



ESCUELA DE DOCTORADO
INTERNACIONAL DE LA USC

Olivia
Fuentes Ríos

Tesis doctoral

NUEVAS ESTRATEGIAS DE ANÁLISIS EN RADIOGENÓMICA DEL CÁNCER DE PRÓSTATA

Santiago de Compostela, 2023

Programa de doctorado en Avances y Nuevas Estrategias en Ciencias Forenses



TESIS DE DOCTORADO

NUEVAS ESTRATEGIAS DE ANÁLISIS EN RADIOGENÓMICA DEL CÁNCER DE PRÓSTATA

Olivia Fuentes Ríos

Directora: Ana Paula Vega Gliemmo

Tutor: Antonio Salas Ellacurriaga

**ESCUELA DE DOCTORADO INTERNACIONAL DE
LA UNIVERSIDAD DE SANTIAGO DE COMPOSTELA**

**PROGRAMA DE DOCTORADO EN
AVANCES Y NUEVAS ESTRATEGIAS EN CIENCIAS FORENSES**

SANTIAGO DE COMPOSTELA

2023



D./Dña. **Olivia Fuentes Ríos**

Título da tese: **NUEVAS ESTRATEGIAS DE ANÁLISIS EN RADIOGENÓMICA DEL
CÁNCER DE PRÓSTA**

Presento mi tesis, siguiendo el procedimiento adecuado al Reglamento y declaro que:

- 1) La tesis abarca los resultados de la elaboración de mi trabajo.
- 2) De ser el caso, en la tesis se hace referencia a las colaboraciones que tuvo este trabajo.
- 3) Confirmando que la tesis no incurre en ningún tipo de plagio de otros autores ni de trabajos presentados por mí para la obtención de otros títulos.
- 4) La tesis es la versión definitiva presentada para su defensa y coincide la versión impresa con la presentada en formato electrónico.

Y me comprometo a presentar el Compromiso Documental de Supervisión en el caso que el original no esté depositado en la Escuela.

En Santiago de Compostela, 23 de noviembre de 2023.

Firma electrónica

Este trabajo fue financiado por el Instituto de Salud Carlos III (ISCIII) y el Fondo Europeo de Desarrollo Regional (FEDER) FIS-PII6/0046. La doctoranda recibió apoyo del ISCIII y del Fondo Social Europeo (FSE) PFIS-FI17/00261 (Contrato predoctoral de formación en investigación en salud PFIS enero 2018 – junio 2022). La realización de este proyecto ha sido posible gracias a la colaboración del Centro de Supercomputación de Galicia (CESGA).



Yo, Olivia Fuentes Ríos, con DNI Y5322937S, declaro no tener ningún conflicto de interés en relación con esta tesis.

En Santiago de Compostela, a 18 de julio de 2023

Firmado: Olivia Fuentes Ríos

RESUMEN

El cáncer de próstata es el cáncer con mayor incidencia en hombres a nivel europeo y el segundo a nivel mundial. Aproximadamente el 50 % de los pacientes con cáncer recibirán radioterapia como parte de su tratamiento. A pesar de que los avances tecnológicos en radioterapia permiten aplicar el tratamiento de forma cada vez más precisa en el tumor, parte del tejido sano será inevitablemente irradiado, lo que puede dar lugar al desarrollo de efectos adversos. Estos efectos impactarán negativamente en la calidad de vida del paciente. Se ha descrito que los factores genéticos desempeñan un papel muy importante en el desarrollo y severidad de los efectos adversos a la radioterapia. Así, el propósito de esta tesis es la identificación de marcadores genéticos asociados con el sangrado rectal, uno de los efectos adversos más relevantes en el tratamiento radioterápico de pacientes con cáncer de próstata.

Para ello realizamos la secuenciación del exoma completo (WES) en 141 pacientes (46 casos y 95 controles) con cáncer de próstata tratados con radioterapia de nuestra cohorte, e incluimos datos de otros 391 pacientes (69 casos y 322 controles) de una cohorte externa de validación. A partir de estos resultados realizamos las siguientes aproximaciones: i) Investigamos variantes y genes previamente asociados con el desarrollo de efectos adversos en pacientes tratados con radioterapia; ii) Realizamos estudios de asociación basados en genes a partir de datos de WES; iii) Realizamos estudios de asociación del transcriptoma (TWAS) y del proteoma (PWAS) como nuevas aproximaciones ómicas; iv) Exploramos las rutas en las que están implicados los principales genes asociados con el sangrado rectal de nuestro estudio.

AGRADECIMIENTOS

Todos los cambios comienzan con un paso. A veces caminamos sin tener un destino fijo; en otras ocasiones, sabemos exactamente el camino que queremos tomar. Hay momentos en los que necesitamos algo más y, en lugar de caminar, damos un salto al infinito con valentía, en busca de ese 'algo', emocionados por probar cosas nuevas y esperando que el mundo nos sorprenda. Aunque implica arriesgar mucho, también sabemos que hay mucho por ganar y aprender. Agradezco a todas las personas que, de alguna manera, me han acompañado a lo largo de este recorrido.

En primer lugar, quiero expresar mi agradecimiento a mi directora de tesis, Ana Vega. Gracias por la oportunidad que me has brindado, por tu confianza, por tu inmensa paciencia, por ser mi guía y por tu apoyo tanto profesional como personal. Tu gran dedicación y entrega a la ciencia inspiran y motivan. Agradezco sinceramente que me hayas permitido formar parte de tu equipo.

Gracias a todos los que forman o formaron parte del grupo de investigación de Genética en Cáncer y Enfermedades Raras. Ana Blanco, Belinda, Javi, Ana C., Carlos López, muchas gracias por tu colaboración en esta tesis. Marta, gracias por tus palabras de ánimo en estos últimos años. Carlos Pérez, gracias por auxiliarme en mis dudas a pesar de que ya no estás por aquí. Miguel, gracias por tu apoyo en el trayecto de esta tesis y por introducirme en el mundo de la radiogenómica. Paloma, mil gracias por tu ayuda a la distancia. Ha sido un placer haber coincidido contigo. Y a Juan Fernández, no tengo cómo agradecer tu apoyo; mil gracias, lo valoro demasiado.

A las chicas de NGS, Bea, Sandra, Fátima y María, gracias por su amable apoyo y por compartir conmigo sus conocimientos y experiencia. A Jorge A., Vanesa y a todos los miembros de la Fundación Pública Galega de Medicina Xenómica, en especial a Ángel Carracedo, por su cercanía y amabilidad que se valora mucho.

Quiero agradecer a Silvia. Gracias por escucharme, por apoyarme, por siempre tener palabras dulces y amables. Eres una gran persona y amiga. Sin ti, me hubiese vuelto más loca. TQM. A Fernando, un chico maravilloso siempre listo para compartir sus conocimientos y brindar mucho ánimo. También a Carmen, por su gran comprensión, por el tiempo juntas y los momentos del café.

Agradezco al departamento de Oncología Radioterápica del Complejo Hospitalario Universitario de Santiago. Esta tesis va dedicada especialmente a todos los pacientes con cáncer de próstata que aceptaron participar en este estudio.

LISTA DE ABREVIATURAS

AJCC	<i>American Joint Committee on Cancer</i>
AUASS	<i>American Urological Association Symptom Score</i>
BH-FDR	<i>Benjamin-Hochberg False Discovery Rate</i>
CADD	<i>Combined Annotation Dependent Depletion</i>
CCI	<i>Cross Cancer Institute</i>
CCI-BT	<i>Cross Cancer Institute–Brachytherapy</i>
CCIEBRT	<i>Cross Cancer Institute–External Beam Radiotherapy</i>
CCV	<i>Credible Causal Variant</i>
CNV	<i>Copy-number variant</i>
CTCAE	<i>Common Terminology Criteria for Adverse Events</i>
EAU	<i>European Association of Urology</i>
EANM	<i>European Association of Nuclear Medicine</i>
EBRT	<i>External Beam Radiation Therapy</i>
EORTC	<i>European Organisation for Research and Treatment of Cancer</i>
ESMO	<i>Sociedad Europea de Oncología Médica</i>
ESTRO	<i>European Society for Radiotherapy and Oncology</i>
ESUR	<i>European Society of Urogenital Radiology</i>
eQTL	<i>Expression quantitative trait loci</i>
FDR	<i>False Discovery Rate</i>
FSHR	<i>Follicle-stimulating hormone receptor</i>
FROG	<i>Florida Radiation Oncology Group</i>
Gene-PARE	<i>Genetic Predictors of Adverse Radiotherapy Effects</i>
GI	<i>Gastrointestinal</i>
GO	<i>Gene Ontology</i>
GTEX	<i>Genotype-Tissue Expression</i>
GU	<i>Genitourinary</i>

gVCF	<i>Genomic Variant Call Format</i>
GWAS	<i>Genome-wide association studies</i> - Estudios de asociación del genoma completo
Gy	Gray, medida de dosis en radioterapia
HNC	<i>Head and neck cancer</i>
HRC	<i>Haplotype Reference Consortium</i>
IARC	<i>International Agency for Research on Cancer</i> - Agencia Internacional para la Investigación del Cáncer
IPSS	<i>International Prostate Symptom Score</i>
ISUP	<i>International Society of Urological Pathology</i> - Sociedad Internacional de Urología Patológica
LD	Desequilibrio de ligamiento
LENT	<i>Late Effects Normal Tissue</i>
LENT-SOMA	<i>Late effects on Normal Tissues-Subjective Objective Management Analytic</i>
MAASTRO	<i>Maastricht Radiation Oncology Clinic</i>
MSEF	<i>Mount Sinai Erectile Function</i>
MSMC	<i>Mount Sinai Medical Center</i>
NGS	<i>Next-Generation Sequencing</i>
NTMC	<i>National Tokyo Medical Center</i>
OM	<i>Oral mucositis</i>
OR	<i>Odds ratio</i>
PRACTICAL	<i>Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome</i>
PSA:	Antígeno prostático específico
PWAS	<i>Proteome Wide Association Study</i>
RADIOGEN	Estudio de la Influencia Genómica en Pacientes Tratados con Radioterapia
RAPPER	<i>Radiogenomics: Assessment of Polymorphisms for Predicting the Effects of Radiotherapy</i>
RectV65	Porcentaje de volumen del recto que recibe al menos 65 Gy
RectV70	Porcentaje de volumen del recto que recibe al menos 70 Gy
REQUIRE	<i>Validating Predictive Models and Biomarkers of Radiotherapy Toxicity to Reduce Side-Effects and Improve Quality of Life in Cancer Survivors</i>

REVEL	<i>Rare exome variant ensemble learner</i>
RGC	<i>RadioGenomics Consortium</i>
RMH	<i>Royal Marsden Hospital</i>
RT	<i>Radioterapia</i>
RTOG	<i>Radiation Therapy Oncology Group</i>
SHIM	<i>Sexual Health Inventory for Men</i>
SKAT-O	<i>Optimal Sequence Kernel Association test</i>
SIOG	<i>International Society of Geriatric Oncology</i>
SNP	<i>Single Nucleotide Polymorphisms</i>
SOMA	<i>Subjective Objective Management Analytic</i>
STAT	<i>Standardized Total Average Toxicity</i>
TBCC	<i>Tom Baker Cancer Centre</i>
TNM	<i>Clasificación del tamaño del tumor (T), la afectación de los ganglios linfáticos (N) y la metástasis (M)</i>
TWAS	<i>Transcriptome-wide association studies</i>
UGhent	<i>University of Ghent</i>
UICC	<i>Union for International Cancer Control</i>
VQSR	<i>Variant quality score recalibration</i>
VCF	<i>Variant call format</i>
WES	<i>Whole exome sequencing</i>
X2	<i>Chi-cuadrada</i>
3D-CRT	<i>Three-dimensional Conformal Radiation Therapy</i>

INDICE

ÍNDICE

1 INTRODUCCIÓN.....	25
1.1 CÁNCER DE PRÓSTATA	27
1.1.1 Epidemiología.....	27
1.1.2 Factores etiológicos y de riesgo	28
1.1.3 Desarrollo y progresión tumoral.....	30
1.1.4 Clasificación histopatológica y estadiaje	31
1.1.5 Tratamiento	33
1.2 RADIOTERAPIA.....	34
1.2.1 Planificación del tratamiento con radioterapia externa	35
1.2.2 Dosis y fraccionamiento de la dosis	36
1.2.3 Mecanismos de acción de la radioterapia	37
1.2.4 Efectos adversos a la radioterapia	39
1.2.4.1 Efectos de la radiación en el tejido sano	39
1.2.4.2 Efectos adversos en pacientes con cáncer de próstata tratados con radioterapia	41
1.3 SUSCEPTIBILIDAD GENÉTICA A LA RADIACION IONIZANTE.....	42
1.3.1 Origen genético de la respuesta a la radiación ionizante.....	42
1.3.2 Radiogenómica.....	45
1.4 ESTUDIOS RADIOGENÓMICOS.....	46
1.4.1 Estudios de asociación gen-candidato.....	46
1.4.2 Estudios de asociación gen-candidato en cáncer de próstata	47
1.4.3 Estudios de asociación de genoma completo	47
1.4.4 Metaanálisis de GWAS	52
1.4.5 Machine learning y deep learning en radiogenómica	53
1.5 NUEVAS TECNOLOGÍAS Y APROXIMACIONES ÓMICAS.....	55
1.5.1 Secuenciación del exoma completo.....	56
1.5.1.1 Estudios basados en genes.....	56

1.5.2	<i>Estudios de asociación del transcriptoma completo</i>	57
1.5.3	<i>Estudios de asociación del proteoma completo</i>	58
2	JUSTIFICACIÓN Y OBJETIVOS	61
2.1	JUSTIFICACIÓN Y OBJETIVOS	63
3	MATERIALES Y MÉTODOS	65
3.1	CONSIDERACIONES ÉTICAS	67
3.2	CÁLCULO DEL PODER DEL ESTUDIO Y TAMAÑO DE LA MUESTRA	67
3.3	COHORTE DE PACIENTES	69
3.3.1	<i>Población de estudio</i>	69
3.3.1.1	Selección de pacientes	69
3.3.2	<i>Características de los pacientes incluidos</i>	70
3.4	PROCESAMIENTO DE LAS MUESTRAS Y SECUENCIACIÓN DE ADN (EXOMAS)	72
3.4.1	<i>Recogida de muestras</i>	72
3.4.2	<i>Extracción de ADN</i>	72
3.4.3	<i>Control de calidad del ADN</i>	72
3.4.3.1	Cuantificación del ADN	72
3.4.3.2	Integridad del ADN	72
3.4.4	<i>Secuenciación de ADN (exomas)</i>	73
3.4.4.1	Creación de librerías	73
3.4.4.2	Secuenciación	74
3.5	ANÁLISIS BIOINFORMÁTICO	74
3.5.1	<i>Control de calidad de los datos crudos FASTQ</i>	75
3.5.2	<i>Mapeo con el genoma de referencia</i>	75
3.5.3	<i>Recalibrado de las variantes y aplicación de umbrales de calidad</i>	75
3.5.4	<i>Marcaje de secuencias duplicadas y recalibración de la calidad de las bases leídas</i>	76
3.5.5	<i>Detección de variantes</i>	76
3.5.6	<i>Fusión de GVCFs y “joint-calling”</i>	77
3.5.7	<i>Anotación de variantes</i>	77
3.5.8	<i>Control de calidad general</i>	78

3.6	SELECCIÓN DE COVARIABLES	78
3.7	ANÁLISIS DE VARIANTES Y GENES PREVIAMENTE DESCRITOS.....	79
3.7.1	<i>Estudios de asociación univariable</i>	79
3.7.2	<i>Estudios de asociación multivariable.....</i>	80
3.7.3	<i>Validación de SNPs previamente asociados con sangrado rectal</i>	80
3.7.4	<i>Nuevas variantes en genes previamente asociados con toxicidad a la radioterapia.</i>	81
3.8	ESTUDIOS DE ASOCIACIÓN BASADOS EN GENES.....	81
3.8.1	<i>Análisis de asociación de genes para el fenotipo sangrado rectal y para el fenotipo de sangrado rectal extremo.....</i>	82
3.8.1.1	<i>Análisis de asociación de genes en RADIOGEN.....</i>	82
3.8.1.2	<i>Validación de genes en la cohorte externa REQUITE.....</i>	84
3.8.2	<i>Evaluación conjunta de variantes comunes y raras</i>	85
3.9	NUEVAS APROXIMACIONES DE ESTUDIOS DE ASOCIACIÓN ÓMICOS APLICADOS A RADIOGENÓMICA Y ANÁLISIS DE RUTAS.....	85
3.9.1	<i>Estudio de asociación del transcriptoma completo</i>	85
3.9.2	<i>Estudios de asociación del proteoma completo.....</i>	86
3.9.3	<i>Análisis de rutas</i>	86
4	RESULTADOS	89
4.1	CAPÍTULO 1. ESTUDIOS DE VARIANTES Y GENES PREVIAMENTE ASOCIADOS CON EFECTOS ADVERSOS A LA RADIOTERAPIA.	91
4.1.1	<i>Análisis de asociación univariable.....</i>	102
4.1.2	<i>Análisis de asociación multivariable.....</i>	102
4.1.3	<i>Validación del SNP rs4804134.....</i>	103
4.1.4	<i>Identificación de nuevas variantes en genes previamente asociados con efectos adversos a la radioterapia</i>	103
4.2	CAPITULO 2. ESTUDIOS DE ASOCIACIÓN BASADOS EN GENES.....	104
4.2.1	<i>Análisis de asociación del fenotipo sangrado rectal</i>	104
4.2.2	<i>Análisis de asociación del fenotipo sangrado rectal extremo</i>	112
4.2.3	<i>Análisis de asociación combinado de variantes comunes y raras</i>	120

4.3	CAPITULO 3. NUEVAS APROXIMACIONES DE ESTUDIOS DE ASOCIACIÓN ÓMICOS APLICADOS A RADIOGENÓMICA Y ANÁLISIS DE RUTAS.....	121
4.3.1	<i>Estudios de asociación del transcriptoma completo</i>	121
4.3.2	<i>Estudios de asociación del proteoma completo</i>	124
4.3.3	<i>Análisis de rutas</i>	125
5	DISCUSIÓN.....	129
6	PERSPECTIVAS FUTURAS.....	143
7	CONCLUSIONES.....	147
8	BIBLIOGRAFÍA.....	151
9	ANEXOS.....	175
	ANEXO 1. CLASIFICACIÓN TNM.....	177
	ANEXO 2. ESTADIOS DEL CÁNCER DE PRÓSTATA.....	179
	ANEXO 3. TRATAMIENTO DEL CÁNCER DE PRÓSTATA LOCALIZADO	181
	ANEXO 4. APROBACIÓN POR EL COMITÉ DE ÉTICA.....	183
	ANEXO 5. ESCALAS DE TOXICIDAD DE ACUERDO CON LOS CTCAE v.4	185
	ANEXO 6. VARIANTES GENÉTICAS IDENTIFICADAS EN GENES PREVIAMENTE ASOCIADOS CON EFECTOS ADVERSOS A LA RADIOTERAPIA.....	189
	ANEXO 7. VARIANTES GENÉTICAS IDENTIFICADAS EN GENES PREVIAMENTE ASOCIADOS CON EFECTOS ADVERSOS A LA RADIOTERAPIA ANALIZADAS CON X2.....	193
	ANEXO 8. TEJIDOS ANALIZADOS EN TWAS.....	195
	ANEXO 9. GENES INCLUIDOS EN EL ANÁLISIS DE RUTAS.....	197
	ANEXO 10. RUTAS ANALIZADAS	199

1 INTRODUCCIÓN

1.1 CÁNCER DE PRÓSTATA

El cáncer de próstata es una neoplasia maligna que afecta a millones de hombres alrededor del mundo. El pronóstico del paciente con este tipo de cáncer depende del estadio clínico, el grado Gleason de la biopsia prostática y de los niveles de antígeno prostático específico (PSA) pretratamiento, entre otros factores. Los pacientes con enfermedad localizada y con riesgo de recurrencia de bajo a intermedio presentan una tasa de supervivencia cercana al 99 % a los diez años¹, siempre y cuando la detección y el tratamiento se realicen en una etapa temprana de la enfermedad. A pesar del buen pronóstico que existe para este tipo de cáncer, se estima que alrededor del 5 % de los pacientes con cáncer de próstata desarrollarán metástasis.

1.1.1 Epidemiología

El cáncer de próstata es el segundo tipo de cáncer más frecuente en hombres a nivel mundial, después del cáncer de pulmón. En el año 2020 la Agencia Internacional para la Investigación del Cáncer (IARC por sus siglas en inglés) reportó 1,4 millones de nuevos casos de cáncer de próstata, con una tasa de mortalidad de 6,8 %. En Europa este tipo de cáncer es la primera causa de cáncer en hombres, con un número estimado de 473 344 nuevos casos (**Figura 1a**) y 108 088 fallecimientos (**Figura 1b**) en 2020². .

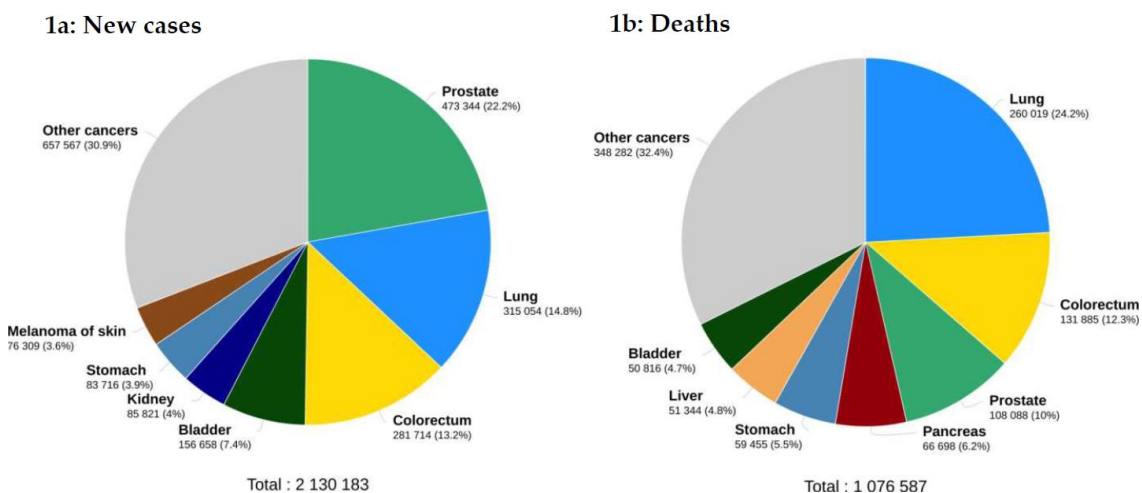


Figura 1. Incidencia (1a) y mortalidad (1b) de cáncer de próstata en Europa en 2020. Permitido por: International Agency for research on Cancer³. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Disponible en: <https://gco.iarc.fr/today>, accessed [21 Mar 2022]³.

De acuerdo con los datos reportados por la IARC se estimaron 34 613 nuevos casos de cáncer de próstata en España en 2020 y 5798 muertes. Sin embargo, debemos considerar que, tanto en España como en el resto del mundo, la pandemia de Covid-19 ha tenido un impacto muy importante sobre los datos de incidencia de cáncer. Se estima que el número de nuevos casos de cáncer registrados en 2020 fue menor a lo esperado, debido a la afectación que tuvieron los programas de cribados, la presión ejercida en el sistema

sanitario que pudo influir en la capacidad diagnóstica del cáncer, aunado al miedo y a las restricciones de acceso a los centros sanitarios. Considerando lo anterior, la Red Española de Registro del Cáncer (REDECAN), destaca que las incidencias de cáncer reportadas para el 2022 se estimaron suponiendo que no hubiera factores que alteraran la información recibida, es decir que no hubiese ocurrido la pandemia. Es así como, para el año 2022 la REDECAN estimó 30 884 nuevos casos de cáncer de próstata⁴.

1.1.2 Factores etiológicos y de riesgo

La etiología del cáncer de próstata es multifactorial, compleja y permanece sin ser bien conocida, sin embargo, existen diversos factores que están relacionados con la predisposición a desarrollar este tipo de enfermedad. Los factores de riesgo se han clasificado en modificables y no modificables. Entre los factores no modificables se encuentra la edad, la ancestralidad y los factores genéticos heredables⁵.

La edad es uno de los factores de riesgo más importante en el desarrollo de cáncer de próstata. Así, más del 85 % de los individuos diagnosticados son mayores de 60 años. Debido a la elevada esperanza de vida en países desarrollados, la incidencia de este cáncer es más elevada en estos países¹.

La ancestralidad también ha sido reportada como factor de riesgo. Diversos estudios relacionan un riesgo más elevado de cáncer de próstata en varones de ancestralidad afroamericana en comparación con individuos de ancestralidad europea⁶. Una comparación interesante es la realizada entre dos grupos de hombres de ascendencia asiática. Los hombres que vivían en Asia presentaron menor riesgo de cáncer de próstata en comparación con aquellos que vivían en Estados Unidos⁷. Estos datos respaldan la hipótesis de la influencia de los factores ambientales en la incidencia de cáncer de próstata.

La herencia juega un papel muy importante en la predisposición al cáncer; se estima que 9 % de los casos de cáncer de próstata son de tipo hereditario¹. Se considera cáncer de próstata hereditario cuando tres o más hombres de la misma familia están afectados, y al menos dos de ellos desarrollaron cáncer de próstata a temprana edad (menores de 55 años)¹. De forma general, se ha estimado que el riesgo relativo (RR) de un varón a desarrollar cáncer de próstata se ve incrementado 1,8 veces cuando hay un familiar de primer grado de consanguinidad afecto con este cáncer⁶. En pacientes con el padre y el hermano afectos, el riesgo aumenta a 5,5 veces⁶. Si hay dos hermanos afectos el riesgo de cáncer de próstata aumenta a 7,7⁶ y cuando son tres hermanos afectos se ha estimado un incremento a 17,7⁸.

Se han descrito variantes genéticas asociadas al cáncer en genes que regulan el crecimiento, la proliferación, la muerte celular y que participan en los mecanismos de reparación del ADN¹. En cáncer de próstata, variantes patogénicas y probablemente patogénicas germinales en los genes de reparación del ADN: *ATM*, *BRCA1*, *BRCA2*, *BRIP1*, *CHEK2*, *FANCA*, *MLH1*, *MSH2*, *MSH6*, *NBN*, *PALB2*, *PMS2*, *RAD51C* y *RAD51D*, se han relacionado con mayor riesgo a desarrollar cáncer, así como con el inicio temprano de la enfermedad y la evolución del tumor a una forma más agresiva^{9,10}. Además, en pacientes con cáncer de próstata metastásico se ha reportado una incidencia aumentada (11,8 – 16 %)^{9,11} de este tipo de variantes. Se ha descrito que variantes deletéreas germinales del gen *BRCA2* se asocian con el riesgo de cáncer de próstata más agresivo (4 %; *Odds ratio* (OR) = 4,29; 95% intervalo de confianza (CI) [3,15 - 5,86]; p-valor = $4,0 \times 10^{-20}$) y metastásico (3,0 %; OR = 5,61; 95% CI [3,73 - 8,44]; p-valor = $1,3 \times 10^{-16}$)¹⁰. Varios estudios han mostrado que el alelo de moderada penetrancia *G84E* rs138213197, localizado en el gen codificador de factores de transcripción *HOXB13*¹²⁻¹⁴, se asocia con el aumento de riesgo de cáncer. Además de estos hallazgos, otros estudios han evidenciado el origen fundador de la variante c.349A>G; p.(Arg117Gly) del gen *CHEK2* y su asociación con el incremento de riesgo a desarrollar cáncer de próstata (OR = 1,9; 95% CI [1,1 - 3,2])¹⁵ (**Figura 2**).

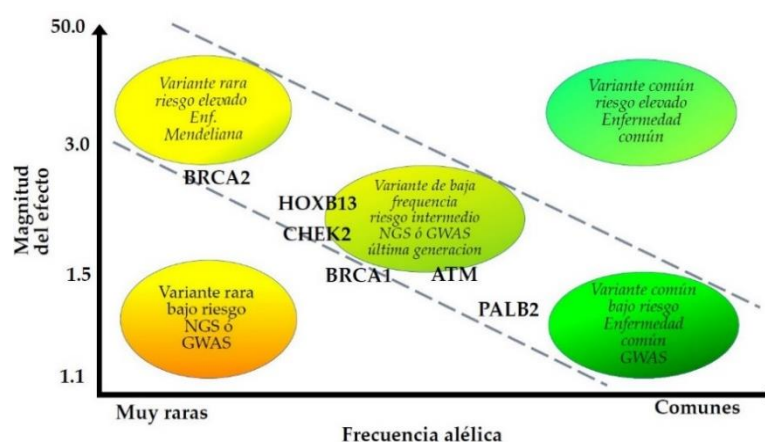


Figura 2. Frecuencia alélica y magnitud del efecto (penetrancia) de variantes genéticas asociadas a cáncer de próstata.

Recientemente se publicaron los resultados de un metaanálisis multiancestral de estudios de asociación del genoma completo (GWAS), que incluyó 107 247 casos y 127 006 controles, provenientes de poblaciones europeas, africanas, asiáticas orientales e hispanas. Este estudio tuvo como objetivo descubrir variantes genéticas relacionadas con cáncer de próstata. Se identificaron 269 variantes asociadas con formas más agresivas de cáncer de próstata, de las cuales 89 no habían sido descritas previamente. Además, se encontraron diferencias en los puntajes de riesgo genético (GRS, *genetic risk score*) de cáncer de próstata para las diferentes ancestralidades. Los hombres de

ascendencia africana presentaron 2,18 veces más riesgo genético que los de origen europeo, mientras que los hombres de ascendencia del este asiático obtuvieron un GRS 0,73 veces menor que los hombres de ascendencia europea. Estos resultados indican que las variantes germinales marcan la diferencia en el riesgo a desarrollar cáncer entre las diferentes poblaciones¹⁶.

Entre los factores de riesgo modificables se ha relacionado a la dieta alta en ácidos grasos saturados y ácidos grasos trans¹⁷. Por otro lado, diversos estudios han asociado la obesidad con un peor pronóstico en pacientes con cáncer de próstata^{18,19,20}. En contraste, factores como hacer ejercicio, el consumo de antioxidantes, ácidos grasos Omega-3 y de fuentes de licopeno se han relacionado con efectos protectores^{17,21}.

1.1.3 Desarrollo y progresión tumoral

La mayoría de los cánceres de próstata son adenocarcinomas provenientes de células epiteliales lumbinales secretoras y células basales^{1,22,23}. Hasta hace algunos años se creía que las células de origen del cáncer de próstata son las células lumbinales, debido a que la enfermedad presenta expansión de éstas y ausencia de células basales²⁴. Sin embargo, en un estudio realizado en un modelo *in vivo* con ratones, Andrew Goldstein et al. (2010)²⁴ demostraron que las células basales del tejido de próstata benigno, pueden dar origen a esta neoplasia. Actualmente, el desarrollo y progresión del cáncer de próstata se asocia con la acumulación de alteraciones genéticas presentes en las células epiteliales de la próstata, que propician el crecimiento maligno. Estas alteraciones pueden incluir la presencia de aneuploidías, pérdida de heterocigosidad, hipermetilación, inactivación de genes supresores de tumores inducida por reordenamientos cromosómicos estructurales y, por último, la activación de oncogenes inducida por una mutación puntual o por reordenamientos estructurales que pueden dar lugar a la fusión de genes²⁵. Las alteraciones que se presentan de forma más frecuente en el cáncer de próstata se localizan en las regiones cromosómicas 8p21(*locus NKX3.1*), 8q24 (*locus C-MYC*), 10q23 (*locus PTEN*), 17q (*locus TP53*), y en el gen del receptor de andrógenos (*AR*)²⁶.

El proceso de desarrollo de células normales a cancerígenas conlleva una serie de cambios histológicos y moleculares (**Figura 3**). En la transformación de una célula normal a una lesión premaligna PIN (*prostatic intraepithelial neoplasia*), se presentan alteraciones que incluyen: 1) deleciones, que conllevan a la pérdida de función de *PTEN*, pérdida de heterocigosidad asociada con desdiferenciación del tejido y pérdida de la capacidad de respuesta a andrógenos en *NKX3.1*²⁷; 2) sobreexpresión en *GSTP1* y *MYC*. La combinación de la pérdida de función de *PTEN* y la sobreexpresión de *MYC* favorece el desarrollo temprano de la neoplasia intraepitelial prostática y del adenocarcinoma^{28,29}, 3) reordenamientos de genes de la familia de eritroblastos específicos de transformación (ETS), que involucran fusiones de genes, por ejemplo, *TMPRSS2:ERG* y 4) mutaciones en el gen *SPOP*³⁰. Previamente, se ha asociado a lesión

premaligna como una lesión precursora del cáncer de próstata³¹. El avance de una lesión premaligna hacia el desarrollo de adenocarcinoma se ha visto influenciado por mutaciones en *FOXA1* y la pérdida de función de *RB1*. Ambas alteraciones también se presentan frecuentemente en el cáncer de próstata resistente a la castración. Puede producirse también la pérdida de función de *PTEN* y la fusión de *TMPRSS2:ERG*³². La progresión a metástasis puede ser conducida por: 1) mutaciones en los genes *FOXA1*, *BRCA1*, *BRCA2*, *ATM*, *ATR*, *RAD51* y en *AR*; 2) ganancia de función en *CXCL12*, *CXCR4*, *RANK-RANKL*, *EZH2*, *MYC* y el coactivador *AR*³³ y 3) pérdida de función de *RB1*, *PTEN* y *SMAD4*.

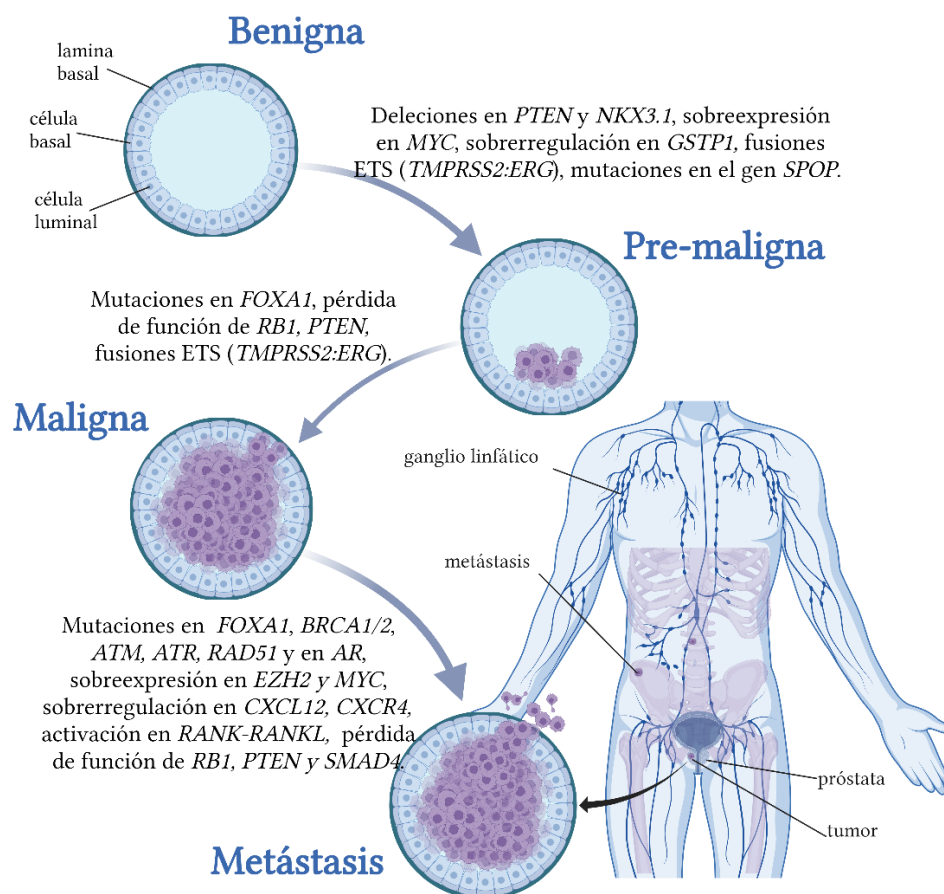


Figura 3. Desarrollo y progresión del cáncer de próstata. Se muestran las modificaciones celulares y los genes implicados en el proceso de desarrollo de la malignidad. Created with BioRender.com

1.1.4 Clasificación histopatológica y estadiaje

El adenocarcinoma es el tipo más frecuente de cáncer de próstata primario, aproximadamente el 95 % de los casos son de este tipo, mientras que el 5 % restante es de carácter heterogéneo.

El grado histológico describe el grado de anormalidad de células y tejidos malignos al ser observados al microscopio y la velocidad de crecimiento y diseminación del cáncer.

