Por otro lado, la hipermetilación se observa con frecuencia en elementos reguladores de la transcripción, como promotores y enhancer de genes [53–55]. La adición de esta marca hace que los promotores se silencien [53–55], presumiblemente por el reclutamiento de represores transcripcionales y enzimas modificadoras de histonas que se reclutan de manera dependiente de la proteína de unión al ADN metilado (del inglés Methyl-CpG-binding domain; MBD) [55].

La metilación del ADN es sólo uno de los muchos mecanismos moleculares de control epigenético. Existen también alteraciones de los patrones normales de las modificaciones covalentes de las histonas que son otra pieza clave y, por lo tanto, otra característica distintiva del cáncer [56]. Las colas de las histonas cuentan con decenas de residuos susceptibles, cada nucleosoma posee ocho histonas, y estos están situados cada 200 pares de bases, lo que (como ya se mencionó) da como resultado una miríada de patrones diferentes. La distribución aberrante de las modificaciones postraduccionales en cáncer se ha documentado en gran medida y se ha relacionado con el mal funcionamiento de la maquinaria enzimática que regula su escritura, borrado y lectura [22]. Específicamente, en relación con la hipometilación de las secuencias repetitivas, se ha descrito una pérdida de acetilación en la lisina 16 y de la trimetilación en la lisina 20 de la histona H4 (H4k16ac y H4K20me3, respectivamente) como marcas de histonas distintivas en cáncer [56]. Identificar qué cambios juegan un papel clave en la tumorigénesis y cuáles son meros espectadores es un objetivo crítico y complicado.

2 Glándula tiroides

En humanos, el desarrollo del tiroides comienza en el día 20 tras la fecundación del ovocito, constituyéndose así la primera glándula del organismo [57]. El proceso comienza a partir de un engrosamiento del endodermo de la parte caudal del segmento faríngeo del intestino primitivo, entre las primeras y segundas bolsas faríngeas [58–61]. La organogénesis de la glándula tiroides (y su posterior mantenimiento en la edad adulta) es el resultado de una compleja interacción entre los factores de transcripción tiroideos. Asimismo, el desarrollo estructural folicular de la glándula coincide con el desarrollo funcional del eje hipotálamo-pituitario-tiroideo y la secreción de hormonas tiroideas a la circulación fetal (entre la semana 10 y la 12 del embarazo) [57]. Tras el nacimiento la propia glándula del neonato, completamente funcional, se encargará de regular el metabolismo y el desarrollo postnatal [57]. Las hormonas tiroideas son secretadas por este órgano especializado y se diseminan sistémicamente para jugar un papel esencial en el desarrollo, crecimiento y homeostasis metabólica, tanto en humanos como en animales.

2.1 Anatomía macro- y microscópica general

En el cuerpo humano la glándula tiroides se sitúa a la altura de la quinta vértebra cervical y la primera vertebra torácica entre el cartílago cricoides y la escotadura supraesternal, y se envuelve sobre la parte anterolateral de la laringe y la tráquea (**Figura 7a**). A pesar de que el vocablo griego para referirse al tiroides significa “forma de escudo”, en realidad es una estructura formada por dos lóbulos unidos por un istmo que se asemeja más a una mariposa (**Figura 7b**). Ambos lados son simétricos y tienen forma cónica con su vértice apuntando caudalmente. Esto es así porque las extremidades inferiores de los lóbulos laterales (denominamos polos inferiores) no hacen proyecciones tan prominentes como los superiores. Del borde lateral de los lóbulos surge una formación fibrosa, llamada ligamento de Berry, que une la glándula al eje laringotraqueal [62–66].

El tiroides es de color pardo, profusamente irrigado y de consistencia blanda [63, 64]. Sus dimensiones varían con la edad y entre individuos, pero generalmente la glándula normal de una persona sin deficiencia de yodo tiene un peso situado entre los

12-25 g, y aproximadamente 5 cm de largo y 2-3 cm en las dimensiones máximas transversal y anteroposterior [62, 63].

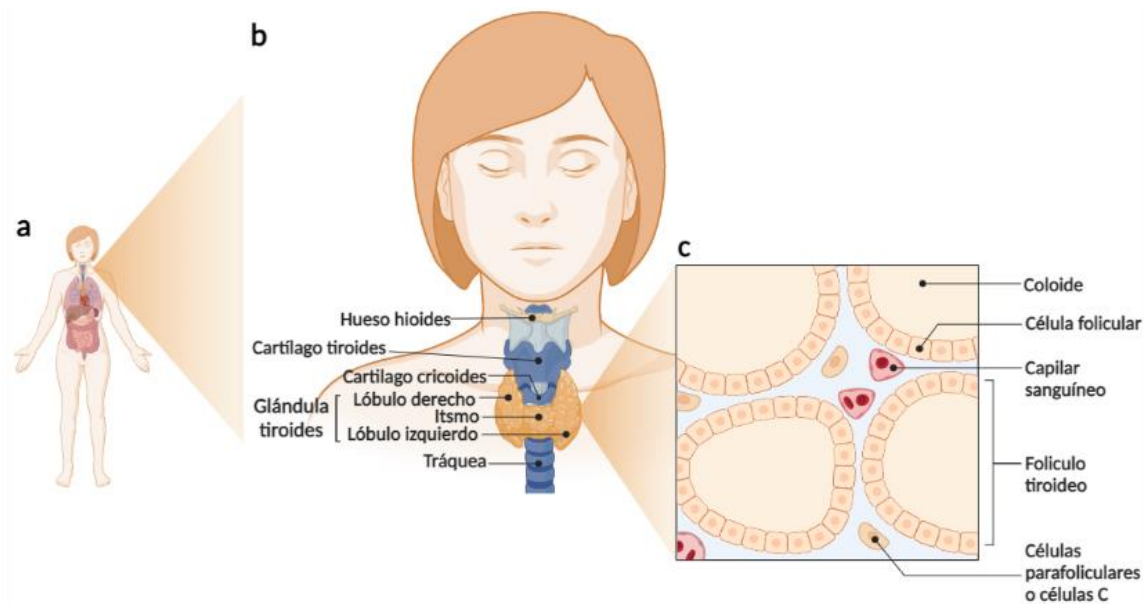


Figura 7 | Localización (a), anatomía (b) e histología (b) de la glándula tiroidea normofuncionante en el cuerpo humano. Imagen creada con BioRender.

Desde el punto de vista microscópico, se encuentra la unidad funcional y estructural de la glándula que es el folículo tiroideo (**Figura 7c**). Cada uno está formado por células foliculares y está rodeado de tejido conectivo en el que se pueden encontrar vasos capilares, fibroblastos dispersos, delgadas fibras de colágeno y un pequeño porcentaje de células parafoliculares o células C [63, 65, 66]. Las células foliculares son células epiteliales polarizadas y son responsables de la producción, almacenamiento y secreción de hormonas tiroideas [63, 65, 66]. Están dispuestas en unidades esféricas conocidas como folículos, y cuya luz contiene una sustancia proteínica producto de la secreción de los tirocitos denominada coloide [63, 65, 66]. Su morfología varía según el estado funcional de la glándula, ya que en periodos de inactividad (en la que los folículos tiroideos están llenos de colide) las células epiteliales son cubicas o planas, mientras que cuando hay gran actividad (cuando se ha liberado el coloide) son cilíndricas altas [63, 65, 66].

2.2 Biogénesis, regulación y función de las hormonas tiroideas

La síntesis y liberación de hormonas tiroideas, tiroxina o tetrayodotironina (T_4) y triyodotironina (T_3), está estrictamente controlada por el eje hipotalámico-pituitario-

tiroideo (**Figura 8**). En respuesta a distintos estímulos, tanto fisiológicos como ambientales, las neuronas del núcleo paraventricular del hipotálamo secretan la hormona liberadora de tirotropina (TRH). Esta hormona estimula a la adenohipófisis para que produzca y secrete la hormona estimulante del tiroides o tirotropina (TSH) al torrente sanguíneo. Este proceso está estrictamente regulado por un sistema de realimentación negativa endocrina clásica por parte de las hormonas tiroideas [67].

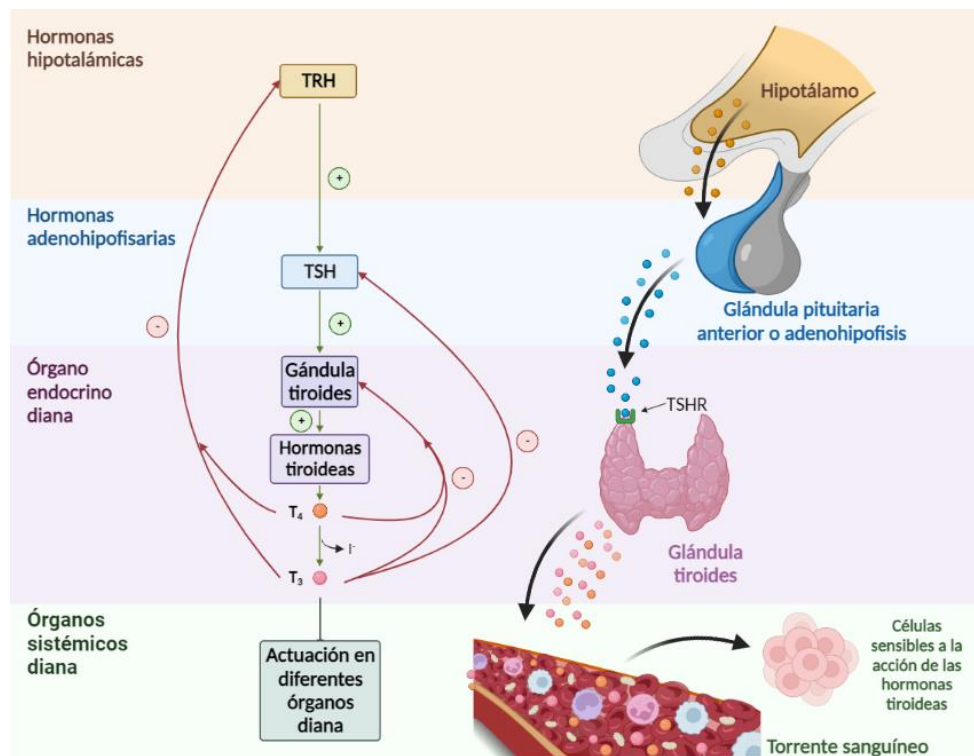


Figura 8 | Descripción general del eje hipotálamo-hipófisis-tiroides, las influencias directas e interacciones de retroalimentación entre los distintos componentes. TRH: hormona liberadora de tirotropina; TSH: tirotropina; TSHR: receptor de tirotropina; I: yoduro; T₄: tiroxina; y T₃: triyodotironina. Imagen creada con BioRender.

La organización folicular y la polaridad de los tirocitos son esenciales para el metabolismo especializado del órgano. El receptor acoplado a proteína G, TSHR, está situado en la membrana basolateral de los tirocitos y se une a la TSH circulante. Su unión desencadena la activación de la adenilil ciclasa (AC), lo que conlleva al aumento de los niveles intracelulares de adenosín monofosfato cíclico (AMPc); y la cascada de fosfolipasa C (PLC) generando inositol 1,4,5-trifosfato (IP3) y diacilglicerol (DAG), moléculas que estimulan la liberación de iones de calcio en el citoplasma [68, 69]. Estos segundos mensajeros generados por la activación inicial del TSHR finalmente resultan en la modulación de la expresión de genes que controlan los aspectos funcionales de la

célula tiroidea como la captación de yodo; la síntesis, yodación, endocitosis y proteólisis de tiroglobulina; la actividad de la yoduro peroxidasa; y la liberación de las hormonas al torrente sanguíneo para su actuación sistémica (**Figura 9**) [68].

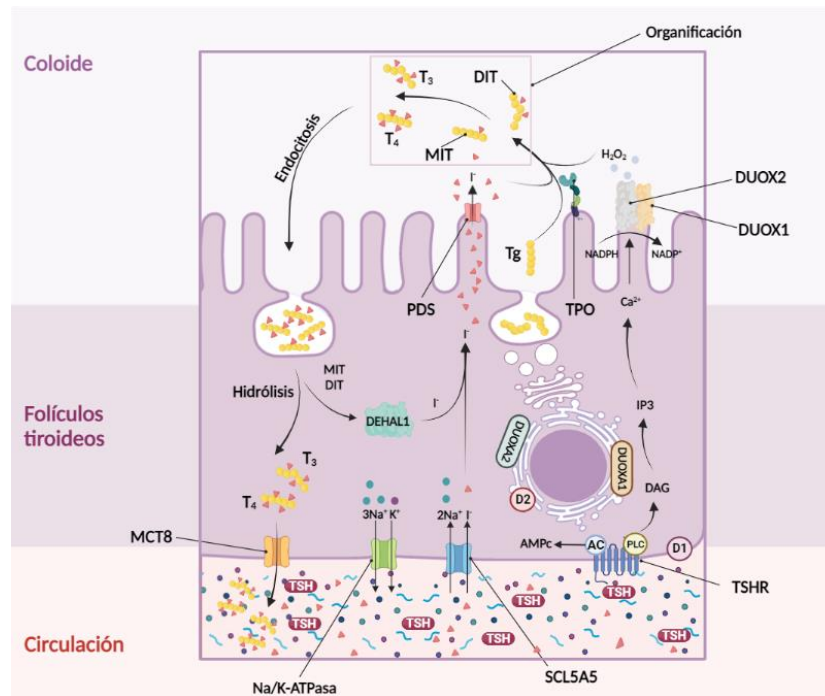


Figura 9 | Representación esquemática de los aspectos clave del transporte de yodo tiroideo y la síntesis de hormona tiroidea en los tirocitos altamente polarizados. AC, Adenil ciclasa; ATPasa, Adenosina trifosfatasa; AMPc, Adenosín monofosfato cíclico ;D1, deyodasa tipo 1; D2, deyodasa tipo 2; DAG, diacilglicerol; DEHAL1, yodotirosina deshalogenasa 1 (IYD); DIT, diyodotirosina; DUOXA1 y 2, factor de maduración de las oxidasas duales; DUOX1 y 2, oxidasa dual; IP3, trifosfato de inositol; MCT8, Transportador de monocarboxilato 8; MIT, monoyodotirosina; NADP+, forma oxidada de nicotinamida adenosina dinucleótido fosfato; NADPH, nicotinamida adenosina dinucleótido fosfato reducido; SCL5A5 o NIS, simportador de yoduro de sodio; PLC, fosfolipasa C; T₃, triyodotironina; T₄, tiroxina; Tg, tiroglobulina; TPO, peroxidasa tiroidea; TSH, tirotropina; TSHR, receptor de tirotropina. Imagen creada con BioRender.

La captación de iones de yodo se produce en la membrana basolateral de los folículos tiroideos, gracias al cotransportador NIS. El yoduro se transloca a través de su membrana apical al coloide de los folículos a través de otros transportadores, como la pendrina, PDS; anoctamina-1, ANO1; el regulador de la conductancia transmembrana de la fibrosis quística, CFTR; y el canal de cloruro Cl⁻/H⁺, CIC-5 [70, 71]. Se produce la señal para que la tiroglobulina (Tg) se sintetice en el retículo endoplasmático rugoso, tras adquirir la conformación apropiada es transferida al aparato de Golgi para su maduración final y, por último, ser exocitada a través de la membrana apical de los tirocitos [71, 72].

El yoduro será oxidado a hipoyodito por los grupos hemo de la peroxidasa tiroidea (TPO) e incorporado a los residuos libres de tirosina de la Tg, para formar 3-monoyodotirosina (MIT) y 3,5-diyodotirosina (DIT) en un proceso denominado organificación (**Figura 10a**) [73]. Adicionalmente, la TPO también cataliza el acoplamiento de dos moléculas de DIT o una de DIT y una de MIT, para dar lugar a las hormonas tiroideas T_4 y T_3 , respectivamente (**Figura 10b**) [71, 73]. Para que la oxidación tenga lugar es imprescindible la presencia de peróxido de hidrógeno (H_2O_2) en el coloide, que es generado por las proteínas DUOX1/2 dependientes de calcio y NADPH [74]. Los factores de maduración DUOXA1/2, proteínas residentes del retículo endoplasmático, son necesarias para la maduración de estas enzimas duales [73, 74].

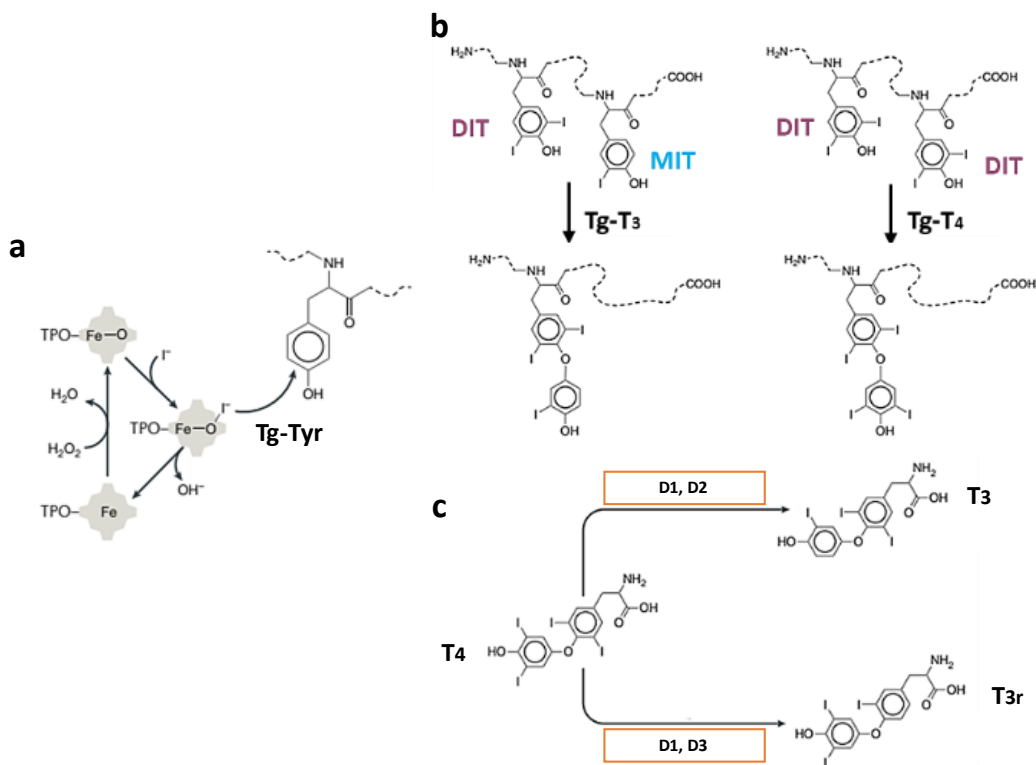


Figura 10 | Fisiología y estructura de las hormonas tiroideas. (a) La TPO cataliza la yodación que conduce al acoplamiento de mono-yodotirosina (MIT) y di-yodotirosina (DIT) dentro de la Tg en un proceso denominado organificación. (b) En el proceso de acoplamiento la TPO une dos moléculas de DIT o una de DIT y una de MIT, para dar lugar a las hormonas tiroideas T_4 y T_3 , respectivamente. (c) Además, las yodotironina desyodasas de tipo 1 y tipo 2 (DIO1 y DIO2) proporcionan actividad de desyodasa del anillo fenólico externo en la molécula de T_4 para producir T_3 . La inactivación de la hormona tiroidea ocurre por desyodación del anillo fenólico interno catalizada principalmente por DIO3 (y secundariamente por DIO1) dando lugar a la T_3r (hormona tiroidea reversa). Figura modificada de Citterio y colaboradores (2019) [71].

MIT, DIT, T₄ y T₃, que todavía forman parte de la molécula de Tg, se almacenan en el coloide de los folículos tiroideos hasta que se produzca la señal que haga que se endociten para ser secretados por el folículo basolateralmente [71]. En el interior del citoplasma se produce la proteólisis de la Tg que libera hormonas tiroideas de su esqueleto polipeptídico [71]. Mientras, DIT y MIT son desyodadas por la yodotirosina deshalogenasa (DEHAL1) para permitir el reciclaje del yoduro [70, 73].

La T₄ sufre un proceso activación a través de la desyodación del anillo fenólico externo (**Figura 10c**) por parte de las selenoproteínas yodotironina desyodasas 1 y 2 (D1 y D2), para dar lugar a la T₃ [71]. Si, por el contrario, la desyodación se produce en el anillo fenólico interno (a través de D1 y D3) el resultado será una T₃ reversa inactiva [71, 75]. Este proceso puede ocurrir tanto en el citosol de células en tejidos diana periféricos como en tirocitos [71]. Se sabe que la entrada y la salida de yodotironinas en las células están mediadas por varias proteínas transportadoras de membrana específicas, incluidos los transportadores de monocarboxilato 8 y 10 (MCT8 y MCT10), el polipéptido 1 cotransportador de aniones orgánicos dependientes de sodio (OATP1) [67, 71, 75].

Las hormonas tiroideas salen de los tirocitos al torrente sanguíneo y son transportadas hasta los órganos diana unidas a globulina fijadora de tiroxina, transtiretina, paraalbumina y una pequeña proporción libre [67]. Cuando llegan a las células, la regulación que ejercen sobre los distintos procesos metabólicos puede ser de dos tipos: (1) genómica o nuclear y (2) no-genómica o extranuclear [76, 77]. La regulación nuclear es el mecanismo de actuación clásico, que implica que la forma activa T₃ ingrese en las células y se una a los receptores de la hormona tiroidea en el núcleo ejerciendo su acción directa sobre la transcripción de genes sensibles [76, 77]. Por el contrario, el mecanismo no-genómico son aquellos que no involucran directamente a los receptores de la hormona tiroidea y son mediados por la membrana celular y/o los sitios de unión mitocondrial y, en ocasiones, provocados por sus catabolitos endógenos [76, 77].

Las hormonas tiroideas tienen efectos pleiotrópicos en muchas vías diferentes y órganos diana, incluyendo cambios en el consumo de oxígeno, proteínas, carbohidratos, lípidos y metabolismo de las vitaminas [78, 79]. También regulan la termogénesis adaptativa, y están involucradas en el desarrollo y mantenimiento de las funciones

cardiaca, renal, pulmonar, hepática y del sistema nervioso central tanto en el feto como en el adulto [57, 78–80]. Por su parte, las células C representan solo un porcentaje relativamente pequeño del tiroides, pero también son responsables de la producción de hormonas que se secretan a la circulación [63, 65, 66]. La calcitonina juega un importante papel en la regulación de la resorción ósea y uno menor en la homeostasis del calcio [66]. Las células parafoliculares también pueden secretar en menor medida somatostatina, péptido liberador de gastrina, katalcalcina y helodermina [66].

3 Nódulos tiroideos

3.1 Epidemiología

La disfunción y las anomalías anatómicas del tiroides se encuentran entre las enfermedades más comunes de las glándulas endocrinas. Entre ellas, la presencia de nódulos tiroideos es cada vez más común en la población general [81]. Su prevalencia depende del método empleado para su evaluación y la población de estudio, diagnosticando un 5 % con palpación y entre 20-35 % con ultrasonido. Sin embargo, los porcentajes de casos reales pueden ser aún más altos, ya que estudios realizados en autopsias sugieren que los nódulos tiroideos ocultos pueden llegar a estar presentes hasta en un 65 % de las personas [81].

Aunque la mayoría de los nódulos tiroideos son lesiones benignas, aproximadamente entre el 10 % y el 15 % de ellos pueden representar cáncer de tiroides [82] y dada la prevalencia es necesaria una evaluación exhaustiva de los mismos para descartar la malignidad. Es importante destacar el sesgo en la incidencia con respecto al sexo, afectando tres veces más a mujeres que a hombres y siendo el séptimo tumor más frecuente en mujeres españolas [83–85]. También la edad es importante, ya que aumenta su frecuencia a partir de los 40 años [83–85].

Asimismo, cabe destacar que la pandemia de Covid-19 tuvo un impacto significativo en el número de diagnósticos de cáncer en el mundo desde 2020 [86], siendo España de los países más afectados durante la primera ola. Todavía no está claro el alcance que ha tenido sobre el diagnóstico de los tumores y las consecuencias que lo acompañaran, pero se calcula que en 2020 se sobreestimaron los casos, ya que se observó una disminución del 20,8 % en pacientes recién diagnosticados [87], en parte, por la presión

asistencial que existía en ese momento. El Centro Internacional de Investigaciones sobre el Cáncer (del inglés International Agency for Research on Cancer; IARC) estimó que en ese año hubo una incidencia aproximada de 18,1 millones en el mundo (excluyendo los tumores cutáneos no melanoma), de los cuales 586.202 correspondieron a nuevos casos de cáncer de tiroides [83]. Según la Asociación Española Contra el Cáncer (AECC) usando los datos proporcionados por la IARC y el padrón continuo del Instituto Nacional de Estadística (INE) español, se calcularon 282.421 casos en 2020 de los cuales 5.063 se atribuyen a los carcinomas en la glándula tiroides [84].

Los datos recogidos por el Área de Tiroides de la Sociedad Española de Endocrinología y Nutrición (SEEN), indican que la incidencia se ha duplicado en los últimos 10 años [88]. En este trabajo se resalta que los nuevos diagnósticos anuales por cada 100.000 habitantes pasó de 3,42 casos en 1997 a 9,40 casos en 2015 [88]. Mientras que la IARC para 2020, muestra que la tasa estimada de incidencia de cáncer de tiroides (ajustada por edad de personas mayores de 20 años a la población mundial) por cada 100.000 habitantes es de 12,3 casos (**Figura 11**) [83].

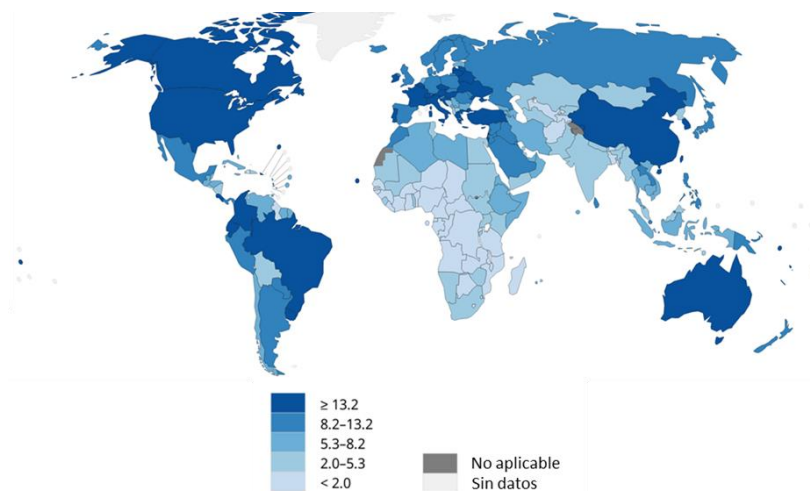


Figura 11 | Incidencia de cáncer de tiroides en el mundo. La tasa se presenta como nuevos casos por cada 100.000 habitantes. Imagen tomada y adaptada de GLOBOCAN 2020 [83].

Según la AECC, la proyección en 2022 es de 290.175 nuevos casos de cáncer en España, de los cuales 5.116 serán de tiroides (**Figura 12a**), similares a las tasas predichas por la Red Española de Registros de Cáncer (REDECAN) [84, 85]. En el caso específico de Asturias se prevén 7.688 nuevas incidencias de cáncer para 2022, de los cuales 119 serán de tiroides (**Figura 12a**). Esto quiere decir que 13 de cada 100.000 personas mayores de 15 años serán diagnosticadas en la comunidad autónoma [84].

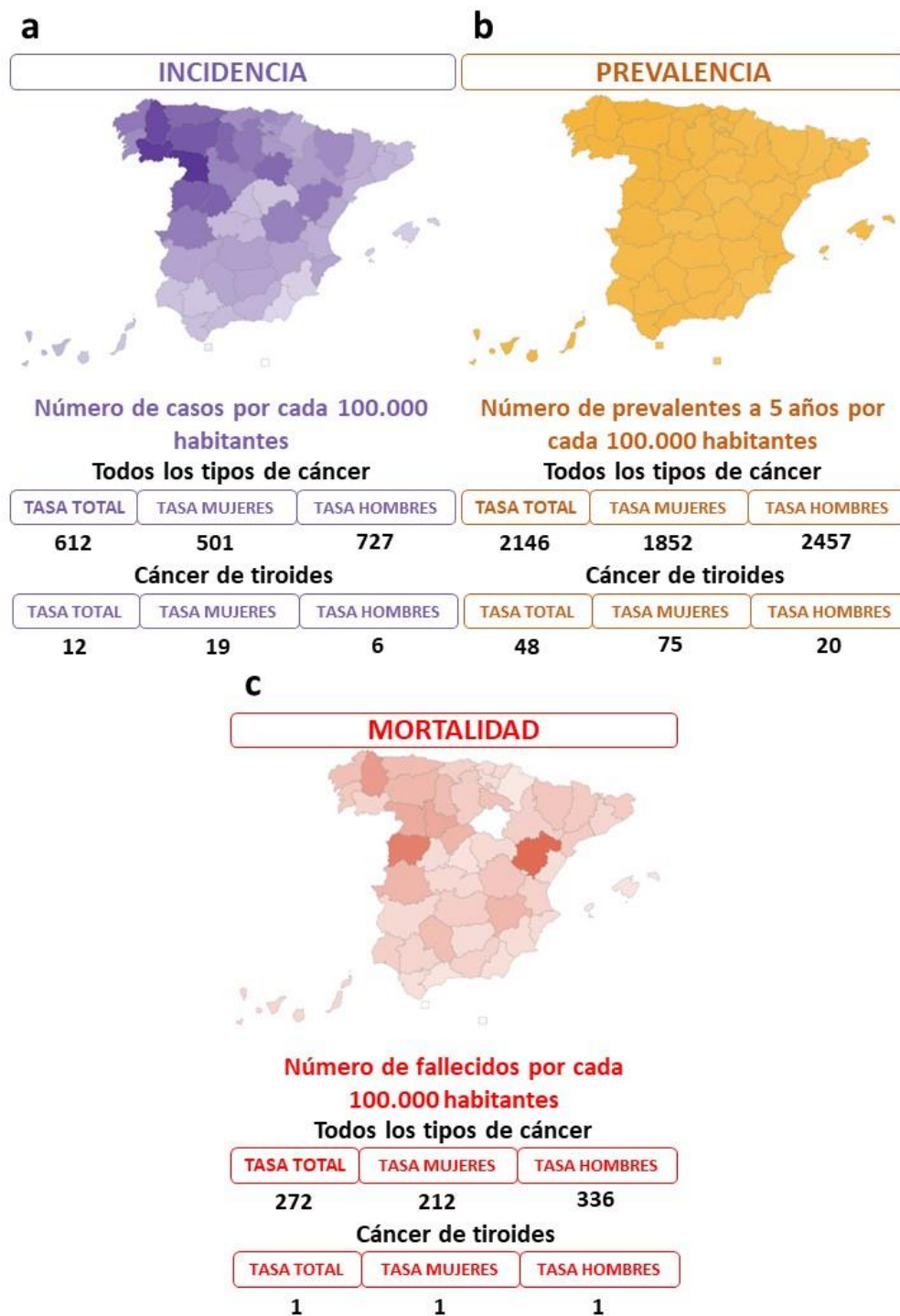


Figura 12 | Representación demográfica de las tasas de incidencia (a), prevalencia (b) y mortalidad (c) del cáncer de tiroides en España por cada 100.000 habitantes. Imágenes tomadas y adaptadas de la página web AECC Observatorio [84]

La supervivencia llega al 98,4 % en Estados Unidos entre el 2012 y el 2018, suponiendo así solo el 0,4 % de todas las muertes por cáncer [89]. Los datos de mortalidad en 2020, muestran que los tumores son una de las principales causas de defunción en España (22,8 %), solo por detrás de las enfermedades del sistema circulatorio [90]. Además, en el año 2022 se estima un aumento una tasa de mortalidad,

llegando a 272 fallecimientos por cada 100.000 habitantes con una edad superior a 15 años (**Figura 12b**) [84]. Sin embargo, si nos centramos en los datos específicos para el cáncer de tiroides, se puede ver que tiene bajos valores de mortalidad y altos de prevalencia. Siendo en este caso la tasa de fallecimientos de 2022 de 1 por cada 100.000 habitantes españoles y la de prevalencia a 5 años de 48 (**Figura 12b-c**) [84].

3.2 Factores de riesgo demográficos y ambientales en cáncer de tiroides

Como ya se ha mencionado, los nódulos tiroideos son muy prevalentes en la población, constituyendo un reservorio de lesiones cancerosas sin investigar. Asimismo, la causa del observado incremento de la incidencia del cáncer de tiroides durante los últimos años no es conocida. Parte del aumento que se ha observado se atribuye a la mejora en los métodos de detección. Si el acrecentamiento de nuevos casos diagnosticados se debiera a esta razón exclusivamente, los distintos tipos tumorales se verían afectados en la misma proporción, sin embargo esto no es siempre así [91]. Por lo tanto, no se puede excluir que existan carcinógenos y/o factores de riesgo específicos que pudieran influir en el aumento de la incidencia, al menos en algunas poblaciones (**Figura 13**).

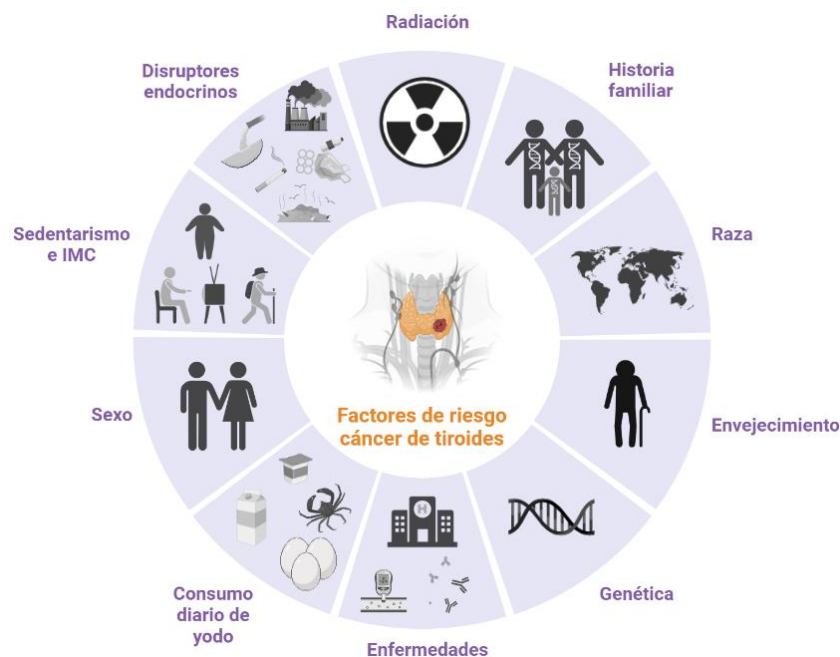


Figura 13 | Factores de riesgo que podrían influenciar el desarrollo de cáncer de tiroides en la población. Imagen creada con BioRender.

Como ya se mencionó anteriormente, las mujeres padecen más cáncer de tiroides que los hombres, convirtiendo al sexo en un claro factor de riesgo. Por otro lado, los hombres presentan un peor pronóstico, asociándose a formas más agresivas de la enfermedad [92]. La causa de las diferencias entre ambos no está clara, pero es probable que las hormonas sexuales de la mujer, sobre todo durante la época fértil, tengan cierto impacto tanto en la promoción como en la limitación del proceso oncogénico [92, 93].

No solo existen diferencias en la incidencia debidas al sexo, si no también parece haber diferencias raciales. Los pacientes americanos de raza negra, reportados por la SEER, presentan una tasa de incidencia menor que el resto de grupos analizados [89]. Una extensa revisión sistemática realizada por Keane y colaboradores en 2017 [94], revela que los hombres caucásicos padecen más cáncer de tiroides en comparación con asiáticos, hispanicos y negros. Los personas caucásicas y negras tienen una mayor proporción de cáncer folicular, mientras que en los hispanos la edad de diagnóstico es anterior al resto de grupos [94]. Las causas pueden ser debidas al sexo del paciente, riesgo genético, factores ambientales o el estilo de vida [94]. Sin embargo, los resultados obtenidos de analizar las posibles diferencias se ven fuerte mente influenciados por el nivel socioeconómico, por lo que los resultados no pueden ser concluyentes [94].

Asimismo, el envejecimiento es un factor inextricablemente relacionado con el desarrollo del cáncer, tanto molecular, como epidemiológicamente [95]. En la mayoría de los tumores se observa un aumento en la incidencia en personas mayores a 60 años, sin embargo, el cáncer de tiroides se diagnostica con mayor frecuencia en personas de 45 a 54 años [89, 95]. Un hecho generalizado al estudiar la epidemiología del cáncer es un aumento de la mortalidad a edades más avanzadas, y lo mismo ocurre en el cáncer de tiroides donde se observa un mayor porcentaje de muertes entre los 75 y los 84 años [89]. Cuando se sobrepasa esta franja de edad se observa un descenso en las defunciones [89], concurrente con la tendencia de mortalidad global asociada al cáncer, debido a la heterogeneidad de edad en las distintas poblaciones [95].

La influencia de la radiación ionizante está bien documentada en el cáncer de tiroides, sobre todo si la exposición se produce en la niñez o durante la adolescencia [91, 93, 96, 97]. Debido a la capacidad que tiene de concentrar yodo, la glándula tiroides puede recibir más radiación que el resto de los tejidos. La exposición durante el

desarrollo en la niñez aumenta en 15 veces la probabilidad de desarrollar adenomas foliculares [98]. Una fuente de radiación son los dispositivos médicos antiguos como radiografías o tomografía computarizada en la zona del cuello y la cabeza, y la terapia radioyodo [91, 93, 96, 97]. En la actualidad, estas herramientas han ido mejorado sus características, estableciendo limitaciones, haciéndolas más seguras y así disminuyendo la probabilidad de desarrollar cáncer ante su exposición [97]. Asimismo, tras la exposición a desastres nucleares como el acontecido en Chernóbil y las bombas de Hiroshima y Nagasaki, se ha observado un aumento de la incidencia de cáncer de tiroides en adultos, sobre todo mujeres [96].

El déficit de yodo es necesario para el correcto funcionamiento de la glándula, afecta a la producción de hormonas tiroideas y tiene múltiples implicaciones sobre la salud de la población. Desde que se introdujo la yodación universal de la sal en la década de 1990, si el yodo contribuye al aumento de la incidencia del cáncer ha sido un tema de controversia y preocupación pública [93]. Tanto la aparición de nódulos tiroideos como de bocio se ha relacionado con el déficit de este micronutriente, sin embargo, la incidencia del cáncer de tiroides no se ve modificada por el estado nutricional de yodo [99]. Algunos estudios sugirieron que la ingesta de yodo puede influir en la distribución de ciertos subtipos de cáncer de tiroides. Puede haber más carcinomas del subtipo papilar en áreas con adecuada ingesta y menos foliculares agresivos. Mientras que en zonas con deficiencia de yodo hay más riesgo de formas más agresivas de carcinoma folicular e incluso anaplásico [99].

La obesidad es una enfermedad multifactorial influenciada por los cambios en la dieta y un estilo de vida sedentario, y desempeñan también un papel importante en el riesgo de desarrollar múltiples enfermedades, como cardiopatías, diabetes mellitus tipo 2 y cáncer [100]. Se desconoce si existe un efecto específico sobre la glándula tiroides y cuál es el mecanismo subyacente, atribuyéndose más a la presencia de diabetes [91, 96, 97]. La obesidad o el sobrepeso se considera ampliamente como el principal impulsor de la diabetes mellitus tipo 2 [100]. Estos niveles elevados de diabetes son simultáneos con tasas crecientes de obesidad, y esto parece reflejar factores ambientales y genéticos comunes que subyacen a ambas condiciones [100]. Parece ser que resistencia a la insulina y la hiperinsulinemia puedan estar involucrados mediada por receptores

específicos sobreexpresados por las células cancerosas que estimulan su proliferación, diferenciación y transformación [91, 93, 96].

Se sabe que existen químicos que actúan como disruptores endocrinos e interfieren con la señalización endocrina a nivel celular y molecular, lo que puede provocar tumores en tejidos y órganos sensibles a las hormonas, incluidos: mama, próstata, testículos y tiroides [101]. En general, la literatura actual sugiere que la exposición a ciertos congéneres de retardantes de llama, bifenilos policlorados y ftalatos, así como a ciertos pesticidas, puede estar potencialmente asociada con un mayor riesgo de cáncer de tiroides [96, 101]. Sin embargo, hasta el momento no se ha establecido una correlación causal entre los contaminantes ambientales y el cáncer de tiroides en humanos [96].

La mayoría de los cánceres de tiroides son esporádicos, lo que quiere decir que no hay un claro factor de riesgo. En algunos casos, sin embargo, puede ser hereditario. El cáncer de tiroides no medular familiar constituye entre un 3-9 % de todos los tumores tiroideos y se presenta como una entidad discreta o como parte de otros síndromes de cáncer familiar como el síndrome de Gardner, la enfermedad de Cowden y otros síndromes raros [102, 103]. Asimismo, también se puede encontrar el carcinoma medular de tiroides familiar, que corresponde a uno de los varios subgrupos clínicos de la neoplasia endocrina múltiple tipo 2 (NEMII), causadas por mutaciones en el protooncogén *RET* que pueden ser heredadas de manera autosómica dominante [103]. El carcinomas medulares suele ser la primera manifestación clínica de NEMII [98].

Las mutaciones pueden pasar de los progenitores a su descendencia, o pueden ser adquiridos por daño genético que ocurre a lo largo de la vida. Parece existir un vínculo entre las exposiciones en el útero, en particular las relacionadas con trastornos de la tiroides materna, deficiencias de yodo, susceptibilidad genética, enfermedades autoinmunes como la tiroiditis de Hashimoto y un mayor riesgo de cáncer de tiroides más adelante en la vida [104].

3.3 Clasificación histopatológica

Por un lado, las hiperplasias son una respuesta fisiológica ante un estímulo que conduce a la proliferación celular normal y al agrandamiento de un órgano [105].

Aunque los tirocitos en condiciones normales tienen una tasa de replicación baja, bajo las señales adecuadas (tanto ambientales como genéticas) son muy proliferativas [106]. Un ejemplo puede ser deficiencia de yodo, que directa o indirectamente, es el principal promotor del potencial replicativo de la glándula [106]. Mientras que las neoplasias se definen como una masa anormal de tejido, cuyo crecimiento es superior, y además no coordinado con el resto del organismo y que persiste en su crecimiento cuando termina el estímulo que lo ha desencadenado [105].

Por otro lado, los tumores son tejidos complejos y heterogéneos compuestos por múltiples tipos de células que interactúan entre sí. Desde el punto de vista clásico, las células inicialmente sanas, evolucionan hacia un estado neoplásico adquiriendo una serie de características y capacidades distintivas, que en última instancia les permitirán diseminarse a otras partes del cuerpo. En específico, la patogenia de los nódulos tiroideos se caracteriza por la acumulación de mutaciones y cambios epigenéticos que conducen a la progresión a través de un proceso de desdiferenciación. En el campo de la patología, se reconoce la desdiferenciación como la pérdida morfológica de la identidad del linaje con la progresión del tumor [107]. Así, las lesiones tiroideas pueden originarse por alteraciones en cualesquiera de los tipos celulares que componen la glándula (células foliculares; células parafoliculares; linfocitos; elementos estromales; y vasculares) e, incluso, puede presentar metástasis procedentes de tumores primarios en otras localizaciones.

La clasificación de los nódulos tiroideos (**Figura 14**) está en continuo desarrollo y actualización, por ello, en este trabajo nos centraremos en la separación de los nódulos en base a los hallazgos histológicos descrita por la Organización Mundial de la Salud (OMS) en 2017⁴ [98]. Se prestará especial atención a las lesiones que se originan a partir de las células foliculares de tiroides, y a continuación se describen con mayor detalle:

⁴ La clasificación completa se detalla en Anexo 1: Tabla S1.

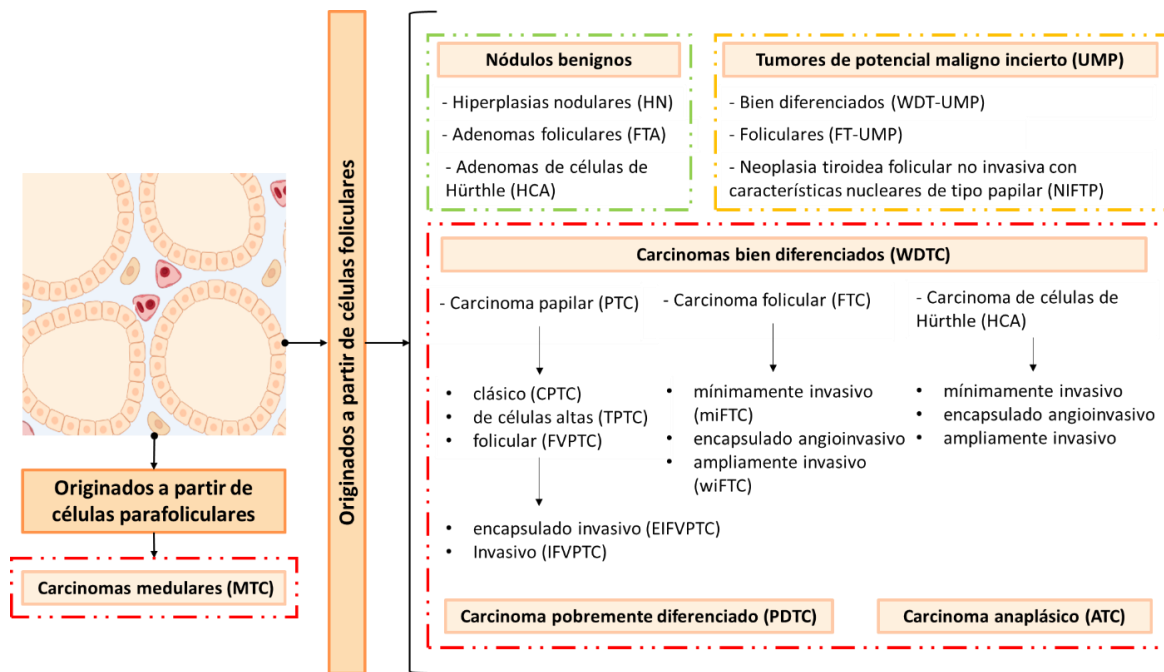


Figura 14 | Esquema representativo de la clasificación de los tumores tiroideos (OMS 2017) referenciados a lo largo de este trabajo.

3.3.1 Hiperplasia nodular

Los llamadas hiperplasias nodulares (NH), son bultos de origen policlonal que pueden presentarse de forma aislada o encontrarse en el contexto de un bocio multinodular [98, 108, 109]. La prevalencia se encuentra entre 1 y 10 %, dependiendo de las ubicaciones geográficas [108]. Presentan tamaños que pueden ir desde 4 mm a 3 cm, o incluso más [98]. Morfológicamente, son lesiones foliculares hiperplásicas definidas por la proliferación epitelial folicular benigna bien circunscrita, que presentan una histología y arquitectura heterogéneas indistinguibles del tejido circundante [105]. Cabe recalcar que los NH pueden presentar en algunas ocasiones pseudocápsula, ya que pueden sufrir cambios regresivos en los que la fibrosis perinodular hace que parezcan a un adenoma folicular. Asimismo, cuando la fibrosis no es homogénea, haciendo que estas irregularidades sean confundidos como invasión y por lo tanto, se pueden diagnosticar erróneamente como carcinomas [110]. El tratamiento de los NH depende de las características clínicas y del contexto en el que aparezcan [108]. Sin embargo, al no ser neoplasias malignas, generalmente no precisan ser operados.

3.3.2 Adenoma folicular

La prevalencia de los adenomas foliculares (FTA) en la población se estima entre un 3-7 % [98]. Se define como una neoplasia benigna, encapsulada, no invasiva, que muestra signos de diferenciación folicular, pero sin características nucleares de carcinoma papilar [98]. Los adenomas son de origen monoclonal, característica que los diferencia de los HN [98, 105, 109]. Además de las características ya mencionadas, presentan rasgos histopatológicos que los diferencian del tejido adyacente [98]. Presentan un estroma escaso con una arquitectura celular que puede ser normal, macrofolicular, microfolicular, sólida y/o trabecular, y que raramente presentan patrones insulares o papilares. Los núcleos en los FTA se presentan de formas y tamaños variables; sin embargo, suelen estar situados en la base, redondeados, con contornos suaves y cromatina uniforme.

Como ocurre con la mayoría de las neoplasias de la tiroides, los FTA suelen ser asintomáticos [98]. Su manejo en la práctica clínica depende múltiples factores, incluyendo el tamaño del nódulo, si mantienen sus dimensiones en el tiempo, de las características ecográficas, los resultados de la biopsia exploratoria y los deseos del paciente [108]. Pueden coexistir con los NH en los bocios multinodulares y aparecer junto a otros nódulos de carácter maligno [105]. Sin embargo, como son tumores benignos el pronóstico de estos nódulos es excelente [98, 108].

3.3.3 Carcinomas de tiroides bien diferenciados

3.3.3.1 Carcinoma folicular de tiroides

Esta categoría representa un 10-15 % de los tumores de tiroides [111]. El carcinoma folicular de tiroides (FTC) es bien diferenciado, delimitado por una cápsula (más bien gruesa) y sin características nucleares papilares, que se origina a partir de las células epiteliales foliculares de la glándula tiroides. La arquitectura celular puede ser normal, macrofolicular, microfolicular, cribiforme, y/o trabecular/sólido [98]. Al ser una neoplasia maligna un rasgo distintivo, que las diferencia de NH y FTA, es la invasión capsular y/o vascular, ya sea más o menos extensiva [98, 111]. Se define angioinvasión o invasión vascular como la penetración de células tumorales en el lumen de los vasos sanguíneos y que deben estar recubiertas por una capa de células endoteliales. Por otro

lado, la invasión de la cápsula se define como la transgresión completa e inequívoca de la cápsula fibrosa por las células tumorales [98].

La extensión de la invasión es relevante desde el punto de vista del pronóstico. De esta forma los FTC se pueden subclasificar en: (1) mínimamente invasivo, (2) encapsulado angioinvasivo y (3) ampliamente invasivo [98]. En los tumores mínimamente invasivos se observa un comportamiento muy indolente, mientras que los tumores ampliamente invasivos y angioinvasivos suelen mostrarse como tumores más agresivos [98]. La tasa de supervivencia libre de enfermedad a los 10 años para el conjunto de FTC mínimamente invasivo y angioinvasivo es del 93,5 %, mientras que las formas más agresivas bajan al 53,6 % [111]. La diseminación linfática es un suceso raro en FTC (1-7 %) [111] y suelen redirigir el diagnóstico hacia carcinoma papilar [98]. Si hay presencia de células tumorales en los ganglios linfáticos es en los FTC más invasivos, generalmente es en combinación con metástasis en otros órganos. Los tumores secundarios con difusión hematogena se desarrollan con mayor frecuencia en pulmón y hueso (6-20 %) [111]. Los pacientes con formas tan avanzadas de la enfermedad tienen una tasa de supervivencia a largo plazo situada entre el 31 % y el 43 % [111]. El manejo de estos pacientes incluye la lobectomía parcial o total de la glándula tiroides y en ocasiones tratamiento con yodo radioactivo [98, 111].

3.3.3.2 Carcinoma papilar de tiroides

Los carcinomas papilares de tiroides (PTC) son los nódulos malignos que los pacientes desarrollan con mayor frecuencia, 80-85 % [98, 105]. Cuando un patólogo hace referencia a la palabra papilar, es para describir a las células que se unen para crear proyecciones de tejido largas en forma de dedos [98]. Las células se ven en la superficie exterior del tejido, mientras que los vasos sanguíneos pequeños se ven en el interior [98].

El PTC un tumor epitelial maligno que muestra evidencia de diferenciación de células foliculares y un conjunto de características nucleares distintivas, como: aspecto de vidrio esmerilado, hipocromáticos, tamaño grande, contorno irregular, hendiduras profundas, nucléolo pequeño, pseudoinclusiones y frecuentemente calcificaciones denominadas cuerpos de psamoma [98, 105]. Estos nódulos también presentan una

cápsula, pero por definición son invasivos para distinguirlos de sus contrapartes indolentes.

En este tipo de neoplasias malignas se pueden encontrar también numerosas variantes, entre ellas las más frecuentes son la clásica (55-65 %) y la folicular (23-41 %) [112]. La variante folicular del carcinoma papilar (FVPTC) representa un grupo de tumores con un patrón de crecimiento exclusivamente folicular, que se subclasifica a su vez en encapsulados infiltrativos y en infiltrativos [98]. Al ser nódulos malignos también presentan alteraciones en la continuidad de la cápsula [98].

En general, son tumores de buen pronóstico, en los que la tasa de supervivencia a 10 años se estima alrededor del 93 % [98, 98, 113]. Además, ambos siguen una historia clínica similar encontrando frecuentemente metástasis ganglionares y más rara vez a otros órganos [98].

3.3.3.3 Carcinomas oncocíticos

Se consideran un grupo aparte en la nueva clasificación de la OMS de 2017, y aquellos tumores con un porcentaje de células oncocíticas superior al 75 % [98]. Los tumores con invasión capsular y/o vascular se conocen como carcinoma de células de Hürthle (HTC), y su contraparte benigna se denomina adenoma de células de Hürthle (HTA) [98].

Los HTC son neoplasias malignas y a su vez se subdividen en los subtipos mínimamente invasivo, encapsulado angioinvasivo y ampliamente invasivo, al igual que los FTC. Representan del 3 % al 4 % de todos los cánceres de tiroides [114]. Sin embargo, hay que tener en cuenta que las células oncocíticas no son específicas de este tipo de nódulos y también se pueden encontrar en el resto, aunque en menor proporción [98, 105]. Los oncocitos son células grandes con grandes núcleos ubicados centralmente con nucléolos prominentes y un citoplasma abundante, granular y eosinofílico, que representan el anormal número de mitocondrias presentes [98, 108]. Por otro lado, los HCA tienen muy buen pronóstico, la tasa de supervivencia de los carcinomas oscila entre el 85 % para lesiones con cápsula prácticamente intacta y alrededor del 10 % para casos muy invasivos [98].

3.3.4 Tumores tiroideos de potencial maligno incierto

Los tumores tiroideos en el límite de la malignidad o también llamados tumores de potencial maligno incierto (UMP), corresponden nódulos donde no está claro su comportamiento biológico, ya que la invasión capsular y/o vascular es cuestionable [98]. Por un lado, una considerable proporción de carcinomas de bajo riesgo (como el FTC mínimamente invasivo) se reasignaron a la categoría de tumor folicular de potencial maligno incierto (FT-UMP). Además, la mayoría de los PTC encapsulados sin invasión clara se reclasificaron en esta categoría, incluidos el tumor bien diferenciado de potencial maligno incierto (WDT-UMP) y la neoplasia tiroidea folicular no invasiva con características nucleares de tipo papilar (NIFTP) [98]. Estos últimos se corresponden con la variante tumoral FVPTC encapsulado no invasiva, que fueron reclasificado en 2016 como una categoría independiente debido a sus características indolentes [98, 115]. Con esta reclasificación se pretendía reducir el sobretratamiento de los tumores de bajo riesgo. La resección de la glándula puede ser evitada en estos casos, pero puede ser una opción para aquellos pacientes que no quieran vigilar periódicamente estos nódulos [115].

3.3.5 Carcinomas desdiferenciados

En este grupo se encuentran los carcinomas foliculares pobremente diferenciados (PDTC) y los anaplásicos (ATC). La evidencia actualmente disponible apoya la suposición de que los carcinomas diferenciados (generalmente PTC) son precursores de PDTC y ATC [98, 116].