La clasificación más usada para determinar el grado histológico es la de Gleason. Este grado identifica los dos patrones glandulares más frecuentes, que se definen como patrones primario y secundario, a cada uno de estos les asigna un grado de diferenciación que va de 1 a 5, donde uno es el grado más diferenciado. La suma de ambos valores da como resultado el grado histológico al que correspondería en Gleason. Posteriormente, en 2014 la Sociedad Internacional de Urología Patológica (ISUP, *International Society of Urological Pathology*) aprobó un sistema de gradación pronóstico. Este sistema comprende grados que van de uno (para tejidos bien diferenciados, de bajo riesgo) a cinco (tejidos poco diferenciados, de alto riesgo) (**Figura 4**). La clasificación tradicional de Gleason y la clasificación ISUP son empleadas en paralelo^{6,34,35}.

		Grado ISUP	Sistema Gleason	
Bien diferenciadas		1	Gleason score ≤ 6	Grado 1. Glándulas pequeñas y uniformes (bien definidas). Glándulas únicas, redondeadas u ovaladas, tamaño medio, separadas y de distribución compacta.
				
Moderadamente diferenciadas		2	Gleason score $3 + 4 = 7$	Grado 2. Más estroma entre glándulas. Nódulos menos definidos y circunscritos. Glándulas únicas, separadas, redondas u ovaladas con variación en tamaño, ligeramente compactadas con separación estromal.
				
Poco diferenciadas		3	Gleason score $4 + 3 = 7$	Grado 3. Bordes claramente infiltrantes, definidos, extensión irregular. Glándulas únicas, separadas, forma y tamaño pequeño, con forma alargada, girada y angular generalmente con una separación estromal amplia.
				
		4	Gleason score $4 + 4 = 8$ $3 + 5 = 8$ $5 + 3 = 8$	Grado 4. Masas irregulares de glándulas neoplásicas. Infiltración irregular. Glándulas fusionadas crean masas, cordones o cadenas, grandes, tamaño irregular, cribiforme.
				
		5	Gleason score $4 + 5 = 9$ $5 + 4 = 9$ $y 5 + 5 = 10$	Grado 5. Sólo forma glándulas ocasionalmente forma poco diferenciada. Comedocarcinoma - Masa papilar, cribiforme con necrosis central.

Figura 4. Clasificación de Gleason y su equivalente en grado ISUP (International Society for Urological Pathology). Adaptado de Kim et al., ³⁶, permitido por: Cancers. La descripción de los grados basada en Humphrey et al. 2004³⁷

Además del grado histológico de Gleason es importante determinar el estadio y la extensión tumoral, para estos últimos se usa el sistema de estadificación TNM de la AJCC (*American Joint Committee on Cancer*) y la UICC (*Union for International Cancer Control*)³⁸. TNM clasifica basándose en el tamaño del tumor (T), en la afectación de los ganglios linfáticos (N) y en la diseminación a otros tejidos o metástasis (M) (Anexo 1).

De acuerdo con la AJCC 8ª edición³⁸ el estadio del cáncer de próstata, se asigna según el sistema de clasificación combinado TNM, el nivel de PSA y el grado basado en el puntaje Gleason (Anexo 2). En base a estos estadios el cáncer de próstata puede clasificarse en: 1) cáncer de próstata localizado, que abarca los estadios I, IIA, IIB, IIC y IIIA; 2) cáncer de próstata localmente avanzado, que incluye a los estadios IIIB y IIIC y 3) cáncer de próstata diseminado o metástasis, que incluye a los estadios IVA y IVB.

Clasificación de pacientes con cáncer de próstata localizado en grupos de riesgo

La estratificación de pacientes en grupos de riesgo se lleva a cabo considerando el nivel de PSA, el Gleason y el estadio tumoral. Entre las diferentes clasificaciones que existen para predecir el riesgo la clasificación de D'Amico³⁹ es una de las más utilizadas (**Tabla 1**)

Tabla 1. Clasificación D'Amico

Grupos de riesgo de cáncer de próstata localizado Clasificación D'Amico	
Bajo riesgo	T1-T2a, Gleason ≤ 6 y PSA ≤ 10
Riesgo intermedio	T2b y/o Gleason = 7 y/o PSA 10-20
Alto riesgo	$\geq$ T2c o Gleason > 7 o PSA > 20

PSA, antígeno prostático específico; T, tamaño del tumor de acuerdo con la clasificación TNM.

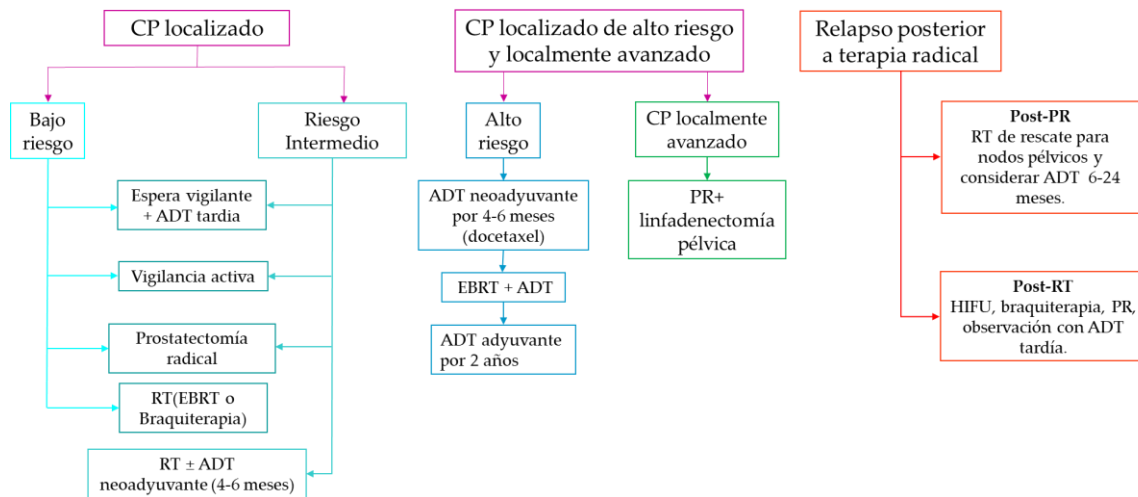
Extraída de Parker et al. (2016)⁴⁰, permitido por: *Creative Commons Attribution License*

Además de la clasificación D'Amico, existen los nomogramas de Kattan, que contribuyen a estimar el riesgo pre y post-prostatectomía radical, a valorar los resultados del tratamiento y a evaluar el riesgo de fallo bioquímico⁴¹. El nomograma preoperatorio de Kattan incorpora variables clínicas como el nivel de PSA en suero antes del tratamiento, el puntaje en la escala de Gleason y el estadio clínico del tumor⁴². Por otra parte, para el nomograma postoperatorio se incluyen, además del PSA y del puntaje Gleason, la invasión capsular prostática, el margen quirúrgico (margen del tejido aparentemente no tumoral que rodea al tumor), la invasión de vesículas seminales y de ganglios linfáticos⁴¹.

1.1.5 Tratamiento

El tratamiento ofrecido al paciente debe considerar las características clínicas, las características moleculares, el estadio tumoral y el riesgo del paciente. Según el algoritmo de la Sociedad Europea de Oncología Médica (ESMO) (**Figura 5**)⁴³, el tratamiento para el cáncer se divide en: tratamiento en estadio localizado con riesgo bajo e intermedio, tratamiento para estadio localizado de alto riesgo y estadio localmente avanzado, tratamiento para pacientes con recidiva posterior a radioterapia (RT) o prostatectomía radical y tratamiento para estadios metastásicos. De acuerdo a cada estadio el tratamiento del cáncer de próstata puede incluir opciones como la espera vigilante, la vigilancia activa, la prostatectomía radical, la radioterapia o braquiterapia, el ultrasonido de alta intensidad y la hormonoterapia.

En el Anexo 3 se presenta información detallada sobre el tratamiento del cáncer de próstata localizado, de acuerdo con las guías de la *European Association of Urology* (EAU)-*European Association of Nuclear Medicine* (EANM)-*European Society for Radiotherapy and Oncology* (ESTRO)-*European Society of Urogenital Radiology* (ESUR)-*International Society of Geriatric Oncology* (SIOG) del 2020⁶.



CP= Cáncer de próstata, ADT(Androgen Deprivation Therapy) = Terapia de privación de andrógenos, RT= Radioterapia, EBRT(External Beam Radiation Therapy) = Radioterapia de haz externo, PR = Prostatectomía radical, HIFU (High Intensity Focused Ultrasound) = Ultrasonido focalizado de alta intensidad

Figura 5. Tratamiento del cáncer de próstata para estadio localizado y localmente avanzado y pacientes con recidiva. Adaptado de Parker et al. (2020)⁴³, permitido por: Creative Commons Attribution License

1.2 RADIOTERAPIA

La radioterapia se utiliza como tratamiento curativo o paliativo en más del 50 % de los pacientes con cáncer. En función de la localización de la fuente de radiación, la radioterapia se puede clasificar en externa (EBRT, *External Beam Radiation Therapy*); interna, también conocida como braquiterapia y radioterapia sistémica (administración de radiofármacos, como el yodo radioactivo).

La EBRT es el tipo de radioterapia más utilizado a nivel global; siendo las partículas más utilizadas los fotones, también conocidos como rayos X. Los fotones son partículas de alta energía que no poseen carga ni masa, lo cual permite que penetren en el cuerpo a mayor profundidad, respetando la piel.

Algunos tipos de radioterapia externa son:

- ❖ Radioterapia conformada en tres dimensiones (3D-CRT, *Three-dimensional Conformal Radiation Therapy*): se basa en la generación de imágenes tridimensionales del tumor, que permiten el ajuste de los haces de radiación hacia el tumor, evitando en la medida de lo posible la exposición de los tejidos sanos circundantes.

- ❖ Radioterapia de intensidad modulada (IMRT, *Intensity modulated radiotherapy*): Es un tipo de RT 3D de rayos X avanzada que modula la intensidad de la radiación mediante haces de intensidad variable. Esto permite administrar y

dirigir dosis más precisas al tumor, disminuyendo la irradiación al tejido normal adyacente.

- ❖ Radioterapia guiada por imagen (IGRT, *Image guided radiotherapy*): en esta RT se lleva a cabo la toma de imágenes del tumor y de la anatomía circundante antes y durante el tratamiento. Esto permite reducir los errores debidos al movimiento de los órganos y la administración de la radiación al tumor de forma más precisa.
- ❖ MRI-linac (*Magnetic Resonance Imaging Guided Linear Accelerator*): combina la resonancia magnética por imagen y la administración simultánea de radioterapia. La resonancia magnética al poseer contraste elevado de tejidos blandos es particularmente útil en algunas áreas anatómicas, como el cerebro y la pelvis.
- ❖ SBRT (*Stereotactic Body Radiotherapy*): consiste en administrar altas dosis de radiación a los tumores de forma más precisa con un daño mínimo a los tejidos sanos circundantes y en menor número de fracciones

Por otra parte, en la braquiterapia, la fuente radiactiva se inserta en el cuerpo, y esto permite una distribución y entrega de dosis más favorable, evitando la exposición de tejidos sanos a la radiación⁴⁴.

1.2.1 Planificación del tratamiento con radioterapia externa

La planificación del tratamiento con radioterapia implica una serie de requerimientos necesarios para suministrar al tejido diana la dosis de radioterapia adecuada. La posición exacta del paciente durante el tratamiento es muy importante. Para ello se pueden realizar tatuajes en la piel o marcadores fiduciales de oro, en conjunto con inmovilizadores de piernas y pies, que ayudarán a una posición adecuada.

La adquisición de imágenes de la tomografía computarizada solas o fusionadas con resonancia magnética nuclear forman parte de la planificación del tratamiento 3D. La creación de un modelo tridimensional a partir de estas imágenes permite delinear el tumor y los órganos de riesgo, que en el caso de RT en la próstata incluye el recto, la vejiga, el ano, el intestino delgado y el bulbo peneano y las cabezas femorales. También se realiza la delimitación de los volúmenes blanco (**Figura 6**), que son:

- ❖ Volumen tumoral macroscópico (GTV, *Gross tumor volume*), contiene la localización y extensión del tumor primario y de sitios metastásicos.

- ❖ Volumen blanco clínico (CTV, Clinical target volume) incluye el GTV más la enfermedad microscópica. Para cáncer de bajo riesgo debe incluir la próstata ± vesículas seminales y para riesgo intermedio y alto incluye la próstata + 5 mm y las vesículas seminales.
- ❖ Volumen blanco de planificación (PTV, Planning target volume). Incluye el CTV y se establecen márgenes en función de los posibles errores de posicionamiento y de los movimientos intrafracción e interfacción de la próstata (aproximadamente 1 cm).

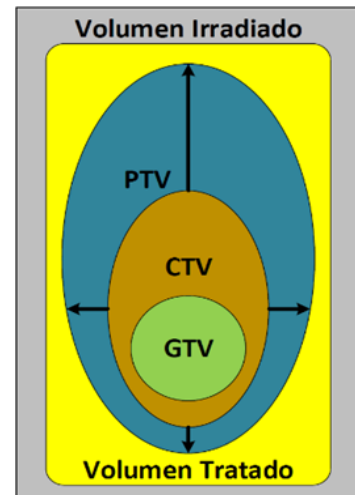


Figura 6. Definición de volúmenes

Un paso importante en el proceso de planificación es la generación de histogramas dosis-volumen, los cuales son utilizados para la evaluación de la dosis de radiación que recibirá tanto el tumor como los tejidos circundantes. Por último, se realiza la simulación-verificación, donde se corrobora el plan de tratamiento sobre una misma tomografía realizada sobre el paciente justo antes de administrar el tratamiento, si dicha simulación es adecuada, esos parámetros se usarán como referencia en todo el tratamiento.

1.2.2 Dosis y fraccionamiento de la dosis

La dosis en radioterapia se mide en Grays (Gy). El Gray es la unidad de medida de la radiación absorbida, que equivale a la absorción de energía de radiación en un *Joule* por kilogramo de masa. La dosis total administrada dependerá del tipo y estadio tumoral, de la intención de tratamiento, así como de la técnica de RT utilizada. La administración fraccionada de la dosis busca aumentar la eficacia sobre el tumor y disminuir los efectos adversos tardíos, manteniendo en niveles similares la toxicidad aguda de los tejidos normales incluidos dentro del volumen irradiado. Existen diferentes pautas de fraccionamiento:

Fraccionamiento convencional: el número de fracciones se calcula dividiendo la dosis total en Gy entre la dosis de la fracción, obteniendo así el número de sesiones requeridas para la administración completa del tratamiento. Por ejemplo, en el caso del tratamiento de la próstata, podían ser administrados > 75 Gy para bajo riesgo y > 78 Gy para riesgo intermedio y alto. Así, el esquema de fraccionamiento implica administrar dosis diarias de 180 - 200 cGy durante cinco días a la semana, hasta alcanzar la dosis total en aproximadamente ocho semanas⁴⁵.

Hiperfraccionamiento: consiste en el aumento del número de fracciones con dosis reducidas por fracción para alcanzar una dosis total mayor en un periodo de

tratamiento convencional. Se la administran al menos dos dosis al día de 115 – 120 cGy, separadas por al menos seis horas, cinco días a la semana.

Hipofraccionamiento y ultrahipofraccionamiento: estos tipos de fraccionamientos implican el aumento de la dosis por fracción y disminución del número de fracciones y de días de tratamiento⁴⁶. Los esquemas de radioterapia varían entre cada centro según la tecnología usada y como se implementen los hipofraccionamientos. Por ejemplo, un esquema de tratamiento con hipofraccionamiento moderado puede consistir en una dosis de 60 Gy en 20 fracciones de 300 cGy o 70 Gy divididas en 28 fracciones de 250 cGy⁴⁷. En el caso de un ultrahipofraccionamiento, la dosis podría ser de 42,7 Gy en siete fracciones de 610 cGy o de 36,25 – 40 Gy divididas cinco fracciones de 725 – 800 cGy⁴⁸. En resumen, los tamaños recomendados por fracción de acuerdo a cada tipo de fraccionamiento son de 250 a 340 cGy para hipofraccionamiento⁶ y de 7,25 - 8Gy en cinco fracciones o 9,5 Gy en cuatro fracciones para ultrahipofraccionamiento.

1.2.3 Mecanismos de acción de la radioterapia

La radioterapia ocasiona daños al ADN, que conducen a la muerte celular a través de diferentes mecanismos, que pueden ocurrir por vía directa o indirecta (**Figura 7**).

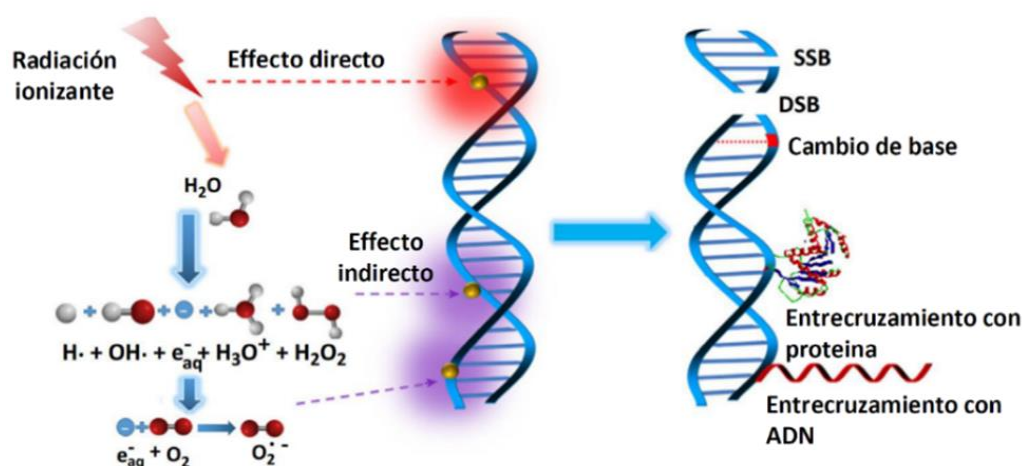


Figura 7. Daño al ADN causado por la radiación. Se observan los efectos directos e indirectos sobre el ADN. Adaptado de Wang et al. (2017)⁴⁹, permitido por: Creative Commons Attribution License.

En la vía directa, la radiación daña biomoléculas, como proteínas y lípidos, además impacta sobre el ADN y provoca alteraciones en su estructura ocasionando daños con roturas de puentes de hidrógeno, roturas de cadena simple (SSB) o doble (DSB)⁴⁹. Esto interrumpe tanto el ciclo como la proliferación celular y origina la muerte celular.

Por otro lado, en la vía indirecta, la radiación ionizante escinde moléculas de agua generando radicales libres, en especial especies reactivas de oxígeno (ROS), cuyo número dependerá de la dosis total de la radiación. Al tener electrones desapareados,

las ROS pueden dañar biomoléculas a través de reacciones químicas como la extracción, adición, desproporción y captura de electrones⁵⁰. El mecanismo indirecto representa aproximadamente el 65 % del daño biológico causado por la radiación. Este es el mecanismo central de la hipótesis de la fijación de oxígeno, que explica la diferencia en la sensibilidad a la radiación entre células aeróbicas o normóxicas e hipóxicas. Se ha descrito que tumores con baja pO₂ (hipóxicos) son generalmente más radiorresistentes y requieren dosis más altas de radiación para el tratamiento, que los tumores bien oxigenados normóxicos⁴⁹.

Otro mecanismo indirecto producido por la radioterapia se da a través del inicio de procesos inflamatorios⁵¹. La reacción proinflamatoria desencadena la liberación de citocinas (IL-1 α y β , IL-6, TNF α , TGF- β), quimiocinas (CXCR-4, CXCL12, CXCL10, SDF-1) y activa las vías de señalización para aumentar la apoptosis celular⁵⁰, lo que a su vez estimula una respuesta inmunológica.

La respuesta inmunológica al daño puede ser de tipo innata, que es la respuesta inmediata al daño, o de tipo adaptativo, que reconoce al agente extraño, lo destruye y crea una memoria inmunológica. Estímulos como la radioterapia y la quimioterapia inducen un tipo peculiar de muerte apoptótica denominado muerte celular inmunogénica (ICD, *immunogenic cell death*)⁵². En ella, las células en proceso de muerte pueden desencadenar respuestas inmunes antitumorales⁵⁰ y se liberan patrones moleculares asociados al daño (DAMP, *damage-associated molecular patterns*)⁵³. Algunos DAMP son; la calreticulina, el ATP extracelular y HMGB1 (*high mobility group box 1*)⁵². Estos DAMPs activan y maduran a las células dendríticas^{50,54}, que posteriormente migran a ganglio linfático en donde ejercen su función como células presentadoras del antígeno. Como resultado se produce la expansión clonal de los linfocitos T citotóxicos que migran hacia el tumor e inician la respuesta contra las células tumorales^{54,55}.

Entre los procesos inmunológicos de respuesta al daño del ADN, destaca la activación de las vías de señalización innata por medio de sensores de ADN. Un ejemplo es el sensor cGAS (*cyclic GMP-AMP synthase*), que activa a STING (*stimulator of interferon genes*), desencadenando el reclutamiento de células dendríticas y preparando a los linfocitos T CD8⁺ antitumorales. Esto resulta en el efecto abscopal, que ocurre cuando se trata un tumor con radioterapia localizada y el efecto de la radiación se da tanto en el área tratada, como en tumores en áreas a distancia que no fueron tratadas⁵¹.

Activación de puntos de control del ciclo celular en respuesta a la radiación

La radiación induce la activación de múltiples cascadas de señalización en respuesta al daño en el ADN, las cuales regulan la detención del ciclo celular, la reparación del daño, así como el destino y la muerte celular.

La fase del ciclo celular en el que se encuentren las células al momento de la irradiación influye en el tipo de muerte celular y en la radiosensibilidad. Las células en fase S serán las más radiorresistentes y las que están en la fase M son las más radiosensibles^{56,57}. Durante el ciclo celular, el daño al ADN inducido por la radiación desencadena la

activación de los puntos de control (*checkpoint*): G1/S y G2/M, controlado por complejos de Ciclina-CDK. La activación de los puntos de control S y G2 requiere de las vías ATR-Chk1 y Wee1, mientras que G1 requiere de p53 (**Figura 8**)⁵⁷.

En el punto de control G1, la radiación activa a p53, lo que induce la transcripción de CDK p21 inhibiendo la actividad de la Ciclina/CDK necesaria para avanzar a la fase S⁵⁷.

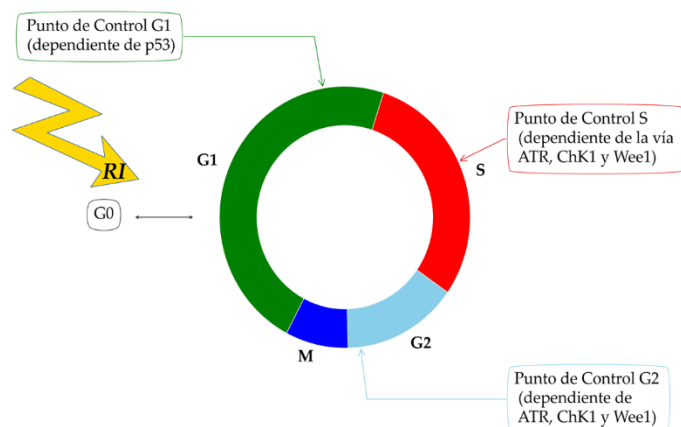


Figura 8. Activación de puntos de control en el ciclo celular como respuesta a la irradiación

Sin embargo, como p53 ha perdido su función normal en la mayoría de las células tumorales, este punto de control se ve afectado, permitiendo que las células dañadas entren en mitosis, resultando en la catástrofe mitótica⁵⁰. Además, mutaciones disruptivas en p53 se asocian con radorresistencia a través de la inhibición de la senescencia. En la fase S se produce una reducción en la replicación del ADN, lo que ralentiza el progreso de las células irradiadas a través de dicha fase.

En respuesta al daño, el ciclo celular es detenido para reparar las roturas de doble cadena, por unión de extremos no homólogos (NHEJ) o por recombinación homóloga (HR). Cuando no se logra reparar el daño, se produce la muerte celular, por apoptosis, catástrofe mitótica, necrosis, senescencia, autofagia, necroptosis o ferroptosis^{58,50}. Es importante destacar que alteraciones en genes como *PLK1*, *ATM* y *CHK1*, implicados en G2/M, alteran la respuesta al daño del ADN inducido por irradiación⁵⁹.

1.2.4 Efectos adversos a la radioterapia

1.2.4.1 Efectos de la radiación en el tejido sano

La radiación ionizante administrada durante la radioterapia, puede alcanzar también al tejido sano y desencadenar reacciones que pueden resultar en la muerte celular, así como el efecto *Bystander* de amplificación del daño e inflamación⁴⁴. La fisiopatología de la toxicidad está determinada principalmente por la dosis recibida, aunque existen diferencias según el tejido irradiado y el estado fisiológico de los órganos en riesgo. El tiempo en el que puede manifestarse el daño causado por la radiación varía dependiendo de cada tejido, presentándose de forma aguda o tardía (**Figura 9**).

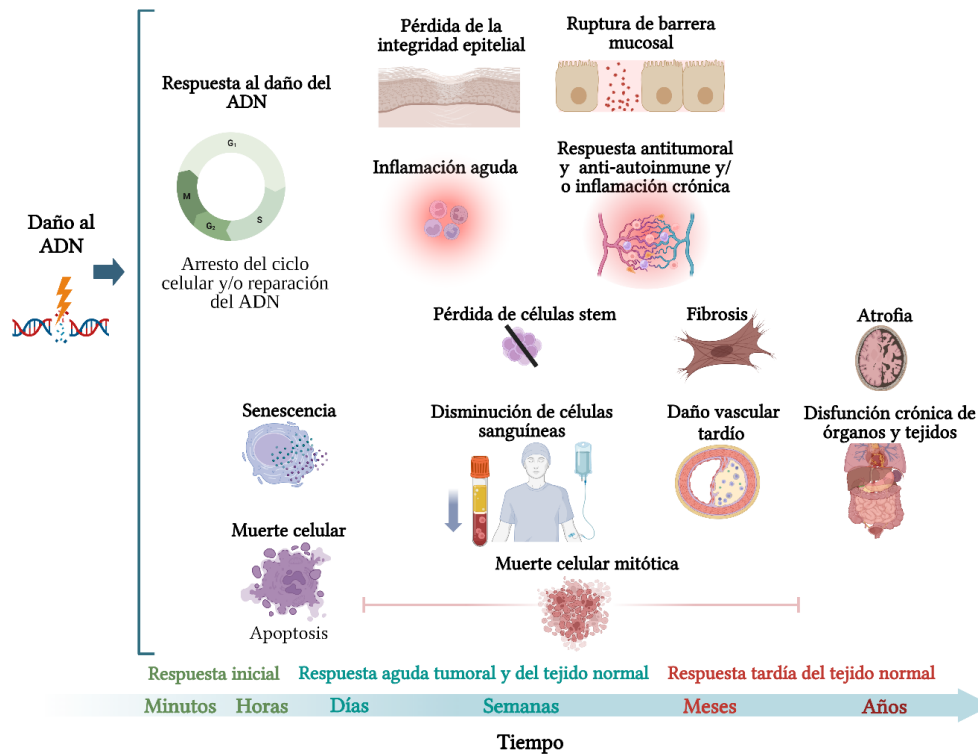


Figura 9. Representación de la respuesta a la radioterapia. Created with BioRender.com

Los efectos adversos agudos pueden presentarse tanto en el transcurso del tratamiento de radioterapia como en las semanas subsiguientes a este. Los tejidos mayoritariamente implicados en la respuesta aguda son aquellos con mayor tasa de proliferación y elevada capacidad de regeneración. Estos incluyen, por ejemplo, los tejidos hematopoyéticos, folículos pilosos, algunas capas del epitelio y la dermis. En respuesta a la irradiación, estos tejidos presentarán procesos apoptóticos mediados por p53 o asociados a una catástrofe mitótica. Los efectos adversos que se presentan en la etapa aguda incluyen, por ejemplo, inflamación (dermatitis, mucositis, proctitis, cistitis), pérdida de cabello y disminución de célula sanguíneas. Estos efectos son reversibles y generalmente los tejidos sanan en un periodo de semanas a meses.

Por otra parte, tejidos de regeneración lenta o estructuralmente importantes como los fibroblastos permanecerán en arresto celular inducido por p53. En comparación con los efectos agudos, los efectos tardíos generalmente son irreversible y progresivos y se presentan meses o años después del tratamiento radioterápico. Algunos efectos tardíos importantes incluyen: fibrosis tisular, daño vascular, deficiencia hormonal, disfunción orgánica, necrosis y riesgo de neoplasias secundarias inducidas por la radiación⁶⁰.

1.2.4.2 Efectos adversos en pacientes con cáncer de próstata tratados con radioterapia

Algunos pacientes tratados con radioterapia desarrollan efectos adversos. Entre éstos, los efectos leves y comunes se presentan entre el 10 y el 50 % de los pacientes tratados⁶¹. El desarrollo de efectos adversos dependerá principalmente de las dosis de radioterapia recibidas. En pacientes con cáncer de próstata, las toxicidades postradioterapia son principalmente de tipo gastrointestinal (GI), genitourinario (GU) y en órganos reproductivos. Se ha reportado que el 15,1 % de los pacientes tratados con 3D-CRT experimentan toxicidades GI y/o GU agudas de grado ≥ 2 , con una incidencia acumulada del 22,0 % a los tres años⁶². El 38 % de los pacientes presentan disfunción eréctil a los tres años postradioterapia⁶³.

Las manifestaciones clínicas de los efectos adversos a la irradiación pélvica en pacientes con cáncer de próstata incluyen:

- ❖ Toxicidad gastrointestinal: en la toxicidad aguda GI, los pacientes pueden desarrollar procesos inflamatorios como proctitis o enteritis, manifestándose con síntomas “rectales” como diarrea, tenesmo, dolor y sangrado^{64,65}. Al rededor del 5 al 30 % de los pacientes presentan toxicidad gastrointestinal aguda (la oscilación del porcentaje es dependiente de la dosis de radiación y del volumen de tratamiento)⁶⁴. Mientras que, en la toxicidad gastrointestinal tardía, las complicaciones incluyen diarrea persistente, tenesmo, urgencia rectal, sangrado rectal y dolor. Además, puede presentarse en menor grado úlceras, incontinencia anal (de grado variable), estenosis y formación de fístulas^{64,66}. Se ha reportado una incidencia a diez años de toxicidad gastrointestinal tardía de grado ≥ 2 del 13 %⁶⁷.

El sangrado rectal tardío es uno de los efectos adversos GI más importantes y comunes que presentan los pacientes con cáncer de próstata tratados con radioterapia. Se considera que el origen de la proctitis por radiación es la isquemia crónica de la mucosa causada por la fibrosis de tejidos y la inflamación grave de las arterias. Este daño en la pared del recto puede fácilmente originar sangrados conduciendo a un estado isquémico crónico y desencadenando episodios de sangrado rectal graves⁶⁸. En un ensayo llevado a cabo en población holandesa, el 28 % de los pacientes desarrollaron toxicidad rectal tardía postradioterapia y el 6 % de los pacientes presentaron hemorragia rectal grave^{67,69}.

- ❖ Toxicidad genitourinaria: la radioterapia puede ocasionar irritación y alteraciones funcionales en el epitelio y la mucosa de la vejiga. Los síntomas agudos son variados e incluyen disuria leve, aumento de la frecuencia urinaria, urgencia urinaria, hematuria, incontinencia urinaria, cistitis y uretritis⁶⁴. La incidencia de las toxicidades urinarias agudas de grado ≥ 2 oscila entre el 10 % y el 16 %⁶⁷. Los efectos tardíos GU abarcan la incontinencia urinaria, la estenosis uretral, la cistitis intersticial, la hematuria y el desarrollo de vejiga rádica⁶⁴. La

frecuencia de toxicidades genitourinarias tardías de grado ≥ 2 a 10 años es del 15 %⁶⁷.

- ❖ Toxicidad en órganos reproductivos (disfunción sexual): esta toxicidad incluye la pérdida de la libido, la disfunción eréctil, la pérdida de la función eyaculatoria y la infertilidad⁶⁴. Se ha escrito que alrededor del 39 % de los pacientes presentan impotencia postradioterapia⁷⁰.

Para evaluar la severidad de los eventos adversos presentados por los pacientes se han desarrollado diferentes escalas de gradación de toxicidad (**Tabla 2**). La escala CTCAE (*Common Terminology Criteria for Adverse Events*) clasifica por grados la severidad de los eventos adversos presentados por los pacientes tratados con terapias contra el cáncer en una escala de 1 a 5, siendo uno el grado más leve y el cinco la muerte relacionada con el evento adverso. Estos criterios definen como evento adverso a cualquier evento desfavorable, signo no intencionado, síntoma o enfermedad asociado temporalmente con el uso de un medicamento, tratamiento o procedimiento, que puede o no estar relacionado con el tratamiento o procedimiento médico. La última versión disponible es la 5.0 (CTCAE v5.0) del año 2017.

Tabla 2. Escalas de toxicidad

Escala	Descripción
WHO (1979) ⁷¹ <i>World Health Organization</i>	Recomendaciones para clasificar la toxicidad aguda
EORTC/RTOG (1984) ⁷² <i>European Organization for Research and Treatment of Cancer and the Radiation Therapy Oncology Group</i>	Escala de clasificación de toxicidad de los efectos secundarios tardíos
LENT/SOMA (1995) ⁷³ <i>Late Effects Normal Tissue/Subjective Objective Management Analytic.</i>	Sistema para clasificar los efectos secundarios después de la radioterapia
CTCAE (versión actual v.5.0 (2017)) ⁷⁴ <i>Common Terminology Criteria for Adverse Events</i> (previamente <i>National Cancer Institute Common Toxicity Criteria</i> NCI CTC 1982)	Escala de clasificación de severidad de efectos adversos en la terapia del cáncer, los ensayos clínicos y otros entornos oncológicos

1.3 SUSCEPTIBILIDAD GENÉTICA A LA RADIACION IONIZANTE

1.3.1 Origen genético de la respuesta a la radiación ionizante

Los efectos adversos desarrollados por los pacientes tratados con radioterapia se asocian con factores relacionados con el estado de salud del propio paciente (nivel de hemoglobina en sangre, comorbilidades como diabetes, hipertensión, etc.), las características del paciente (edad, tabaquismo, etc.), los tratamientos adicionales recibidos, el sitio de irradiación y la física del tratamiento radioterápico⁶⁰. Este último incluye la dosis total de radiación, dosis por fracción, el volumen del tejido u órgano

irradiado, el lugar de irradiación y la homogeneidad en la dosis. Sin embargo, todos estos factores explican solo un porcentaje del desarrollo de la toxicidad.

Se ha estimado que al rededor del 80 % de la variación de la respuesta a la radiación se debe a otro tipo de factores no conocidos, siendo estos probablemente de origen genético⁷⁵. Basándose en esto, diversos estudios se han enfocado en examinar la relación de la variación genética en la respuesta del tejido normal a la radiación.

Los primeros reportes de casos con respuesta severa a la radioterapia se presentaron en pacientes con ataxia telangiectasia (AT). Este síndrome es causado por mutaciones heredadas de forma autosómica recesiva en el gen *ATM*⁷⁶. Además de AT, se han reportado síndromes relacionados con la inestabilidad genómica como Li-Fraumeni. Esta enfermedad se origina por alteraciones en el gen *TP53*, que se transmite de forma autosómica dominante. Pacientes con Li-Fraumeni, presentan mayor probabilidad a desarrollar segundas neoplasias postratamiento⁷⁷. En estos pacientes la radioterapia debe evitarse o adaptarse mediante la reducción de los volúmenes irradiados considerando también los riesgos de toxicidad tisular, la radiosensibilidad, la radorresistencia y la radiocurabilidad⁷⁷.

Otros genes implicados en la radiosensibilidad identificados a partir de síndromes hereditarios son: *MRE11A* (mutado en individuos con trastornos similares a AT; asociado con radiosensibilidad clínica y celular; es un gen involucrado en la detección e inicio de la reparación de daños al ADN)⁷⁸; *LIG4* (mutado en individuos con síndrome de *LIG4*; asociado a radiosensibilidad clínica; participa en la reparación NHEJ); *NBN* (mutado en individuos con síndrome de rotura de Nijmegen; asociado con radiosensibilidad clínica y celular; el gen forma parte del complejo MRN involucrado en la detección e inicio de la respuesta al daño en el ADN); *RAD50* (mutado en individuos con trastorno similar al síndrome de rotura de Nijmegen; asociado con radiosensibilidad celular; forma parte del complejo MRN), *PRKDC* (codifica para ADN-PKcs; síndrome ADN-PK-RS-SCID; asociada con radiosensibilidad celular e involucrado en NHEJ).

Además de los genes relacionados con síndromes hereditarios, se han propuesto genes que, debido a su función y su participación en diversos mecanismos, se encuentran probablemente implicados en la variabilidad de respuesta a la radioterapia. Estos genes y mecanismos incluyen:

- ❖ Reparación del ADN: 1) En detección del daño; *MRE11A*, *RAD50*, *NBN*, *H2AFX*, *TP53BP1*, *BRCA1*, *MDC1*. 2) Proteínas mediadoras: *RB1*, ciclinas (por ejemplo, *CCNE1*, *CCND1*, *CCNB1*), *CDKs* (por ejemplo, *CDK7*, *CDK10*). 3) *Checkpoints* del ciclo celular: inhibidores *CDK* (*CDKN2C*, *CCKND2*), *ATM*, *ATR*, *TP53*, *CHEK2*. 4) Reparación no homóloga (NHEJ): *XRCC6*, *XRCC5*, *PRKDC*, *XRCC4*, *LIG4*. 5) Reparación homóloga (HR): *RAD51*, *BRCA1*, *BRCA2*, *XRCC3*. 6) Reparación por escisión de base: *XRCC1*, *APEX*, *OGG1*.