3.3.5.1 Carcinoma pobremente diferenciado

El carcinoma pobremente diferenciado se encuentra biológicamente en una posición intermedia entre los carcinomas bien diferenciados y los tumores más agresivos, los ATC [98, 117]. La prevalencia de los PDTC varía según la región geográfica, entre 0,3 % y 7 % [98]. Se caracterizan por la presencia de crecimiento sólido/trabecular/insular en más del 50 % del área del tumor; células pequeñas sin características de PTC pero con núcleos hipercromáticos contorneados; mitosis frecuentes; y necrosis extensa [98, 117]. Además, se trata de tumores voluminosos (>5

cm) y muy invasivos, extendiéndose hacia tejidos blandos (60-70 %), vasos sanguíneos (60 %) y formando tumores secundarios tanto ganglionares (15-65 %) como a distancia (40 -70 %; pulmón, hueso, cerebro, hígado, piel, bazo y riñón) [98]. Tienen peor prognosis que los tumores anteriormente mencionados con una tasa de supervivencia 60–70 % a los 5 años [117].

3.3.5.2 Carcinoma anaplásico

Los carcinomas anaplásicos son tumores raros (< 5 %) pero son los más indiferenciados y de peor pronóstico [98]. La supervivencia media de los pacientes se encuentra en los 6 meses, y solo el 10-15 % de los pacientes sobrevive 2 años después del diagnóstico [98, 113]. Son tumores muy agresivos que presentan extensa invasión de las estructuras adyacentes a la glándula (tráquea, esófago, músculos, nervio laríngeo recurrente), ganglios linfáticos cervicales y/u órganos a distancia (principalmente pulmón, hueso y cerebro).

Tabla 1 | Resumen de las características clínicas y epigenéticas en las lesiones foliculares de tiroides. Salvo especificado de otra manera, las alteraciones epigenéticas representadas en esta tabla hacen referencia a aquellas que ocurren con respecto al tejido normal de referencia.

	FTA	PTC	FVPTC	FTC	PDTC	ATC
Frecuencia	1-10 %	≈ 80-85 %	≈ 20 %	≈ 10-15 %	≈ 7 %	< 5 %
Diferenciación	Buena	Buena o moderada	Buena o moderada	Buena o moderada	Baja	Ninguna
Mortalidad	99-100 %	93 %	93 %	90 %	60-70 %	10-15 %
Frecuencia de mutaciones	Baja	Baja	Baja	Baja	Alta	Más alta
Patrones de metilación	Intermedio con respecto a FTC	Hipometilación	Hipometilación	Hipermetilación	Muy alterado, hipometilación	Muy alterado, hipometilación
Patrones de hidroximetilación	-	Disminuido	Disminuido	-	-	Disminuido
H3K18ac	Elevado	Elevado	Elevado	Elevado	-	Disminuido
H3K9–K14ac	Elevado	Elevado	Elevado	Elevado	-	Elevados
H4K12ac	Elevado	No hay cambios	No hay cambios	No hay cambios	-	No hay cambios
H3K27me3	Moderado	Elevado	Elevado	Elevado	Muy elevado	Muy elevado

3.4 Alteraciones moleculares en las lesiones foliculares de tiroides

En este apartado se revisa el conocimiento actual de las alteraciones moleculares en el cáncer de tiroides desde una perspectiva genómica y epigenómica, destacando los mecanismos mutacionales involucrados en la iniciación y progresión de las lesiones

tiroideas, las modificaciones postraduccionales de las histonas y el papel de las alteraciones en los niveles de metilación e hidroximetilación sobre el genoma (Tabla 1).

3.4.1 Perfil de alteraciones genéticas

La disponibilidad de avanzadas tecnologías de secuenciación del genoma ha llevado a un gran progreso en la aclaración de los mecanismos moleculares que subyacen al cáncer de tiroides, una enfermedad genéticamente simple con una carga de mutaciones somáticas relativamente baja [118–121]. Las neoplasias clonales de tiroides pueden adquirir mutaciones conductoras que les proporcionan una ventaja fundamental en su transformación hacia carcinomas. Estas alteraciones que otorgan una ventaja selectiva son mutuamente excluyentes en carcinomas bien diferenciados, mientras que en los tumores indiferenciados además de las mutaciones conductoras acumulan otras adicionales posteriores, denominadas eventos tardíos [116, 119].

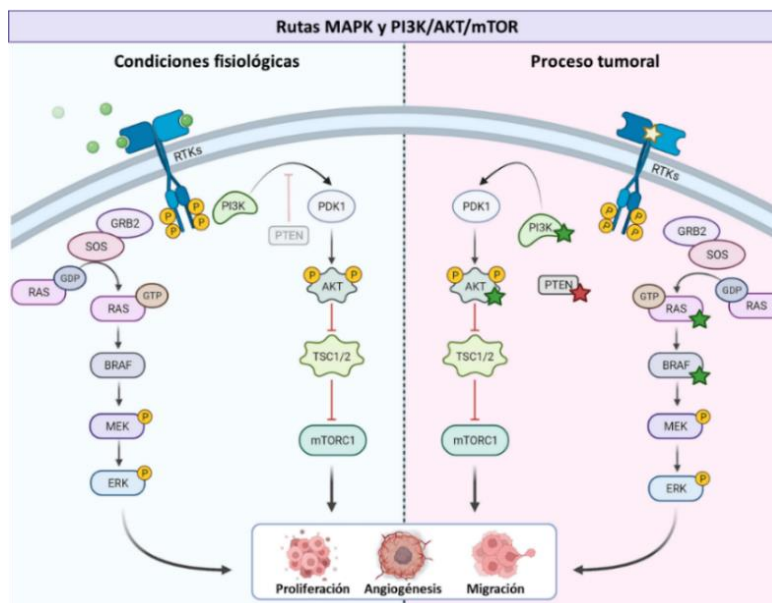


Figura 15 | Vías MAPK y PI3K/AKT/mTOR en el cáncer de tiroides versus en condiciones normales. Tras la activación de varias tirosín quinazas (RTK) a través de ligandos extracelulares, se dimerizan y activan la vía MAPK a través de la activación de proteínas RAS y también estimulan la cascada PI3K-AKT a través de PI3K. Ambas vías estimulan la proliferación, la angiogénesis y la migración de células. En el cáncer de tiroides, varios genes de fusión o mutaciones simulan la activación fisiológica de los RTK, incluso en ausencia de ligandos, lo que conduce a una señalización constitutivamente activa de MAPK y PI3K/AKT. Se destacan las fusiones que han acontecido en los genes RET y NTRK1/3 (estrella amarilla), las mutaciones activadoras de PI3KCA, AKT, RAS y BRAF (estrella verde), así como las mutaciones nocivas de PTEN (estrella roja). Imagen tomada y adaptada de Balock y colaboradores (2022) [122].

A pesar de la frecuencia baja de mutaciones somáticas en los tumores bien diferenciados, la naturaleza de las vías implicadas en la tumorigénesis tiroidea es

compleja, con numerosos componentes, múltiples puntos de interferencia, diferentes localizaciones subcelulares y con la capacidad de regular, potencialmente, muchos procesos celulares [123]. Específicamente, las mutaciones puntuales y las fusiones de genes que conducen a la activación de las vías de señalización de la proteína quinasa activada por mitógenos (MAPK) y la fosfoinositol 3-quinasa/proteína quinasa B/diana de rapamicina en células de mamífero (PI3K/AKT/mTOR) son cruciales para la tumorigénesis y la progresión de los tumores tiroideos (**Figura 15**).

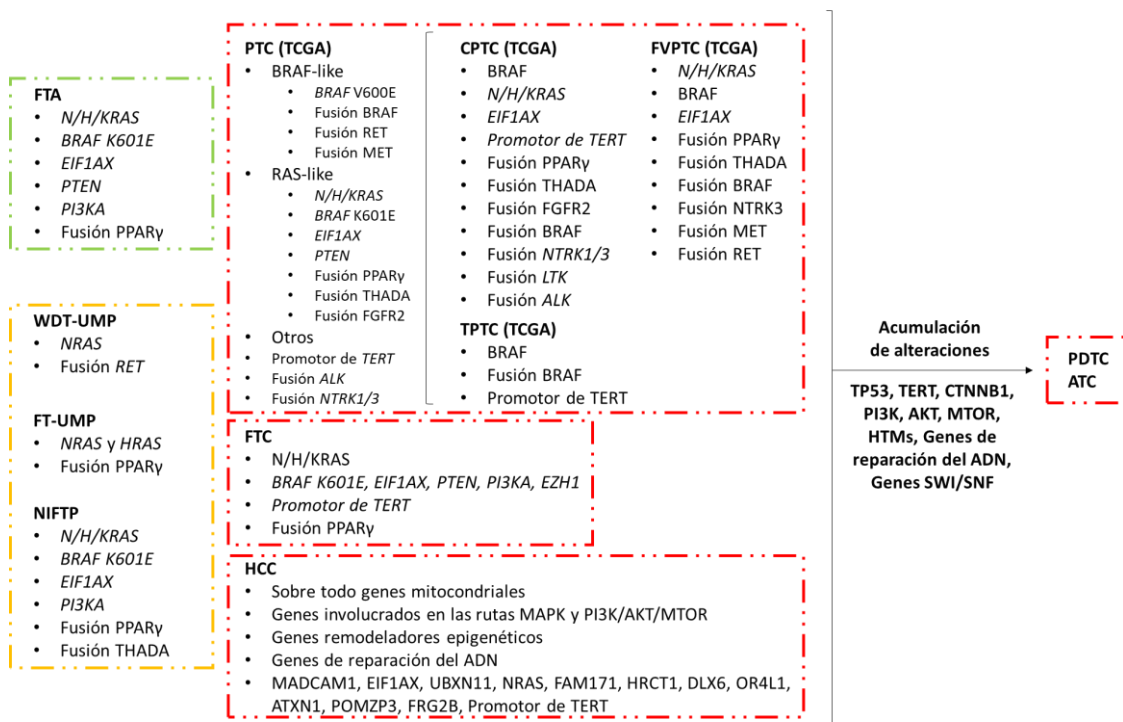


Figura 16 | Resumen de las mutaciones más frecuentes de neoplasias derivadas de células foliculares. Imagen tomada y adaptada de Kakudo y colaboradores (2018) [117].

Por un lado, la cascada de señalización MAPK se asocia al desarrollo de PTC y los elementos afectados incluyen mutaciones puntuales de los genes RAS y BRAF y reordenamientos RET/PTC [117, 123]. Por otro, tanto la activación de MAPK como de la vía PI3K/AKT/mTOR, se relacionan con el desarrollo de FTC, mediante mutaciones puntuales en RAS, PIK3CA, AKT1 y PTEN [117, 123]. Ambas vías de señalización controlan una amplia variedad de procesos celulares esenciales, y su hiperactivación aberrante causa una mayor proliferación, angiogénesis y migración celular, así como a la regulación negativa de la transcripción de genes responsables de la diferenciación de la glándula tiroidea [123]. A pesar de que se ha mencionado que las alteraciones en ambas rutas se relacionan con el desarrollo de PTC o FTC, también se encuentran en diferentes

proporciones en los distintos subtipos tumorales anteriormente descritos en la **Sección 3.3 (Figura 16)**.

RAS es el oncogén mutado con mayor frecuencia en cáncer y se encuentra constitutivamente activado en un tercio de todos los tumores humanos [123]. La familia *RAS* comprende tres isoformas principales codificadas por genes diferentes, a saber, *H*, *K* y *NRAS*. La frecuencia de mutación en los tumores varía según la isoforma siendo la más común *KRAS* (86 %), seguido por *NRAS* (11 %), y finalmente *HRAS*, que es relativamente raro (3 %) [123]. Debido al parecido entre los tres genes cabría esperar funciones similares, sin embargo, *KRAS* activa la vía de las MAPK de manera preferencial [123]. Mientras que *NRAS*, y sobre todo *HRAS*, activan la cascada de señalización PI3K-AKT [123]. Específicamente en las neoplasias de tiroides, las mutaciones en el gen *RAS* ocurren en FTA (30 %), NIFTP (30-50 %), HCA (16 %), PTC clásico (0-35 %) (CPTC), FVPTC (30-45 %), FTC (30-50 %), HCC (19 %), PDTC (20-50 %) y ATC (10-50 %) [98, 116, 117, 124]. Mutaciones en otros puntos de la vía, como *PIK3CA* ocurren tanto tumores benignos (FTA), como malignos (FTC, PTC hobnail, PDTC, ATC) de tiroides [98, 113, 117]. Diferentes fosfatasas ejercen un control negativo sobre la vía, especialmente *PTEN*, que se opone a la función de PI3K al promover la desfosforilación de fosfatidilinositol (3,4,5)-trifosfato [123]. Las mutaciones inactivadoras de *PTEN* se pueden encontrar en tanto en FTA, como en FTC, PDTC y ATC [98, 113, 116, 117].

Por otro lado, las mutaciones en *BRAF* mantienen activa la vía de las MAPK independientemente de señalización extracelular, teniendo un impacto directo sobre el desarrollo tumoral. La alteración más común consiste en la transversión T1799A, que conlleva la sustitución del aminoácido neutro valina por ácido glutámico en la posición 600, dando lugar a la variante V600E [98]. Comúnmente es una alteración encontrada en PTC (30-60 %), siendo más frecuente en las variantes clásica (45-80 %) y de células altas (60-95 %) (TPTC) y menos en FVPTC (5-25 %) [98, 125]. También puede detectarse en las formas más indiferenciadas del cáncer de tiroides ATC y PDTC. Generalmente, son raras en FTC (<2 %) [126, 127], e incluso más en NIFTP, UMPs y FTA. Curiosamente, la sustitución de una adenina por una guanina en la 1801 da como resultado una sustitución de lisina por ácido glutámico en el aminoácido 601, generando la variante *BRAF^{K601E}*. Esta variante rara de *BRAF* puede encontrarse alterada en neoplasias con

patrón folicular tanto malignas (FVPTC, FTC) como benignas (FTA y NIFTP) [98, 117, 128, 129].

Curiosamente, en 2014 se presentaron los resultados del proyecto El Atlas del Genoma del Cáncer (del inglés, *The Cancer Genome Atlas*; TCGA), en el que se analiza una de las cortes más grandes de cáncer de tiroides (en inglés *Thyroid Cancer Atlas*; THCA) [119]. Esta iniciativa cataloga los cambios moleculares de importancia biológica que acontecen en 496 PTC, (incluidos CPTC, FVPTC y TPTC) [119]. Gracias a este trabajo se pueden resumir las alteraciones que ocurren en PTC en un 75 % de mutaciones puntuales, un 15 % de fusiones de genes y un 7 % de variaciones en el número de copias [117]. No solo eso, en este trabajo se describieron dos grupos moleculares en función de sus perfiles de expresión génica, los denominados *BRAF-like* y *RAS-like*, cada uno con características clínicas diferentes [119]. Los tumores *BRAF-like* muestran un patrón de crecimiento papilar, correspondiendo generalmente con los CPTCs y TPTC. Las mutaciones más comunes en este subtipo son $BRAF^{V600E}$ y reordenamientos génicos en *BRAF*, *RET* y *MET* [119]. Por el contrario, los tumores *RAS-like* presentan se correspondían en un 80 % de los casos FVPTC, con mutaciones en los genes *RAS*, *EIF1AX* y *PTEN*, y reordenamiento de los genes *PPARG*, *FGFR2* y *THADA* [117, 119]. Adicionalmente, mutaciones en el promotor de *TERT* y reordenamientos en *ALK*, *NTRK1*, *NTRK3* pueden ocurrir en ambos subtipos moleculares [119]. Los tumores WDT-UMP y NIFTP, al haber sido reclasificados, también pueden presentar un perfil mutacional *RAS-like* [98]. En 2016, Yoo y colaboradores describen en una cohorte, que también incluye FTA y FTC, una tercera clase que denominaron *Non-BRAF-Non-RAS*. Este grupo presentaba mutaciones conductoras principalmente en *DICER1*, *EIF1AX*, *IDH1*, *PTEN*, *PAX8-PPARG* y *THADA-IGF2BP3*, e incluía sobre todo a FTA, FTC y rara vez FVPTC o CPTC [125]. Asimismo, la alteraciones de *EIF1AX* son concomitantes con las mutaciones conductoras de la vía MAPK [119], se puede encontrar generalmente en neoplasias como FTA, NIFTP, FTC, PTC (generalmente FVPTC), HCC, PDTTC y ATC, y raramente en hiperplasias [98, 117, 130].

En cuanto a las reorganizaciones génicas, la fusión de los genes *PAX8* y *PPAR γ* es la lesión más comúnmente asociada con FTC (30-35 %) [123]. Como se comentó, *PAX8* es un factor de transcripción tiroideo importante, ya que participa en la diferenciación de

las células foliculares y en la regulación de una serie de genes específicos del tiroides [59]. Mientras que *PPAR γ* es mejor conocido por su implicación en la adipogénesis y la sensibilización a la insulina [131]. La expresión queda controlada por el gen *PAX8*, y aunque la función de este reordenamiento todavía es objeto de estudio, generalmente estimula la proliferación celular, inhibe la apoptosis y altera el estado de diferenciación celular [131]. La fusión *PAX8/PPAR γ* , aunque con menor frecuencia, se puede encontrar en también FVPTC, PDTC, HCC, NIFTP y FTA [116, 125, 117, 123, 132–136].

Otros genes que pueden ser diana de reorganizaciones en cáncer de tiroides son *RET* y *THADA*. El primero es una tirosina quinasa, que sirve como receptor para ligandos de la familia del factor neurotrófico derivado de la glía. La proteína del protooncogén *RET* se dimeriza y sufre autofosforilación en residuos de tirosina intracelular, lo que permite el reclutamiento de proteínas adaptadoras y la activación de cascadas de señalización como MAPK y PI3K, entre otras [123]. Se han descrito numerosas fusiones, sin embargo, las más comunes ocurren con los genes *CCDC6* y *NCOA*, denominados *RET/PCT1* y *RET/PTC3*, respectivamente [123]. Preferentemente ocurren en PCT, pero también se pueden encontrar en FTA [123]. No parece estar relacionado con la progresión de PDTC o ATC, aunque en el primero sí que se han registrado fusiones *RET/PTC* [116, 123]. Por otro lado, el *THADA* se ha relacionado tanto con el metabolismo como con la adaptación al clima [137]. La fusión de este gen con el *IGF2BP3* conduce al aumento de la traducción de *IGF2* y la señalización del receptor de IGF1 a través de las cascadas PI3K y MAPK, y promueve la proliferación, invasión y transformación celular [138]. Se puede encontrar en FVPTC, NIFTP y FTA [139]

Por otro lado, el 55 % neoplasias de células oncocíticas también sufren mutaciones conductoras en las vías de señalización MAPK y PI3K/AKT/mTOR [124, 140]. Sin embargo, estas neoplasias, tanto adenomas como carcinomas, tienen un perfil molecular distintivo que generalmente la diferencia del resto de los subtipos [114, 124, 140, 141]. Las hiperplasias nodulares con componente de células de Hürthle, al ser no clonales, no presentan alteraciones moleculares, mientras que los HCA no suelen presentar más de una mutación conductora. En cuanto los HCC, las mutaciones más comunes se pueden encontrar en *MADCAM1*, *EIF1AX*, *UBXN11*, *TERT*, *DAXX*, y *NF1*, siendo infrecuentes las mutaciones en *BRAF* (0-5 %) y las diferentes isoformas de *RAS*

(*NRAS*, 9-12 %; *HRAS*, 0-8 %; y *KRAS*, 0 %) [124, 140]. Además, la variación en el número de copias es la alteración genética más habitual en HCC, ocurriendo también en HCA en menor frecuencia [124]. Las mutaciones del promotor de *TERT* y las mutaciones de *TP53* se encontraron más comúnmente en los subtipos más invasivos [124, 140, 141]. Cabe destacar que se encontraron genes alterados relacionados con el daño del ADN, las vías de reparación del genoma, mutaciones en genes relacionados con la remodelación epigenética y mutaciones mitocondriales [140]. Estas últimas explicarían la base de la disfunción mitocondrial que es el sello distintivo de este subtipo tumoral de tiroides.

Por último, mientras que los carcinomas papilares tienen frecuencias bajas de mutaciones somáticas, los ATC se encuentran en el extremo opuesto superando los 100 cambios por megabase [38]. Un peor pronóstico e incluso la progresión del cáncer de tiroides diferenciado a cáncer de mayor grado (PDTC y ATC) se asocia con la acumulación de mutaciones en otros genes oncogénicos, como *TP53*, *PIK3CA*, *AKT1* y el promotor de *TERT* [98, 113, 116]. Alteraciones genéticas en las vías MAPK y PI3K/AKT son concurrentes en este tipo de tumores malignos e impulsan el carácter agresivo de los tumores indiferenciados. Asimismo, las diferencias entre los perfiles moleculares *BRAF-like* y *RAS-like* persisten en PDTC, pero se pierden en gran medida en ATC [116].

3.4.2 Perfil de modificaciones epigenéticas

Se sabe que una única lesión génica, como las mutaciones puntuales, puede alterar la expresión de genes supresores de tumores u oncogenes, siendo clave en la formación de un tumor. Por ello, se consideró en un principio que el cáncer tenía un origen genético. Gracias al avance en el campo de la epigenética, actualmente se sabe que la acumulación de alteraciones epigenéticas también puede inducir cambios en la regulación de la expresión de genes relacionados, por lo que pueden ser también partícipes de la biología tumoral, incluido el cáncer de tiroides

Más adelante, se describe en detalle cómo alteraciones en los perfiles de metilación del ADN se relacionan con el cáncer de tiroides. No obstante, no es éste el único mecanismo epigenético que se ve alterado en las células foliculares. Determinados cambios en las modificaciones postraduccionales de histonas se han observado también durante la carcinogénesis tiroidea. La investigación del patrón global

de acetilación de histonas reveló que los niveles de H3K9-K14Ac son más altos en los tumores tiroideos FTA, PTC, FTC y PDTC con respecto al tejido normal adyacente [142]. Curiosamente, sólo los FTA presentaron valores más elevados de H4K12ac en comparación con el resto de las muestras. Este hecho se destacó porque podría ser usado como marcador para distinguir los adenomas del resto de neoplasias foliculares en punciones citológicas [142]. Asimismo, los niveles de H3K18ac se encontraron más bajos en los PDTC que en los tumores bien diferenciados, por lo que proponen que la reducción de esta marca podría jugar un papel durante la progresión del cáncer de tiroides [142], como ya se ha observado en otros tumores, como carcinomas de mama, colon, pulmón, hígado y páncreas [143]. Por otro lado, los niveles de H3K27me3 podría estar asociado con la agresividad y la desdiferenciación en el cáncer de tiroides, ya que se ha visto sobreexpresado en tumores tiroideos malignos, especialmente en los indiferenciados (PDTC y ATC) [144]. Además, se relaciona con una mayor frecuencia de extensión extratiroidea, invasión linfovascular y metástasis en ganglios linfáticos en carcinomas tiroideos bien diferenciados [144]. A pesar de la existencia de diferentes estudios, se dispone de escasa información sobre el impacto que tiene la remodelación de las modificaciones postraduccionales sobre la expresión génica en la tumorigénesis tiroidea. Son necesarios más estudios para comprender con mayor profundidad las implicaciones de estas alteraciones y explorar, en base a ello, su utilidad pronóstica y diagnóstica en el manejo de estos pacientes.

Asimismo, se estudios recientes han demostrado que los PTC, tanto las variantes clásica y folicular, presentan niveles de 5hmC globales significativamente más bajos en comparación con el tejido normal [145–147]. Por otro lado, en las formas más agresivas del PTC los niveles de 5hmC se asocian con la presencia de metástasis en los ganglios linfáticos [146]. Los ATC, los tumores de tiroides más agresivos y desdiferenciados, exhiben niveles de hidroximetilación marcadamente reducidos en relación con el tejido tiroideo normal [148]. Cabe destacar que, si la información relativa a los niveles de hidroximetilación en tumores tiroideos es reducida, es aún menor en las formas benignas de la enfermedad y los resultados disponibles no son suficientes para ser concluyentes. Mientras que los FTA muestran un estado intermedio de 5hmC entre el tejido normal adyacente y las muestras de PTC, pero sin alcanzar significación estadística

[145]. Por otro lado, 5hmC aumenta significativamente en NH en comparación con PTC [147].

La metilación del ADN es una de las marcas epigenéticas más estudiadas en cáncer de tiroides. Como ya se mencionó, la hipometilación global es una característica distintiva de los tumores malignos, sin embargo, se sabe poco sobre este evento en este tipo de tumores [38]. Los niveles totales de 5mC se pueden cuantificar con precisión mediante cromatografía líquida de alta resolución, sin embargo, se requiere una gran cantidad de muestra, así como un equipo costoso y especializado [38]. En su lugar, se han desarrollado técnicas alternativas, entre ellas los arrays de metilación, que permiten establecer la disposición de patrones específicos en los genomas tumorales.

Estudios pan-cáncer muestran que los PTC exhiben unas frecuencias de eventos de hipometilación e hipermetilación en regiones del genoma próximas a los sitios de inicio de la transcripción de genes más bajas en comparación con otros carcinomas [149, 150]. Otro trabajo similar muestra que los PTC están significativamente hipometilados a lo largo de todo el genoma y en todas las regiones funcionales [151]. Resultados equivalentes se han registrado en proyectos con cohortes independientes, en los que se observó que los PTC se caracterizan por una mayor frecuencia regiones hipometiladas en comparación con el tejido normal o benigno (**Tabla 2**) [152–159]. Solo un trabajo, realizado por nuestro laboratorio en 2013, observa mayor número de sondas diferencialmente hipermetiladas que hipometiladas [160]. Esta discrepancia puede ser debida a que solo se analizaron dos PTC y no se hizo distinción entre variantes.

Por otro lado, aunque numerosos estudios indican la presencia de FVPTC en sus cohortes (**Tabla 2**), pocos son los que estudian sus diferencias con respecto al PTC convencional. Ellis y colaboradores, observaron una fuerte separación en los PTC que presentaban mutaciones somáticas, especialmente evidente para FVPTC, donde la mutación aparentemente estaba estrechamente asociada con su patrón de metilación y la posterior agrupación (2014) [152]. Los FVPTC presentaban más sondas diferencialmente hipometiladas con respecto al tejido no tumoral, pero en comparación con los CPTC esa proporción era seis veces menor. Mancikova y colaboradores (2014), demostraron que los metilomas de los FVPTC eran más similares a FTA y FTC que a CPTC [153]. En cuanto al perfil de metilación de su contraparte benigna, los NIFTP, hay poca

información. Tras su reclasificación en la versión de la OMS de 2017, solo el estudio de Park y colaboradores (2020) [159] ha evaluado las diferencias con respecto al resto de tumores. Observando que estos tumores de potencial maligno incierto albergan más sondas específicamente hipermetiladas que hipometiladas, en comparación con PTC y FVPTC encapsulado invasivo.

En contraposición a PTC, los datos indican que el patrón de metilación de FTC, albergan más hipermetilaciones que sondas hipometiladas (**Tabla 2**) [153, 156, 161, 162]. De manera similar, los FTA presentan una metilación del ADN alterada más modesta con respecto al FTC [153, 156, 161, 162]. Sus niveles de metilación son heterogéneos y se distribuyen en una posición intermedia entre las muestras normales y los carcinomas foliculares, no pudiendo ser completamente separados cuando se realizan análisis de agrupamiento [153, 156, 161, 162].

En relación con los carcinomas más desdiferenciados, nuestro el laboratorio fue el primero en caracterizar una fuerte pérdida de metilación en comparación con el resto de los tumores tiroideos (entre 13-23 veces más sondas hipometiladas que el resto de tumores con respecto al tejido normal) (**Tabla 2**) [160]. Estudios posteriores, realizados con versiones más modernas de los arrays de metilación confirmaron que la hipometilación es un evento común, y que puede estar asociada con la inestabilidad genómica en conjunción con la complejidad de las alteraciones observadas en ATC [156, 163]. Por el contrario, la mayoría de las localizaciones hipermetiladas estaban en islas CpG, lo que indica efectos específicos sobre la expresión génica [163].

De todos los histotipos nombrados, el perfil de metilación de las neoplasias de células oncocíticas es el menos estudiado hasta el momento (**Tabla 2**). La razón principal puede ser que fueran estudiados como variantes de otros tumores tiroideos, y en 2017 se empezaran a considerar como categoría independiente [117]. Las mutaciones ya sea en enzimas relacionadas en la remodelación de las marcas que ocurren sobre la cromatina (55 %) o modificación del ADN (9 %) son comunes en los HCC [140]. Por lo que cabría esperar que si su expresión esta alterada, la remodelación epigenética pueda jugar un papel importante en su desarrollo como tumor. Hay que tener en cuenta algunas limitaciones de este tipo de estudios a la hora de analizar las muestras, ya que su caracterización histopatológica, el grado y estadio del tumor no son fáciles de

INTRODUCCIÓN

comparar entre distintos carcinomas puesto que cada uno tiene diferentes sistemas de progresión tumoral.

Tabla 2 | Resumen de estudios en cáncer de tiroides utilizando métodos para el análisis de la metilación del ADN en todo el genoma. Los métodos para la medición se basaron en arrays de metilación de Illumina que identifican 27.578 (27K), 485.512 (450K) y 866.836 (EPIC) sitios CpG en el genoma y Secuenciación de bisulfito de representación reducida (RRBS) en todo el genoma. Si los códigos de acceso estaban disponibles así se indican en la tabla. Se muestran los sitios diferencialmente metilados tanto hipermetilados (hiper.) como hipometilados (hipo.) separados por el subtipo tumoral según los criterios establecidos en cada publicación referenciada. ATC: carcinoma anaplásico, BTL: lesión nodular benigna; HR: Hazard Ratio; IC: Intervalo de confianza; LT: linfocitosis tiroidea, NT: tejido normal y TC: cáncer de tiroides. (1/3)

Estudio	Técnica	Cohorte (n)	Hiper.	Hipo.	Genes/Clasificadores de interés identificados en las publicaciones	Ref.
1	27K	NT (10), PTC (14)	-	-	<i>HIST1H3J, POU4F2, SHOX2, PHKG2, TLX3 y HOXA7</i>	[150] 2013
2	27K (trabajo de laboratorio)	NT (2)	-	-	<i>ADAMTS8, HOXB4, ZIC1, KISS1R, INSL4, DPPA2, TCL1B, NOTCH4 y MAP17</i>	[160] 2013
		PTC (2)	309	14		
		FTC (2)	408	24		
		ATC (2)	114	328		
		MTC (2)	148	490		
		Líneas celulares (4)	548	637		
3	27K (GSE51090)	NT (8)	-	-	<i>COL4A2, DLEC1, KLK10, E124 y WT1</i> Los niveles de metilación de <i>E124</i> y <i>WT1</i> como marcadores de pronóstico ($p < 0,004$; HR 52,08; IC 95 %: 1,262–3,445 y $p < 0,006$; HR 51,64; IC 51,149–2,335, respectivamente)	[153] 2014
		FTA (18)	89	9		
		PTC (42), FVPTC (5)	39	53		
		FTC (18)	460	83		
4	450K	NT (8)	-	-	Hipermetilados: <i>ECAD, DAPK, NIS y TIMP3</i>	[152] 2014
		CPTC (29)	252	2582		
		FVPTC (15)	164	405		
		PTC rec (7)	1023	2796		
5	450K (TCGA)	NT (56)	-	-	Patrones de metilación en PTC agrupan las muestras en 4 grupos: <i>classical 1, classical 2, Meth-follicular</i> y <i>Meth-CpG Island</i>	[119] 2014
		cPTC (324)	-	-		
		FVPTC (99), TPTC (35)	-	-		
		otros PTC (38)	-	-		
6	450K (GSE53051)	NT (12)	-	-	Los bloques hipometilados son una característica universal del cáncer humano	[164] 2014
		FTA (11), HA (10), HCC (8), FVPTC (10), PTC (10), FTC (10)	24130	19857		
7	450K (TCGA)	NT (49)	-	-	<i>CDKN1B, hsa-miR-146, PDGF, SERPINA1, TGFBI, TPO, DUSP5, ERBB3, FGF1, FGFR2, GABRB2, HMGA2.</i> Clasificador de diagnóstico: 93 % de sensibilidad	[154] 2016
		CPTC (49)	165	1061		
8	450K (GSE86961)	NT (41)	-	-	Cuerpo de gen: <i>ERBB3, FGF1, GABRB2, HMGA2</i> ; promotor: <i>FGFR2, RHD5</i>	[155] 2017
		CPTC (29), FVPTC (7), otros PTC (5)	645	5425		

Tabla 2 | Continuación tabla (2/3)

Estudio	Técnica	Cohorte (n)	Hiper.	Hipo.	Genes/Clasificadores de interés identificados en las publicaciones	Ref.
9	450K (GSE97466)	NT (50)	-	-	<p>Hipometilados: <i>OR6K2, LAIR2, PFKFB2, OR4K15, NLRP11, OR9G4, THSD7B, OR2M3, OR52B2, OR2T6, FFAR2, TN3, OR52M1, DCD, ADGRE2, HRH1, CXXC5</i>; Hipermetilados: <i>GPR21, MBP, YPEL4, ATP6VOC</i>.</p> <p>Clasificador de pronóstico: Cohorte de descubrimiento: 63 % sensibilidad, 92 % especificidad; Cohorte de validación (TCGA): 64 % sensibilidad, 88 % especificidad</p>	[156] 2017
		FTA (8), NH (6), LT (3)	1531	222		
		CPTC (47), FVPTC (9), otros PTC (4)	242	2773		
		FTC (8), HCC (2)	4100	1475		
		PDTC (1), ATC (3)	6195	28252		
10	450K	NT (7)	-	-	<p>Los genes hipermetilados con FTC se enriquecieron para los objetivos del complejo represor transcripcional Polycomb y las marcas específicas de histona H3 (H3K4me2/me3-H3K27me3). 7,9 % de las DMP se asocian con genes diferencialmente expresados</p>	[161] 2019
		FTA (10)	0	0		
		FTC (11)	2979	585		
11	450K (GSE97466)	NT (50)	-	-	<p>cg26307926 (intergénico: <i>LINC01143</i>), cg18395809 (cuerpo de gen: <i>PTPRN2</i>), cg24240626 (5'UTR: <i>REG3A</i>), cg06533895 (5'UTR: <i>DCUN1D3</i>), cg05471495 (intergénico: <i>closest gene NPAS2</i>), cg25236791 (cuerpo de gen: <i>WDR82</i>)</p> <p>Clasificador de diagnóstico: Cohorte de descubrimiento: especificidad 82,4 %, sensibilidad 94,3 %, VPP 95,7 %, VPV 77,8 %; Cohorte de validación (GSE53051): especificidad 87,5 %, sensibilidad 97,4 %, VPP 90,2 %, VPV 96,6 %. Cohorte de pirosecuenciación: sensibilidad 90,7 % y especificidad 75,4 %, VPP 89,6 %, y VPV 77,6 %</p>	[165] 2019
		FTA (8), NH (6), LT (3)	-	-		
		CPTC (47), FVPTC (9), otros PTC (4)	302	2713		
		FTC (8), HCC (2)	4126	1449		
		PDTC (1), ATC (3)	-	-		
12	450K (GSE86961, GSE97466 y TCGA)	NT (41)	-	-	<p>cg11228682 (<i>ATXN1</i>), cg01291854 (<i>ELK3</i>), cg06778183 (<i>MCTP1</i>), cg01668008 (intergénico) y cg01702055 (intergénico)</p> <p>Clasificador de diagnóstico: AUC (0,980 en GSE86961, 0,981 en GSE97466, 0,950 en TCGA), PPV 100 %</p>	[166] 2019
		CPTC (29), FVPTC (7), otros PTC (5)	90750			
13	450K (GSE86961 y GSE97466)	NT (91)	-	-	<p><i>hsa-miR-199a-5p, DCN, CXCL12, LRP4 y CA12</i></p>	[158] 2019
		CPTC (76), FVPTC (16), otros PTC (9)	212	2698		
14	EPIC (GSE121377)	NT (7)	-	-	<p>cg10705422 (<i>MICAL2</i>), cg15441605 (<i>β LURAP1L-A</i>), cg24327132 (<i>PKM2</i>), cg16336556 (<i>LTBP1</i>), cg17707274 (<i>MMP7</i>), cg00567113, cg06034194 (<i>LURAP1L-AS1</i>), cg21341586 (<i>EIF4E</i>), cg26849382 (<i>DIAPH1</i>), cg05763918 (<i>LOC100507487</i>)</p> <p>Clasificador de diagnóstico y pronóstico: cg10705422 (AUC 0,83, sensibilidad 76 %, especificidad 80 %), cg17707274 (AUC 0,83, sensibilidad 76 %, especificidad 79 %), cg26849382 (AUC 0,8, sensibilidad 74 %, especificidad 78 %)</p>	[159] 2020
		NIFTP (6), CPTC (11), FVPTC (3), TPTC (7)	1173	3606		
15	EPIC (GSE146003)	NT (4)	-	-	<p><i>MTOR, NOTCH1, HIF1A, MAGI1, TSHR, SCL26A7, TWIST</i></p>	[163] 2020
		ATC (10)	1993	11849		
16	450K (TCGA)	NT (56)	-	-	<p>51 genes impulsados por la metilación del ADN. <i>TREM1, CDH16, BIRC7 y SLC26A7</i> (de riesgo). <i>LPAR5, RDH5 y LIPH</i> (protectores)</p> <p>Clasificador de pronóstico: puntuación de riesgo <i>RDH5, TREM1, BIRC7 y SLC26A7</i>. AUC 0,84 y HR = 1,422, IC95 % 1,081–1,869, p 0,012</p>	[167] 2020
		Cáncer de tiroides (514)	-	-		

Tabla 2 | Continuación tabla (3/3)

Estudio	Técnica	Cohorte (n)	Hiper.	Hipo.	Genes/Clasificadores de interés identificados en las publicaciones	Ref.
17	EPIC (GSE197861)	BTL (10)	-	-	5 CpG como candidatos para biomarcadores entre FTC y nódulos benignos (cg06447474, cg06928209, cg17874802, cg17876578, cg24160381) Clasificador de diagnóstico: Perfecto en la cohorte de descubrimiento. Cohorte de validación: cg06447474 (AUC 0,53; sensibilidad 90 %; especificidad 40 %), cg06928209 (AUC 0,77; sensibilidad 90 %; especificidad 80 %), cg17874802 (AUC 0,62; sensibilidad 100 %; especificidad 40 %), cg17876578 (AUC 0,56; sensibilidad 100 %; especificidad 50 %), cg24160381 (-)	[168] 2022
		FTC (4)	6902	6253		
18	RRBS	NT (3)	-	-	CXCL12, FBLN7, FAM3B, PROX1, COL23A1, GJB3, LAD1, PIWIL1, DNASE1L2, EVPL, COX4I2, LCN12, AGPAT2, DOK7	[157] 2017
		PTC (3)	639	907		
19	RRBS	NT (42)	-	-	Firma de metilación del ADN o DDMS (373 localizaciones genómicas) Clasificador de diagnóstico: Cohorte de descubrimiento: especificidad 96 %, sensibilidad 94 %, VPP 97 %, VPB 92 %; Cohorte de validación: especificidad 97 %, sensibilidad 100 %, VPP 97 %, VPB 100 %	[169] 2019
		BTL (28)	-	-		
		PTC (39)	-	-		
20	RRBS (GSE107738)	NT	-	-	Bloque de haplotipos de metilación (MHB); 70 marcadores de diagnóstico. Clasificador de diagnóstico: AUC 0,994, especificidad 100 % y sensibilidad 92,3 %	[162] 2022
		FTA (36), HCA (10)	7408	315		
		FTC (35), HCC (11)	7582	712		

4 Evaluación diagnóstica inicial de los nódulos tiroideos

El procedimiento estándar implica la valoración bioquímica y física del paciente, y la determinación las características del nódulo tiroideo mediante ecografía, para valorar si se considera necesario o no tomar una biopsia mediante Punción por Aspiración con Aguja Fina (PAAF).

La PAAF es una prueba diagnóstica que se usa rutinariamente en la evaluación inicial de múltiples enfermedades [170], incluidos los nódulos descubiertos en la glándula tiroides. Los primeros reportes sobre el uso de agujas con fines terapéuticos datan de la Andalucía del siglo XI, y se encuentran recogidos en el libro más influyente de la medicina medieval árabe (*Kitab al-Tasrif*) redactado por el médico y científico Abū 'l Qāsim Khalaf ibn 'Abbās [171]. En él, Albucasis (nombre latinizado para el anteriormente mencionado «padre de la cirugía moderna»), describe punciones realizadas en la glándula tiroides con instrumentos que se asemejan a las agujas actuales. Sin embargo, no es hasta el siglo XIX que la punción vuelve a ser utilizada, en este caso por el Barón Guillaume Dupuytren, para diagnosticar quistes hidatídicos [171]. Kün en 1847 describe por primera vez “una aguja exploratoria, que tiene en su extremo

una pequeña depresión con bordes cortantes” usada para diagnosticar tumores [172]. Posteriormente, el desarrollo real de la técnica moderna se lleva a cabo en el siglo XX.

En 1930, Martin y Ellis (**Figura 17**) [173] publican un artículo en el que describen el protocolo seguido, utilizando agujas de calibre 18, para diagnosticar nódulos situados en la cavidad oral, mama, cuello, pulmones, estómago y hueso, así como enfermedades no neoplásicas y otras patologías que categorizaron como misceláneos. Los mismos autores, en 1934, publican un segundo artículo con una cohorte de pacientes más extensa, y en el advocan aún más por la utilidad de la técnica [171]. No obstante, el primero en sugerir el uso de agujas más finas con un calibre 22, es Manheim en 1931 [171]. El establecimiento y la expansión mundial debe atribuirse a los representantes de la Escuela Sueca de Citopatología en 1960 [174] y después se generalizó en América del Norte en la década de 1980 [175]. Hasta este momento la comunidad médica era reticente a usar la técnica debido a varios factores como el temor a los implantes malignos en el trayecto de la aguja, los resultados falsos negativos y las complicaciones graves [171].

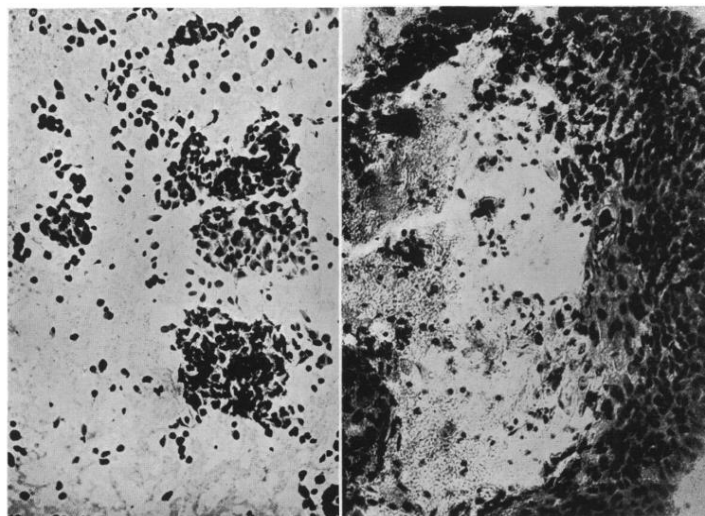


FIG. 7.—Carcinoma of breast. Smear of aspirated material (Case 15). FIG. 8.—Epidermoid carcinoma of metastatic neck node. Smear of aspirated material (Case 24).

Figura 17 | Ejemplos de punciones realizadas por Martin y Ellis en 1930. Las imágenes muestran un frotis de un aspirado tomado de un tumor con carcinoma de mama (izquierda) y una masa metastásica epidermoide procedente de un nódulo del cuello del paciente (derecha). Tomadas del artículo publicado por Martin y Ellis (1930) [173].

En la actualidad, esta técnica es ampliamente utilizada y reconocida como proceder de carácter orientador y diagnóstico en diversos órganos y tejidos, pues proporciona múltiples ventajas tanto al doctor que la realiza como al paciente [170]. Por

ello, la citología juega un papel fundamental en el diagnóstico preoperatorio y tratamiento de los nódulos tiroideos. En términos generales consiste en la extracción, mediante agujas muy finas (calibres de 22 a 25) de una muestra celular que será extendida en un portaobjetos, fijada, teñida e inspeccionada mediante un microscopio por un médico patólogo. La glándula tiroides es un órgano superficial, por lo que es fácilmente accesible para realizar la PAAF y, además, al sumar que puede ser una punción ecoguiada por ultrasonidos, se trata de una técnica sencilla y económica que puede realizarse de forma rutinaria en la consulta sin reportar graves consecuencias para los pacientes [176].

Tabla 3 | El Sistema Bethesda 2017 para informar la citopatología tiroidea: riesgo de malignidad (%) y recomendaciones clínicas. Se distingue entre la inclusión de los NIFTP como carcinomas malignos (CA) y ser considerados tumores indolentes. Tomada y adaptada de Cibas y Ali, 2017 [177].

Categoría	Riesgo de malignidad NIFTP ≠ CA	Riesgo de malignidad NIFTP = CA	Manejo del paciente
No diagnóstico	5-10 %	5-10 %	Repetir PAAF
Benigno	0-3 %	0-3 %	Seguimiento clínico
AUS/FLUS	6-18 %	10-30 %	Repetir PAAF, prueba molecular o lobectomía
FN/FSN	10-40 %	25-40 %	Prueba molecular o lobectomía
SM	45-60 %	50-75 %	Tiroidectomía o lobectomía casi total
M	94-96 %	97-99 %	Tiroidectomía o lobectomía casi total

Finalmente, las células aspiradas se clasifican en una escala de seis grados según el sistema Bethesda (del inglés Bethesda System for Reporting Thyroid Cytopathology; TBSRTC) para proporcionar un informe de la citología tiroidea, con un riesgo creciente de malignidad asociado para cada categoría [176–178] (**Tabla 3**). Esta guía fue publicada por primera vez en 2010 [176], y actualmente es mundialmente aceptada con los siguientes grupos de diagnóstico: Bethesda I o insatisfactoria/indeterminada; Bethesda II o benigna; Bethesda III o atipia/lesión folicular de significado indeterminado (AUS/FLUS); Bethesda IV o neoplasia folicular (FN)/sospechosa de neoplasia folicular (FSN); Bethesda V o lesión sospechosa de malignidad (SM); y Bethesda VI o maligna [176, 177].

Las categorías están en continua revisión, siendo susceptibles a cambios según la experiencia de los citopatólogos y los diagnósticos histológicos postcirugía. A continuación, se describen brevemente las distintas categorías:

❖ Bethesda I. Insatisfactoria/No diagnóstico:

En esta categoría se incluyen las biopsias insatisfactorias, es decir, aquellas que no son válidas para el diagnóstico. Esto puede suceder cuando se aspira un quiste o son muestras hemorrágicas; cuando tienen mala preservación celular o presentan material artefacto y cuando en el aspirado hay insuficientes células foliculares para ser evaluadas. Para que una muestra de PAAF de tiroides sea susceptible para la evaluación, se requieren al menos seis grupos de células foliculares benignas, cada grupo compuesto por al menos 10 células. Por lo general, si esto no se cumple se necesita repetir la biopsia después de la primera [176–178].

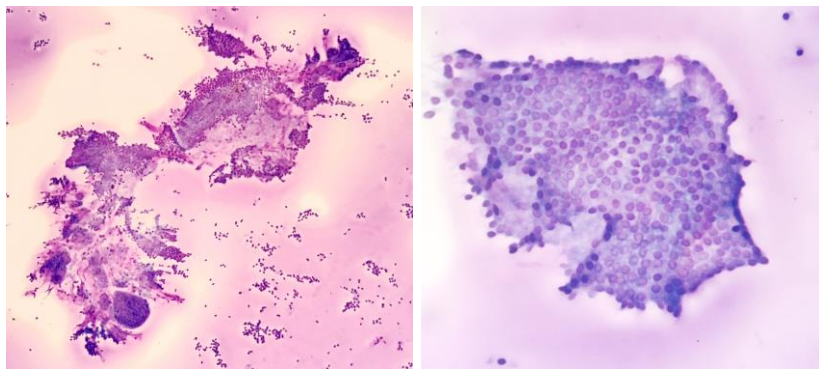


Figura 18 | Imágenes representativas de PAAF con clasificación Bethesda II. Fondo hemático coloide acompañado de aislados macrófagos cargados y escasos grupos de células foliculares en su mayoría de morfología macrofolicular; compatible con hiperplasia nodular. Cedido por el Servicio de Anatomía Patológica del HUCA.

❖ Bethesda II. Benignas:

Esta categoría hace referencia a las punciones aptas su evaluación (**Figura 18**). Son las muestras más comunes en la citopatología tiroidea y abarca a las lesiones con características histológicas benignas que no revisten mayor gravedad: nódulos en bocio nodular, nódulos hiperplásicos o adenomatoides, nódulos coloides, tiroiditis linfocitarias, nódulos compatibles con la enfermedad de Graves y los adenomas macrofoliculares [176–178].

❖ Bethesda III. Atipia/Lesión folicular de significado incierto (AUS/FLUS):

Es el grupo de PAAF indeterminado por excelencia (**Figura 19**). Son aquellos aspirados que, aunque son aptos para su evaluación, presentan algún problema técnico o morfológico para ser clasificada en el resto de las categorías.

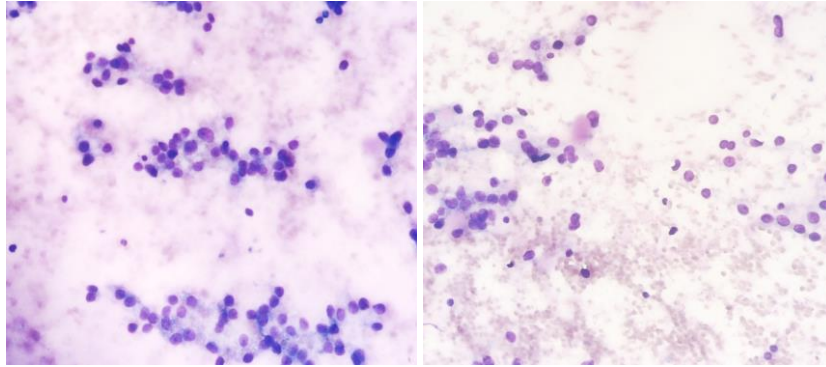


Figura 19 | Imágenes representativas de PAAF con clasificación Bethesda III. En su mayoría macrofolículos, aunque también observando la presencia de microfolículos en moderada cantidad. También se visualiza la presencia de macrófagos muy cercanos a los microfolículos. Los hallazgos citopatológicos plantean el diagnóstico diferencial entre una hiperplasia nodular y una atipia de significado incierto, decantándose por esta última. Se recomendó una nueva PAAF a los tres meses. Cedido por el Servicio de Anatomía Patológica del HUCA.

Debido a la heterogeneidad encontrada, se debe subclasificar la atipia en (1) citológica (en su mayoría nuclear), (2) arquitectónica (microfolicular), (3) citológica y arquitectónica, (4) aspirados de células de Hürthle, (5) atipia no especificada de otro modo y (6) células linfoides (en el que es necesario descartar un linfoma) [176–178]. Por lo general, estas punciones se presentan con pocas células con núcleos agrandados, cromatina pálida, a veces surcos o vacuolización nuclear vaga (que imita pseudoinclusiones nucleares) [179]. Además, algunos casos son claramente benignos clínicamente, pero tienen un componente microfolicular que no se puede ignorar. Los pacientes que sean diagnosticados dentro de este grupo deben ser reevaluados a través de una segunda punción. En este caso la cirugía es una alternativa aceptable cuando la siguientes PAAF realizadas son indeterminadas también [180].

❖ **Bethesda IV. (Sospechoso de) neoplasia folicular o de neoplasia de células de Hürthle (FN/FSN o FN/FSN-HTC):**

Esta categoría indeterminada se define como aquellas PAAF que están compuestas de células foliculares con atipia nuclear leve, la mayoría de las cuales están dispuestas en un patrón folicular arquitectónico alterado, caracterizado por agrupamiento celular, microfoliculos/trabéculas y células aisladas dispersas en un coloide escaso o ausente [177, 178] (**Figura 20**).

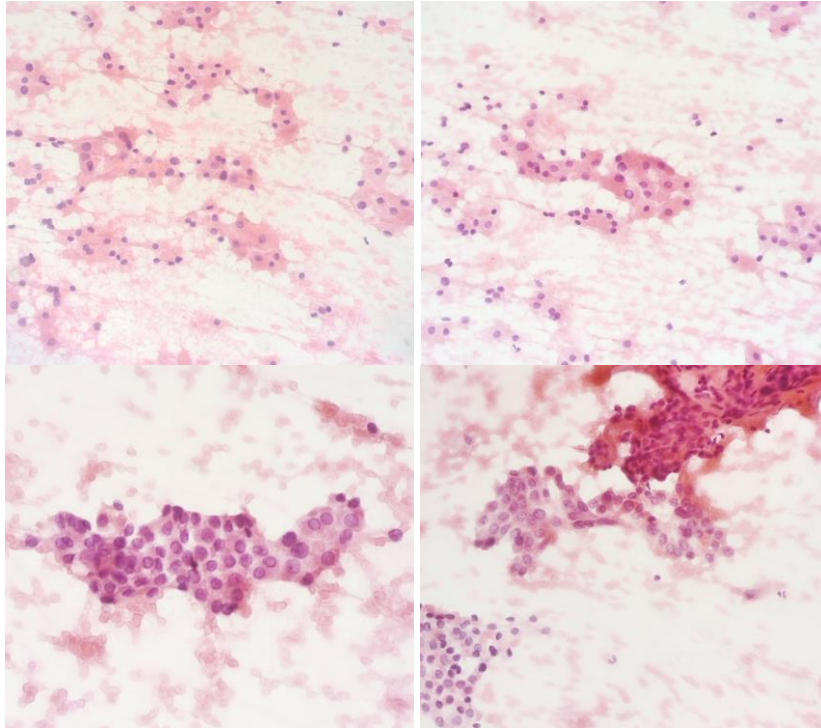


Figura 20 | Panel superior: Imágenes representativas de PAAF con clasificación Bethesda IV. En la muestra se observa un fondo hemático-coloide, sobre el cual se identifica infiltrado mononuclear, apreciando también áreas hiperceulares, constituida por una celularidad epitelial de mediano tamaño, con citoplasmas amplios eosinófilos y núcleos centrales hiper cromáticos ligeramente y cariomegálicos, con cambios oncocíticos. Dicha celularidad se organiza en pequeños grupos macrofoliculares, aunque preferentemente digregados. Se plantea diagnóstico diferencial entre tiroiditis y neoplasia oncocítica, decantándose por esta última. **Panel inferior: Imágenes representativas de PAAF con clasificación Bethesda V.** En la muestra examinada se muestra un fondo hemático-coloide, sobre el cual se evidencia hiper celularidad, dada por la presencia de células foliculares, muchas de ellas formando macrofoliculos, aunque también se observa mucho microfoliculo y células disgregadas. Presenta atipia dada por pseudoinclusiones y hendiduras nucleares. Sospechoso de malignidad para PTC. Cedido por el Servicio de Anatomía Patológica del HUCA. Cedido por el Servicio de Anatomía Patológica del HUCA.

Las células presentan un tamaño normal o ligeramente agrandado, con citoplasma moderado o escaso. Los núcleos bajo tinción son ligeramente hiper cromáticos con nucleolos discretos. Si presentan papilas verdaderas o pseudoinclusiones intranucleares, estas muestras deben ser excluidas de esta categoría [177, 178]. En este grupo también se incluyen las neoplasias de células oncocíticas (presencia de al menos un 75 %), sin embargo, es necesario diferenciarlas en FN/FSN-HTC debido al distinto origen citológico y genético de las mismas [180]. Estas punciones presentan células con citoplasmas granuloso, núcleos agrandados localizados céntrica o excéntrica y nucleolos prominentes.