- ❖ Fibrogénesis por radiación: 1) Apoptosis: *TP53*, *BCL2*, *CASP3*, *BIRC5*, *RHOB*, *TP53INP1*. 2) Citosinas proinflamatorias: *TNF*, *IL1A*, *IL6*. 3) Proteínas profibróticas: *TGFB1*, *TGFB2*, *TGFB3*, *CTF*. 4) Vía de señalización de Smad: *SMAD1*, *SMAD2*, *SMAD3*, *SMAD5*, *SMAD8*. 5) Factores de transcripción, co-reguladores y co-receptores de unión al ADN: Receptor tipo I (*SMAD4*), inhibidores *SMAD* (*SMAD6*, *SMAD7*), *NFKB*. 6) Aumento de la deposición de colágeno en la matriz extracelular: receptores transmembrana *TGF*; *TGFB1* y *TGFB2*.
- ❖ Estrés oxidativo: 1) Enzimas antioxidantes: Superóxido dismutasas (*SOD1*). 2) Sistema renina-angiotensina (*AGT*).
- ❖ Daño celular endotelial: 1) Quimiocinas reclutadoras de monocitos: proteínas Rho, quinasas Rho, *CTGF*, *HSP27*, *ZYX*. 2) Citocinas (*FGF2*). 3) Factores de crecimiento (*VEGF*)⁶⁰.

Aunque como mencionábamos previamente, existen casos de enfermedades raras, en donde la susceptibilidad a la radiación ionizante puede relacionarse con alteraciones en un solo gen (*ATM*, *NBN*, etc.), la mayor parte de los pacientes que presentan efectos adversos después del tratamiento con radioterapia, no presentan estas enfermedades. Por ello, la susceptibilidad genética a la radiación ionizante se considera un fenotipo complejo, originado por el efecto combinado de múltiples variantes genéticas. Existen dos hipótesis sobre como los factores genéticos contribuyen en los fenotipos complejos. La primera hipótesis se denomina enfermedad común – variante común, esta hipótesis aplicada a la radioterapia sugiere que la respuesta a la radiación se debe a un conjunto de variantes con frecuencias poblacionales relativamente altas en un número limitado de genes. Por otra parte, la segunda hipótesis sostiene que es poco probable que las variantes comunes afecten a los rasgos complejos e hipotetiza que son múltiples variantes raras las que contribuyen a una mayor susceptibilidad⁷⁹. Estas dos hipótesis dan lugar a un escenario adicional, el cual considera que los rasgos complejos están compuestos de una combinación de variantes genéticas raras y comunes⁸⁰. Debido a que la radiosensibilidad puede verse influenciada por un conjunto de variantes genéticas, Andreassen et al. (2002)⁸⁰ proponen un modelo hipotético que explica cómo la variación genética puede influir en la sensibilidad a la radioterapia en diferentes pacientes. En este modelo se plantea que algunas variantes genéticas se expresan selectivamente en tejidos específicos o se asocian a reacciones adversas concretas y se describe además la existencia de variantes genéticas que afectan de forma más generalizada o global la radiosensibilidad del paciente (**Figura 10**).



Figura 10. Modelo hipotético de la influencia de la variación genética en la radiosensibilidad en pacientes con tejido sano. Cada letra representa un gen con variantes genéticas que afectan la radiosensibilidad. La altura de cada caja representa contribución de las variantes a la radiosensibilidad. Algunos de los genes (C-I) se asocian a reacciones adversas específicas, mientras que otros genes (A y B) afectan de forma más global. El recuadro de la derecha muestra como una mutación de alta penetrancia en el gen A puede ocasionar aumento en la radiosensibilidad general. Adaptado de Andreassen et al. (2002)⁷², permitido por: Elsevier y creado con BioRender.

1.3.2 Radiogenómica

El término radiogenómica hace referencia a dos contextos principales. Por un lado, se utiliza para referirse al estudio de la variación genética asociada con la respuesta a la radiación (*Radiation Genomics*), es decir a la genómica de la radiación. En este sentido fue utilizado por primera vez por Andreassen et al. 2002⁸⁰ como analogía de la farmacogenética⁸⁰. Aquí hace referencia a la respuesta del tejido normal a la radiación ionizante. El estudio de la variación genética se ha basado principalmente en el análisis de polimorfismos de un solo nucleótido (SNPs) y su asociación con el riesgo de desarrollar efectos adversos después del tratamiento radioterápico⁸¹⁻⁸³. En el mismo contexto se utiliza en relación a la respuesta tumoral a la radioterapia⁸⁴⁻⁸⁶, es decir, evalúa la respuesta del tejido tumoral a la radiación ionizante. En este marco, el propósito final de la radiogenómica es desarrollar un perfil de riesgo genético que permita predecir la respuesta de los pacientes a la radiación, y así personalizar y optimizar el tratamiento⁶⁰.

En el año 2009, con la finalidad de fomentar la colaboración multicéntrica entre investigadores cuyo enfoque es vincular las variantes genéticas con la respuesta a la radioterapia se estableció el Consorcio Internacional de Radiogenómica (RGC, *RadioGenomics Consortium* <https://epi.grants.cancer.gov/radiogenomics/>). Esta iniciativa tiene como propósito desarrollar un perfil de riesgo genético capaz de predecir qué

pacientes con cáncer tendrán mayor probabilidad de desarrollar efectos adversos a la radiación. Otro propósito es la obtención de información que contribuya a dilucidar las vías moleculares implicadas en las toxicidades inducidas por la radiación en tejido normal, por medio de la identificación de SNPs asociados a efectos adversos radioinducidos⁸⁷. A la fecha, el consorcio de Radiogenómica ha publicado numerosos estudios en los que se han identificado variantes genéticas asociadas con efectos adversos a la radioterapia en distintos tipos de cáncer, tales como próstata, mama, pulmón y cabeza y cuello. Entre sus investigaciones destacan los estudios de asociación GWAS, METAGWAS y de riesgo poligénico.

El segundo contexto del termino radiogenómica, ha sido utilizado posteriormente para estudiar la correlación entre las características de las imágenes del cáncer y el perfil genómico, es decir, hace referencia a la genómica de la imagen⁸⁸⁻⁹⁰.

1.4 ESTUDIOS RADIOGENÓMICOS

1.4.1 Estudios de asociación gen-candidato

Los estudios de asociación de gen-candidato (CGS, *Candidate Gene Studies*) analizan genes que se seleccionan de acuerdo al conocimiento *a priori* de su impacto funcional en relación con el fenotipo de estudio. Después de seleccionar los genes o regiones, se eligen las variantes que se incluirán en el estudio, que generalmente son SNPs. Para el estudio de asociación, los CGS comparan las frecuencias alélicas entre casos y controles, empleando generalmente para ello el test de Chi cuadrado (χ^2). Los resultados obtenidos de χ^2 darán un p-valor para cada SNP analizado. Además, suele estimarse la magnitud del riesgo, a través del *odds ratio* con su intervalo de confianza (95% CI). El p-valor del resultado indicará el grado de asociación SNP-fenotipo y el nivel de significancia, mientras que el OR indicará si un alelo confiere riesgo ($OR > 1$) o protección ($OR < 1$).

Los primeros estudios en radiogenómica basaron su enfoque en el análisis de genes candidatos, principalmente en aquellos implicados en las vías de detección del daño y reparación del ADN (*ATM*, *BRCA1*, *BRCA2*, *TP53*, *XRCC1*, *XRCC3* y *APEX*), eliminación de especies reactivas de oxígeno (*SOD2* y *GSTP1*) y de respuesta antiinflamatoria (*TGFB1*)⁹¹⁻⁹³. Un alto porcentaje de los estudios reportaron resultados de asociación significativos, sin embargo, la mayoría de ellos carecían de validación, tal ha sido el caso de las variantes comunes en el gen *TGFB1*, previamente asociadas con radiotoxicidad, que no han podido ser replicadas^{82,94}.

Las limitaciones de los estudios CGS se basan principalmente en deficiencias metodológicas, al uso de pequeños tamaños de muestra que derivan en falta de poder estadístico para detectar los efectos pequeños, que normalmente son esperados para los SNPs⁹¹. Además, al basarse en el conocimiento *a priori* para seleccionar los genes a

estudiar, omiten aquellos genes no conocidos, que pudieran estar relacionados con el fenotipo y la naturaleza poligénica de la radiosensibilidad⁹³.

1.4.2 Estudios de asociación gen-candidato en cáncer de próstata

En radiogenómica del cáncer de próstata se han realizado numerosos estudios CGS. La mayor parte de ellos se basó en el estudio de variantes en genes participantes en vías de la reparación del ADN. En la **Tabla 3** se presentan los SNPs y los genes analizados en este tipo de estudios y que han presentado resultados de asociación con radiotoxicidades.

Tabla 3. Estudios gen-candidato asociados con toxicidad postradioterapia

Gen	SNP	OR /HR*(95%CI)	p-valor	Localización tumoral	Referencia
<i>MLH1</i>	rs1799977	2,86 (1,15-7,11)*	0,024**	próstata	Damaraju et al. 2006 ⁹⁵
<i>XRCC3</i>	rs1799794	4,83 (1,67-13,99)*	0,004**	próstata	Damaraju et al. 2006 ⁹⁵
<i>CYP2D6*4</i>	rs1800716 (G>A)	2,63 (1,23-5,63)*	0,013**	próstata	Damaraju et al. 2006 ⁹⁵
<i>ERCC2</i>	rs1052555	2,41 (0,91-6,39)*	0,078**	próstata	Damaraju et al. 2006 ⁹⁵
<i>MSH</i>	rs1800935	2,10 (0,99-4,46)*	0,053**	próstata	Damaraju et al. 2006 ⁹⁵
<i>XRCC3</i>	rs1799794	4,83 (1,67-13,99)*	0,004**	próstata	Damaraju et al. 2006 ⁹⁵
<i>NBN</i>	rs1805794	2,89 (0,69-12,19)*	0,150**	próstata	Damaraju et al. 2006 ⁹⁵
<i>XRCC3</i>	rs861539	NA	0,12	próstata	Burri et al. 2008 ⁹⁶
<i>SOD</i>	rs4880	NA	0,07	próstata	Burri et al. 2008 ⁹⁶
<i>XRCC1</i>	rs25489	0,284(0,051-0,956)	0,043	próstata	Langsenlehner et al. 2011 ⁹⁷
<i>XRCC3</i>	rs1799794 (G/G)	5,65(1,95-16,38)	0,03	próstata	Fachal et al. 2012 ⁹⁸
<i>XRCC3</i>	rs1799794 (G/A)	2,75(1,25-6,05)	0,03	próstata	Fachal et al. 2012 ⁹⁸
<i>TP53</i>	rs35117667	0,007(0,0002-0,227)	0,003	próstata	Cintra et al. 2013 ⁹⁹
<i>TP53</i>	rs17883323	0,071(0,006-0,784)	0,032	próstata	Cintra et al. 2013 ⁹⁹
<i>ATM</i>	rs1801516	1, 20(1, 04-1, 38)	NA	Próstata	Andreassen et al. 2016 ⁹¹
<i>ERCC2</i>	rs13181	0,71(0, 55-0, 93)	0,01	próstata	Song et al. 2015 ¹⁰⁰
<i>TGFBR1</i>	rs10512263	3,80 (1, 78-8, 12)	0,0006***	próstata	Guhlich et al. 2021 ¹⁰¹

SNP, polimorfismo de un solo nucleótido; OR, *odds ratio*; CI, intervalo de confianza; HR*, *Hazard ratio*, ** Cox P valor; *** p-valor corregido por Bonferroni

1.4.3 Estudios de asociación de genoma completo

Los estudios de asociación de genoma completo (GWAS - *Genome-wide association studies*) son estudios sin hipótesis previa, es decir no seleccionan variantes o genes candidatos, sino que buscan descubrir nuevas asociaciones, a través del estudio de

variantes distribuidas por todo el genoma. El objetivo de los GWAS es identificar mediante un análisis de frecuencias alélicas, la asociación entre los genotipos y los fenotipos del estudio. En este análisis se incluyen y se comparan las variantes en individuos que son fenotípicamente diferentes. En los GWAS además de analizarse SNPs, también pueden analizarse otros tipos de marcadores genéticos, incluidas las variantes en el número de copias (CNV, *Copy-number variants*) u otro tipo de variantes de secuencia genómica¹⁰². Los GWAS permiten el análisis de cientos de miles de SNPs distribuidos a lo largo de todo el genoma. Para ello se emplean matrices de genotipado (*arrays*) comerciales, donde el número de SNPs a analizar dependerá del *array* utilizado. Como hemos mencionado en este apartado, el análisis se realiza mediante un análisis de asociación entre los SNPs y el fenotipo a estudio.

Los SNPs incluidos en los arrays se seleccionan considerando el desequilibrio de ligamiento que hay entre ellos, incluyendo aquellos más significativos y representativos del bloque haplotípico, que se denominan tag-SNPs. Después del genotipado se realiza el control de calidad de los datos, que incluye considerar la calidad del genotipado, evaluar la contaminación de las muestras, la concordancia del sexo genético, así como la existencia de subestructura poblacional, entre otros. Posteriormente se suele realizar una imputación de variantes, que consiste en emplear paneles de referencia para aumentar el número de variantes a estudiar. Este aumento se lleva a cabo mediante la imputación de variantes incluidas en el panel de referencia usando como esqueleto los datos de genotipado. Entre los paneles de referencia más utilizados se encuentran el proyecto HapMap (actualmente casi en desuso)¹⁰³, 1000 Genomas¹⁰⁴, o las más recientes del Consorcio de Haplotipos de Referencia (HRC, *Haplotype Reference Consortium*)¹⁰⁵ y TopMED¹⁰⁶. Los análisis de asociación emplean un modelo de regresión logística para rasgos binarios, o una regresión lineal para rasgos cuantitativos. Debido a que en un estudio de GWAS se analizan millones de variantes, para que la asociación se considere significativa, el p-valor obtenido debe de ser inferior a 5×10^{-8} , que sería la corrección de Bonferroni cuando analizamos un millón de variantes independientes (calculado para $\alpha = 0,05$, en un estudio en el que se realizan 10^6 tests $0,05/10^6$). Debemos mencionar que aquellas variantes que sean significativas con un p-valor inferior a 5×10^{-8} en un estudio GWAS deben de ser validadas posteriormente en un estudio réplica o a través de un metaanálisis para que finalmente puedan ser consideradas como realmente asociadas al fenotipo de estudio. Este hecho hace que los estudios actuales de GWAS empleen dos fases, una de descubrimiento (donde se determina el nivel de asociación de las variantes con el fenotipo) y una de réplica (donde las variantes significativamente asociadas con el fenotipo son validadas, si también se asocian significativamente con el fenotipo en esta cohorte).

El primer GWAS llevado a cabo en radiogenómica incluyó a pacientes afroamericanos con cáncer de próstata¹⁰⁷. El fenotipo estudiado fue la disfunción eréctil posterior al tratamiento del cáncer con radioterapia externa EBTR. Se incluyeron 79 pacientes (27 casos y 52 controles) en el estudio. El grado de disfunción eréctil se evaluó basándose

en el cuestionario del Inventario de Salud Sexual para Hombres (SHIM, *The Sexual Health Inventory for Men*)¹⁰⁸. El análisis de GWAS incluyó 512 497 SNPs. El estudio identificó al SNP rs2268363, localizado en el gen del receptor hormonal folículo estimulante (FSHR) asociado significativamente (OR = 7,03; 95% CI [3,4 - 14,7]; p-valor = $5,46 \times 10^{-08}$ (ajustado por Bonferroni 0,028)) con disfunción erétil.

El segundo GWAS en radiogenómica, se publicó en el año 2013. Este GWAS, al igual que el primero, dirigió su estudio a la asociación de disfunción erétil en pacientes con cáncer de próstata tratados con radioterapia¹⁰⁹. El estudio se realizó en dos etapas, dividiendo la cohorte de forma aleatoria en dos grupos. La cohorte de descubrimiento incluyó un total de 235 pacientes (132 casos y 103 controles). Los 940 SNPs más significativos fueron analizados en la cohorte de réplica, que incluyó a 230 pacientes (128 casos y 102 controles). El fenotipo fue definido utilizando la puntuación Función Erétil -Mont Sinai (MSEF, *Mount Sinai Erectile Function*)¹¹⁰ para el 42 % de los pacientes, o los cuestionarios SHIM, para el resto de los pacientes. Como resultado doce SNPs (rs322895, rs11693002, rs3749191, rs11747037, rs6557362, rs766838, rs1486147, rs10764930, rs895255, rs9595967, rs11648233 y rs7245988) fueron identificados, replicados y asociados con disfunción erétil postradioterapia. Sin embargo, ninguno de esos SNPs resultó ser estadísticamente significativo a un umbral de 10^{-07} o 10^{-08} . Los doce SNPs se combinaron en un puntaje acumulado de alelos de riesgo, obteniendo un OR = 2,2; 95% CI [1,9 - 2,6]; p-valor = $2,1 \times 10^{-19}$.

En el año 2013 Kerns et al., publicaron otros dos estudios de GWAS. El primer estudio se enfocó en la búsqueda de SNPs asociados con síntomas urinarios relacionados con la RT¹¹¹ y el segundo en la asociación de sangrado rectal con la RT¹¹².

El estudio de Kerns et al.¹¹¹, sobre toxicidad urinaria postradioterapia, incluyó un total de 723 pacientes con cáncer de próstata. Los pacientes fueron divididos aleatoriamente en dos cohortes; la cohorte de descubrimiento (346 pacientes) y la cohorte de réplica (377 pacientes). Para la definición del fenotipo se consideró la puntuación de síntomas de la Asociación Americana de Urología (AUASS, *American Urological Association Symptom Score*)¹¹³. Treinta y dos SNPs presentes en 17 regiones genómicas, previamente identificados en la fase de descubrimiento fueron estudiados en la cohorte de replicación. La mayor asociación entre toxicidad urinaria general y radiotoxicidad fue encontrada en ocho SNPs localizados en el *locus* 9p21.2, que forman un haplotipo. De ellos, el SNP rs17779457 localizado en el gen *IFNK* mostró la significancia más alta en ambas cohortes (p-valor combinado = $6,5 \times 10^{-07}$).

En el segundo estudio GWAS de Kerns et al. del 2013¹¹², el *fenotipo* estudiado fue el sangrado rectal tardío en pacientes con cáncer de próstata postradioterapia. El estudio fue realizado en dos fases. En la cohorte de descubrimiento se incluyeron 368 pacientes (79 casos y 289 controles) del Mount Sinai Medical Center (MSMC) EEUU, y se analizaron 614 453 SNPs. Los SNPs más significativos de la asociación fueron analizados en la cohorte de replicación. Esta fase incluyó a 781 pacientes (108 casos y 673 controles) del

Hospital Clínico Universitario de Santiago de Compostela en España (Cohorte RADIOGEN - Estudio de la Influencia Genómica en Pacientes Tratados con Radioterapia), del *Maastricht Radiation Oncology Clinic* (MAASTRO) Países Bajos, del *Cross Cancer Institute (CCI) and Tom Baker Cancer Centre* (TBCC) Canadá y Florida *Radiation Oncology Group* en (FROG) EE. UU. De 491 SNPs analizados en la cohorte de replicación, los SNPs rs7120482 y rs17630638 localizados en el *locus* 11q14.3, se encontraron asociados con sangrado rectal postradioterapia, con un p-valor combinado = $5,4 \times 10^{-08}$ y $6,9 \times 10^{-07}$ respectivamente. El fenotipo se definió de la siguiente forma: MSMC y FROG utilizaron los esquemas de puntuación de morbilidad para radiación tardía del Grupo de Oncología de Radioterapia (RTOG, *Radiation Therapy Oncology Group*), RADIOGEN y CCI utilizaron CTCAE v3.0 y MAASTRO utilizó un autoinforme armonizado con RTOG. Pacientes con sangrado rectal grado 2 o superior se consideraron casos y pacientes con grado 0 o 1 se consideraron controles. Se consideró como sangrado rectal tardío aquel presentado 90 días postradioterapia. En este estudio, también se analizaron 18 SNPs que en estudios previos se asociaron con toxicidad rectal, y de los que tenían datos disponibles de algunos pacientes. Dos SNPs, el rs1800935 en *MSH6* y el rs4803455 en *TGFB1*, presentaron un p-valor menor de 0,05, que después de las correcciones no permanecieron significativos.

El primer estudio GWAS en radiogenómica que incluyó un gran número de pacientes¹¹⁴, analizó y evidenció la asociación de variantes comunes con toxicidad tardía a la radioterapia. Este estudio se llevó a cabo en dos fases. En la primera fase se incluyeron 1850 pacientes de la cohorte RAPPER (*Radiogenomics: Assessment of Polymorphisms for Predicting the Effects of Radiotherapy*) de Reino Unido, 633 pacientes con cáncer de próstata y 1276 pacientes con cáncer de mama. La toxicidad global fue definida con la puntuación STAT¹¹⁵ (*Standardized Total Average Toxicity*). Para cáncer de mama, las toxicidades incluidas fueron dolor, telangiectasia y encogimiento de la mama. Para cáncer de próstata se incluyó incontinencia urinaria, disminución del flujo de orina, frecuencia urinaria, nocturia, proctitis, sangrado e incontinencia rectales. La fase de réplica se llevó a cabo en tres cohortes y con los dos tipos de cáncer e incluyó: 355 pacientes con cáncer de mama de la cohorte LeND de Reino Unido y para cáncer de próstata se incluyeron 516 pacientes de la cohorte RADIOGEN y 862 de la cohorte de Estados Unidos Gene-PARE (*Genetic Predictors of Adverse Radiotherapy Effects*). Los resultados de toxicidad global no replicaron, sin embargo, se encontró fuerte asociación en las toxicidades individuales.

En 2014, se publicó un GWAS llevado a cabo por nuestro grupo de investigación. En este estudio de tres fases se identificó el *locus* 2q24.1 de susceptibilidad asociado al desarrollo de toxicidad tardía a la radioterapia en pacientes con cáncer de próstata. El estudio fue realizado en su primera fase en 741 pacientes con cáncer de próstata de la cohorte RADIOGEN de España. En su segunda fase incluyó 633 pacientes de la cohorte RAPPER y en la tercera fase se incluyó a 368 pacientes de la cohorte Gene-PARE. Se analizaron para toxicidad tardía 417 pacientes de RADIOGEN, 579 de RAPPER y 270

de Gen-PARE. Las toxicidades evaluadas fueron: diarrea, hemorragia gastrointestinal, proctitis, incontinencia gastrointestinal, frecuencia/urgencia genitourinaria, cistitis, incontinencia y retención genitourinarias. El fenotipo fue definido como toxicidad global (STAT)¹¹⁵ aguda (de la semana 1-20) y tardía (de dos a cinco años postratamiento). Todos los datos fueron homogeneizados y normalizados. Siete SNPs fueron asociados con toxicidad tardía global a la radioterapia (p-valor combinado = $4,64 \times 10^{-11}$). Estos SNPs se encontraron localizados en el gen *TANC1* (*locus* 2q24.1). Destacó entre ellos el SNP rs264663, que podría estar constituyendo un eQTL y podría regular la expresión del gen *TANC1*¹¹⁶.

La mayoría de los GWAS realizados en radiogenómica han incluido pacientes con cáncer de próstata. Sin embargo, recientemente se han publicado cuatro estudios GWAS, que asocian toxicidades en pacientes con cáncer de cabeza y cuello con radioterapia.

En un estudio de 2942 pacientes con cáncer de cabeza y cuello, Wang et al (2019)¹¹⁷ asociaron las variantes rs17111237 y rs162171 en *CEP128* (*hazard ratio* (HR) = 1.45; 95% CI [1.26 to 1.66]; p-valor combinado = $3,18 \times 10^{-07}$ y HR = 1.46; 95% CI [1.29–1.66]; p-valor combinado = $6,17 \times 10^{-09}$ respectivamente) con la susceptibilidad a lesión cerebral inducida por radiación. Otro GWAS incluyó 1467 pacientes con carcinoma nasofaríngeo e identificó 50 *loci* y 64 genes asociados con mucositis oral aguda inducida por radioterapia¹¹⁸. Los SNPs más significativos fueron: rs9484606 (OR = 1,70; 95% CI [1,36 - 2,13]; p-valor $2,98 \times 10^{-06}$), rs16876733 (OR = 1,95; 95% CI [1,47–2,59]; p-valor $3,05 \times 10^{-06}$) y rs117157809 (OR = 3,72; 95% CI [2,10 - 6,57]; p-valor $6,33 \times 10^{-06}$). Análisis posteriores de las variantes asociadas realizados con la plataforma de interpretación funcional de variantes de GWAS FUMA¹¹⁹, identificaron conjuntos de genes involucrados en la regulación de la actividad de los telómeros y en la vía de señalización Wnt, que previamente se ha relacionado como promotora de la radiorresistencia¹²⁰. El tercer GWAS tuvo como objetivo identificar SNPs asociados con toxicidades agudas inducidas por radiación (mucositis, disfagia, xerostomía (sequedad en la boca y/o saliva espesa)¹²¹ e incluyó 1279 pacientes con cáncer de cabeza y cuello. El estudio se llevó a cabo en dos fases. En la fase de descubrimiento, 16 SNPs en 5q21.3 se asociaron con xerostomía aguda, sin embargo, estos resultados no se asociaron en la validación. El cuarto estudio GWAS, y el más reciente en radiogenómica, incluyó 1780 pacientes europeos con cáncer de cabeza y cuello en laringe y oro/hipofaringe, que fueron tratados con radioterapia. En este estudio se identificó en el cromosoma 5 un *locus* de susceptibilidad asociado con mucositis inducida por radioterapia¹²². Las toxicidades fueron normalizadas y combinadas, estimando para esto las puntuaciones STAT¹¹⁵. El fenotipo fue definido como STAT_{agudo} (disfagia aguda y mucositis), STAT_{tardío} (disfagia tardía, xerostomía, fibrosis y atrofia) y STAT_{global} (STAT_{agudo} + STAT_{tardío}). El *locus* en el cromosoma 5 en el gen *STING1* replicó en el metaanálisis (rs1131769*C (OR = 1,95; 95% CI [1,48 - 2,41] p-valor combinado = $4,34 \times 10^{-16}$) y rs28419191*C (OR = 1,96; 95% CI [1,45 - 2,46]; p-valor combinado = $3,67 \times 10^{-14}$)). La validación adicional se realizó con datos

de secuenciación del exoma completo de una cohorte independiente, sin embargo, el SNP rs1131769 no validó, y otros dos SNPs significativos (rs28419191 y rs10875554) no estaban disponibles en los datos del exoma por tratarse de variantes intergénicas.

1.4.4 Metaanálisis de GWAS

Los GWAS han permitido un sinnúmero de asociaciones entre genotipo-fenotipo. Sin embargo, a pesar de la identificación de nuevos genes, de variantes y de vías biológicas implicadas con el fenotipo, los GWAS tienen limitaciones importantes. Por ejemplo, la baja potencia para detectar toda la heredabilidad explicada por los SNPs¹²³, para identificar variantes raras o los efectos genéticos de pequeño tamaño¹²⁴. Una forma eficaz de aumentar el poder de los GWAS y subsanar algunas de estas limitaciones se logra a través del aumento del tamaño de la muestra. En este punto, los metaanálisis juegan un rol fundamental debido a su capacidad para analizar conjuntamente distintas cohortes y detectar asociaciones más consistentes¹²⁴. El aumento del tamaño de la muestra generalmente se da en el marco de consorcios internacionales^{123,124}, ya que a través de ellos se puede tener acceso a diferentes cohortes que pueden utilizarse para llevar a cabo la realización de metaanálisis. La ventaja de los metaanálisis es que, al poder realizarse de modo acumulativo, permiten la incorporación de nuevos conjuntos de datos de forma secuencial, y por lo tanto el tamaño muestral aumenta y consecuentemente el poder estadístico para descubrir nuevas variantes¹²⁴.

Hasta el momento de la escritura de esta tesis, se han publicado dos metaanálisis de GWAS realizados dentro del Consorcio Internacional de Radiogenómica. El primero de ellos se realizó con datos de GWAS de cuatro cohortes de pacientes con cáncer de próstata tratados con radioterapia⁶¹. Se incluyó un total de 1564 pacientes de las cohortes; RAPPER (n = 527), RADIOGEN (n = 597), Gene-PARE (n = 290) y CCI (n = 150). Las toxicidades tardías fueron medidas en un periodo de 1,5 a 2,5 años postradioterapia. Las toxicidades registradas fueron: sangrado rectal, frecuencia urinaria, disminución del flujo de orina y toxicidad global. Para la evaluación de las toxicidades se usaron diferentes sistemas de puntuación de toxicidad. RAPPER utilizó RMH (*Royal Marsden Hospital*) RTOG y LENT-SOMA (*Late effects on Normal Tissues-Subjective Objective Management Analytic*)⁷³; RADIOGEN y CCI emplearon CTCAE v.3.0 y Gene-PARE uso IPSS (*International Prostate Symptom Score*) y RTOG. Todos los datos de toxicidad fueron armonizados y la toxicidad global se estandarizó con STAT. Los GWAS se analizaron individualmente para cada toxicidad, y se ajustaron de acuerdo con sus factores de riesgo no genéticos. El metaanálisis usando el modelo de efectos fijos identificó al SNPs rs17599026 en 5q31.2 asociado con frecuencia urinaria (OR = 3,12; 95% CI [2,08 - 4,69]; p-meta $4,16 \times 10^{-08}$) y al SNP rs7720298 en 5p15.2 asociado con disminución del flujo de orina (OR = 2,71; 95% CI [1,90 - 3,86]; p-meta $3,21 \times 10^{-08}$). Los SNPs identificados en este metaanálisis están localizados en genes que se expresan en tejidos (vejiga, recto e intestino delgado) afectados por la radioterapia pélvica. Además, siete SNPs del locus 2q24.1 previamente publicados por RADIOGEN, fueron replicados

en la cohorte CCI. Todos los SNPs mostraron una dirección consistente de asociación con la toxicidad global, de ellos el SNP rs264651 presentó asociación significativa. Al comprobar la asociación de este *locus* con las tres toxicidades se observó asociación con toxicidad urinaria. También se identificó asociación consistente en seis de los siete SNPs con hemorragia rectal, sin embargo, ninguno alcanzó la significancia estadística. Los resultados obtenidos sugieren que la asociación de la toxicidad global podría estar siendo impulsada por toxicidades urinarias.

El segundo metaanálisis incluyó seis GWAS de 3871 pacientes europeos con cáncer de próstata tratados con radioterapia¹²⁵. Las cohortes incluidas fueron: RAPPER, RADIOGEN, GenePARE, UGhent (University of Ghent) CCI-BT (*Cross Cancer Institute–Brachytherapy*), CCIEBRT (*Cross Cancer Institute–External Beam Radiotherapy*). Además, para la réplica de los resultados se incluyeron 962 pacientes de tres cohortes japonesas (PRRG-fotón, PRRG-ion carbón y NTMC (*National Tokyo Medical Center*)). Los grados de toxicidad fueron armonizados de acuerdo con los criterios CTCAE. Las toxicidades evaluadas fueron: el aumento de la frecuencia urinaria, la disminución del flujo de la orina, hematuria, sangrado rectal y la toxicidad global calculada con la puntuación STAT. El fenotipo fue definido como el periodo de tiempo comprendido desde el inicio de la radiación hasta la primera aparición de alguna toxicidad grado dos o superior. Como resultado del metaanálisis de efectos fijos se identificaron tres nuevos SNPs: rs17055178 asociado con sangrado rectal (HR = 1,95; 95% CI [1,58 - 2,40]; p-meta = $6,2 \times 10^{-10}$), rs10969913 asociado con disminución del flujo urinario (HR = 3,92; 95% CI [2,57 - 6,00]; p-meta = $2,9 \times 10^{-10}$) y rs11122573 asociado con hematuria (HR = 1,92; 95% CI [1,53 - 2,42]; p-meta = $1,8 \times 10^{-08}$). Estas tres regiones fueron mapeadas con *fine-scale mapping*, identificándose el SNP rs147121532 asociado con hematuria (p-condicional = $4,7 \times 10^{-06}$). Con la herramienta *Variant Effect Predictor* se anotaron las variantes causales creíbles (CCV, *Credible Causal Variants*), para determinar su efecto en los genes, en la transcripción o en la secuencia proteica¹²⁶. Las CCV se identificaron en regiones reguladoras y de expresión en los genes. Dos de ellas se asociaron con hematuria; la primera señal asociada con rs11122573 contiene 47 CCVs, los alelos de esta disminuyen la expresión de los genes *AGT* y *COG2* (con actividad en tejidos y arterias). La segunda señal contiene diez CCVs asociadas con el rs147121532, que disminuye la expresión de los genes *CAPN9* y *ARV1* (con actividad en tejidos gastrointestinales). Por otro lado, las variantes rs17599026, rs7720298 y rs1801516, previamente asociadas con el aumento de la frecuencia urinaria, la disminución del flujo urinario y con la toxicidad global, respectivamente, no replicaron en las cohortes japonesas, aunque presentaron la misma dirección del efecto que las cohortes europeas.

1.4.5 Machine learning y deep learning en radiogenómica

Los modelos entrenados de inteligencia artificial como *machine learning* y *deep learning* han sido empleados recientemente en estudios de radiogenómica^{127,128, 129,130,131}. *Machine learning* es un aprendizaje automático, que evolucionó a partir de la inteligencia

artificial, la informática y la inferencia estadística. Se basa en el descubrimiento de patrones en los datos que permitan hacer predicciones¹³². Esto permite que los modelos de *machine learning* sean, normalmente, generalizables a distintos tipos de problema, en lugar de ser problema-específicos.

El uso de *machine learning* en estudios de radiogenómica fue realizado por primera vez en pacientes (n = 368) con cáncer de próstata tratados con radioterapia¹²⁷. En este estudio se hipotetizó que los métodos de *machine learning* que emplean los resultados obtenidos de estudios GWAS podrían identificar pacientes con diferentes riesgos de efectos adversos y ayudar a la identificación de fuentes biológicas claves asociadas a las toxicidades (sangrado rectal y disfunción eréctil). Se desarrolló un algoritmo de aprendizaje denominado *pre-conditioned random forest regression* (PRFR), que permite construir modelos de riesgo poligénico utilizando cientos de SNPs. Para la identificación de los procesos biológicos clave y las proteínas, se realizaron análisis de enriquecimiento ontológico y de interacción proteína-proteína, respectivamente. Se encontró a los genes *CACNA1D* y *CXCR5* implicados en procesos biológicos asociados con ambos fenotipos. Los análisis de reconocimiento ontológico revelaron que los procesos biológicos involucrados en el transporte de iones se encuentran relacionados con sangrado rectal. La deficiencia del transporte de iones es causada por la inflamación y la diarrea. El sangrado rectal a menudo es acompañado con diarrea, lo que sugiere una conexión entre el sangrado rectal y la deficiencia en el transporte de iones. Los resultados para disfunción eréctil relacionaron a la disfunción del endotelio vascular como su posible causa. Esta se origina debido a la disminución del flujo sanguíneo que se desencadena cuando existe mal funcionamiento vascular. El estudio concluyó que el empleo de métodos de *machine learning* puede generar modelos predictivos que ayuden a la estratificación de los pacientes en función del riesgo a desarrollar toxicidad y a identificar los procesos biológicos involucrados tanto en el daño ocasionado por la radiación como en la reparación de tejidos.

Otro estudio que utilizó *machine learning* en radiogenómica incluyó un GWAS con 324 pacientes con cáncer de próstata tratados con radioterapia, para predecir la toxicidad genitourinaria tardía. Aquí también aplicaron el método PRFR. Se analizaron la frecuencia/urgencia urinaria, nocturia, disminución del flujo urinario, obstrucción del flujo y el vaciado incompleto de la vejiga. Solo la disminución del flujo de orina alcanzó el nivel de significancia en el área bajo la curva de 0,7 (95% CI [0,54 - 0,86]; p-valor = 0,01). Se identificaron a la neurogénesis y al transporte de iones como los procesos biológicos clave relacionados con las toxicidades genitourinarias¹²⁸.

Deep learning es un tipo de *machine learning* que se caracteriza por el uso de redes neuronales artificiales en las cuales existen múltiples capas de neuronas que son usadas para extraer de manera progresiva características de los datos analizados. Una estructura tipo de estas redes naturales se caracteriza por tres tipos de capas. La capa de entrada (*input layer*), docenas o centenas de capas ocultas (*hidden layer*) y capas de

salida (*output layer*). En la primera capa se realiza el reconocimiento de la información, en las capas ocultas se realiza la combinación de la información y el mejor resultado de las combinaciones se pasa a la última capa para determinar el resultado¹³².

El estudio más reciente en radiogenómica que involucra el uso de *deep learning*¹²⁹, fue el realizado en 1401 pacientes con cáncer de próstata del estudio REQUITE (*Validating Predictive Models and Biomarkers of Radiotherapy Toxicity to Reduce Side-Effects and Improve Quality of Life in Cancer Survivors*). Los fenotipos de toxicidad tardía estudiados fueron: sangrado rectal, hematuria y disminución del flujo urinario, todos $\geq$ grado 1 (CTCAE v4.0). Además, se consideraron también la frecuencia urinaria y la nocturia, ambos $\geq$ grado 2 (CTCAE v4.0). Se analizaron 43 SNPs previamente reportados en la literatura asociados con estas toxicidades. Se usó el algoritmo *Deep Sparse AutoEncoders* (DSAE), para seleccionar las características más representativas de los datos de entrada, identificar a los SNPs que contribuyen con un mayor riesgo de toxicidad y validar su efecto en la toxicidad tardía a la radioterapia. Veinte SNPs se asociaron con la misma toxicidad a la descrita en la literatura (nueve para toxicidades urinarias y once para toxicidad global) y cuatro se asociaron con otra toxicidad. Estos resultados apoyan el uso de *deep learning* para validar SNPs asociados con toxicidad postradioterapia.

1.5 NUEVAS TECNOLOGÍAS Y APROXIMACIONES ÓMICAS

Las tecnologías de secuenciación de nueva generación (NGS, *Next-Generation Sequencing*), han revolucionado el diagnóstico, pronóstico y toma de decisiones terapéuticas en diversas enfermedades de origen genético¹³³. Estas tecnologías avanzadas ofrecen numerosas ventajas y han tenido un impacto significativo en la investigación, brindando la oportunidad de identificar nuevos *loci* causales o de riesgo. Su aplicación ha ampliado nuestra comprensión de la genética de las enfermedades y ha abierto nuevas perspectivas en la medicina personalizada y la terapia dirigida.

Las tecnologías NGS presentan la capacidad de secuenciar millones de lecturas en paralelo¹³⁴, de forma rápida, eficiente y a un coste más bajo respecto a la secuenciación Sanger¹³⁵. Además, la NGS permite secuenciar diferentes tipos de muestras y regiones genómicas, incluyendo paneles específicos de genes (*target-sequence*), las regiones codificantes de todos los genes del genoma (*Whole exome sequencing*), el ARN codificante y no codificante (*Whole Transcriptome*), modificaciones a nivel de cromatina (metil-seq, ChIP-seq y ATAC-seq) o el genoma completo (*Whole genome sequencing*). La selección del tipo de secuenciación dependerá del enfoque específico de cada estudio, lo que permite su aplicación en diversas áreas, como el descubrimiento de variantes raras, la caracterización de perfiles transcriptómicos o la identificación de marcadores epigenéticos¹³⁶. Estas capacidades versátiles de la NGS han revolucionado la genómica y han abierto nuevas oportunidades de investigación en la medicina personalizada.

1.5.1 Secuenciación del exoma completo

La secuenciación del exoma completo (WES, *Whole exome sequencing*) es una aplicación altamente valiosa de la tecnología NGS que se enfoca principalmente en las regiones codificantes del genoma. Dependiendo de la librería usada también puede incluir regiones no traducidas (UTRs, *untranslated region*). Es importante destacar que las regiones codificantes representan menos del 2 % del genoma completo, pero albergan aproximadamente el 95 % de las mutaciones causantes de enfermedades^{137,138}.

En el ámbito de la investigación, la WES es una herramienta ideal para explorar fenotipos complejos y descubrir nuevas asociaciones genotipo-fenotipo. La WES ha demostrado ser especialmente valiosa al permitir el análisis de variantes tanto comunes como raras. Esta capacidad de identificar variantes raras ha ampliado nuestra comprensión de la contribución genética en diversas enfermedades y ha proporcionado nuevas perspectivas para el diagnóstico y la investigación clínica.

1.5.1.1 Estudios basados en genes

Como mencionamos previamente, el análisis del exoma brinda la oportunidad de identificar y analizar un gran número de variantes raras y de baja frecuencia en regiones codificantes. Esto ha permitido establecer vínculos más directos entre la función del gen y la patogénesis de la enfermedad¹³⁹ o de rasgos complejos.

Sin embargo, debido al gran tamaño de muestra requerido para la detección de asociaciones verdaderas de variantes raras analizadas de forma individual, se han desarrollado diversos análisis con enfoques basados en genes. Estos análisis evalúan el efecto acumulado de múltiples variantes en un mismo gen o región de interés. Algunos de los test que pueden utilizarse para este tipo de enfoque son el Burden^{140,141} test, SKAT (*Sequence Kernel Association Test*) y SKAT-O (*Optimal Sequence Kernel Association test*)¹⁴².

El Burden test se basa en el colapso de las variantes raras dentro de una región o gen, asigna una puntuación génica única y analiza la asociación entre esta puntuación con el rasgo de interés¹⁴¹. Este test es más potente cuando gran proporción de las variantes son causales y sus efectos van en la misma dirección. Por otro lado, el test SKAT¹⁴⁰ es una prueba del componente de varianza basada en puntajes. Este test agrega estadísticas de puntaje individual por variante, evalúa la distribución de estas en cada región o gen y su efecto en conjunto sobre el fenotipo¹⁴¹. SKAT es potente cuando hay variantes que aumentan y disminuyen los rasgos o sólo hay una pequeña fracción de variantes causales. En cuanto a SKAT-O¹⁴², es un test que combina la prueba de Burden y las estadísticas SKAT. Este test modifica automáticamente su comportamiento para aplicar entre los dos tests, el método más poderoso y adecuado para cada situación. Un test interesante que forma parte de este tipo de análisis es el *Combined test of common rare variants*¹⁴³, este test permite analizar de forma combinada el efecto de variantes comunes y raras dentro de un mismo gen y asociarlas con el fenotipo de interés.

Una estrategia implementada para aumentar el poder y la sensibilidad de los estudios de variantes raras y de baja frecuencia es la selección de fenotipos extremos. La estrategia de selección de muestras con fenotipos extremos ayuda a reducir el tamaño de la muestra requerido hasta cuatro veces, aumenta el poder para descubrir variantes raras y el tamaño del efecto puede ser mayor (según algunos estudios hasta dos veces)¹⁴⁴.

Algunos estudios realizados en exomas han implementado estrategias de selección de fenotipos extremos de riesgo, incluyendo en algunos casos pequeños tamaños de muestra por grupo (entre 20 – 50 pacientes), así como la priorización de variantes de impacto funcional y han sido capaces de identificar variantes o genes candidatos^{145,146}. Por otro lado, recomiendan la réplica en una segunda cohorte de fenotipos extremos para mejorar el poder del estudio⁶⁵.

El uso de WES, junto con la implementación de estrategias como el análisis de fenotipos extremos, la priorización de variantes y el uso de test que evalúan el efecto acumulado de variantes dentro de un mismo gen, contribuye a potenciar la capacidad de identificación de biomarcadores asociados con fenotipos de interés.

1.5.2 Estudios de asociación del transcriptoma completo

Los estudios GWAS han permitido importantes avances en la identificación de variantes asociadas con enfermedades, rasgos complejos y toxicidades¹⁴⁷. Sin embargo, la mayoría de los SNPs identificados se encuentran localizados en regiones no codificantes, lo que dificulta la interpretación de su función y limita el conocimiento sobre los mecanismos en los que se encuentran involucrados¹⁴⁸.

Por otro lado, diversos estudios han reportado que muchos de los SNPs identificados mediante GWAS tienen alta probabilidad de ser *loci* cuantitativos de rasgos de expresión (eQTL, *expression quantitative trait loci*)¹⁴⁹, los cuales influyen en la expresión génica¹⁵⁰. Esto ha permitido la aplicación y el desarrollo de análisis bioinformáticos que integran las asociaciones de GWAS y los estudios de eQTL¹⁵¹.

En este contexto, destacan los estudios de asociación del transcriptoma completo (TWAS, *Transcriptome-wide association studies*). Los estudios TWAS son una aproximación basada en genes que investiga las asociaciones entre la expresión génica regulada genéticamente y enfermedades complejas o rasgos de interés^{147,152,153}. Estos estudios consideran que uno o varios eQTLs regulan las actividades transcripcionales de un gen, y que las alteraciones genéticas en los niveles de expresión génica dan como resultado cambios en el riesgo de la enfermedad¹⁴⁷.

El primer paso de los análisis TWAS consiste en la construcción de modelos predictivos de expresión génica a partir de los datos genéticos y de expresión génica a nivel individual. Debido al hecho de que se necesitan los datos genéticos individuales se han

creado repositorios públicos donde se han depositado estos modelos para distintos tipos de muestras. Uno de los más utilizados es el repositorio de PredictDB (<https://predictdb.org/>) que alberga modelos de predicción de expresión génica para los tejidos incluidos dentro del repositorio *Genotype-Tissue Expression* (GTEx)¹⁵⁴ y el proyecto *Depression Genes and Networks* (DGN)¹⁵⁵. El siguiente paso es utilizar estos modelos predictivos de expresión en base a genotipo para integrarlo con los resultados obtenidos en un estudio GWAS (tanto a resultados individuales o a las estadísticas de asociación resumidas “*summary-statistics*”) para estimar las asociaciones entre la expresión génica predicha y el rasgo de interés^{153,156}.

El enfoque de estos estudios permite una mejor comprensión funcional de los mecanismos implicados en la sensibilidad y desarrollo de las enfermedades, lo que constituye una de sus principales fortalezas.

Además, en el primer paso, los modelos pueden realizarse en los tejidos de interés tanto de forma individual como en conjunto, lo que ofrece flexibilidad en el diseño experimental. Otro aspecto positivo es que los dos pasos incluidos en este estudio son independientes, lo que conlleva un ajuste de la carga por pruebas múltiples menor en comparación con un GWAS tradicional. Este ajuste se aplica sobre la cantidad de genes y no por variantes individuales. Además de los análisis de TWAS que se realizan por tejido, existe la posibilidad de hacer análisis multi-tejido, lo que equivaldría a un metaanálisis de tejidos.

Una amplia variedad de estudios ha incorporado el análisis TWAS en sus investigaciones. Estos estudios han abordado distintas enfermedades como trastornos neurológicos^{157,158}; distintos tipos de cáncer, incluyendo el cáncer de próstata^{159,160} y algunas toxicidades, en su mayoría relacionadas con fármacos¹⁶¹.

1.5.3 Estudios de asociación del proteoma completo

Varios estudios han observado que un amplio número de los SNPs asociados en GWAS también son eQTLs. Este hallazgo también ha sido observado en relación con los *loci* de rasgos cuantitativos de proteínas (pQTLs, *protein quantitative trait loci*), lo que ha llevado a sugerir un enriquecimiento de pQTLs en los estudios GWAS¹⁶². Además, varios estudios han destacado la superposición entre eQTLs y pQTLs^{163–165}. Sin embargo, es importante tener en cuenta el amplio rango de variabilidad descrito para estas superposiciones, con estimaciones que van desde el 35 % hasta el 80 %. Los estudios de pQTLs desempeñan un papel crucial al permitirnos identificar variantes genéticas que regulan directamente la expresión de proteínas.

Actualmente se cuenta con un nuevo método capaz de detectar asociaciones genotipo-fenotipo mediadas por alteraciones en la función de proteínas, que se denomina estudio de asociación del proteoma completo (PWAS, *Proteome Wide Association Study*). Este enfoque de análisis considera que las variantes causales presentes en regiones

codificantes afectan al fenotipo al modificar las funciones de los productos proteicos de los genes. Estas modificaciones podrían resultar en cambios en las vías de señalización, en el transporte de sustancias, en la actividad enzimática o en la capacidad de unión de la proteína entre otras¹⁶⁶. Para llevar a cabo esta tarea, PWAS integra la información de todas las variantes que afectan de manera simultánea a un gen codificador de proteínas y evalúa el impacto global de estas en la función de la proteína mediante el uso de *machine learning* y modelos probabilísticos. Posteriormente, examina si el gen presenta variabilidad funcional entre individuos que se asocie con el fenotipo de interés¹⁶⁶.

Al igual que en los TWAS, el análisis PWAS implica el uso de modelos predictivos, en este caso son modelos predictores del funcionamiento de la proteína. Luego, estos modelos se correlacionan con los datos obtenidos en estudios de GWAS, ya sea a nivel individual o utilizando los resultados de las estadísticas de asociación por SNP resumidas (*summary-statistics*)¹⁶⁷, con el fin de estimar las asociaciones con el rasgo de interés. En el caso de fenotipos binarios, una correlación significativa indicaría diferencia entre casos y controles, lo que implica que la proteína presenta niveles de alteración más elevados o más bajos en los individuos afectados¹⁶⁶.

En resumen, los enfoques novedosos de estudios multiómicos que integran datos de GWAS, *loci* cuantitativos de expresión y *loci* cuantitativos de proteínas están contribuyendo a una comprensión más profunda de los mecanismos moleculares subyacentes de los rasgos complejos. Estos enfoques permiten un análisis integral de múltiples capas de información genética y molecular, lo que ayuda a desentrañar las interacciones complejas y proporcionar una visión más completa de los procesos biológicos involucrados en los rasgos de interés.

2 JUSTIFICACIÓN Y OBJETIVOS

2.1 JUSTIFICACIÓN Y OBJETIVOS

El cáncer de próstata es el segundo cáncer con mayor frecuencia en hombres a nivel mundial². Se estima que aproximadamente la mitad de los pacientes con cáncer serán tratados con radioterapia en algún momento de su enfermedad¹⁶⁸. El éxito de este tratamiento depende de la dosis administrada y existen límites de dosis que deben considerarse en función de la tolerancia de los tejidos. Aunque la radiación se dirige al tumor, también afecta al tejido adyacente, lo que puede dar lugar al desarrollo de efectos adversos. En la mayor parte de los pacientes, estos efectos adversos son agudos, pero en un pequeño número las toxicidades persisten, afectando significativamente la calidad de vida de los pacientes que las padecen. Aquí destaca el sangrado rectal, una de las radiotoxicidades más frecuentes y la segunda causa de derivación del paciente al servicio de gastroenterología. Esta toxicidad impacta negativamente en las actividades cotidianas del paciente, en su comodidad y funcionalidad. Además, es una de las toxicidades que normalmente requiere intervención médica para su manejo y puede suponer un riesgo vital para el paciente.

La manifestación de los efectos adversos puede variar enormemente. Parte de esta variación puede ser explicada por factores como la edad, la comorbilidad o factores dosimétricos. Sin embargo, estos factores no explican todas las diferencias observadas entre los pacientes. Por tal motivo, se considera que parte de la variabilidad podría estar relacionada con factores genéticos. La sensibilidad a la radiación puede considerarse como un fenotipo complejo influenciado por variantes genéticas tanto comunes como raras. En este sentido, se han realizado diversos estudios de asociación de variantes comunes, incluyendo análisis de genes candidatos y estudios GWAS que han contribuido a la identificación de posibles marcadores asociados a la radiotoxicidad. Sin embargo, la replicación de los resultados, la validación mediante estudios funcionales y el conocimiento de los mecanismos moleculares que subyacen a la susceptibilidad a la radiación, siguen siendo un desafío importante, ya que a día de hoy sólo se consideran asociaciones verdaderas un número limitado de genes y variantes.

Las nuevas tecnologías de secuenciación NGS han abierto las puertas a la identificación de nuevos *loci* causales o de riesgo a enfermedades y rasgos complejos. Además de la implementación de la NGS, los avances tecnológicos, el constante enriquecimiento de diferentes bases de datos y el desarrollo de nuevas e innovadoras aproximaciones ómicas, se presentan como una oportunidad para explorar nuevas estrategias de análisis en el área de radiogenómica.

Así, el objetivo general de esta tesis es identificar biomarcadores en línea germinal asociados con el sangrado rectal como efecto adverso en el tratamiento radioterápico de pacientes diagnosticados de cáncer de próstata. Para ello se proponen los siguientes objetivos específicos:

Objetivos específicos

1. Realizar la secuenciación de exomas (WES) de pacientes con cáncer de próstata tratados con radioterapia, seleccionados de la cohorte RADIOGEN.
2. Analizar variantes genéticas y genes asociados con el desarrollo de radiotoxicidad independientemente de la localización tumoral, validar variantes genéticas asociadas con sangrado rectal en pacientes con cáncer de próstata e identificar nuevas variantes en genes previamente descritos.
3. Identificar nuevos genes asociados con sangrado rectal como efecto adverso en pacientes tratados con radioterapia, mediante estudios de asociación basados en genes.
4. Identificar biomarcadores de susceptibilidad a sangrado rectal mediante nuevas aproximaciones ómicas como son el estudio del transcriptoma completo y del proteoma completo.
5. Explorar, a partir de los resultados más significativos, las rutas metabólicas en las que están implicados los principales genes que se han identificado asociados con sangrado rectal de nuestro estudio.

3 MATERIALES Y MÉTODOS

3.1 CONSIDERACIONES ÉTICAS

Este proyecto ha sido realizado de acuerdo con los principios éticos expresados en la Declaración de Helsinki de la (Asociación Médica Mundial), el Convenio del Consejo de Europa relativo a los derechos Humanos y a la Biomedicina, a la Declaración Universal de la UNESCO sobre el Genoma Humano y los Derechos Humanos. Además, cumple los requisitos establecidos en la Legislación Española en el ámbito de la Investigación Biomédica (Ley 14/2007 de Investigación Biomédica) y la protección de datos de carácter personal (Ley Orgánica 7/2021, de 26 de mayo). El protocolo cuenta con la aprobación del el Comité de Investigación Clínica de Galicia, códigos de registro CEIC de Galicia 2008/092, 2014/107 y 2016/597) (Anexo 4).

Todos los participantes de este estudio fueron debidamente informados sobre la naturaleza y el objetivo del estudio y han otorgado la firma del consentimiento informado tanto para la utilización de sus datos clínicos, como para la obtención de muestras biología y posteriores análisis. El participante pudo negar selectivamente su participación para uno u otro objetivo del estudio. Los datos clínicos y personales fueron recogidos y analizados de manera disociada, garantizando con esto el anonimato de los participantes. Las muestras biológicas se encuentran almacenadas y custodiadas en la Fundación Pública Gallega de Medicina Xenómica, que garantizó su uso para los fines del presente proyecto en las condiciones establecidas por la Ley Orgánica 15/1999 de Protección de Datos de Carácter Personal.

3.2 CÁLCULO DEL PODER DEL ESTUDIO Y TAMAÑO DE LA MUESTRA

El cálculo del poder del estudio se realizó con el paquete *genpwr*¹⁶⁹. Este programa permite calcular la potencia para estudios de asociación en el entorno y lenguaje de programación de R¹⁷⁰, para un tamaño de muestra de 141 pacientes, con una proporción de casos en la muestra (*case.rate*) de 0,3262, un alfa de 0,05 y 5×10^{-5} , MAF de 0,01, 0,05, 0,1 y 0,2 y OR de 1,5, 3, 5 y 7 (**Figura 11**). Para calcular el número de muestra necesaria para alcanzar un poder del 80 % se utilizó el programa *Power and Sample Size*¹⁷¹. El cálculo se realizó considerando una incidencia de sangrado rectal en nuestra muestra del 32.6 %, a diferentes OR (1,5, 3, 5 y 7) y alfas (0,05, 5×10^{-6} y 5×10^{-8}). La **Tabla 4** muestra el número de pacientes necesarios en función de los ORs y de la tasa de error alfa, considerando una proporción entre controles y casos de 2:1.

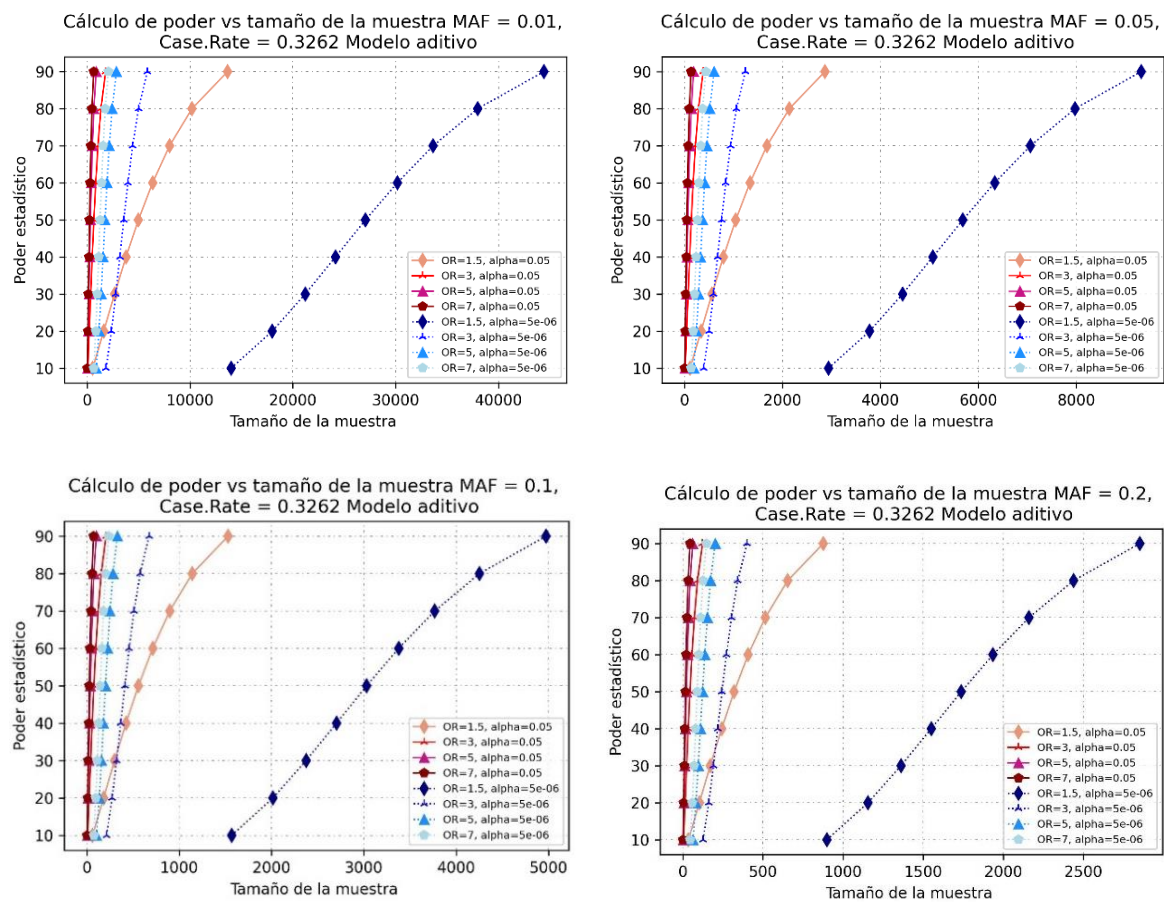


Figura 11. Cálculo de poder del estudio (MAF de 0,01, 0,05, 0,10 y 0,20; OR= 1,5, 3, 5 y 7, a un alfa de 0,05 y 5×10^{-6})

Tabla 4. Cálculo de tamaño de la muestra para un poder del 80 %

Tamaño de muestra a diferente alfa y odds ratio

		Alfa		
		0,05	5×10^{-6}	5×10^{-8}
Odds ratio	1.5	915	3399	4602
	3	120	450	609
	5	57	219	297
	7	42	159	216

3.3 COHORTE DE PACIENTES

3.3.1 Población de estudio

El estudio se llevó a cabo en pacientes seleccionados de la cohorte RADIOGEN-próstata constituida por 1113 pacientes diagnosticados con adenocarcinoma de próstata, incluidos de forma prospectiva por el Servicio de Oncología Radioterápica del Complejo Hospitalario Universitario de Santiago de Compostela y tratados con radioterapia externa entre los años 2005 y 2018¹¹⁶. De todos los pacientes incluidos en la cohorte se dispone de muestras de sangre preriferica extraídas previamente al tratamiento con 3D-CRT, así como de datos dosimétricos, clínicos y de toxicidad, además del consentimiento informado (ver apartado 3.1). El seguimiento incluyó una revisión semanal durante el tratamiento, cada tres meses durante el primer año, cada seis meses los cuatro años siguientes y anualmente *a posteriori*.

3.3.1.1 Selección de pacientes

3.3.1.1.1 Grupo casos y casos extremos

Se revisaron los efectos adversos tardíos que presentaron los pacientes de nuestra cohorte. Estos incluyen: 1) Toxicidades genitourinarias (GU): incontinencia/retención urinaria, frecuencia/urgencia urinaria, estenosis genitourinaria y cistitis. 2) Toxicidades gastrointestinales (GI): proctitis, diarrea, constipación, sangrado rectal, incontinencia gastrointestinal y úlceras. Entre estos destacó por su frecuencia el sangrado rectal.

Se seleccionaron pacientes que a partir del primer año postradioterapia presentaron sangrado rectal con grado ≥ 2 de acuerdo con los CTCAE versión 4.0 (Anexo 5). En total fueron seleccionados: 47 pacientes con grado de toxicidad ≥ 2 (grupo casos), de los cuales 30 presentaron toxicidad severa grado ≥ 3 (grupo casos extremos).

3.3.1.1.2 Grupo control y control extremo

En este grupo se incluyeron a los pacientes que tuvieron al menos cinco años de seguimiento desde la finalización del tratamiento con RT y que no presentaron sangrado rectal ni ningún otro tipo de toxicidad genitourinaria ni gastrointestinal, durante el periodo de estudio. En total se incluyeron 95 pacientes con toxicidades CTCAE = 0 (grupo control).

Para la selección de pacientes controles del análisis del fenotipo de sangrado rectal extremo se consideró también la variable RectV70 (porcentaje de volumen de recto que recibe una dosis de al menos 70 Gy). Esta covariable se encontró asociada al sangrado rectal mediante la evaluación por regresión logística multivariada que incluyó variables clínicas, epidemiológicas, dosimétricas y patológicas. Es decir, se consideraron controles

extremos a aquellos pacientes que, a pesar de haber tenido un porcentaje alto de volumen de RectV70 ($\geq 17,8$ Gy) no presentaron ninguna toxicidad de acuerdo con los CTCAE v.4.0 durante el periodo de estudio. En total, se incluyeron 30 pacientes en este grupo. La **Figura 12** resume la información de los fenotipos de estudio.

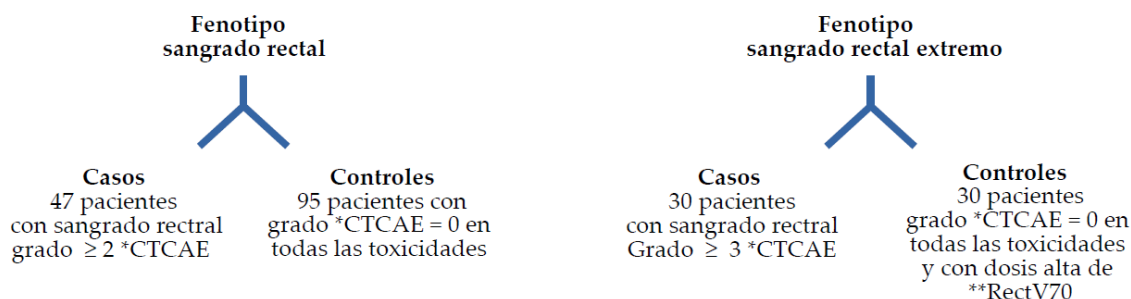


Figura 12. Fenotipos de toxicidad estudiados. *CTCAE, Criterios Comunes de Terminología para Efectos Adversos CTCAE versión 4.0. **RectV70 porcentaje alto de volumen de recto expuesto a una dosis de 70Gy.

3.3.2 Características de los pacientes incluidos

Se incluyeron un total de 142 pacientes con cáncer de próstata de la cohorte RADIOGEN tratados con 3D-CRT. Cuarenta y siete (33,1 %) de estos pacientes presentaron sangrado rectal después del primer año. La edad media de los pacientes al inicio de la radioterapia fue de 71 años. La **Tabla 5** resume las características de los pacientes incluidos.

Tabla 5. Características de los pacientes de la cohorte RADIOGEN-próstata incluidos en el estudio

	Media	Desviación estándar
Edad al tratamiento	71,52	6,58
PSA al diagnóstico	21,13	21,07
	Número	Porcentaje [%]
Sangrado	47	33,1
TURP	9	6,43
Hemorroides	16	11,51
Diabetes	29	20,42
Prostatectomía previa	6	4,23
Gleason score		
≤ 6	42	29,58
7	55	38,73
≥ 8	45	31,69
Estadio al diagnóstico		
T0	1	0,7
T1	14	9,86
T2	85	59,86
T3	40	28,17
T4	2	1,41

TURP, resección transuretral de la próstata; PSA, antígeno prostático específico.
Número, número de pacientes y porcentaje, porcentaje de pacientes.

Respecto al tratamiento, el número de dosis promedio administrada fue de 37 y la dosis promedio por fracción de 1,98 Gy. Alrededor del 90 % de los pacientes recibieron además tratamiento hormonal. La información del tratamiento con radioterapia recibido por los pacientes de la cohorte de RADIOGEN incluidos en este estudio se describe en la **Tabla 6**.

Tabla 6. Características del tratamiento de radioterapia de los pacientes de la cohorte RADIOGEN-próstata incluidos en el estudio

	Media	Desviación estándar
Número de dosis	37,31	1,81
Dosis por fracción (Gy)	1,98	0,05
BED	150,58	6,87
Dosis media de recto (Gy)	46,19	10,45
Volumen de recto (cm3)	57,05	43,81
V50 de recto (%)	48,48	18,39
V60 de recto (%)	31,17	12,71
V65 de recto (%)	22,72	8,81
V70 de recto (%)	13,82	5,99
V74 de recto (%)	3,93	4,12
V75 de recto (%)	1,87	2,97
Volumen de vejiga (cm3)	339,89	164,65
Dosis media de vejiga (Gy)	35,2	13,54
V50 de vejiga (%)	30,32	18,92
V60 de vejiga (%)	20,74	13,28
V70 de vejiga (%)	11,58	8,29
V74 de vejiga (%)	4,6	4,62
Volumen de intestino (cm3)	780,33	421,34
Dosis media de intestino (Gy)	2,91	4,48
V15 de intestino (%)	7,11	12,64
V45 de intestino (%)	0,86	2,32
	Número	Porcentaje [%]
Tratamiento hormonal	128	90,14

BED, dosis biológicamente efectiva. V50, V60, V65, V70, V74 y V75 de recto, porcentaje del volumen del recto que recibe 50, 60, 65, 70, 74 y 75 Gy, respectivamente; V50, V60, V70 y V74 de vejiga, porcentaje del volumen de vejiga que recibe 50, 60, 70 y 74 Gy, respectivamente; V15 y V45 de intestino, porcentaje del volumen de intestino que recibe 15 y 45 Gy, respectivamente. Número, número de pacientes y porcentaje, porcentaje de pacientes.

3.4 PROCESAMIENTO DE LAS MUESTRAS Y SECUENCIACIÓN DE ADN (EXOMAS).

3.4.1 Recogida de muestras

A todos los participantes incluidos en el estudio se les extrajo una muestra de sangre periférica, que fue recolectada en tubos de 10mL con ácido etilendiaminotetraacético (EDTA) como anticoagulante. Todas las muestras se almacenaron a 4°C hasta el momento de la extracción del ADN, que debe realizarse como máximo tres días posteriores a la extracción.

3.4.2 Extracción de ADN

La extracción del ADN de las muestras se llevó a cabo de forma automatizada con el equipo *Chemagic™ Magnetic Separation Module (MSM) I* (Chemagen Biopolymer-Technologie AG) utilizando el kit *Chemagic DNA Blood 100 Kit* (PerkinElmer Inc, Massachusetts, USA).

3.4.3 Control de calidad del ADN

3.4.3.1 Cuantificación del ADN

La cuantificación del ADN se llevó a cabo mediante métodos de espectrometría y fluorimetría. Para la medición por espectrometría se utilizó el equipo *NanoDrop® 1000 Spectrophotometer* (Thermo Fisher Scientific, Massachusetts, USA), este equipo se fundamenta en la Ley de Beer-Lambert de óptica, que se traduciría como la cantidad de luz absorbida es proporcional a la concentración de la muestra. La concentración es calculada mediante absorbancia a 260 nm y la pureza se analiza en los ratios de 260/280 ~ 1,8 y 260/230 ~ 2,2. Por otra parte, la medición por fluorimetría fue realizada usando el equipo *Qubit™ 3.0 fluorometer* (Thermo Fisher Scientific, Invitrogen™), empleando el *Qubit™ dsDNA BR Assay o Qubit® dsDNA HS Kit* (by Thermo Fisher Scientific, Invitrogen™), este equipo detecta la fluorescencia emitida por moléculas tras su unión con el ADN, así la intensidad de la fluorescencia va a estar directamente relacionada con la cantidad de ADN de la muestra, que se compara con la recta patrón.

3.4.3.2 Integridad del ADN

La integridad de las muestras de ADN fue medida con el sistema de electroforesis automatizado *Agilent 4200 TapeStation* (Agilent Technologies) y se usó el kit *Genomic DNA ScreenTape Analysis* (Agilent Technologies). Este sistema nos brinda un informe completo de la calidad de las muestras proporcionando el tamaño, la concentración y la integridad de estas (**Figura 13**).

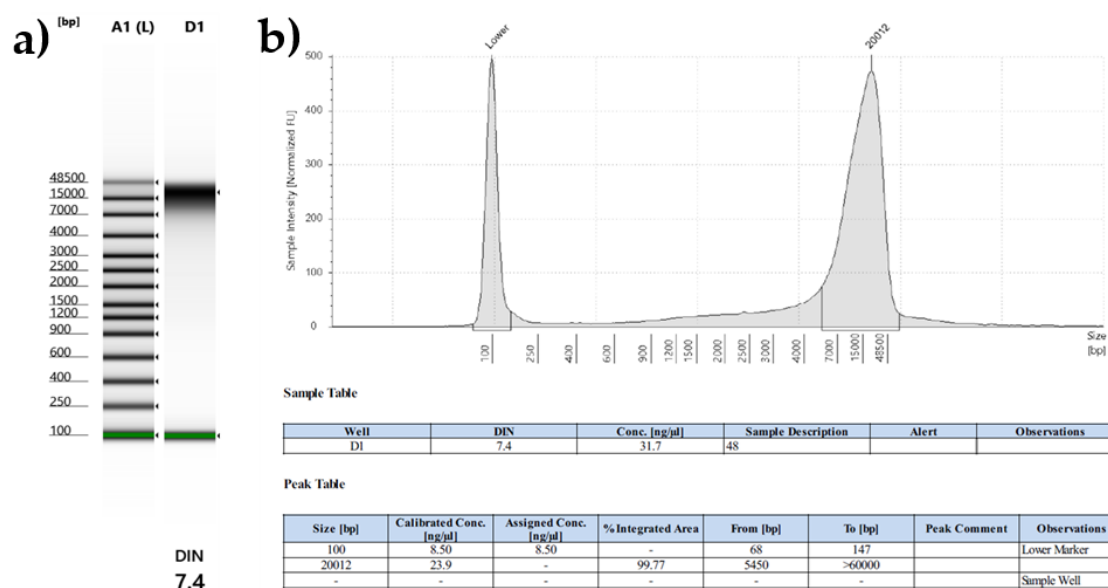


Figura 13. Resultado del análisis del control de calidad de ADN genómico realizado con Agilent TapeStation 4200. a) Electroforesis de la muestra de ADN con su valor de integridad del ADN (DIN). b) grafica del tamaño de los fragmentos de ADN en pares de bases y tablas con los resultados obtenidos del análisis (información sobre el DIN, la concentración en ng/μl, el tamaño de los fragmentos).

3.4.4 Secuenciación de ADN (exomas)

La preparación de librerías se realizó de acuerdo con el protocolo *SureSelect QXT Target Enrichment for Multiplexed Sequencing* (Agilent Technologies, Santa Clara, CA USA), para el enriquecimiento de las librerías se usó el *SureSelect All Exon V6+UTR* de Agilent (90,69 Mpb de captura) (Agilent Technologies, Santa Clara, CA USA). La secuenciación se llevó a cabo en el sistema *NovaSeq 6000 de Illumina* (Illumina, Inc.) y se obtuvo una cobertura promedio de 74x.

3.4.4.1 Creación de librerías

La preparación de librerías inicia con la fragmentación enzimática del ADN genómico mediado por trasposasas y la unión de adaptadores en los extremos de los fragmentos. Como paso siguiente se purificaron los fragmentos ligados con los adaptadores con *AMPure XP beads* (Beckman Coulter™), este sistema utiliza esferas paramagnéticas que sirven para seleccionar el ADN por tamaño y eliminar los contaminantes. Seguido del paso anterior se realiza la amplificación, purificación y control de calidad de las librerías unidas a los adaptadores, las librerías en este paso deberán tener entre 245 – 325 pb (pares de bases). El proceso de purificación con *AMPure XP beads* y el control de calidad de las librerías (kits *dsDNA HS Assay* y *D1000 ScreenTape Assay*) se repitió entre pasos para garantizar la calidad de las muestras. A continuación, se lleva a cabo la hibridación y la captura de las librerías. La hibridación de la librería se realiza con el *SureSelect All Exon V6+UTR*, este kit está constituido por un conjunto de sondas marcadas con biotina, que posteriormente son capturadas por

perlas magnéticas recubiertas con streptavidina, en este caso usamos las *Dynabeads My One Streptavidin T1*. Una vez realizada la captura de las librerías se procede a realizar el indexado de las muestras mediante PCR, estos index identificarán a cada muestra en el pool, posteriormente las librerías indexadas fueron purificadas y se realizó el control de calidad final usando para esto los kits *dsDNA HS Assay* y *D1000 ScreenTape Assay*. Las librerías indexadas que pasaron el control de calidad deberían estar posicionadas entre 325 y 450 bp.