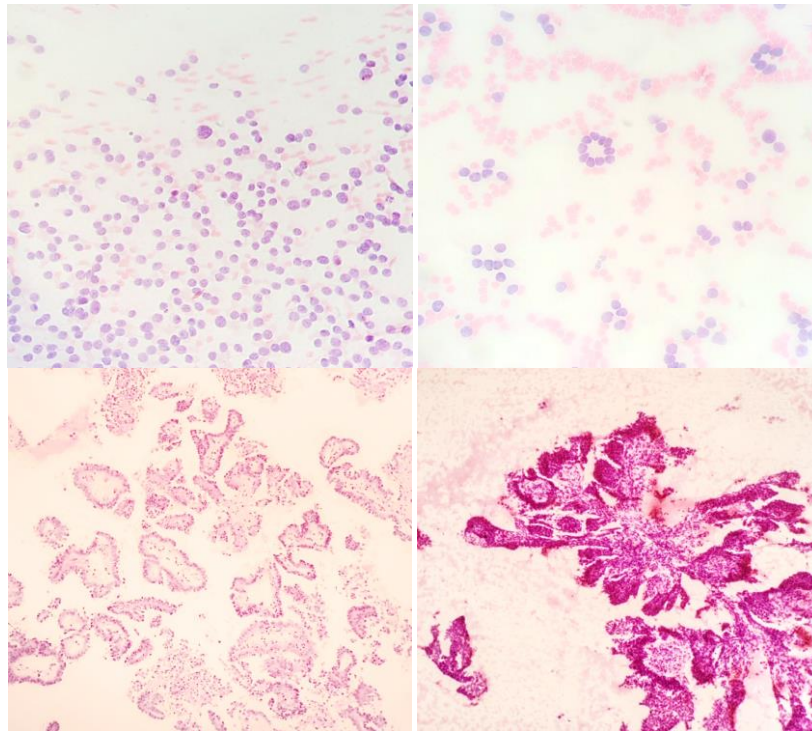


Figura 21 | Imágenes representativa de PAAF con clasificación Bethesda VI. Panel superior: marcado aumento de la celularidad a expensas de células foliculares con distribución suelta o formando microfoliculos. Inicialmente marcada como Bethesda IV, pero dados los antecedentes y las características se considera un Bethesda VI compatible con FTC. Panel inferior: en la muestra examinada se observa un fondo hemático sobre el cual se identifican abundante celularidad, constituida principalmente por células foliculares con morfología papilaroide, en la cual se observan ejes fibrovasculares y, sobre los mismo, la presencia de dicha celularidad. Las células foliculares presentan atipia dada por perdida de la relación núcleo-citoplasma, hiper cromasia, irregularidad nuclear, con presencia de frecuentes hendiduras nucleares y pseudoinclusiones. Representa un Bethesda VI compatible con un PTC. Cedido por el Servicio de Anatomía Patológica del Hospital Universitario de Asturias.

❖ Bethesda V. Sospechoso de malignidad:

Son muestras sospechosas de malignidad (SN) (**Figura 20**), pero no presentan hallazgos cuantitativa y/o cualitativamente suficientes para concluir el diagnóstico definitivamente [177, 178, 180]. En esta categoría se incluyen las lesiones sospechosas de PTC, MTC, metástasis, linfomas y aquellas con necrosis [177, 178, 180]. Se incluyen también aspirados de nódulos que pueden ser diagnosticados como FVPTC, pero se excluyen de esta categoría las neoplasias de células de Hürthle [177, 178, 180].

❖ Bethesda VI. Maligno:

Todas aquellas punciones con características citomorfológicas concluyentes de malignidad [176, 177] (**Figura 21** **Figura 21**). Presentan extendidos con abundante celularidad y escaso coloide. También se encuentran núcleos grandes con aclaramiento

de cromatina y membrana nuclear marcada [177]. En esta categoría se incluyen el PTC y sus variantes, FTC, MTC, PDTC, ATC, metástasis y linfomas. Los pacientes cuya PAAF sea diagnosticada como maligna, deben someterse a cirugía para retirar parcial o totalmente la glándula tiroides.

4.1 Limitación de las punciones en el diagnóstico exploratorio de los nódulos tiroideos

La PAAF es una técnica precisa que alcanza valores de sensibilidad y especificidad de 65-99 % y 72-100 %, respectivamente [181]. La precisión general de la PAAF se sitúa en el 95 % [182]. Sin embargo, aunque es cierto que es el *gold standard* para evaluar la malignidad de los nódulos tiroideos, también se sabe que presenta ciertas limitaciones.

Las causas de las limitaciones pueden ser extrínsecas, debido a la heterogeneidad que puede existir en la forma en que los citopatólogos perciben, interpretan e informan los aspirados [183]. La decisión de asignar una punción a una u otra categoría es subjetiva y depende de la capacitación, la experiencia y el control de calidad del hospital donde resida el especialista. Sin embargo, el mayor problema es el intrínseco específico del tejido que se está muestreando [183]. Las punciones que entran dentro de la llamada “*zona gris*” no pueden ser diferenciados con claridad, ya que las características citológicas se superponen [184, 185]. La guía Bethesda clasifica a las categorías III y IV como citologías indeterminadas, donde no es posible especificar si un nódulo es benigno o maligno [176, 177]. Aproximadamente el 20-30 % de todas las punciones de tiroides no dan un diagnóstico citológico definitivo [182].

En la mayoría de los casos, la citología tiroidea funciona como una regla de inclusión (del inglés *rule-in*) para los PTC, mientras que para las neoplasias con patrón folicular como una regla de exclusión (del inglés *rule-out*), siendo una prueba de detección más que una diagnóstica [181]. Para todos los nódulos difíciles de categorizar, el diagnóstico final se hace tras la resección de la glándula tiroides, ya que es necesario incluir el nódulo completamente para vislumbrar la presencia o ausencia de cápsula discontinuada y/o invasión vascular (**Figura 22**) [114, 186].

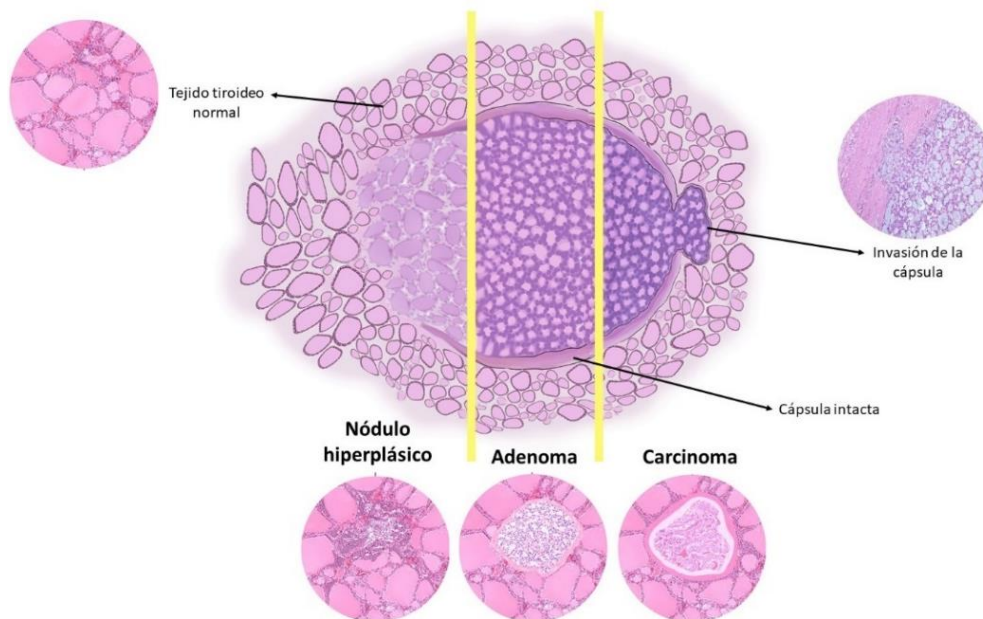


Figura 22 | Imagen representativa de la transición desde nódulo hiperplásico, a adenoma y finalmente carcinoma de tiroides. Tejido normal tiroideo que presenta folículos de tamaño variable según actividad, rodeados de un fino estroma de tejido conectivo rico en capilares fenestrados (junto con los nervios simpáticos que los inervan) y linfáticos. Los nódulos hiperplásicos se caracterizan por una proliferación epitelial folicular benigna bien circunscrita, sin o con pseudocápsula (regresivos) que presentan una citología y una arquitectura heterogéneas, indistinguibles del parénquima circundante. Adenoma folicular representa una neoplasia benigna, encapsulada, no invasiva, que muestra signos de diferenciación folicular con respecto al tejido adyacente. Carcinoma tejido tiroideo maligno con alteraciones biológicas tanto morfológicas como moleculares suficientes para tener un carácter más invasivo (con implicaciones graves para los pacientes), causando irregularidades en la cápsula que los rodea. Las características nucleares y arquitectónicas serán las típicas de cada variante tumoral particular.

En este momento la biopsia postoperatoria es la única técnica diagnóstica que permite diferenciar realmente los distintos tipos histológicos de los tumores tiroideos. Dada la tendencia actual hacia los tratamientos conservadores, el diagnóstico final debe confirmarse mediante técnicas diagnósticas más invasivas y los pacientes suelen ser derivados a cirugía, lo que aumenta el número de tiroidectomías realizadas [178, 181, 186]. En una encuesta llevada a cabo en diferentes hospitales españoles sobre el diagnóstico y tratamiento de los nódulos tiroideos, el 50 % de los profesionales en caso de resultado Bethesda III repetirían la PAAF, mientras que el 35,5 % optaría por que el paciente realizara la cirugía [187]. Además, si la punción era clasificada como Bethesda IV, el 95,8 % optaría por la resección parcial o total de la glándula [187]. Sin embargo, la mayoría de estos nódulos indeterminados (30-75 % de ellos) se clasifican como benignos después de la evaluación histológica final [188–190]. En retrospectiva, para estos

pacientes la cirugía es innecesaria y los pone en riesgo de complicaciones quirúrgicas, hipocalcemia permanente, cambios de voz y terapia con levotiroxina exógena de por vida, entre otros [191, 192].

Por ello, numerosos esfuerzos se están dirigiendo a la búsqueda de marcadores de distinta naturaleza cuyo propósito es ayudar en el diagnóstico de aquellas punciones con citología indeterminada. Algunas técnicas complementarias pueden ser de gran utilidad para definir mejor o confirmar algunos diagnósticos, como la histoquímica, inmunohistoquímica y/o los marcadores moleculares. Dado el impacto económico de estos enfoques en los sistemas nacionales de atención médica y, sobre todo, las consecuencias sobre la calidad de vida de los pacientes, existe una necesidad crítica de desarrollar métodos novedosos y rentables para mejorar la evaluación diagnóstica preoperatoria de los nódulos tiroideos.

5 Marcadores para el diagnóstico de cáncer de tiroides

La profesión médica persigue promover y restablecer la salud, diagnosticar y curar enfermedades aplicando conocimiento especializado, que requiere la mejora y superación permanente de las barreras del conocimiento para prestar la mejor asistencia posible de forma ética y competente [193]. Este compromiso, junto con el continuo avance de la tecnología de biología molecular, han propiciado el cambio de mentalidad desde una medicina más generalista a la medicina personalizada [194]. Esta última, específicamente se refiere al ajuste de los tratamientos a individuos o subgrupos de pacientes en función del uso de biomarcadores específicos de la enfermedad.

Los biomarcadores se pueden definir como casi cualquier medida que refleje una interacción entre un sistema biológico y un peligro potencial, que puede ser químico, físico o biológico. La respuesta medida puede ser funcional y fisiológica, bioquímica a nivel celular o una interacción molecular” [195]. Son desarrollados para cubrir las necesidades diarias de la práctica clínica, siendo su fin último lograr optimizar las estrategias de prevención, detección y tratamiento de la enfermedad que el paciente padece o puede llegar a tener [196]. Además, por definición son objetivos y, por lo tanto, no necesariamente reflejan o caracterizan cómo se siente, funciona o sobrevive un sujeto [195]. Las aplicaciones generales de los biomarcadores son: (1) de detección

temprana para evaluar a los pacientes para detectar la enfermedad en una etapa temprana; (2) de diagnóstico que se utilizan para determinar el trastorno de salud específico del paciente; (3) de pronóstico que ayudan a trazar el curso probable de la enfermedad; (4) predictivos que indican la respuesta probable a un tratamiento particular; y (5) de predisposición que muestran el riesgo de desarrollar una enfermedad [196, 197].

Preferiblemente deben ser binarios, además de fácilmente medibles en una muestra biológica accesible, adaptables a la práctica clínica habitual y con un tiempo de procesamiento corto [197]. Además, existe un compromiso entre sensibilidad (verdadero positivo) y especificidad (verdadero negativo), es decir, si se modifica el valor de corte para obtener mayor sensibilidad, sólo puede hacerse a expensas de disminuir al mismo tiempo la especificidad. La solución será el equilibrio entre los falsos positivos (VPP) y los falsos negativos (VPN), que dependerá de la finalidad de la prueba, es decir, existen situaciones en las que se requiere disponer de una prueba diagnóstica altamente sensible (*p. ej.* tamizaje de enfermedades) o altamente específico (*p. ej.* confirmación de enfermedades). Cuando la prueba diagnóstica pretenda predecir benignidad requerirá un VPN alto, mientras que para predecir malignidad necesitará un VPP alto. Considerando un rango de prevalencia de cáncer de 20 a 40 %, una especificidad de 80 % o más podría considerarse como un estándar óptimo para una prueba *rule-in*, mientras que la sensibilidad mínima para un marcador *rule-out* robusto sería de 90 % [198]. Además, el área bajo la curva (AUC) se utiliza generalmente como resumen de rendimiento del modelo. Se considera que cuanto más esté hacia la izquierda la curva, más área habrá contenida bajo ella y, por ende, mejor será el clasificador. Valores de AUC superiores al 75 % se consideran aceptables y 90 % excelentes [199].

Una ventaja del desarrollo de biomarcadores es que se pueden estudiar a partir de una variedad de tipos de muestras. El tejido directo del lugar afectado es el más analizado hasta el momento, pero también se pueden valorar en sangre, orina, heces y, con menos frecuencia, aliento exhalado, saliva/hisopos bucales, líquido cefalorraquídeo, esputo y otros fluidos corporales [196]. De todos los biomarcadores que se pueden encontrar, en esta tesis se prestará especial atención a aquellos enfoques cuyo fin es mejorar o apoyar el diagnóstico de los pacientes con PAAF indeterminadas.

Hasta la fecha, aunque hay pruebas que muestran un gran rendimiento, no existe un biomarcador molecular confiable para diferenciar los nódulos tiroideos indeterminados con una sensibilidad y especificidad conjunta cercana al 100 % [117].

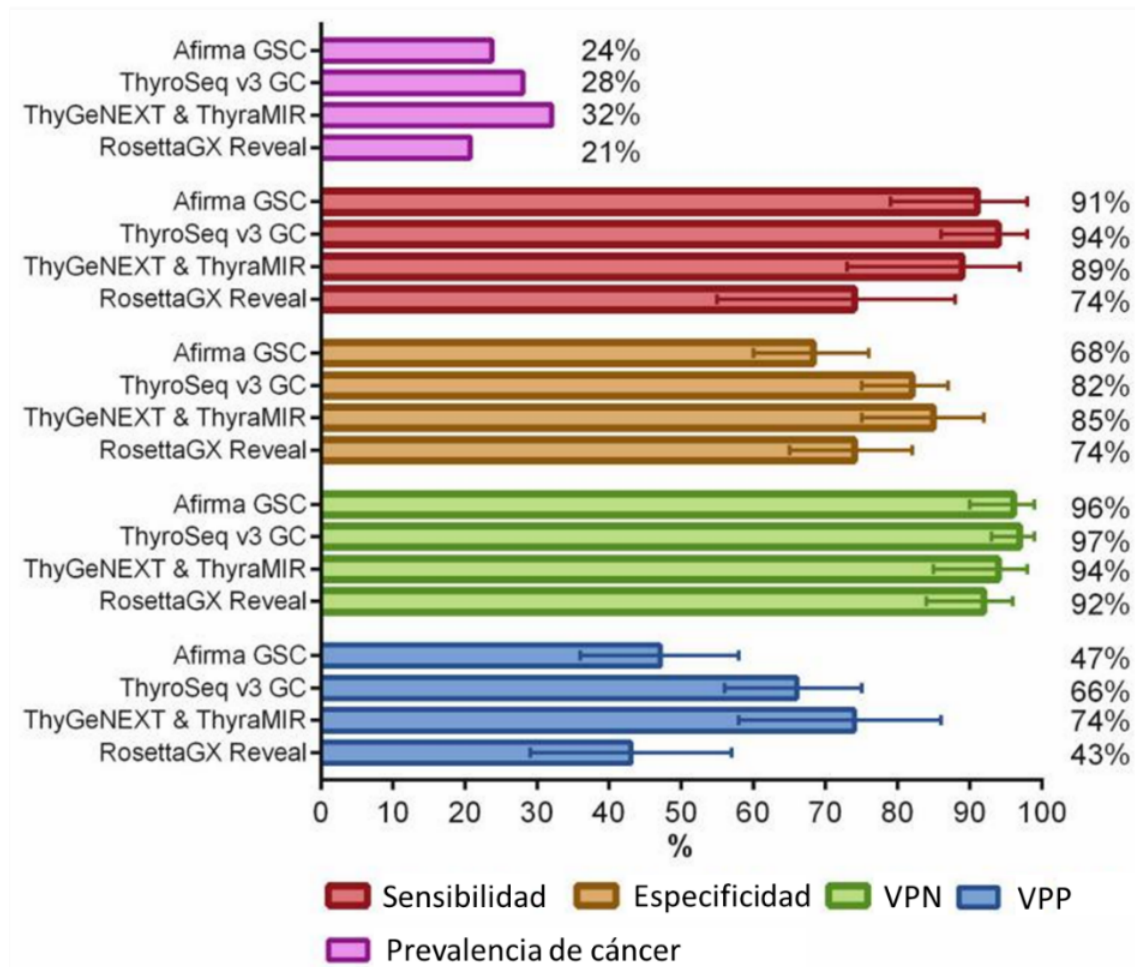


Figura 23 | Resumen del rendimiento diagnóstico para los clasificadores moleculares de nódulos tiroideos con citología indeterminada disponibles en el mercado. La longitud de las barras de error es un intervalo de confianza del 95 %. VPN, valor predictivo negativo; y VPP, valor predictivo positivo. Imagen modificada de Agarwal y colaboradores (2021) [200]

5.1 Marcadores moleculares de diagnóstico en cáncer de tiroides

Las alteraciones genéticas relacionadas con el cáncer de tiroides se han utilizado como marcadores moleculares para distinguir entre lesiones benignas y malignas. Las pruebas comerciales desarrolladas para este propósito, durante los últimos 14 años, se basan en 3 enfoques principales: pruebas de mutaciones somáticas, evaluación de la

expresión génica y clasificadores basados en microARN⁵ (**Figura 23**). Asimismo, los marcadores que se han desarrollado se pueden agrupar en dos categorías diferentes de pruebas moleculares: los métodos *rule-out*, que tienen como objetivo reducir el tratamiento evitable de los nódulos benignos; y las pruebas de *rule-in*, que tienen el propósito de optimizar el manejo quirúrgico.

Como ya se comentó, en cáncer de tiroides la mayoría de las mutaciones conductoras son eventos mutuamente excluyentes, sin embargo, cuando se utilizan de forma independiente no tienen sensibilidad y especificidad suficiente para ser clínicamente relevantes [201]. *BRAF*_{V600E} es altamente predictivo de malignidad, ya que ocurre con una frecuencia relativamente baja en nódulos benignos. Sin embargo, las mutaciones *RAS* se encuentran tanto en adenomas como tumores malignos, por lo que tienen un valor predictivo limitado. En 2009, Nikiforov y colaboradores desarrollaron el primer panel de mutaciones (*BRAF*, *RAS*, *RET/PTC* y *PAX8/PPARγ*) que mejoraba las tasas de sensibilidad y especificidad en el diagnóstico de PAAF indeterminadas [202]. Estos autores reportaron una ganancia en sensibilidad del 44 % en citologías, al 80 % con el análisis molecular; y precisión una del 93,3 % al 97,4 %. Posteriormente, crearon el test molecular denominado ThyroSeq que se ha ido actualizando desde que se publicó por primera vez en 2013 (v1), hasta 2017 que se presentó la última versión (v3). ThyroSeq v3 es una prueba *rule-in* que utiliza un panel de 112 mutaciones puntuales, inserciones/deleciones (InDel), alteraciones del número de copias, fusiones y expresión para clasificar tumores según el material citológico. Es una prueba que requiere de poca cantidad de material genético para poder realizarse y que también ha probado su utilidad en extensiones citológicas de archivo [203, 204]. En muestras PAAF, con resultado quirúrgico conocido, el rendimiento del clasificador posee una sensibilidad del 98,0 %, una especificidad del 81,8 % y una precisión del 90,9 % [203]. Sin embargo, solo se incluyeron 1 FTC, 17 FTA y 2 HCC en el conjunto de validación [203]. Aunque el rendimiento de la prueba puede variar en función de la institución que realice el diagnóstico de las muestras, los desarrolladores de la prueba reportan en los nódulos Bethesda III y IV (combinados) una sensibilidad del 94 % y una especificidad del 82 %

⁵ Los microARN son pequeñas moléculas de ARN no codificante (~22 pares de bases) que pueden modular la expresión génica de forma postranscripcional y, en consecuencia, controlar varios procesos biológicos.

[205]. El 23,43 % de las muestras benignas analizadas fueron positivas tras realizar el ThyroSeq(v3) en ellas, incluyendo nódulos que tras el análisis postoperatorio fueron diagnosticados como HN, FTA, HCA o NIFTP. Mientras que 5 (7,8 %) de las PAAF con resultado negativo, tras ser evaluados los nódulos resultaron ser muestras malignas (PTC y miFTC) [205].

Con el desarrollo de los microarrays capaces de analizar un gran número de genes, surgieron los clasificadores basados en las diferencias de expresión. Uno de ellos es el Clasificador de Expresión Génica Afirma (de inglés *Afirma Gene Expression Classifier*, GEC) que mide la expresión del ARN mensajero de 167 genes para clasificar las PAAF como benignas o mantenerlas como sospechosas [206]. De hecho, esta prueba *rule-out* logró identificar un 90 % de los carcinomas y un 52 % de los nódulos benignos en citologías indeterminadas [206]. Si bien solo el 8,2 % de las PAAF que finalmente se diagnosticaron como nódulos malignos fueron inicialmente clasificadas como benignas. Cuando las muestras benignas fueron analizadas 48,3 % se mantuvieron como sospechosas de malignidad, incluyendo nódulos que tras la operación se diagnosticaron como HN (16,6 %), FTA (15 %), UMPs (6 %) y HCA (9 %) [206]. La versión más avanzada de esta prueba se denomina Clasificador de Secuenciación Genómica de Afirma (del inglés *Afirma Gene Sequencing Classifier*; GSC) e interroga 10.196 genes nucleares y mitocondriales para la medición de recuentos de expresión génica, variantes de secuencia y cambios en el número de copias genómicas (incluida la pérdida de heterocigosis). GSC identificó con precisión el 91 % de los tumores malignos (es decir, sensibilidad) y el 68 % de los nódulos benignos (es decir, especificidad) con citología Bethesda III o IV indeterminada [207]. En este caso, de acuerdo con la categoría histológica, el 31,7 % de los nódulos fue clasificado como sospechoso y el 8,8 % como benigno siendo muestras malignas. La mitad de las veces el clasificador tenía problemas con los FTA, FT-UMP y WDT-UMP, catalogándolas como sospechosas [207]. Para complementar el GSC en aquellos resultados que permanecieran como sospechosas tras analizar los biomarcadores o aquellas con categorías Bethesda V y VI, la compañía desarrollo otra prueba llamada Afirma Xpression Atlas (XA) [208]. Esta incluye 905 variantes genómicas y 235 reordenamientos, cubriendo 593 genes más. Al ser, test *rule-out* intentan contestar a la pregunta de si los nódulos son benignos y, por ello, favorecen

una buena sensibilidad a expensas de la especificidad para asegurar la seguridad de la prueba.

La combinación de ThyGeNEXT® + ThyraMIR®, aúna tanto la evaluación de las mutaciones somáticas como la expresión de genes, fusiones y expresión de microRNAs. Los resultados iniciales del estudio muestran una sensibilidad y especificidad relativamente altas, 89 % y 85 % respectivamente [209]. El procedimiento general (versión del 2022), es que si se diagnóstica una punción de tiroides indeterminada se ejecuta inicialmente la prueba ThyGeNEXT que identifica: (1) por un lado, las mutaciones *ALK*, *BRAF*, *GNAS*, *HRAS*, *KRAS*, *NRAS*, *PIK3CA*, *PTEN*, *RET* y *TERT*; y (2) el panel de ARN identifica el número de fusiones: *ALK*, *BRAF*, *NTRK*, *PPAR γ* , *RET*, *THADA*. Si el resultado de la prueba ThyGeNEXT es negativo para malignidad, entonces se utiliza ThyraMIR para aumentar la sensibilidad y la especificidad generales. Este último analiza 11 microRNAs y en combinación con el panel de alteraciones genéticas dará un resultado agrupado según categorías de riesgo: positivo, riesgo moderado o negativo para malignidad [210]. Los resultados mejoraron con respecto a la versión anterior, obteniendo 99 % VPN en nódulos benignos y 96 % VPP en muestras malignas con respecto a 94 % y 74 % obtenidos anteriormente, respectivamente [209, 211]. Incrementando también la AUC de 0,85 a 0,97 [210].

Por último, la prueba comercial RosettaGX Reveal mide la expresión de 24 miRNAs a partir del material extraído de frotis citológico de la punción tiroidea, para clasificar los nódulos tiroideos indeterminados en: benignos, sospechosos de malignidad o positivos para carcinoma medular. Alcanzan valores de VPN del 92 %, VPP del 43 %, sensibilidad del 74 % y especificidad del 74 %. Los autores apuntan que, si todos los patólogos coincidían en el diagnóstico histológico, se alcanzaba una sensibilidad del 98 % y una especificidad del 78 % en nódulos con Bethesda III-IV. La mejora en el rendimiento del clasificador se debió en parte a la exclusión de 14 FVPTC encapsulados (se consideraron malignos) de la cohorte de validación, 5 de las cuales fueron clasificadas como benignas.

5.2 Marcadores epigenéticos de diagnóstico en cáncer de tiroides

Las modificaciones epigenéticas provocan cambios en el fenotipo de las células sin que estos afecten a la secuencia del ADN. Sin embargo, perturban la estructura y estabilidad del genoma y juegan un papel importante en la carcinogénesis, lo que los convierte en buenos candidatos como biomarcadores.

Específicamente, la metilación del ADN es una de las principales modificaciones epigenéticas y su papel en el inicio, desarrollo y metástasis de cánceres humanos ha sido ampliamente descrito. Por ello, se espera que en un futuro próximo haya un aumento significativo en el número de pruebas basadas en biomarcadores epigenéticos, debido al aumento de la inversión en investigación y al desarrollo de empresas especializadas en ofrecer soluciones para el cuidado de la salud [212].

En la actualidad, en el mercado ya se pueden encontrar más de 100 ensayos clínicos y más de 30 test que han sido aprobados para su comercialización [212]. Han demostrado su eficacia en múltiples aplicaciones clínicas potenciales e incluso para el cribado poblacional, teniendo en cuenta que el uso directo de fluidos corporales, incluso los obtenidos por autorecolección. Ejemplo de ello son las herramientas diagnosticas Cologuard® (Exact Sciences Company) [213, 214] y Epi proColon® (Epigenomics, Inc.) [215]. Cologuard® utiliza un enfoque multiobjetivo para diagnosticar a través de muestras de heces el cáncer colorrectal. En esta aproximación detectan la metilación aberrante en los promotores de los genes *BMP3* y *NDRG4*, la hemoglobina en las heces, la mutación de *KRAS* y *β -actina* (un gen de referencia para la cantidad de ADN humano) [213, 214]. Si un paciente recibía un resultado negativo en la prueba, había un 99,94 % de posibilidades de que no hubiera un carcinoma. Por otro lado, Epi proColon® 2.0 CE es una prueba diseñada para ayudar en la detección temprana del cáncer colorrectal y mide la metilación del gen *SEPT9* a partir de muestras de sangre periférica [215]. Sin embargo, un resultado positivo obtenido (sensibilidad 75–81 % y especificidad 96–99 %) siempre debe ser verificado por otros procedimientos más invasivos, como la colonoscopia o la sigmoidoscopia. Otros ejemplos de aplicación real de marcadores epigenéticos, aunque ninguno está todavía disponible fuera de los ensayos clínicos, pueden ser la prueba diagnóstica PanSeer y el test de Galleri™ que se diseñaron para la detección y localización de múltiples cánceres utilizando firmas de metilación en el ADN

tumoral circulante [216, 217]. El primero, PanSeer, detecta cinco tipos comunes de cáncer (pulmón, colorrectal, esófago, hígado y estómago) en el 88 % de los pacientes después del diagnóstico con una especificidad del 96 %, incluso en 95 % de individuos asintomáticos que fueron diagnosticados hasta cuatro años después [217]. El test de Galleri es capaz de detectar presencia de malignidad en más de 50 tipos de cáncer, con una especificidad del 99 % pero con una sensibilidad pobre del 51 % [218]. Este estudio presenta una gran variabilidad en la sensibilidad dependiendo del tipo de tumor (0-94,4 %), siendo el carcinoma de tiroides indetectable [218].

Si bien varios estudios han demostrado el papel de la metilación aberrante del ADN en la tumorigénesis tiroidea, se han descrito pocos biomarcadores de diagnóstico con un rendimiento aceptable que puedan solventar la limitación que tienen las PAAF para discernir benignidad o malignidad en las punciones indeterminadas. En 2014, Zhang y colaboradores reportaron por primera vez la utilización de los niveles de 5mC como biomarcador en cáncer de tiroides [219]. Combinaron la metilación del promotor de los genes supresores de tumores *CALCA*, *RAR-beta*, *DAPK1*, *TIMP3* y *RASSF1A*, junto con el estado de mutación *BRAF*^{V600E}. Este enfoque pudo distinguir las PAAF benignas de las malignas con una sensibilidad (78 %) y una especificidad (85 %) relativamente altas. Además, la cohorte de validación con estas mismas condiciones se desempeñó aún mejor, alcanzando una sensibilidad del 85 % y una especificidad del 94 %. Posteriormente, Kartal y colaboradores (2014), investigaron la posibilidad de utilizar el estado de metilación del promotor *TSHR* en PAAF como marcador tumoral para diferenciar entre carcinomas tiroideos bien diferenciados y neoplasias benignas [220]. Sin embargo, la sensibilidad (60 %) y la especificidad (66 %) de este enfoque fueron bajas. Un estudio posterior probó *RASSF1A* en punciones benignas [221], y aunque se correlacionaron los niveles de metilación bajos en PAAF de PTC con un tamaño de nódulo más grande, no se muestra precisión y no se proporcionan datos sobre su rendimiento como clasificador. Estudios llevados a cabo en los años subsiguientes, evaluaron el estado de metilación de *PTGS2*, *HOXA1*, *TMEFF2*, *p16* y *PTEN* para comprobar su rendimiento en el diagnóstico de citologías tiroideas indeterminadas. Sin embargo, demostraron baja sensibilidad (45 %) y especificidad (75 %) [222]. Se mostró una mejora cuando se agregó al clasificador el diagnóstico de citología y el estado de

mutación BRAF. Sin embargo, en este intento no logró superar la sensibilidad y especificidad del diagnóstico citológico convencional por sí solo [222].

Por lo general, los biomarcadores basados en la metilación del ADN para establecer clasificadores se obtienen de estudios basados en publicaciones sobre tumorigénesis tiroidea. Sin embargo, los estudios que buscan nuevos marcadores a nivel de metilación global carecen de validaciones posteriores en punciones, limitación que no asegura la aplicabilidad clínica (**Tabla 2**). Un clasificador desarrollado en 2019 por Yim y colaboradores [169], utiliza una firma de metilación del ADN basado en 373 localizaciones genómicas para diferenciar nódulos benignos de malignos. La cohorte en la que está probado y validado este estudio es en PTC con un rendimiento muy bueno en la cohorte de validación con una de especificidad 97 % (80–100 %), una de sensibilidad 100 % (86–100 %), PPV de 97 % (82–100 %) y NPV de 100 % (85–100 %) [169]. Los autores desarrollan una técnica de PCR múltiple para aplicar su clasificador sin necesidad de mucho material genético, pero no se ha probado en otros subtipos tumorales de tiroides ni en PAAF [169]. El único estudio basado en arrays de metilación que tratan de solventar este problema directamente aplicando en PAAF, es el llevado a cabo por Barros-Filho y colaboradores también en 2019 [165]. Identificaron seis sitios CpG que distinguían los nódulos benignos de los carcinomas de tiroides, ya sea diferenciados o no, con una sensibilidad (94 %) y una especificidad (82 %) relativamente altas. La metilación de cinco, de las CpG anteriormente mencionadas, se cuantificó mediante pirosecuenciación con bisulfito de ADN en PAAF y se alcanzó una sensibilidad y especificidad similares a las obtenidas en biopsias de tiroides indeterminadas o malignas [165]. Curiosamente, su aplicación alcanzó una sensibilidad del 100 % y especificidad de 83 % en PAAF indeterminadas, clasificando el 89,5 % de las punciones con Bethesda II como benignas.

HIPÓTESIS Y OBJETIVOS

Debido a las limitaciones de las punciones por aspiración con aguja fina, el manejo óptimo de los nódulos tiroideos con citología indeterminada presenta un dilema clínico. El manejo habitual de estos pacientes recomienda la resección quirúrgica de muchos de estos nódulos por el riesgo de cáncer, aunque la mayoría resultan ser benignos. En retrospectiva, en pacientes con nódulos indolentes, la cirugía conlleva riesgos innecesarios. Dado el impacto no solo sobre la salud de los pacientes, si no también sobre los sistemas nacionales de atención médica, existe una necesidad crítica de desarrollar métodos novedosos y rentables para mejorar la evaluación diagnóstica. Las pruebas moleculares se han convertido en una poderosa herramienta con el potencial de mejorar la evaluación diagnóstica preoperatoria, si bien presentan un gran rendimiento ninguna parece ser totalmente resolutive.

La metilación del ADN es una de las modificaciones epigenéticas más estudiadas. Su papel en el inicio, desarrollo y metástasis de cánceres humanos ha sido ampliamente descrito, incluido en cáncer de tiroides. Dado que los patrones de metilación del ADN pueden ser indicativos de la plasticidad molecular que surge en la tumorigénesis temprana, planteamos **la hipótesis** de que esta marca epigenética podría funcionar como un indicador para la discriminación de nódulos benignos frente a malignos en casos difíciles de diagnosticar. En base a esta hipótesis, **el objetivo general** de esta tesis es la identificación de alteraciones de la metilación del ADN asociados a características diagnósticas que nos permitan complementar el potencial de las técnicas actualmente disponibles y priorizar las intervenciones quirúrgicas. A continuación, se detallan **los objetivos específicos** mediante los cuales se pretende abordar la hipótesis y alcanzar el objetivo general.

- 1. Analizar el perfil mutacional de las lesiones nodulares originadas a partir de células foliculares.**
- 2. Evaluar las alteraciones epigenéticas mediante *arrays* de metilación del ADN.**
- 3. Identificar nuevos biomarcadores que permitan el desarrollo de un clasificador de tipo diagnóstico mediante aproximaciones bioinformáticas.**

MATERIAL Y MÉTODOS

1 Cohortes de pacientes

Las muestras y los datos de los pacientes analizados en este estudio se obtuvieron retrospectivamente de la Red Nacional de Biobancos de España. Los pacientes que participaron en la donación a los Biobancos otorgaron de forma escrita su consentimiento informado y, además, los protocolos y manuales de funcionamiento general fueron aprobados por los Comités externos, científico y ético. Para verificar el diagnóstico y el contenido tumoral de las neoplasias recolectadas por los distintos centros, secciones de tejido adyacente al entregado fueron revisados por, al menos, dos expertos anatomopatólogos en sus centros de referencia y otros dos más en el Hospital Universitario Central de Asturias (HUCA). Las muestras biológicas incluidas en el estudio fueron proporcionadas por los Biobancos que se detallan en la **Tabla 4**, en esta se recogen el número de muestras empleadas de cada uno de los centros involucrados; y en la **Tabla 5** las características de los pacientes que constituyen las distintas cohortes analizadas.

Tabla 4 | Biobancos pertenecientes a la Red Nacional de Biobancos de España que han participado en el estudio.

Centros/Biobancos	Comunidad Autónoma
Instituto de Investigación Biomédica de Salamanca; Biobanco en Red de Enfermedades Oncológicas de Castilla y León	Castilla y León / Salamanca
Instituto de Investigación Biomédica de Bellvitge	Cataluña/Barcelona
Instituto de Investigación Biomédica de Girona	Cataluña/Girona
Instituto de Investigación Biomédica de Lérida	Cataluña/Lleida
Hospital Universitario de La Princesa; Fundación de Investigación Biomédica	Madrid
Instituto de Investigación Hospital Universitario La Paz	Madrid
Hospital Universitario Ramón y Cajal; Instituto Ramón y Cajal de Investigación Sanitaria	Madrid
Hospital Clínico San Carlos (BHCS); Instituto de Investigación Sanitaria del Hospital Clínico San Carlos	Madrid
Biobanco Vasco. Fundación Vasca de Innovación e Investigación Sanitaria	País Vasco
Hospital Universitario Central de Asturias; Instituto de Investigación Sanitaria del Principado de Asturias	Principado de Asturias

Simultáneamente, se realizaron los perfiles genómicos y epigenómicos en un grupo inicial de pacientes que se denominó cohorte de descubrimiento (período de recolección 2001-2017). Esta estuvo formada por un total de 54 muestras procedentes de tumores tiroideos congelados, que incluían muestras de FTA (n = 15), NH (n = 22), FTC (n = 10) y la

MATERIAL Y MÉTODOS

variante folicular de PTC (FVPTC, n = 7). Subsecuentemente, se empleó una cohorte de validación compuesta por 103 muestras de tiroides ultracongeladas (período de recolección 2001–2021: FTA, n = 22; NH, n = 48; FTC, n = 11 y FVPTC n = 22), para corroborar el rendimiento de los marcadores de metilación del ADN seleccionados mediante pirosecuenciación con bisulfito. Asimismo, se ampliaron los grupos de tumores tiroideos estudiados para incluir muestras, como PTC (n = 15), HCA (n = 6) y HCC (n = 1), para evaluar el rendimiento de este sistema de clasificación epigenética en otros escenarios.

Tabla 5 | Datos de pacientes para las cohortes de entrenamiento, validación y punciones de tiroides.

Diagnóstico	n	Edad	Tamaño de nódulo	Puntuación de malignidad
Cohorte de descubrimiento				
NH	22	58,3 ± 11,7	2,3 ± 0,5	0,05 ± 0,15
FTA	15	47,1 ± 14,0	3,4 ± 0,8	0,29 ± 0,29
FTC	10	57,8 ± 18,1	6,5 ± 4,3	0,87 ± 0,19
FVPTC	7	46,8 ± 18,3	2,2 ± 1,1	0,84 ± 0,29
Total	54	53,6 ± 15,1	3,4 ± 1,9	0,39 ± 0,40
Cohorte de validación				
NH	48	53,1 ± 14,0	2,4 ± 0,8	0,12 ± 0,13
FTA	22	48,9 ± 15,5	3,3 ± 1,2	0,29 ± 0,38
FTC	11	45,6 ± 15,1	4,0 ± 2,3	0,68 ± 0,26
FVPTC	22	46,4 ± 17,1	4,0 ± 2,3	0,87 ± 0,17
HTA	6	57,1 ± 9,3	3,0 ± 0,9	0,21 ± 0,38
HTC	1	87 ± 0,0	4 ± 0,0	0,99 ± 0,0
micro-PTC	1	83 ± 0,0	0,6 ± 0,0	0,37 ± 0,0
PTC	14	56 ± 14,2	2,5 ± 1,5	0,77 ± 0,23
Total	125	51,1 ± 15,4	2,4 ± 1,1	0,42 ± 0,39
Cohorte de punciones tiroideas				
Pruebas de extracción de material genético de punciones de tiroides				
PAAF	24	-	-	-
Citologías	5	-	-	-
PAAF + Citología	23	-	-	-
Total	52	-	-	-
Pirosecuenciación de PAAF				
Bethesda II	20	52,4 ± 11,8	-	0,21 ± 0,19
Bethesda III	8	49,5 ± 10,0	-	0,37 ± 0,27
Bethesda IV	1	60,0 ± 0,0	-	0,38 ± 0,0
Bethesda V/VI	4	58,0 ± 13,9	-	0,74 ± 0,47
Total	33	52,6 ± 11,4	-	0,32 ± 0,30

Finalmente, para validar la utilidad del clasificador y evaluar su viabilidad en la práctica clínica, se recolectaron PAAF (período de recolección 2018-2021; n = 85). En este grupo, se almacenaron muestras citológicas de pacientes que acudieron al Servicio de Endocrinología y posteriormente se derivaron a los Servicios de Anatomía Patológica y Otorrinolaringología del HUCA, para la realización de la PAAF, clasificación y seguimiento. Se obtuvo el consentimiento informado por escrito de todos los participantes y el estudio se realizó de acuerdo con los principios de la Declaración de Helsinki para la investigación en humanos. El protocolo fue aprobado por el Comité Ético del Hospital Universitario Central de Asturias (Proyecto nº 98/18, Oviedo, Asturias, España). Se evaluaron y clasificaron las citologías de cada nódulo prospectivamente (n = 85), de acuerdo con el TBSRTC [177] obteniendo las siguientes clasificaciones: Bethesda II (n = 72), Bethesda III (n = 8), Bethesda IV (n = 1), Bethesda V (n = 2) y Bethesda VI (n = 2). El diagnóstico anatomopatológico de las PAAF recogidas se incluye en la **Tabla 5**.

2 Procesado de las muestras embebidas en el compuesto preservador frente la congelación

Se solicitó a los biobancos una biopsia *punch*⁶ con un tamaño mínimo de 2 mm de diámetro o 4-5 cortes histológicos realizados con el microtomo. Todas las muestras de la cohorte de descubrimiento y validación se encontraban embebidas en el compuesto preservador frente a la congelación (del inglés *Optimal Cutting Temperature compound*; OCT). Por lo tanto, para ser empleadas en las diferentes técnicas primero se procedió a retirar la mayor cantidad posible de OCT de las muestras. Para ello, se sumergieron en tampón fosfato salino, se posicionaron en un termobloque a 37 °C y 500 r.p.m. de agitación durante 30 min. Pasado este tiempo, las biopsias se centrifugaron a 2000 r.p.m. y se retiró el OCT que estaba disuelto en el sobrenadante. Finalmente, las muestras se trasladaron a un nuevo tubo donde se realizó un segundo lavado con tampón fosfato salino y se centrifugaron a 2000 r.p.m. para retirar el sobrenadante. El material se almacenó a -80 °C hasta su utilización.

⁶ Biopsia punch o biopsia en sacabocados se realiza para extraer un pequeño crecimiento extirpando un cilindro de tejido con un sacabocados.

3 Procesado de las punciones por aspiración con aguja fina

Las Punciones por Aspiración con Aguja Fina (PAAF, n = 80) se recogieron bajo el consentimiento del paciente y sin comprometer el diagnóstico final de este. Cuando fue posible la obtención de la muestra se realizó con un pinchazo extra dentro de los cuatro recomendados dentro del protocolo del HUCA. A continuación, la PAAF del paciente junto con el lavado de la aguja se almacenaron inmediatamente a -20 °C y, posteriormente, se trasladaron al laboratorio para ser guardadas hasta su utilización a -80 °C. Si las PAAF eran recogidas en un volumen superior a 500 µl, estas se separaban en varios tubos del mismo volumen y el material genético se juntaba en el último paso del protocolo que a continuación se detalla.

4 Procesado de las muestras de las extensiones citológicas

Las extensiones citológicas de archivo procedentes de PAAF de tiroides (n = 28) se recogieron bajo el consentimiento ético otorgado por el HUCA para la realización de este proyecto. Las preparaciones histológicas antiguas en todos los casos fueron diagnosticadas previamente por patólogos especializados. Las extensiones citológicas estaban procesadas y teñidas según el protocolo del hospital y se almacenaron a temperatura ambiente hasta su utilización.

Para despegar los cubreobjetos, los cristales se sumergieron en xileno a temperatura ambiente. Se intentó no dejar la muestra durante periodos largos de tiempo, y por ello, se revisaban las muestras cada 24 h, tratando de despegarlo. El tiempo total fue entre 24 y 60 h dependiendo de la citología. Para retirar el cristal se despegaba con un bisturí uno de los lados cortos del cubreobjetos y a partir de ahí continuar con el movimiento de vaivén, hasta llegar al otro extremo, con la hoja introducida prácticamente en toda su longitud. Una vez nos quedamos solo con el portaobjetos, con el mismo bisturí se raspo la superficie del cristal para recoger el material en un tubo limpio. Para facilitar el proceso y reducir el riesgo de perder muestra, estas se humedecían con una gota del mismo tampón que se usó para la extracción de ADN.

5 Extracción del ADN genómico

La extracción de ADN genómico de las biopsias pertenecientes a las distintas cohortes de estudio se llevó a cabo empleando un protocolo estándar. Inicialmente, el punch de tejido procesado o los 500 μ l de la citología celular se resuspendieron en 500 μ l de tampón de lisis (SDS al 0,5 %, EDTA 25 mM; Tris 20 mM pH 8; NaCl 0,1 M) (**Figura 24**).

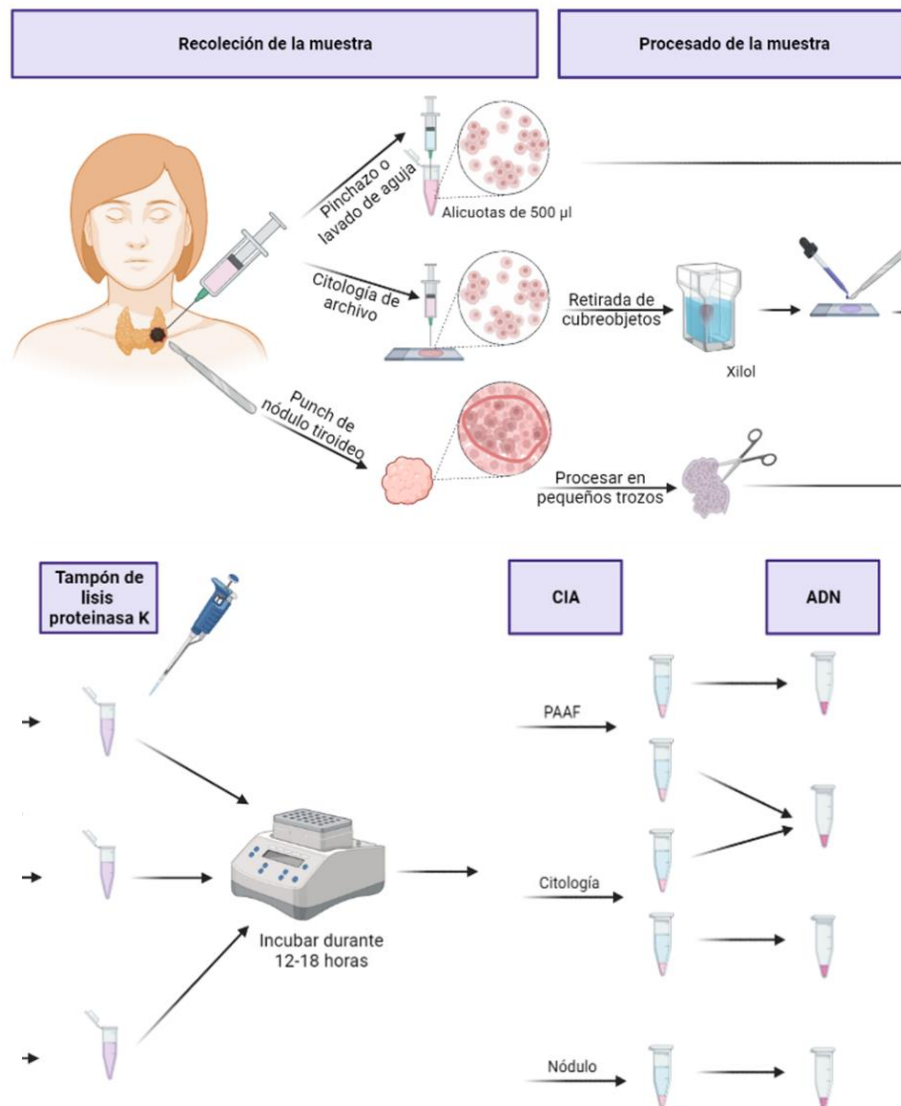


Figura 24 | Resumen del proceso de extracción de ADN genómico según los distintos tipos de muestras biológicas. La obtención del material genético de las punciones se obtuvo a partir de lavados de aguja, citologías de archivo o una combinación de ambos. Para la cohorte de biopsias nodulares se obtuvo el material genético a partir de los punches previamente proporcionados por los biobancos.

Posteriormente, se añadió 15 μ l de proteinasa K (10 mg/ml) y se incubaron de 12 a 18 h a 55 °C. Se aisló el ADN genómico de los residuos proteicos y demás impurezas mediante una extracción estándar con fenol:cloroformo:alcohol isoamílico (CIA). A cada tubo *Phase Lock Gel* (Quantabio, Beverly, Massachusetts, Estados Unidos) se le añadieron 500 μ l de CIA (25:24:1)

y se situaron en una noria durante 15-30 minutos. Posteriormente, se centrifugaron durante 5 minutos a 12.000 r.p.m., provocando la separación de los componentes. En la capa superior, la fase acuosa, se encuentra el ADN mientras que, en la capa orgánica situada por debajo de la interfase quedan las proteínas. Tras esto se retiró la fase acuosa a un nuevo tubo de 1,5 ml donde se llevó a cabo la precipitación y purificación del material genético. Para ello, se añadió 1 μ l de glicógeno (20 mg/ml) y 50 μ l de acetato sódico (3 M, pH 5,3). A continuación, se adicionó a la disolución 1 ml de etanol absoluto frío (4 °C) para eliminar la concentración residual de sales y promover la precipitación del ADN, ya que al alcohol va a deshidratar el ADN. La precipitación podría ser inmediata, pero cuando no es así, se puede facilitar manteniendo los tubos a -80 °C. Se centrifugó (20 minutos a 13.000 r.p.m.) el tubo con etanol para recuperar el ADN. El *pellet* que se obtuvo se lavó con 1 ml etanol al 70 % (tras eliminar el sobrenadante) para tratar de retirar todas las sales que permanezcan en la solución. Se volvió a centrifugar (5 minutos a 13.000 r.p.m.) para poder retirar el sobrenadante, dejando secar las trazas de etanol que pudieran quedar. Por último, dependiendo del tamaño del *pellet* obtenido se añadieron de 25 a 100 μ l de agua ultrapura autoclavada. Para facilitar la disolución del ADN genómico en solución acuosa, se atemperaron los tubos 15 minutos a 37 °C. El ADN se mantuvo a -20 °C hasta su utilización.

6 Cuantificación y calidad del ADN

La calidad y cuantificación del ADN se midió mediante dos técnicas diferentes. Primero, con el espectrofotómetro *NanoDrop™ 1000* (Thermo Scientific, Waltham, Massachusetts, Estados Unidos). El instrumento mide la densidad óptica a 260 nm y 280 nm, los valores obtenidos deben estar comprendidos entre 1,8 y 2, para poder trabajar con ellos (valores por encima significa contaminación por sales y por debajo por proteínas o solventes orgánicos, como el fenol). Finalmente, se realizó una cuantificación mediante el ensayo *Qubit® dsDNA BR* en el fluorómetro *Qubit® 2.0* (Life Technologies, Carlsbad, California, Estados Unidos) de acuerdo con las instrucciones del fabricante. Ambos instrumentos nos dan la concentración de ADN en ng/ μ l y las cuantificaciones obtenidas con dichas técnicas se utilizaron para enviar el material genético a las empresas encargadas de las caracterizaciones de los perfiles mutacionales y epigenéticos.

7 Caracterización del perfil mutacional

7.1 Secuenciación del exoma

La secuenciación de exoma completo fue realizada por Macrogen, Inc (Seúl) usando el protocolo *SureSelect Human All Exon v6 Target* y la plataforma de *Illumina Novaseq 6000* 150PE (100 x 2 pares de bases). Para la realización de la secuenciación se partió de 200 µg de ADN genómico siguiendo las especificaciones del fabricante para muestras con poca cantidad de material. Los subsecuentes pasos del análisis se realizaron de acuerdo con las recomendaciones de buenas prácticas de *GATK* (v.4.1.8) [223]. Se muestra un resumen gráfico de los procedimientos realizados para la caracterización del perfil mutacional de las muestras de la cohorte de descubrimiento en la **Figura 25**.

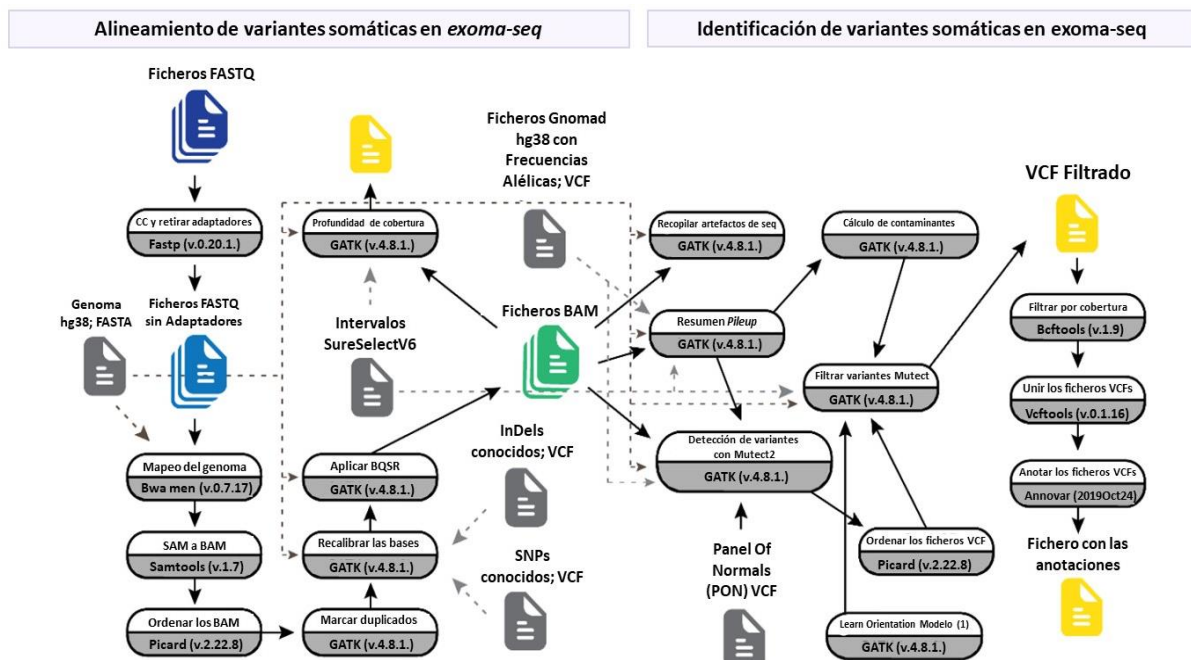


Figura 25 | Información adicional relacionada con el protocolo seguido para la caracterización mutacional mediante secuenciación del exoma en las muestras de la cohorte de descubrimiento. El gráfico representa el procedimiento computacional y el software utilizado para la identificación de mutaciones nocivas en muestras de cáncer de tiroides.