Para la preparación de los pools se agregó 5 µl de cada muestra de las librerías previamente normalizada a 10 nM. Los pools de las librerías fueron desnaturalizados con 0,2 NaOH, la reacción se neutralizó con 400 mM de Tris-HCl y posteriormente se realizó la secuenciación.

3.4.4.2 Secuenciación

La secuenciación se llevó a cabo en el sistema *NovaSeq 6000 de Illumina* (Illumina, Inc.), basado en un proceso químico de secuenciación por síntesis (SBS). Como paso inicial se realiza la generación de *clusters* en la superficie de las celdas de flujo, estos *clusters* son generados a partir de la amplificación clonal de fragmentos individuales de ADN por medio de una PCR en puentes. A continuación, se realizó la secuenciación por síntesis basada en el uso de desoxirribonucleótidos trifosfato (*dNTPs*) marcados fluorescentemente (terminador reversible), estos nucleótidos son detectados a medida que se incorporan a la cadena de ADN. Debido a la intensidad y longitud de onda que emite cada fluoróforo los nucleótidos van a ser identificados en el punto de incorporación en cada ciclo de síntesis del ADN. Al ser solo cuatro nucleótidos los que compiten durante cada ciclo por el sitio el porcentaje de seguridad de la incorporación del nucleótido correcto es alto.

3.5 ANÁLISIS BIOINFORMÁTICO

Esta sección abarca el control de calidad de los archivos crudos FASTQ, la detección de variantes a partir de la *pipeline* estándar de análisis genómico GATK versión 4.0 (*Genome Analysis Toolkit*) (**Figura 15**)¹⁷²⁻¹⁷⁴, la anotación de las variantes con ANNOVAR¹⁷⁵ y el control de calidad con PLINK¹⁷⁶ (**Figura 14**).

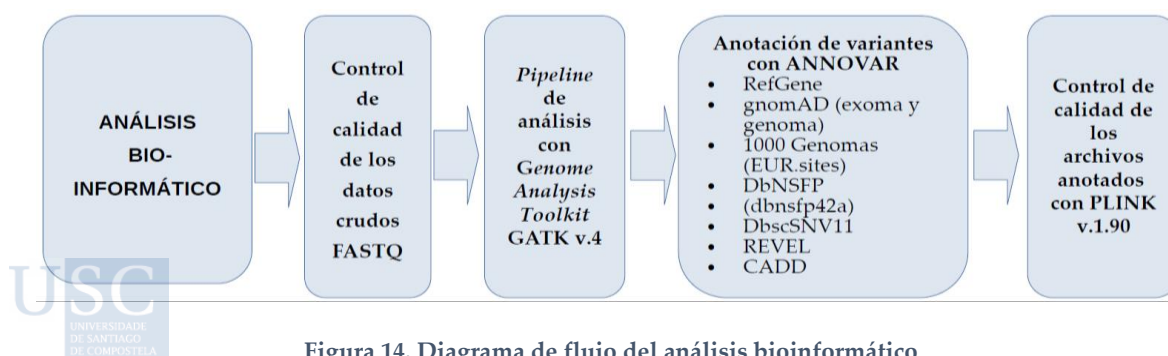


Figura 14. Diagrama de flujo del análisis bioinformático

3.5.1 Control de calidad de los datos crudos FASTQ

Una vez terminada la secuenciación se obtuvieron los datos crudos de las secuencias almacenados en archivos FASTQ. Estos datos se preprocesaron con el programa *FastP*¹⁷⁷, con el cual se realizó el control de calidad de las secuencias y la eliminación de los adaptadores añadidos durante la preparación de las librerías.

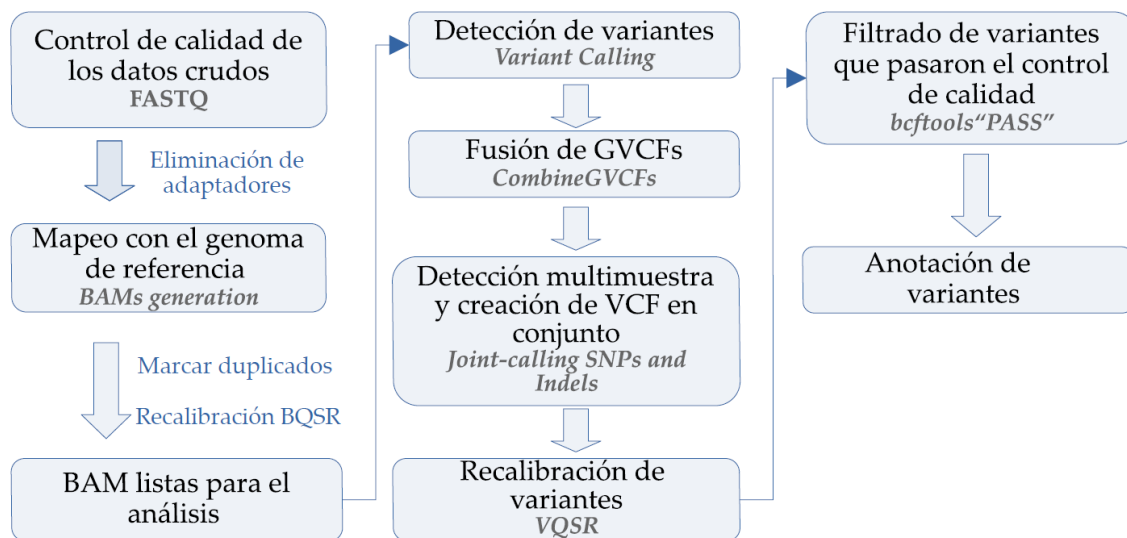


Figura 15. Flujo de trabajo de acuerdo con Genome Analysis Toolkit GATK v4.0 para la detección de variantes puntuales (SNPs e Indels) en línea germinal. La *pipeline* parte de los datos FASTQ preprocesados limpios que son alineados con el genoma de referencia, hasta la obtención de un VCF con las variantes que pasaron el control de calidad, estas variantes posteriormente pueden ser anotadas

3.5.2 Mapeo con el genoma de referencia

Los archivos FASTQ resultantes del preprocesado fueron alineados con el genoma de referencia hg19/GRCh37 (human_g1k_v37.fasta) por medio de la herramienta *Burrows-Wheeler Aligner* (BWA)¹⁷⁸, obteniendo como resultado un archivo BAM (*Binary Alignment Map*) para cada muestra.

3.5.3 Recalibrado de las variantes y aplicación de umbrales de calidad

Un paso importante en el procedimiento es la revisión de la calidad de las variantes. Para esta tesis dicha revisión se llevó a cabo primero recalibrando la calidad de las variantes con la herramienta de recalibración estándar de puntaje de calidad de variante (*Variant quality score recalibration* VQSR¹⁷³).

La herramienta VQSR¹⁷³, ajusta y mejora la calidad de las variantes detectadas, creando un subconjunto de *calls* que posee el nivel de calidad, el equilibrio de precisión y de sensibilidad estándar adecuados. Del filtrado se obtiene un archivo VCF (*Variant call*

format) recalibrado en el que cada variante se anota con su VQSLOD score (*variant quality score log-odds*, razón de probabilidad logarítmica de que una variante sea verdadera frente a que sea falsa). Si la puntuación está por debajo del nivel de calidad deseado será marcado como filtrado. El filtrado VQSR se realizó en dos pasos: en primer lugar, se utilizó la herramienta *VariantRecalibrator*, con la cual se crea un modelo de recalibración para puntuar la calidad de las variantes asignando el VQSLOD; posteriormente con la herramienta *ApplyRecalibration*, se filtraron las variantes en base a los puntos de corte (*score cutoffs*). Es decir, se aplicó un umbral de filtrado del 99,9 % de sensibilidad y sólo las variantes con un score superior al umbral pasaron este filtro, permitiendo con esto equilibrar la sensibilidad (variantes reales) y la especificidad (limitación de falsos positivos). Al final de este filtrado se obtuvo un fichero VCF con las muestras recalibradas. De estas, se seleccionaron solo aquellas variantes que pasaron el filtro de VQSR “PASS”.

Como paso final, con la herramienta *bcftools*^{179,180} se eliminaron las variantes monomórficas.

3.5.4 Marcaje de secuencias duplicadas y recalibración de la calidad de las bases leídas

A continuación, con la herramienta *MarkDuplicates* del programa Picard¹⁸¹ se marcaron los duplicados originados durante la amplificación por PCR de las librerías, esta herramienta diferencia la lectura primaria de las duplicadas mediante un algoritmo que clasifica la suma de sus puntuaciones de calidad, la lectura con la suma de calidad de bases más alta se determinará como lectura primaria el resto se marcarán como duplicados y serán descartadas. La recalibración de la calidad de las bases (*Base Quality Score Recalibration* BQSR) se llevó a cabo a través de GATK^{172–174}. En este paso se detectan los errores cometidos por el secuenciador cuando estima la precisión de cada llamada de base, debido a esto ajusta los valores de calidad de toda la base para mejorar la precisión de las llamadas de variantes.

3.5.5 Detección de variantes

En el siguiente paso se realizó la detección de variantes o “*variant calling*” por muestra, por medio de un montaje local *de-novo* de haplotipos, usando el programa *HaplotypeCaller*¹⁷². Este programa identifica variantes (SNPs e Indels) y las define como regiones activas (regiones donde se detecta variación), en ese sitio excluye la información del mapeo existente e identifica probables haplotipos que se encuentren en las lecturas, reensambla de nuevo las lecturas en esa región, ordenando cada haplotipo con el haplotipo de referencia e identificando con esto a las posibles variantes. De aquí, se obtienen las probabilidades de alelos para cada sitio de variante candidata, que son evaluadas con el teorema de probabilidades de Bayes para determinar el genotipo más

probable de la muestra en cada sitio. De este procedimiento se obtiene un archivo intermedio gVCF (*Genomic Variant Call Format*) por muestra.

3.5.6 Fusión de GVCFs y “joint-calling”

Con el programa *CombineGVCFs*¹⁷³ se fusionaron los 142 archivos GVCFs en un solo GVCF multimuestra. La detección multimuestra “*joint-calling*” se llevó a cabo con la herramienta *GenotypeGVCFs*. La herramienta calcula las probabilidades de detección correctas, reajusta el registro recién fusionado y atribuye filtros de calidad, para generar un archivo VCF que contiene los SNPs e Indels conjuntamente (*joint-called SNP and Indels*). Este análisis permite aumentar la sensibilidad de detección de variantes que incluso se encuentren en sitios difíciles de detectar, además proporcionará información sobre todos los sitios de interés de todas las muestras.

3.5.7 Anotación de variantes

Después del filtrado, se realizó la anotación de variantes con el programa *table_annoar.pl* de la herramienta *ANNOVAR* (V.2020)¹⁷⁵. Esta herramienta permite revisar diferentes bases de datos y con ello anotar variantes de un solo nucleótido (*single nucleotide variants* SNVs), pequeñas inserciones /deleciones y evaluar el impacto biológico de las variantes y de los genes en los que se encuentran estas variantes. Las bases de datos que utilizamos para la anotación fueron: i) RefGene v.2¹⁸², esta base anota las variantes en los genes y especifica los genes codificantes y no codificantes de proteínas en humanos; ii) gnomAD v.2.1.1¹⁸³, cuenta con información de secuenciación de exomas y genomas de diferentes proyectos en conjunto y proporciona información de la frecuencia de las variantes; iii) 1000 Genomas-EUR.sites v.2015¹⁰⁴, esta base posee información de la frecuencia alélica de variantes en el genoma de la población europea; iv) dbNSFP v.4.2a(dbnsfp42a)¹⁸⁴, esta base es usada para anotar predicciones funcionales, entre los predictores que incluye se encuentran: PolyPhen2 HDIV; PolyPhen2 HVAR; LRT; MutationTaster, MutationAssessor; FATHMM (FATHMM-XF, predice mutaciones patogénicas puntuales)¹⁸⁵, CADD, GERP++ y DANN entre otras; v) dbSNP v.1.1, permite la predicción de sitios que alteran el *splicing* mediante los puntajes de Ada Boost y Random Forest, vi) CADD v.1.6 (*Combined Annotation Dependent Depletion*) prioriza cuantitativamente las variantes funcionales, perjudiciales y causales de enfermedades¹⁸⁶; vii) REVEL (*Rare exome variant ensemble learner*) este algoritmo incorpora 18 predictores de patogenicidad de variantes genéticas, combina múltiples características de las variantes, como la conservación evolutiva, la anotación funcional y la frecuencia poblacional¹⁸⁷.

3.5.8 Control de calidad general

Previo a la realización de los estudios de asociación se llevó a cabo el control de calidad de los archivos anotados. El control de calidad se llevo a cabo con el programa PLINK v.1.90¹⁷⁶ y con el lenguaje de programación de R¹⁸⁸.

El control de calidad consiste en:

1. Exclusión de las variantes faltantes en más del 5 % de los pacientes. Las variantes con tasas de detección faltantes fueron eliminadas por medio del filtro `--geno`, utilizando el valor `--geno 0,05`.
2. Exclusión de individuos con alta tasa de variantes faltantes. Los individuos con pérdidas de variantes mayores al 5 %, fueron descartados del estudio. Filtro usado `--mind 0,05`.
3. Eliminación de variantes que se desvían del equilibrio Hardy-Weinberg, debido a que pueden indicar problemas en la detección de las variantes. El filtro utilizado en nuestro análisis fue de `--hwe 1e-06`.
4. Exclusión de individuos con alta o baja tasa de heterocigosidad (± 3 desviaciones estándar). En este punto se excluyen individuos con valores superiores de ± 3 desviaciones estándar de la tasa media de heterocigosis.
5. Grado de parentesco, en este paso se calcula la identidad por descendencia (*Identity-by-descent* IBD) de todas las muestras en pares usando `--genome`, además con el filtro `--min` se establece un umbral y se crea un listado con individuos con parentesco familiar por encima del umbral deseado, para este estudio se eligió un $\pi\text{-hat} > 0,25$.
6. En este punto se evaluó la existencia de estratificación en la población de estudio mediante el análisis de componentes principales (*Principal Components Analysis* PCA)^{189,190}. El componente principal 1 representa la mayor varianza de los datos y se usa para excluir a aquellos individuos con desviaciones significativas a la media. En este paso se eliminan los individuos que fallaron el componente principal 1 (CP1) con ± 5 desviaciones estándar.

Los individuos y las variantes que pasaron el control de calidad de PLINK fueron incluidos para la realización de los análisis de esta tesis. En total se incluyeron 141 pacientes (46 casos y 95 controles) y 1 143 963 variantes.

3.6 SELECCIÓN DE COVARIABLES

Con el fin de seleccionar de covariables relacionadas con el sangrado rectal, se evaluaron 122 variables de la cohorte de próstata de RADIOGEN. Entre ellas se incluyen la antropometría, comorbilidades y su manejo terapéutico, datos patológicos

del tumor, tratamiento oncológico y datos dosimétricos de radioterapia. Primero se eliminaron las covariables con varianza igual a cero, luego se realizó un análisis de correlación, a partir del cual se eliminaron las variables relacionadas entre sí (*cutoff* de 0,3). Se aplicó un modelo de regresión logística univariante y un modelo de selección automática paso a paso de variables predictoras (*stepwise bidirectional elimination*). Esta regresión es una combinación del modelo *forward selection* y *backward elimination*. El modelo de *forward selection* prueba cada variable a medida que se agrega, manteniendo sólo las que se considera estadísticamente significativas; por su parte, el *backward elimination* incluye todas las variables independientes, elimina una a la vez y analiza el modelo para ver si la variable eliminada es estadísticamente significativa. Se consideraron como covariables aquellas variables con $p\text{-valor} \leq 0,05$. Este análisis identificó como asociada al fenotipo al volumen de recto 65 RectV65 ($P\text{-valor}$ 0,04), que es el porcentaje del volumen del recto que recibe al menos 65 Gy. Además de la covariable RectV65, se incluyó también en los análisis la covariable edad, debido a que anteriores estudios han reportado a esta covariable relacionada con mayor predisposición de toxicidad a la radioterapia.

3.7 ANÁLISIS DE VARIANTES Y GENES PREVIAMENTE DESCRITOS

Se realizó una búsqueda en la base de datos de PubMed de publicaciones que incluyeran genes y variantes genéticas asociados con efectos adversos a la radiación ionizante. Las palabras clave utilizadas para la búsqueda fueron: “*radiation*”, “*radiotherapy*”, “*toxicity*”, “*adverse effects*”, “*SNP*”, “*genes*”, “*prostate*”, “*cancer*”, “*rectal toxicity*”, “*prostate*”, “*rectal bleeding*”, “*overall toxicity*”.

Se identificaron alrededor de 114 artículos, 40 de ellos incluían estudios de asociación de variantes genéticas, de ellos se seleccionaron las variantes con $p\text{-valor}$ corregido por Bonferroni menor a 0,05 o aquellas con $p\text{-valor}$ al metaanálisis o combinado menor de 10^{-05} . Posteriormente se revisó en nuestros datos de exomas la presencia de las variantes que cumplieron con el $p\text{-valor}$ requerido para incluirlas en los análisis de variantes previamente descritas. Los diferentes estudios de asociación de variantes y genes previamente descritos incluyeron a 141 pacientes de la cohorte de próstata de RADIOGEN (46 casos y 95 controles).

3.7.1 Estudios de asociación univariable

El análisis de asociación de las variantes previamente asociadas a radiotoxicidad y presentes en nuestros datos de exomas (ver **Tabla 8**), se realizó con el test de Chi-cuadrada (X^2) de la herramienta PLINK v.1.9. La prueba de X^2 es una prueba especial para rasgos binarios. PLINK realizó una prueba alélica de X^2 con un grado de libertad para comparar la frecuencia de alelos entre casos y controles. Los resultados de la prueba de X^2 se ajustaron por test múltiple.

La fórmula para calcular de Chi-cuadrada es:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

En donde:

Σ = la suma

O = frecuencia observada

E = frecuencia esperada

$i = \{1, \dots, n\}$ n número de variantes

3.7.2 Estudios de asociación multivariable

Al igual que para el análisis univariable (ver apartado 3.7.1), para este análisis se incluyeron las variantes previamente asociadas a radiotoxicidad presentes en nuestros datos de secuenciación de exomas (ver **Tabla 8**).

El test de asociación se realizó mediante un modelo de regresión logística multivariable glm, que analiza un SNP a la vez atribuyendo a cada uno un p-valor. Se seleccionó este modelo de regresión logística debido a que permite la inclusión de covariables y es adecuado para fenotipos binarios, como es el nuestro. Las covariables incluidas en el modelo fueron el volumen de recto 65 (RectV65) y la edad. Los resultados fueron corregidos por test múltiple. Tanto el análisis de asociación como las correcciones múltiples se realizaron con la herramienta PLINK v 2.0¹⁷⁶.

Modelo glm para fenotipos binarios

$$G\beta_G + X\beta_X$$

En donde:

G es la matriz de genotipo/dosis para la variante actual

X es la matriz de covariables fijas.

3.7.3 Validación de SNPs previamente asociados con sangrado rectal

Para validar las variantes asociadas con sangrado rectal previamente descritas, investigamos dichas variantes en los datos de la secuenciación de los exomas (ver **Tabla 8**). En este trabajo se considera validación al proceso de confirmar y respaldar los hallazgos iniciales obtenidos de un estudio de asociación, independientemente de la ancestralidad de las poblaciones de estudio, de la definición del fenotipo, de la

estrategia de reclutamiento o del muestreo al momento de la investigación¹⁹¹. Para la validación, se realizó un análisis de regresión logística *glm* multivariable. Las covariables usadas fueron las descritas en los estudios de referencia. El análisis de asociación y las correcciones múltiples se realizaron con la herramienta PLINK v 2.0.

3.7.4 Nuevas variantes en genes previamente asociados con toxicidad a la radioterapia.

Considerando las 152 variantes previamente identificadas en la bibliografía asociadas con toxicidades a la radioterapia (ver apartado 3.7), seleccionamos únicamente aquellos genes o regiones que tuvieran variantes estadísticamente significativas con un p-valor al metaanálisis o combinado menor que 5×10^{-08} o que tuvieran un p-valor corregido por Bonferroni $< 0,05$ (ver **Tabla 12**).

Para el análisis de los genes incluidos se consideraron únicamente las variantes de impacto funcional. La priorización de variantes se realizó seleccionando en primer lugar variantes raras y de baja frecuencia, refiriéndonos a ellas a partir de ahora en esta tesis como variantes de baja frecuencia. Se consideraron variantes de baja frecuencia aquellas con frecuencia menor a 0,05 en la población europea, de acuerdo con las bases gnomAD_exoma, gnomAD_genoma y 1000G EUR.sites. A partir de estas variantes se seleccionaron aquellas de impacto funcional, considerando variantes *nonsense*, *frameshift Indels*, *missense* y de *splicing* ± 2 bp. Para priorizar estas variantes se usaron los predictores REVEL (seleccionamos variantes con scores $> 0,5$), CADD (seleccionamos $CADD_{phred} \geq 20$), FATHM-XF (seleccionamos variantes dañinas “D”)¹⁸⁵, y ADA_SCORE (seleccionamos variantes con scores $\geq 0,6$). Se incluyeron las variantes que cumplieran con alguna de las condiciones de los predictores.

El análisis de asociación de las variantes de impacto funcional con el fenotipo sangrado rectal se llevó a cabo mediante un modelo de regresión logística multivariable *glm*, utilizando como covariables RectV65 y edad. Los resultados de cada análisis fueron corregidos por test múltiple. El análisis de asociación y las correcciones por test múltiple se realizaron con la herramienta PLINK v.2.0.

3.8 ESTUDIOS DE ASOCIACIÓN BASADOS EN GENES

El análisis WES se realizó considerando las variantes de baja frecuencia ($< 0,05$) como se indica en la **Figura 16**. Estudiamos dos fenotipos, uno de sangrado rectal, (casos grado ≥ 2 CTCAE v.4.0 y controles grado = 0 CTCAE v.4.0) y otro fenotipo en el que se seleccionaron los individuos más extremos en cada uno de los grupos analizados, y que llamamos fenotipo sangrado rectal extremo (casos grado ≥ 3 CTCAE v.4.0 y grado = 0 CTCAE v.4.0 + RectV70 controles). Los estudios fueron realizados en dos fases, la fase de descubrimiento y la fase de validación de los genes más asociados con el fenotipo. Ambas fases incluyeron análisis con tres diferentes estrategias de filtrado de variantes.

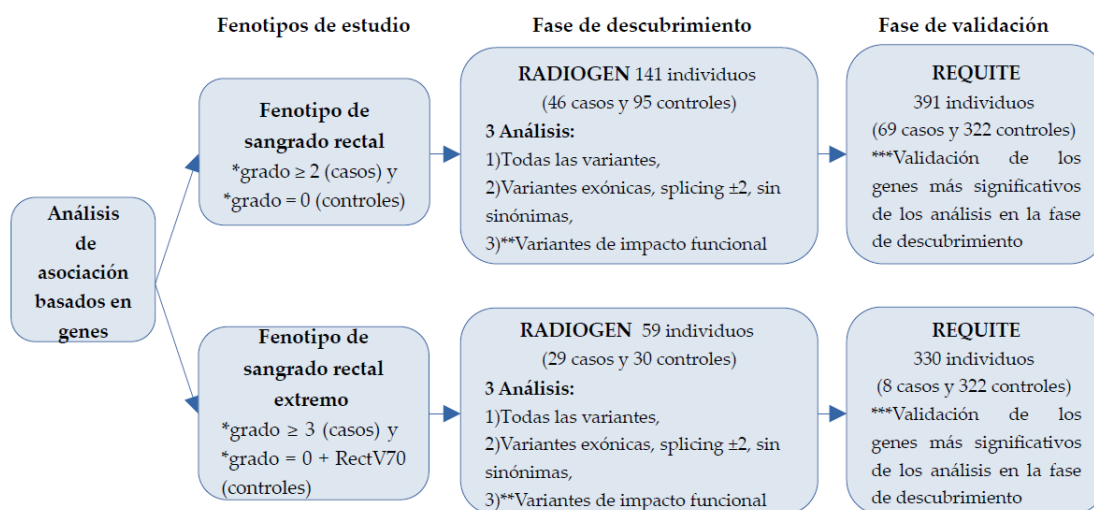


Figura 16. Esquema de análisis. Primero se realizaron los análisis en la fase de descubrimiento en la cohorte RADIOGEN. Posteriormente se realizó la validación de los genes más significativos en la cohorte REQUITE. *CTCAE v.4.0, **el análisis de asociación incluye variantes de impacto funcional en exomas y en sitios de *splicing* canónicos ± 2 pb y excluye variantes sinónimas, ***Los análisis de la fase de validación se llevaron a cabo siguiendo los mismos criterios que en sus respectivas fases de descubrimiento.

3.8.1 Análisis de asociación de genes para el fenotipo sangrado rectal y para el fenotipo de sangrado rectal extremo

El análisis de asociación para el estudio del fenotipo sangrado rectal se realizó en dos fases: la fase de descubrimiento en la cohorte RADIOGEN y la fase de validación en la cohorte externa REQUITE¹⁹².

3.8.1.1 Análisis de asociación de genes en RADIOGEN

Los análisis de asociación de la fase de descubrimiento en la cohorte RADIOGEN incluyeron un total de 141 pacientes (46 casos CTCAE v.4 ≥ 2 y 95 controles CTCAE v.4 = 0) para el fenotipo sangrado rectal y 59 pacientes (29 casos CTCAE v.4 ≥ 3 y 30 controles CTCAE v.4 = 0 y RectV70) en el caso del fenotipo de sangrado rectal extremo. En todos los análisis se consideraron exclusivamente genes que tenían disponibles al menos dos variantes de baja frecuencia por gen. Al igual que en el apartado 3.7.4 se consideraron como variantes de baja frecuencia a aquellas que en población europea tuvieron una frecuencia menor a 0,05 de acuerdo con las bases gnomAD_exoma, gnomAD_genoma y 1000G (EUR.sites). La asociación de los genes con el sangrado rectal como efecto adverso a la radioterapia se estudió siguiendo tres estrategias de filtrado de variantes:

- 1) En la primera estrategia se incluyeron las variantes presentes en todas las regiones secuenciadas.

- 2) En el segundo análisis se incluyeron únicamente variantes exónicas (excluyendo las variantes sinónimas) y en sitios canónicos de *splicing* ± 2 pb.
- 3) En la tercera estrategia, a partir de las variantes del segundo análisis, se priorizaron las variantes de alto impacto funcional. La priorización incluyó variantes *nonsense*, *frameshift indels*, *missense* y *splicing* ± 2 pb. Los filtros utilizados fueron: REVEL ($> 0,5$), CADD (CADD_phred ≥ 20), FATHM-XF (variantes dañinas “D”) y ADA_SCORE ($\geq 0,6$). Se incluyeron las variantes que cumplieran con al menos uno de los filtros utilizados.

La priorización y filtrado de las variantes se llevó a cabo con el *software* MATLAB v.R2021b¹⁹³. A partir de los archivos VCF filtrados para cada estrategia se crearon los archivos *bfiles* con la herramienta PLINK v.2.0, que se utilizaron para el análisis de SKAT-O. Para llevar a cabo los tests de asociación, se usó el programa SKAT y el entorno de R v.4.1.1.

SKAT-O evalúa el efecto acumulado de un conjunto de variantes raras en un mismo gen. Este test es una combinación de la prueba de colapso de variantes de Burden y de las estadísticas de puntaje de variante individual de SKAT, que evalúa los promedios ponderados de SKAT y pruebas de Burden. SKAT-O se comporta automáticamente como un Burden test cuando éste es más poderoso que SKAT (cuando una gran proporción de las variantes son causales y sus efectos van en la misma dirección) y como SKAT, cuando éste es más poderoso (cuando hay presencia de variantes que aumentan y disminuyen los rasgos o sólo hay una pequeña fracción de variantes causales)¹⁴². La ecuación de SKAT-O es:

$$Q_{\rho} = (1 - \rho)Q_S + \rho Q_B$$

En donde:

Q_S es el estadístico del test SKAT

Q_B es el estadístico del Burden test

P es el p valor

Los análisis de asociación con SKAT-O utilizaron el método optimizado para muestras pequeñas “optimal.adj”. Además, se seleccionó el modelo para fenotipos binarios y se usaron los pesos beta predeterminados (*weights.beta* = c [1, 25]). Las covariables “edad” y “RectV65” fueron incluidas en todos los análisis. Después del análisis, revisamos los resultados obtenidos para excluir aquellos genes que, a pesar de haber sido incluidos por presentar dos o más variantes, solo una de ellas había sido utilizada en el test de asociación. La corrección de los resultados se llevó a cabo con el test Benjamini-Hochberg FDR (BH-FDR). Todos los análisis de asociación se realizaron usando los mismos criterios.

3.8.1.2 Validación de genes en la cohorte externa REQUITE

Los genes más significativos (con un p-valor $< 10^{-04}$) obtenidos de los análisis de asociación para el fenotipo de sangrado rectal y el fenotipo de sangrado rectal extremo, estudiados en la fase de descubrimiento en la cohorte RADIOGEN, se validaron en una cohorte externa. En este caso se utilizó la cohorte de pacientes de cáncer de próstata del estudio REQUITE (*validating predictive models and biomarkers of radiotherapy toxicity to reduce side effects and improve QUALITY of life in cancer survivors*)¹⁹².

REQUITE es un estudio internacional-multicéntrico cuyo objetivo fue validar modelos y biomarcadores que permitan predecir el riesgo de sufrir efectos adversos posteriores a la radioterapia en pacientes con cáncer de próstata, mama y pulmón. En este estudio participaron pacientes de 26 hospitales en ocho países (Bélgica, Francia, Alemania, Italia, Países Bajos, España, Reino Unido, EE. UU.). El estudio incluyó 2069 pacientes con cáncer de mama, 1808 de próstata y 561 de pulmón. Las muestras se genotiparon con el chip Illumina OncoArray-500K-C (480 083 SNPs), y los genotipos fueron imputados utilizando como panel de referencia 1000 Genomas (versión 3)^{104,194}. Así, la cohorte dispone de genotipos imputados para 4223 pacientes con ascendencia europea (1948 de mama, 1728 de próstata, 547 de pulmón). Además, se realizó el seguimiento prospectivo pre y post radioterapia, con el que se recopilaron datos de comorbilidades, tratamientos (regímenes de RT, dosis en órganos de riesgo) y toxicidades a la RT (medidas de acuerdo con los CTCAE v4.0), entre otros.

Para la validación de los resultados de la fase de descubrimiento del fenotipo sangrado rectal se incluyeron 391 pacientes (69 casos, pacientes que presentaron sangrado rectal ≥ 2 CTCAE v.4.0 y 322 controles, pacientes sin toxicidad CTCAE v.4.0 grado = 0). En cuanto a la validación de los resultados del fenotipo de sangrado rectal extremo, fueron incluidos un total de 330 pacientes (8 casos, pacientes que presentaron sangrado rectal ≥ 3 CTCAE v.4.0 y 322 controles, pacientes sin toxicidad CTCAE v.4.0 grado = 0). Para el análisis, utilizamos los datos de genotipado que pasaron el control de calidad descrito por Seibold et al. (2019)¹⁹².

Para cada gen, se seleccionaron las variantes genotipadas/imputadas localizadas dentro de las mismas regiones secuenciadas en nuestros exomas, para lo que consideramos los BED files de nuestra librería de exomas (*SureSelect Human All Exon V6+UTR*), para el genoma de referencia *H. sapiens* hg19. Las variantes que se localizan dentro de las regiones de los genes de interés ± 1500 pb fueron filtradas por la frecuencia alélica $< 0,05$.

La preparación de los de los bfiles y los análisis de validación con SKAT-O se llevarón a cabo siguiendo las mismas estrategias de análisis, el mismo método, los mismos pesos (*weights.beta* = c [1, 25]) y las mismas covariables utilizadas en la cohorte de descubrimiento (ver apartado 3.8.1.1). Se consideró como genes validados a aquellos que presentaron p-valor menores a 0,05 y que tuvieran al menos dos variantes analizadas.

3.8.2 Evaluación conjunta de variantes comunes y raras

Este análisis se realizó con pacientes de la cohorte RADIOGEN. Se incluyeron se incluyeron 141 pacientes y 1 143 365 variantes.

Para analizar la posible asociación del efecto combinado de las variantes raras y comunes en un mismo gen con nuestro fenotipo de toxicidad, se utilizó la función *Combined test of common rare variants* del paquete SKAT¹⁴³ en el entorno de R. Para este análisis primero se seleccionó un punto de cohorte para definir las variantes en variantes raras ($MAF \leq 0,05$) y en comunes ($MAF > 0,05$). Posteriormente se llevó a cabo el análisis de las variantes raras y de las comunes, con el análisis SKAT-C. El test evalúa por separado la contribución de las variantes raras y comunes y posteriormente combina los p-valores usando diferentes pesos para cada tipo de variante. Los pesos usados para el test combinado fueron: para las variantes raras (*weights.beta.rare* = c [1, 25]) y para las variantes comunes (*weights.beta.common* = c [0,5, 0,5]). Para el análisis sólo se incluyeron genes que tuviesen disponibles al menos dos variantes. Posteriormente, de los resultados obtenidos sólo se consideran aquellos con al menos dos variantes analizadas. Los datos resultantes de este análisis se corrigieron con el test BH-FDR.

3.9 NUEVAS APROXIMACIONES DE ESTUDIOS DE ASOCIACIÓN ÓMICOS APLICADOS A RADIOGENÓMICA y ANÁLISIS DE RUTAS

3.9.1 Estudio de asociación del transcriptoma completo

En este estudio se incluyeron 141 pacientes de la cohorte de RADIOGEN (46 casos y 95 controles).

El análisis de TWAS se realizó en cuatro etapas: en primer lugar, se realizó el estudio de asociación de las variantes comunes ($MAF > 0,05$) con el test de chi-cuadrado de la herramienta PLINK v.1.9. Se incluyeron 244 983 variantes.

A continuación, se realizaron los análisis de asociación del transcriptoma completo por tejido con la herramienta S-Predixcan¹⁹⁵. Para ello, se utilizaron 46 modelos de predicción del transcriptoma calculados mediante el método de red elástica (*Elastic Net*) de GTEx v8¹⁵⁴, obtenidos del repositorio de datos de PredictDB¹⁹⁶ (<https://predictdb.org/>). GTEx v8 cuenta con modelos de expresión para 49 tejidos, con datos de 838 donantes sanos y 15,201 muestras de ARN. Para nuestro análisis, utilizamos 46 tejidos de 715 donantes europeoamericanos y 12 423 muestras de ARN. Del mismo repositorio se obtuvieron los archivos de covarianza de los SNPs incluidos en cada modelo génico. Estos, son archivos de referencia del desequilibrio de ligamiento (LD). Los modelos predictivos de expresión por tejido fueron aplicados a los resultados de asociación resumidos por variante (*summary-statistics*) con la herramienta S-Predixcan¹⁹⁵.

Posteriormente, se llevó a cabo un análisis multi-tejido con el fin de correlacionar la expresión génica entre los tejidos, por lo que este análisis puede considerarse un metaanálisis entre tejidos¹⁵¹. Para esto se utilizó la herramienta S-MultiXcan¹⁵¹. El modelo y el archivo de covarianza se obtuvieron de PredictDB. Aquí el modelo predictivo *Elastic Net* se aplicó a los resultados de asociación uni-tejido, incluyendo solo los tejidos adyacentes a la próstata o componentes sanguíneos para los que existiera modelos de predicción y archivos de covarianza. Los modelos de tejidos incluidos fueron: próstata, colon sigmoide, intestino delgado-íleon, piel suprapúbica, testículos, omento adiposo visceral y sangre completa.

Por último, se realizó un segundo análisis multi-tejido, utilizando el mismo modelo de predicción y de covarianza que para el análisis anterior. En este caso, el modelo predictivo fue aplicado a los resultados de asociación uni-tejido de 46 diferentes tejidos de GTEx V8 (ANEXO 8), excluyendo de nuestro análisis a los tejidos específicos de órganos femeninos. Los resultados de los análisis con S-MultiXcan fueron corregidos con Bonferroni. Además, se aplicó un ajuste adicional recomendado ($p_{i_best} \leq 10^{-04}$)¹⁵¹ para reducir el número de falsos positivos debido a la especificación errónea de desequilibrio de ligamiento (LD) derivada del uso de estadísticas de asociación resumidas a nivel de SNPs.