Brevemente, el control de calidad y el recorte de los adaptadores de las lecturas de los archivos FASTQ sin procesar se realizaron con *Fastp* (v.0.20.1) [224]. Las lecturas de secuenciación resultantes se alinearon contra el genoma de referencia humano (hg38) utilizando *Burrows-Wheeler Aligner Bwamem* (v.0.7.17) [225]. Los archivos *Sequence Alignment Map* (SAM) se convirtieron a archivos *Binary Alignment Map* (BAM) utilizando *Samtools* (v.1.7) [226] y estos últimos se ordenaron utilizando el kit *PicardTools* (v.2.22.8)

[227]. Se marcaron los duplicados de la Reacción en Cadena de la Polimerasa (del inglés, polymerase chain reaction; PCR) y se realizó un paso de recalibración de la calidad de las bases utilizando datos de inserciones y deleciones conocidas (1000G InDel de referencia) y polimorfismos de un solo nucleótido (del inglés, *Single-Nucleotide Variants*; SNP) conocidos (dbSNP 138) obtenidos del paquete de recursos *GATK*. Las métricas de cobertura, los artefactos de secuenciación, el resumen de los recuentos de lectura que respaldan los alelos de referencia alternativos y otros para sitios determinados y el cálculo de los contaminantes en la muestra se realizaron para proceder con los análisis subsecuentes.

Los SNP se identificaron utilizando *Mutect2* (v.4.1.8) [228] por medio de los archivos BAM para cada paciente. Este programa está diseñado para llamar variantes somáticas solamente y para ello se le proporcionó los datos recogidos en la base de datos curada *gnomAD*, que almacena las frecuencias alélicas de datos de secuenciación de exomas y genomas de varios estudios genéticos de población y de enfermedades. Además, se también se cruzó con las variantes somáticas que se recogen en el denominado *Panel of Normal*⁷ (PON), obtenidos del proyecto 1000 Genomas. Los ficheros *Variant Call Format* (VCF) resultantes se ordenaron y se calculó el modelo de orientación de lectura (del inglés, *Learn Orientation Model*) para su uso posterior en el paso de filtrado de variantes.

Las variaciones detectadas por *Mutect2* se filtraron utilizando los resultados del paso de contaminación calculado, el modelo de orientación de lectura y la lista de intervalos de SureSelect V6. Se aplicaron filtros simples a los archivos VCF resultantes. Brevemente, se analizaron las variantes que pasaban los filtros estándar en *MuTect2* y, para fines posteriores, se utilizó *bcftools* (v.1.9) [229] para seleccionar aquellas entradas con una cobertura de profundidad mínima > 35 y una frecuencia alélica mínima > 0,15. Los archivos VCF filtrados de todas las muestras se combinaron con *vcftools* (v.0.1.16) [230] y las variantes resultantes se anotaron con *ANNOVAR* (v.2019Oct24) [231].

Para definir las variantes patógenas se utilizaron SNV no sinónimos con ganancia de regiones no traducidas (del inglés, *untranslated trailer*; UTR) con un MAF < 0,01 utilizando anotaciones del proyecto 1000 Genomas (versión de agosto de 2015). Los efectos funcionales de estas variantes potencialmente nocivas se evaluaron con los algoritmos *Polyphen2* [232],

⁷ Panel of Normal o PON es un tipo de recurso utilizado en el análisis de variantes somáticas. Lo que todos los PON tienen en común es que (1) están generados a partir de muestras normales (en este contexto, "normal" significa derivado de tejido sano que se cree que no tiene alteraciones somáticas) y (2) su propósito principal es capturar artefactos técnicos recurrentes para mejorar los resultados del análisis de variantes somáticas.

MutationTaster [233] *SIFT* [234] y se incluyeron también las variantes identificadas en la base de datos COSMIC (v70) [235].

Adicionalmente, se investigaron las mutaciones en el gen *BRAF* manualmente a través de la herramienta *Integrative Genomics Viewer* (IGV), ya que ninguna de estas alteraciones se observó inicialmente debido al uso de criterios de filtrado estrictos. Las inserciones y deleciones que causaban cambio en el marco de lectura, así como aquellas variantes identificadas como mutaciones perjudiciales en los análisis anteriores, se consideraron como posibles mutaciones patógenas para análisis posteriores.

7.2 Determinación de mutaciones somáticas: BRAF, KRAS, NRAS y HRAS

Para detectar las mutaciones somáticas en los genes *BRAF*, *KRAS*, *NRAS* y *HRAS* se llevó a cabo una PCR cuantitativa a tiempo real (RT-qPCR) en un equipo 7500 Applied Biosystems con el *Thyroid Cancer Mutation Detection Kit* (Entrogen, Los Ángeles, California, Estados Unidos) y siguiendo las instrucciones del fabricante. Este kit detecta cambios en nucleótidos de una base causando las siguientes mutaciones: *BRAF* p.V600E; *KRAS* p.G12A, p.G12D, p.G12V, p.G12C, p.G12S, p.G13D; *NRAS* p.Q61H, p.Q61L, p.Q61K, p.Q61R and *HRAS* p.G12V, p.G13R and p.Q61R. Brevemente, el ADN genómico de la cohorte de validación extraído anteriormente (FTC, n = 11; FVPTC, n = 22 y FTA, n = 22), fue cuantificado específicamente mediante fluorimetría empleando el equipo *Quantus Fluorometer* (Promega, Madison, Wisconsin, Estados Unidos). Tras ello, se utilizaron 25 ng de ADN genómico de cada muestra para llevar a cabo cada una de las carreras a una concentración de 5 ng por reacción. Las mutaciones anteriormente mencionadas se detectan en cinco grupos definidos por cinco reacciones, y no se distingue ninguna mutación específica dentro de cada conjunto. Se debe mencionar que el límite de detección (del inglés *Limit of detection*, LOD) del ensayo varía según cada mutación en un rango de 0,25 % a 10 %.

8 Caracterización del perfil de metilación

8.1 Procesado de los datos brutos

El perfil de metilación del ADN se caracterizó empleando la plataforma *Illumina's High Infinium HumanMethylationEPIC 850K Beadchip* [236] siguiendo el protocolo de metilación de *Illumina Infinium HD*. Los servicios de hibridación fueron proporcionados por Oxford Genomics

Center (The Wellcome Center for Human Genetics, Universidad de Oxford, Reino Unido). En la **Figura 26(a-b)** se muestra el flujo de trabajo detallado con la estrategia seguida para analizar los datos de metilación.

Las muestras que se sometieron a este análisis fueron las biopsias de tiroides obtenidas de los biobancos (FTC, n = 10; FVPTC, n = 7; FTA, n = 15 y NH, n = 22), muestras de tejido normal⁸ procedentes del proyecto ENCODE (NT, n = 4) [237], muestras de origen hematopoyético (células madre hematopoyéticas, HCS; células B, LB; células T CD8, LTC; células T CD4, LTH; células asesinas naturales, NK; monocitos, M; neutrófilos; N; en total n = 7) [238] y por motivos de validación posterior los datos experimentales de distintos subtipos tumorales de tiroides (GSE121377: PTC, n = 11; IEFVPTC, n = 3; TCVPTC, n = 7 y NT, n = 37) [159].

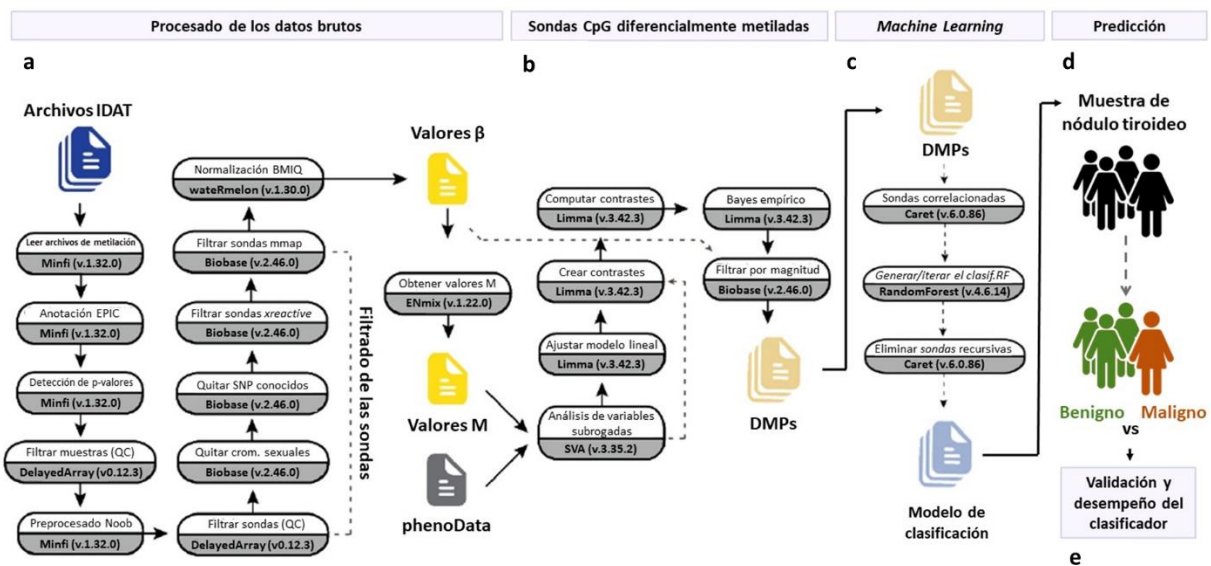


Figura 26 | Esquema representativo del procedimiento llevado a cabo para realizar la caracterización del perfil de metilación mediante el experimento con Infinium HumanMethylationEPIC 850K Beadchip y la generación del clasificador empleando técnicas de aprendizaje automático. (a) Gráfico que ilustra la canalización computacional y el software utilizado para la identificación de sitios CpG diferencialmente metilados (DMPs) en las muestras de cáncer de tiroides. (b) Gráfico que muestra la estrategia que se siguió para la identificación de los sitios CpG que podrían ser utilizados para la discriminación de nódulos tiroideos benignos de los malignos. (c) Procedimiento seguido para establecer cuáles de las DMPs conformarían el clasificador. (d) Predicción de benignidad/malignidad en las muestras tiroideas. (e) Validación del modelo mediante técnicas metodológicas alternativas y valoración del rendimiento.

Los datos de los archivos crudos (del inglés *Intensity Data Files*; IDAT) de la plataforma HumanMethylationEPIC Beadchip se procesaron utilizando el paquete R/Bioconductor *minfi*

⁸ Se entiende como normal a tejido tiroideo procedente de pacientes sin patología dicha glándula conocida. Los IDAT de estas muestras fueron descargados de la plataforma ENCODE.

(v_1.32.0) [239]. Las señales rojas y verdes de los datos sin procesar se corrigieron utilizando el algoritmo *ssNOOB* utilizando los parámetros predeterminados. Las sondas que superponen variantes genéticas, las sondas ubicadas en los cromosomas sexuales, las sondas de reacción cruzada y de mapeo múltiple, y las sondas con al menos una muestra con un p-valor de detección $> 0,01$ se descartaron de los análisis antes de proseguir con los análisis. Posteriormente, utilizando el paquete R/Bioconductor *wateRmelon* (v.1.30.0) [240], se aplicó el método de normalización denominado *Beta-Mixture Quantile* (BMIQ) [241] para la corrección del sesgo de las sondas. Subsecuentemente, se calcularon los valores β y valores M (**Figura 26b**), y se emplearon para establecer la metilación diferencial en los pasos siguientes.

8.2 Detección de sondas CpG diferencialmente metiladas

Para tener en cuenta posibles efectos de lote o variables de confusión, se realizó un análisis de variables subrogadas utilizando el algoritmo *sva* (v.3.35.2) [242]. Para cada comparación, los coeficientes de las variables subrogadas detectadas se estimaron utilizando la función *n.sv* en el contexto de los valores M, y las variables subrogadas *n.sv -1* se incluyeron en la definición del modelo para la detección de las CpG diferencialmente metiladas (dmCpG). La significación estadística de las sondas de metilación del ADN se calculó con el test-t implementado en *limma* (v_3.38.3) [243]. Se ajustó a un modelo lineal usando los niveles de metilación (valores M) como respuesta y el tipo de muestra (por ejemplo, maligno/normal, benigno/normal o tiroides/sangre) como la principal covariable de interés. Además, las variables subrogadas generadas en el análisis se incluyeron en la definición del modelo. Los valores de significación resultantes, obtenidos de los diferentes contrastes se corrigieron para realizar pruebas de comparación múltiples utilizando el método de Benjamini-Hochberg. Se utilizó un umbral de tasa de descubrimientos falsos (del inglés *False Discovery Rate*, FDR) inferior a 0,05 y una diferencia absoluta entre los valores medios de metilación del ADN de casos y controles de al menos el 25 % de cambio para ser considerados como dmCpG significativas (**Figura 26b**). Sin embargo, este límite se bajó al 10 % en las comparaciones de sangre versus glándula tiroides para descartar aquellas sondas con fluctuaciones mínimas entre los diferentes tipos de tejido.

9 Desarrollo del clasificador de tumores tiroideos mediante aprendizaje automático

El clasificador basado en Bosques Aleatorios (del inglés *Random Forest*, término que se usará en el resto del manuscrito) se creó utilizando el paquete R/CRAN *randomForest* (v.4.6.14). Se trata de un algoritmo de aprendizaje automático supervisado en el que los datos de entrenamiento o descubrimiento ya están clasificados, es decir, se conoce el grupo al que pertenecen a priori. La información detallada sobre el enfoque seguido para realizar el aprendizaje automático sobre la cohorte de descubrimiento se muestra en la **Figura 26c**.

Primero fue necesario establecer el conjunto de sondas que proporcionaran más información acerca de las alteraciones de metilación encontradas en los distintos subtipos de lesiones tiroideas estudiadas. Por ello, para el descubrimiento del clasificador se le proporcionaron al algoritmo aquellas dmCpG resultantes de la suma/intersección de las comparaciones de NH, FTA, FTC y FVPTC versus la glándula tiroides normal, excluyendo aquellas sondas que eran compartidas, en términos significación y dirección de cambio, con el tejido sanguíneo.

Para la generación correcta del modelo, se redujo la cantidad de variables de entrada mediante un filtro adicional utilizando el paquete R/CRAN *Caret* (v.6.0.86). Tras la aplicación de esta herramienta solo se incluyeron aquellas dmCpG altamente correlacionadas (punto de corte 0,9).

Posteriormente, se efectuó un remuestreo aleatorio (del inglés *bootstrapping*) sobre la cohorte de descubrimiento para crear el modelo utilizando la función *RandomForest*. Utilizando una semilla inicial, se crearon de forma iterativa 100 modelos, con los parámetros por defecto de la función, exceptuando la cantidad de árboles que se estableció en 5000 y el valor de corte para la predicción de la clase benigna que se fijó en 0,7 votos (para una mejor sensibilidad/especificidad). Para cada modelo creado en cada iteración, la importancia de las variables se clasificó de acuerdo con la disminución de la media del coeficiente de Gini (del inglés, *Mean Decrease Gini*) otorgada a cada sonda del modelo. Se emplearon las 10 variables que más veces aparecieron como más informativas en el total de los modelos creados. Finalmente, de las variables seleccionadas se comprobó el set mínimo necesario para clasificar las muestras de tiroides en las clases benigno o maligno, y para ello se utilizó el método de

validación cruzada (10 veces), junto con el método de eliminación de características recursivas dentro del paquete R/CRAN *Caret* y el algoritmo *rfe*.

La puntuación de malignidad para una muestra determinada se calculó en función de la proporción de votos malignos sobre el número total de votos (5000) pronosticados utilizando el modelo mencionado anteriormente, la matriz con los valores de metilación del ADN del set mínimo de los sitios CpG seleccionados y la función de *predict* que se encuentra en el paquete *stats* de R (v.4.0.2) (**Figura 26c-d**). A continuación, se muestra la **Ecuación 1** para determinación de la malignidad de las muestras:

$$MS = \frac{M \text{ votos}}{M \text{ votos} + B \text{ votos}}$$

Ecuación 1 | Ecuación para la determinación de la puntuación de malignidad. Donde MS significa Molecular Malignancy Score, y M y B representan el número de votos malignos o benignos identificados en los diferentes árboles generados durante la creación del modelo mediante Random Forest.

Las puntuaciones obtenidas a través de la **Ecuación 1** oscilan entre 0 y 1, y las más altas representan un estado de malignidad molecular más alto en una muestra determinada. El poder estadístico del estudio considerando diferentes tamaños de muestra, se calculó con el marco *PowerTools* [244] usando la matriz de metilación de ADN (valores M) que incluyen el set mínimo de CpG tras la generación del clasificador y las muestras antes mencionadas del conjunto de datos de descubrimiento (**Figura 27**). Para el cálculo se incluyó un rango de entre 5 y 39 muestras, un total de 100 iteraciones por variable y un umbral de significación de 0,001.

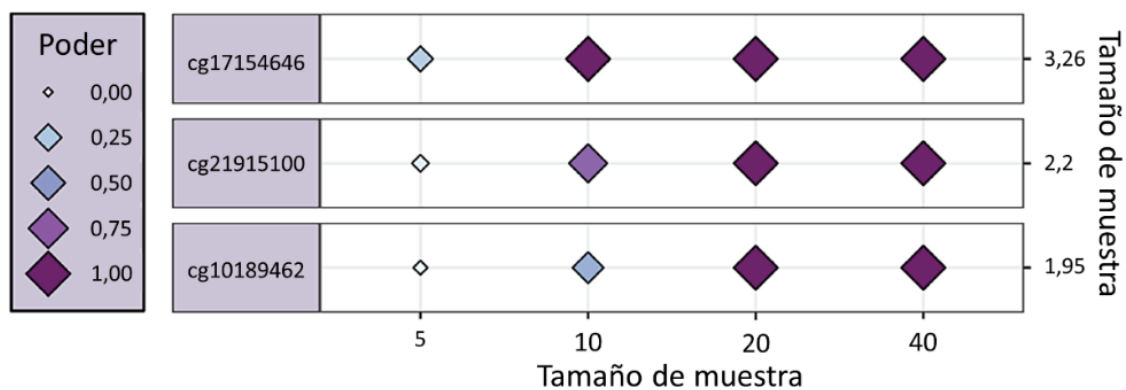


Figura 27 | Evaluación del rendimiento del sistema de clasificación Random Forest utilizando diferentes tamaños de muestra. Los cálculos de potencia y tamaño del efecto se obtuvieron utilizando el marco *PowerTools*.

10 Validación del rendimiento del clasificador de tumores tiroideos

Este sistema de clasificación fue posteriormente validado mediante diferentes enfoques ortogonales. Para ello, se analizó el estado de metilación locus específico mediante pirosecuenciación con bisulfito, en aquellas sondas obtenidas en el proceso de desarrollo mediante algoritmos de aprendizaje automático. Por otro lado, también se inspeccionó la literatura existente en búsqueda de bases de datos similares basadas en arrays de metilación.

10.1 Validación mediante pirosecuenciación con bisulfito

10.1.1 Genes de estudio

Los criterios para la selección de los genes se basaron en los resultados obtenidos en los estudios previos a nivel genómico total, con la *plataforma Infinium HumanMethylationEPIC 850K Beadchip*.

Los cebadores se diseñaron para llevar a cabo la validación del set mínimo de sondas necesarias para separar las muestras en benigno y maligno utilizando la técnica de pirosecuenciación con bisulfito en una cohorte más amplia y diversa. La selección final de genes se muestra en la **Tabla 6**.

Tabla 6 | Resumen de las características de los genes seleccionados para la validación del clasificador mediante pirosecuenciación con bisulfito.

Identificador de Illumina	Nombre común	Localización cromosómica	Contexto genómico	Hebra genómica
cg21915100	NPC2	chr14: 74.957.674	<i>Shell</i>	-
cg17154646	MBNL2	chr13: 97.858.492	<i>Open sea</i>	+
cg10189462	GPCPD1	chr20: 5.672.830	<i>Open sea</i>	+

10.1.2 Diseño de cebadores

Los cebadores se diseñaron con el programa *PyroMark Design Software* (v. 2.0.01.15). Brevemente, se introducen las secuencias génicas de interés y el software encuentra las posiciones CpG más relevantes, incluyéndose la de interés. Entonces, el programa genera automáticamente conjuntos de cebadores, incluidos tanto los necesarios para realizar la amplificación de la zona mediante PCR como los de secuenciación para la pirosecuenciación. Idealmente, se seleccionan aquellos cebadores que no presenten dinucleótidos CpG solapantes, de manera que la amplificación del ADN genómico modificado no sea sesgada.

La **Tabla 7** muestra los detalles de las condiciones de los cebadores y PCR usados en este estudio. A cada conjunto se le asigna una puntuación y la calidad, lo que refleja la idoneidad para el análisis; se escogieron por defecto, el conjunto de cebadores con puntuación más alta y temperaturas de anillamiento más cercanas entre el cebador directo e inverso. Además, uno de los cebadores fue marcado con biotina para permitir la inmovilización con las bolas de estreptavidina.

Tabla 7 | Secuencias de los cebadores directo, inverso y secuenciaciones correspondientes a cada gen. Tm °C es la temperatura de hibridación utilizada en la PCR en el caso de cebadores directo y reverso, y para la pirosecuenciación en el caso de los cebadores de secuenciación. BIO: biotinilado.

Gen	Posición de la CpG	Cebador	Secuencia (5'-3')	Tm °C	Producto de PCR (pares de bases)
NPC2	chr14: 74.957.673-74.957.674	Directo	AAGGTGGGTGTAGAGTGGAATAT	60°C	325
		Reverso-BIO	ATAACTCATTCCATCCTCACAACTATC		
		Secuenciación	GGAAATTTAAAGTGGTTATAAA	80°C	
MBNL2	chr13: 97.858.492- 97.858.493	Directo-BIO	AAGATTGAAAGAATTGGGTATGGAA	63°C	120
		Reverso	TAACCCAACAAAATTTACTCCCCTATC	80°C	
		Secuenciación	CCTCAAAAACAACCAC		
GPCPD1	chr20: 5.672.830-5.672.831	Directo	TGAAGGGAGTGAGGGTTTATTAG	60°C	114
		Reverso-BIO	CTAACCACTTCACAAAACCTTTTACTTAT	80°C	
		Secuenciación	AGATTAGAGTAGGTGTTG		

10.1.3 Modificación del ADN mediante bisulfito sódico

El tratamiento con bisulfito sódico (Bs) es usado habitualmente para detectar la metilación del ADN. El Bs es capaz de desaminar las citosinas a uracilos, que se sustituyen por timinas en el proceso de amplificación. La 5mC es resistente a la desaminación, cambiándose por citosinas tras la PCR (**Figura 28**) [245]. Tras esta modificación y posterior pirosecuenciación estamos midiendo los niveles de 5mC.

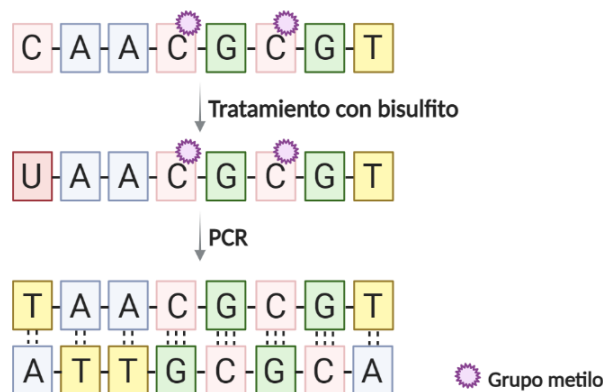


Figura 28 | Esquema del proceso de modificación del ADN genómico mediante bisulfito sódico. A: adeninas; T: timinas; C: citosinas; G: guaninas y U: uracilos. Imagen creada con BioRender.

Para llevar a cabo este procedimiento se utilizó un kit *EZ DNA Methylation Gold* (Zymo Research, Orange, California, Estados Unidos), en la que la preparación de las muestras, reactivos y protocolo de purificación se hizo tal y como indica el fabricante. A continuación, se detallan brevemente los pasos para llevar a cabo la modificación del material genético de todas las cohortes. Se parte de 500 ng de ADN genómico en un volumen de reacción final de 20 μ l de agua ultrapura autoclavada. En un tubo individual de se añaden 300 μ l del reactivo que contiene el Bs para realizar la conversión de las citosinas. Las muestras junto con el BS se homogenizan y se centrifugan a 13.000 r.p.m. durante un minuto. Para que ocurra el proceso de conversión las muestras son introducidas en un termociclador, sucediéndose los siguientes pasos: (1) 10 minutos a 98 °C; (2) 2,5 horas a 64 °C y (3) mantener a 4 °C infinito. A continuación, se lleva a cabo la desulfonación y limpieza de las muestras, mediante la sucesión de los reactivos *M-Binding Buffer*, *M-Wash Buffer*, *M-Desulphonation Buffer* y un lavado final *M-Wash Buffer* en las columnas *Zymo-Spin™ IC Column* pertenecientes al kit comercial. Finalmente, se eluye el ADN modificado M-Elution Buffer en un volumen final de 10 μ l. El material se conserva a -20 °C hasta su uso.

10.1.4 Amplificación por Reacción en Cadena de la Polimerasa

Se optimizaron las condiciones de la PCR, usando ADN genómico modificado de las líneas celulares comerciales de tiroides FTC-133 (FTC), K1 (PTC) y 8305C (ATC). El volumen de reacción final se estableció en 30 μ l y se procedió a realizar las reacciones de amplificación usando el kit *IMMOLASE™ DNA polymerase* para PCR (Bioline Reagents, Londres, Reino Unido) bajo las condiciones que siguen: 0,3 μ l de cada cebador específico directo e inverso, a 10 mM; 3 μ l de tampón *ImmoBuffer*; 0,9 μ l de MgCl₂ (1,5 mM); 0,9 μ l de dNTPs (0,2 mM) y 0,2 μ l de ADN polimerasa *IMMOLASE*; además de 2 μ l de muestra y 2 μ l de agua ultrapura autoclavada en el caso de los blancos. Se programa el termociclador de la siguiente forma: (1) 10 minutos a 95°C para realizar la desnaturalización del ADN; (2) 30 segundos a 95 °C; (3) 30 segundos a la temperatura de hibridación para el alineamiento de los cebadores; (4) 30 segundos a 72 °C para la elongación; se repitieron los pasos 2, 3 y 4 un total de 50 ciclos; (5) 30 segundos a 72 °C; (6) 30 segundos a 72 °C de elongación final; (7) mantener a 4 °C hasta recoger la muestra para su almacenamiento. Conservar el producto de PCR a -20 °C hasta su uso. Una vez establecidas las temperaturas de hibridación adecuadas para cada cebador se procedió a la amplificación de todas las muestras implicadas en la validación de los resultados. El esquema

del proceso de PCR llevado a cabo para amplificar los genes seleccionados se muestra en la **Figura 29**.

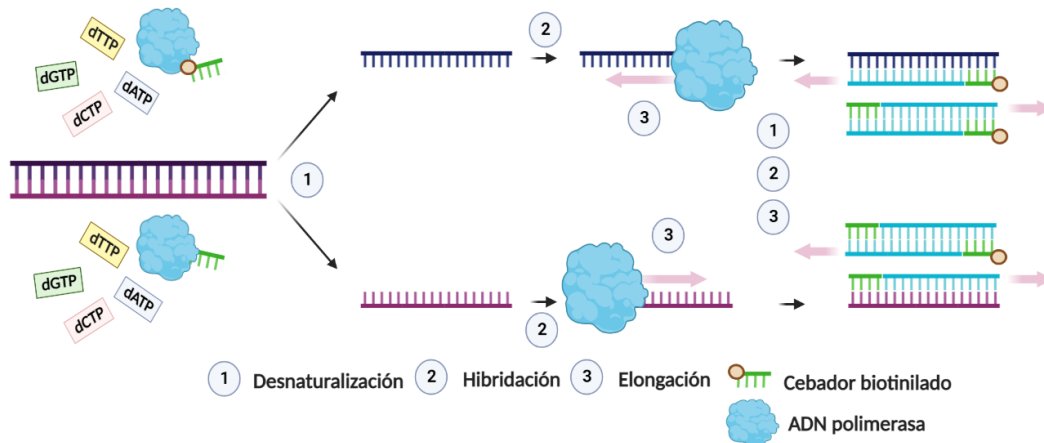


Figura 29 |Esquema del proceso llevado a cabo para realizar la amplificación de los genes seleccionados mediante PCR. Imagen creada con BioRender.

Previo a la pirosecuenciación, es necesario comprobar si el producto de PCR está presente tras la amplificación llevada a cabo. Por lo tanto, de cada reacción de PCR se emplearon 5 μ l y se resolvieron por electroforesis en gel de agarosa al 2 % en de Tris Borato EDTA al 0,5 X. Como agente intercalante se usaron 2 μ l de SYBR™ Safe DNA Gel Stain (Thermo Scientific, Waltham, Massachusetts, Estados Unidos) y en un pocillo independiente se agregaron 2 μ l como marcador de tamaño de amplicón de 50 pares de bases. El resto del producto de amplificación (25 μ l) se destina a la pirosecuenciación.

10.1.5 Pirosecuenciación

Se empleó la pirosecuenciación con bisulfito para medir de forma directa el estado de metilación de las muestras de la cohorte de validación y en conjunto de PAAF. Esta técnica se basa en la cuantificación del ratio C/T de los nucleótidos CpG susceptibles de ser metilados en los fragmentos cortos de ADN genómico previamente amplificados. Para que se pueda llevar a cabo esta determinación, es necesario que tenga lugar una cascada enzimática que generalmente se sucede de la siguiente forma (**Figura 30**) [246]: (1) la ADN polimerasa, incorpora dinucleótidos trifosfato (dNTPs) a la cadena molde, liberándose una cantidad equimolar de pirofosfato (PPi) al número de adiciones que se produzcan; (2) la ATP sulfúrilasa cataliza la conversión de PPi en ATP al reaccionar con el sustrato adenosina 5'-fosfosulfato (APS); (3) el ATP liberado se usa en la reacción de oxidación de la luciferasa, generándose un pulso de luz cuya intensidad es proporcional a la cantidad de dNTPs adicionados

anteriormente; (4) finalmente, la apirasa degrada el ATP y los dNTPs no incorporados. La adición consecutiva de nucleótidos genera un pirograma. Éste representa la dispensación secuencial de dNTPs y la intensidad lumínica asociada, que permite su cuantificación.

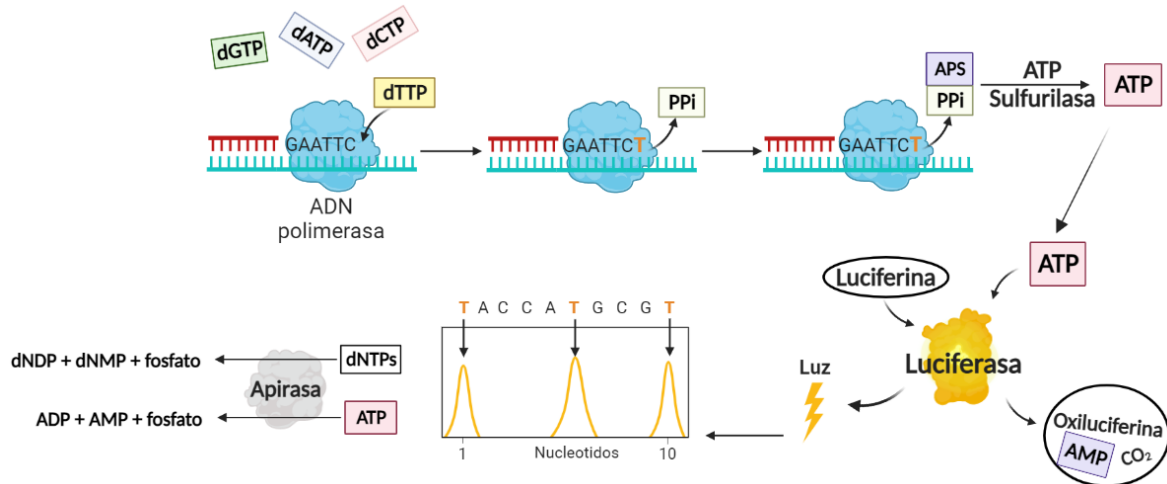


Figura 30 | Esquema del funcionamiento de la pirosecuenciación de ADN genómico modificado con bisulfito. (a) Proceso llevado a cabo en la estación de trabajo para seleccionar las hebras de ADN con la citosina de interés. (b) Proceso de pirosecuenciación llevado a cabo en el pirosecuenciador PyroMark Q24 para producir los pirogramas. Imagen creada con BioRender.

Para la generación de los pirogramas se siguió el siguiente protocolo: se añade 1 μ l sefarosa (Streptavidin Sepharose High Performance) (GE Health-care Bio-Sciences, Uppsala, Sweden), 40 μ l *Binding Buffer* (Qiagen, Qilden, Alemania), 19 μ l de agua ultrapura autoclavada y 20 μ l del producto de PCR por cada reacción. Los cebadores están biotinilados, por lo que el extremo 5' se unirá a la estreptavidina de las esferas de sefarosa. Esta disolución debe dejarse en agitación continua (14.000 r.p.m.) durante 10 minutos. Durante este tiempo se prepara la placa de pirosecuenciación: se añaden a cada pocillo 2,5 μ l de cebador de secuenciación (3 μ M) correspondiente, complementarios a la hebra molde que se desea estudiar (**Tabla 7**), junto con 22,5 μ l de *Annealing Buffer* (Qiagen, Qilden, Alemania). Transcurridos los 10 minutos se procede a las sucesivas fases de lavado y desnaturalización, la sefarosa con el ADN unido se mezcla con los cebadores de secuenciación en la placa de pirosecuenciación. Esta se incubará 2 minutos a 80°C en un termociclador, para que se produzca la hibridación del cebador. Se deja atemperar la placa 5 minutos en el pirosecuenciador y para finalizar se coloca el cartucho cargado con los volúmenes indicados de sustrato, enzima y dNTPs. Las placas con las muestras de interés se incorporaron al pirosecuenciador PyroMark Q24 (Qiagen, Qilden, Alemania) y así conseguir generar el pirograma que dará el porcentaje de 5mC para cada CpG analizada.

10.2 Análisis del rendimiento del clasificador mediante bases de datos externas

Obtuvimos datos sin procesar de HumanMethylationEPIC de un trabajo reciente centrado en la identificación de biomarcadores de cáncer de tiroides [159]. La base de datos con código de acceso GSE121377, se descargó de la plataforma Gene Expression Omnibus (GEO). Los datos se procesados y normalizados tal como se indica en la **Sección 8.1 de Material y Métodos**. Tras calcular los valores β y valores M, aquellas sondas identificadas en el proceso de identificación de biomarcadores fueron evaluadas. Se otorgaron puntuaciones de malignidad a los tumores siguiendo el mismo protocolo que el empleado para las muestras de la cohorte de descubrimiento descrito en la **Sección 9 de Material y Métodos**.

11 Disponibilidad y accesibilidad de los datos

Los oligonucleótidos de ADN utilizados en este estudio se proporcionan en la **Tabla 6**. Los datos sin procesar de la secuenciación completa del exoma de los nódulos tiroideos se han depositado en el servicio European Genome-phenome Archive (EGA) con el número de acceso EGAS00001005561. Los datos brutos del análisis de metilación del ADN usando la plataforma HumanMethylationEPIC, incluidos los archivos IDAT, se han depositado en ArrayExpress con el número de acceso E-MTAB-10906. Los datos correspondientes a los arrays de metilación de las diferentes poblaciones de células sanguíneas se obtuvieron de la entrada E-MTAB-6315 de ArrayExpress [238]. Los IDAT sin procesar de las cuatro muestras control (ENCLB141TUI, ENCLB867YHK, ENCLB610AGH y ENCLB305LHQ) procedentes de la glándula tiroides se obtuvieron del proyecto ENCODE [237]. Con fines de validación usando bases de datos externa, se descargaron archivos HumanMethylationEPIC del trabajo publicado por Park y colaboradores en 2020 (entrada GEO GSE121377).

RESULTADOS

1 Perfil mutacional de las lesiones tiroideas con patrón folicular

La acumulación de alteraciones genéticas y epigenéticas juega un papel clave en el desarrollo de firmas oncogénicas en los cánceres humanos. Para identificar las alteraciones moleculares que pueden explicar la potencial tumorigénica de las lesiones tiroideas benignas y malignas con fenotipo folicular (asociadas con mayor frecuencia en las categorías de PAAF indeterminadas) primero se exploró el perfil mutacional mediante la secuenciación del exoma completo (Figura 31).

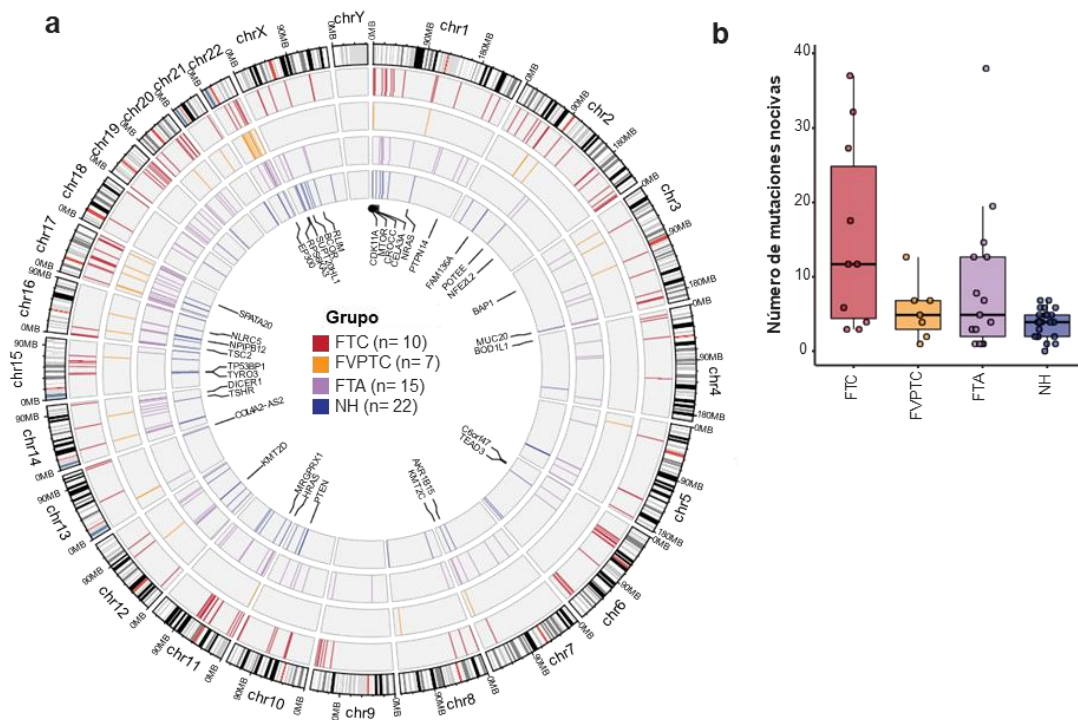


Figura 31 | La secuenciación del exoma identifica alteraciones genéticas en lesiones tiroideas foliculares benignas y malignas. a) Diagrama de circos que ilustra el panorama mutacional de las muestras de tiroides analizadas en este estudio. La figura exterior muestra el ideograma de los cromosomas humanos (hg38), incluida la ubicación y la posición de las bandas citogenéticas conocidas. Las diferentes circunferencias concéntricas muestran la ubicación genómica de las mutaciones (líneas) identificadas en FTC (rojo), FVPTC (naranja), FTA (morado) y NH (azul). Los genes con mutaciones recurrentes en más de una muestra o los genes con implicaciones conocidas para el cáncer de tiroides se destacan en el recuadro central. También se destaca el número de casos analizados por cada subtipo muestral. b) Diagrama de cajas que indica el número de mutaciones nocivas observadas en los pacientes con nódulos tiroideos incluidos en la cohorte de descubrimiento. Los puntos individuales indican el número de mutaciones identificadas en una muestra determinada.

Para cubrir el espectro de lesiones tiroideas con patrón folicular, se analizaron biopsias de 10 FTC, 7 FVPTC, 15 FTA y 22 NH con una profundidad media de cobertura de 132X (rango de 68 a 277X, Anexo 1: Tabla S2, Anexo 2: Figura S1). Este análisis reveló la presencia de mutaciones potencialmente nocivas a lo largo de todo el genoma en las muestras de la cohorte

de descubrimiento, encontrándose 152 en FTC, 38 en FVPTC, en 133 FTA y en 77 NH (**Figura 31a**). Las muestras de FTC y NH presentaron el mayor y menor número de mutaciones por muestra, respectivamente (**Figura 31b**). Mientras que los FTA, segunda neoplasia con mayor número de mutaciones dañinas, presentaron un número variable de mutaciones por paciente (**Figura 31b**). Curiosamente, los FVPTC no presentaban tantas mutaciones como sus homólogos malignos los FTC (media de alteraciones 5 y 11,5, respectivamente; **Figura 31b**), lo que sugiere que los mecanismos moleculares que gobiernan cada tipo de carcinoma de tiroides pueden resultar en una carga mutacional tumoral diferente.

Se observaron mutaciones recurrentes en los genes involucrados en rutas clásicas que participan en el desarrollo del cáncer de tiroides en nuestro conjunto de descubrimiento, incluidos MAPK (*HRAS*, *NRAS*, *MAP2K3*, *RPS6KA3*, *TAB1*, *DUSP22*, *MAPK8IP3*, *CACNA1H*); mTOR (*HIF1A*, *MTOR*, *RPS6KA3*, *TSC2*); vía de la insulina (*HRAS*, *NRAS*, *INPPL1*, *MTOR*, *TSC2*, *PRKCZ*, *HK3*), y el metabolismo y sistema de señalización de inositol fosfato (*DGKG*, *INPPL1*, *INPP1*, *INPP5B*, *INPP5J*, *MTMR3*, *PI4KA*, *PIK4K2A*) (**Figura 32a**).

Una inspección más detallada a los genes más frecuentemente alterados en este tipo de neoplasias muestra que tanto *BRAF* como *HRAS* solo se observan en los tumores malignos (en total 23,5 % y 17,6 %, respectivamente), mientras que *NRAS* está mutado en todos los subtipos (FTC 30 %, FVPTC 14,23 %, FTA 6,6 % y NH 4,5 %) (**Figura 32b**). Asimismo, se estudiaron las mutaciones *BRAF*, *KRAS*, *NRAS* y *HRAS* mediante RT-qPCR en la cohorte de validación (exceptuando los NH) (**Figura 32c**). Un 41 % de los tumores presentaron alguna mutación conductora. Como cabría esperar, los FVPTC fueron los únicos tumores que exhibieron mutaciones somáticas en *BRAF* (36,36 %), además de *NRAS* (13,63 %) y *HRAS* (13,63 %). Los FTA exhibieron mutaciones tanto en *NRAS* como en *HRAS* (21,73 % y 8,69 %, respectivamente), y que su contraparte maligna, los FTC, solo en *NRAS* (25 %). Teniendo en cuenta ambas técnicas la prevalencia estas mutaciones en los tumores malignos analizadas es de 10,71 % y 14,29 % para *BRAF*, *RAS*, respectivamente. Mientras que en los nódulos benignos mutaciones en *RAS* activadoras de la cascada de señalización PI3K-AKT en 8 %.

No se pudieron definir clústeres significativos asociados al estado de malignidad, ni a ninguno de los grupos de diagnóstico. Estos resultados indican que las alteraciones mutacionales en el genoma son complejas y subóptimas en nuestra cohorte para establecer una estimación diagnóstica apropiada de las lesiones foliculares.

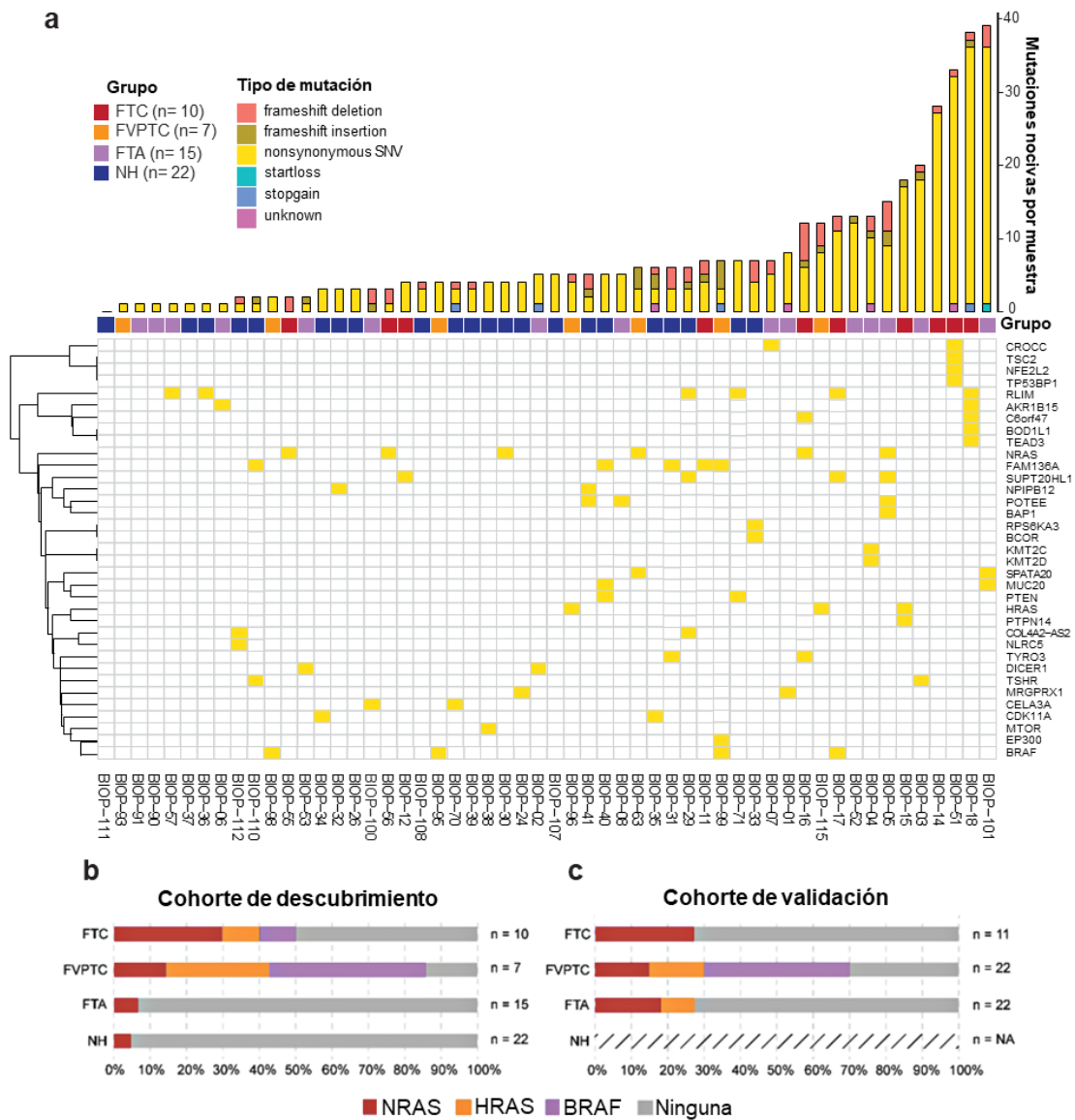


Figura 32 | Representación detallada de los genes mutados identificados en este estudio. (a) Heatmap que muestra solo aquellas mutaciones que ocurren en al menos dos pacientes o que han sido descritas previamente en el cáncer de tiroides en la cohorte de descubrimiento. El histograma superior refleja el número total y el tipo de mutaciones nocivas observadas para cada muestra, tal como se codifica en la leyenda de la figura. Los cuadrados de colores ilustran el diagnóstico de estas muestras como se indica en el recuadro: en FTC (rojo), FVPTC (naranja), FTA (morado) y NH (azul). (b-c) Prevalencia de mutaciones NRAS, HRAS y BRAF en las cohortes de descubrimiento y validación identificadas por exoma completo (b) o RT-PCR (c), respectivamente.

2 Alteraciones epigenómicas en lesiones tiroideas con patrón folicular

Para poder superar las limitaciones de los estudios genéticos, se realizó un análisis epigenético de las mismas muestras analizadas en el apartado anterior mediante arrays de metilación.

Aunque los cambios de metilación pueden ser globales, se exploraron las posibles diferencias entre las distintas muestras de la cohorte de descubrimiento. Los análisis de agrupamiento (**Figura 33a**) usando los 32.202 sitios dmCpG encontrados entre las muestras malignas y las benignas revelaron tres clústeres diferentes: un gran grupo formado por los NH y la gran mayoría de los FTA; un segundo en el que se encontraban tanto adenomas como FTC y un tercero en el que se encontraban los FVPTC y el resto de las muestras malignas. En general, las biopsias tendieron a agruparse según su categoría diagnóstica, con excepción de los FTA que mostraron perfiles de metilación similares tanto a tumores benignos como malignos.

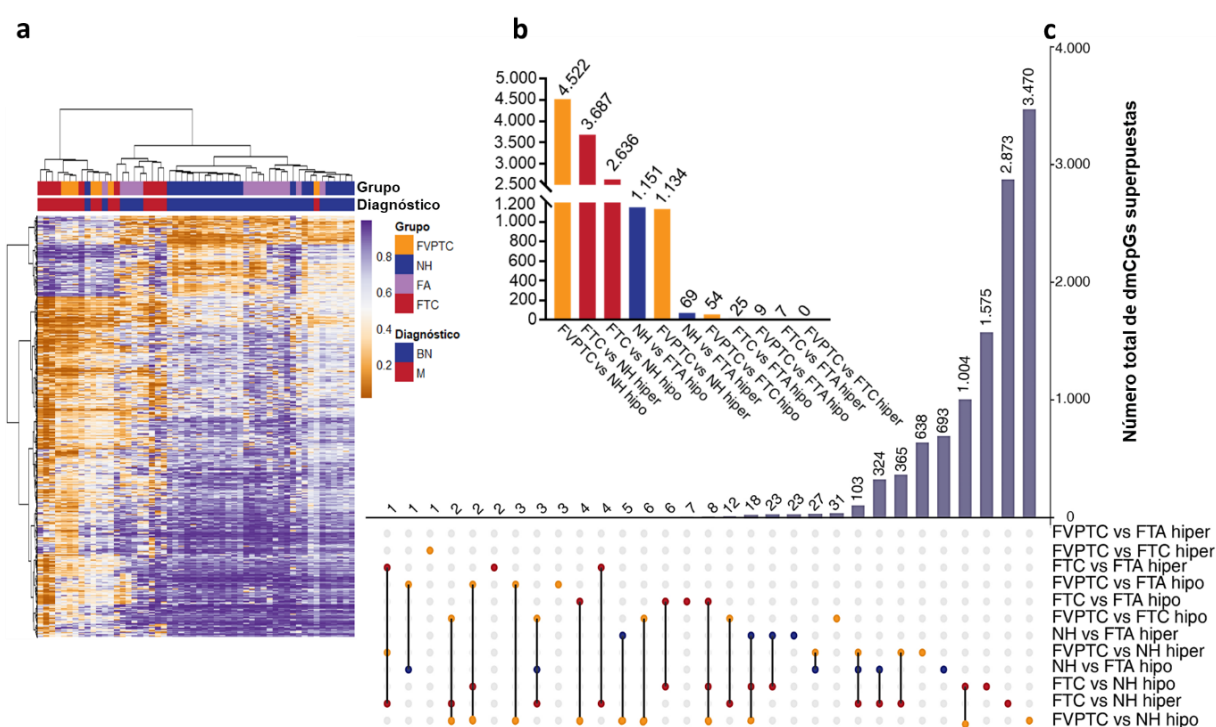


Figura 33 | Identificación de patrones aberrantes de metilación del ADN en los tumores tiroideos de origen folicular en la cohorte de descubrimiento. (a) Gráfico de agrupamiento jerárquico que muestra los patrones aberrantes de metilación del ADN de las muestras malignas (FTC y FVPTC) frente a las benignas (FTA y NH) (32.202 sondas). (b) Gráficos de barras que indican el número total de sitios CpG diferencialmente metilados, separados por hiper e hipometilación, identificados en cada una de las comparaciones posibles en la cohorte de descubrimiento. (c) Gráfico de barras que muestra el número de cambios de hipermetilación e hipometilación particulares (puntos individuales) y superpuestos (líneas conectadas) en las diferentes comparaciones analizadas. Se indica el número total de dmCpG por categoría encima de cada barra.

A continuación, de acuerdo con los criterios establecidos en la metodología, se identificaron los dmCpG en las diferentes comparativas y se clasificaron según el estado de metilación (**Figura 33b**). Como era de esperar, los carcinomas mostraron más cambios con respecto a los nódulos hiperplásicos, presentando más hipometilaciones los FVPTC (4.522) y

más hipermetilaciones los FTC (3.687) con respecto a los NH. Por otro lado, los adenomas mostraron pocas diferencias con respecto a los tumores malignos, pero sí que se encontraron cambios con respecto a las hiperplasias; sobre todo, sondas diferencialmente hipermetiladas en FTA con respecto a las NH. Curiosamente, el 41 % de los sitios diferencialmente metilados en los tumores malignos versus NH ya están alterados en los adenomas considerados benignos (Figura 33c).

Es importante mencionar, que en las muestras de tiroides se pueden encontrar no solo células foliculares si no también, entre otros: glóbulos blancos, músculo esquelético, vasos sanguíneos o tejido normal adyacente. Por ello, para superar las dificultades de diagnóstico asociadas con la contaminación de células no neoplásicas, se utilizaron las muestras de glándula tiroides del proyecto ENCODE como referencia de tejido normal. Asimismo, para asegurar que los cambios eran exclusivos de las células tiroideas, se realizaron análisis diferenciales de metilación de tejido tiroideo contra un *pool* de diferentes poblaciones celulares hematopoyéticas (HCS, LB, LTC, LTH, NK, M y N).

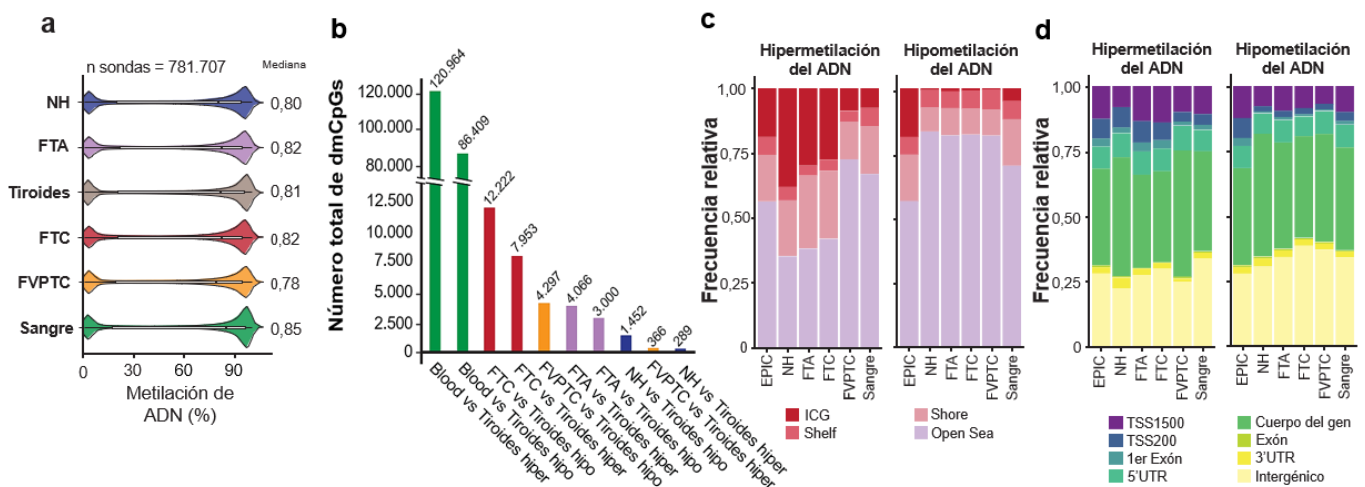


Figura 34 |Metilación global del ADN, sitios CpG diferencialmente metilados y ubicación en el genoma en las muestras analizadas en la cohorte de descubrimiento mediante arrays de metilación. (a) Gráficos de violín que muestran los niveles globales de metilación del ADN de los sitios CpG. El gráfico muestra la distribución promedio de metilación de CpG de todas las muestras agrupadas por sangre, nódulos benignos (tiroides normal, NH y FTA) y tumores malignos (FTC y FVPTC). Se utilizaron un total de 781.707 sitios CpG, que fueron detectados a través de la plataforma Human Methylation EPIC (850k). Se indica la mediana de metilación de CpG por grupo. (b) Gráficos de barras que indican el número total de sitios CpG diferencialmente metilados, separados por hiper e hipometilación, identificados en cada una de las comparaciones frente al tejido tiroideo normal (ENCODE). (c y d) Gráficos de barras apilados que muestran la frecuencia relativa de sitios CpG hipermetilados o hipometilados observados en cada una de las diferentes comparaciones frente a una muestra de tiroides normal en relación con su contexto de CpG (c) o ubicación de CpG (d) en el genoma. Se indica la distribución genómica de fondo de la plataforma de metilación EPIC como referencia.