3.9.2 Estudios de asociación del proteoma completo

El análisis PWAS se llevó a cabo con pacientes de la cohorte RADIOGEN. Se incluyeron 141 individuos (46 casos y 95 controles). Se utilizaron los mismos resultados de asociación resumidos por variante “*summary-statistics*”, utilizados en el análisis de TWAS (ver apartado 3.9.1).

Los modelos predictivos utilizados y el archivo de covarianza fueron los de proteínas plasmáticas de Zhang¹⁶², obtenidos de <https://zenodo.org/record/6332981> y que cuenta con datos de 4657 proteínas plasmáticas de 7.213 individuos europeos americanos y 1871 afroamericanos. Dichos modelos fueron aplicados a nuestros “*summary-statistics*” con la herramienta S-Predixcan en el entorno de Python. Los resultados del análisis de PWAS fueron corregidos por FDR (*False Discovery Rate*) ($q < 0,05$).

3.9.3 Análisis de rutas

El análisis de rutas se llevó a cabo para identificar las principales vías en las que se encuentran relacionados los genes con los p-valor nominales más significativos de los análisis de asociación SKAT-O, SKAT-C- *common-rare*, TWAS y PWAS (Anexo 9). Para ello, utilizamos dos estrategias de análisis.

En la primera estrategia examinamos si estos genes formaban parte de alguna ruta de interés previamente relacionada con radiotoxicidad. Las rutas investigadas fueron la ruta de la inflamación, la ruta inmunológica y la ruta de reparación del ADN. Incluimos

además la ruta de la hematoquecia, la ruta de miogénesis y por último la ruta de fusión de mioblastos (Anexo 10).

Para la segunda estrategia utilizamos la plataforma online FUMA^{119,197}. La función GENE2FUNC de FUMA toma una lista de genes y los anota en un contexto biológico. Es decir, realiza pruebas hipergeométricas de enriquecimiento para determinar si los genes de interés están sobre representados en los conjuntos de genes predefinidos en la base de datos de firmas moleculares (*Molecular Signatures Database*, MSigDB)¹⁹⁸, que incluye entre otros datos de Kyoto Encyclopedia of Genes and Genomes (KEGG)¹⁹⁹, Reactome^{200,201} y GeneOntology (GO)^{202,203}. La corrección de los resultados se realiza por BH-FDR para cada categoría por separado.

4 RESULTADOS

4.1 CAPÍTULO 1. ESTUDIOS DE VARIANTES Y GENES PREVIAMENTE ASOCIADOS CON EFECTOS ADVERSOS A LA RADIOTERAPIA.

Tras revisar la bibliografía en busca de variantes asociadas con efectos adversos a la radioterapia, se identificaron alrededor de 315 variantes. De ellas, un total de 152 variantes presentes en 156 genes cumplieron con los criterios de selección, presentando p-valores corregidos por Bonferroni menores a 0,05 o p-valores al metaanálisis o combinado menor de 10^{-05} (**Tabla 7**).

Al revisar la presencia de las 152 variantes en nuestros datos de la secuenciación de exomas, identificamos siete variantes dentro de las regiones secuenciadas (**Tabla 8**). La mayoría de las variantes identificadas en estudios previos se localizan en regiones no codificantes, debido a esto solo pudimos analizar un número reducido de variantes.

Tabla 7. Variantes genéticas asociadas con efectos adversos a la radioterapia reportadas en la literatura

Referencia	Tipo de cáncer	Gen	rsID	Fenotipo	Covariables	Resultados
Kerns et al. 2011 ¹⁰⁷	Próstata	<i>FSHR</i>	rs2268363	Disfunción eréctil		OR=7,03; BF-Pvalor=0,028
Kerns et al. 2013 ¹⁰⁹	Próstata	<i>CDCP1</i>	rs3749191	Disfunción eréctil	Edad, terapia de privación de andrógenos (ADT), Tipo de tratamiento y PCs	Pcombinado=2,1x10 ⁻⁰⁵
		<i>MYT1L</i>	rs11693002			Pcombinado=3,1x10 ⁻⁰⁵
		<i>GLRX3</i>	rs10764930			Pcombinado=3,2x10 ⁻⁰⁵
		<i>HSD17B2</i>	rs11648233			Pcombinado=9,1x10 ⁻⁰⁵
Kerns et al. 2013 ¹¹¹	Próstata	<i>IFNK</i>	rs17779457	Toxicidad urinaria		Pcombinado 6,5 x 10 ⁻⁰⁷
Kerns et al. 2013 ¹¹²	Próstata	<i>MTNR1B/ SLC36A4</i>	rs7120482	Sangrado rectal tardío	5 componentes principales	Pcombinado=5,4x10 ⁻⁰⁸
		<i>MTNR1B/ SLC36A4</i>	rs17630638			Pcombinado=6,9x10 ⁻⁰⁷
		<i>LOC646316/PCDH18</i>	rs10519410			Pcombinado=1,3x10 ⁻⁰⁶
		<i>ACN9/TAC1</i>	rs10255878			Pcombinado=3,6x10 ⁻⁰⁶
		<i>SLC39A11</i>	rs4969040			Pcombinado=5,0x10 ⁻⁰⁵
		<i>ST6GALNAC3</i>	rs2706183			Pcombinado=7,8x10 ⁻⁰⁵
		<i>TBC1D9</i>	rs7698088			Pcombinado=8,8x10 ⁻⁰⁵
Barnett et al. 2014 ¹¹⁴	Próstata/ mama	<i>C10orf113</i>	rs12243039	Toxicidad global		B=0,81; SE=0,18; Pcombinado=6,05x10 ⁻⁰⁶
		<i>GABRB3</i>	rs718304			B=-0,44; SE=0,1; Pcombinado=2,86x10 ⁻⁰⁵
		<i>CCRN4L</i>	rs13116075			B=0,17; SE=0,04; Pcombinado=5,80x10 ⁻⁰⁵
		<i>LINC02024</i>	rs4234649			B=-0,2; SE=0,05; Pcombinado=7,29x10 ⁻⁰⁵
	Próstata	<i>KCND3</i>	rs2788612	Incontinencia rectal		B=0,66; SE=0,09; Pcombinado=1,05x10 ⁻¹²
		<i>ERLIN1</i>	rs11595238			B=0,39; SE=0,08; Pcombinado=1,93x10 ⁻⁰⁶
		<i>KCND3</i>	rs12025303			B=0,2; SE=0,05; Pcombinado=7,40x10 ⁻⁰⁵

Referencia	Tipo de cáncer	Gen	rsID	Fenotipo	Covariables	Resultados
Barnett et al. 2014 ¹¹⁴	Próstata	<i>RBMS3</i>	rs9832625	Incontinencia rectal		B=0,24; SE=0,06; Pcombinado=7,78x10 ⁻⁰⁵
		<i>LOC105376108</i>	rs12353488	Frecuencia urinaria		B=0,21; SE=0,05; Pcombinado=3,57x10 ⁻⁰⁵
		<i>UBR4</i>	rs7527580			B=0,41; SE=0,1; Pcombinado=7,67x10 ⁻⁰⁵
		<i>API5</i>	rs12576830	Incontinencia urinaria		B=0,32; SE=0,08; Pcombinado=8,70x10 ⁻⁰⁵
	Mama	<i>SEMA6D</i>	rs16958536	Telangiectasia		B=-0,24; SE=0,05; Pcombinado=3,87x10 ⁻⁰⁶
		<i>TMTC1</i>	rs17549779			B=-0,19; SE=0,04; Pcombinado=1,43x10 ⁻⁰⁵
		<i>TMEM187</i>	rs6643653			B=-0,26; SE=0,06; Pcombinado=1,60x10 ⁻⁰⁵
		<i>SEMA6D</i>	rs11854033			B=0,22; SE=0,05; Pcombinado=3,02x10 ⁻⁰⁵
		<i>PLXNA2</i>	rs12565978			B=0,26; SE=0,06; Pcombinado=4,09x10 ⁻⁰⁵
		<i>INO80D</i>	rs16837908			B=-0,64; SE=0,16; Pcombinado=5,75x10 ⁻⁰⁵
		<i>PTPRN2</i>	rs221277			B=-0,47; SE=0,12; Pcombinado=6,99x10 ⁻⁰⁵
		<i>SLFN14</i>	rs2840044	Toxicidad global		B=-0,19; SE=0,04; Pcombinado=7,90x10 ⁻⁰⁶
		<i>ANKH</i>	rs10060885			B=0,7; SE=0,16; Pcombinado=1,49x10 ⁻⁰⁵
		<i>LRFN5</i>	rs8004909			B=-0,78; SE=0,18; Pcombinado=2,34x10 ⁻⁰⁵
		<i>SATB2</i>	rs2881208			B=0,18; SE=0,04; Pcombinado=2,83x10 ⁻⁰⁵
		<i>CCDC129</i>	rs882460			B=0,19; SE=0,05; Pcombinado=5,70x10 ⁻⁰⁵
		<i>MTPN</i>	rs12531679	Dolor		B=0,31; SE=0,07; Pcombinado=3,20x10 ⁻⁰⁵
Fachal et al. 2014 ¹¹⁶	Próstata	<i>TANC1</i>	rs7582141	Toxicidad tardía	Recto dosis media	B=0,37; Pmeta=4,64x10 ⁻¹¹

Referencia	Tipo de cáncer	Gen	rsID	Fenotipo	Covariables	Resultados
Fachal et al. 2014 ¹¹⁶	Próstata	TANC1	rs10497203	Toxicidad tardía	Recto dosis media	B=0,39; Pmeta=8,84x10 ⁻¹¹
			rs6432512			B=0,35; Pmeta=1,97x10 ⁻¹⁰
			rs264588			B=0,37; Pmeta=3,08x10 ⁻¹⁰
			rs264631			B=0,36; Pmeta=6,40x10 ⁻¹⁰
Wang et al. 2019 ¹¹⁷	Cabeza y cuello	CEP128	rs162171	Daño en el lóbulo temporal	Edad al diagnóstico de NPC, estado del tumor, técnica de radioterapia	Pcombinado=6,17x10 ⁻⁰⁹
			rs17111237			Pcombinado=3,18x10 ⁻⁰⁷
Kerns et al. 2016 ⁶¹	Próstata	ROBO2	rs12497518	Sangrado rectal	Edad al tratamiento, diabetes, volumen rectal y BED	OR=0,60; 95%CI(0,48 - 0,74); Pmeta=1,41x10 ⁻⁰⁶
		PLS3	rs141044160			OR=2,68; 95%CI(1,78 - 4,04); Pmeta=2,26x10 ⁻⁰⁶
		LINC01192	rs7432328			OR=3,36; 95%CI(2,02 - 5,59); Pmeta=3,23x10 ⁻⁰⁶
		LOC101927040/ LINC01607	rs6999859			OR=1,73; 95%CI(1,37 - 2,17); Pmeta=3,42x10 ⁻⁰⁶
		PDE4A	rs4804134			OR=0,60; 95%CI(0,48 - 0,74); Pmeta=3,47x10 ⁻⁰⁶
		ENAH	rs360071			OR=1,61; 95%CI(1,32 - 1,97); Pmeta=3,62x10 ⁻⁰⁶
		KDM3B	rs17599026	Frecuencia urinaria	Edad al tratamiento, resección prostática transuretral (TURP) previo a la radioterapia, síntomas previos al tratamiento y BED	OR=3,12; 95%CI(2,08 - 4,69); Pmeta=4,16x10 ⁻⁰⁸
		ITGB7	rs11574532			OR=3,25; 95%CI(2,08 - 5,07); Pmeta=2,11x10 ⁻⁰⁷
		KLHL8/AFF1	rs342442			OR=0,51; 95%CI(0,39 - 0,66); Pmeta=3,86x10 ⁻⁰⁷
		CLEC4D	rs4534636			OR=1,96; 95%CI(1,49 - 2,59); Pmeta=1,67x10 ⁻⁰⁶
		EPG5/PSTPIP2	rs8098701			OR=2,41; 95%CI(1,68 - 3,47); Pmeta=2,11x10 ⁻⁰⁶
		CENPF	rs7366282			OR=3,20; 95%CI(1,98 - 5,16); Pmeta=2,03x10 ⁻⁰⁶
		LOC105373899	rs10209697			OR=2,66; 95%CI(1,77 - 3,99); Pmeta=2,27x10 ⁻⁰⁶
		LOC105377876/UBE3D	rs4997823			OR=0,49; 95%CI(0,36 - 0,66); Pmeta=2,35x10 ⁻⁰⁶

Referencia	Tipo de cáncer	Gen	rsID	Fenotipo	Covariables	Resultados
Kerns et al. 2016 ⁶¹	Próstata	<i>HIVEP1/EDN1</i>	rs7356945	Frecuencia urinaria	Edad al tratamiento, resección prostática transuretral (TURP) previo a la radioterapia, síntomas previos al tratamiento y BED	OR=1,74; 95%CI(1,38 - 2,21); Pmeta=3,71x10 ⁻⁰⁶
		<i>LOC100652871</i>	rs6003982			OR=0,51; 95%CI(0,38 - 0,68); Pmeta=4,28x10 ⁻⁰⁶
		<i>MROH5</i>	rs10101158			OR=1,80; 95%CI(1,40 - 2,31); Pmeta=4,39x10 ⁻⁰⁶
		<i>DNAH5</i>	rs7720298	Disminución del flujo urinario	Edad al tratamiento, resección prostática transuretral (TURP) previo a la radioterapia, síntomas previos al tratamiento y BED	OR=2,71; 95%CI(1,90 - 3,86); Pmeta=3,21x10 ⁻⁰⁸
		<i>MCPH1/CSMD1</i>	rs17362923			OR=2,70; 95%CI(1,83 - 4,00); Pmeta=6,79x10 ⁻⁰⁷
		<i>PTPRC</i>	rs76273496			OR=3,68; 95%CI(2,13 - 6,33); Pmeta=2,71x10 ⁻⁰⁶
		<i>BRWD3</i>	rs2203205			OR=1,74; 95%CI(1,38 - 2,20); Pmeta=2,86x10 ⁻⁰⁶
		<i>P3H2/CLDN16</i>	rs144596911			OR=3,60; 95%CI(2,11 - 6,17); Pmeta=2,94x10 ⁻⁰⁶
		<i>LINC01925</i>	rs62091368			OR=4,36; 95%CI(2,33 - 8,14); Pmeta=3,95x10 ⁻⁰⁶
		<i>LOC105379168</i>	rs141342719			OR=3,50; 95%CI(2,05 - 5,95); Pmeta=3,97x10 ⁻⁰⁶
		<i>SMCHD1/EMILIN2</i>	rs673783			OR=2,49; 95%CI(1,69 - 3,67); Pmeta=4,33x10 ⁻⁰⁶
		<i>MS4A5</i>	rs11230328	Toxicidad global	Edad al tratamiento, resección prostática transuretral (TURP) previo a la radioterapia, síntomas previos al tratamiento y BED	OR=0,31; 95%CI(0,21 - 0,41); Pmeta=6,27x10 ⁻¹⁰
		<i>LINC02670 / ELOCP3</i>	rs147596965			OR=0,67; 95%CI(0,43 - 0,91); Pmeta=6,19x10 ⁻⁰⁸
		<i>WIPF1</i>	rs77530448			OR=0,36; 95%CI(0,23 - 0,49); Pmeta=7,36x10 ⁻⁰⁸
		<i>ATP10A</i>	rs4906759			OR=0,55; 95%CI(0,34 - 0,75); Pmeta=1,55x10 ⁻⁰⁷
		<i>RBPJ</i>	rs71610881			OR=0,60; 95%CI(0,36 - 0,83); Pmeta=5,41x10 ⁻⁰⁷
		<i>NUMBL</i>	rs141799618			OR=0,44; 95%CI(0,26 - 0,62); Pmeta=1,22x10 ⁻⁰⁶
		<i>C10orf90</i>	rs2842169			OR=0,28; 95%CI(0,17 - 0,40); Pmeta=1,45x10 ⁻⁰⁶
		<i>LOC105369540</i>	rs11219068			OR=0,28; 95%CI(0,17 - 0,40); Pmeta=1,74x10 ⁻⁰⁶
		<i>PITPNM3</i>	rs8075565			OR=0,28; 95%CI(0,16 - 0,39); Pmeta=2,20x10 ⁻⁰⁶
		<i>LOC105377436</i>	rs6535028			OR=0,29; 95%CI(0,17 - 0,40); Pmeta=2,70x10 ⁻⁰⁶

Referencia	Tipo de cáncer	Gen	rsID	Fenotipo	Covariables	Resultados
Kerns et al. 2016 ⁶¹	Próstata	<i>SEMA6D</i>	rs4775602	Toxicidad global	Edad al tratamiento, resección prostática transuretral (TURP) previo a la radioterapia, síntomas previos al tratamiento y BED	OR=0,23; 95%CI(0,13 - 0,32); Pmeta=3,02x10 ⁻⁰⁶
		<i>ELAPOR1</i>	rs60481745			OR=0,37; 95%CI(0,22 - 0,53); Pmeta=3,53x10 ⁻⁰⁶
		<i>CSMD1</i>	rs7829759			OR=0,33; 95%CI(0,19 - 0,47); Pmeta=3,84x10 ⁻⁰⁶
		<i>SNX33</i>	rs79604958			OR=0,47; 95%CI(0,27 - 0,68); Pmeta=4,33x10 ⁻⁰⁶
		<i>LOC124902872</i>	rs10770857			OR=0,17; 95%CI(0,10 - 0,25); Pmeta=4,98x10 ⁻⁰⁶
		<i>MINAR1/MTHSF</i>	rs12591436			OR=0,18; 95%CI(0,10 - 0,26); Pmeta=5,66x10 ⁻⁰⁶
Kerns et al. 2020 ¹²⁵	Próstata	<i>RPS26P2/FTLP4</i>	rs10969913	Disminución del flujo urinario	Edad al momento del tratamiento, prostatectomía radical, terapia de privación de andrógenos y total BED	HR=3,92; 95%CI(2,57 - 6,00); Pmeta=2,9x10 ⁻¹⁰
		<i>AGMO</i>	rs60424486			HR=1,84; 95%CI(1,45 - 2,33); Pmeta=4,0x10 ⁻⁰⁷
		<i>NKX2-8</i>	rs11624322			HR=1,51; 95%CI(1,29 - 1,77); Pmeta=4,2x10 ⁻⁰⁷
		<i>AGT</i>	rs11122573	Hematuria	Edad al momento del tratamiento, prostatectomía radical, terapia de privación de andrógenos y total BED	HR=1,92; 95%CI(1,53 - 2,42); Pmeta=1,8x10 ⁻⁰⁸
		<i>COG2</i>	rs11122572			HR=1,86; 95%CI(1,48 - 2,34); Pmeta=9,8x10 ⁻⁰⁸
		<i>LOC101927213</i>	rs17190422			HR=2,57; 95%CI(1,79 - 3,68); Pmeta=3,1x10 ⁻⁰⁷
		<i>CD70</i>	rs72993079			HR=2,14; 95%CI(1,60 - 2,87); Pmeta=3,5x10 ⁻⁰⁷
		<i>SLCO2A1</i>	rs9832989			HR=4,45; 95%CI(2,50 - 7,92); Pmeta=3,8x10 ⁻⁰⁷
		<i>CADM1/BUD13</i>	rs368141164	Frecuencia urinaria	Edad al momento del tratamiento, prostatectomía radical, terapia de privación de andrógenos y total BED	HR=1,87; 95%CI(1,48 - 2,36); Pmeta=1,8x10 ⁻⁰⁷
		<i>TBL1XR1/LINC00578</i>	rs6791846			HR=1,47; 95%CI(1,26 - 1,71); Pmeta=5,0x10 ⁻⁰⁷
		<i>RNU2-48P/CLINT1</i>	rs17055178	Sangrado rectal	Edad al momento del tratamiento, prostatectomía radical, terapia de privación de andrógenos y total BED	HR=1,95; 95%CI(1,58 - 2,40); Pmeta=6,2x10 ⁻¹⁰
		<i>KHDRBS3</i>	rs9644474			HR=2,15; 95%CI(1,64 - 2,81); Pmeta=7,9x10 ⁻⁰⁸
		<i>MAP3K8/MTPAP</i>	rs2031925			HR=2,43; 95%CI(1,73 - 3,40); Pmeta=2,4x10 ⁻⁰⁷

Referencia	Tipo de cáncer	Gen	rsID	Fenotipo	Covariables	Resultados
Kerns et al. 2020 ¹²⁵	Próstata	<i>PLOD2</i>	rs74346764	Sangrado rectal	Edad al momento del tratamiento, prostatectomía radical, terapia de privación de andrógenos y total BED	HR=2,52; 95%CI(1,77 - 3,58); Pmeta=3,1x10 ⁻⁰⁷
		<i>PLPP4</i>	rs61871726			HR=1,98; 95%CI(1,52 - 2,59); Pmeta=4,8x10 ⁻⁰⁷
		<i>LAMB1</i>	rs2237706			HR=1,94; 95%CI(1,50 - 2,52); Pmeta=4,8x10 ⁻⁰⁷
		<i>RPGR</i>	rs75759941	Toxicidad global	Edad al momento del tratamiento, prostatectomía radical, terapia de privación de andrógenos y total BED	HR=0,25; 95%CI(0,16 - 0,34); Pmeta=9,3x10 ⁻⁰⁸
		<i>PLK5</i>	rs180958289	Disminución del flujo urinario	Edad al momento del tratamiento, prostatectomía radical, terapia de privación de andrógenos y total BED	HR=7,78; 95%CI(3,92 - 15,5); Pmeta=4,6x10 ⁻⁰⁹
		<i>FHIT/LOC105377113</i>	rs139239158			HR=3,69; 95%CI(2,32 - 5,87); Pmeta=3,4x10 ⁻⁰⁸
		<i>LRP1B</i>	rs75988504			HR=2,01; 95%CI(1,56 - 2,58); Pmeta=4,9x10 ⁻⁰⁸
		<i>CAMTA1</i>	rs112193369			HR=4,09; 95%CI(2,43 - 6,90); Pmeta=1,2x10 ⁻⁰⁷
		<i>DNM2</i>	rs77581414			HR=6,07; 95%CI(3,04 - 12,1); Pmeta=3,1x10 ⁻⁰⁷
		<i>PTPRS/KDM4B</i>	rs78166464			HR=2,83; 95%CI(1,90 - 4,21); Pmeta=3,4x10 ⁻⁰⁷
		<i>COL6A3/COPS8</i>	rs11687040			HR=3,37; 95%CI(2,11 - 5,40); Pmeta=4,2x10 ⁻⁰⁷
		<i>ZNF385D</i>	rs73046248			HR=2,40; 95%CI(1,71 - 3,36); Pmeta=4,2x10 ⁻⁰⁷
		<i>LRPAP1</i>	rs113370662			HR=2,78; 95%CI(1,87 - 4,15); Pmeta=4,5x10 ⁻⁰⁷
		<i>SNED1</i>	rs13403657	Hematuria	Edad al momento del tratamiento, prostatectomía radical, terapia de privación de andrógenos y total BED	HR=3,26; 95%CI(2,17 - 4,90); Pmeta=1,4x10 ⁻⁰⁸
		<i>CACNA2D3</i>	rs139882217			HR=4,78; 95%CI(2,71 - 8,44); Pmeta=6,9x10 ⁻⁰⁸
		<i>ATXN7/PSMD6-AS2</i>	rs4688181			HR=3,22; 95%CI(2,09 - 4,98); Pmeta=1,3x10 ⁻⁰⁷
		<i>SPATA6L</i>	rs3739643			HR=3,44; 95%CI(2,17 - 5,47); Pmeta=1,7x10 ⁻⁰⁷

Referencia	Tipo de cáncer	Gen	rsID	Fenotipo	Covariables	Resultados
Kerns et al. 2020 ¹²⁵	Próstata	CCN6	rs73539559	Hematuria	Edad al momento del tratamiento, prostatectomía radical, terapia de privación de andrógenos y total BED	HR=2,86; 95%CI(1,93 - 4,26); Pmeta=2,0x10 ⁻⁰⁷
		FRMD3	rs190601686			HR=9,13; 95%CI(3,95 - 21,1); Pmeta=2,3x10 ⁻⁰⁷
		LINC02880	rs490393			HR=2,96; 95%CI(1,96 - 4,47); Pmeta=2,4x10 ⁻⁰⁷
		PLEK	rs144214859			HR=2,58; 95%CI(1,80 - 3,70); Pmeta=2,5x10 ⁻⁰⁷
		PRELID2	rs76661052			HR=2,33; 95%CI(1,68 - 3,22); Pmeta=3,3x10 ⁻⁰⁷
		TLE4	rs138731641			HR=5,46; 95%CI(2,82 - 10,6); Pmeta=4,8x10 ⁻⁰⁷
	Próstata	TDRP/ERICH1	rs148048756	Sangrado rectal	Edad al momento del tratamiento, prostatectomía radical, terapia de privación de andrógenos y total BED	HR=4,83; 95%CI(2,75 - 8,47); Pmeta=4,0x10 ⁻⁰⁸
		DLGAP2	rs139288166			HR=4,30; 95%CI(2,53 - 7,31); Pmeta=6,9x10 ⁻⁰⁸
		LRRC3B/NEK10	rs149927798			HR=3,98; 95%CI(2,36 - 6,71); Pmeta=2,3x10 ⁻⁰⁷
		FAM135B	rs73712257			HR=4,72; 95%CI(2,62 - 8,50); Pmeta=2,4x10 ⁻⁰⁷
		GABRG3	rs192744896			HR=3,72; 95%CI(2,23 - 6,18); Pmeta=4,3x10 ⁻⁰⁷
		LOC105369309	rs72915971			HR=2,16; 95%CI(1,60 - 2,91); Pmeta=5,0x10 ⁻⁰⁷
		GSR	rs149176864			HR=2,99; 95%CI(1,99 - 4,49); Pmeta=1,4x10 ⁻⁰⁷
		TRAPPC11	rs191705561	Frecuencia urinaria	Edad al momento del tratamiento, prostatectomía radical, terapia de privación de andrógenos y total BED	HR=3,05; 95%CI(2,05 - 4,53); Pmeta=3,7x10 ⁻⁰⁸
	Próstata	ABCC8	rs61415111			HR=4,60; 95%CI(2,57 - 8,24); Pmeta=3,0x10 ⁻⁰⁷
		SPATA5	rs186353960			HR=2,44; 95%CI(1,73 - 3,44); Pmeta=3,4x10 ⁻⁰⁷
	Próstata	STAC/ARPP21	rs141203061			HR=2,71; 95%CI(1,84 - 4,00); Pmeta=4,3x10 ⁻⁰⁷
Naderi et al. 2021 ¹²¹	Cabeza y cuello	EFNA5/FBXL17/FER	rs1592907	Xerostomía	Edad, sexo, tipo de RT, quimioterapia concomitante, sitio del tumor, toxicidad basal y 10 componentes principales (eigenvectors)	B=0,132; SE=0,026; Pmeta=2,58x10 ⁻⁰⁷
		EFNA5/FBXL17/FER	rs249983			B=0,135; SE=0,026; Pmeta=3,48x10 ⁻⁰⁷
		EFNA5/FBXL17/FER	rs35542			B=0,133; SE=0,026; Pmeta=5,31x10 ⁻⁰⁷
		EFNA5/FBXL17/FER	rs12691430			B=0,132; SE=0,026; Pmeta=6,29x10 ⁻⁰⁷

Referencia	Tipo de cáncer	Gen	rsID	Fenotipo	Covariables	Resultados
Naderi et al. 2021 ¹²¹	Cabeza y cuello	<i>EFNA5/FBXL17/FER</i>	rs67533472	Xerostomía	Edad, sexo, tipo de RT, quimioterapia concomitante, sitio del tumor, toxicidad basal y 10 componentes principales (eigenvectors)	B=0,132; SE=0,026; Pmeta=6,29x10 ⁻⁰⁷
		<i>EFNA5/FBXL17/FER</i>	rs1490829			B=0,131; SE=0,026; Pmeta=6,97x10 ⁻⁰⁷
		<i>EFNA5/FBXL17/FER</i>	rs35420			B=0,131; SE=0,026; Pmeta=7,16x10 ⁻⁰⁷
		<i>EFNA5/FBXL17/FER</i>	rs249981			B=0,131; SE=0,026; Pmeta=7,53x10 ⁻⁰⁷
		<i>EFNA5/FBXL17/FER</i>	rs9328017			B=0,130; SE=0,026; Pmeta=8,95x10 ⁻⁰⁷
		<i>EFNA5/FBXL17/FER</i>	rs1844705			B=0,129; SE=0,026; Pmeta=1,08x10 ⁻⁰⁶
		<i>EFNA5/FBXL17/FER</i>	rs193899			B=0,129; SE=0,026; Pmeta=1,12x10 ⁻⁰⁶
		<i>EFNA5/FBXL17/FER</i>	rs9328016			B=0,127; SE=0,027; Pmeta=1,77x10 ⁻⁰⁶
		<i>EFNA5/FBXL17/FER</i>	rs1990917			B=0,126; SE=0,026; Pmeta=1,81x10 ⁻⁰⁶
		<i>EFNA5/FBXL17/FER</i>	rs712634			B=0,127; SE=0,027; Pmeta=1,89x10 ⁻⁰⁶
		<i>EFNA5/FBXL17/FER</i>	rs1990918			B=0,125; SE=0,026; Pmeta=2,28x10 ⁻⁰⁶
		<i>EFNA5/FBXL17/FER</i>	rs249974			B=0,124; SE=0,026; Pmeta=2,65x10 ⁻⁰⁶
Guhlich et al. 2021 ¹⁰¹	Próstata	<i>TGFBR1</i>	rs10512263	Toxicidad aguda		OR=3,80; 95% CI(1,78 - 8,12); BF-pvalor=0,0006
Schack et al. 2022 ¹²²	Cabeza y cuello	<i>ECSCR</i>	rs28419191	Mucositis	Dosis total de RT, quimioterapia concomitante y un marcador sustituto para el volumen irradiado estimado a partir del sitio del cáncer y el estadio TNM	OR=1,958; SE=0,259; Pmeta=3,6x10 ⁻¹⁴
		<i>STING1</i>	rs1131769			OR=1,946; SE=0,239; Pmeta=4,3x10 ⁻¹⁶

rsID, identificado de variante; OR, *odds ratio*; B, beta; SE, desviación estandar; CI, intervalo de confianza; HR, *hazard ratio*; BF-Pvalor, p valor corregido por Bonferroni; Pcombinado, p-valores combinados de estudios divididos en dos fases, Pmeta, p valor del metaanálisis.

Seis de las variantes descritas en la literatura y presentes en los exomas de los pacientes secuenciados, fueron identificadas en estudios de asociación en pacientes con cáncer de próstata, mientras que sólo una variante, la rs1131769 fue identificada en un estudio de pacientes con cáncer de cabeza y cuello (Schack et al. 2022)¹²². El fenotipo analizado en este estudio fue mucositis, mientras que en los estudios de pacientes con cáncer de próstata se consideraron diferentes fenotipos (**Tabla 8**).

Tabla 8. Variantes descritas en la literatura asociadas con toxicidad a la radioterapia presentes en los exomas analizados

Referencia	Tipo de cáncer	Gen	rsID	Fenotipo estudiado	OR (95%CI)	P-valor*
Kerns et al. 2013 ¹⁰⁰	Próstata	<i>CDCP1</i>	rs3749191	‡Disfunción eréctil	NA	2,1×10 ⁻⁰⁵
Kerns et al. 2016 ⁶¹	Próstata	<i>KDM3B</i>	rs17599026	£Frecuencia urinaria	3,12(2,08-4,69)	4,16×10 ⁻⁰⁸
		<i>ITGB7</i>	rs11574532		3,25(2,08-5,07)	2,11×10 ⁻⁰⁷
		<i>PDE4A</i>	rs4804134	‡Sangrado rectal	0,60(0,48-0,74)	3,47×10 ⁻⁰⁶
Kerns et al. 2020 ¹²⁵	Próstata	<i>PLEK</i>	rs144214859	‡Hematuria	NA	2,5×10 ⁻⁰⁷
		<i>SPATA6L</i>	rs3739643		NA	1,7×10 ⁻⁰⁷
Schack et al. 2022 ¹⁰¹	Cabeza y cuello	<i>STING1</i>	rs1131769	‡Mucositis	1,95	4,3 ×10 ⁻¹⁶

rsID, identificador de variante; A1, alelo de riesgo; OR, *odds ratio*; CI, intervalo de confianza; P-valor*, P-valor al metaanálisis/combinado. ‡ Covariables usadas en el estudio: edad, terapia de privación de andrógenos, tratamiento y los cinco primeros componentes principales (CPs). £ Covariables usadas en el estudio: edad al tratamiento, resección prostática transuretral (TURP) previo a la radioterapia, síntomas previos al tratamiento y BED. ‡ Covariables usadas en el estudio: edad al tratamiento, diabetes, volumen rectal y BED. ‡ Covariables usadas en el estudio: edad al momento del tratamiento, prostatectomía radical, terapia de privación de andrógenos y BED. ‡ Covariables usadas en el estudio: dosis total de RT, quimioterapia concomitante y un marcador sustituto para el volumen irradiado estimado a partir del sitio del cáncer y el estadio TNM.

4.1.1 Análisis de asociación univariable

Analizamos las siete variantes previamente asociadas con radiotoxicidad en nuestra cohorte de 141 pacientes, incluyendo 46 casos y 95 controles. De todas las variantes estudiadas, sólo el SNP rs144214859, localizado en el gen *PLEK*, mostró una asociación significativa con un p-valor nominal de 0,04. Esta asociación no se mantuvo significativa después de aplicar la corrección por test múltiple. Los resultados de asociación para los SNPs analizados se presentan en la **Tabla 9**.

Tabla 9. Resultados de asociación en nuestra cohorte de estudio de variantes previamente descritas en la literatura

Chr	Posición	rsID	Gen	A1	OR (95%CI)	SE	P-valor
2	68624610	rs144214859	<i>PLEK</i>	A	NA	NA	0,0414
9	4600633	rs3739643	<i>SPATA6L</i>	T	NA	NA	0,15
19	10568519	rs4804134	<i>PDE4A</i>	G	0,69(0,41-1,17)	0,266	0,171
5	138857919	rs1131769	<i>STING1</i>	C	1,6 (0,69-3,67)	0,427	0,274
5	137763798	rs17599026	<i>KDM3B</i>	T	1,30(0,59-2,87)	0,405	0,519
12	53594438	rs11574532	<i>ITGB7</i>	T	0,81(0,25-2,65)	0,605	0,726
3	45134822	rs3749191	<i>CDCP1</i>	C	1,09(0,65-1,85)	0,267	0,729

Chr, cromosoma; rsID, identificador de variante; A1, alelo de riesgo; A2 alelo de referencia; OR, *odds ratio*; CI, intervalo de confianza; SE, desviación estándar; P-valor*, P-valor al metaanálisis/combinado.

4.1.2 Análisis de asociación multivariable

Analizamos las variantes identificadas en la literatura en nuestro modelo de regresión logística ajustando por las covariables “edad” y “RectV65”, asociadas con el fenotipo sangrado rectal en la cohorte RADIOGEN-Próstata. Los resultados de este análisis no revelaron ninguna asociación significativa para ninguno de los SNPs analizados (**Tabla 10**).

Tabla 10. Resultados del análisis multivariable de variantes previamente relacionadas con efectos adversos a la radioterapia en nuestra cohorte de estudio

Chr	Posición	rsID	Gen	A1	OR (95%CI)	SE	Z_STAT	P-valor
19	10568519	rs4804134	<i>PDE4A</i>	G	0,73 (0,45-1,17)	0,242	-1,31	0,191
5	138857919	rs1131769	<i>STING1</i>	T	0,63 (0,26-1,54)	0,454	-1	0,318
5	137763798	rs17599026	<i>KDM3B</i>	T	1,38 (0,60-3,13)	0,419	0,76	0,446
3	45134822	rs3749191	<i>CDCP1</i>	T	0,96 (0,55-1,66)	0,281	-0,15	0,877
12	53594438	rs11574532	<i>ITGB7</i>	T	0,97 (0,32-2,92)	0,563	-0,06	0,955
2	68624610	rs144214859	<i>PLEK</i>	A	NA	NA	NA	NA
9	4600633	rs3739643	<i>SPATA6L</i>	T	NA	NA	NA	NA

Chr, cromosoma; rsID, identificador de variante; A1, alelo de riesgo; OR, *odds ratio*; CI, Intervalo de confianza; SE, desviación estándar; Z_STAT, puntaje Z de Wald para regresión logística.

4.1.3 Validación del SNP rs4804134

Una de las variantes identificadas en la literatura y presente en nuestros exomas se había descrito asociada con sangrado rectal. Esta variante corresponde al SNP rs4804134 localizado en las proximidades del gen *PDE4A*, identificado en el estudio de Kerns et al. (2016)⁶¹. En dicho estudio el fenotipo sangrado rectal se definió como el incremento de al menos un grado en la toxicidad en comparación con los datos basales. Las covariables usadas en nuestro análisis fueron las descritas en el estudio de referencia⁶¹, que incluyen: edad al tratamiento, diabetes, volumen de recto y BED. No identificamos al SNP rs4804134 asociado con el fenotipo de sangrado rectal en nuestra cohorte de pacientes (**Tabla 11**).