En este sentido, la metilación global del ADN osciló entre valores de 0,78 (FVPTC) y 0,82 (FTA y FTC), con el *pool* de muestras sanguíneas mostrando una hipermetilación sustancial (0,85) (**Figura 34a**). Como era de esperar, la comparativa de las muestras sanguíneas frente a los epigenomas de glándulas tiroideas normales presentó el mayor número de dmCpGs (207.373). Curiosamente, el número de alteraciones epigenéticas mantuvo una relación estrecha con el estado diagnóstico de las lesiones tiroideas, con FTC y NH mostrando el mayor (20.175) y menor (1.741) número de dmCpGs, respectivamente (**Figura 34b**). La distribución genómica de las sondas diferenciales mostró una hipermetilación significativa en las ICG y las orillas de las ICG, particularmente en los grupos NH, FTC y FTA (p -valor $< 0,001$). En contraste, la hipometilación estaba fuertemente asociada a localizaciones de mar abierto (del inglés, *open sea CpG*) y sitios intergénicos en todas las comparaciones testadas (p -valor $< 0,001$) (**Figura 34c y d**).

Para evitar posibles efectos de confusión, nos enfocamos únicamente en aquellas dmCpG específicas que no se comparten, o que al menos muestran una direccionalidad opuesta, entre las muestras de sangre y cualquiera de las comparaciones de lesiones tiroideas mencionadas anteriormente (**Figura 35a**; las barras amarillas), lo que resulta en la identificación de 12.845 sitios dmCpG. Los análisis de clúster jerárquico en la cohorte de descubrimiento usando correlaciones por pares y la información de la metilación del ADN codificada en esos sitios altamente específicos revelaron la presencia de diferentes grupos (**Figura 35b**). Así, el análisis de agrupamiento fue útil para discriminar el diagnóstico de las lesiones tiroideas, lo que resultó en una clara diferenciación entre muestras benignas (NH, sangre y tiroides normal) y malignas (FTC y FVPTC). Además, se observó que las muestras correspondientes al mismo subtipo de lesión tiroidea tienden a agruparse a excepción, una vez más, del histotipo benigno FTA que muestra un patrón disperso en varios grupos. Este hecho refuerza la hipótesis de que los adenomas no están completamente definidos y pueden representar un continuo hacia estados malignos (también a nivel molecular), estando potencialmente asociados con un pronóstico benigno o maligno dependiendo de sus características epigenéticas intrínsecas.

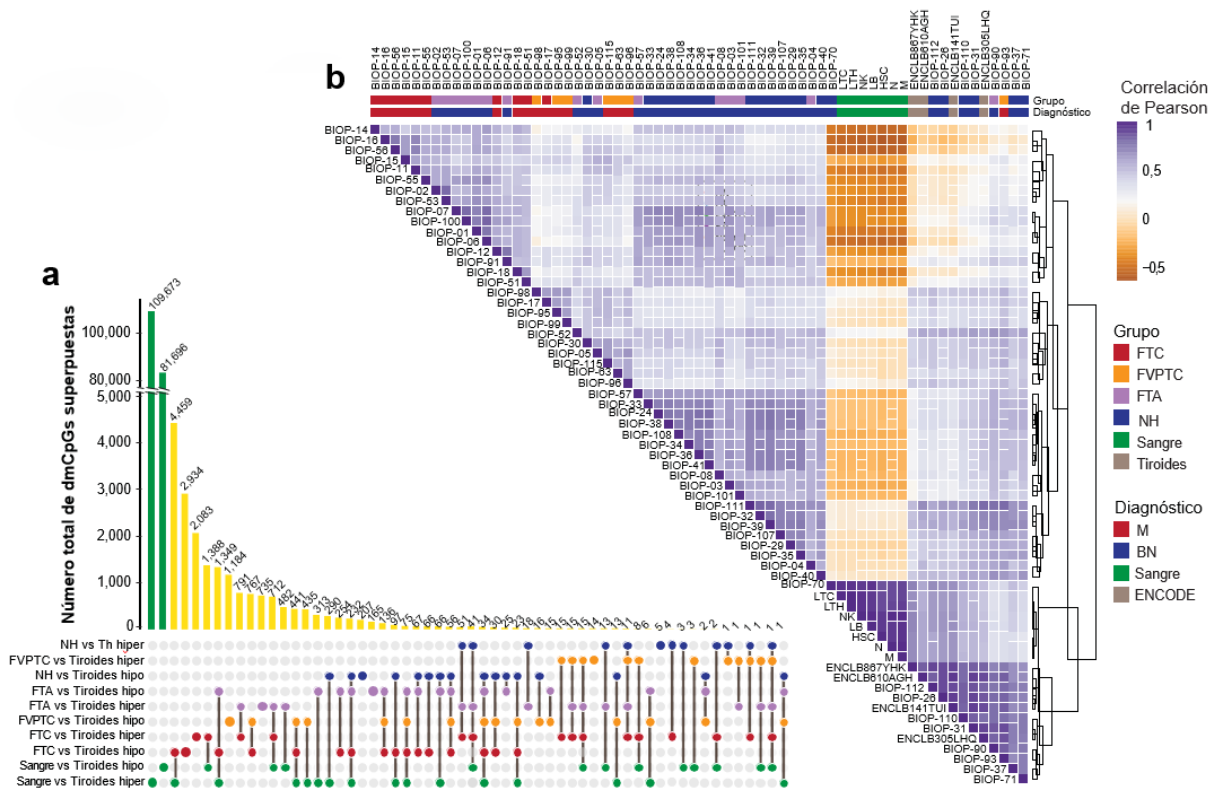


Figura 35 | Identificación de patrones aberrantes de metilación del ADN en el cáncer folicular de tiroides con respecto al tejido normal. (a) Gráfico de barras que muestra el número de cambios de hipermetilación e hipometilación particulares (puntos individuales) y superpuestos (líneas conectadas) en las diferentes comparaciones analizadas. Se indica el número total de dmCpG por categoría encima de cada barra. Debido al alto número de dmCpGs identificados entre las muestras de sangre y de tiroides, los gráficos incluyen una ruptura del eje para una mejor visualización. (b) Análisis de correlación por pares entre las muestras incluidas en este estudio. La escala de colores indica los valores de correlación de Pearson para los pares de correlación correspondientes. En la parte superior, la leyenda de color indica la categoría de cáncer y el diagnóstico de cada muestra individual.

3 Desarrollo de un sistema de clasificación de diagnóstico de lesiones tiroideas malignas basado en biomarcadores epigenéticos

Mediante los análisis epigenéticos anteriormente mencionados, se identificó con precisión la heterogeneidad presente en las lesiones de tiroides. Por ello, se desarrolló una estrategia, mediante técnicas de aprendizaje automático, para la identificación de biomarcadores moleculares que permitan establecer el estado de malignidad de los nódulos tiroideos examinados. Para el entrenamiento inicial del modelo, se emplearon las 22 muestras benignas de NH y las 17 malignas (FTC y FVPTC) de la cohorte de descubrimiento. Los FTA no fueron incluidos en el desarrollo del clasificador para evitar posibles efectos de confusión. Esta decisión se tomó en base a los resultados obtenidos en el análisis de metilación diferencial, ya

que se observó que los adenomas podían agruparse en varios grupos y, en particular, junto con los carcinomas.

Se utilizaron las 12.845 dmCpGs obtenidas a partir del filtrado restrictivo del estudio de metilación del ADN (**Figura 35a**) para construir un modelo robusto de clasificación *Random Forest* para tratar de discriminar de forma correcta entre las lesiones tiroideas benignas y las malignas. Asimismo, los gráficos de importancia de variables (disminución media de precisión o *mean decrease accuracy*; y disminución media del coeficiente de Gini o *mean decrease Gini*) son un resultado fundamental del algoritmo de *Random Forest* y muestran, para cada variable, qué tan importante son individualmente para clasificar los datos. La disminución media del coeficiente de Gini es una medida que hace referencia a cómo cada variable contribuye a la homogeneidad de los nodos y las hojas en el bosque resultante. Cuanto mayor sea la puntuación, mayor será la importancia de la variable en el modelo.

De este modo, como se comentó en el **apartado 8 de Material y Métodos**, se crearon de forma iterativa 100 modelos *Random Forest* y para cada uno de ellos se anotó la cantidad de veces que un sitio CpG se identificó como una de las 10 variables más informativas (**Figura 36a**).

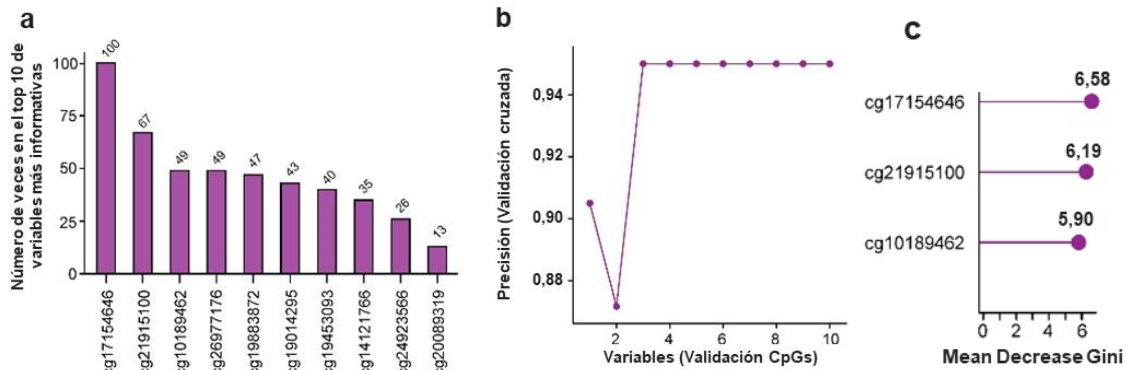


Figura 36 | Generación del modelo de clasificación de lesiones tiroideas mediante 12845 dmCpG utilizando el algoritmo Random Forest. a) Gráfico de barras que muestra los sitios CpG más informativos para discriminar entre los nódulos benignos y los malignos. Los valores indican el número de veces que un sitio CpG se identificó como una de las 10 variables más informativas utilizando un enfoque iterativo de selección aleatoria de bosques (100 semillas diferentes, 5000 árboles cada una). b) Diagrama de líneas que ilustra el número óptimo de variables (CpG) requeridas para el desarrollo de un modelo preciso por medio de un enfoque recursivo de eliminación de características. La precisión alcanza su nivel máximo cuando se retienen 3 características en el modelo. c) Histograma que representa la importancia relativa (Gini de disminución media) de los tres sitios CpG incluidos en el modelo final.

A continuación, utilizando la técnica de validación cruzada de K iteraciones se evaluaron los resultados obtenidos anteriormente. Esta técnica permitió distinguir, de entre

las CpG más informativas con cuantas se adquiere la mayor precisión. Además, la herramienta proporcionó la combinación mínima de sondas a utilizar para clasificar las muestras entre benigno y maligno (**Figura 36b**). De esta forma se seleccionaron las dmCpG más informativas y robustas, y se obtuvo un sistema de clasificación mínimo constituido por tres sondas CpG: cg17154646, cg21915100 y cg10189462 (**Figura 36c**). Estos sitios CpG se sitúan en las proximidades de los genes *MBNL2*, *NPC2* y *GPCPD1*, respectivamente (**Tabla 6**).

Individualmente, utilizando cualquiera de los sitios de CpG, se observó una discriminación significativa de muestras benignas contra malignas en el conjunto de descubrimiento (**Figura 37a**). Cuando estas tres sondas se implementaron conjuntamente en el clasificador *Random Forest* utilizando un enfoque de validación cruzada, se obtuvo una precisión del 95 %, con una especificidad del 90 % (20 de 22 muestras benignas) y una sensibilidad del 100 % (17 de 17 muestras malignas) en el conjunto de descubrimiento. Asimismo, los valores VPN y VPP fueron considerablemente altos (1 y 0,89, respectivamente) (**Figura 37b; Anexo 1: Tabla S3**).

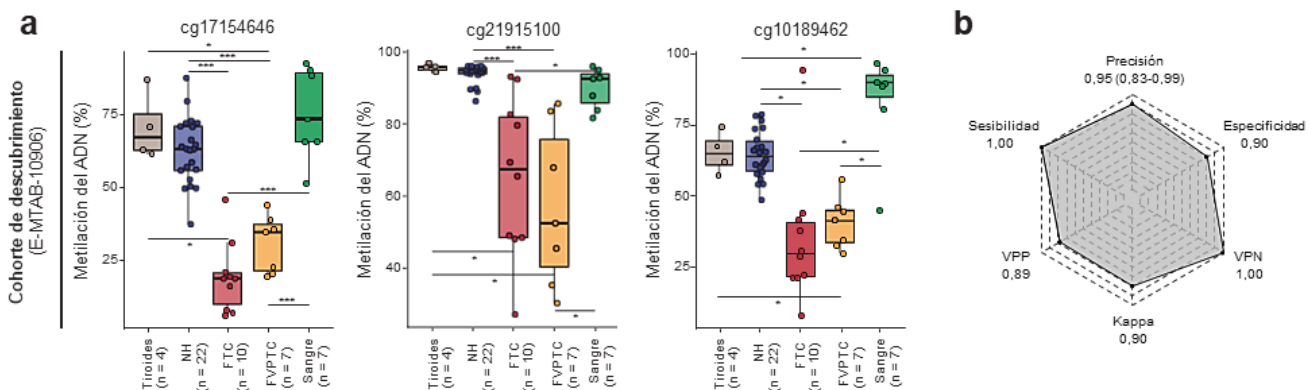


Figura 37 | Clasificador epigenético de tres CpG discrimina entre nódulos tiroideos benignos y malignos en la cohorte de descubrimiento. (a) Diagramas de cajas que indican el valor de metilación del ADN de los tres sitios CpG en los diferentes histotipos (los puntos representan muestras individuales). (b) Gráfico de radar que ilustra el rendimiento del modelo de tres CpG propuesto en el contexto de la cohorte de descubrimiento. La significación estadística entre las condiciones se calculó mediante una prueba no paramétrica (test de Wilcoxon; * p-valor < 0,05, ** p-valor < 0,01, *** p-valor < 0,001).

4 Rendimiento del sistema de clasificación de diagnóstico de lesiones tiroideas malignas mediante bases de datos externas

Paralelamente, además de evaluar el desempeño en la cohorte de descubrimiento, este sistema de clasificación mínima fue validado mediante diferentes enfoques.

RESULTADOS

Inicialmente, se recopilaron datos de un estudio recientemente publicado por Park y colaboradores en 2020 [159]. En él se centraron en la identificación de biomarcadores de diagnóstico del cáncer de tiroides, entrenando su modelo con sondas procedentes de la misma plataforma de metilación de ADN (GSE121377) que se emplea en el presente trabajo. La cohorte se compone de 13 nódulos identificados como benignos (tejido tiroideo no tumoral n = 7; y NIFTP, n = 6) y 21 carcinomas de tiroides (CPTC, n = 11; IEFVPTC, n = 3; y TPTC, n = 7).

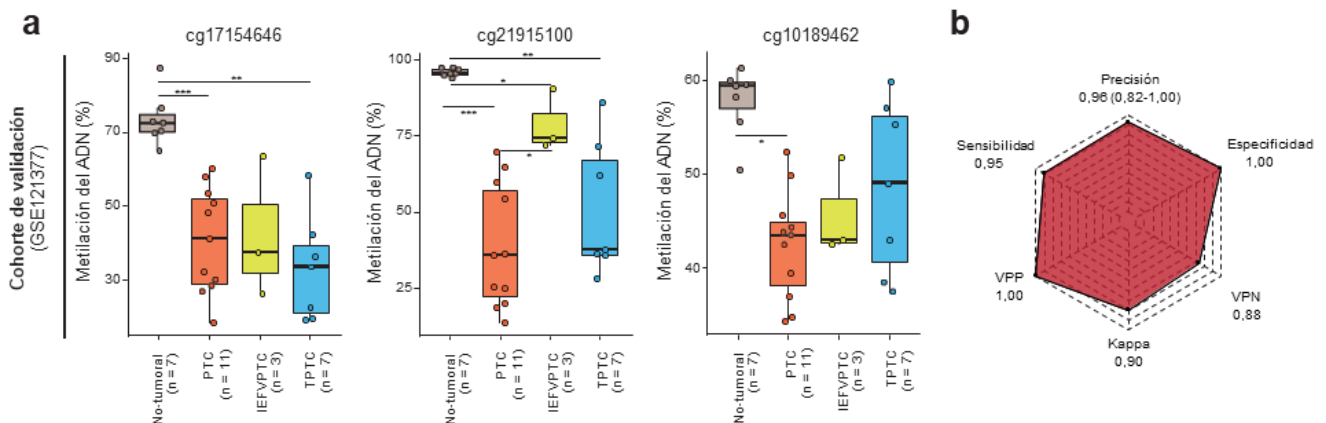


Figura 38 | Clasificador epigenético de tres CpG (cg17154646, cg21915100 y cg10189462) que discriminan entre nódulos tiroideos benignos y malignos en la cohorte de validación GSE121377. (a) Diagramas de cajas que indican el valor de metilación del ADN de los tres sitios CpG en los diferentes histotipos (los puntos representan muestras individuales). (b) Gráfico de radar que ilustra el rendimiento del modelo de tres CpG propuesto en el contexto de la cohorte de validación GSE121377 tras retirar los tumores NIFTP. La significación estadística entre las condiciones se calculó mediante una prueba no paramétrica (test de Wilcoxon; * p-valor < 0,05, ** p-valor < 0,01, *** p-valor < 0,001).

Analizando la metilación de las tres sondas identificadas por nuestro modelo para discernir benignidad de malignidad, se encontraron diferencias significativas entre los niveles del tejido no tumoral y los en carcinomas, sobre todo en la sonda que hace referencia al gen *NPC2* (Figura 38a; Anexo 1: Tabla S4). Curiosamente, la combinación de nuestros 3 biomarcadores con el modelo *Random Forest* clasificó sistemáticamente a los NIFTP como lesiones malignas. Al investigar las 3 CpGs se puede ver que, al igual que los carcinomas, existe una reducción significativa en los niveles de metilación con respecto al tejido no tumoral del estudio de Park y colaboradores (Anexo 2: Figura S2). Los NIFTP están clasificados como tumores de potencial maligno incierto según la cuarta versión de la OMS (Anexo 1: Tabla S1) [98]. A pesar de su buen pronóstico y de estar encapsulados, los resultados parecen sugerir que su contenido molecular puede asemejarse más a las características epigenéticas de las células malignas. Por ello, al retirar estos tumores de potencial maligno incierto de la cohorte de validación, nuestro sistema de clasificación mostró un gran rendimiento, siendo capaz de

discriminar entre los nódulos no tumorales y los carcinomas con una precisión de 96 %, una sensibilidad de 95 % y una especificidad de 100 % (Figura 38b).

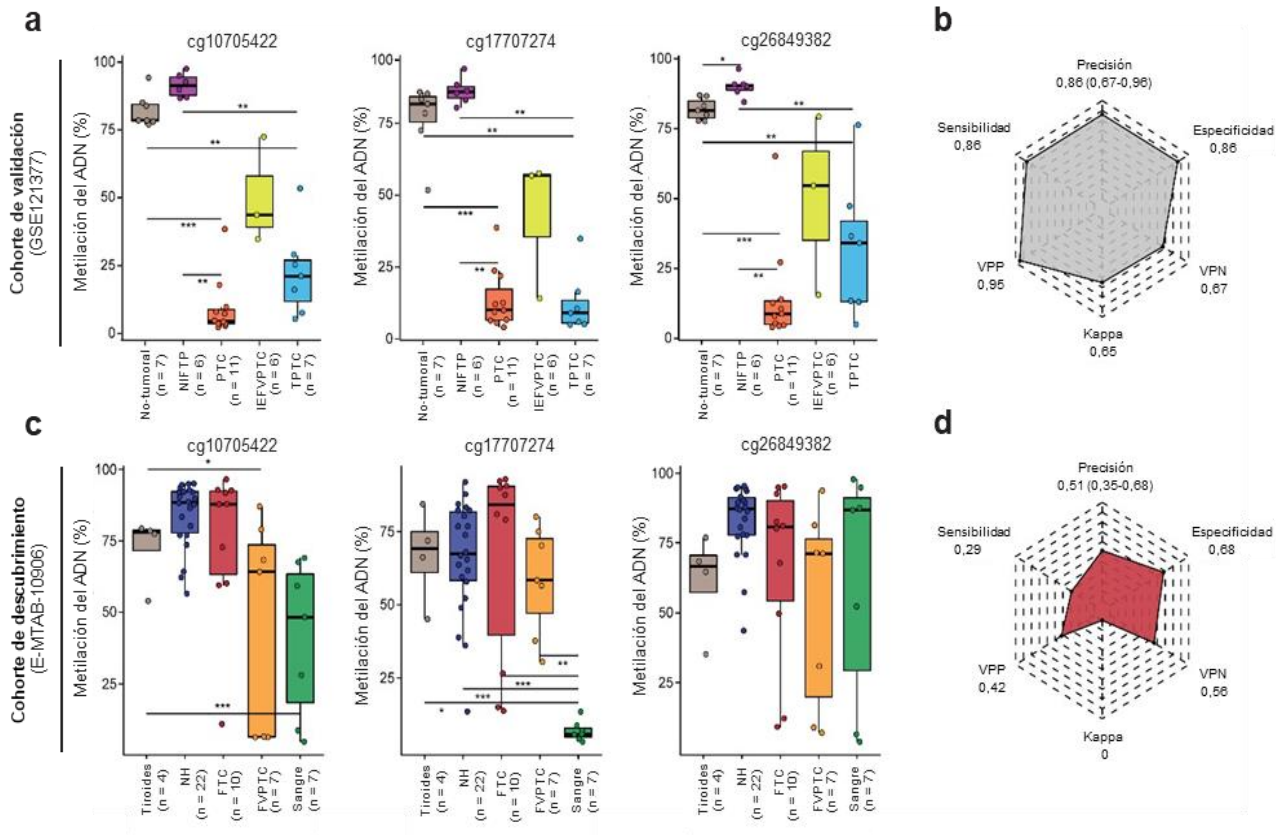


Figura 39 | Generación del modelo de clasificación de lesiones tiroideas mediante las sondas cg10705422, cg17707274 y cg26849382 utilizando el algoritmo Random Forest. Se desarrollo un clasificador basado en aprendizaje automático usando esos tres sitios CpG y se aplicó en las diferentes cohortes de descubrimiento y validación. (a) Diagramas de caja que ilustra los valores de metilación del ADN de estos tres sitios CpG en el contexto de las muestras del estudio GSE121377 (conjunto de datos de entrenamiento, excluyendo NIFTP). (b) Gráfico de radar que ilustran el rendimiento del modelo de tres CpG obtenido de GSE121377 en el contexto del conjunto de datos de ese mismo trabajo (c) Diagrama de cajas que muestra los valores de metilación del ADN de estos tres sitios CpG de Park y colaboradores utilizando las lesiones foliculares de la cohorte de descubrimiento. (d) Gráfico de radar que ilustran el rendimiento del modelo de tres CpG obtenido de GSE121377 en el contexto del conjunto de la cohorte de descubrimiento de esta tesis. La significación estadística entre las condiciones se calculó mediante una prueba no paramétrica (test de Wilcoxon; * p-valor < 0,05, ** p-valor < 0,01, *** p-valor < 0,001).

Además, para evaluar la capacidad de discriminación del sistema de clasificación propuesto en la literatura por Park y colaboradores [159], realizamos un análisis análogo utilizando las alteraciones epigenéticas descritas en este trabajo tanto en su grupo de muestras como en nuestra cohorte de descubrimiento. Para la discriminación del estado de malignidad, se entrenó un modelo *Random Forest* utilizando la información obtenida de los tres marcadores de metilación del ADN (cg10705422, cg17707274, cg26849382), candidatos

descritos a partir de las muestras de la base de datos GSE121377 [159]. Como era de esperar, se logró una discriminación significativa de las lesiones tiroideas benignas contra malignas de forma independiente con cualquiera de sus 3 sitios CpG en el conjunto de la cohorte de Park y colaboradores (**Figura 39a**). En ausencia del subtipo NIFTP, la implementación de estos 3 CpG de forma conjunta arrojó una precisión de clasificación, sensibilidad y especificidad del 86 %, respectivamente en los tres estadísticos mencionados (**Figura 39b**).

Curiosamente, cuando se evaluaron las muestras de nuestra cohorte de descubrimiento no se detectaron diferencias significativas entre el tejido no tumoral y el resto nódulos, salvo a excepción de la cg10705422 con los FVPTC (con un rango intercuartílico muy amplio) (**Figura 39c**). Asimismo, cabe destacar que cuando se implementó este modelo de tres CpG de Park y colaboradores para clasificar las muestras de nuestra cohorte de descubrimiento, se observó una dramática reducción del rendimiento con respecto al anteriormente mostrado con los nódulos con arquitectura papilar. Al discriminar el estado de malignidad de las lesiones de tipo folicular solo se pudo alcanzar una precisión de 51 %, una sensibilidad de 29 % y una especificidad de 68 % (**Figura 39d**).

5 Traslación del sistema de clasificación de diagnóstico de lesiones tiroideas malignas a un modelo desarrollado mediante pirosecuenciación con bisulfito

Para demostrar la aplicabilidad del sistema de clasificación en otras condiciones experimentales más rentables que permitieran una posible traslación a la práctica clínica, se validaron los resultados obtenidos con las 3 sondas CpG (cg17154646, cg21915100 y cg10189462) identificadas en la cohorte de descubrimiento utilizando los análisis de pirosecuenciación clásicos.

Se pudo observar una correlación fuerte y significativa entre los niveles de metilación del ADN obtenidos de los arrays de metilación con los resultados de la técnica de pirosecuenciación con bisulfito (coeficiente de correlación promedio = 0,98, p-valor < 0,001) (**Anexo 1: Tabla S5; Anexo 2: Figura S3**). Lo que indica que el estado de metilación de estas muestras también puede abordado por un enfoque experimental menos costoso y más rápido.

Para validar la aplicación del clasificador de tres CpG en este escenario, se analizó el estado de metilación de los nódulos de la cohorte de descubrimiento y se pudo observar que

la hipometilación significativa en las muestras malignas se mantenía en todas las sondas estudiadas (**Figura 40a**). Asimismo, se reconstruyó el modelo *Random Forest* en nuestro conjunto de datos de descubrimiento, pero utilizando los valores de metilación del ADN obtenidos tras la pirosecuenciación con bisulfito. La validación cruzada de este modelo arrojó una precisión de clasificación (90 %), sensibilidad (94 %) y especificidad (86 %) similares a los anteriormente obtenidos con la cohorte de descubrimiento (**Figura 40b**).

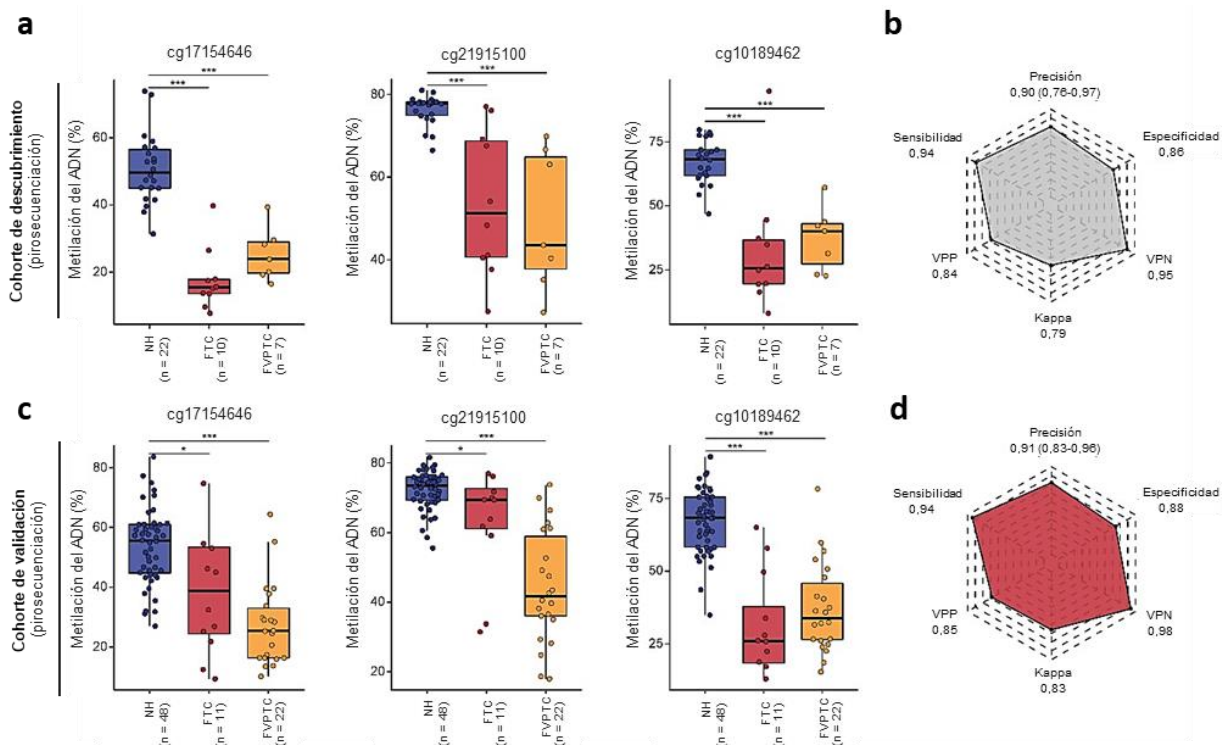


Figura 40 | Validación del clasificador epigenético de tres CpG (cg17154646, cg21915100 y cg10189462) mediante la técnica de pirosecuenciación con bisulfito. Panel de la izquierda: (a-c) Diagramas de caja que indican el valor de metilación del ADN de tres sitios CpG en diferentes conjuntos de muestras y categorías de cáncer de tiroides con el modelo *Random Forest* entrenado con los valores de metilación mediante la técnica de pirosecuenciación con bisulfito en la cohorte de descubrimiento (a) y el conjunto de datos de validación adicional (c). Panel de la derecha: (b-d) Gráficos de radar que ilustran el rendimiento del modelo de tres CpG obtenido en el contexto del conjunto de datos de descubrimiento (b) y el conjunto de datos de validación adicional (d). La significación estadística entre las condiciones se calculó mediante una prueba no paramétrica (test de Wilcoxon; * p-valor < 0,05, ** p-valor < 0,01, *** p-valor < 0,001).

Viendo que el modelo se mantenía robusto en estas condiciones se procedió a ampliar la cohorte de muestras para corroborar estas observaciones. Se realizó una última ronda de validación de nuestro clasificador mínimo de CpG mediante ensayos de pirosecuenciación de ADN modificado utilizando una cohorte independiente de 82 muestras (48 NH, 12 FTC y 22 FVPTC) (**Anexo 1: Tabla S6**). Una vez más, independientemente de las tres sondas testadas, los carcinomas presentaron niveles de metilación significativamente menores que los NH (**Figura 40c**). Para esta nueva cohorte se determinó el estado de malignidad con el modelo

Random Forest y el clasificador alcanzó una precisión del 91 %, con una especificidad del 88 % (42 de 48 muestras benignas) y una sensibilidad del 94 % (31 de 33 muestras malignas) (**Figura 40d**).

Los resultados confirman el poder de discriminación de nuestro sistema mínimo de clasificación epigenética, superando a los sistemas de clasificación basados en biomarcadores de metilación anteriores y haciendo viable el uso de nuestro clasificador mínimo de tres CpG a otras técnicas más rentables como la pirosecuenciación con bisulfito.

6 Puntuaciones de malignidad en los nódulos tiroideos según el sistema mínimo de clasificación epigenético

Una vez desarrollado el modelo mediante el algoritmo *Random Forest* y validado el rendimiento del clasificador epigenético utilizando una técnica más rentable basada en pirosecuenciación con bisulfito, se procedió a predecir el comportamiento de las muestras de las diferentes cohortes investigadas en función de la metilación de las tres CpG del clasificador (cg17154646, cg21915100 y cg10189462).

Inicialmente, para tratar de ajustar el rendimiento del clasificador desarrollado y evitar que las muestras malignas se diagnosticaran erróneamente, se calcularon las puntuaciones de malignidad (MS) para cada muestra analizada tal como se detalla en la **sección 8 de Material y Métodos con la Ecuación 1**. En el conjunto de datos de descubrimiento, empleando los valores de metilación de los arrays 850k, la mayoría de las muestras de NH que originalmente se clasificaron como tejido benigno mostraron MS por debajo de 0,3 (19/22; media 0,09) (**Figura 41a; Anexo 2: Figura S4**). Por ello, se estableció este valor como punto de corte para separar las muestras benignas de las malignas en todas las interpretaciones que a continuación se detallan.

La mayoría de las muestras malignas mostraron MS por encima del punto de corte establecido (FTC y FVPTC; MS promedio de 0,87 y 0,85, respectivamente) (**Figura 41a; Anexo 1: Tabla S3 y Anexo 2: Figura S4**). Únicamente un FVPTC de la cohorte de descubrimiento presentó valores inferiores (0,19) y, por lo tanto, se clasificó incorrectamente como benigno. Además, también se otorgaron puntuaciones a las muestras de la cohorte de Park y colaboradores (**Anexo 1: Tabla S4**) [159]. En este caso, todas las biopsias correspondientes a tejido no tumoral fueron clasificadas correctamente con valores MS inferiores 0,003. Únicamente una de las muestras malignas, diagnosticado como TPTC, fue clasificada como

benigna con un valor MS de 0,28; mientras que el resto de los tumores presentaron puntuaciones superiores a 0,45. Como ya se mencionó el clasificador designó a los NIFTP de este grupo de muestras como malignos, y cuando se les otorgó una puntuación MS obtuvieron valores comprendidos entre 0,65 y 1 (**Anexo 1: Tabla S4**).

Curiosamente, ya fuera en el contexto de los arrays a nivel de todo el genoma o los experimentos de pirosecuenciación, cuando se analizaron los FTA de la cohorte de descubrimiento a nivel de los tres sitios CpG no se observaron diferencias significativas con respecto a los NH (**Anexo 1: Tabla S3 y Tabla S6; Anexo 2: Figura S5a y b**). Las biopsias correspondientes a los FTA se dispersaron a lo largo de un amplio rango de valores MS en la cohorte de descubrimiento (0 y 0,76) (**Figura 41**), resultados que concuerdan con las dispersiones anteriormente reportadas en el nivel del epigenoma o en el nivel de CpG única. Es decir, los perfiles de metilación de las muestras de FTA parecen representar un continuo entre estados tumorales benigno y maligno. De hecho, la inclusión de estos casos de FTA (n = 15) en el desarrollo del clasificador de *Random Forest* resultó en una disminución del rendimiento del modelo debido a una clasificación incorrecta de algunas muestras como nódulos malignos. Este resultado indica que al menos el 53,3 % de las muestras de FTA (8 de 15) pueden clasificarse como benignos si tomamos como punto de corte el MS más alto de los NH. Siendo conservadores, con un MS por debajo de 0,05, el 40 % se diagnosticarían con precisión como casos verdaderamente benignos. Estos resultados sugieren que un número significativo de pacientes con FTA no presentan características malignas a nivel molecular y pueden beneficiarse de la utilización del clasificador y podrían evaluar cirugías innecesarias.

Para verificar estas observaciones en una cohorte independiente, validamos estos datos utilizando muestras benignas (FTA, n = 22; y NH, n = 48) y malignas (FTC n = 11; y FVPTC n = 22) adicionales mediante ensayos de pirosecuenciación con bisulfito (**Anexo 1: Tabla S6**). De nuevo, la mayoría de los nódulos hiperplásicos mostraron MS por debajo de 0,3 (42/48; promedio 0,12), mientras que la mayoría de las muestras de FTC y FVPTC obtuvieron valores más altos que este límite (promedios de 0,67 y 0,87, respectivamente) (**Figura 41b**). En este caso, con un punto de corte de 0,3, solo un FTC fue clasificado como benigno (falso negativo) con una puntuación límite de 0,29. Por otro lado, no se observaron diferencias significativas entre las muestras de NH y FTA en los tres CpG analizados (**Anexo 2: Figura S5c**). Asimismo, en los nódulos benignos, las puntuaciones de FTA fueron muy heterogéneas, con valores comprendidos entre 0 y 1. Como se puede observar (**Figura 41c**), con un punto de corte 0,3,

de un total de 70 muestras benignas al menos el 81 % de ellas (42 NH y 15 FTA) fueron correctamente diagnosticadas con nuestras tres CpG clasificador. Un tercio de las muestras FTA mostraron características malignas claras, lo que indica que la clasificación basada en el aprendizaje automático podría ayudar a priorizar las intervenciones quirúrgicas y, particularmente, en aquellos casos asociados con lesiones tiroideas citológicamente indeterminadas.

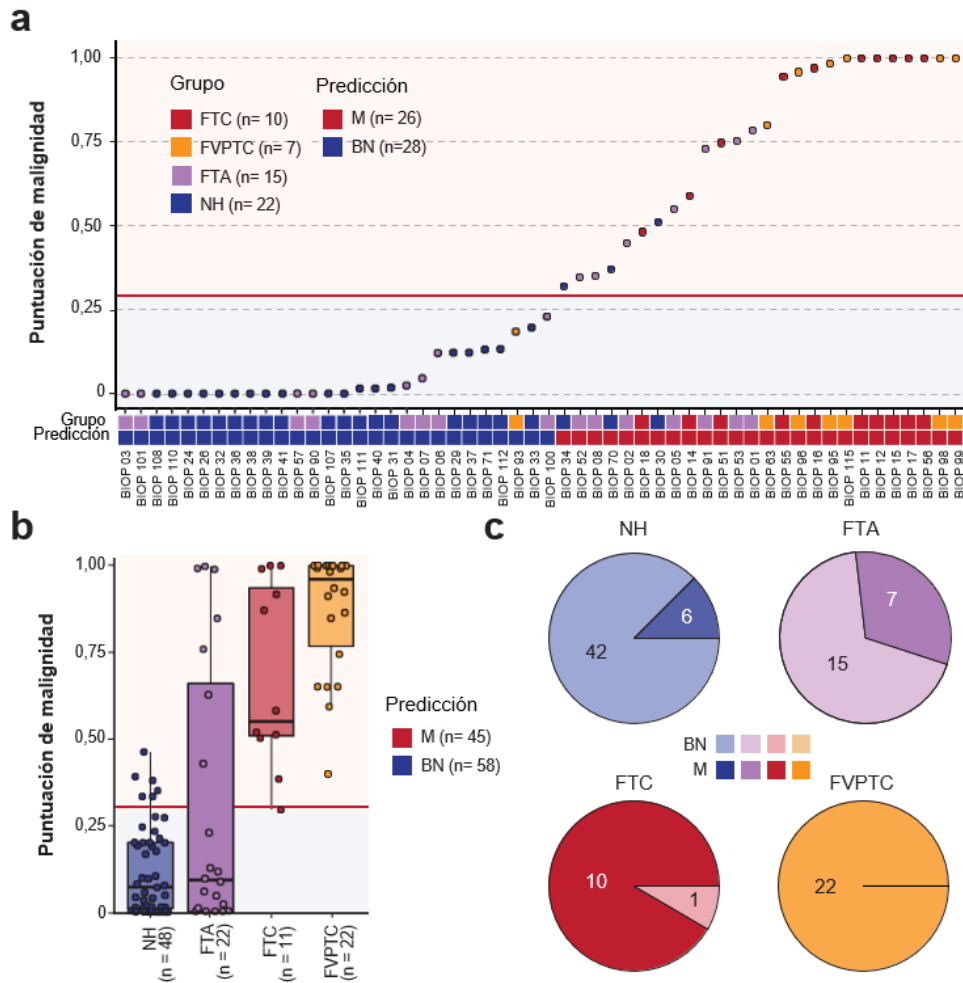


Figura 41 | Puntuaciones de malignidad inferidas para cada nódulo tiroideo a través del clasificador desarrollado con 3 CpG. (a) Diagrama de puntos que muestra las puntuaciones MS de las muestras utilizadas en el conjunto de datos de entrenamiento. Se indica el tipo de cáncer y la predicción diagnóstica (benigno, BN o maligno, M) de cada muestra utilizando el clasificador de bosque aleatorio de tres CpG. (b) Diagrama de caja que indica las puntuaciones MS inferidas para cada muestra utilizadas en el conjunto de datos de validación. (c) Gráficos circulares que muestran el número de muestras clasificadas como malignas o benignas en los diferentes subgrupos de nódulos tiroideos del conjunto de datos de validación.

Asimismo, habiendo probado su eficacia en lesiones tumorales no foliculares con los arrays de metilación, se comprobó también su rendimiento en una cohorte independiente de PTC (n = 15) y un pequeño grupo de tumores oncocíticos (HTC, n = 1; y HTA, n = 6) (**Anexo 1:**

Tabla S6). Los resultados indicaron que sistema de clasificación también es consistente en esta cohorte independiente. Como era de esperar, las muestras PTC mostraron valores MS elevados, a excepción de uno. Curiosamente, tras la reevaluación clínica del paciente presentó en el mismo lóbulo un NH, sin embargo, al no poseer muestra de los dos tumores no podemos descartar un posible cruce de muestras.

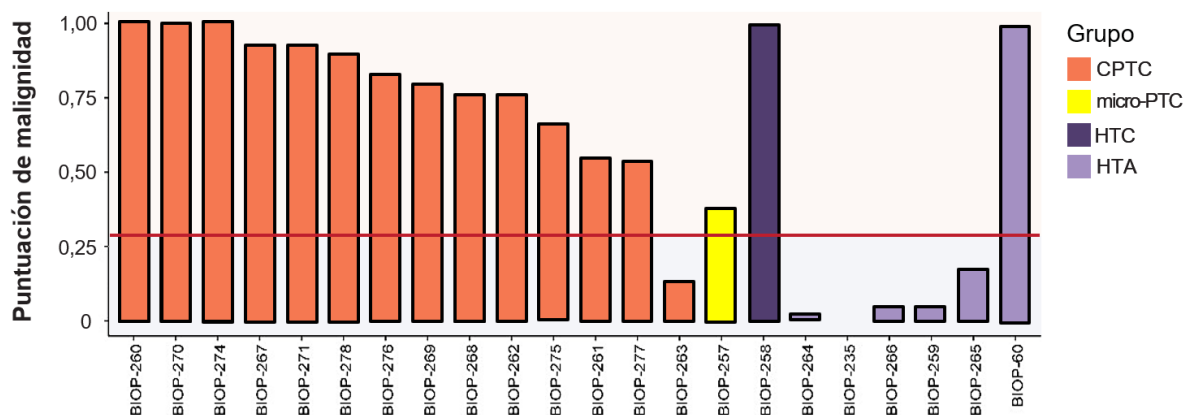


Figura 42 | Puntuaciones de malignidad inferidas para la cohorte de validación de tumores papilares y oncocíticos a través del clasificador desarrollado con 3 CpG. Gráfico de barras que indica los MS una cohorte adicional de muestras de PTC (CPTC y microcarcinoma), HTA y HTC. El punto de cohorte entre benignidad y malignidad (línea roja) se fijó en 0,3 en todas las gráficas y cohortes analizadas.

Por otro lado, los biobancos reportaron la presencia de células oncocíticas en algunos de los tumores de la cohorte de entrenamiento, por lo que el clasificador también podría distinguir el diagnóstico de este tipo de muestras. Para comprobar su utilidad en este tipo de tumores se evaluó una cohorte de 7 biopsias con una cantidad de células oncocíticas superior al 75 %. La mayoría de los nódulos diagnosticados histológicamente como HTA tras la operación (5 de 6), mostraron valores MS relacionados con un comportamiento más indolente (**Figura 42**). Mientras que el HTC mostró una clara tendencia maligna, lo que sugiere que este sistema de clasificación epigenética puede ser válido para otros subtipos de cáncer de tiroides. Sin embargo, aunque los resultados son prometedores, la cohorte de tumores oncocíticos es reducida y sería necesario ampliar el tamaño muestral en futuros estudios.

Considerando un rango de prevalencia de cáncer de 20-40 %, una especificidad del 80 % y sensibilidad del 90 % se considera como una prueba robusta [198]. Con una prevalencia del 39 % en la cohorte de validación el clasificador descrito podría describirse como tal, ya que la inclusión de todas las muestras analizadas en la cohorte de validación tanto benignas (NH, FTA y HCA) como malignas (FTC, FVPTC, PTC y HCC) en función de las puntuaciones MS obtenidas y un punto de cohorte de 0,03, dieron como resultado general: una sensibilidad del

96 % (IC 95 %: 90-100 %), una especificidad del 82 % (IC 95 %: 73-90 %) y unos VPP y VPN del 77 % (IC 95 %: 66-88 %) y del 97 % (IC 95 %: 93-100 %), respectivamente.

7 Validación del sistema de clasificación epigenética en punciones de nódulos tiroideos

El material genético de las PAAF se puede obtener a partir de punciones frescas, mediante el lavado final de la aguja con suero salino o a través de un pinchazo extra que el paciente voluntariamente accede a proporcionar (sin que se comprometa su diagnóstico). Además, en la bibliografía existe evidencia de que es posible extraer material de la extensión citológica de archivo del aspirado tras ser teñida y analizada por los anatomopatólogos [204, 247]. Por ello, antes de probar la capacidad de clasificación del sistema de diagnóstico desarrollado, se comprobó que tipo de muestra de partida era la más idónea para ser tratada y sometida a los ensayos de pirosecuenciación con bisulfito. Para ello se partió de PAAF (ya fuera uno extra o el lavado de la aguja) (n = 24), citologías de archivo (n = 5) o punciones frescas junto con su frotis (n = 23), y se realizó la extracción de ADN mediante el método convencional CIA. Posteriormente, se comprobó el rendimiento de la extracción (**Anexo 2: Figura S6**) y se observó que el material genético obtenido directamente a partir de los aspirados era el más adecuado para realizar las validaciones subsecuentes, ya que presentó menos contaminaciones. Dados los resultados se decidió hacer la validación del clasificador directamente del material obtenido de las PAAF.

Finalmente, probamos la capacidad de diagnóstico del sistema de clasificación *Random Forest* en un panel de muestras de PAAF sometidas a ensayos de pirosecuenciación del ADN. El análisis se realizó en una cohorte adicional de 33 PAAF, en la que se analizaron los niveles de metilación de las 3 CpG que componen el clasificador (**Anexo 1: Tabla S7**). Las muestras se agruparon según el riesgo de malignidad reportado por la clasificación TBSRTC: Benigno o Bethesda II, Indeterminado o Bethesda III/IV y Maligno o Bethesda V/IV (**Tabla 3**).

Cuando se investigaron las diferencias de metilación para cada CpG entre los distintos grupos Bethesda, solo se observó un descenso significativo en la que correspondía al gen *NCP2* en las categorías Indeterminado y Maligno (**Anexo 2: Figura S7**). Para la determinación de benignidad o malignidad se estableció el valor MS de 0,3 como punto de corte tal como se comentó en el apartado anterior (**Figura 43a**). Así, las predicciones obtenidas con las tres CpG del modelo clasificaron correctamente el 75 % de las muestras que normalmente son

derivadas a cirugía (3 de 4; incluidos Bethesda V y VI) y del 80 % de los casos citológicos que se concluyeron como benignos (16 de 20 Bethesda II) tras el análisis citopatológico (**Figura 43b**).

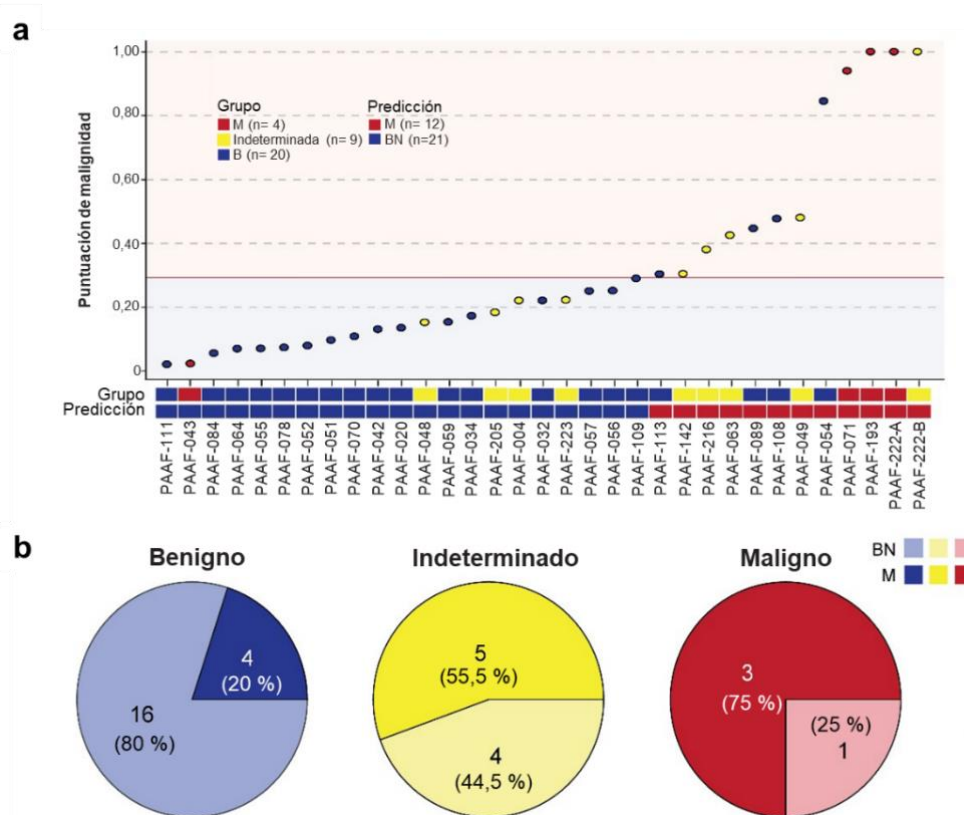


Figura 43 | Puntuaciones de malignidad inferidas para cada PAAF a través del clasificador desarrollado con 3 CpG. (a) Diagrama de puntos que muestra las puntuaciones de malignidad (MS) del cáncer inferidas de las muestras utilizadas en el conjunto de datos de PAAF. Se indica el tipo de cáncer y la predicción diagnóstica (BN benigno o M maligno) de cada muestra utilizando el clasificador Random Forest de tres CpG. (b) Gráficos circulares que muestran el número de muestras clasificadas como malignas o benignas en las PAAF agrupadas según el riesgo de malignidad de la clasificación TBSRTC (Benigno o Bethesda II, Indeterminado o Bethesda III/IV, Maligno V/VI)

En el grupo de PAAF indeterminadas (AUS/FLUS y FN/FSN), 5 muestras se clasificaron como malignas y 4 como benignas (**Figura 43b**). Los nódulos que se derivaron a cirugía fueron evaluados y diagnosticados por anatomopatólogos especializados (**Anexo 1: Tabla S7**). De tal manera que los nódulos con MS asociados a tumores malignos se clasificaron como 2 PTC (microcarcinoma y clásico), 2 FVPTC y 1 FTA, mientras que los benignos se confirmaron como tales.

En la **Figura 44 (a-e)** se pueden observar algunos ejemplos de las PAAF analizadas y la relación entre el diagnóstico citológico y la puntuación de malignidad obtenida con el clasificador. Cabe destacar la PAAF-205, que inicialmente se diagnosticó como indeterminada según los criterios de TBSRTC. Mientras que se predijo como benigna con nuestro clasificador,

RESULTADOS

presentando un MS de 0,18 (**Figura 44c**). Tras el tiempo recomendado por las guías oficiales, al paciente se le realizó una segunda PAAF, presentando en este caso un diagnóstico benigno Bethesda II.

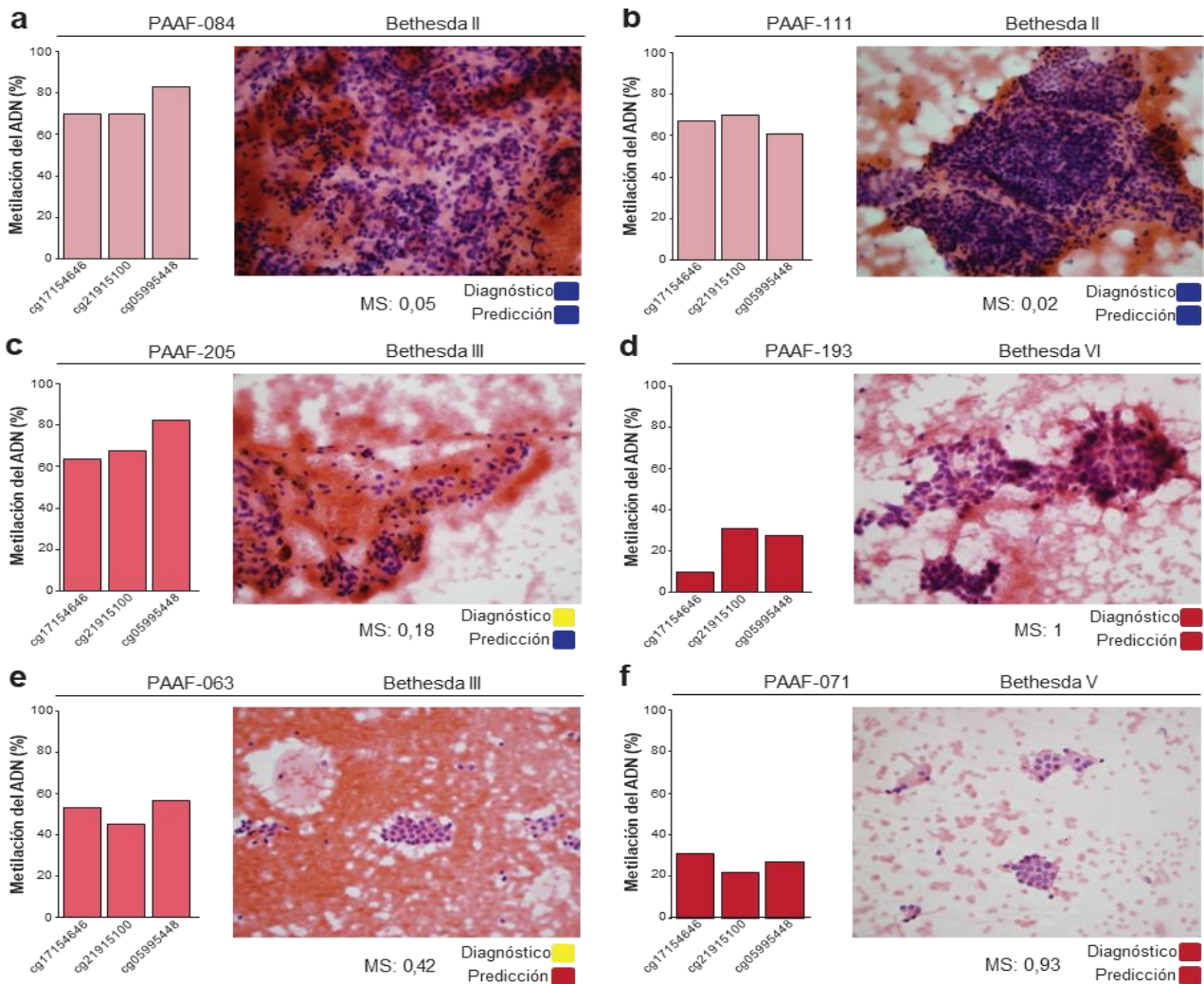


Figura 44 | Ejemplos del rendimiento del clasificador en la cohorte de PAAF analizada. (a-f) Gráficos de barras que muestran el porcentaje de metilación para cada CpG, en seis ejemplificaciones de PAAF de pacientes. La puntuación de malignidad asignada por el clasificador se indica como MS debajo de cada imagen. Asimismo, se muestra el diagnóstico inicial (benigno (azul, BN), indeterminado (amarillo) y maligno (rojo, M)) como el predicho por el clasificador (benigno (azul, BN) y maligno (rojo, M)). (a y b) PAAF-084 y PAAF-111, respectivamente: fondo hemático que muestra buena representación de células foliculares redondeadas sin atipia con distribución macro y microfolicular. Bethesda II. (c) PAAF-205: fondo hemático coloide acompañado de macrófagos y cúmulos de células predominantemente microfoliculares. Sin atipia. Dado que se observaron macrofoliculos aislados, se optó por Bethesda III. (d) PAAF-193: fondo hemático con abundante celularidad epitelial, de aspecto folicular con morfología papilar, en la que se observan ejes fibrovasculares. Atipia celular por: pérdida de la relación núcleo-citoplasma, hiper cromasia, irregularidad nuclear, frecuentes hendiduras nucleares y ocasionales pseudo inclusiones. Bethesda VI. (e-f) Mismo paciente en distintas tomas de muestra. (e) PAAF-063: fondo hemático que presenta ligera hiper celularidad constituida por células foliculares organizadas mayoritariamente en macrofoliculos, aunque también se observaban algunos microfoliculos. Atipia debido a cariomegalia y superposición ocasional. Compatible con Bethesda III. (f) PAAF-071: fondo hemático con celularidad aumentada; Células foliculares con atipia (cariomegalia, irregularidad nuclear, superposición y hendiduras ocasionales). Bethesda V. La significación estadística entre las condiciones se calculó mediante una prueba no paramétrica (test de Wilcoxon; * p-valor < 0,05, ** p-valor < 0,01, *** p-valor < 0,001).

Curiosamente, dos de las muestras incluidas en esta cohorte pertenecían al mismo paciente. Inicialmente, en nódulo tiroideo sospechoso presentó un Bethesda III al realizar la punción. Debido a al diagnóstico indeterminado y ante la sospecha clínica de malignidad, se realizó una segunda PAAF (**Figura 44e y f**), con una separación de 19 días entre toma de muestra. En este momento cuando la citología fue analizada por el profesional sanitario el diagnóstico concluyó que era un Bethesda VI, y por lo tanto fue remitido a cirugía. Cuando se analizaron estas muestras con el clasificador, la primera PAAF indeterminada presentó un MS de 0,42 (calificada como maligna), mientras que la segunda ya reportada con una sospecha de malignidad alta, mostró un MS de 0,93. Finalmente, el paciente fue derivado a cirugía y tras el postoperatorio el tumor fue finalmente categorizado como un FVPTC. Estos resultados destacan la utilidad de nuestro sistema de clasificación en muestras PAAF en la toma de decisiones en la práctica clínica, ya que proporciona puntuaciones de malignidad que están altamente correlacionadas con el grado y la agresividad de un nódulo determinado.

DISCUSIÓN

1 Las limitaciones de la punción por aspiración con aguja fina en el diagnóstico preoperatorio de los nódulos tiroideos ponen de manifiesto la necesidad de nuevos biomarcadores

Los nódulos tiroideos se encuentran entre las enfermedades más comunes del sistema endocrino [81]. Aunque la mayoría son benignos, un 5-15% de los nódulos pueden ser carcinomas en función de la edad, el sexo, los antecedentes de exposición a la radiación, los antecedentes familiares y otros factores [248, 249]. La incidencia del cáncer de tiroides en España ha ido en aumento en los últimos 25 años, pasando de 3,42 casos en 1997, a 9,40 casos en 2015 y a 12 casos en 2022 por cada 100.000 habitantes [84, 88]. Sin embargo, son tumores que en general tienen un buen pronóstico, con tasas de supervivencia cercanas al 100 % [89]. A pesar de ello, la gran prevalencia de la enfermedad hace necesaria un triaje exhaustivo de los nódulos.

La Punción por Aspiración con Aguja Fina es una técnica precisa ampliamente utilizada y reconocida como proceder de carácter orientador y diagnóstico en los nódulos tiroideos. Sin embargo, el 25-30 % de todas las punciones de tiroides no dan un diagnóstico citológico definitivo siendo designadas como indeterminadas [182]. Principalmente, esta limitación es resultado de la incapacidad que tienen las PAAF para distinguir entre los distintos subtipos de neoplasias foliculares [184, 185]. Las características citológicas frecuentemente se superponen, no siendo posible establecer la malignidad del nódulo evaluando la presencia de invasión vascular o capsular. Ante la posibilidad de malignidad, el diagnóstico final debe confirmarse mediante técnicas diagnósticas más invasivas y los pacientes suelen ser derivados a cirugía [178, 181, 186]. Tras la operación, solo alrededor del 25 % de las lesiones resultan malignas [182] por lo que el riesgo de cáncer no es lo suficientemente alto como para respaldar definitivamente la cirugía. Este hecho da lugar a tiroidectomías innecesarias, que generan complicaciones que acompañan a los pacientes de por vida [191, 192].

Por lo tanto, existe una urgente necesidad de desarrollar métodos efectivos para la detección de PAAF indeterminadas. En consecuencia, el descubrimiento de nuevos biomarcadores podría ayudar en la toma de decisiones del médico tratante para manejar con mayor certeza los nódulos tiroideos. El diseño experimental para alcanzar este objetivo general se ha basado en dos estrategias: por un lado, el análisis de las alteraciones moleculares y, por otro, la caracterización de los perfiles de metilación de neoplasias foliculares que con

mayor frecuencia son categorizadas incorrectamente tras la realización de la PAAF. Derivado de esta última aproximación, surge la identificación de alteraciones de metilación específicas, susceptibles de ser utilizadas como biomarcadores. A continuación, se discutirán los distintos aspectos abordados en este trabajo.

1.1 Análisis el perfil mutacional de las lesiones nodulares originadas a partir de células foliculares

La reciente disponibilidad de tecnologías de secuenciación del genoma ha permitido un progreso significativo en la comprensión de los mecanismos moleculares que subyacen al cáncer de tiroides [189]. En general, los carcinomas diferenciados son genéticamente simples, con una carga de mutaciones somáticas relativamente baja [38, 250]. El Atlas del Genoma del Cáncer identificó las mutaciones y otras alteraciones genéticas más frecuentes en el PTC, y de esta manera se cubrió más del 90 % de los cánceres de tiroides, lo que lo convierte en uno de los cánceres humanos mejor caracterizados desde un punto de vista genético [119]. Sin embargo, aunque esto sea así existe un sesgo hacia un único histotipo y sus variantes. Por ello, múltiples trabajos han tratado de rellenar los huecos y sentar las bases de las alteraciones más frecuentes que ocurren en el resto de subtipos tumorales de tiroides [125, 126, 140, 251, 252].

Las alteraciones en los genes implicados en las vías de señalización MAPK y PI3K-AKT son frecuentes en los tumores tiroideos y, como cabría esperar, también se pueden observar en nuestras cohortes analizadas. Los datos de secuenciación muestran alteraciones en *NRAS* y *HRAS* en al menos el 30 % (27 % en la cohorte de validación) y 10 % de los que los FTC, respectivamente. Nuestros resultados están en concordancia con trabajos publicados previamente en los que *NRAS* se puede encontrar en el 18-50 % de este subtipo tumoral, mientras que *HRAS* entre el 7-10 % [125, 126, 251]. Además, también se observaron en la cohorte de descubrimiento mutaciones en *NRAS*, *HRAS* en el 14 y 28 % de los FVPTC, respectivamente (13 % de mutaciones en ambos genes en la cohorte de validación), lo que confirma que tiene un perfil genómico similar a FTC como se informó anteriormente [125, 253]. Los FVPTC presentan mutaciones en *H/K/NRAS* en el 30-48 % [119, 125], pero también son muy frecuentes las alteraciones en *BRAF*. Normalmente, los valores esperados para esta mutación están comprendidos entre el 5-25 % [119, 125], mientras que en nuestras cohortes se puede observar una prevalencia del 38 % (11 de 29 FVPTCs).

Los primeros trabajos relacionados con la evaluación molecular se centraron en el análisis de BRAF^{V600E} [254]. Si bien el estado mutacional de este gen se puede considerar como un marcador de pronóstico relativamente sensible para los PTC, no se puede usar de forma aislada para este propósito en tumores tiroideos [189, 254]. La ausencia de mutaciones en *BRAF* no es suficiente para descartar la malignidad de los tumores y, por ello, la sensibilidad de las pruebas moleculares mejoró al introducir los paneles de genes. El primer panel aportado por Nikiforov y colaboradores en 2011 (ThyroSeq[®] v0), fue una prueba molecular de 7 alteraciones compuesta tanto por mutaciones como fusiones de genes (*BRAF*, *HRAS*, *KRAS*, *NRAS*, *PAX8/PPAR γ* , *RET/PTC1* y *RET/PTC3*). Sin embargo, la sensibilidad era baja, situada entre el 57 % y el 63 % en PAAF (AUS/FLUS y FN/FSN) [202, 255]. Posteriormente, el advenimiento de la tecnología de secuenciación de nueva generación promovió la expansión de los paneles de genotipado para PAAFs de tiroides. Las sucesivas versiones de ThyroSeq[®] analizan un número cada vez mayor de alteraciones genéticas, con una mayor sensibilidad (94 % en su última versión (v3)) [205].

La presencia de mutaciones comunes entre formas benignas y malignas [125, 126, 251] y la falta de variantes genéticas comunes en una proporción considerable de carcinomas de tiroides [256, 257] conducen a una pérdida significativa de sensibilidad en estos métodos. Un ejemplo, son las mutaciones ocurrentes en *RAS*. Como ocurre con *BRAF*, esta mutación por sí sola tiene baja sensibilidad y especificidad, y para tener un impacto clínico las alteraciones en este gen en las neoplasias tiroideas no deben considerarse de forma aislada, sino en conjunto con otras anomalías genéticas en un contexto más general [256]. Asimismo, la detección de mutaciones *RAS* está relacionada con la arquitectura folicular y son detectables en todas las etapas de la tumorigénesis tiroidea, encontrándose en los adenomas, los carcinomas foliculares diferenciados y los carcinomas indiferenciados [98, 113, 117]. Parece que se requieren alteraciones genéticas adicionales para transformar las células tiroideas, pero la mutación *RAS* sola con expresión fisiológica, no puede [258]. Nuestra cohorte no es una excepción y se pueden encontrar mutaciones en todos los subtipos analizados, tanto benignos como malignos. Cabe destacar, aunque en diferentes proporciones, ambos grupos de muestras benignas (FTA y NH) de la cohorte de descubrimiento mostraron alteraciones genómicas, siendo el total de mutaciones en FTA comparable al de FTC. Los niveles

comparables evidenciados entre estos dos subgrupos no permitieron la identificación de firmas moleculares.

En una visión prospectiva, un clon de células mutantes en un individuo sano puede morir, persistir sin expandirse, formar un tumor benigno no invasivo o progresar a cáncer. Por ello, sobre todo si los estudios van dirigidos a la prevención o la detección precoz del cáncer, es necesario mirar hacia adelante en el tiempo y considerar resultados con distintas trayectorias. El concepto tradicional de mutaciones conductoras y pasajeras es atractivo porque sugiere que se encontrado la "causa" y, sin embargo, pueden llevar a malentendidos sobre el desarrollo tumoral, con consecuencias no deseadas que incluyen sobrediagnóstico, sobretratamiento y falta de identificación de las verdaderas fuentes de riesgo [259]. Como ya se ha comentado, es ampliamente aceptado que incluso los tejidos normales pueden albergar mutaciones que antes se consideraban características definitorias de las neoplasias malignas [260, 259]. Aún más, pueden ocurrir alteraciones consideradas de "bajo riesgo" en genes concretos, ya que su propagación depende del azar, pero la activación de la ruta biológica en su conjunto puede explicar una proporción sustancial de la predisposición para desarrollar un tumor maligno [259]. Asimismo, carcinomas con fenotipos similares pueden albergar cientos de mutaciones cada uno, pero estas tienen poco solapamiento entre sí [261]. En este escenario, la dicotomía canónica de mutaciones conductoras y pasajeras podría no reflejar necesariamente el complejo paisaje de los genomas del cáncer. Estos términos se podrían refinar: considerándose mutación "ventajosa" cuando confieren características que favorecen la supervivencia o crecimiento somático en un entorno determinado (de modo que las células que lo poseen aumentan en número a expensas de otras células) o "desventajosas" cuando conducen a una reducción del crecimiento somático o la supervivencia [259]. Una mutación puede denominarse "predisponente" cuando tiende hacia un fenotipo de cáncer (invasión de tejidos y metástasis), ya sea al conferir una de las características de malignidad o al desestabilizar el genoma y, por lo tanto, promover la aparición de otras mutaciones [259].

El examen del genoma del cáncer ha demostrado la complejidad y la heterogeneidad entre y dentro de las localizaciones de los órganos y, sobre todo, dentro de los propios tumores. La disección de las alteraciones genéticas que acontecen en un modelo de progresión del cáncer podría ser más adecuado para la comprensión actual de la biología tumoral [262]. Uno que incluya lesiones indolentes y aquellas que se diseminan temprana o

tardíamente. Mientras tanto, los términos conductor y pasajero deben utilizarse con prudencia, sobre todo en el contexto de la prevención y la detección precoz del cáncer, en el que el sobrediagnóstico y el sobretratamiento se reconocen cada vez más como retos importantes.

En la actualidad, no existe una única prueba molecular óptima que pueda confirmar o descartar definitivamente la malignidad en todos los casos de citología indeterminada, y se necesitan datos de resultados a largo plazo que demuestren la utilidad clínica frente a la PAAF que se utiliza de rutina. Teniendo en cuenta el elevado coste asociado, tanto de las cirugías como de los clasificadores comerciales con amplios paneles de alteraciones [263], el desarrollo de alternativas novedosas y rentables representa un área de estudio prometedora.

1.2 Evaluación de las alteraciones epigenómicas mediante *arrays* de metilación del ADN

La identificación de firmas epigenéticas específicas de tejido para el diagnóstico de nódulos tiroideos es un campo de investigación prometedor. Como ya se ha comentado, múltiples trabajos han respondido importantes preguntas sobre la metilación del ADN e investigado su potencial como biomarcadores en el diagnóstico de las lesiones tumorales tiroideas.

En la bibliografía científica actual, al menos en lo que respecta a *arrays* de metilación o la secuenciación de bisulfito de representación reducida, el número de PTC estudiados supera casi en 18 veces a las lesiones foliculares como FTC y FTA (**Tabla 2**). Sin embargo, el diagnóstico diferencial FTC/FTA constituye actualmente un desafío ya que los métodos de detección rutinarios (como los ultrasonidos o las PAAFs) no pueden distinguir entre estos nódulos malignos y los benignos [176, 186]. Por ello, en este trabajo se utilizaron *arrays* de metilación del ADN para determinar el perfil de metilación de las muestras con patrón folicular: NH, FTA, FTC y FVPTC.

Inicialmente, se valoraron las diferencias y similitudes en el número de sondas diferencialmente metiladas en una cohorte de descubrimiento. Como era de esperar, los tumores malignos presentaron un mayor número de sitios CpG diferencialmente metilados frente a los nódulos hiperplásicos y al tejido tiroideo normal. En el caso de los FVPTC se observaron un mayor número de sondas hipometiladas con respecto a ambos tipos de muestras. Aunque los resultados concuerdan con lo reportado en la literatura científica [152,

153, 155, 156, 158, 159, 165], hay que tener en cuenta que los FVPTC se estudian en conjunto con otras variantes de PTC (**Tabla 2**). Por el contrario, los FTC presentan un mayor número de sondas hipermetiladas con respecto a los NH. Yao y colaboradores (2022), usando la misma plataforma que en este estudio, encuentran también un mayor número de sondas con niveles más elevados de metilación en FTC versus lesiones tiroideas benignas [168]. Sin embargo, cuando se enfrentaron los FTC contra el tejido normal del proyecto ENCODE se encontraron un mayor número de sondas hipometiladas que de hipermetiladas. Estos resultados contrastan con los reportados para este tipo de carcinoma con los arrays de metilación 450K [153, 156, 161, 162, 165]. La discrepancia observada en nuestros resultados puede deberse a la diferencia en el número de sondas existente entre las plataformas utilizadas, ya que a los 450.000 sitios originales se agregan 413.745 adicionales (un total de 853.307 sitios CpG) [236]. Debido a que, según nuestro conocimiento, este es el primer estudio con este tipo de array que analiza las diferencias entre el tejido normal y los carcinomas foliculares de tiroides, no se pueden sacar conclusiones sobre la direccionalidad de los cambios en este momento. Serían necesarios más estudios que pudieran añadir más información al respecto. Cabe mencionar que se ha descrito que grandes secciones del genoma se encuentran hipometilados en los tumores benignos de tiroides y que aumentan con la progresión hacia estadios más agresivos de la enfermedad [164]. Curiosamente, estudios que analizan los cambios existentes entre los carcinomas más desdiferenciados y el tejido normal encuentran niveles de metilación equiparables a los observados en este estudio [156, 163]. Por ello, sería esclarecedor determinar también los subtipos del grupo de FTC (sobre todo los ampliamente invasivos) y comprobar las similitudes con las formas más agresivas de las variantes foliculares.

Por otro lado, en nuestro trabajo se describió un estado de metilación del ADN global similar para las muestras de tiroides FTC y FTA. Estas similitudes también se mantuvieron en el contexto de la distribución genómica de las sondas dmCpG. Los análisis iniciales indicaron que los adenomas presentaban el mismo porcentaje de metilación global que los FTC y, además, que un 65,8 % de las dmCpG de los FTA contra el tejido normal y un 41 % con respecto a las hiperplasias ya se encuentran alteradas en estas muestras analizadas. Aún más, cabe remarcar que cuando se realizaron los análisis de agrupación los FTA mostraron un patrón disperso en varios grupos. Los análisis de clusterización mostraron que los adenomas se posicionaron en múltiples grupos, tanto con nódulos benignos como con malignos. Estudios previos desarrollados por Bisarro dos Reis y colaboradores (2017) han informado que, además

de las similitudes morfológicas y genéticas, los FTA se agrupan con los FTC cuando se utilizan enfoques de clusterización no supervisado [156]. Asimismo, Affinito y colaboradores (2019) se encontraron que los FTA mostraban un perfil de metilación del ADN intermedio entre los FTC y el tejido tiroideo normal [161]. De acuerdo con nuestras observaciones, también podemos observar que los FTC muestran un perfil de metilación del ADN profusamente alterado en comparación con el tejido tiroideo normal. Por otro lado, Mancikova y colaboradores (2014) encontraron pocas diferencias en la metilación entre adenomas y muestras de tejido tiroideo normal que también se observaron en FTC. Demostraron que la hipermetilación del promotor asociado con los FTA también era evidente en los FTC, lo que sugiere la ganancia progresiva de los niveles de 5mC sobre el ADN durante el proceso tumorigénico de adenomas a carcinomas [153]. De la misma forma, Timp y colaboradores (2014) observaron igual tendencia de cambios de bloques de hipometilación en los FTA que en los cánceres completamente desarrollados [164]. El número de bloques y la magnitud de los cambios aumentaban en los tumores tiroideos más agresivos. Estos resultados sugieren que, potencialmente, a pesar de que los FTA se han identificado clásicamente como lesiones tiroideas benignas, una proporción de las muestras pueden tener alteraciones epigenéticas que les confieren un fenotipo molecular maligno.

1.3 Identificación nuevos biomarcadores que permitan el desarrollo de un clasificador de tipo diagnóstico mediante técnicas de *aprendizaje automático*

El diagnóstico de malignidad requiere una evaluación patológica postoperatoria, en la que se investigue la invasión capsular y vascular. El diagnóstico preciso es importante para mejorar la calidad de vida de los pacientes, por lo tanto, existe una necesidad urgente de desarrollar herramientas de diagnóstico preoperatorio eficaces para las lesiones foliculares.

Para favorecer la discriminación correcta de las lesiones tiroideas benignas y malignas, explotamos el potencial de los métodos de clasificación basados en el aprendizaje automático para construir un clasificador *Random Forest* utilizando las 12.845 sondas diferencialmente metiladas encontrados en las muestras FTC, FVPTC y NH en comparación con el tejido tiroideo normal. Mediante el uso de un modelo mínimo compuesto por solo tres sondas CpG (cg17154646, cg21915100 y cg10189462) se logró una alta sensibilidad y especificidad en los conjuntos de datos de descubrimiento (100 % y 90 %, respectivamente) y validación (97 % y

88 %, respectivamente), lo que destaca la capacidad de nuestro modelo para diferenciar las formas benignas de las malignas de las lesiones tiroideas, incluidos los PTC (95 % y 100 %, respectivamente). Asimismo, adicionalmente se exploró el rendimiento del clasificador en una pequeña cohorte de tumores oncocíticos. Los resultados preliminares parecen indicar que el conjunto de los tres biomarcadores de metilación podría diferenciar también este tipo de neoplasias. Sin embargo, no se pueden hacer asunciones certeras ya que la proporción de muestras estudiadas no es suficiente para asegurar su utilidad en este subtipo tumoral.

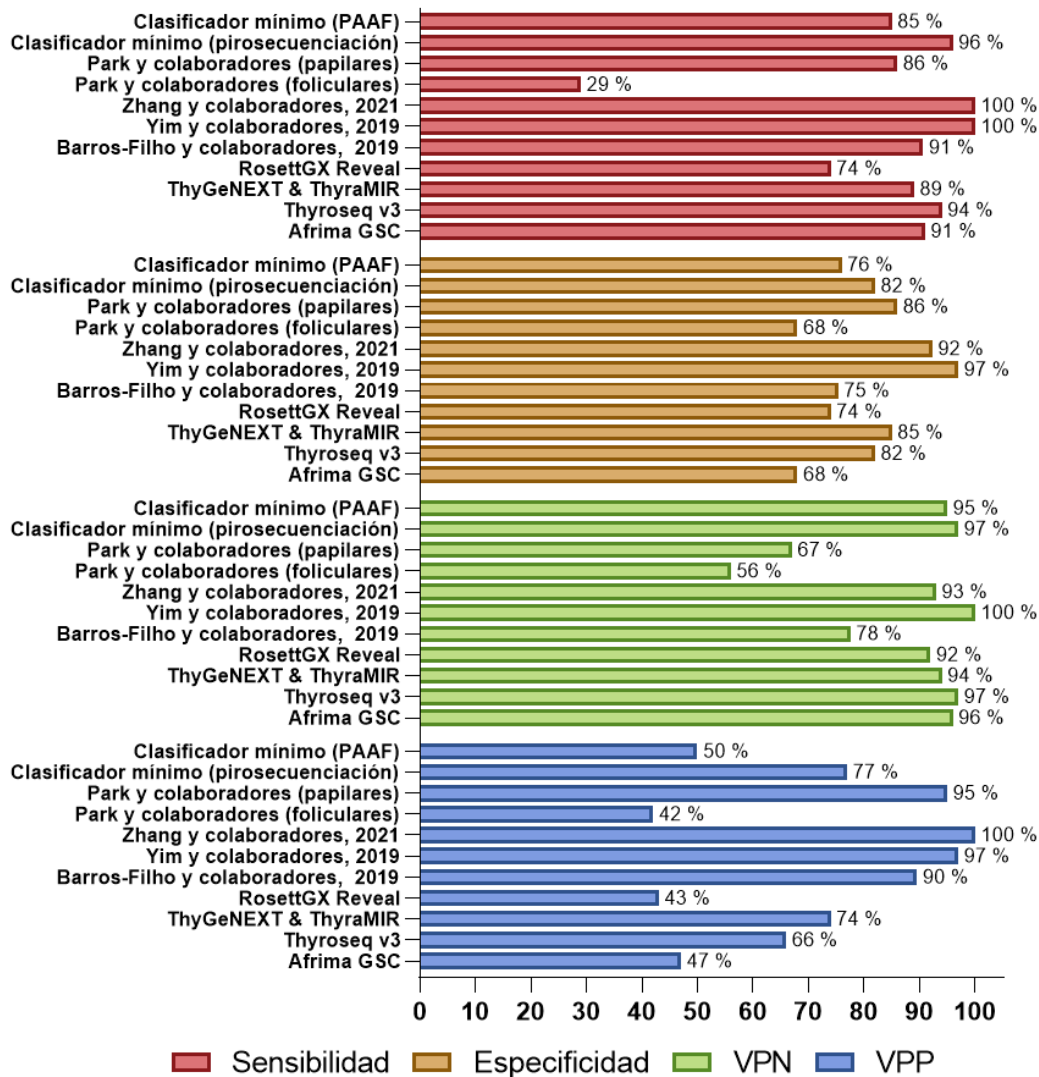


Figura 45 | Resumen del rendimiento de los clasificadores diagnósticos encontrados en la bibliografía en comparación con los desarrollados por nuestro laboratorio. Los datos referentes a “Park y colaboradores” hacen referencia al rendimiento de las 3 CpG nombradas en su trabajo usadas en combinación en un modelo Random Forest. Los valores se refieren tanto para las muestras con arquitectura papilar usadas en su estudio como para las muestras foliculares de nuestra cohorte de descubrimiento. Asimismo, los valores del “clasificador mínimo” son los obtenidos en este estudio para la cohorte de validación en este presente trabajo (pirosecuenciación en la cohorte de validación completa y PAAF). El resto de valores se referencian en la Figura 23 y la Tabla 2.

En general, en la cohorte completa de validación el rendimiento del clasificador alcanza una sensibilidad del 96 %, una especificidad de 82 % y unos valores predictivos positivos y negativos del 77 % y 97 %, respectivamente. Así, el rendimiento del clasificador mínimo utilizando la técnica de pirosecuenciación con bisulfito clasificó correctamente un 95 % de los carcinomas e incorrectamente un 18 % de las lesiones benignas de la cohorte de validación completa. Estudios similares han logrado describir clasificadores epigenéticos con un gran rendimiento (**Figura 45**). Yim y colaboradores (2019), han identificado con éxito una firma de metilación del ADN de diagnóstico basada en la técnica RRBS [169]. A pesar de las limitaciones intrínsecas de este enfoque (como la variación experimental entre librerías, que se manifiesta como diferencias en la cobertura y profundidad genómica) [264], los autores lograron una alta especificidad (97 %) y sensibilidad (100 %) en la discriminación de nódulos tiroideos benignos frente a malignos, principalmente formas clásicas de los PTC. Zhang y colaboradores (2021), llevaron a cabo una aproximación similar centrándose solo en los FTC, UMP y FTA (incluidos los tumores oncocíticos) [162]. Con su modelo lograron separar los FTC de los adenomas, con una gran especificidad 100 % y sensibilidad 92,3 % y solo con un carcinoma clasificado como benigno. Sin embargo, dado el elevado número de sitios CpG necesarios para sus estimaciones (373 localizaciones genómicas y 70 bloques de haplotipos, respectivamente), así como los recursos computacionales, mano de obra necesarios para el correcto análisis de este tipo de datos, y que no han probado su rendimiento en las PAAF, su aplicabilidad en la práctica clínica habitual para el diagnóstico preoperatorio tiene un alcance limitado. Igualmente, las pruebas moleculares comerciales como Thyroseq v3 o Afirma GSC, a pesar de tener un gran rendimiento(**Figura 45**) y haber probado su utilidad en punciones, el coste asociado a un diagnóstico correcto en Estados Unidos puede llegar a alcanzar los 14.277 \$ y 17.873 \$, respectivamente [263, 265].

Un trabajo llevado a cabo Barros-Filho y colaboradores (2019), emplean arrays de metilación para encontrar marcadores de metilación candidatos y posteriormente utilizan la técnica de pirosecuenciación con bisulfito para clasificar tumores tiroideos utilizando 5 sitios CpG, obteniendo en las muestras de validación una sensibilidad de 90,7 % y especificidad de 75,4 %, comparables a los presentados en nuestro trabajo (96 % y 82 % respectivamente) [165] (**Figura 45**). En PAAF, su clasificador mostró, en tumores malignos confirmados, una sensibilidad del 86,7 % y una especificidad del 89,5 %. Mientras que, en casos con citología indeterminada, 100 % y 89 %, respectivamente. Las PAAF de nuestra cohorte en el modelo

desarrollado presentaron un porcentaje de verdaderos carcinomas del 85 % y de verdaderos benignos del 76 %. Aunque el valor predictivo negativo es alto 95 %, y el positivo es del 50 %. El clasificador epigenético basado en pirosecuenciación descrito en este trabajo, representa un método rentable y rápido que puede ser aplicado en PAAF's para complementar el diagnóstico preoperatorio de los nódulos tiroideos utilizando solo 3 sitios CpG. Una aproximación similar llevaron a cabo Park y colaboradores, en su trabajo se seleccionaron 3 marcadores de metilación del ADN candidatos para discriminar NIFTP y el carcinoma papilar de tiroides, a saber: cg10705422 AUC 0,83, sensibilidad 76 %, especificidad 80 %; cg17707274 AUC 0,83, sensibilidad 76 %, especificidad 79 %; y cg26849382 AUC 0,8; sensibilidad 74 %, especificidad 78 % [159]. Sin embargo, solo se encontraron diferencias significativas de metilación con los PTC y no se analizaron PAAF. Curiosamente, el rendimiento se vio muy reducido cuando en nuestro trabajo se generó un modelo *Random Forest*, con las tres CpG anteriormente mencionadas, para clasificar las lesiones tiroideas de nuestra cohorte de descubrimiento. Asimismo, como en su trabajo, no se encontraron diferencias significativas entre los tumores foliculares en las diferentes CpG. Los PTC presentan perfiles de metilación diferentes a los FTC [153, 156, 165], por lo que cabría esperar que los marcadores epigenéticos, al no estar entrenado para detectar carcinomas foliculares, no puedan aplicarse a ambos subtipos por igual. Esta posibilidad se evaluó con las tres CpG (cg17154646, cg21915100 y cg10189462) del clasificador mínimo epigenético desarrollado en esta tesis y los tumores papilares de la cohorte de Park y colaboradores obteniendo una precisión de discriminación del 95 %.

Cabe mencionar, que en el proceso inicial de desarrollo de nuestro clasificador se excluyeron las sondas dmCpG en sangre, sin embargo, las muestras tumorales son heterogéneas y pueden contener células distintas de las epiteliales foliculares, dando lugar a la ausencia de firmas epigenéticas específicas [169]. Esta contaminación no solo se puede encontrar en las biopsias postquirúrgicas de los nódulos, sino también en las PAAF. En las punciones es frecuente observar distintas proporciones de sangre y células, como glóbulos blancos, linfocitos infiltrantes de tejidos, fibroblastos estromales o tejido normal adyacente [169, 203, 266]. Por ello, se debe enfatizar que los análisis de las PAAF en nuestro estudio pueden estar influenciados por ciertas limitaciones técnicas, ya que las muestras se obtuvieron de los lavados de las agujas y la calidad de las punciones secundarias puede ser una restricción importante en términos del rendimiento del algoritmo. Por lo que sería

necesario realizar un estudio de verificación analítica que pruebe si el conjunto de tres CpGs puede mantener la misma validez clínica crítica, así como el mismo rendimiento en condiciones de variabilidad técnica e interferencias potencialmente encontradas en un entorno real. Este estudio incluiría dos componentes clave: sensibilidad y especificidad analítica. El primero de ellos, la sensibilidad analítica o el límite de detección de la prueba, hace referencia al contenido tumoral mínimo aceptable que produce un resultado preciso [267, 322]. Por otro lado, la especificidad analítica es la capacidad del ensayo para medir, en una muestra, un organismo o sustancia en particular en lugar de otros [267]. En este caso, evaluando hasta qué punto una muestra contaminada con sangre puede seguir siendo fiable. Así, Afirma GSC sigue siendo una prueba precisa en presencia de hasta un 75 % de sangre y una dilución del 75 % de material maligno en PAAF [322]. Mientras que Thyroseq v3 aseguran que con su panel de mutaciones la clasificación de una muestra se mantiene precisa hasta el 88 % del contenido de sangre y, aproximadamente, un 12 % de las células tumorales en un contexto de células tiroideas no neoplásicas [267]. Por ello, sugerimos que, además de ampliar el número de muestras, estos resultados se complementen con la evaluación de otros parámetros, como la celularidad de la PAAF, el porcentaje relativo de células sanguíneas, la calidad y cantidad del ADN genómico mínimo necesario en las muestras o el análisis de duplicados técnicos, que pueden mejorar aún más el rendimiento del modelo propuesto.

1.4 Discriminación pronóstica de las lesiones foliculares indolentes según los biomarcadores epigenéticos

La definición oficial de FTA establece que se trata de una neoplasia benigna, encapsulada y no invasiva, sin características nucleares de PTC [98]. Sin embargo, como ya se ha discutido en apartados anteriores la presencia de tanto de mutaciones comunes como alteraciones epigenéticas entre los adenomas y las lesiones foliculares, conduce no solo a que los adenomas se agrupen junto a los tumores malignos, sino también a una pérdida significativa de sensibilidad de los métodos de diagnóstico basados en alteraciones génicas.

Por lo tanto, planteamos la hipótesis de que nuestro sistema de clasificación de tres CpG puede proporcionar algunas pistas en relación con el estado de malignidad de un adenoma determinado en función de la proporción de votos benignos/malignos calculados durante el paso de predicción. De esta manera, los FTA fueron evaluados usando el clasificador de 3 CpGs y se otorgaron puntuaciones de malignidad para predecir el comportamiento de

estos tumores. Cuando se predijeron las puntuaciones de malignidad para los FTA, estas fueron muy heterogéneas y comprendieron valores entre 0 y 1. Bisarro dos Reis (2017), describen un modelo de metilación del ADN para predecir el pronóstico de los carcinomas bien diferenciados, en términos de recurrencia [156]. Curiosamente, el 47 % de muestras benignas presentaban valores intermedios en el clasificador generado por estos autores. Posteriormente, el mismo grupo de investigadores genera un clasificador, esta vez para la detección preoperatoria de nódulos tiroideos, y observan que aproximadamente el 24 % de las lesiones tiroideas benignas analizadas presentan puntuaciones propias de nódulos malignos [165]. Por otro lado, Zhang y colaboradores (2021) construyeron un modelo de predicción de FTC utilizando 70 bloques de haplotipos de metilación como marcadores, que distinguían entre tejidos FTA y FTC con un alto grado de precisión (**Figura 45**) [162]. Los FTA presentaban puntuaciones de malignidad comprendidas entre 0-0,5, mientras que los valores de los FTC se encontraban 0,46 y 1. Este clasificador también se utilizó con el propósito de ayudar a determinar el pronóstico de tumores de potencial maligno incierto, y encontraron que puntuaciones más altas de metilación mostraron mayor concordancia con una alta frecuencia de variaciones genéticas.

Las alteraciones en la expresión de genes y proteínas pueden dar lugar a cambios en la morfología y función de las células, lo que se conoce como fenotipo celular. Sin embargo, los cambios moleculares pueden acontecer antes de que se de este cambio observable por el ojo humano. Aunque no existe un modelo aceptado para la patogenia de los carcinomas foliculares, recientemente se demostró mediante el perfil de expresión global [268] que los FTA pueden representar un intermediario biológico entre el tejido normal y FTC. La transición de un adenoma a un carcinoma folicular en varios pasos, similar a la que ocurre en el cáncer de colon y otros adenocarcinomas, no es universalmente aceptado porque no se reconoce el carcinoma folicular in situ y la evidencia de la evolución del adenoma a formas malignas es escasa [269, 270]. Sin embargo, un considerable número de autores parecen plantear la hipótesis de que un continuo hacia carcinoma en este tipo de tumores es plausible [125, 153, 161, 251, 270, 271]. Los resultados presentados en nuestro trabajo refuerzan la noción de que los FTA pueden presentar cierto potencial para malignizar, y que las similitudes morfológicas, genéticas y epigenéticas observadas con los FTC concurren con la dificultad del patólogo para distinguir entre nódulos benignos y malignos al evaluar las PAAF. Las coincidencias sugieren que los FTA y FTC están etiológicamente relacionados, o más específicamente, que los FTA son

lesiones que pueden (en algunos casos, pero no en probablemente no en todos) progresar hacia carcinomas [269, 270].

Asimismo, cabe señalar que los NIFTP del conjunto de datos de validación externa [159], fueron clasificados sistemáticamente por nuestro modelo de 3 CpG como tejido maligno. Inicialmente, Nikiforov y colaboradores (2016) propusieron revisar la nomenclatura de los FVPTC encapsulados no invasivos y renombrarlos como NIFTP [272]. Este planteamiento se basó en un estudio que realizaron en una cohorte de 109 pacientes en 2016, en el que comparaban la evolución de tumores FVPTC encapsulados invasivos y los que no lo eran. Tras aproximadamente 13 años de seguimiento, los individuos evaluados seguían vivos y sin evidencia de enfermedad. Tras este trabajo, a pesar de que se publicaron estudios que confirmaron estos resultados, muchos autores mostraron su escepticismo ante esta nueva nomenclatura [273]. Aunque los NIFTP tienen un comportamiento indolente, se ha descrito que algunas de estas lesiones pueden estar asociadas a metástasis [274], y notificando eventos oncológicos adversos en al menos un 6 % de los casos. Otro estudio, registro que el 9,4 % de los pacientes bajo la nueva nomenclatura de NIFTP padecieron fracaso tumoral locorregional [275].

Una de las críticas más importantes sobre esta clasificación, además de la posible variabilidad entre los diagnósticos de los distintos evaluadores, es que los patólogos que participaron en el estudio original no desconocían las alteraciones moleculares asociados a los tumores que se revisaban, y que estas se utilizaron para guiar el desarrollo de los criterios diagnósticos del NIFTP [275]. Curiosamente, pruebas moleculares realizadas en este tipo de tumores demostraron resultados “sospechosos” mediante Afirma, y las pruebas de ThyroSeq mostraron alteraciones moleculares que se encontraban también en lesiones malignas [132, 276, 277]. Además, se ha descrito que existen dos tipos de NIFTP altamente dependientes de la presencia de diferentes mutaciones (sobre todo *RAS* y *BRAF*): un tipo más “benigno” con perfiles de expresión similares a FTA y uno más “maligno” con perfiles de expresión similares a FVPTC invasivo [278]. Nuestro clasificador parece indicar que las células de los tumores NIFTP ya presentan características moleculares propias de tumores malignos en sus epigenomas (**Figura 46**) [272]. Lo que puede significar que este tipo de tumores de potencial maligno incierto sean un precursor temprano de las formas invasivas del FVPTC. No cabe duda de que el pronóstico relativamente bueno de los tumores NIFTP puede estar asociado con la presencia de una cápsula bien definida que rodea la masa tumoral interna. Sin embargo, el

riesgo de desarrollar resultados adversos no es despreciable. Por ello, los autores del artículo inicial en el que se describió este subtipo, y debido a la controversia generada, advierten a los investigadores que no se debe malinterpretar el retirar la palabra “cáncer” de la definición como sinónimo de neoplasia benigna [273]. Así se recoge en la cuarta clasificación de la OMS en la surgió el grupo de los tumores de potencial maligno incierto, para destacar su buen pronóstico y desalentando así un tratamiento excesivamente agresivo e incentivando el seguimiento [98, 117].

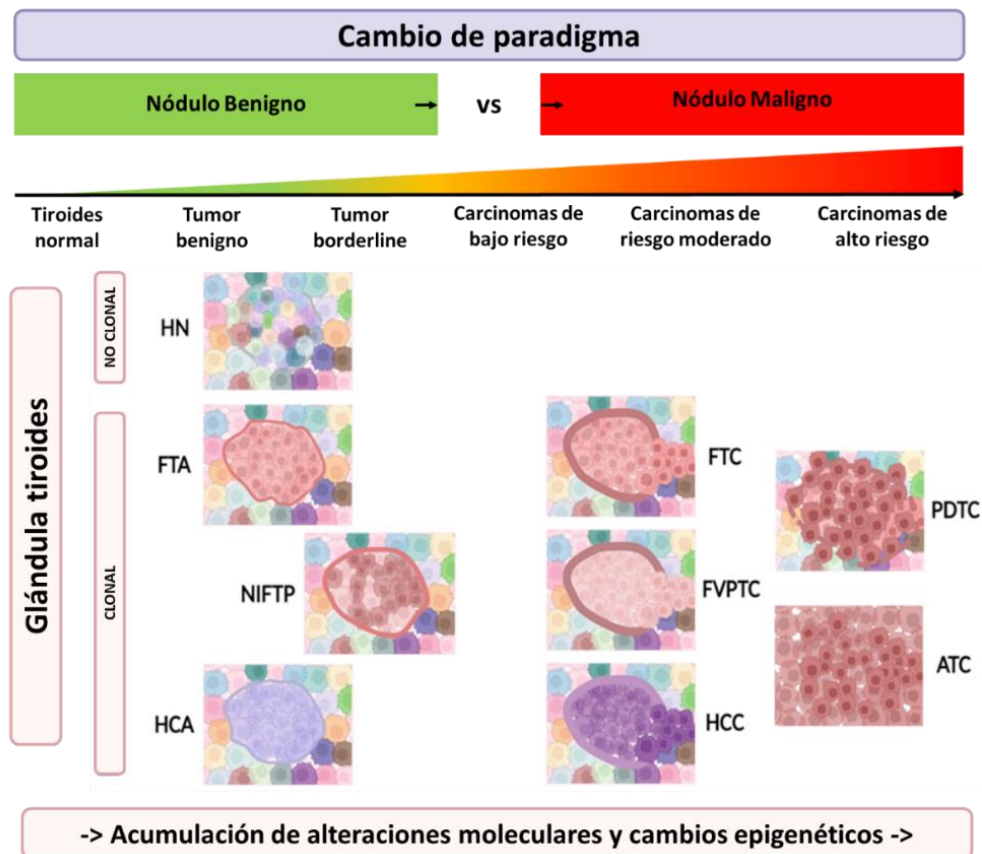


Figura 46 | Cambio de paradigma en el diagnóstico de los nódulos tiroideos. Comúnmente, la clasificación se basa en la distinción de nódulos tiroideos benignos y malignos que se encontraba en la tercera clasificación de la OMS para los tumores endocrinos. Aproximadamente el 20 % de los diagnósticos de tumores tiroideos tenían discrepancias con este esquema. Según la cuarta edición, todos los tumores tiroideos tienen cierto potencial para desarrollar metástasis y se elimina la distinción entre benignos y malignos. La estratificación del riesgo de los tumores tiroideos desde un riesgo muy bajo de recurrencia hasta un riesgo alto de recurrencia se muestra como un espectro continuo, desde un tumor benigno hasta un cáncer de alto riesgo. Los tumores adquieren mutaciones y cambios epigenéticos que confieren un crecimiento somático o una ventaja de supervivencia en un entorno dado.

Convencionalmente, las neoplasias se dividen en dos grupos, benigno y maligno (Figura 46). La principal característica que los separa es la capacidad de invadir y dar lugar a metástasis, que es exclusiva de los tumores malignos [260]. Dadas las implicaciones, los

clínicos no pueden esperar a que se produzca un evento adverso en los pacientes para diagnosticar la malignidad. Por ello, se deben apoyar en otras particularidades clinicopatológicas de los tumores para la diferenciación ente entidades. Por ejemplo, las neoplasias benignas muestran con mayor frecuencia límites bien definidos y tienden a recapitular la misma morfología del tejido de origen, con una mínima la atipia celular (o estando ausente) y sin hemorragias o áreas de necrosis [260]. Sin embargo, el comportamiento de los tumores benignos puede variar de indolente a significativamente mórbido y, en ocasiones, incluso fatal [279]. Además, el hecho de que un tumor no recurra o produzca metástasis tras la resección, no prueba que los nódulos sean realmente benignos [269]. La operación puede haber interrumpido la historia natural de los tumores, hacia formas más agresivas que hubieran culminado en metástasis [269].

Muchas neoplasias no encajan perfectamente en esta clasificación binaria. La distinción dicotómica sigue siendo esquivada para categorías enteras de tumores, incluidos los carcinomas de tiroides [260, 279]. Por ello, se han desarrollado sistemas de clasificación basados en el comportamiento y estratifican a los tumores en función del riesgo para guiar el manejo clínico. Las categorías intermedias se utilizan para designar aquellos procesos patológicos que no pueden determinarse como completamente benignos, pero que no cumple con todos los criterios de malignidad. Las neoplasias de comportamiento incierto de órganos urinarios, tracto genital, cavidad bucal y aparato digestivo, son tumores que muchas veces pueden incluirse en esta categoría [262, 280]. En el caso de los tumores de la glándula tiroides, a diferencia de la tercera edición de la OMS, la cuarta versión recoge que los de tumores de órganos endocrinos tienen algún potencial metastásico. Asimismo la quinta edición parece indicar que la clasificación de los tumores tiroideos seguirá desarrollándose en esta línea y los tumores se clasificarán en función del riesgo [122]. El reto consiste en redefinir el cáncer en función de su comportamiento, utilizar los términos cuando proceda y cambiar los métodos de comunicación. La adaptación puede implicar cambios en el enfoque del cribado y el tratamiento de los nódulos tiroideos, pues puede personalizarse en consecuencia para maximizar el beneficio para el paciente individual y la población [262, 281].

Es importante mencionar que este cambio en la estratificación de los pacientes no soluciona las limitaciones actuales que tienen las PAAF para discernir benignidad de malignidad en las categorías indeterminadas. Sin embargo, después de considerar las características clínicas y ecográficas, se pueden usar pruebas moleculares para complementar

los datos de evaluación del riesgo de malignidad en lugar de proceder directamente con la cirugía. Es interesante mencionar que, para desarrollar estas pruebas moleculares, los clasificadores se nutren de los tumores con los que se entrenan. Por ello, la continua reevaluación de la evidencia científica sobre la patogenia de los tumores tiroideos dará lugar a una mayor comprensión y a clasificadores más certeros. Esto significa que las pruebas moleculares pueden no solo ayudar a diagnosticar a los pacientes con nódulos indeterminados si no también en esta estratificación de riesgo y asistir en el seguimiento de los pacientes.

A pesar de las limitaciones, nuestro modelo es capaz de asignar puntuaciones de malignidad molecular a aquellas lesiones foliculares difíciles de diagnosticar, y también observamos una alta consistencia entre los resultados esperados y el diagnóstico patológico de PAAF, incluso con muestras obtenidas en diferentes puntos de tiempo. Aunque, la efectividad real de esta nueva herramienta de diagnóstico debe verificarse en un estudio multicéntrico prospectivo, creemos que este tipo de clasificador epigenético podría asistir en la toma de decisiones en el manejo de los pacientes con nódulos tiroideos.

CONCLUSIONES

De acuerdo con los objetivos presentados y a los resultados obtenidos en esta tesis, las conclusiones son:

1. Las alteraciones genéticas son complejas y aunque se encontraron mutaciones propias de los nódulos tiroideos, tanto benignos como malignos, no pudo establecerse una firma distintiva.
2. El empleo de arrays 850K de Illumina en el estudio de los cambios de metilación del ADN constituye una aproximación fiable, asequible y altamente reproducible, que permite establecer las similitudes y diferencias existentes entre las lesiones tiroideas.
3. Los niveles de metilación observados en los adenomas foliculares parecen indicar que estos tumores presentan perfiles intermedios entre los nódulos benignos y los realmente malignos.
4. El empleo algoritmos de aprendizaje automático basados en *Random Forest* constituye una aproximación fiable (concordante con los resultados de pirosecuenciación) para la identificación de biomarcadores en cáncer de tiroides.
5. El sistema de clasificación epigenético con las sondas diferencialmente metiladas, cg21915100, cg17154646 y cg10189462, permite otorgar puntuaciones de malignidad que se asocian con precisión al estado de malignidad de las biopsias interrogadas.
6. Las puntuaciones de malignidad de los adenomas foliculares y las neoplasias tiroideas foliculares no invasivas con características nucleares de tipo papilar indican que algunos nódulos ya presentan alteraciones moleculares propias de tumores malignos.
7. Los biomarcadores identificados tienen potencial relevancia clínica para el diagnóstico, evaluación de progresión de los nódulos tiroideos, y pueden complementar las técnicas de atención estándar actuales evitando un número razonable de intervenciones quirúrgicas.

REFERENCIAS

1. Roy AL, Conroy RS (2018) Toward mapping the human body at a cellular resolution. *MBoC* 29(15):1779–1785. <https://doi.org/10.1091/mbc.E18-04-0260>
2. Hemberger M, Dean W, Reik W (2009) Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington’s canal. *Nat Rev Mol Cell Biol* 10(8):526–537. <https://doi.org/10.1038/nrm2727>
3. Harper JW, Bennett EJ (2016) Proteome complexity and the forces that drive proteome imbalance. *Nature* 537(7620):328–338. <https://doi.org/10.1038/nature19947>
4. Esteve-Puig R, Bueno-Costa A, Esteller M (2020) Writers, readers and erasers of RNA modifications in cancer. *Cancer Letters* 474:127–137. <https://doi.org/10.1016/j.canlet.2020.01.021>
5. Feero WG, Guttmacher AE, Collins FS (2010) Genomic Medicine — An Updated Primer. *New England Journal of Medicine* 362(21):2001–2011. <https://doi.org/10.1056/NEJMra0907175>
6. Waddington CH (2012) The epigenotype. 1942. *Int J Epidemiol* 41(1):10–13. <https://doi.org/10.1093/ije/dyr184>
7. Ecker S, Pancaldi V, Valencia A, Beck S, Paul DS (2018) Epigenetic and Transcriptional Variability Shape Phenotypic Plasticity. *BioEssays* 40(2):1700148. <https://doi.org/10.1002/bies.201700148>
8. Reik W (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447(7143):425–432. <https://doi.org/10.1038/nature05918>
9. Kanherkar RR, Bhatia-Dey N, Csoka AB (2014) Epigenetics across the human lifespan. *Front Cell Dev Biol* 2. <https://doi.org/10.3389/fcell.2014.00049>
10. Esteller M (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 8(4):286–298. <https://doi.org/10.1038/nrg2005>
11. Fraga MF, Esteller M (2007) Epigenetics and aging: the targets and the marks. *Trends Genet* 23(8):413–418. <https://doi.org/10.1016/j.tig.2007.05.008>
12. Kulis M, Queirós AC, Beekman R, Martín-Subero JI (2013) Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1829(11):1161–1174
13. Gavrillov AA, Razin SV (2015) Compartmentalization of the cell nucleus and spatial organization of the genome. *Mol Biol* 49(1):21–39. <https://doi.org/10.1134/S0026893315010033>
14. Dekker J, Mirny L (2016) The 3D genome as moderator of chromosomal communication. *Cell* 164(6):1110–1121. <https://doi.org/10.1016/j.cell.2016.02.007>
15. Krefting J, Andrade-Navarro MA, Ibn-Salem J (2018) Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol* 16. <https://doi.org/10.1186/s12915-018-0556-x>
16. Soler-Oliva ME, Guerrero-Martínez JA, Bachetti V, Reyes JC (2017) Analysis of the relationship between coexpression domains and chromatin 3D organization. *PLOS Computational Biology* 13(9):e1005708. <https://doi.org/10.1371/journal.pcbi.1005708>
17. Mariño-Ramírez L, Jordan IK, Landsman D (2006) Multiple independent evolutionary solutions to core histone gene regulation. *Genome Biol* 7(12):R122. <https://doi.org/10.1186/gb-2006-7-12-r122>
18. Bach SV, Hegde AN (2016) The proteasome and epigenetics: zooming in on histone modifications. *Biomolecular Concepts* 7(4):215–227. <https://doi.org/10.1515/bmc-2016-0016>
19. Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res* 21(3):381–395. <https://doi.org/10.1038/cr.2011.22>
20. Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128(4):693–705. <https://doi.org/10.1016/j.cell.2007.02.005>
21. Smith ER, Cayrou C, Huang R, Lane WS, Côté J, Lucchesi JC (2005) A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol Cell Biol* 25(21):9175–9188.