Tabla 11. Validación del SNP rs4804134

Chr	Posición	rsID	Gen	A1	OR (95%CI)	SE	Z_STAT	P-valor
19	10568519	rs4804134	<i>PDE4A</i>	G	0,71 (0,45-1,17)	0,25	-1,308	0,176

Chr, cromosoma; rsID, identificador de variante; A1, alelo de riesgo; OR, *odds ratio*; CI, intervalo de confianza; SE, desviación estándar; Z_STAT, puntaje Z de Wald para regresión logística.

4.1.4 Identificación de nuevas variantes en genes previamente asociados con efectos adversos a la radioterapia

Se exploraron nuevas variantes en genes en los que previamente se habían identificado variantes candidatas asociadas con efectos adversos a la radioterapia. De las 152 variantes identificadas previamente (ver apartado 3.7 y **Tabla 7**), encontramos 29 localizadas en 23 genes, que presentaron un p-valor combinado o al metaanálisis menor a 5×10^{-08} , o un p-valor corregido por Bonferroni menor a 0,05 (**Tabla 12**).

Entre los genes incluidos, ocho (*PLK5*, *FHIT*, *LOC105377113*, *LRP1B*, *SNED1*, *TDRP*, *ERICH1* y *TRAPPC11*) presentaron variantes de baja frecuencia (MAF < 0,05).

Tabla 12. Genes/regiones con variantes previamente relacionados con toxicidad a la radioterapia

Gen/gen-cercano	Fenotipo	Tipo de cáncer	Referencia
<i>FSHR</i>	Disfunción eréctil	Próstata	Kerns et al. 2011 ¹⁰⁷
<i>MTNR1B/SLC36A4</i>	Sangrado rectal	Próstata	Kerns et al. 2013 ¹¹²
<i>TANC1</i>	Toxicidad global	Próstata	Fachal et al. 2014 ¹¹⁶
<i>KDM3B</i>	Frecuencia Urinaria		
<i>DNAH5</i>	Disminución del flujo urinario	Próstata	Kerns et al. 2016 ⁶¹
<i>MS4A5</i>	Toxicidad global		
<i>CEP128</i>	Daño en el lóbulo temporal	Cabeza y cuello	Wang et al. 2019 ¹¹⁷
<i>RNU2-48P/CLINT1</i>	Sangrado rectal		
<i>RPS26P2/FTLP4</i>	Disminución del flujo urinario		
<i>AGT</i>	Hematuria		
* <i>PLK5</i> , <i>FHIT/LOC105377113</i> , <i>LRP1B</i>	Disminución del flujo urinario	Próstata	Kerns et al. 2020 ¹²⁵
* <i>SNED1</i>	Hematuria		
* <i>TDRP/ERICH1</i>	Sangrado rectal		
* <i>TRAPPC11</i>	Frecuencia urinaria		
<i>ECSCR</i> , <i>STING1</i>	Mucositis	Cabeza y cuello	Schack et al. 2022 ¹²²

* Genes con variantes de baja frecuencia < 5 % asociadas con radiotoxicidad, analizadas en datos de genotipado.

Se identificaron 60 variantes de impacto funcional presentes en los 23 genes previamente asociados con efectos adversos a la radioterapia (ver apartado 3.7.4). Estas variantes se analizaron mediante regresión logística multivariable (ANEXO 6). Trece no mostraron asociación con el fenotipo estudiado (**Tabla 13**), mientras que 47 no pudieron ser analizadas debido a que tenían una frecuencia de alelo menor muy baja o su ausencia en alguno de los grupos (caso o control) impidió su análisis.

Tabla 13. Nuevas variantes identificadas en genes previamente descritos

Chr	Posición	ID	Gen	A1	OR (95%CI)	SE	Z_STAT	P-valor
2	160076227	rs144753108	TANC1	G	7,91 (0,62-100,15)	1,29	1,59	0,11
2	140990879	2:140990879:T:C	LRP1B	C	2,59 (0,30-22,30)	1,10	0,87	0,383
2	241974013	rs200219556	SNED1	A	3,51 (0,18-68,11)	1,51	0,82	0,406
4	184585165	rs141909783	TRAPPC11	C	0,54 (0,10-2,77)	0,83	-0,73	0,464
2	242009463	rs183525741	SNED1	T	2,89 (0,16-53,05)	1,48	0,71	0,474
2	141707812	rs141827692	LRP1B	C	2,42 (0,15-40,07)	1,43	0,61	0,536
5	13701422	rs113425437	DNAH5	T	1,66 (0,12-22,96)	1,34	0,37	0,705
5	13721137	rs140832239	DNAH5	T	1,59 (0,09-28,52)	1,47	0,31	0,752
2	242004296	rs150081750	SNED1	A	0,87 (0,08-10,05)	1,24	-0,1	0,913
5	137688497	rs200772506	KDM3B	A	0,95 (0,26-3,51)	0,66	-0,08	0,935
5	137754808	rs7706614	KDM3B	A	0,97 (0,31-3,04)	0,58	-0,04	0,961
2	141812703	2:141812703:C:T	LRP1B	T	1,07 (0,06-18,29)	1,45	0,04	0,964
4	184615166	rs75176151	TRAPPC11	C	1,02 (0,08-12,51)	1,27	0,01	0,987

Chr, cromosoma; ID, identificador de variante; A1, alelo de riesgo; OR, *odds ratio*; CI, intervalo de confianza; SE, desviación estándar; Z_STAT, puntaje Z de Wald para regresión logística.

4.2 CAPITULO 2. ESTUDIOS DE ASOCIACIÓN BASADOS EN GENES

Para evaluar los efectos acumulativos de múltiples variantes raras y de baja frecuencia en un gen realizamos el análisis de asociación basado en genes utilizando el test SKAT-O. Estos análisis se llevaron a cabo de forma separada, según el grado de fenotipos de estudio, como se detalla en la **Figura 16**.

4.2.1 Análisis de asociación del fenotipo sangrado rectal

En la fase de descubrimiento del fenotipo sangrado rectal, los análisis de asociación utilizando el test SKAT-O (ver apartado 3.8.1.1), incluyeron diferentes estrategias de filtrado de variantes. El número de variantes y de genes incluidos en el análisis dependió de cada estrategia de filtrado:

- Al considerar las variantes de baja frecuencia poblacional de todas las regiones secuenciadas, se incluyeron en el análisis 898 443 variantes y se obtuvieron resultados de 21 018 genes.

- ii) Con el filtrado de variantes exónicas (excluyendo las sinónimas) y variantes de splicing canónicas ± 2 pb, se obtienen 114 453 variantes y 11 106 genes.
- iii) Con el filtrado de variantes de impacto funcional, priorizadas a partir de variantes exónicas (excluyendo las sinónimas) y variantes de splicing canónicas ± 2 pb, se obtienen 59 480 variantes y 6787 genes.

Siguiendo estas estrategias de análisis para la fase de descubrimiento de la cohorte RADIOGEN, identificamos un total de 32 genes asociados, con p-valores nominales menores a 10^{-04} (entre $1,07 \times 10^{-04}$ y $9,7 \times 10^{-04}$), todos los genes tuvieron al menos dos variantes analizadas. Cuando se consideraron todas las variantes de baja frecuencia poblacional identificamos 18 genes, al considerar variantes exónicas (excluyendo sinónimas) y de *splicing* canónicas ± 2 pb, identificamos nueve genes, y al considerar únicamente las variantes de impacto funcional se identificaron cinco genes. Cinco de los 32 genes se encontraron asociados en más de uno de los análisis realizados (estrategias i, ii e iii). El gen *LRRC70* se encontró asociado en los tres análisis, aunque el resultado del análisis en variantes de impacto funcional no se consideró por presentar sólo una variante analizada. Los genes *MYBPC2*, *NUP188*, *PSMB8* y *ZCCHC2* presentaron asociación significativa en el análisis que incluyó variantes exónicas (excluyendo las sinónimas) y de *splicing* canónicas ± 2 pb (ii) y en el análisis de variantes de impacto funcional (iii), todos estos genes presentaron mayor número de variantes en casos que en controles (**Tabla 14**). Otro de los genes para los que se encontró asociación es *LRP1B*.

Los 32 genes más significativos identificados en la fase de descubrimiento fueron validados en la cohorte externa REQUITE (ver apartado 3.8.1.2). Como resultado de la validación del análisis que incluyó a todas las variantes de baja frecuencia secuenciadas (estrategia i), encontramos asociados nominalmente a los genes *PRB1* y *RANBP10* con p-valor de 0,024 y 0,046, respectivamente. Al explorar los resultados de los genes que validaron, observamos que en la fase de descubrimiento se analizaron para *PRB1* trece variantes, de las cuales diez eran variantes intrónicas, dos *frameshift deletion* y una variante *inframe deletion*. El mismo número de variantes se incluyó en la validación, aunque aquí se analizaron siete variantes *missense*, cuatro sinónimas y dos intrónicas. En el caso del gen *RANBP10*, el análisis en la cohorte de descubrimiento incluyó 30 variantes (cuatro variantes *missense*, cuatro sinónimas, doce intrónicas, cinco UTRs, tres *upstream* y dos *downstream* del gen), mientras que en la validación se incluyeron doce variantes (dos variantes sinónimas, tres intrónicas y siete en regiones UTRs). Por otra parte, ningún gen validó los resultados de los análisis llevados a cabo con la estrategia de filtrado que incluyó variantes exónicas (excluyendo las sinónimas) y variantes de *splicing* canónicas ± 2 pb (ii), ni con la estrategia de filtrado de variantes de impacto funcional (iii).

En la **Tabla 14** se presentan los genes más significativos de los análisis en la fase de descubrimiento y de la validación con sus respectivos p-valores nominales. La **Figura 17** presenta los genes analizados con las diferentes estrategias de filtrado de variantes (i, ii, iii), en la fase de descubrimiento del fenotipo sangrado rectal. Ninguna de las asociaciones se mantuvo después de la corrección con BH-FDR.

Tabla 14. Genes asociados con sangrado rectal como efecto adverso a la radioterapia en pacientes con cáncer de próstata (fenotipo sangrado rectal)

Análisis	Gen	Chr	Descripción	ΦFASE DE DESCUBRIMIENTO (RADIOGEN)			ψFASE DE VALIDACIÓN (REQUITE)		
				V. incl.	Descripción de variante	P-valor	V. incl.	Descripción de variante	P-valor
1) Todas las variantes analizadas	<i>MBP</i>	18	<i>myelin basic protein</i>	67	2 variantes <i>missense</i> , 2 sinónimas, 38 intrónicas y 25 en UTRs	1,07x10 ⁻⁰⁴	25	1 variante <i>stoploss</i> , 1 <i>missense</i> , 1 sinónima, 8 intrónicas y 14 en UTRs	0,184
	<i>PRB1</i>	12	<i>proline rich protein BstNI subfamily 1</i>	13	2 variantes <i>frameshift deletion</i> , 1 <i>inframe deletion</i> y 10 intrónicas	1,37x10 ⁻⁰⁴	13	7 variantes <i>missense</i> , 4 sinónimas, 2 intrónicas	0,024
	<i>PRB2</i>	12	<i>proline rich protein BstNI subfamily 2</i>	9	1 variante <i>frameshift deletion</i> , 1 <i>inframe deletion</i> y 7 intrónicas	1,77x10 ⁻⁰⁴	0	sin información	NA
	<i>CACNA1C-AS1</i>	12	<i>CACNA1C antisense RNA 1</i>	15	4 variantes ncRNA-exónicas y 11 variantes ncRNA-intrónicas	2,09x10 ⁻⁰⁴	6	4 variantes intrónicas, 1 <i>upstream</i> y 1 ncRNA-exónica	0,781
	<i>**LRRC70</i>	5	<i>leucine rich repeat containing 70</i>	2	2 variantes <i>missense</i>	2,68x10 ⁻⁰⁴	2	1 variante <i>missense</i> y 1 variante en UTR	0,853
	<i>TNFRSF10D</i>	8	<i>TNF receptor superfamily member 10d</i>	41	4 variantes sinónimas, 3 <i>missense</i> , 12 intrónicas, 13 intergénicas y 9 en UTRs	2,84x10 ⁻⁰⁴	4	1 variante sinónima, 1 nosinónima y 2 UTRs	0,664

Análisis	Gen	Chr	Descripción	ΦFASE DE DESCUBRIMIENTO (RADIOGEN)			ΨFASE DE VALIDACIÓN (REQUITE)		
				V. incl.	Descripción de variante	P-valor	V. incl.	Descripción de variante	P-valor
1) Todas las variantes analizadas	<i>PLEKHD1</i>	14	<i>pleckstrin homology and coiled-coil domain containing D1</i>	21	3 variantes <i>missense</i> , 1 sinónima, 16 intrónicas y 1 <i>upstream</i>	3,45x10 ⁻⁰⁴	4	4 variantes intrónicas	0,291
	<i>RANBP10</i>	16	<i>RAN binding protein 10</i>	30	4 variantes <i>missense</i> , 4 sinónimas, 12 intrónicas, 5 UTRs, 3 <i>upstream</i> y 2 <i>downstream</i>	4,20x10 ⁻⁰⁴	12	2 variantes sinónimas, 3 intrónicas y 7 en UTRs	0,015
	<i>GLO1</i>	6	<i>glyoxalase I</i>	12	1 variante <i>missense</i> , 4 intrónicas, 2 <i>downstream</i> , 4 en UTRs y 1 <i>upstream</i>	4,47x10 ⁻⁰⁴	3	1 variante <i>missense</i> y 2 en UTRs	0,867
	<i>LCN15</i>	9	<i>lipocalin 15</i>	13	2 variantes <i>missense</i> , 8 intrónicas y 3 en UTRs	4,65x10 ⁻⁰⁴	1	1 variante intrónica	0,461
	<i>CACNA1C-AS2</i>	12	<i>CACNA1C antisense RNA 2</i>	2	1 variante ncRNA-exónica y 1 ncRNA-intrónica	4,67x10 ⁻⁰⁴	0	sin información	1
	<i>R3HDM2</i>	12	<i>R3H domain containing 2</i>	31	9 variantes <i>missense</i> , 2 sinónimas, 17 intrónicas y 3 UTRs	5,34x10 ⁻⁰⁴	8	2 variantes <i>missense</i> , 2 sinónimas y 4 intrónicas	0,844

Análisis	Gen	Chr	Descripción	ΦFASE DE DESCUBRIMIENTO (RADIOGEN)			ψFASE DE VALIDACIÓN (REQUIRE)		
				V. incl.	Descripción de variante	P-valor	V. incl.	Descripción de variante	P-valor
1) Todas las variantes analizadas	<i>LRP1B</i>	2	<i>LDL receptor related protein 1B</i>	152	108 variantes intrónicas, 23 <i>missense</i> , 15 sinónimas, 5 en UTRs y 1 <i>upstream</i>	6,44x10 ⁻⁰⁴	42	18 variantes <i>missense</i> , 6 sinónimas, 14 intrónicas, 3 <i>upstream</i> y 1 en UTR	0,486
	<i>NCOA2</i>	8	<i>nuclear receptor coactivator 2</i>	54	37 variantes intrónicas, 6 <i>missense</i> , 2 sinónimas, 8 en UTRs y 4 <i>upstream</i>	7,17x10 ⁻⁰⁴	12	1 variante <i>missense</i> , 1 sinónima, 3 intrónicas y 7 en UTRs	0,590
	<i>TNFRSF10C</i>	8	<i>TNF receptor superfamily member 10c</i>	21	3 variantes sinónimas, 1 <i>frameshift deletion</i> , 2 intrónicas, 13 variantes intergénicas y 2 en UTRs	7,22x10 ⁻⁰⁴	8	1 variante <i>missense</i> , 2 sinónimas, 1 <i>frameshift deletion</i> , 2 en UTRs, 1 <i>downstream</i> y 1 intrónica	0,156
	<i>PSMC1</i>	14	<i>proteasome 26S subunit, ATPase 1</i>	7	7 variantes intrónicas	7,68x10 ⁻⁰⁴	1	1 variante intrónica	0,315
	<i>TSSC2</i>	11	<i>tumor suppressing subtransferable candidate 2 (pseudogene)</i>	8	6 variantes ncRNA-intrónicas y 2 intergénicas	9,61x10 ⁻⁰⁴	2	2 variantes ncRNA-intrónicas	0,671

Análisis	Gen	Chr	Descripción	ΦFASE DE DESCUBRIMIENTO (RADIOGEN)			ψFASE DE VALIDACIÓN (REQUITE)		
				V. incl.	Descripción de variante	P-valor	V. incl.	Descripción de variante	P-valor
2)Variantes exónicas, (exclusión de sinónimas) y de splicing canónicas ± 2pb	*MYBPC2	19	<i>myosin binding protein C2</i>	15	14 variantes <i>missense</i> y 1 <i>stopgain</i>	6,92x10 ⁻⁰⁵	8	8 variantes <i>missense</i>	0,213
	*NUP188	9	<i>nucleoporin 188</i>	9	9 variantes <i>missense</i>	9,21x10 ⁻⁰⁵	4	4 variantes <i>missense</i>	0,577
	*LRRC70	5	<i>leucine rich repeat containing 70</i>	2	2 variantes <i>missense</i>	2,08x10 ⁻⁰⁴	1	1 variante <i>missense</i>	1
	CBLC	19	<i>Cbl proto-oncogene C</i>	8	7 variantes <i>missense</i> y 1 <i>frameshift insertion</i>	3,28x10 ⁻⁰⁴	6	5 variantes <i>missense</i> y 1 <i>frameshift insertion</i>	0,875
	*PSMB8	6	<i>proteasome 20S subunit beta 8</i>	3	3 variantes <i>missense</i>	5,08x10 ⁻⁰⁴	1	1 variante <i>missense</i>	0,283
	PLEKHA1	10	<i>pleckstrin homology domain containing A1</i>	4	4 variantes <i>missense</i>	5,45x10 ⁻⁰⁴	0	sin información	1
	*ZCCHC2	18	<i>zinc finger CCHC-type containing 2</i>	4	4 variantes <i>missense</i>	7,35x10 ⁻⁰⁴	1	1 variante <i>missense</i>	0,598
	RAVER1	19	<i>ribonucleoprotein, PTB binding 1</i>	3	3 variantes <i>missense</i>	8,25x10 ⁻⁰⁴	0	sin información	1
	UPK1A	19	<i>uropod protein 1A</i>	4	3 variantes <i>missense</i> y 1 <i>frameshift deletion</i>	9,24x10 ⁻⁰⁴	2	2 variantes <i>missense</i>	0,088

Análisis	Gen	Chr	Descripción	ΦFASE DE DESCUBRIMIENTO (RADIOGEN)			ΨFASE DE VALIDACIÓN (REQUIRE)		
				V. incl.	Descripción de variante	P-valor	V. incl.	Descripción de variante	P-valor
3) Variantes de impacto funcional en regiones exónicas (exclusión de sinónimas) y de splicing canónicas ±2 pb	*MYBPC2	19	<i>myosin binding protein C2</i>	8	7 variantes <i>missense</i> y 1 <i>stopgain</i>	4,52x10 ⁻⁰⁵	2	2 variantes <i>missense</i>	0,447
	*NUP188	9	<i>nucleoporin 188</i>	6	6 variantes <i>missense</i>	2,59x10 ⁻⁰⁴	2	2 variantes <i>missense</i>	0,436
	CRNKL1	20	<i>crooked neck pre-mRNA splicing factor 1</i>	2	1 variante <i>missense</i> y 1 <i>stopgain</i>	2,99x10 ⁻⁰⁴	1	1 variante <i>stopgain</i>	0,243
	*ZCCHC2	18	<i>zinc finger CCHC-type containing 2</i>	2	2 variantes <i>missense</i>	5,98x10 ⁻⁰⁴	1	1 variante <i>missense</i>	0,604
	*PSMB8	6	<i>proteasome 20S subunit beta 8</i>	3	3 variantes <i>missense</i>	6,28x10 ⁻⁰⁴	1	1 variante <i>missense</i>	0,259

*Estos genes se encuentran asociados en dos de los análisis; los genes que validaron se presentan en negrita; Chr, cromosoma; V. incl., variantes incluidas; Φ fase de descubrimiento incluyó 141 pacientes con cáncer de próstata de la cohorte RADIOGEN; Ψ fase de validación incluyó a 391 pacientes de la cohorte de REQUIRE; UTR, *untranslated region*; ncRNA, *non-coding RNA*; *upstream*, variante ubicada aguas arriba del gen; *downstream*, variante ubicada aguas abajo del gen.

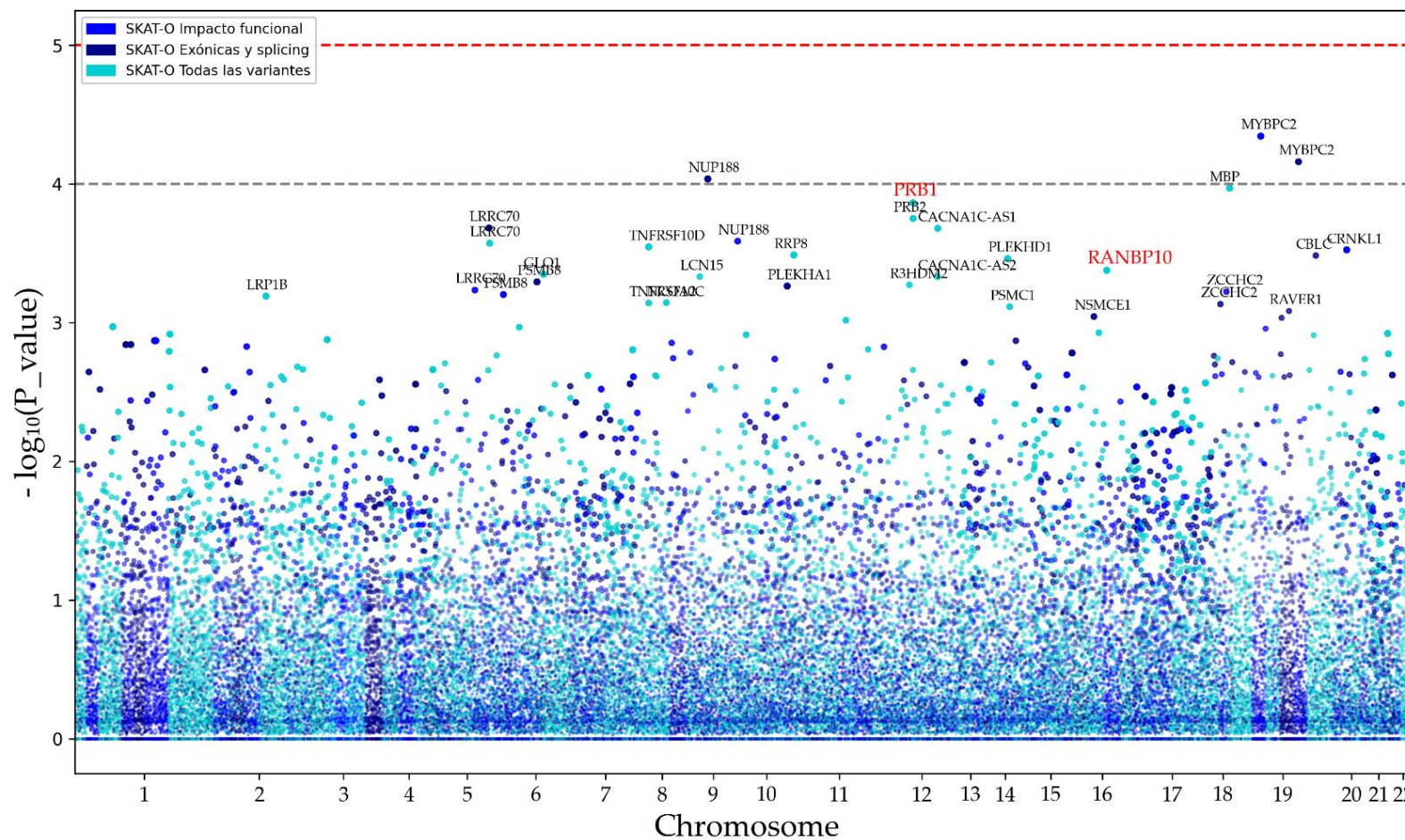


Figura 17. Manhattan plot de los genes asociados en los diferentes análisis SKAT-O, realizados en la fase de descubrimiento para el fenotipo sangrado rectal extremo. Los p-valores fueron transformados a-log. Se indican los nombres de los genes más asociados nominalmente, con color rojo se escribe el nombre de los genes que validaron en la cohorte de REQUITE. La línea gris representa el umbral de $-\log_{10}$ del p-valor 1×10^{-4} y la línea roja representa el $-\log_{10}$ del p-valor sugestivo 1×10^{-5} .

4.2.2 Análisis de asociación del fenotipo sangrado rectal extremo

Los análisis de asociación para el fenotipo sangrado rectal extremo se realizaron con el test SKAT-O. El número de variantes y de genes incluidos en el análisis (ver apartado 3.8.1.1) dependió de la estrategia de filtrado:

- i) Al considerar las variantes de todas las regiones secuenciadas, se incluyeron en el análisis 244 056 variantes y se obtuvieron resultados de 19 318 genes.
- ii) Con el filtrado de variantes exónicas (excluyendo las sinónimas) y de splicing canónicas ± 2 pb, 32 550 variantes fueron incluidas y se obtuvieron 7308 genes resultantes del análisis.
- iii) La tercera estrategia, la dirigimos a variantes de impacto funcional priorizadas a partir de variantes exónicas (excluyendo las sinónimas) y variantes en sitios de splicing canónicos ± 2 pb. Aquí partimos de 15 413 variantes y obtuvimos resultados de 3365 genes.

Como resultado de los análisis realizados con estas estrategias de filtrado, identificamos 32 genes en la fase de descubrimiento en la cohorte RADIOGEN, estos genes presentaron valores de significancia nominal entre $3,23 \times 10^{-05}$ y $9,99 \times 10^{-04}$, todos los genes tuvieron al menos dos variantes analizadas. Veinte genes fueron identificados cuando se consideró en el análisis a las variantes de todas las regiones secuenciadas (estrategia i), doce genes resultaron asociados en el análisis de las variantes exónicas (excluyendo las sinónimas) y de *splicing* canónicas ± 2 pb (estrategia ii) y sólo dos genes resultaron asociados en el análisis de las variantes de impacto funcional (estrategia iii). De ellos, dos genes presentaron asociación en dos de los análisis llevados a cabo con las diferentes estrategias de filtrado. El gen *ADRM1* presentó asociación nominal tanto en el análisis que incluyó a las variantes de todas las regiones secuenciadas como en el análisis de las variantes exónicas y de *splicing* ± 2 pb. El gen *METTL25* también presentó asociación nominal en dos de los análisis, específicamente en el análisis de las variantes exónicas (excluyendo las sinónimas) y de *splicing* ± 2 pb y en el análisis de las variantes de impacto funcional. Además, encontramos asociado nominalmente al gen *ZNF385D* (p-valor $4,34 \times 10^{-04}$). Los resultados de los análisis de asociación de la fase de descubrimiento se muestran en la **Tabla 15**. La **Figura 18** muestra los genes analizados con las diferentes estrategias de filtrado de la fase de descubrimiento.

Posteriormente, en la cohorte de REQUITE se realizó la validación de los 32 genes más asociados identificados en la fase de descubrimiento. En estos análisis, el gen *ADRM1* con p-valor nominal de 0,048 validó los resultados del análisis en RADIOGEN, el cuál evaluó a las variantes de baja frecuencia de todas las regiones secuenciadas (i) (**Tabla 15**).

Al explorar en los resultados del gen *ADRM1*, observamos que en la fase de descubrimiento se analizaron cinco variantes (dos variantes *missense*, dos sinónimas y una variante *inframe insertion*). En contraste, en la fase de validación se analizaron para este gen 17 variantes (tres variantes *missense*, una *inframe insertion*, dos sinónimas, tres intrónicas y ocho *downstream*). Por otra parte, el gen *THEGL* con p-valor 0,034 no fue considerado como gen validado debido a que solo se analizó una variante en el análisis llevado a cabo con las variantes exónicas (sin sinónimas) y variantes de *splicing* canónicas ± 2 pb (ii). Ningún gen validó los resultados de los análisis llevados a cabo con la estrategia de filtrado de variantes de impacto funcional (iii). Después de la corrección por test múltiple ninguna de las asociaciones de la fase de descubrimiento o de la validación se mantuvo asociada.

Tabla 15. Genes asociados con sangrado rectal como efecto adverso a la radioterapia en pacientes con cáncer de próstata (fenotipo sangrado rectal extremo)

Análisis	Gen	Chr	Descripción	φFASE DE DESCUBRIMIENTO (RADIOGEN)			ψFASE DE VALIDACIÓN (REQUIRE)		
				V. incl.	Descripción de variante	P-valor	V. incl.	Descripción de variante	P-valor
1) Todas las variantes analizadas	SULT1C2P1	2	sulfotransferase family, cytosolic, 1C, member 2 pseudogene 1	4	1 variante ncRNA exónica y 3 ncRNA intrónicas	3,23x10 ⁻⁰⁵	2	1 variante ncRNA exónica y 1 ncRNA intrónica	1
	FLNC-AS1	7	FLNC antisense RNA 1	7	1 variante ncRNA exónica y 6 ncRNA intrónicas	4,52x10 ⁻⁰⁵	18	1 variante ncRNA exónica, 12 ncRNA intrónicas, 3 downstream y 2 upstream	0,192
	MIR4289	9	microRNA 4289	4	1 variante ncRNA exónica, 2 variantes intergénicas y 1 downstream	8,99x10 ⁻⁰⁵	2	2 variantes ncRNA exónicas	0,504
	SLC37A3	7	solute carrier family 37 member 3	13	1 variante missense, 11 intrónicas y 1 en UTR	2,42x10 ⁻⁰⁴	12	1 variante missense, 8 intrónicas y 2 en UTRs, 1 upstream	0,385
	*ADRM1	20	ADRM1 26S proteasome ubiquitin receptor	5	2 variantes missense, 2 sinónimas y 1 inframe insertion	2,59x10 ⁻⁰⁴	17	3 variantes missense, 1 inframe insertion, 2 sinónimas, 3 intrónicas y 8 downstream	0,048
	PRPF19	11	pre-mRNA processing factor 19	13	2 variantes sinónimas, 7 intrónicas y 4 en UTRs	3,06x10 ⁻⁰⁴	5	3 variantes intrónicas y 2 en UTRs	0,220

Análisis	Gen	Chr	Descripción	φFASE DE DESCUBRIMIENTO (RADIOGEN)			ψFASE DE VALIDACIÓN (REQUITE)		
				V. incl.	Descripción de variante	P-valor	V. incl.	Descripción de variante	P-valor
1) Todas las variantes analizadas	SGSM1	22	small G protein signaling modulator 1	49	5 variantes missense, 5 sinónimas, 26 intrónicas y 13 en UTRs	3,32x10 ⁻⁰⁴	24	3 variantes missense, 4 sinónimas, 13 intrónicas y 4 en UTRs	1
	PRR7	5	proline rich 7, synaptic	4	1 variante inframe insertion y 3 en UTRs	3,40x10 ⁻⁰⁴	4	3 variantes intrónicas y 1 en UTR	0,408
	RGS8	1	regulator of G protein signaling 8	14	1 variante missense, 1 sinónima, 8 intrónicas y 4 en UTRs	3,61x10 ⁻⁰⁴	11	1 variante missense, 1 sinónima, 4 intrónicas y 5 en UTRs	0,374
	TMTC4	13	transmembrane O-mannosyltransferase targeting cadherins 4	8	8 variantes intrónicas	4,02x10 ⁻⁰⁴	7	1 variante missense, 2 sinónimas, 1 splicing y 3 intrónicas,	0,370
	NFE2	12	nuclear factor, erythroid 2	5	1 variante sinónima, 3 intrónicas y 1 en UTRs	4,08x10 ⁻⁰⁴	3	2 variantes intrónicas y 1 en UTR	1
	ZNF385D	3	zinc finger protein 385D	12	2 variantes missense, 2 sinónima, 4 intrónicas y 4 en UTRs	4,34x10 ⁻⁰⁴	7	1 variante missense, 1 sinónima, 3 intrónicas y 2 en UTRs	0,360
	KCNE1B	21	potassium voltage-gated channel subfamily E regulatory subunit 1B	9	1 variante missense, 1 sinónima, 1 intrónica y 6 en UTRs	4,75x10 ⁻⁰⁴	14	1 variante missense, 2 intrónicas y 11 en UTRs	0,298

Análisis	Gen	Chr	Descripción	φFASE DE DESCUBRIMIENTO (RADIOGEN)			ψFASE DE VALIDACIÓN (REQUITE)		
				V. incl.	Descripción de variante	P-valor	V. incl.	Descripción de variante	P-valor
1) Todas las variantes analizadas	GLCE	15	glucuronic acid epimerase	7	2 variantes sinónima y 5 en UTRs	4,98x10 ⁻⁰⁴	3	1 variante missense, 1 sinónimas y 1 en UTRs	0,149
	CACNA1C-AS1	12	CACNA1C antisense RNA 1	11	4 variantes ncRNA exónicas y 7 ncRNA intrónicas	5,12x10 ⁻⁰⁴	0	sin información	1
	TIGD7	16	tigger transposable element derived 7	7	1 variante sinónima, 2 missense y 5 en UTRs	5,49x10 ⁻⁰⁴	1	1 variante en UTRs	0,290
	LRRC57	15	leucine rich repeat containing 57 fibroblast	5	1 variante missense, 2 intrónicas y 2 en UTRs	6,79x10 ⁻⁰⁴	4	1 variante intrónica y 3 en UTRs	0,339
	FGFBP3	10	growth factor binding protein 3	4	2 variantes missense y 2 en UTRs	8,27x10 ⁻⁰⁴	2	1 variante missense y 1 sinónimas	1
	SIX1	14	SIX homeobox 1	14	1 variante intrónica, 3 intergénicas, 1 upstream, 1 downstream y 8 en UTRs	9,04x10 ⁻⁰⁴	0	sin información	1
	DNAI4	1	dynein axonemal intermediate chain 4	12	1 variante frameshift deletion, 3 sinónimas, 7 intrónicas y 1 downstream	9,99x10 ⁻⁰⁴	6	1 variante missense, 2 sinónimas y 3 intrónicas	0,298

Análisis	Gen	Chr	Descripción	ΦFASE DE DESCUBRIMIENTO (RADIOGEN)			ψFASE DE VALIDACIÓN (REQUIRE)		
				V. incl.	Descripción de variante	P-valor	V. incl.	Descripción de variante	P-valor
2) Variantes exónicas (exclusión de sinónimas) y de <i>splicing</i> canónicas ± 2pb	<i>THEGL</i>	4	<i>theg spermatid protein like</i>	4	4 variantes <i>missense</i>	2,76x10 ⁻⁰⁴	1	1 variante <i>missense</i>	0,034
	<i>C9orf131</i>	9	<i>chromosome 9 open reading frame 131</i>	4	4 variantes <i>missense</i>	3,90x10 ⁻⁰⁴	4	4 variantes <i>missense</i>	0,409
	<i>THBS2</i>	6	<i>thrombospondin 2</i>	3	3 variantes <i>missense</i>	4,13x10 ⁻⁰⁴	1	1 variante <i>missense</i>	1
	* <i>BMPER</i>	7	<i>BMP binding endothelial regulator</i>	6	6 variantes <i>missense</i>	4,17x10 ⁻⁰⁴	1	1 variante <i>missense</i>	0,271
	<i>BSG</i>	19	<i>basigin (Ok blood group)</i>	2	2 variantes <i>missense</i>	4,78x10 ⁻⁰⁴	0	sin información	1
	* <i>METTL25</i>	12	<i>methyltransferase like 25</i>	4	4 variantes <i>missense</i>	5,84x10 ⁻⁰⁴	4	4 variantes <i>missense</i>	1
	<i>DVL1</i>	1	<i>dishevelled segment polarity protein 1</i>	3	3 variantes <i>missense</i>	6,64x10 ⁻⁰⁴	0	sin información	1
	<i>HMMR</i>	5	<i>hyaluronan mediated motility receptor</i>	2	2 variantes <i>missense</i>	7,89x10 ⁻⁰⁴	3	3 variantes <i>missense</i>	0,212
	<i>SPDL1</i>	5	<i>spindle apparatus coiled-coil protein 1</i>	2	2 variantes <i>missense</i>	9,23x10 ⁻⁰⁴	5	5 variantes <i>missense</i>	0,562
	* <i>ADRM1</i>	20	<i>ADRM1 26S proteasome ubiquitin receptor</i>	3	2 variantes <i>missense</i> y 1 <i>inframe insertion</i>	9,23x10 ⁻⁰⁴	6	6 variantes <i>missense</i>	0,080

Análisis	Gen	Chr	Descripción	φFASE DE DESCUBRIMIENTO (RADIOGEN)			ψFASE DE VALIDACIÓN (REQUIRE)		
				V. incl.	Descripción de variante	P-valor	V. incl.	Descripción de variante	P-valor
3) Variantes de impacto funcional en regiones exónicas (exclusión de sinónimas) y de <i>splicing</i> canónicas ±2 pb	<i>PLEKHH3</i>	17	<i>pleckstrin homology, MyTH4 and FERM domain containing H3</i>	3	3 variantes <i>missense</i>	2,88x10 ⁻⁰⁴	1	1 variante <i>missense</i>	0,421
	<i>*METTL25</i>	12	<i>methyltransferase like 25</i>	3	3 variantes <i>missense</i>	9,32x10 ⁻⁰⁴	1	1 variante <i>missense</i>	1

*Estos genes se encuentran asociados en dos de los análisis; el gen que validó se presenta en negrita; Chr, cromosoma; V. incl., variantes incluidas; φRADIOGEN, cohorte de descubrimiento; ψREQUIRE, cohorte de réplica. UTR, *untranslated region*; ncRNA, *non-coding RNA*; *upstream*, variante ubicada aguas arriba del gen; *downstream*, variante ubicada aguas abajo del gen.