- <https://doi.org/10.1128/MCB.25.21.9175-9188.2005>
22. Sawan C, Herceg Z (2010) Histone Modifications and Cancer. In: *Advances in Genetics*. Elsevier, pp 57–85
 23. Li B, Carey M, Workman JL (2007) The Role of Chromatin during Transcription. *Cell* 128(4):707–719. <https://doi.org/10.1016/j.cell.2007.01.015>
 24. Martin C, Zhang Y (2005) The diverse functions of histone lysine methylation. *Nature Reviews Molecular Cell Biology* 6(11):838–849. <https://doi.org/10.1038/nrm1761>
 25. Simon RP, Robaa D, Alhalabi Z, Sippl W, Jung M (2016) KATching-Up on Small Molecule Modulators of Lysine Acetyltransferases. *Journal of Medicinal Chemistry* 59(4):1249–1270. <https://doi.org/10.1021/acs.jmedchem.5b01502>
 26. Roth SY, Denu JM, Allis CD (2001) Histone acetyltransferases. *Annu Rev Biochem* 70:81–120. <https://doi.org/10.1146/annurev.biochem.70.1.81>
 27. Sterner DE, Berger SL (2000) Acetylation of histones and transcription-related factors. *Microbiol Mol Biol Rev* 64(2):435–459
 28. Teperino R, Schoonjans K, Auwerx J (2010) Histone methyl-transferases and demethylases; can they link metabolism and transcription? *Cell Metab* 12(4):321–327. <https://doi.org/10.1016/j.cmet.2010.09.004>
 29. Jones PA, Takai D (2001) The role of DNA methylation in mammalian epigenetics. *Science* 293(5532):1068–1070. <https://doi.org/10.1126/science.1063852>
 30. Bird AP (1986) CpG-rich islands and the function of DNA methylation. *Nature* 321(6067):209–213. <https://doi.org/10.1038/321209a0>
 31. Moore LD, Le T, Fan G (2013) DNA Methylation and Its Basic Function. *Neuropsychopharmacol* 38(1):23–38. <https://doi.org/10.1038/npp.2012.112>
 32. Antequera F (2003) Structure, function and evolution of CpG island promoters. *CMLS, Cell Mol Life Sci* 60(8):1647–1658. <https://doi.org/10.1007/s00018-003-3088-6>
 33. Brenet F, Moh M, Funk P, et al (2011) DNA Methylation of the First Exon Is Tightly Linked to Transcriptional Silencing. *PLoS One* 6(1):e14524. <https://doi.org/10.1371/journal.pone.0014524>
 34. Anastasiadi D, Esteve-Codina A, Piferrer F (2018) Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics & Chromatin* 11(1):37. <https://doi.org/10.1186/s13072-018-0205-1>
 35. Jjingo D, Conley AB, Yi SV, Lunyak VV, Jordan IK (2012) On the presence and role of human gene-body DNA methylation. *Oncotarget* 3(4):462–474. <https://doi.org/10.18632/oncotarget.497>
 36. Subramaniam D, Thombre R, Dhar A, Anant S (2014) DNA Methyltransferases: A Novel Target for Prevention and Therapy. *Front Oncol* 4. <https://doi.org/10.3389/fonc.2014.00080>
 37. Gowher H, Jeltsch A (2018) Mammalian DNA methyltransferases: new discoveries and open questions. *Biochemical Society Transactions* 46(5):1191–1202. <https://doi.org/10.1042/BST20170574>
 38. Zafón Llopis C, Gil J, Pérez-González B, Jordà M (2019) DNA methylation in thyroid cancer. *Endocr Relat Cancer*. <https://doi.org/10.1530/ERC-19-0093>
 39. López V, Fernández AF, Fraga MF (2017) The role of 5-hydroxymethylcytosine in development, aging and age-related diseases. *Ageing Res Rev* 37:28–38. <https://doi.org/10.1016/j.arr.2017.05.002>
 40. Shen L, Zhang Y (2013) 5-Hydroxymethylcytosine: generation, fate, and genomic distribution. *Curr Opin Cell Biol* 25(3):289–296. <https://doi.org/10.1016/j.ceb.2013.02.017>
 41. Kriaucionis S, Tahiliani M (2014) Expanding the epigenetic landscape: novel modifications of cytosine in genomic DNA. *Cold Spring Harb Perspect Biol* 6(10):a018630. <https://doi.org/10.1101/cshperspect.a018630>
 42. Laird A, Thomson JP, Harrison DJ, Meehan RR (2013) 5-hydroxymethylcytosine profiling as an indicator of cellular state. *Epigenomics* 5(6):655–669. <https://doi.org/10.2217/epi.13.69>
 43. Guo JU, Szulwach KE, Su Y, et al (2014) Genome-wide antagonism between 5-

- hydroxymethylcytosine and DNA methylation in the adult mouse brain. *Front Biol (Beijing)* 9(1):66–74. <https://doi.org/10.1007/s11515-014-1295-1>
44. Jin S-G, Jiang Y, Qiu R, et al (2011) 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res* 71(24):7360–7365. <https://doi.org/10.1158/0008-5472.CAN-11-2023>
45. Wen L, Tang F (2014) Genomic distribution and possible functions of DNA hydroxymethylation in the brain. *Genomics* 104(5):341–346. <https://doi.org/10.1016/j.ygeno.2014.08.020>
46. Wu X, Zhang Y (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet* 18(9):517–534. <https://doi.org/10.1038/nrg.2017.33>
47. Branco MR, Ficuz G, Reik W (2011) Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet* 13(1):7–13. <https://doi.org/10.1038/nrg3080>
48. Hanahan D, Weinberg RA (2000) The Hallmarks of Cancer. *Cell* 100(1):57–70. [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9)
49. Hanahan D, Weinberg RA (2011) Hallmarks of Cancer: The Next Generation. *Cell* 144(5):646–674. <https://doi.org/10.1016/j.cell.2011.02.013>
50. Hanahan D (2022) Hallmarks of Cancer: New Dimensions. *Cancer Discovery* 12(1):31–46. <https://doi.org/10.1158/2159-8290.CD-21-1059>
51. Singh Nanda J, Kumar R, Raghava GPS (2016) dbEM: A database of epigenetic modifiers curated from cancerous and normal genomes. *Scientific Reports* 6(1). <https://doi.org/10.1038/srep19340>
52. Copeland RA, Olhava EJ, Scott MP (2010) Targeting epigenetic enzymes for drug discovery. *Current Opinion in Chemical Biology* 14(4):505–510. <https://doi.org/10.1016/j.cbpa.2010.06.174>
53. Kulis M, Esteller M (2010) DNA Methylation and Cancer. In: *Advances in Genetics*. Elsevier, pp 27–56
54. Ehrlich M (2009) DNA hypomethylation in cancer cells. *Epigenomics* 1(2):239–259. <https://doi.org/10.2217/epi.09.33>
55. Nishiyama A, Nakanishi M (2021) Navigating the DNA methylation landscape of cancer. *Trends in Genetics* 37(11):1012–1027. <https://doi.org/10.1016/j.tig.2021.05.002>
56. Fraga MF, Ballestar E, Villar-Garea A, et al (2005) Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nature Genetics* 37(4):391–400. <https://doi.org/10.1038/ng1531>
57. Forhead AJ, Fowden AL (2014) Thyroid hormones in fetal growth and parturition maturation. *Journal of Endocrinology* 221(3):R87–R103. <https://doi.org/10.1530/JOE-14-0025>
58. Macchia PE (2000) Recent advances in understanding the molecular basis of primary congenital hypothyroidism. *Molecular Medicine Today* 6(1):36–42. [https://doi.org/10.1016/S1357-4310\(99\)01620-2](https://doi.org/10.1016/S1357-4310(99)01620-2)
59. Fernández LP, López-Márquez A, Santisteban P (2015) Thyroid transcription factors in development, differentiation and disease. *Nat Rev Endocrinol* 11(1):29–42. <https://doi.org/10.1038/nrendo.2014.186>
60. López-Márquez A, Carrasco-López C, Fernández-Méndez C, Santisteban P (2021) Unraveling the Complex Interplay Between Transcription Factors and Signaling Molecules in Thyroid Differentiation and Function, From Embryos to Adults. *Front Endocrinol (Lausanne)* 12:654569. <https://doi.org/10.3389/fendo.2021.654569>
61. Heredia-Jara A, Ortiz-Hidalgo C (2021) Anatomía microscópica normal de la glándula tiroides. Principios básicos para el residente de Endocrinología y anatomía patológica con una breve nota histórica. *Revista Colombiana de Endocrinología, Diabetes & Metabolismo* 8(4). <https://doi.org/10.53853/encr.8.4.688>
62. Bliss RD, Gauger PG, Delbridge LW (2000) Surgeon's Approach to the Thyroid Gland: Surgical Anatomy and the Importance of Technique. *World J Surg* 24(8):891–897. <https://doi.org/10.1007/s002680010173>
63. Jameson JL, Mandel SJ, Weetman AP (2018) Trastornos de la glándula tiroides. In: Jameson JL,

- Fauci AS, Kasper DL, Hauser SL, Longo DL, Loscalzo J (eds) Harrison. Principios de Medicina Interna, 20e. McGraw-Hill Education, New York, NY
64. Gil Carcedo-Sañudo E, Heras-Flórez P de las, Herrero-Calvo D, et al (2020) Anatomía quirúrgica de las glándulas tiroideas y paratiroides. *Revista ORL* 11(2):161–178. <https://doi.org/10.14201/orl.21494>
65. Sternberg SS (1997) *Histology for Pathologists*
66. Ontogeny, Anatomy, Metabolism and Physiology of the Thyroid. In: *Thyroid Disease Manager*. <https://www.thyroidmanager.org/chapter/ontology-anatomy-metabolism-and-physiology-of-the-thyroid/>. Accessed 5 Oct 2022
67. Kim H-Y, Mohan S (2013) Role and Mechanisms of Actions of Thyroid Hormone on the Skeletal Development. *Bone Res* 1(1):146–161. <https://doi.org/10.4248/BR201302004>
68. Tuncel M (2017) Thyroid Stimulating Hormone Receptor. *Mol Imaging Radionucl Ther* 26(Suppl 1):87–91. <https://doi.org/10.4274/2017.26.suppl.10>
69. Vieira IH, Rodrigues D, Paiva I (2022) The Mysterious Universe of the TSH Receptor. *Front Endocrinol (Lausanne)* 13:944715. <https://doi.org/10.3389/fendo.2022.944715>
70. Sorrenti S, Baldini E, Pironi D, et al (2021) Iodine: Its Role in Thyroid Hormone Biosynthesis and Beyond. *Nutrients* 13(12):4469. <https://doi.org/10.3390/nu13124469>
71. Citterio CE, Targovnik HM, Arvan P (2019) The role of thyroglobulin in thyroid hormonogenesis. *Nat Rev Endocrinol* 15(6):323–338. <https://doi.org/10.1038/s41574-019-0184-8>
72. Luo Y, Ishido Y, Hiroi N, Ishii N, Suzuki K (2014) The Emerging Roles of Thyroglobulin. *Advances in Endocrinology* 2014:e189194. <https://doi.org/10.1155/2014/189194>
73. Rousset B, Dupuy C, Miot F, Dumont J (2000) Chapter 2 Thyroid Hormone Synthesis And Secretion. In: Feingold KR, Anawalt B, Boyce A, et al (eds) *Endotext*. MDText.com, Inc., South Dartmouth (MA)
74. Szanto I, Pusztaszeri M, Mavromati M (2019) H2O2 Metabolism in Normal Thyroid Cells and in Thyroid Tumorigenesis: Focus on NADPH Oxidases. *Antioxidants (Basel)* 8(5):126. <https://doi.org/10.3390/antiox8050126>
75. Visser TJ (2000) Cellular Uptake of Thyroid Hormones. In: Feingold KR, Anawalt B, Boyce A, et al (eds) *Endotext*. MDText.com, Inc., South Dartmouth (MA)
76. Cayrol F, Sterle HA, Díaz Flaqué MC, Barreiro Arcos ML, Cremaschi GA (2019) Non-genomic Actions of Thyroid Hormones Regulate the Growth and Angiogenesis of T Cell Lymphomas. *Frontiers in Endocrinology* 10
77. Giammanco M, Di Liegro CM, Schiera G, Di Liegro I (2020) Genomic and Non-Genomic Mechanisms of Action of Thyroid Hormones and Their Catabolite 3,5-Diiodo-L-Thyronine in Mammals. *International Journal of Molecular Sciences* 21(11):4140. <https://doi.org/10.3390/ijms21114140>
78. Sinha R, Yen PM (2000) Cellular Action of Thyroid Hormone. In: Feingold KR, Anawalt B, Boyce A, et al (eds) *Endotext*. MDText.com, Inc., South Dartmouth (MA)
79. Liu Y-Y, Brent GA (2010) Thyroid Hormone Crosstalk with Nuclear Receptor Signaling in Metabolic Regulation. *Trends Endocrinol Metab* 21(3):166–173. <https://doi.org/10.1016/j.tem.2009.11.004>
80. Prezioso G, Giannini C, Chiarelli F (2018) Effect of Thyroid Hormones on Neurons and Neurodevelopment. *Horm Res Paediatr* 90(2):73–81. <https://doi.org/10.1159/000492129>
81. Dean DS, Gharib H (2008) Epidemiology of thyroid nodules. *Best Practice & Research Clinical Endocrinology & Metabolism* 22(6):901–911. <https://doi.org/10.1016/j.beem.2008.09.019>
82. Kamran SC, Marqusee E, Kim MI, et al (2013) Thyroid Nodule Size and Prediction of Cancer. *The Journal of Clinical Endocrinology & Metabolism* 98(2):564–570. <https://doi.org/10.1210/jc.2012-2968>
83. Cancer (IARC) TIA for R on Global Cancer Observatory. <https://gco.iarc.fr/>. Accessed 13 Oct 2022

84. Dimensiones del cáncer | AECC Observatorio. <https://observatorio.contraelcancer.es/explora/dimensiones-del-cancer>. Accessed 13 Oct 2022
85. Proyectos | Redecan. <https://redecan.org/es/proyectos>. Accessed 13 Oct 2022
86. Neamțiu L, Martos C, Giusti F, et al (2022) Impact of the first wave of the COVID-19 pandemic on cancer registration and cancer care: a European survey. *European Journal of Public Health* 32(2):311–315. <https://doi.org/10.1093/eurpub/ckab214>
87. Amador M, Matias-Guiu X, Sancho-Pardo G, et al (2021) Impact of the COVID-19 pandemic on the care of cancer patients in Spain. *ESMO Open* 6(3). <https://doi.org/10.1016/j.esmoop.2021.100157>
88. Sandi FJS, Arri EA, Ferrater JCG, Lluch AR, González EN (2019) Estudio epidemiológico de la incidencia de cáncer de tiroides en España: situación actual y evolución durante los últimos 20 años. *Endocrinología, Diabetes y Nutrición* 66(Extra 2):181–190
89. Cancer of the Thyroid - Cancer Stat Facts. In: SEER. <https://seer.cancer.gov/statfacts/html/thyro.html>. Accessed 13 Oct 2022
90. INEbase / Sociedad /Salud /Estadística de defunciones según la causa de muerte / Últimos datos. In: INE. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176780&menu=ultiDatos&idp=1254735573175. Accessed 13 Oct 2022
91. Kitahara CM, Sosa JA (2016) The changing incidence of thyroid cancer. *Nat Rev Endocrinol* 12(11):646–653. <https://doi.org/10.1038/nrendo.2016.110>
92. Suteau V, Munier M, Briet C, Rodien P (2021) Sex Bias in Differentiated Thyroid Cancer. *Int J Mol Sci* 22(23):12992. <https://doi.org/10.3390/ijms222312992>
93. Liu Y, Su L, Xiao H (2017) Review of Factors Related to the Thyroid Cancer Epidemic. *Int J Endocrinol* 2017:5308635. <https://doi.org/10.1155/2017/5308635>
94. Keane E, Francis EC, Catháin ÉÓ, Rowley H (2017) The role of race in thyroid cancer: systematic review. *J Laryngol Otol* 131(6):480–486. <https://doi.org/10.1017/S0022215117000688>
95. de Magalhães JP (2013) How ageing processes influence cancer. *Nat Rev Cancer* 13(5):357–365. <https://doi.org/10.1038/nrc3497>
96. Pellegriti G, Frasca F, Regalbuto C, Squatrito S, Vigneri R (2013) Worldwide Increasing Incidence of Thyroid Cancer: Update on Epidemiology and Risk Factors. *J Cancer Epidemiol* 2013:965212. <https://doi.org/10.1155/2013/965212>
97. Holt EH (2021) Current Evaluation of Thyroid Nodules. *Medical Clinics of North America* 105(6):1017–1031. <https://doi.org/10.1016/j.mcna.2021.06.006>
98. Lloyd RV, Osamura RY, Klöppel G, Rosai J (2017) WHO classification of tumours of endocrine organs. WHO–OMS. WHO Classification of Tumours of Endocrine Glands
99. Feldt-Rasmussen U (2001) Iodine and Cancer. *Thyroid* 11(5):483–486. <https://doi.org/10.1089/105072501300176435>
100. Morales-Sánchez P, Pérez RF, Santamarina P, Rodríguez-Rodero S, Fernandez-Fernandez A, Fraga MF (2019) Epigenetics: At the Crossroads Between Genetic and Environmental Determinants of Disease. In: Miszkiewicz JJ, Brennan-Olsen SL, Riancho JA (eds) Bone Health: A Reflection of the Social Mosaic. Springer Singapore, Singapore, pp 105–128
101. Alsen M, Sinclair C, Cooke P, Ziadkhanpour K, Genden E, van Gerwen M (2021) Endocrine Disrupting Chemicals and Thyroid Cancer: An Overview. *Toxics* 9(1):14. <https://doi.org/10.3390/toxics9010014>
102. Yang SP, Ngeow J (2016) Familial non-medullary thyroid cancer: unraveling the genetic maze. *Endocrine-Related Cancer* 23(12):R577–R595. <https://doi.org/10.1530/ERC-16-0067>
103. Alsanea O, Clark OH (2001) Familial thyroid cancer: Current Opinion in Oncology 13(1):44–51. <https://doi.org/10.1097/00001622-200101000-00009>
104. Kitahara CM, Slettebø Daltveit D, Ekbohm A, et al (2021) Maternal health, in-utero, and perinatal exposures and risk of thyroid cancer in offspring: a Nordic population-based nested case-control study. *The Lancet Diabetes & Endocrinology*

- 9(2):94–105. [https://doi.org/10.1016/S2213-8587\(20\)30399-5](https://doi.org/10.1016/S2213-8587(20)30399-5)
105. Mete O, Asa SL (2012) Pitfalls in the Diagnosis of Follicular Epithelial Proliferations of the Thyroid. *Advances in Anatomic Pathology* 19(6):363–373. <https://doi.org/10.1097/PAP.0b013e318271a5ac>
106. Salabè GB (2001) Pathogenesis of thyroid nodules: histological classification? *Biomedicine & Pharmacotherapy* 55(1):39–53. [https://doi.org/10.1016/S0753-3322\(00\)00010-X](https://doi.org/10.1016/S0753-3322(00)00010-X)
107. Yamada Y, Haga H, Yamada Y (2014) Concise Review: Dedifferentiation Meets Cancer Development: Proof of Concept for Epigenetic Cancer. *Stem Cells Transl Med* 3(10):1182–1187. <https://doi.org/10.5966/sctm.2014-0090>
108. Schmitt F (2017) *Cytopathology*. Springer International Publishing
109. Kakudo K, Li Y, Taniguchi E, Liu Z (2018) Follicular Neoplasms in the 4th Edition WHO Classification of Endocrine Organs. *Basic & Clinical Medicine* 7
110. Suster S (2019) Controversies Regarding the Interpretation of Follicular Thyroid Nodules. *Arch Pathol Lab Med* 143(12):1472–1476. <https://doi.org/10.5858/arpa.2019-0301-RA>
111. Staubitz JI, Musholt PB, Musholt TJ (2019) The surgical dilemma of primary surgery for follicular thyroid neoplasms. *Best Practice & Research Clinical Endocrinology & Metabolism* 33(4):101292. <https://doi.org/10.1016/j.beem.2019.101292>
112. Henke LE, Pfeifer JD, Baranski TJ, DeWees T, Grigsby PW (2018) Long-term outcomes of follicular variant vs classic papillary thyroid carcinoma. *Endocrine Connections* 7(12):1226–1235. <https://doi.org/10.1530/EC-18-0264>
113. Pizzimenti C, Fiorentino V, Ieni A, et al (2022) Aggressive variants of follicular cell-derived thyroid carcinoma: an overview. *Endocrine* 78(1):1–12. <https://doi.org/10.1007/s12020-022-03146-0>
114. Ganly I, Ricarte Filho J, Eng S, et al (2013) Genomic Dissection of Hurthle Cell Carcinoma Reveals a Unique Class of Thyroid Malignancy. *The Journal of Clinical Endocrinology & Metabolism* 98(5):E962–E972. <https://doi.org/10.1210/jc.2012-3539>
115. Zajkowska K, Kopczyński J, Gózdź S, Kowalska A (2020) Noninvasive follicular thyroid neoplasm with papillary-like nuclear features: a problematic entity. *Endocr Connect* 9(3):R47–R58. <https://doi.org/10.1530/EC-19-0566>
116. Landa I, Ibrahimasic T, Boucai L, et al (2016) Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. *J Clin Invest* 126(3):1052–1066. <https://doi.org/10.1172/JCI85271>
117. Kakudo K, Bychkov A, Bai Y, Li Y, Liu Z, Jung CK (2018) The new 4th edition World Health Organization classification for thyroid tumors, Asian perspectives. *Pathology International* 68(12):641–664. <https://doi.org/10.1111/pin.12737>
118. Lawrence MS, Stojanov P, Polak P, et al (2013) Mutational heterogeneity in cancer and the search for new cancer genes. *Nature* 499(7457):214–218. <https://doi.org/10.1038/nature12213>
119. Agrawal N, Akbani R, Aksoy BA, et al (2014) Integrated Genomic Characterization of Papillary Thyroid Carcinoma. *Cell* 159(3):676–690. <https://doi.org/10.1016/j.cell.2014.09.050>
120. Fang Y, Ma X, Zeng J, et al (2018) The Profile of Genetic Mutations in Papillary Thyroid Cancer Detected by Whole Exome Sequencing. *Cell Physiol Biochem* 50(1):169–178. <https://doi.org/10.1159/000493966>
121. Erinjeri NJ, Nicolson NG, Deyholos C, Korah R, Carling T (2018) Whole-Exome Sequencing Identifies Two Discrete Druggable Signaling Pathways in Follicular Thyroid Cancer. *J Am Coll Surg* 226(6):950–959.e5. <https://doi.org/10.1016/j.jamcollsurg.2018.01.059>
122. Baloch ZW, Asa SL, Barletta JA, et al (2022) Overview of the 2022 WHO Classification of Thyroid Neoplasms. *Endocr Pathol* 33(1):27–63. <https://doi.org/10.1007/s12022-022-09707-3>
123. Zaballos MA, Santisteban P (2017) Key signaling pathways in thyroid cancer. *Journal of Endocrinology* 235(2):R43–R61. <https://doi.org/10.1530/JOE-17-0266>

124. Doerfler WR, Nikitski AV, Morariu EM, et al (2021) Molecular alterations in Hürthle cell nodules and preoperative cancer risk. *Endocrine-Related Cancer* 28(5):301–309. <https://doi.org/10.1530/ERC-20-0435>
125. Yoo S-K, Lee S, Kim S-J, et al (2016) Comprehensive Analysis of the Transcriptional and Mutational Landscape of Follicular and Papillary Thyroid Cancers. *PLoS Genet* 12(8):e1006239. <https://doi.org/10.1371/journal.pgen.1006239>
126. Duan H, Liu X, Ren X, Zhang H, Wu H, Liang Z (2019) Mutation profiles of follicular thyroid tumors by targeted sequencing. *Diagn Pathol* 14. <https://doi.org/10.1186/s13000-019-0817-1>
127. Kebebew E, Weng J, Bauer J, et al (2007) The prevalence and prognostic value of BRAF mutation in thyroid cancer. *Ann Surg* 246(3):466–470; discussion 470–471. <https://doi.org/10.1097/SLA.0b013e318148563d>
128. Afkhami M, Karunamurthy A, Chiosea S, et al (2016) Histopathologic and Clinical Characterization of Thyroid Tumors Carrying the BRAF(K601E) Mutation. *Thyroid* 26(2):242–247. <https://doi.org/10.1089/thy.2015.0227>
129. Taneja C, Yip L, Morariu EM, et al (2022) Clinicopathologic Characteristics and Postsurgical Follow-Up of Noninvasive Follicular Thyroid Neoplasm with Papillary-Like Nuclear Features in the Postnomenclature Revision Era. *Thyroid*. <https://doi.org/10.1089/thy.2022.0337>
130. Karunamurthy A, Panebianco F, Hsiao SJ, et al (2016) Prevalence and phenotypic correlations of EIF1AX mutations in thyroid nodules. *Endocr Relat Cancer* 23(4):295–301. <https://doi.org/10.1530/ERC-16-0043>
131. Placzkowski KA, Reddi HV, Grebe SKG, Eberhardt NL, McIver B (2008) The Role of the PAX8/PPAR Fusion Oncogene in Thyroid Cancer. *PPAR Research* 2008:672829. <https://doi.org/10.1155/2008/672829>
132. Brandler TC, Liu CZ, Cho M, et al (2018) Does Noninvasive Follicular Thyroid Neoplasm With Papillary-Like Nuclear Features (NIFTP) Have a Unique Molecular Profile? *American Journal of Clinical Pathology* 150(5):451–460. <https://doi.org/10.1093/ajcp/aqy075>
133. Nikiforova MN, Biddinger PW, Caudill CM, Kroll TG, Nikiforov YE (2002) PAX8-PPAR γ Rearrangement in Thyroid Tumors: RT-PCR and Immunohistochemical Analyses. *The American Journal of Surgical Pathology* 26(8):1016–1023. <https://doi.org/10.1097/00000478-200208000-00006>
134. Marques AR, Espadinha C, Catarino AL, et al (2002) Expression of PAX8-PPAR gamma 1 rearrangements in both follicular thyroid carcinomas and adenomas. *J Clin Endocrinol Metab* 87(8):3947–3952. <https://doi.org/10.1210/jcem.87.8.8756>
135. Nikiforova MN, Lynch RA, Biddinger PW, et al (2003) RAS Point Mutations and PAX8-PPAR γ Rearrangement in Thyroid Tumors: Evidence for Distinct Molecular Pathways in Thyroid Follicular Carcinoma. *The Journal of Clinical Endocrinology & Metabolism* 88(5):2318–2326. <https://doi.org/10.1210/jc.2002-021907>
136. Bim LV, Carneiro TNR, Buzatto VC, et al (2021) Molecular Signature Expands the Landscape of Driver Negative Thyroid Cancers. *Cancers (Basel)* 13(20):5184. <https://doi.org/10.3390/cancers13205184>
137. Moraru A, Cakan-Akdogan G, Strassburger K, et al (2017) THADA Regulates the Organismal Balance between Energy Storage and Heat Production. *Developmental Cell* 41(1):72–81.e6. <https://doi.org/10.1016/j.devcel.2017.03.016>
138. Panebianco F, Kelly LM, Liu P, et al (2017) THADA fusion is a mechanism of IGF2BP3 activation and IGF1R signaling in thyroid cancer. *Proc Natl Acad Sci U S A* 114(9):2307–2312. <https://doi.org/10.1073/pnas.1614265114>
139. Morariu EM, McCoy KL, Chiosea SI, et al (2021) Clinicopathologic Characteristics of Thyroid Nodules Positive for the THADA-IGF2BP3 Fusion on Preoperative Molecular Analysis. *Thyroid* 31(8):1212–1218. <https://doi.org/10.1089/thy.2020.0589>
140. Ganly I, Makarov V, Deraje S, et al (2018) Integrated genomic analysis of Hürthle cell cancer reveals oncogenic drivers, recurrent mitochondrial mutations, and unique chromosomal landscapes. *Cancer Cell* 34(2):256–270.e5. <https://doi.org/10.1016/j.ccell.2018.07.002>

141. Gopal RK, Kübler K, Calvo SE, et al (2018) Widespread chromosomal losses and mitochondrial DNA alterations as genetic drivers in Hürthle cell Carcinoma. *Cancer Cell* 34(2):242-255.e5. <https://doi.org/10.1016/j.ccell.2018.06.013>
142. Puppin C, Passon N, Lavarone E, et al (2011) Levels of histone acetylation in thyroid tumors. *Biochemical and Biophysical Research Communications* 411(4):679–683. <https://doi.org/10.1016/j.bbrc.2011.06.182>
143. Hałasa M, Wawruszak A, Przybyszewska A, et al (2019) H3K18Ac as a Marker of Cancer Progression and Potential Target of Anti-Cancer Therapy. *Cells* 8(5):485. <https://doi.org/10.3390/cells8050485>
144. Tsai C-C, Chien M-N, Chang Y-C, Lee J-J, Dai S-H, Cheng S-P (2019) Overexpression of Histone H3 Lysine 27 Trimethylation Is Associated with Aggressiveness and Dedifferentiation of Thyroid Cancer. *Endocr Pathol* 30(4):305–311. <https://doi.org/10.1007/s12022-019-09586-1>
145. Iancu IV, Botezatu A, Plesa A, et al (2020) Alterations of regulatory factors and DNA methylation pattern in thyroid cancer. *Cancer Biomark*. <https://doi.org/10.3233/CBM-190871>
146. Tong M, Gao S, Qi W, et al (2019) 5-Hydroxymethylcytosine as a potential epigenetic biomarker in papillary thyroid carcinoma. *Oncol Lett* 18(3):2304–2309. <https://doi.org/10.3892/ol.2019.10531>
147. Yu S, Yin Y, Hong S, et al (2020) TET1 is a Tumor Suppressor That Inhibits Papillary Thyroid Carcinoma Cell Migration and Invasion. *Int J Endocrinol* 2020. <https://doi.org/10.1155/2020/3909610>
148. Seok JY, Astvatsaturyan K, Peralta-Venturina MD, Lai J, Fan X (2021) TROP-2, 5hmC, and IDH1 Expression in Anaplastic Thyroid Carcinoma. *Int J Surg Pathol* 29(4):368–377. <https://doi.org/10.1177/1066896920978597>
149. Saghafinia S, Mina M, Riggi N, Hanahan D, Ciriello G (2018) Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors. *Cell Rep* 25(4):1066-1080.e8. <https://doi.org/10.1016/j.celrep.2018.09.082>
150. Kikuchi Y, Tsuji E, Yagi K, et al (2013) Aberrantly methylated genes in human papillary thyroid cancer and their association with BRAF/RAS mutation. *Front Genet* 4. <https://doi.org/10.3389/fgene.2013.00271>
151. Yang X, Gao L, Zhang S (2017) Comparative pan-cancer DNA methylation analysis reveals cancer common and specific patterns. *Briefings in Bioinformatics* 18(5):761–773. <https://doi.org/10.1093/bib/bbw063>
152. Ellis RJ, Wang Y, Stevenson HS, et al (2014) Genome-Wide Methylation Patterns in Papillary Thyroid Cancer Are Distinct Based on Histological Subtype and Tumor Genotype. *J Clin Endocrinol Metab* 99(2):E329–E337. <https://doi.org/10.1210/jc.2013-2749>
153. Mancikova V, Buj R, Castelblanco E, et al (2014) DNA methylation profiling of well-differentiated thyroid cancer uncovers markers of recurrence free survival. *International Journal of Cancer* 135(3):598–610. <https://doi.org/10.1002/ijc.28703>
154. White MG, Nagar S, Aschebrook-Kilfoy B, et al (2016) Epigenetic Alterations and Canonical Pathway Disruption in Papillary Thyroid Cancer: A Genome-wide Methylation Analysis. *Ann Surg Oncol* 23(7):2302–2309. <https://doi.org/10.1245/s10434-016-5185-4>
155. Beltrami CM, dos Reis MB, Barros-Filho MC, et al (2017) Integrated data analysis reveals potential drivers and pathways disrupted by DNA methylation in papillary thyroid carcinomas. *Clinical Epigenetics* 9(1):45. <https://doi.org/10.1186/s13148-017-0346-2>
156. Bisarro Dos Reis M, Barros-Filho MC, Marchi FA, et al (2017) Prognostic Classifier Based on Genome-Wide DNA Methylation Profiling in Well-Differentiated Thyroid Tumors. *J Clin Endocrinol Metab* 102(11):4089–4099. <https://doi.org/10.1210/jc.2017-00881>
157. Zhang S, Wang Y, Chen M, et al (2017) CXCL12 methylation-mediated epigenetic regulation of gene expression in papillary thyroid carcinoma. *Sci Rep* 7(1):44033. <https://doi.org/10.1038/srep44033>
158. Wei S, Yun X, Ruan X, Wei X, Zheng X, Gao M (2019) Identification of potential pathogenic candidates or diagnostic biomarkers in papillary thyroid carcinoma using expression and methylation profiles. *Oncol Lett* 18(6):6670–6678. <https://doi.org/10.3892/ol.2019.11059>

159. Park J-L, Jeon S, Seo E-H, et al (2020) Comprehensive DNA Methylation Profiling Identifies Novel Diagnostic Biomarkers for Thyroid Cancer. *Thyroid*. <https://doi.org/10.1089/thy.2019.0011>
160. Rodríguez-Rodero S, Fernández AF, Fernández-Morera JL, et al (2013) DNA Methylation Signatures Identify Biologically Distinct Thyroid Cancer Subtypes. *J Clin Endocrinol Metab* 98(7):2811–2821. <https://doi.org/10.1210/jc.2012-3566>
161. Affinito O, Salerno P, D’Alessio AM, et al (2019) Association between DNA methylation profile and malignancy in follicular-patterned thyroid neoplasms. *Endocr Relat Cancer*. <https://doi.org/10.1530/ERC-18-0308>
162. Zhang H, Zhang Z, Liu X, et al (2021) DNA methylation haplotype block markers efficiently discriminate follicular thyroid carcinoma from follicular adenoma. *J Clin Endocrinol Metab*. <https://doi.org/10.1210/clinem/dgaa950>
163. Ravi N, Yang M, Mylona N, Wennerberg J, Paulsson K (2020) Global RNA Expression and DNA Methylation Patterns in Primary Anaplastic Thyroid Cancer. *Cancers (Basel)* 12(3). <https://doi.org/10.3390/cancers12030680>
164. Timp W, Bravo HC, McDonald OG, et al (2014) Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Medicine* 6(8):61. <https://doi.org/10.1186/s13073-014-0061-y>
165. Barros-Filho MC, dos Reis MB, Beltrami CM, et al (2019) DNA methylation-based method to differentiate malignant from benign thyroid lesions. *Thyroid*. <https://doi.org/10.1089/thy.2018.0458>
166. Jia X, Zhao Q, Zhang Y, et al (2019) Identification of a Five-CpG Signature with Diagnostic Value in Thyroid Cancer. *J Comput Biol*. <https://doi.org/10.1089/cmb.2019.0165>
167. Lv L, Cao L, Hu G, Shen Q, Wu J (2020) Methylation-Driven Genes Identified as Novel Prognostic Indicators for Thyroid Carcinoma. *Front Genet* 11:294. <https://doi.org/10.3389/fgene.2020.00294>
168. Yao Y, Xu P, Ying T, et al (2022) Integrative Analysis of DNA Methylation and Gene Expression Identified Follicular Thyroid Cancer-Specific Diagnostic Biomarkers. *Front Endocrinol (Lausanne)* 12:736068. <https://doi.org/10.3389/fendo.2021.736068>
169. Yim JH, Choi AH, Li AX, et al (2019) Identification of Tissue-Specific DNA Methylation Signatures for Thyroid Nodule Diagnostics. *Clin Cancer Res* 25(2):544–551. <https://doi.org/10.1158/1078-0432.CCR-18-0841>
170. Buley ID, Roskell DE (2000) Fine-Needle Aspiration Cytology in Tumour Diagnosis: Uses and Limitations. *Clinical Oncology* 12(3):166–171. <https://doi.org/10.1053/clon.2000.9144>
171. Diamantis A, Magiorkinis E, Koutselini H (2009) Fine-needle aspiration (FNA) biopsy: historical aspects. *Folia Histochem Cytobiol* 47(2):191–197. <https://doi.org/10.2478/v10042-009-0027-x>
172. Long SR, Cohen MB (1996) Classics in cytology VII: Kün, Lebert, and early efforts at fine-needle aspiration biopsy. *Diagnostic Cytopathology* 14(2):182–183. [https://doi.org/10.1002/\(SICI\)1097-0339\(199603\)14:2<182::AID-DC15>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-0339(199603)14:2<182::AID-DC15>3.0.CO;2-I)
173. Martin HE, Ellis EB (1930) BIOPSY BY NEEDLE PUNCTURE AND ASPIRATION. *Ann Surg* 92(2):169–181. <https://doi.org/10.1097/00000658-193008000-00002>
174. Einhorn J, Franzén S (1962) Thin-Needle Biopsy in the Diagnosis of Thyroid Disease. *Acta Radiologica* 58(5):321–336. <https://doi.org/10.1177/028418516205800501>
175. Bäckdahl M, Wallin G, Löwhagen T, Auer G, Granberg PO (1987) Fine-needle biopsy cytology and DNA analysis. Their place in the evaluation and treatment of patients with thyroid neoplasms. *Surg Clin North Am* 67(2):197–211. [https://doi.org/10.1016/s0039-6109\(16\)44179-4](https://doi.org/10.1016/s0039-6109(16)44179-4)
176. Ali SZ, Cibas ES (2010) The Bethesda System for Reporting Thyroid Cytopathology. Springer US, Boston, MA
177. Cibas ES, Ali SZ (2017) The 2017 Bethesda System for Reporting Thyroid Cytopathology. *J Am Soc Cytopathol* 6(6):217–222. <https://doi.org/10.1016/j.jasc.2017.09.002>
178. Hoda RS, Rao R, Scognamiglio T (2020) Atlas of Thyroid Cytopathology on Liquid-Based Preparations: Correlation with Clinical,

- Radiological, Molecular Tests and Histopathology. Springer International Publishing, Cham
179. Bode-Lesniewska B, Cochand-Priollet B, Straccia P, Fadda G, Bongiovanni M (2019) Management of thyroid cytological material, preanalytical procedures and bio-banking. *Cytopathology* 30(1):7–16. <https://doi.org/10.1111/cyt.12586>
180. Pusztaszeri M (2019) Introduction to the Second Edition of the Bethesda System for Reporting Thyroid Cytopathology. In: Kakudo K (ed) *Thyroid FNA Cytology: Differential Diagnoses and Pitfalls*. Springer, Singapore, pp 59–68
181. Malheiros DC, Canberk S, Poller DN, Schmitt F (2018) Thyroid FNAC: Causes of false-positive results. *Cytopathology* 29(5):407–417. <https://doi.org/10.1111/cyt.12575>
182. Cantara S, Marzocchi C, Pilli T, et al (2017) Molecular Signature of Indeterminate Thyroid Lesions: Current Methods to Improve Fine Needle Aspiration Cytology (FNAC) Diagnosis. *Int J Mol Sci* 18(4). <https://doi.org/10.3390/ijms18040775>
183. Orell SR (2003) Pitfalls in fine needle aspiration cytology. *Cytopathology* 14(4):173–182. <https://doi.org/10.1046/j.1365-2303.2003.00078.x>
184. Turk AT, Wenig BM (2019) Thyroid tumors with follicular architecture. *Annals of Diagnostic Pathology* 38:51–58. <https://doi.org/10.1016/j.anndiagpath.2018.10.007>
185. Rossi ED, Adeniran AJ, Faquin WC (2019) Pitfalls in Thyroid Cytopathology. *Surgical Pathology Clinics* 12(4):865–881. <https://doi.org/10.1016/j.path.2019.08.001>
186. Schmitt F (2019) False Positive in Thyroid FNA: Causes and How to Avoid Them. In: Kakudo K (ed) *Thyroid FNA Cytology*. Springer Singapore, Singapore, pp 27–36
187. Gómez-Pérez AM, Fernández-García JC, Iglesias P, et al (2020) Diagnosis and treatment of thyroid nodules in Spain. Results of a national survey. *Endocrinol Diabetes Nutr (Engl Ed)* 67(7):438–445. <https://doi.org/10.1016/j.endinu.2019.10.003>
188. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW (2012) The Bethesda System for Reporting Thyroid Cytopathology: A Meta-Analysis. *ACY* 56(4):333–339. <https://doi.org/10.1159/000339959>
189. Haugen BR, Alexander EK, Bible KC, et al (2016) 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* 26(1):1–133. <https://doi.org/10.1089/thy.2015.0020>
190. Vuong HG, Ngo HTT, Bychkov A, et al (2020) Differences in surgical resection rate and risk of malignancy in thyroid cytopathology practice between Western and Asian countries: A systematic review and meta-analysis. *Cancer Cytopathol* 128(4):238–249. <https://doi.org/10.1002/cncy.22228>
191. Sitges-Serra A (2021) Etiology and Diagnosis of Permanent Hypoparathyroidism after Total Thyroidectomy. *J Clin Med* 10(3):543. <https://doi.org/10.3390/jcm10030543>
192. Enomoto K, Uchino S, Watanabe S, Enomoto Y, Noguchi S (2014) Recurrent laryngeal nerve palsy during surgery for benign thyroid diseases: risk factors and outcome analysis. *Surgery* 155(3):522–528. <https://doi.org/10.1016/j.surg.2013.11.005>
193. Sendín R, José J (2010) Definición de “Profesión médica”, “Profesional médico/a” y “Profesionalismo médico.” *Educación Médica* 13(2):63–66
194. Ashley EA (2016) Towards precision medicine. *Nat Rev Genet* 17(9):507–522. <https://doi.org/10.1038/nrg.2016.86>
195. Strimbu K, Tavel JA (2010) What are Biomarkers? *Curr Opin HIV AIDS* 5(6):463–466. <https://doi.org/10.1097/COH.0b013e32833ed177>
196. Sarhadi VK, Armengol G (2022) Molecular Biomarkers in Cancer. *Biomolecules* 12(8):1021. <https://doi.org/10.3390/biom12081021>
197. Bensalah K, Montorsi F, Shariat SF (2007) Challenges of Cancer Biomarker Profiling. *European Urology* 52(6):1601–1609. <https://doi.org/10.1016/j.eururo.2007.09.036>

198. Vargas-Salas S, Martínez JR, Urra S, et al (2017) Genetic testing for indeterminate thyroid cytology: review and meta-analysis. *Endocr Relat Cancer* 25(3):R163–R177. <https://doi.org/10.1530/ERC-17-0405>
199. Ray P, Manach YL, Riou B, Houle TT, Warner DS (2010) Statistical Evaluation of a Biomarker. *Anesthesiology* 112(4):1023–1040. <https://doi.org/10.1097/ALN.0b013e3181d47604>
200. Agarwal S, Bychkov A, Jung C-K (2021) Emerging Biomarkers in Thyroid Practice and Research. *Cancers (Basel)* 14(1):204. <https://doi.org/10.3390/cancers14010204>
201. Ferraz C, Eszlinger M, Paschke R (2011) Current State and Future Perspective of Molecular Diagnosis of Fine-Needle Aspiration Biopsy of Thyroid Nodules. *The Journal of Clinical Endocrinology & Metabolism* 96(7):2016–2026. <https://doi.org/10.1210/jc.2010-2567>
202. Nikiforov YE, Steward DL, Robinson-Smith TM, et al (2009) Molecular Testing for Mutations in Improving the Fine-Needle Aspiration Diagnosis of Thyroid Nodules. *J Clin Endocrinol Metab* 94(6):2092–2098. <https://doi.org/10.1210/jc.2009-0247>
203. Nikiforova MN, Mercurio S, Wald AI, et al (2018) Analytical performance of ThyroSeq v3 Genomic Classifier for cancer diagnosis in thyroid nodules. *Cancer* 124(8):1682–1690. <https://doi.org/10.1002/cncr.31245>
204. Nikiforova MN, Lepe M, Tolino LA, et al (2021) Thyroid cytology smear slides: An untapped resource for ThyroSeq testing. *Cancer Cytopathology* 129(1):33–42. <https://doi.org/10.1002/cncy.22331>
205. Steward DL, Carty SE, Sippel RS, et al (2019) Performance of a Multigene Genomic Classifier in Thyroid Nodules With Indeterminate Cytology: A Prospective Blinded Multicenter Study. *JAMA Oncol* 5(2):204–212. <https://doi.org/10.1001/jamaoncol.2018.4616>
206. Alexander EK, Kennedy GC, Baloch ZW, et al (2012) Preoperative Diagnosis of Benign Thyroid Nodules with Indeterminate Cytology. In: <http://dx.doi.org/10.1056/NEJMoa1203208>. https://www.nejm.org/doi/10.1056/NEJMoa1203208?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat
207. Patel KN, Angell TE, Babiarz J, et al (2018) Performance of a Genomic Sequencing Classifier for the Preoperative Diagnosis of Cytologically Indeterminate Thyroid Nodules. *JAMA Surg* 153(9):817–824. <https://doi.org/10.1001/jamasurg.2018.1153>
208. Hu MI, Waguespack SG, Dosiou C, et al (2021) Afirma Genomic Sequencing Classifier and Xpression Atlas Molecular Findings in Consecutive Bethesda III-VI Thyroid Nodules. *J Clin Endocrinol Metab* 106(8):2198–2207. <https://doi.org/10.1210/clinem/dgab304>
209. Labourier E, Shifrin A, Busseniers AE, et al (2015) Molecular Testing for miRNA, mRNA, and DNA on Fine-Needle Aspiration Improves the Preoperative Diagnosis of Thyroid Nodules With Indeterminate Cytology. *J Clin Endocrinol Metab* 100(7):2743–2750. <https://doi.org/10.1210/jc.2015-1158>
210. Finkelstein SD, Sistrunk JW, Malchoff C, et al (2022) A Retrospective Evaluation of the Diagnostic Performance of an Interdependent Pairwise MicroRNA Expression Analysis with a Mutation Panel in Indeterminate Thyroid Nodules. *Thyroid*. <https://doi.org/10.1089/thy.2022.0124>
211. Performance. In: ThyGeNEXT® + ThyraMIR®. <https://thygenext-thyramir.com/performance/>. Accessed 24 Oct 2022
212. Davalos V, Esteller M Cancer epigenetics in clinical practice. CA: A Cancer Journal for Clinicians n/a(n/a). <https://doi.org/10.3322/caac.21765>
213. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al (2014) Multitarget Stool DNA Testing for Colorectal-Cancer Screening. *New England Journal of Medicine* 370(14):1287–1297. <https://doi.org/10.1056/NEJMoa1311194>
214. Ahlquist DA (2015) Multi-Target Stool DNA Test: A New High Bar for Noninvasive Screening. *Dig Dis Sci* 60(3):623–633. <https://doi.org/10.1007/s10620-014-3451-5>
215. Lamb YN, Dhillon S (2017) Epi proColon® 2.0 CE: A Blood-Based Screening Test for Colorectal Cancer. *Mol Diagn Ther* 21(2):225–232. <https://doi.org/10.1007/s40291-017-0259-y>

REFERENCIAS

216. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV (2020) Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol* 31(6):745–759. <https://doi.org/10.1016/j.annonc.2020.02.011>
217. Chen X, Gole J, Gore A, et al (2020) Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. *Nat Commun* 11:3475. <https://doi.org/10.1038/s41467-020-17316-z>
218. Klein EA, Richards D, Cohn A, et al (2021) Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Annals of Oncology* 32(9):1167–1177. <https://doi.org/10.1016/j.annonc.2021.05.806>
219. Zhang B, Liu S, Zhang Z, et al (2014) Analysis of BRAFV600E mutation and DNA methylation improves the diagnostics of thyroid fine needle aspiration biopsies. *Diagn Pathol* 9:45. <https://doi.org/10.1186/1746-1596-9-45>
220. Kartal K, Onder S, Kosemehmetoglu K, Kilickap S, Tezel YG, Kaynaroglu V (2015) Methylation status of TSHr in well-differentiated thyroid cancer by using cytologic material. *BMC Cancer* 15(1):824. <https://doi.org/10.1186/s12885-015-1861-1>
221. Koziołek M, Bińczak-Kuleta A, Stepaniuk M, et al (2015) Frequency assessment of BRAF mutation, KRas mutation, and RASSF1A methylation in nodular goitre based on fine-needle aspiration cytology specimens. *Endokrynologia Polska* 66(5):384–393. <https://doi.org/10.5603/EP.2015.0048>
222. Chang H, Shin BK, Kim A, Kim HK, Kim BH (2016) DNA methylation analysis for the diagnosis of thyroid nodules – a pilot study with reference to BRAFV600E mutation and cytopathology results. *Cytopathology* 27(2):122–130. <https://doi.org/10.1111/cyt.12248>
223. A framework for variation discovery and genotyping using next-generation DNA sequencing data | *Nature Genetics*. <https://www.nature.com/articles/ng.806>. Accessed 18 Aug 2022
224. Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34(17):i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
225. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
226. Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
227. biotools:picard_tools (2019) Picard Toolkit
228. Cibulskis K, Lawrence MS, Carter SL, et al (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 31(3):213–219. <https://doi.org/10.1038/nbt.2514>
229. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27(21):2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
230. Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
231. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38(16):e164. <https://doi.org/10.1093/nar/gkq603>
232. Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* Chapter 7:Unit7.20. <https://doi.org/10.1002/0471142905.hg0720s76>
233. Schwarz JM, Rödelberger C, Schuelke M, Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 7(8):575–576. <https://doi.org/10.1038/nmeth0810-575>
234. Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31(13):3812–3814. <https://doi.org/10.1093/nar/gkg509>
235. Tate JG, Bamford S, Jubb HC, et al (2019) COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* 47(D1):D941–D947. <https://doi.org/10.1093/nar/gky1015>

236. Moran S, Arribas C, Esteller M (2016) Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* 8(3):389–399. <https://doi.org/10.2217/epi.15.114>
237. ENCODE Project Consortium, Moore JE, Purcaro MJ, et al (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 583(7818):699–710. <https://doi.org/10.1038/s41586-020-2493-4>
238. Tejedor JR, Bueno C, Cobo I, et al (2018) Epigenome-wide analysis reveals specific DNA hypermethylation of T cells during human hematopoietic differentiation. *Epigenomics* 10(7):903–923. <https://doi.org/10.2217/epi-2017-0163>
239. Fortin J-P, Triche TJ, Hansen KD (2017) Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* 33(4):558–560. <https://doi.org/10.1093/bioinformatics/btw691>
240. A data-driven approach to preprocessing Illumina 450K methylation array data | BMC Genomics | Full Text. <https://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-293>. Accessed 19 Aug 2022
241. Teschendorff AE, Marabita F, Lechner M, et al (2013) A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* 29(2):189–196. <https://doi.org/10.1093/bioinformatics/bts680>
242. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28(6):882–883. <https://doi.org/10.1093/bioinformatics/bts034>
243. Ritchie ME, Phipson B, Wu D, et al (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43(7):e47. <https://doi.org/10.1093/nar/gkv007>
244. Acharjee A, Larkman J, Xu Y, Cardoso VR, Gkoutos GV (2020) A random forest based biomarker discovery and power analysis framework for diagnostics research. *BMC Medical Genomics* 13(1):178. <https://doi.org/10.1186/s12920-020-00826-6>
245. Frommer M, McDonald LE, Millar DS, et al (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences* 89(5):1827–1831
246. Tost J, Gut IG (2007) DNA methylation analysis by pyrosequencing. *Nature Protocols* 2(9):2265–2275. <https://doi.org/10.1038/nprot.2007.314>
247. Mattu R, Sorbara L, Filie AC, et al (2004) Utilization of polymerase chain reaction on archival cytologic material: a comparison with fresh material with special emphasis on cerebrospinal fluids. *Mod Pathol* 17(10):1295–1301. <https://doi.org/10.1038/modpathol.3800185>
248. Haugen BR, Alexander EK, Bible KC, et al (2015) 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* 26(1):1–133. <https://doi.org/10.1089/thy.2015.0020>
249. Cooper DS, Doherty GM, Haugen BR, et al (2009) Revised American Thyroid Association Management Guidelines for Patients with Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* 19(11):1167–1214. <https://doi.org/10.1089/thy.2009.0110>
250. Lawrence MS, Stojanov P, Polak P, et al (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499(7457):214–218. <https://doi.org/10.1038/nature12213>
251. Jung S-H, Kim MS, Jung CK, et al (2016) Mutational burdens and evolutionary ages of thyroid follicular adenoma are comparable to those of follicular carcinoma. *Oncotarget* 7(43):69638–69648. <https://doi.org/10.18632/oncotarget.11922>
252. Qi T, Rong X, Feng Q, et al (2021) Somatic Mutation Profiling of Papillary Thyroid Carcinomas by Whole-exome Sequencing and Its Relationship with Clinical Characteristics. *Int J Med Sci* 18(12):2532–2544. <https://doi.org/10.7150/ijms.50916>

253. Fukahori M, Yoshida A, Hayashi H, et al (2012) The associations between RAS mutations and clinical characteristics in follicular thyroid tumors: new insights from a single center and a large patient cohort. *Thyroid* 22(7):683–689. <https://doi.org/10.1089/thy.2011.0261>
254. Muzza M, Colombo C, Pogliaghi G, Karapanou O, Fugazzola L (2020) Molecular markers for the classification of cytologically indeterminate thyroid nodules. *J Endocrinol Invest* 43(6):703–716. <https://doi.org/10.1007/s40618-019-01164-w>
255. Nikiforov YE, Otori NP, Hodak SP, et al (2011) Impact of Mutational Testing on the Diagnosis and Management of Patients with Cytologically Indeterminate Thyroid Nodules: A Prospective Analysis of 1056 FNA Samples. *The Journal of Clinical Endocrinology & Metabolism* 96(11):3390–3397. <https://doi.org/10.1210/jc.2011-1469>
256. Marotta V, Bifulco M, Vitale M (2021) Significance of RAS Mutations in Thyroid Benign Nodules and Non-Medullary Thyroid Cancer. *Cancers (Basel)* 13(15):3785. <https://doi.org/10.3390/cancers13153785>
257. Song YS, Park YJ (2019) Genomic Characterization of Differentiated Thyroid Carcinoma. *Endocrinol Metab (Seoul)* 34(1):1–10. <https://doi.org/10.3803/EnM.2019.34.1.1>
258. Xing M (2010) Genetic Alterations in the Phosphatidylinositol-3 Kinase/Akt Pathway in Thyroid Cancer. *Thyroid* 20(7):697–706. <https://doi.org/10.1089/thy.2010.1646>
259. Kuhner MK, Kostadinov R, Reid BJ (2016) Limitations of the Driver/Passenger Model in Cancer Prevention. *Cancer Prevention Research* 9(5):335–338. <https://doi.org/10.1158/1940-6207.CAPR-15-0343>
260. Gomes CC (2022) Recurrent driver mutations in benign tumors. *Mutation Research/Reviews in Mutation Research* 789:108412. <https://doi.org/10.1016/j.mrrev.2022.108412>
261. Kumar S, Warrell J, Li S, et al (2020) Passenger Mutations in More Than 2,500 Cancer Genomes: Overall Molecular Functional Impact and Consequences. *Cell* 180(5):915–927.e16. <https://doi.org/10.1016/j.cell.2020.01.032>
262. Esserman LJ, Thompson IM, Reid B, et al (2014) Addressing overdiagnosis and overtreatment in cancer: a prescription for change. *Lancet Oncol* 15(6):e234–e242. [https://doi.org/10.1016/S1470-2045\(13\)70598-9](https://doi.org/10.1016/S1470-2045(13)70598-9)
263. Nicholson KJ, Roberts MS, McCoy KL, Carty SE, Yip L (2019) Molecular Testing Versus Diagnostic Lobectomy in Bethesda III/IV Thyroid Nodules: A Cost-Effectiveness Analysis. *Thyroid* 29(9):1237–1243. <https://doi.org/10.1089/thy.2018.0779>
264. Carmona JJ, Accomando WP, Binder AM, et al (2017) Empirical comparison of reduced representation bisulfite sequencing and Infinium BeadChip reproducibility and coverage of DNA methylation in humans. *npj Genomic Med* 2(1):1–10. <https://doi.org/10.1038/s41525-017-0012-9>
265. Sahli ZT, Smith PW, Umbricht CB, Zeiger MA (2018) Preoperative Molecular Markers in Thyroid Nodules. *Frontiers in Endocrinology* 9
266. Hao Y, Choi Y, Babiarz JE, et al (2019) Analytical Verification Performance of Afirma Genomic Sequencing Classifier in the Diagnosis of Cytologically Indeterminate Thyroid Nodules. *Front Endocrinol* 10. <https://doi.org/10.3389/fendo.2019.00438>
267. Saah AJ, Hoover DR (1998) [Sensitivity and specificity revisited: significance of the terms in analytic and diagnostic language]. *Ann Dermatol Venereol* 125(4):291–294
268. Dom G, Frank S, Floor S, et al (2017) Thyroid follicular adenomas and carcinomas: molecular profiling provides evidence for a continuous evolution. *Oncotarget* 9(12):10343–10359. <https://doi.org/10.18632/oncotarget.23130>
269. Melmed S, Koenig R, Rosen CJ, Auchus R, Goldfine AB (2019) *Williams Textbook of Endocrinology*. Elsevier
270. Arora N, Scognamiglio T, Zhu B, Fahey TJ (2008) Do benign thyroid nodules have malignant potential? An evidence-based review. *World J Surg* 32(7):1237–1246. <https://doi.org/10.1007/s00268-008-9484-1>
271. Wojtas B, Pfeifer A, Oczko-Wojciechowska M, et al (2017) Gene Expression (mRNA) Markers for Differentiating between Malignant and Benign

- Follicular Thyroid Tumours. *Int J Mol Sci* 18(6). <https://doi.org/10.3390/ijms18061184>
272. Nikiforov YE, Seethala RR, Tallini G, et al (2016) Nomenclature Revision for Encapsulated Follicular Variant of Papillary Thyroid Carcinoma: A Paradigm Shift to Reduce Overtreatment of Indolent Tumors. *JAMA Oncol* 2(8):1023–1029. <https://doi.org/10.1001/jamaoncol.2016.0386>
273. Lloyd RV, Asa SL, LiVolsi VA, et al (2018) The evolving diagnosis of noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP). *Human Pathology* 74:1–4. <https://doi.org/10.1016/j.humpath.2017.12.027>
274. Parente DN, Kluijfhout WP, Bongers PJ, et al (2018) Clinical Safety of Renaming Encapsulated Follicular Variant of Papillary Thyroid Carcinoma: Is NIFTP Truly Benign? *World J Surg* 42(2):321–326. <https://doi.org/10.1007/s00268-017-4182-5>
275. Eskander A, Hall SF, Manduch M, Griffiths R, Irish JC (2019) A Population-Based Study on NIFTP Incidence and Survival: Is NIFTP Really a “Benign” Disease? *Ann Surg Oncol* 26(5):1376–1384. <https://doi.org/10.1245/s10434-019-07187-0>
276. Jug RC, Datto MB, Jiang XS (2018) Molecular testing for indeterminate thyroid nodules: Performance of the Afirma gene expression classifier and ThyroSeq panel. *Cancer Cytopathol* 126(7):471–480. <https://doi.org/10.1002/cncy.21993>
277. Silaghi CA, Lozovanu V, Georgescu CE, et al (2021) Thyroseq v3, Afirma GSC, and microRNA Panels Versus Previous Molecular Tests in the Preoperative Diagnosis of Indeterminate Thyroid Nodules: A Systematic Review and Meta-Analysis. *Frontiers in Endocrinology* 12
278. Giannini R, Ugolini C, Poma AM, et al (2017) Identification of Two Distinct Molecular Subtypes of Non-Invasive Follicular Neoplasm with Papillary-Like Nuclear Features by Digital RNA Counting. *Thyroid* 27(10):1267–1276. <https://doi.org/10.1089/thy.2016.0605>
279. Marino-Enriquez A, Fletcher CDM (2014) Shouldn't we care about the biology of benign tumours? *Nat Rev Cancer* 14(11):701–702. <https://doi.org/10.1038/nrc3845>
280. D Codes | SEER Training. <https://training.seer.cancer.gov/icd10cm/neoplasm/d-codes.html#uncertain>. Accessed 14 Nov 2022
281. Kakudo K (2018) How to handle borderline/precursor thyroid tumors in management of patients with thyroid nodules. *Gland Surgery* 7(Suppl 1):S8. <https://doi.org/10.21037/gs.2017.08.02>
282. Bai Y, Kakudo K, Jung CK (2020) Updates in the Pathologic Classification of Thyroid Neoplasms: A Review of the World Health Organization Classification. *Endocrinol Metab (Seoul)* 35(4):696–715. <https://doi.org/10.3803/EnM.2020.807>

ANEXO 1: TABLAS SUPLEMENTARIAS

- **Tabla suplementaria 1**

Tabla S1| Clasificación de tumores tiroideos de la Organización Mundial de la Salud (OMS) de 2017. CIE-10: Clasificación internacional de enfermedades, 10.ª edición; D34: Neoplasia benigna de tiroides; D44.0: Neoplasia de comportamiento incierto de glándula tiroides; C73: Neoplasia maligna de la glándula tiroides. CIE-O: Clasificación Internacional de Enfermedades para Oncología, los códigos de comportamiento /0, /1, /2 y /3 significan carcinoma in situ benigno, limítrofe o incierto, neoplasia intraepitelial de grado III y tumores malignos, respectivamente. Tabla modificada de Bai y colaboradores (2020) [282]

Tipos histológicos y variantes	CIE-10	CIE-O
Originados a partir de células foliculares		
Adenoma folicular (FTA)	D34	8330/0
Variante hiperfuncionante: adenoma tóxico o caliente. Variantes morfológicas: con hiperplasia papilar, con núcleos bizarros, células en anillo de sello, células claras, células fusiformes, FTA negro y lipoadenoma	D34	
Tumor trabecular hialinizante	D44.0	8336/1
Otros tumores tiroideos encapsulados con patrón folicular		
Tumor folicular de potencial de malignidad incierto (FT-UMP)	D44.0	8335/1
Tumor bien diferenciado de potencial de malignidad incierto (WT-UMP)	D44.0	8348/1
Neoplasia tiroidea folicular no invasiva con características nucleares de tipo papilar (NIFTP)	D44.0	8349/1
Carcinoma papilar de tiroides (PTC)	C73	8260/3
Variante folicular del PTC (FVPTC)	C73	8340/3
Variante encapsulada del PTC	C73	8343/3
Microcarcinoma del PTC	C73	8341/3
Variante columnar del PTC	C73	8344/3
Variante oncocítica del PTC	C73	8342/3
Otras variantes: esclerosante difuso, células altas, cribiforme-morular, PTC con estroma similar a fibromatosis/fascitis, sólido-trabecular, células fusiformes, células claras y variantes similares a Warthin	C73	
Carcinoma folicular de tiroides (FTC)	C73	8330/3
FTC mínimamente invasivo	C73	8335/3
FTC encapsulado angioinvasivo	C73	8339/3
FTC ampliamente invasivo	C73	8330/3
Variantes: de células claras, tipo de células en anillo de sello y FTC con patrón glomeruloide	C73	
Tumores de Hürthle u oncocíticos		
Adenoma de células de Hürthle (HCA)	D34	8290/0
Carcinoma de células de Hürthle (HCC): mínimamente invasivo, encapsulado angioinvasivo, ampliamente invasivo	C73	8290/3
Carcinoma folicular pobremente diferenciado (PDTC)	C73	8337/3
Carcinoma oncocítico pobremente diferenciado	C73	8337/3
Carcinoma anaplásico de tiroides (ATC)	C73	8020/3
Variantes: sarcomatoide, de células gigantes, epitelial, paucicelular, tipo linfoepitelioma y de células pequeñas	C73	
Carcinoma de células escamosas	C73	8070/3
Carcinoma mucoepidermoide	C73	8430/3
Carcinoma mucinoso	C73	8480/3
Timoma ectópico	C73	8580/3
Tumor epitelial fusiforme con diferenciación tipo timo	C73	8588/3
Carcinoma tímico intratiroideo	C73	8589/3
Subtipos: carcinoma de células escamosas, linfoepitelioma, carcinoma neuroendocrino		
Originados a partir de las células parafoliculares		
Carcinoma medular de tiroides (MTC)	C73	8345/3
Variantes: papilar, folicular (tubular/glandular), de células fusiformes, de células gigantes, de células claras, oncocítica, melanótica, escamosa, anficrina, tipo paraganglioma, tipo angiosarcoma y variantes de células pequeñas. Formas mixtas de MTC, PTC, FTC, PDTC y ATC		
Hiperplasia primaria de células C		
Microcarcinoma medular	C73	8345/3

ANEXO 1: TABLAS SUPLEMENTARIAS

- *Tabla suplementaria 2*

Tabla S2 | Resumen de los resultados del experimento de secuenciación del exoma completo correspondientes a las lesiones tiroideas humanas analizadas en la cohorte de descubrimiento.

Muestra	Cobertura media	Tamaño de librería estimado	Número total de lecturas	Porcentaje de mapeo
BIOP-01	261,07	629.521.638	278.223.942	99,9594452
BIOP-02	231,6	537.137.496	270.041.014	99,9662117
BIOP-03	251,72	674.214.172	254.095.518	99,9710775
BIOP-04	129,81	346.892.154	144.046.370	99,9562749
BIOP-05	255,54	590.144.794	276.791.376	99,9686772
BIOP-06	266,1	650.637.531	283.787.808	99,9704194
BIOP-07	263,89	678.602.857	270.049.111	99,9675501
BIOP-08	125,77	377.971.236	129.838.220	99,9571897
BIOP-11	263,1	615.493.531	284.480.197	99,9642346
BIOP-12	265,78	658.241.202	279.500.974	99,964849
BIOP-14	137,57	369.793.764	138.850.443	99,9600773
BIOP-15	249,79	593.586.409	258.476.221	99,9670763
BIOP-16	269,83	621.630.612	272.543.746	99,9693523
BIOP-17	145,01	390.721.401	152.294.320	99,9665799
BIOP-18	140,75	402.751.975	145.131.582	99,9497585
BIOP-24	129,18	410.884.762	133.548.698	99,9698688
BIOP-26	90,04	472.948.332	134.319.356	99,9627043
BIOP-29	119,7	358.311.107	126.049.771	99,9594547
BIOP-30	135,11	438.236.700	135.711.316	99,9661712
BIOP-31	106,75	515.225.568	141.562.207	99,9596432
BIOP-32	277,79	627.127.797	278.360.765	99,9766731
BIOP-33	143,23	299.321.623	143.263.563	99,968759
BIOP-34	135,2	264.041.150	156.430.325	99,9696149
BIOP-35	113,03	221.800.182	153.691.535	99,9681762
BIOP-36	126,31	272.945.476	145.453.473	99,9706171
BIOP-37	90,85	253.433.412	144.422.256	99,9692109
BIOP-38	142,48	349.963.288	139.689.143	99,9711478
BIOP-39	101,7	174.646.826	169.305.342	99,9725884
BIOP-40	124,63	381.778.233	129.776.070	99,9696839
BIOP-41	148,87	299.148.106	159.369.797	99,9691223
BIOP-51	127,52	278.199.998	158.350.644	99,9678602
BIOP-52	128,61	304.973.440	147.468.053	99,9683767
BIOP-53	132,38	389.774.495	140.607.104	99,9718671
BIOP-55	127,62	399.234.280	130.871.972	99,9709762
BIOP-56	129,51	415.617.340	129.133.121	99,9711861
BIOP-57	133,71	427.665.808	137.497.376	99,9652994
BIOP-63	127,95	350.295.369	132.076.632	99,9669037
BIOP-70	196,21	397.445.743	199.611.388	99,975548
BIOP-71	75,67	210.670.855	148.131.705	99,969027
BIOP-90	134,98	410.799.357	138.203.730	99,9687682
BIOP-91	130,38	380.293.821	137.634.668	99,967017
BIOP-93	80,45	331.349.667	157.551.567	99,9658977
BIOP-95	68,61	227.311.471	156.266.209	99,9699507
BIOP-96	169,7	394.746.707	172.845.441	99,9791938

Tabla S2 | Continuación tabla (2/2)

Muestra	Cobertura media	Tamaño de librería estimado	Número total de lecturas	Porcentaje de mapeo
BIOP-98	139,23	397.970.952	142.632.982	99,9648662
BIOP-99	123,31	334.034.651	134.451.453	99,9680544
BIOP-100	252,81	588.807.564	268.566.137	99,9694939
BIOP-101	125,55	398.447.879	130.739.844	99,9688692
BIOP-107	127,89	376.992.494	134.492.389	99,9700303
BIOP-108	127,76	348.132.658	136.316.595	99,9670054
BIOP-110	131,97	417.329.771	135.687.853	99,9678652
BIOP-111	130,57	438.891.431	131.444.168	99,9704692
BIOP-112	135,52	383.231.264	138.713.570	99,9673462
BIOP-115	132,52	302.143.531	136.259.508	99,9668551

- Tabla suplementaria 3

Tabla S3 | Información clínico-patológica y puntuaciones moleculares de malignidad calculados usando los valores de metilación del ADN obtenidos del array de metilación y el algoritmo de predicción en la cohorte de descubrimiento. Se muestran los valores de metilación para cada una de las 3 CpGs del modelo, así como el grupo diagnóstico al que pertenecen y la predicción de benignidad (BN) o malignidad (M) en función de la puntuación de malignidad (MS) obtenida. (1/2)

Muestra	Grupo	cg17154646	cg21915100	cg10189462	MS	Predicción
BIOP-101	FTA	0,9238	0,9542	0,7764	0,0000	BN
BIOP-57	FTA	0,6304	0,9659	0,6404	0,0000	BN
BIOP-90	FTA	0,6632	0,9458	0,6569	0,0000	BN
BIOP-03	FTA	0,8287	0,9364	0,6274	0,0032	BN
BIOP-04	FTA	0,9285	0,9690	0,5592	0,0190	BN
BIOP-07	FTA	0,8878	0,9561	0,5638	0,0190	BN
BIOP-08	FTA	0,7750	0,9409	0,4645	0,2350	BN
BIOP-100	FTA	0,4672	0,9558	0,5921	0,2644	BN
BIOP-06	FTA	0,4840	0,9573	0,4583	0,3112	M
BIOP-52	FTA	0,9158	0,9270	0,4427	0,3842	M
BIOP-02	FTA	0,6963	0,9545	0,2260	0,3966	M
BIOP-05	FTA	0,5483	0,7125	0,4958	0,4428	M
BIOP-53	FTA	0,2968	0,9417	0,2775	0,6988	M
BIOP-01	FTA	0,1882	0,9522	0,2790	0,7194	M
BIOP-91	FTA	0,3949	0,9108	0,4112	0,7616	M
BIOP-18	FTC	0,4566	0,9242	0,2222	0,6495	M
BIOP-14	FTC	0,3078	0,4809	0,9441	0,6782	M
BIOP-51	FTC	0,1854	0,9312	0,2858	0,7935	M
BIOP-55	FTC	0,0766	0,8270	0,4370	0,9921	M
BIOP-16	FTC	0,0605	0,7966	0,0790	0,9966	M
BIOP-12	FTC	0,1587	0,6939	0,3767	0,9994	M
BIOP-11	FTC	0,0658	0,4903	0,2111	1,0000	M
BIOP-15	FTC	0,2086	0,6553	0,3069	1,0000	M
BIOP-17	FTC	0,1911	0,2715	0,4138	1,0000	M
BIOP-56	FTC	0,1845	0,4827	0,2107	1,0000	M
BIOP-93	FVPTC	0,4407	0,8360	0,5594	0,4872	M
BIOP-63	FVPTC	0,2232	0,8569	0,4551	0,8457	M

ANEXO 1: TABLAS SUPLEMENTARIAS

Tabla S3 | Continuación tabla (2/2)

Muestra	Grupo	cg17154646	cg21915100	cg10189462	MS	Predicción
BIOP-115	FVPTC	0,3899	0,4550	0,4114	0,9477	M
BIOP-96	FVPTC	0,3563	0,6793	0,4407	0,9871	M
BIOP-95	FVPTC	0,3455	0,3028	0,3271	0,9904	M
BIOP-98	FVPTC	0,1951	0,3514	0,3414	1,0000	M
BIOP-99	FVPTC	0,2008	0,5248	0,2980	1,0000	M
BIOP-108	NH	0,8765	0,9486	0,6123	0,0000	BN
BIOP-24	NH	0,6454	0,9562	0,6571	0,0000	BN
BIOP-26	NH	0,6386	0,9522	0,7352	0,0000	BN
BIOP-31	NH	0,5255	0,9433	0,6072	0,0000	BN
BIOP-32	NH	0,7080	0,9561	0,7846	0,0000	BN
BIOP-35	NH	0,5595	0,9421	0,7389	0,0000	BN
BIOP-36	NH	0,6624	0,9505	0,7651	0,0000	BN
BIOP-38	NH	0,7270	0,9607	0,6648	0,0000	BN
BIOP-39	NH	0,6301	0,9597	0,6697	0,0000	BN
BIOP-41	NH	0,7977	0,9529	0,6241	0,0000	BN
BIOP-110	NH	0,7209	0,9377	0,6526	0,0023	BN
BIOP-107	NH	0,5563	0,9613	0,5801	0,0033	BN
BIOP-29	NH	0,4947	0,9538	0,7811	0,0081	BN
BIOP-71	NH	0,6309	0,9472	0,5362	0,0116	BN
BIOP-111	NH	0,7186	0,9476	0,5369	0,0118	BN
BIOP-33	NH	0,5016	0,9391	0,5742	0,0158	BN
BIOP-112	NH	0,5850	0,8902	0,6983	0,0162	BN
BIOP-40	NH	0,7100	0,9590	0,5560	0,0238	BN
BIOP-70	NH	0,6315	0,8971	0,5966	0,0255	BN
BIOP-37	NH	0,5688	0,8992	0,6047	0,0267	BN
BIOP-34	NH	0,3712	0,9503	0,6495	0,4965	M
BIOP-30	NH	0,4953	0,8636	0,4803	0,5907	M

- Tabla suplementaria 4

Tabla S4 | Información clínico-patológica y puntuaciones moleculares de malignidad calculados usando los valores de metilación del ADN obtenidos del array de metilación y el algoritmo de predicción en la cohorte de validación externa. Se muestran los valores de metilación para cada una de las 3 CpGs del modelo, así como el grupo diagnóstico al que pertenecen y la predicción de benignidad (BN) o malignidad (M) en función de la puntuación de malignidad (MS) obtenida. Las muestras corresponden al trabajo de Park y colaboradores de (2020) [159]. (1/2)

Muestras	Grupo	cg17154646	cg21915100	cg10189462	MS	Predicción
GSM3433515	Normal	0,6971	0,9515	0,6144	0	BN
GSM3433522	Normal	0,7654	0,9719	0,5961	0	BN
GSM3433527	Normal	0,7232	0,9709	0,6004	0	BN
GSM3433535	Normal	0,8725	0,9527	0,5829	0,0004	BN
GSM3433518	Normal	0,6473	0,9361	0,5947	0,0032	BN
GSM3433533	Normal	0,7019	0,9464	0,5048	0,0072	BN
GSM3433504	Normal	0,7267	0,9656	0,5558	0,0092	BN
GSM3433513	TPTC	0,5844	0,8580	0,5530	0,282	BN
GSM3433512	PTC	0,5806	0,1870	0,4985	0,4522	M
GSM3433520	FVPTC	0,3739	0,9061	0,5186	0,4534	M

Tabla S4 | Continuación tabla (2/2)

Muestras	Grupo	cg17154646	cg21915100	cg10189462	MS	Predicción
GSM3433532	PTC	0,5352	0,5963	0,4554	0,6434	M
GSM3433505	FVPTC	0,6358	0,7207	0,4292	0,6458	M
GSM3433514	PTC	0,6002	0,1391	0,4342	0,6482	M
GSM3433530	PTC	0,5078	0,2564	0,4374	0,6632	M
GSM3433525	PTC	0,4836	0,2501	0,4428	0,673	M
GSM3433517	TPTC	0,4240	0,3570	0,5710	0,8152	M
GSM3433507	TPTC	0,3358	0,6223	0,5992	0,8496	M
GSM3433537	PTC	0,4117	0,6494	0,5244	0,8724	M
GSM3433526	TPTC	0,2253	0,2815	0,4906	0,9276	M
GSM3433529	TPTC	0,3652	0,7148	0,4292	0,9602	M
GSM3433508	PTC	0,3214	0,3579	0,3473	0,9986	M
GSM3433511	PTC	0,2971	0,6965	0,4234	0,9988	M
GSM3433528	FVPTC	0,2636	0,7419	0,4248	0,9994	M
GSM3433509	TPTC	0,1967	0,3771	0,3835	1	M
GSM3433510	TPTC	0,1934	0,3626	0,3731	1	M
GSM3433521	PTC	0,2668	0,5426	0,3935	1	M
GSM3433524	PTC	0,2824	0,3625	0,3423	1	M
GSM3433531	PTC	0,1865	0,2003	0,3688	1	M
GSM3433506	NIFTP	0,3456	0,7399	0,3665	0,9978	M
GSM3433516	NIFTP	0,4756	0,3015	0,3057	0,88	M
GSM3433519	NIFTP	0,2433	0,9026	0,4687	0,6584	M
GSM3433523	NIFTP	0,0917	0,9529	0,2794	0,7444	M
GSM3433534	NIFTP	0,1898	0,7086	0,3059	1	M
GSM3433536	NIFTP	0,2697	0,7962	0,4820	0,9184	M

- Tabla suplementaria 5

Tabla S5 | Información clínico-patológica y puntuaciones moleculares de malignidad calculados usando los valores de metilación del ADN obtenidos mediante pirosecuenciación con bisulfito y el algoritmo de predicción en la cohorte de descubrimiento. Se muestran los valores de metilación para cada una de las 3 CpGs del modelo, así como el grupo diagnóstico al que pertenecen y la predicción de benignidad (BN) o malignidad (M) en función de la puntuación de malignidad (MS) obtenida. (1/2)

Muestra	Grupo	cg17154646	cg21915100	cg10189462	MS	Predicción
BIOP-01	FTA	0,1435	0,8197	0,2267	0,7870	M
BIOP-02	FTA	0,5634	0,7793	0,1662	0,4524	M
BIOP-03	FTA	0,6809	0,7833	0,6927	0,0000	BN
BIOP-04	FTA	0,8665	0,76	0,5773	0,0270	BN
BIOP-05	FTA	0,4185	0,6109	0,4859	0,5530	M
BIOP-06	FTA	0,4698	0,8172	0,4627	0,1240	BN
BIOP-07	FTA	0,7884	0,7853	0,5479	0,0470	BN
BIOP-08	FTA	0,5602	0,7436	0,428	0,3542	M
BIOP-100	FTA	0,4012	0,7654	0,661	0,2328	BN
BIOP-101	FTA	0,8711	0,7766	0,8405	0,0002	BN
BIOP-52	FTA	0,811	0,7282	0,4347	0,3500	M
BIOP-53	FTA	0,255	0,7764	0,2097	0,7564	M
BIOP-57	FTA	0,5116	0,8251	0,6825	0,0000	BN

ANEXO 1: TABLAS SUPLEMENTARIAS

Tabla S5 | Continuación tabla (2/2)

Muestra	Grupo	cg17154646	cg21915100	cg10189462	MS	Predicción
BIOP-90	FTA	0,5333	0,7313	0,6815	0,0000	BN
BIOP-91	FTA	0,3078	0,7109	0,3749	0,7316	M
BIOP-11	FTC	0,1357	0,3768	0,197	1,0000	M
BIOP-12	FTC	0,1563	0,542	0,3493	1,0000	M
BIOP-14	FTC	0,2649	0,411	0,9484	0,5908	M
BIOP-15	FTC	0,1795	0,4835	0,2626	1,0000	M
BIOP-16	FTC	0,0776	0,6916	0,0806	0,9714	M
BIOP-17	FTC	0,1745	0,2747	0,3725	1,0000	M
BIOP-18	FTC	0,3974	0,761	0,1949	0,4847	M
BIOP-51	FTC	0,1374	0,7696	0,2488	0,7493	M
BIOP-55	FTC	0,0965	0,6759	0,4459	0,9456	M
BIOP-56	FTC	0,1518	0,405	0,1638	1,0000	M
BIOP-115	FVPTC	0,2392	0,4353	0,4019	0,9995	M
BIOP-63	FVPTC	0,1935	0,6978	0,4224	0,8007	M
BIOP-93	FVPTC	0,3938	0,6667	0,5719	0,1877	BN
BIOP-95	FVPTC	0,2958	0,2726	0,2266	0,9847	M
BIOP-96	FVPTC	0,2839	0,6306	0,4389	0,9594	M
BIOP-98	FVPTC	0,2011	0,3518	0,3145	1,0000	M
BIOP-99	FVPTC	0,1651	0,4035	0,2322	1,0000	M
BIOP-107	NH	0,4516	0,8048	0,6293	0,0005	BN
BIOP-108	NH	0,7287	0,7818	0,6495	0,0000	BN
BIOP-110	NH	0,5722	0,7781	0,712	0,0000	BN
BIOP-111	NH	0,529	0,7779	0,5813	0,0181	BN
BIOP-112	NH	0,4177	0,7598	0,7188	0,1362	BN
BIOP-24	NH	0,5293	0,7377	0,7018	0,0000	BN
BIOP-26	NH	0,4871	0,7786	0,7783	0,0000	BN
BIOP-29	NH	0,396	0,7818	0,7762	0,1255	BN
BIOP-30	NH	0,4517	0,7002	0,4689	0,5144	M
BIOP-31	NH	0,4152	0,7798	0,6472	0,0205	BN
BIOP-32	NH	0,5378	0,7786	0,788	0,0000	BN
BIOP-33	NH	0,3791	0,7817	0,6175	0,1997	BN
BIOP-34	NH	0,314	0,7775	0,6861	0,3234	M
BIOP-35	NH	0,4738	0,7422	0,7724	0,0005	BN
BIOP-36	NH	0,5897	0,7882	0,7973	0,0000	BN
BIOP-37	NH	0,4494	0,6969	0,6218	0,1250	BN
BIOP-38	NH	0,5693	0,7761	0,6995	0,0000	BN
BIOP-39	NH	0,5528	0,7481	0,7196	0,0000	BN
BIOP-40	NH	0,6045	0,7873	0,5783	0,0183	BN
BIOP-41	NH	0,7386	0,8097	0,679	0,0000	BN
BIOP-70	NH	0,4718	0,6647	0,6087	0,3734	M
BIOP-71	NH	0,5063	0,7516	0,5434	0,1338	BN

- Tabla suplementaria 6

Tabla S6 | Información clínico-patológica y puntuaciones moleculares de malignidad calculados usando los valores de metilación del ADN obtenidos mediante pirosecuenciación con bisulfito y el algoritmo de predicción en la cohorte de validación. Se muestran los valores de metilación para cada una de las 3 CpGs del modelo, así como el grupo diagnóstico al que pertenecen y la predicción de benignidad (BN) o malignidad (M) en función de la puntuación de malignidad (MS) obtenida. (1/3)

Muestra	Grupo	cg17154646	cg21915100	cg10189462	MS	Predicción
BIOP-09	FTA	0,1738	0,645	0,2189	0,9936	M
BIOP-10	FTA	0,5208	0,3263	0,3625	0,6278	M
BIOP-54	FTA	0,6303	0,7439	0,7338	0	BN
BIOP-243	FTA	0,1354	0,5047	0,1789	1	M
BIOP-210	FTA	0,7382	0,7064	0,5141	0,0896	BN
BIOP-227	FTA	0,4774	0,7458	0,6243	0,0006	BN
BIOP-236	FTA	0,2136	0,6698	0,566	0,848	M
BIOP-238	FTA	0,5587	0,6905	0,7499	0,0974	BN
BIOP-241	FTA	0,5807	0,7331	0,5268	0,0488	BN
BIOP-242	FTA	0,6113	0,6861	0,6799	0,1176	BN
BIOP-244	FTA	0,4103	0,7268	0,8701	0,0606	BN
BIOP-201	FTA	0,2538	0,6707	0,4131	0,9894	M
BIOP-235	FTA	0,9023	0,7417	0,7141	0	BN
BIOP-58	FTA	0,3828	0,7936	0,5583	0,23	BN
BIOP-59	FTA	0,417	0,7216	0,436	0,4292	M
BIOP-178	FTA	0,7301	0,7415	0,8717	0,0236	BN
BIOP-180	FTA	0,4691	0,7258	0,6595	0,0004	BN
BIOP-120	FTA	0,5078	0,7171	0,606	0,013	BN
BIOP-117	FTA	0,4982	0,7339	0,8566	0,0014	BN
BIOP-119	FTA	0,6414	0,7332	0,4748	0,1284	BN
BIOP-124	FTA	0,4456	0,7972	0,7783	0,0002	BN
BIOP-127	FTA	0,2604	0,3226	0,7286	0,7592	M
BIOP-118	FTA	0,5743	0,7497	0,6675	0,0006	BN
BIOP-13	FTC	0,2688	0,6383	0,3384	0,9914	M
BIOP-19	FTC	0,7467	0,6937	0,2242	0,5834	M
BIOP-194	FTC	0,5456	0,7173	0,2611	0,5138	M
BIOP-237	FTC	0,4501	0,6175	0,4972	0,3866	M
BIOP-182	FTC	0,3244	0,6946	0,281	0,8716	M
BIOP-148	FTC	0,2193	0,7613	0,6502	0,504	M
BIOP-150	FTC	0,0939	0,7561	0,1297	0,9176	M
BIOP-151	FTC	0,5301	0,5914	0,5792	0,2982	BN
BIOP-154	FTC	0,6195	0,7689	0,137	0,5208	M
BIOP-156	FTC	0,4619	0,6997	0,1717	0,5208	M
BIOP-152	FTC	0,2522	0,3151	0,258	0,9998	M
BIOP-116	FTC	0,1252	0,3375	0,1876	1	M
BIOP-23	FVPTC	0,1376	0,7376	0,5075	0,5948	M
BIOP-97	FVPTC	0,1358	0,6999	0,3159	0,936	M
BIOP-233	FVPTC	0,2978	0,4056	0,2479	0,9944	M
BIOP-232	FVPTC	0,246	0,3649	0,4133	0,9998	M
BIOP-64	FVPTC	0,1741	0,4344	0,265	1	M
BIOP-65	FVPTC	0,1636	0,6107	0,3239	0,9936	M
BIOP-191	FVPTC	0,5525	0,1866	0,3218	0,652	M
BIOP-188	FVPTC	0,1641	0,4749	0,2666	1	M

ANEXO 1: TABLAS SUPLEMENTARIAS
Tabla S6 | Continuación tabla (2/3)

Muestra	Grupo	cg17154646	cg21915100	cg10189462	MS	Predicción
BIOP-193	FVPTC	0,2069	0,6642	0,7992	0,6516	M
BIOP-176	FVPTC	0,1601	0,4914	0,2255	1	M
BIOP-189	FVPTC	0,3954	0,6278	0,5982	0,4052	M
BIOP-192	FVPTC	0,2534	0,2823	0,3625	0,9998	M
BIOP-94	FVPTC	0,29	0,6124	0,3747	0,9824	M
BIOP-157	FVPTC	0,6444	0,3605	0,2399	0,652	M
BIOP-159	FVPTC	0,1642	0,1795	0,2612	1	M
BIOP-165	FVPTC	0,2841	0,3818	0,4784	0,9244	M
BIOP-167	FVPTC	0,102	0,2931	0,1539	1	M
BIOP-173	FVPTC	0,2902	0,5261	0,5691	0,865	M
BIOP-161	FVPTC	0,2549	0,2479	0,3585	0,9998	M
BIOP-169	FVPTC	0,3959	0,4274	0,4047	0,8488	M
BIOP-171	FVPTC	0,3379	0,3969	0,5401	0,7464	M
BIOP-162	FVPTC	0,378	0,3502	0,1853	0,912	M
BIOP-259	HTA	0,4855	0,7621	0,6148	0,0426	BN
BIOP-264	HTA	0,5061	0,7118	0,7204	0,0152	BN
BIOP-265	HTA	0,8105	0,6959	0,5619	0,1698	BN
BIOP-266	HTA	0,4902	0,6963	0,664	0,0448	BN
BIOP-60	HTA	0,0892	0,6796	0,1735	0,9982	M
BIOP-258	HTC	0,144	0,5475	0,4427	0,9974	M
BIOP-257	micro-CPTC	0,8029	0,8249	0,7668	0,3766	M
BIOP-25	NH	0,5886	0,7604	0,832	0,0146	BN
BIOP-27	NH	0,4979	0,7397	0,6783	0	BN
BIOP-28	NH	0,6584	0,7125	0,7374	0,015	BN
BIOP-43	NH	0,4661	0,7947	0,6823	0,0002	BN
BIOP-44	NH	0,6138	0,5554	0,5334	0,3806	M
BIOP-45	NH	0,6093	0,7579	0,8184	0,014	BN
BIOP-46	NH	0,7489	0,7936	0,7564	0	BN
BIOP-47	NH	0,5736	0,7795	0,6899	0	BN
BIOP-48	NH	0,3114	0,7654	0,7272	0,2762	BN
BIOP-49	NH	0,4415	0,6952	0,5829	0,1048	BN
BIOP-50	NH	0,7245	0,8178	0,7886	0	BN
BIOP-195	NH	0,5167	0,6059	0,6992	0,2022	BN
BIOP-196	NH	0,5782	0,6417	0,7107	0,202	BN
BIOP-197	NH	0,3957	0,6647	0,7024	0,3342	M
BIOP-198	NH	0,6096	0,6905	0,5514	0,2132	BN
BIOP-199	NH	0,4557	0,6375	0,6381	0,2024	BN
BIOP-200	NH	0,7723	0,7056	0,582	0,0416	BN
BIOP-202	NH	0,5947	0,7071	0,7924	0,019	BN
BIOP-204	NH	0,5677	0,6872	0,6834	0,0992	BN
BIOP-205	NH	0,4664	0,6673	0,5119	0,3512	M
BIOP-68	NH	0,4323	0,5856	0,6195	0,273	BN
BIOP-69	NH	0,3548	0,7466	0,6269	0,1932	BN
BIOP-72	NH	0,487	0,6442	0,7267	0,202	BN
BIOP-74	NH	0,4324	0,727	0,7068	0,0046	BN
BIOP-75	NH	0,6108	0,7518	0,8931	0,0722	BN
BIOP-76	NH	0,4411	0,7283	0,7542	0,001	BN
BIOP-77	NH	0,6151	0,7182	0,5515	0,057	BN

Tabla S6 | Continuación tabla (3/3)

Muestra	Grupo	cg17154646	cg21915100	cg10189462	MS	Predicción
BIOP-78	NH	0,5922	0,7084	0,3491	0,463	M
BIOP-79	NH	0,8359	0,732	0,7048	0	BN
BIOP-80	NH	0,5017	0,694	0,6672	0,0974	BN
BIOP-81	NH	0,3192	0,7833	0,582	0,1904	BN
BIOP-83	NH	0,7069	0,7606	0,838	0,0438	BN
BIOP-84	NH	0,4561	0,688	0,5696	0,2336	BN
BIOP-85	NH	0,2699	0,7773	0,636	0,3346	M
BIOP-86	NH	0,5779	0,7536	0,5494	0,0496	BN
BIOP-87	NH	0,6018	0,7635	0,4359	0,3918	M
BIOP-186	NH	0,5284	0,6812	0,6452	0,168	BN
BIOP-133	NH	0,3791	0,7291	0,7773	0,1766	BN
BIOP-135	NH	0,5566	0,7813	0,7878	0	BN
BIOP-137	NH	0,6509	0,7093	0,6049	0,016	BN
BIOP-138	NH	0,5721	0,7461	0,6592	0	BN
BIOP-143	NH	0,5535	0,7648	0,576	0,0742	BN
BIOP-103	NH	0,4493	0,7451	0,5684	0,0818	BN
BIOP-104	NH	0,5984	0,7591	0,7572	0,0144	BN
BIOP-106	NH	0,422	0,7522	0,7501	0,035	BN
BIOP-109	NH	0,7002	0,7393	0,6367	0,0006	BN
BIOP-113	NH	0,3206	0,7602	0,7692	0,2464	BN
BIOP-114	NH	0,549	0,7866	0,7842	0	BN
BIOP-245	PTC	0,2448	0,4402	0,486	0,9262	M
BIOP-246	PTC	0,2776	0,4404	0,6222	0,7634	M
BIOP-247	PTC	0,098	0,7746	0,1675	0,7998	M
BIOP-248	PTC	0,1976	0,233	0,3656	1	M
BIOP-249	PTC	0,3276	0,1443	0,3636	0,9244	M
BIOP-250	PTC	0,5793	0,7368	0,5389	0,048	BN
BIOP-251	PTC	0,51	0,7458	0,6898	0	BN
BIOP-252	PTC	0,127	0,1448	0,301	1	M
BIOP-253	PTC	0,6613	0,2382	0,1728	0,652	M
BIOP-254	PTC	0,3061	0,4674	0,4645	0,8298	M
BIOP-255	PTC	0,3574	0,5347	0,7141	0,538	M
BIOP-256	PTC	0,1582	0,5023	0,5191	0,8952	M
BIOP-260	PTC	0,1917	0,2617	0,3459	1	M
BIOP-261	PTC	0,317	0,5936	0,6089	0,5488	M
BIOP-262	PTC	0,258	0,3005	0,6412	0,7606	M
BIOP-263	PTC	0,4282	0,6889	0,7551	0,1338	BN

ANEXO 1: TABLAS SUPLEMENTARIAS

- Tabla suplementaria 7

Tabla S7 Información clínico-patológica y puntuaciones moleculares de malignidad calculados usando los valores de metilación del ADN obtenidos mediante pirosecuenciación con bisulfito y el algoritmo de predicción en la cohorte de PAAF. Se muestran los valores de metilación para cada una de las 3 CpGs del modelo, así como el grupo diagnóstico al que pertenecen y la predicción de benignidad (BN) o malignidad (M) en función de la puntuación de malignidad (MS) obtenida. Se adjunta también la información relativa a los diagnósticos anatomopatológicos postcirugía (PostQx)

Paciente	cg17154646	cg21915100	cg10189462	B	Predicción	MS	PostQx
PAAF-004	0,49	0,63	0,79	B3	BN	0,220	FTA
PAAF-020	0,49	0,66	0,77	B2	BN	0,136	NH
PAAF-032	0,62	0,64	0,82	B2	BN	0,220	NH
PAAF-034	0,41	0,69	0,66	B2	BN	0,174	NH
PAAF-042	0,45	0,68	0,74	B2	BN	0,132	NH
PAAF-043	0,56	0,70	0,80	B6	BN	0,024	PTC
PAAF-048	0,65	0,66	0,83	B3	BN	0,153	NH
PAAF-049	0,34	0,63	0,76	B3	M	0,480	2ª PAAF: BII
PAAF-051	0,63	0,69	0,70	B2	BN	0,097	NH
PAAF-052	0,64	0,71	0,52	B2	BN	0,080	NA
PAAF-054	0,40	0,63	0,41	B2	M	0,845	NA
PAAF-055	0,62	0,72	0,89	B2	BN	0,071	NA
PAAF-056	0,69	0,68	0,93	B2	BN	0,251	NA
PAAF-057	0,59	0,67	0,88	B2	BN	0,250	NA
PAAF-059	0,54	0,65	0,79	B2	BN	0,154	NA
PAAF-063	0,53	0,45	0,57	B3	M	0,425	PAAF-071
PAAF-064	0,66	0,73	0,91	B2	BN	0,071	NA
PAAF-070	0,64	0,69	0,80	B2	BN	0,109	NA
PAAF-071	0,31	0,22	0,27	B5	M	0,940	FVPTC
PAAF-078	0,42	0,70	0,83	B2	BN	0,074	NA
PAAF-084	0,70	0,70	0,83	B2	BN	0,057	NA
PAAF-089	0,38	0,64	0,83	B2	M	0,446	NA
PAAF-108	0,41	0,71	0,41	B2	M	0,477	NA
PAAF-109	0,42	0,63	0,61	B2	BN	0,289	NA
PAAF-111	0,67	0,70	0,61	B2	BN	0,022	NA
PAAF-113	0,33	0,74	0,88	B2	M	0,303	NA
PAAF-142	0,42	0,59	0,74	B3	M	0,304	micro-PTC
PAAF-193	0,10	0,31	0,27	B6	M	1,000	PTC
PAAF-205	0,64	0,68	0,82	B3	BN	0,184	2ª PAAF: BII
PAAF-216	0,52	0,41	0,93	B4	M	0,380	FTA
PAAF-222_A	0,14	0,22	0,21	B5	M	1,000	FTC
PAAF-222_B	0,20	0,25	0,21	B3	M	1,000	CPTC
PAAF-223	0,53	0,65	0,90	B3	BN	0,222	Pendiente

ANEXO 2: FIGURAS SUPLEMENTARIAS

• Figura suplementaria 1

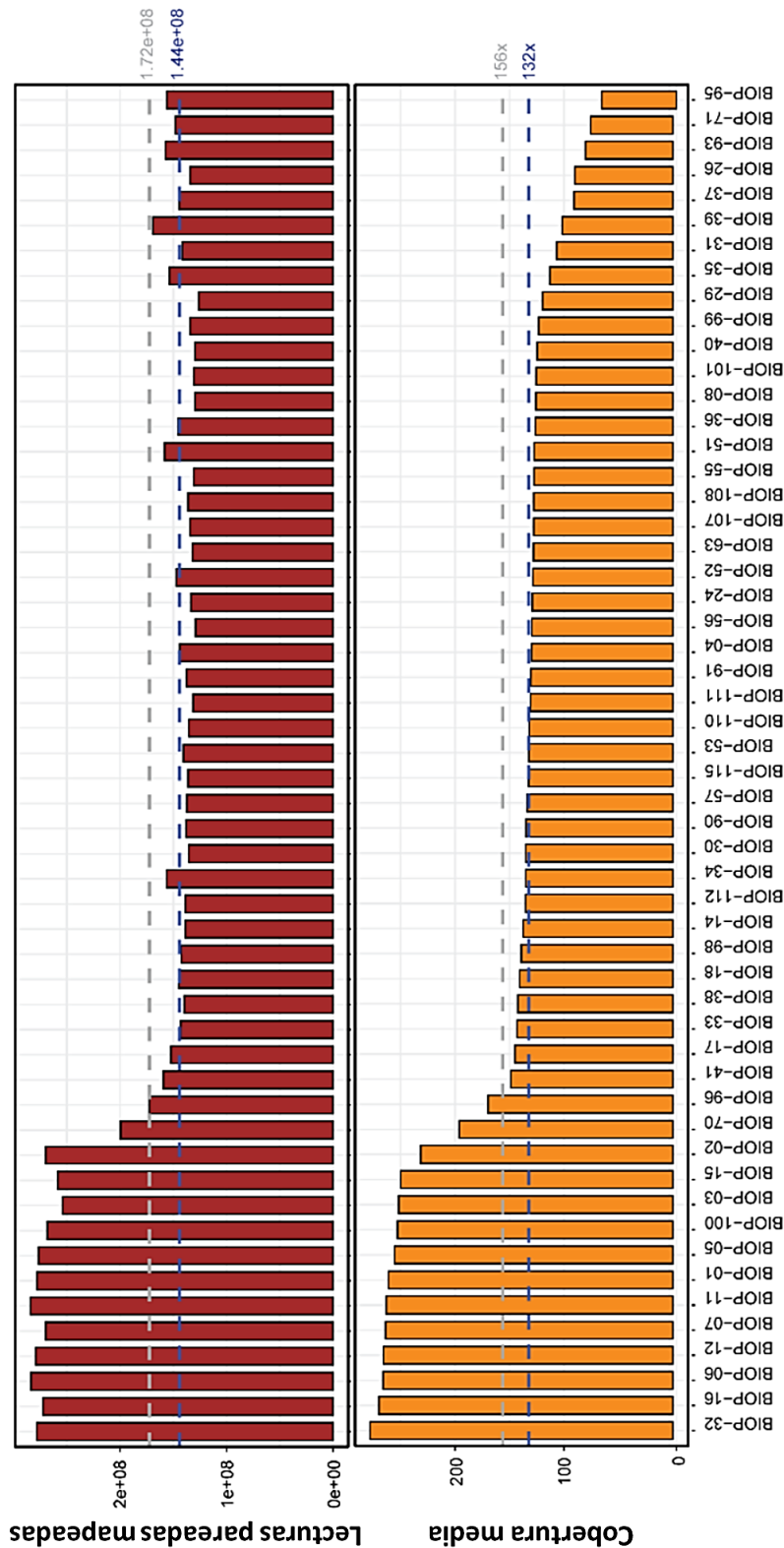


Figura S1| Información adicional relacionada con el experimento de secuenciación del exoma. Gráficos de barras que reflejan el número de lecturas finales emparejadas de secuenciación asignadas al genoma humano (hg38, gráfico superior) o la cobertura media del exoma (gráfico inferior) en el experimento de secuenciación del exoma. Las líneas grises y azules representan, respectivamente, las puntuaciones media y mediana de todas las muestras

• Figura suplementaria 2

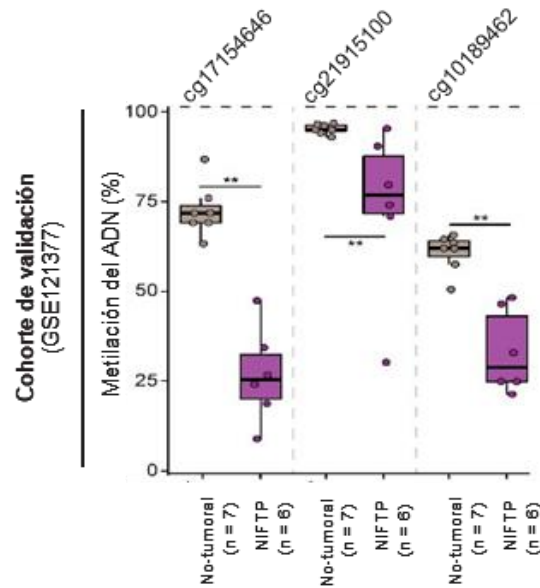


Figura S2 | Niveles de metilación de las 3 CpG (cg17154646, cg21915100 y cg10189462) en NIFTP con respecto a los nódulos no tumorales. Diagramas de caja que ilustran los valores de metilación del ADN de los CpG indicados en el contexto de muestras benignas y NIFTP de un conjunto de datos de validación GSE121377

• Figura suplementaria 3

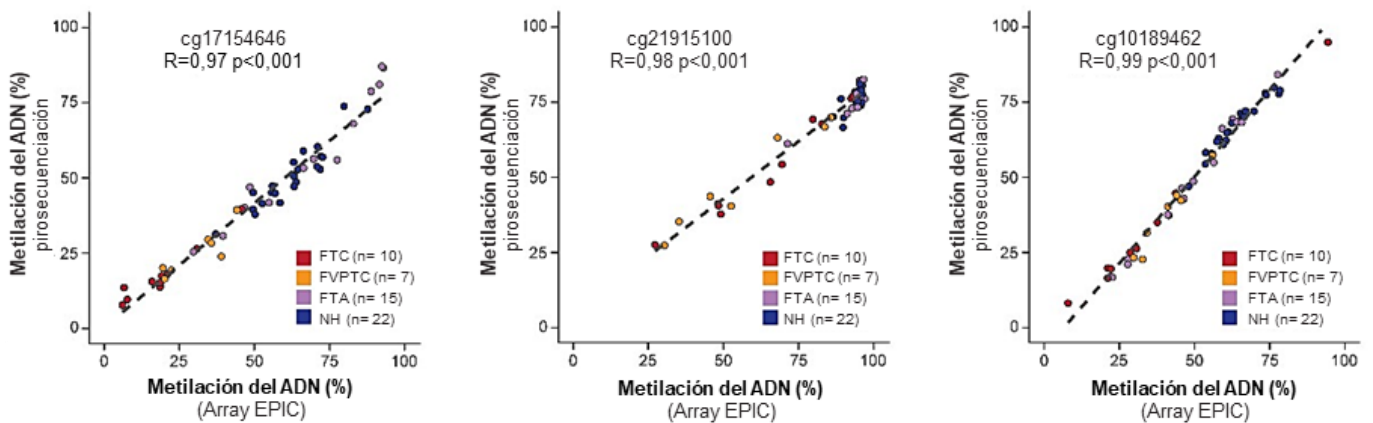


Figura S3 | Correlación entre los valores de metilación mediante pirosecuenciación con bisulfito y los arrays de metilación para cada una de las 3 CpGs (cg17154646, cg21915100 y cg10189462) que componen el clasificador de lesiones tiroideas de la cohorte de validación. Se muestra la puntuación de la correlación de Pearson, la significación estadística y el número de casos incluidos en los diferentes subtipos histopatológicos.

• Figura suplementaria 4

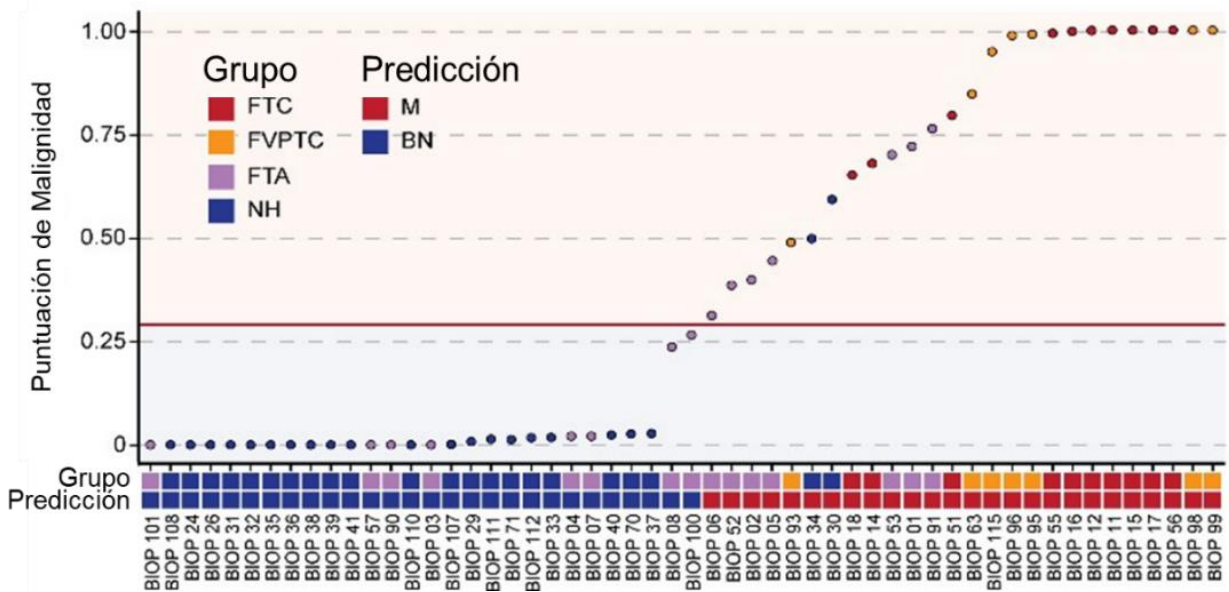


Figura S4 | Puntuaciones de malignidad del conjunto de datos de descubrimiento utilizando arrays de metilación EPIC. Gráfico de puntos que muestra las puntuaciones de malignidad de los tumores tiroideos inferidas de las muestras utilizadas en el conjunto de datos de entrenamiento. Se indica el tipo de cáncer y la predicción diagnóstica (BN benigno o M maligno) de cada muestra utilizando el clasificador de bosque aleatorio de tres CpG.

• Figura suplementaria 5

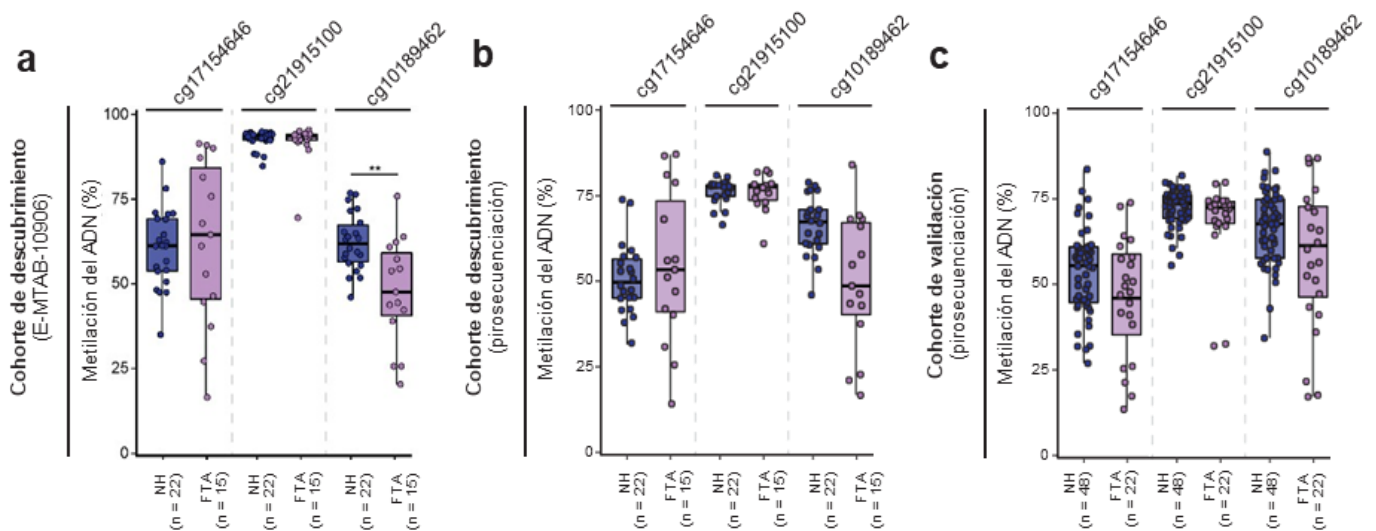


Figura S5 | Diferencias estadísticas entre los niveles de metilación en las muestras NH y FTA de las cohortes de descubrimiento y validación Diagrama de caja que indica los valores de metilación del

ADN de las muestras de NH y FTA en el contexto de los tres sitios CpG utilizados para el modelado de la malignidad molecular del cáncer de tiroides mediante pirosecuenciación.

• **Figura suplementaria 6**

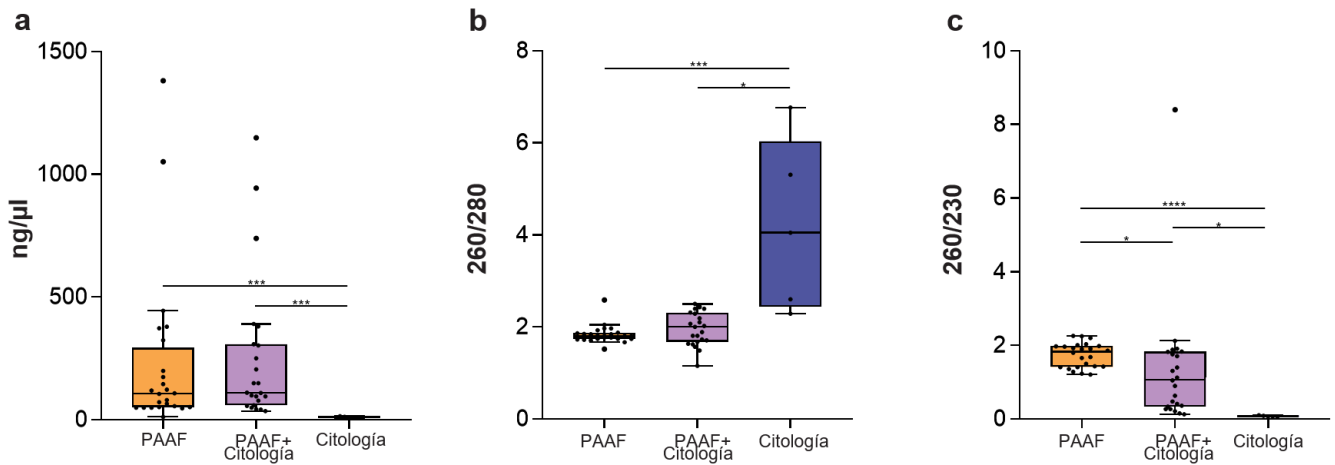


Figura S6 | Comparación de la calidad de las muestras tras la extracción de material genético a partir de PAAF (n =24), citologías de archivo (n =5) o la combinación de ambos (n = 23). Diagrama de cajas que muestran: (a) Concentración del ADN (ng/μl) para cada muestra testada. (b) La relación de absorbancia a 260 nm y 280 nm. Se utiliza para evaluar la pureza del ADN y el ARN. Una proporción de ~1.8 generalmente se acepta como "pura" para el ADN. Valores inferiores suelen indicar que una muestra está contaminada con fenol residual, guanidina u otro reactivo utilizado en el protocolo de extracción. También se pueden encontrar proporciones imprecisas a concentraciones muy bajas (< 10 ng/μl) de ácidos nucleicos. (c) La relación 260/230 se utiliza como medida secundaria de la pureza de los ácidos nucleicos. Los valores para considerar una muestra "pura", suelen ser más altos que los respectivos valores del ratio 260/280. Los valores esperados están comúnmente en el rango de 2.0-2.2. Si la relación es apreciablemente más baja de lo esperado, puede indicar la presencia de contaminantes que absorben a 230 nm. La guanidina HCL, los carbohidratos y el fenol tienen una absorbancia cercana a los 230 nm. La significación estadística entre las condiciones se calculó mediante una prueba no paramétrica (test de Wilcoxon; * p-valor < 0,05, ** p-valor < 0,01, *** p-valor < 0,001)

• **Figura suplementaria 7**

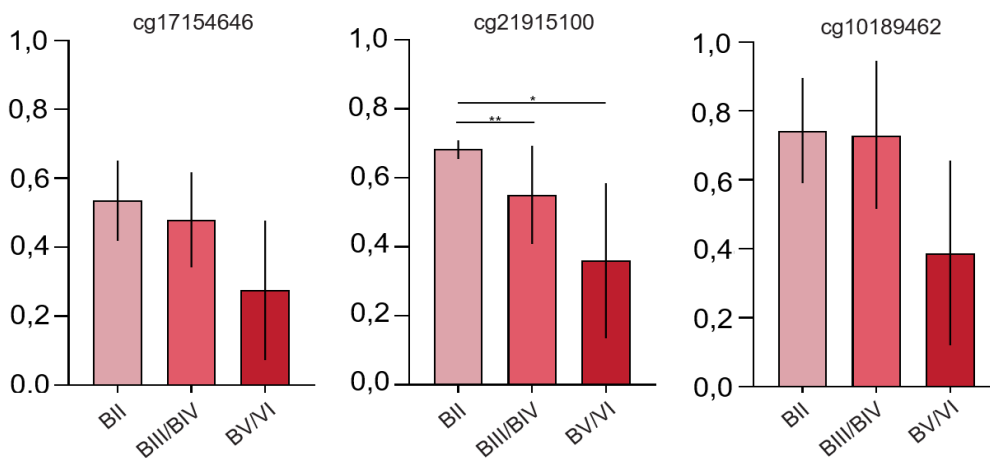


Figura S7 | Discriminación de PAAF mediante el clasificador epigenético. Gráficos de barras que muestran el porcentaje de metilación de los tres CpG analizados por pirosecuenciación por grupo de muestras Bethesda II, III/IV y V/VI.



Universidad de Oviedo

**Programa Oficial de Doctorado
de Ciencias de la Salud**