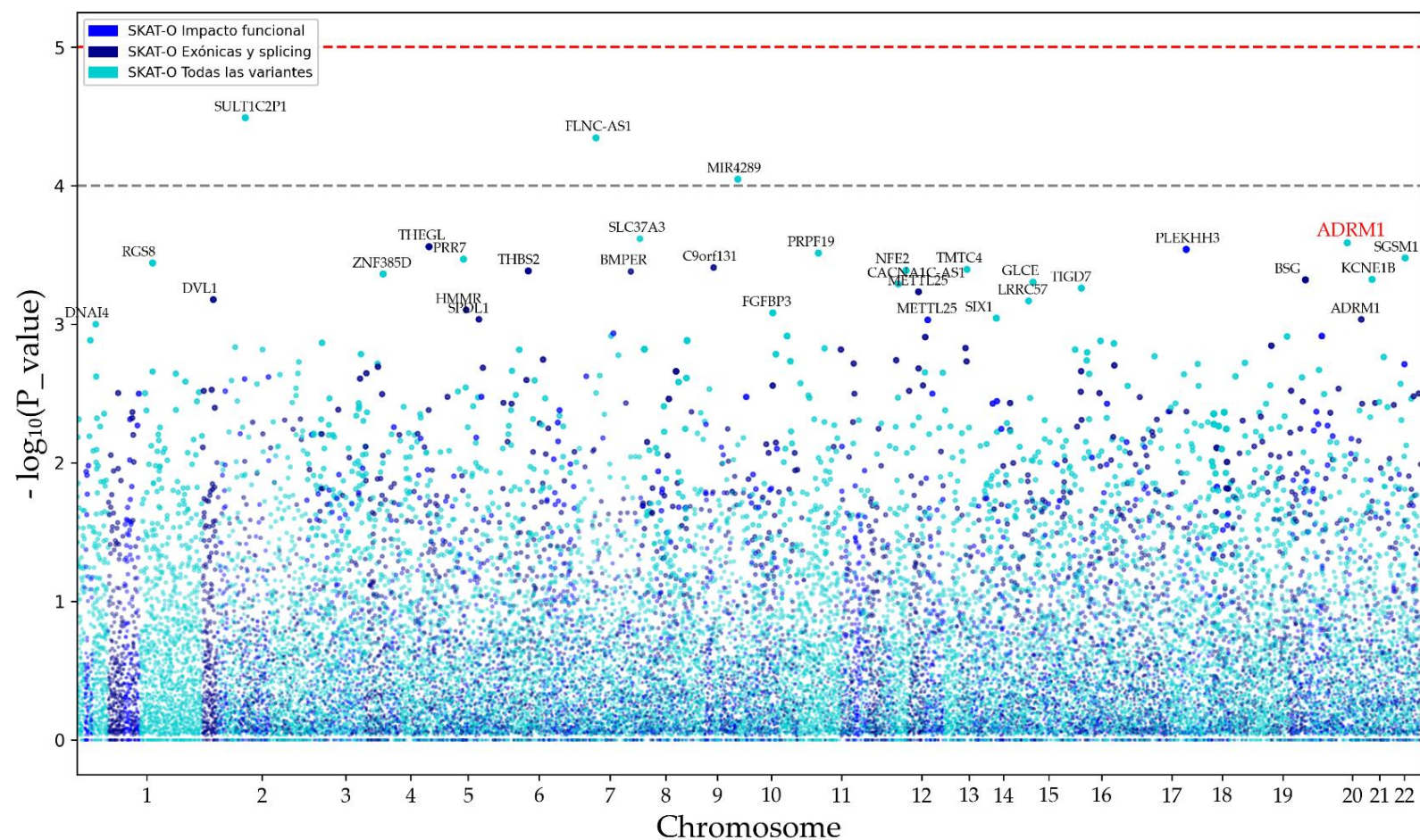


Figura 18. Manhattan plot de los genes asociados en los diferentes análisis SKAT-O, realizados en la fase de descubrimiento para el fenotipo sangrado rectal extremo. Los p-valores fueron transformados a-log. Se indican los nombres de los genes más asociados nominalmente, con color rojo se escriben el nombre del gene que validó en la cohorte de REQUITE. La línea gris representa el umbral de $-\log_{10}$ del p-valor 1×10^{-4} y la línea roja representa el $-\log_{10}$ del p-valor sugestivo 1×10^{-5} .

4.2.3 Análisis de asociación combinado de variantes comunes y raras

El análisis de asociación realizado incluyó 1 143 365 variantes comunes y raras. De los resultados obtenidos, se consideraron solo aquellos genes que tuvieran al menos dos marcadores analizados lo que resultó en un total de 21 629 genes. Se encontraron 27 genes asociados nominalmente con sangrado rectal, tres de ellos (*MBP*, *LRP1* y *PGPEP1L*) presentaron asociaciones nominales menores a $8,6 \times 10^{-05}$: el gen *MBP* obtuvo un p-valor de $2,87 \times 10^{-05}$, 112 variantes fueron analizadas para este gen, de las cuales 68 se clasificaron como variantes raras y 44 como variantes comunes. El segundo gen es *LRP1* (p-valor = $6,81 \times 10^{-05}$), se analizaron 140 variantes presentes en este gen, de ellas 112 eran variantes raras y 28 variantes comunes, el último gen es *PGPEP1L* cuyo p-valor fue de $8,52 \times 10^{-05}$, el análisis de este gen incluyó 25 variantes, las cuales 18 fueron variantes raras y siete variantes comunes. Además, cinco genes (*MBP*, *CACNA1C-AS2*, *LRRC70*, *RPB2* y *CACNA1C-AS1*) que presentaron asociaciones en este análisis, también presentaron asociaciones en los análisis de SKAT-O (ver **Tabla 14** y **Tabla 15**). Tras la corrección con BH-FDR ninguna de estas asociaciones se mantuvo significativa. Los genes más asociados se muestran en la **Tabla 16**. La representación grafica de los resultados de los genes analizados se muestra en la **Figura 19**.

Tabla 16. Genes asociados con sangrado rectal identificados en el estudio combinado de variantes comunes y raras.

Gen	Localización	Nº variantes analizadas	Nº variantes raras	Nº variantes comunes	P-valor
<i>MBP</i>	18q23	112	68	44	$2,87 \times 10^{-05}$
<i>LRP1</i>	12q13.3	140	112	28	$6,81 \times 10^{-05}$
<i>PGPEP1L</i>	15q26.3	25	18	7	$8,52 \times 10^{-05}$
<i>ZNF135</i>	19q13.43	34	17	17	$1,17 \times 10^{-04}$
<i>PRDX2</i>	19p13.13	17	11	6	$1,43 \times 10^{-04}$
<i>CACNA1C-AS2</i>	12p13.33	2	2	0	$1,57 \times 10^{-04}$
<i>RAET1L</i>	6q25.1	9	6	3	$1,92 \times 10^{-04}$
<i>TTC28</i>	22q12.1	43	31	12	$1,96 \times 10^{-04}$
<i>RASGRP3</i>	2p22.3	51	36	15	$2,54 \times 10^{-04}$
<i>HSCB</i>	22q12.1	11	7	4	$2,95 \times 10^{-04}$
<i>P2RX6</i>	22q11.21	7	4	3	$3,33 \times 10^{-04}$
<i>TMEM151A</i>	11q13.2	17	13	4	$3,59 \times 10^{-04}$
<i>CACNA1C-AS1</i>	12p13.33	16	12	4	$4,08 \times 10^{-04}$
<i>LRRC70</i>	5q12.1	2	2	0	$4,50 \times 10^{-04}$
<i>SLC22A3</i>	6q25.3	55	40	15	$4,64 \times 10^{-04}$
<i>BUB1</i>	2q13	38	32	6	$5,39 \times 10^{-04}$
<i>AFF3</i>	2q11.2	52	36	16	$6,13 \times 10^{-04}$
<i>MIR3663</i>	10q25.3	5	2	3	$6,29 \times 10^{-04}$
<i>TMEM104</i>	17q25.1	50	34	16	$6,37 \times 10^{-04}$
<i>MBNL1</i>	3q25.1-q25.2	29	20	9	$6,46 \times 10^{-04}$
<i>KRT18P55</i>	17q11.2	17	12	5	$6,65 \times 10^{-04}$
<i>PRB2</i>	12p13.2	9	8	1	$7,39 \times 10^{-04}$
<i>ATP2A1-AS1</i>	16p11.2	3	1	2	$7,83 \times 10^{-04}$
<i>CCDC117</i>	22q12.1	43	30	13	$8,03 \times 10^{-04}$
<i>NPIP9</i>	16p11.2	2	1	1	$8,27 \times 10^{-04}$
<i>TPH2</i>	12q21.1	28	18	10	$8,62 \times 10^{-04}$
<i>TRPM4</i>	19q13.33	93	69	24	$9,68 \times 10^{-04}$

Nº= número. Los genes más asociados se presentan en negrita

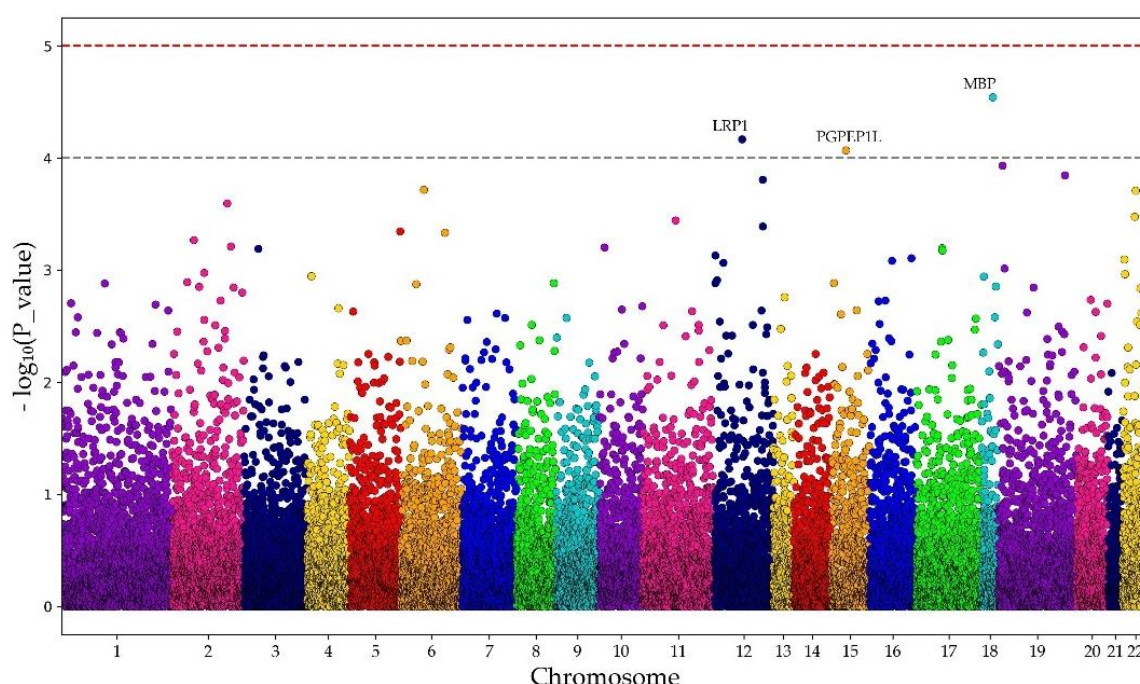


Figura 19. Manhattan plot de los genes asociados del análisis combinado de variantes comunes y raras. Los p-valores fueron transformados a $-\log_{10}$. Se indican los nombres de los genes más asociados nominalmente. La línea gris representa el umbral de $-\log_{10}$ del p-valor 1×10^{-04} y la línea roja representa el $-\log_{10}$ del p-valor sugestivo 1×10^{-05} .

4.3 CAPITULO 3. NUEVAS APROXIMACIONES DE ESTUDIOS DE ASOCIACIÓN ÓMICOS APLICADOS A RADIOGENÓMICA y ANÁLISIS DE RUTAS

4.3.1 Estudios de asociación del transcriptoma completo

Se realizaron dos tipos de análisis de asociación de TWAS, uno uni-tejido y otro multi-tejido. En el caso de los análisis multi-tejido, se llevaron a cabo dos análisis diferentes. El primero consideró seis tejidos cercanos a la próstata y la sangre total, mientras que el segundo incluyó 46 tejidos distintos presentes en GTEx V8 (ver apartado 3.9.1). Todos los análisis se llevaron a cabo utilizando una muestra de 141 individuos (46 casos y 95 controles), y se consideraron significativos aquellos resultados que superaron el umbral de Bonferroni.

En el análisis de tejidos cercanos a la próstata, se evaluaron un total 11 753 genes, y se encontraron asociaciones estadísticamente significativas ($BF < 4,2 \times 10^{-06}$) para los genes *AXL* (p-valor = $6,38 \times 10^{-17}$), *ITSN1* (p-valor = $2,93 \times 10^{-12}$) y *WWC1* (p-valor = $2,41 \times 10^{-08}$). Al explorar nuestros resultados encontramos que dos de los genes más significativos (*AXL* y *LTBP4*) se encontraban en el mismo *locus* a una distancia menor de 1 Mb. Por

otra parte, en el análisis que incluyó los 46 tejidos, se analizaron 12 883 genes y se identificaron asociaciones significativas $BF < 3,8 \times 10^{-06}$ para los genes *CPN1* (p-valor = $1,35 \times 10^{-16}$) y *MLEC* (p-valor = $1,02 \times 10^{-08}$). Los resultados de ambos análisis se presentan en la **Tabla 17**.

El miami plot (**Figura 20**) presenta los genes asociados en los análisis de TWAS multi-tejido, llevados a cabo en siete tejidos diferentes incluida la sangre total y de los análisis de 46 tejidos diferentes.

Tabla 17. Resultados del estudio TWAS multi-tejido de asociación de genes con sangrado rectal

Gen	Chr	Start	End	P-valor	n	z_mean	Análisis
<i>AXL</i>	19	41219223	41261766	$6,38 \times 10^{-17}$	3	0,69	*Tejidos cercanos a la próstata
<i>ITSN1</i>	21	33642400	33899861	$2,93 \times 10^{-12}$	2	1,51	*Tejidos cercanos a la próstata
<i>WWC1</i>	5	168291645	168472303	$2,41 \times 10^{-08}$	3	-0,44	*Tejidos cercanos a la próstata
<i>CYBRD1</i>	2	171522247	171558129	$5,14 \times 10^{-06}$	6	-0,16	*Tejidos cercanos a la próstata
<i>GDPD1</i>	17	59220467	59275970	$6,58 \times 10^{-06}$	2	0,74	*Tejidos cercanos a la próstata
<i>LIMS2</i>	2	127638381	127681786	$1,10 \times 10^{-05}$	5	0,76	*Tejidos cercanos a la próstata
<i>LTBP4</i>	19	40592883	40629818	$5,14 \times 10^{-05}$	2	-0,95	*Tejidos cercanos a la próstata
<i>CPN1</i>	10	100042193	100081869	$1,35 \times 10^{-16}$	2	1,71	46 tejidos
<i>MLEC</i>	12	120687149	120701859	$1,02 \times 10^{-08}$	15	1,11	46 tejidos
<i>QPRT</i>	16	29663279	29698699	$9,07 \times 10^{-06}$	30	-0,68	46 tejidos
<i>FGFR1OP</i>	6	166999317	167094789	$2,03 \times 10^{-05}$	25	-1,67	46 tejidos
<i>TRIM22</i>	11	5689697	5737089	$7,18 \times 10^{-05}$	30	-0,55	46 tejidos

Chr, cromosoma; Start, posición donde inicia el gen; End, posición final del gen; n, número de tejidos utilizados para el análisis S-MultiXcan; z_mean, es el valor medio de z. Un valor de z mayor o menor a cero indica el aumento o disminución de la expresión génica respectivamente, comparada con la expresión génica media en el grupo control. *seis tejidos cercanos a la próstata y la sangre total. Los genes más asociados se presentan en negrita

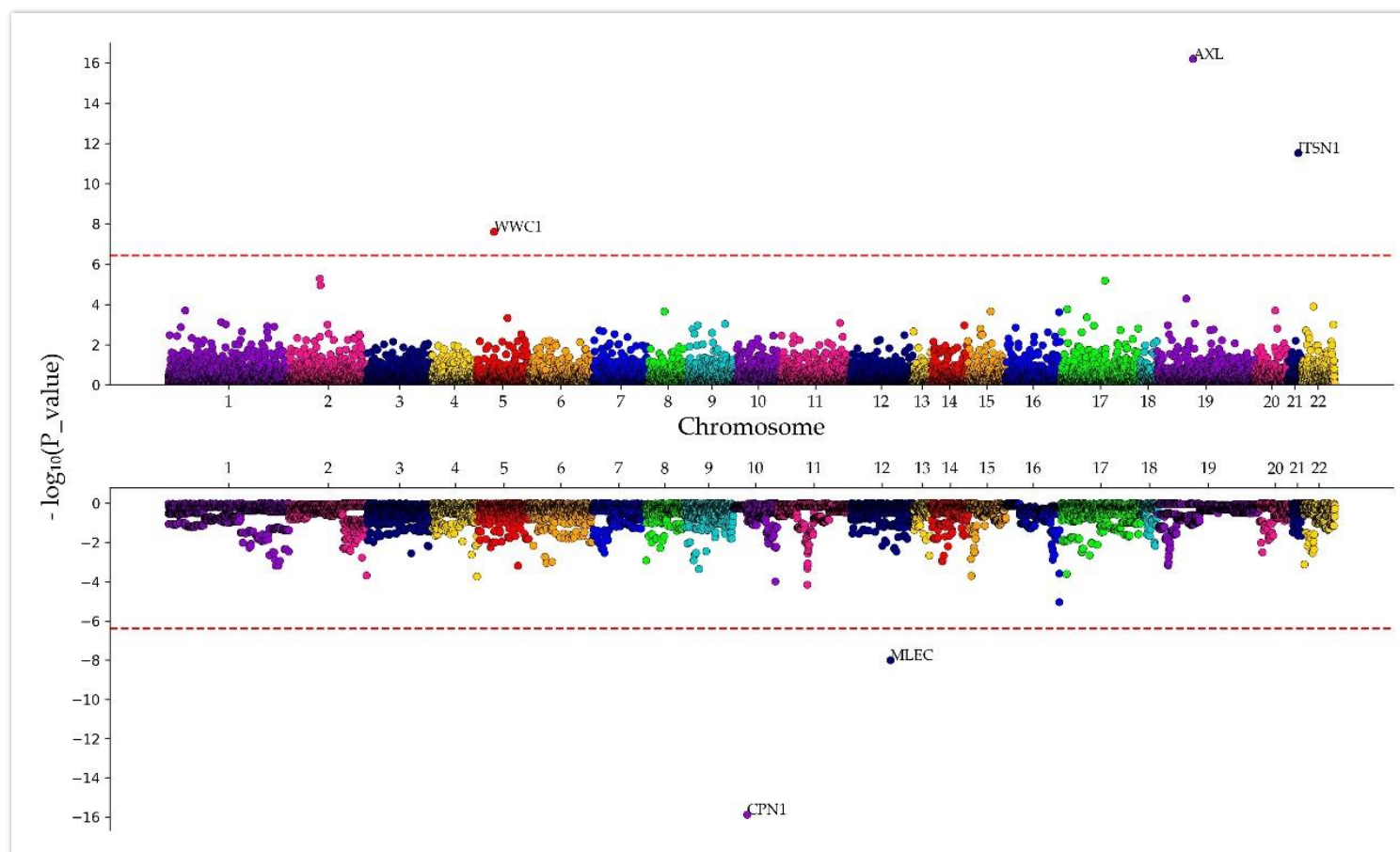


Figura 20. Miami plot de los análisis multi-tejido de TWAS. En la parte superior se muestran los genes estudiados en los tejidos cercanos a la próstata y la sangre total y en la parte inferior los genes analizados en 46 diferentes tejidos. Los p-valores fueron transformados a $-\log_{10}$. Se encuentran etiquetados los genes estadísticamente significativos. Las líneas rojas representan el $-\log_{10}$ del umbral de significancia de Bonferroni para cada estudio, $BF < 4,2 \times 10^{-6}$ para la parte superior y $BF < 3,8 \times 10^{-6}$ en la parte inferior

4.3.2 Estudios de asociación del proteoma completo

Para el análisis de asociación del proteoma completo (PWAS), se incluyeron 141 individuos de nuestra cohorte de estudio (46 casos y 95 controles). En nuestro análisis utilizamos el modelo de proteínas plasmáticas propuesto por Zhang et al., (2022)¹⁶² (ver apartado 3.9.2). En total se analizaron 1105 genes.

Entre los genes analizados, los que presentaron mejores valores de asociación estadística para nuestro fenotipo fueron: *NPTX2* (p-valor = $1,48 \times 10^{-03}$) y *THBS2* (p-valor = $2,52 \times 10^{-03}$). De estos genes, *THBS2* presenta un *zscore* de -3,02 lo que sugiere que la proteína presenta niveles más bajos en los individuos del grupo casos en comparación con el grupo control. Ninguna de estas asociaciones se mantuvo significativa después de aplicar la corrección por test múltiple con FDR.

Los resultados de los genes más significativos obtenidos de este análisis se presentan en la **Tabla 18**.

Tabla 18. Resultados del estudio de asociación PWAS de asociación de genes con sangrado rectal

Gen	Chr	Start	End	n	zscore	effect_size	P-valor
<i>NPTX2</i>	7	98617285	98629869	3	3,18	37,68	$1,48 \times 10^{-03}$
<i>THBS2</i>	6	169215780	169254050	3	-3,02	-14,66	$2,52 \times 10^{-03}$
<i>ALCAM</i>	3	105366909	105576900	2	-2,95	-27,26	$3,18 \times 10^{-03}$
<i>CXCL1</i>	4	73869393	73871308	2	-2,87	-7,56	$4,17 \times 10^{-03}$
<i>AGFG1</i>	2	227472152	227561217	1	-2,65	-1066,25	$7,96 \times 10^{-03}$
<i>STX8</i>	17	9250471	9576591	1	-2,63	-370,76	$8,45 \times 10^{-03}$
<i>EGF</i>	4	109912883	110013766	1	2,63	223,31	$8,45 \times 10^{-03}$

Chr, cromosoma; Start, posición de inicio del gen; End, posición final del gen; n, número de variantes analizadas; zscore, es el valor de z, una correlación significativa indica que existe diferencia entre casos y controles, lo que implica que la proteína presenta niveles de alteración más elevados o más bajos en los individuos afectados; effect_size, tamaño del efecto estimado.

En el gráfico de Manhattan plot (**Figura 21**) se muestran los dos genes más significativos identificados en el estudio de asociación del proteoma completo.

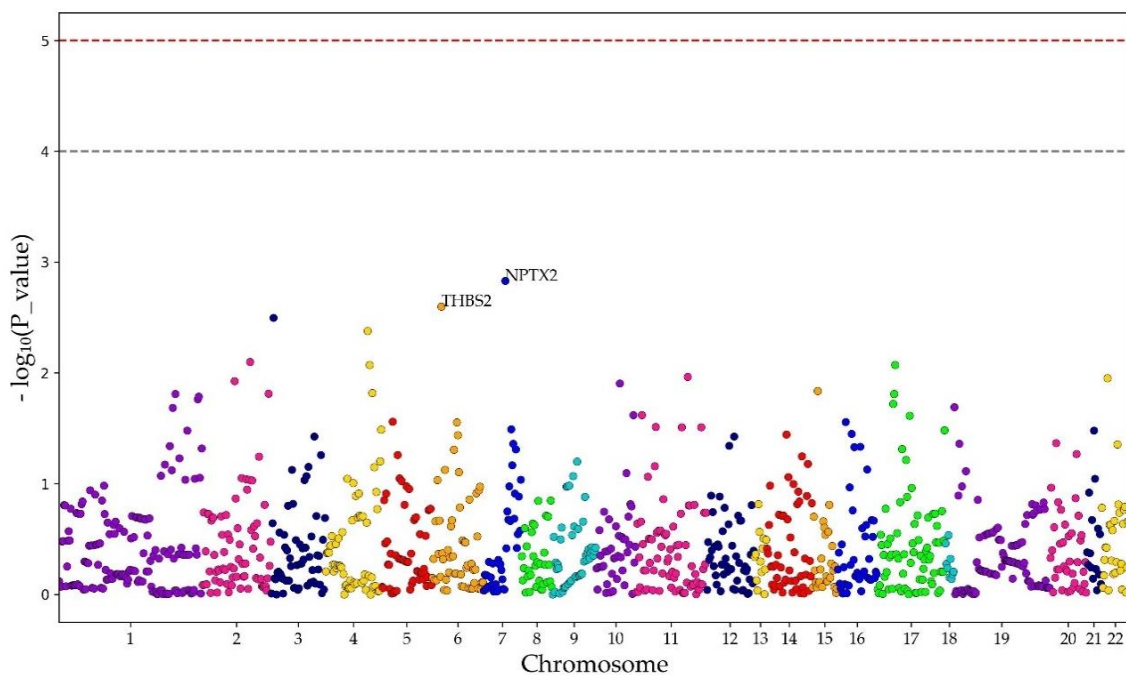


Figura 21. Manhattan plot de los genes asociados del análisis PWAS. Los p-valores fueron transformados a $-\log_{10}$. Se indican los nombres de los genes más asociados nominalmente. La línea gris representa el umbral de $-\log_{10}$ del p-valor 1×10^{-4} y la línea roja representa el $-\log_{10}$ del p-valor sugestivo 1×10^{-5}

4.3.3 Análisis de rutas

Para el análisis de rutas se consideraron 96 genes seleccionados a partir de los resultados obtenidos en los análisis SKAT-O, SKAT-C- *common-rare*, TWAS Y PWAS. El análisis se llevó a cabo siguiendo dos estrategias. En la primera estrategia examinamos si los genes incluidos formaban parte de alguna de las rutas interés previamente relacionadas con radiotoxicidad (ver apartado 3.9.3). De seis rutas analizadas, se identificaron cinco genes en tres de ellas (**Tabla 19**).

Tabla 19. Genes localizados en rutas conocidas

Gen	Ruta	Localización	*P-valor	ESTUDIO DE REFERENCIA
<i>PRPF19</i>	Reparación del ADN	11q12.2	$3,06 \times 10^{-4}$	SKAT-O-Fenotipo de sangrado extremo (filtrado todas las variantes)
<i>CXCL1</i>	Inflamación Inmunológica	4q13.3	$4,17 \times 10^{-3}$	PWAS
<i>MBP</i>	Inmunológica	18q23	$1,07 \times 10^{-4}$ / $2,87 \times 10^{-5}$	SKAT-O-Fenotipo de sangrado (filtrado todas las variantes) / SKATC-Common-rare
<i>PSMB8</i>	Inmunológica	6p21.32	$5,08 \times 10^{-4}$ / $6,28 \times 10^{-4}$	SKAT-O-Fenotipo de sangrado (filtrado de variantes exónicas- <i>splicing</i> e impacto funcional)
<i>TNFRSF10C</i>	Inmunológica	8p21.3	$7,22 \times 10^{-4}$	SKAT-O-Fenotipo de sangrado (filtrado todas las variantes)

*p-valor: Es el p valor obtenido en el estudio de referencia.

En la segunda estrategia, para el análisis de enriquecimiento de conjuntos de genes usamos la herramienta GENE2FUNC de FUMA. Se encontraron tres conjuntos de genes con asociaciones significativas en la categoría que examina la expresión de alteraciones genéticas y químicas. Entre ellos destaca *“Warters response to IR skin”*. Además, encontramos señales de enriquecimiento en conjuntos de genes para la ruta de Reactome *TP53 regulates transcription of death receptors and ligands* y para rutas implicadas en la fase de mitosis del ciclo celular (*Reactome cell cycle mitotic* y *GO regulation of mitotic cell cycle*). Para estas tres rutas las asociaciones fueron nominales. Los resultados más significativos se muestran en la **Tabla 20**.

Tabla 20. Resultados más significativos de los *sets* de genes/rutas con FUMA

Categoría	GeneSet	N. genes overlap	P-valor	P-valor ajustado	Genes
Chemical and Genetic perturbation	Boquest stem cell up	7	5,47x10 ⁻⁰⁶	1,81x10 ⁻⁰²	LRP1, MBP, AXL, LRP1B, CYBRD1, SLC22A3, THBS2
Chemical and Genetic perturbation	Charafe breast cancer luminal vs basal dn	8	2,43x10 ⁻⁰⁵	4,01x10 ⁻⁰²	TRIM22, MBP, AXL, CYBRD1, MBNL1, CXCL1, PSMB8, TNFRSF10D
Chemical and Genetic perturbation	Warters response to IR skin	4	6,01x10 ⁻⁰⁵	4,99x10 ⁻⁰²	FGFBP3, TRIM22, ALCAM, TNFRSF10C
Chemical and Genetic perturbation	Dacosta UV response via ERCC3 dn	10	7,57x10 ⁻⁰⁵	5,00x10 ⁻⁰²	R3HDM2, AGFG1, ITSN1, TTC28, ALCAM, MBNL1, HMMR, WWC1, THBS2, NCOA2
Reactome	Reactome TP53 regulates transcription of death receptors and ligands	2	4,30x10 ⁻⁰⁴	2,06x10 ⁻⁰¹	TNFRSF10C, TNFRSF10D
Reactome	Reactome degradation of DVL	3	4,44x10 ⁻⁰⁴	2,06x10 ⁻⁰¹	DVL1, PSMC1, PSMB8
Reactome	Reactome cell cycle mitotic	7	4,90x10 ⁻⁰⁴	2,06x10 ⁻⁰¹	PSMC1, BUB1, HMMR, SPD1, PSMB8, FGFR1OP, NUP188
Reactome	Reactome M phase	6	5,49x10 ⁻⁰⁴	2,06x10 ⁻⁰¹	PSMC1, BUB1, SPD1, PSMB8, FGFR1OP, NUP188
GO bp	GO positive regulation of signaling	15	6,21x10 ⁻⁰⁵	1,81x10 ⁻⁰¹	DVL1, FGFBP3, TRIM22, LRP1, PSMC1, PRDX2, AXL, TRPM4, LIMS2, ITSN1, EGF, WWC1, PRR7, PSMB8, BMPER
GO bp	GO chemokine C C motif ligand 2 secretion	2	9,88x10 ⁻⁰⁵	1,81x10 ⁻⁰¹	LRP1, MBP
GO bp	GO positive regulation of molecular function	14	1,50x10 ⁻⁰⁴	1,81x10 ⁻⁰¹	RGS8, TRIM22, LRP1, MBP, PRDX2, RASGRP3, AGFG1, ADRM1, ITSN1, SGSM1, HSCB, CXCL1, EGF, PRR7
GO bp	GO signal transduction by protein phosphorylation	10	1,80x10 ⁻⁰⁴	1,81x10 ⁻⁰¹	LRP1, PSMC1, MBP, PRDX2, CBLC, RASGRP3, EGF, WWC1, PSMB8, BMPER

Categoría	GeneSet	N. genes overlap	P-valor	P-valor ajustado	Genes
GO bp	<i>GO cell surface receptor signaling pathway involved in cell cell signaling</i>	8	1,93x10 ⁻⁰⁴	1,81x10 ⁻⁰¹	<i>DVL1, LRP1, PSMC1, AXL, TRPM4, P2RX6, EGF, PSMB8</i>
GO bp	<i>GO positive regulation of catalytic activity</i>	12	2,58x10 ⁻⁰⁴	1,81x10 ⁻⁰¹	<i>RGS8, MBP, PRDX2, RASGRP3, AGFG1, ADRM1, ITSN1, SGSM1, HSCB, CXCL1, EGF, PRR7</i>
GO bp	<i>GO cell-cell signaling</i>	13	3,02x10 ⁻⁰⁴	1,81x10 ⁻⁰¹	<i>DVL1, FGFBP3, LRP1, PSMC1, MBP, LTBP4, AXL, TRPM4, P2RX6, EGF, PRR7, PSMB8, NPTX2</i>
GO bp	<i>GO regulation of Wnt signaling pathway</i>	6	3,11x10 ⁻⁰⁴	1,81x10 ⁻⁰¹	<i>DVL1, LRP1, PSMC1, TRPM4, EGF, PSMB8</i>
GO bp	<i>GO regulation of mitotic cell cycle</i>	8	3,21x10 ⁻⁰⁴	1,81x10 ⁻⁰¹	<i>PSMC1, BUB1, TTC28, EGF, HMMR, SPD1, PSMB8, FGFR1OP</i>
Hallmark gene sets	<i>Hallmark heme metabolism</i>	4	1,85x10 ⁻⁰³	9,25x10 ⁻⁰²	<i>NFE2, RANBP10, BSG, PRDX2</i>
Hallmark gene sets	<i>Hallmark reactive oxygen species pathway</i>	2	7,21x10 ⁻⁰³	1,80x10 ⁻⁰¹	<i>MBP, PRDX2</i>
Hallmark gene sets	<i>Hallmark epithelial mesenchymal transition</i>	3	1,51x10 ⁻⁰²	1,91x10 ⁻⁰¹	<i>LRP1, CXCL1, THBS2</i>
Hallmark gene sets	<i>Hallmark G2M checkpoint</i>	3	1,53x10 ⁻⁰²	1,91x10 ⁻⁰¹	<i>BUB1, EGF, HMMR</i>
Hallmark gene sets	<i>Hallmark coagulation</i>	2	4,93x10 ⁻⁰²	3,96x10 ⁻⁰¹	<i>CPN1, LRP1</i>
Hallmark gene sets	<i>Hallmark UV response up</i>	2	6,34x10 ⁻⁰²	3,96x10 ⁻⁰¹	<i>BSG, NPTX2</i>

Gene Set, conjunto de genes; N. genes overlap, número de genes que se superponen, Chemical and Genetic perturbation, conjunto de genes que representan firmas de expresión de perturbaciones genéticas y químicas; Reactome, es una base de datos de código abierto de rutas; GO bp, Gene Ontology de procesos biológicos; Hallmark gene sets, representan estados o procesos biológicos específicos bien definidos. Con negrita se presentan los conjuntos de genes con asociaciones estadísticamente significativas.

5 DISCUSIÓN

La sensibilidad a la radioterapia puede ser definida como un rasgo genéticamente complejo. Los estudios de asociación gen-candidato y los GWAS han permitido la identificación de algunas regiones y variantes genéticas asociadas con el desarrollo de efectos adversos a la radioterapia. La mayoría de estas variantes son frecuentes y están localizadas en regiones no codificantes del genoma, lo que dificulta su interpretación y la de los mecanismos moleculares subyacentes. Además, determinar la composición alélica de los rasgos complejos no es una tarea sencilla, debido a que pueden participar tanto variantes comunes como raras a diferentes niveles²⁰⁴. La incorporación de las tecnologías de secuenciación NGS nos brinda la oportunidad de analizar ambos tipos de variantes de forma conjunta y nos acerca a una mejor comprensión de su implicación con el fenotipo.

Conocer las bases moleculares de la radiosensibilidad nos permitirá, entender mejor su contribución en el desarrollo de efectos adversos, así como su posible aplicación en la búsqueda de tratamientos individualizados destinados a minimizar el riesgo de experimentar dichos efectos. Esto es de especial relevancia para los pacientes con cáncer de próstata, los cuales presentan alta tasa de supervivencia, y para quienes los efectos adversos impactan significativamente en su calidad de vida. Uno de los efectos adversos más relevantes en estos pacientes es el sangrado rectal, que en sus grados más severos puede llegar a poner en riesgo la vida del paciente.

Con el fin de identificar nuevas variantes y genes de susceptibilidad a la radiación ionizante, analizamos el sangrado rectal como efecto adverso a la radioterapia, a partir de la secuenciación del exoma completo en la cohorte de pacientes con cáncer de próstata RADIOGEN (141 pacientes; 46 casos y 95 controles). Consideramos diferentes estrategias de análisis que incluyen: el análisis y la validación de variantes genéticas previamente descritas, la búsqueda de nuevas variantes en genes en los que previamente se identificaron variantes candidatas y la identificación de nuevos genes relacionados con el fenotipo a estudio, así como la validación de los genes asociados en una cohorte externa. Además, exploramos un fenotipo extremo de toxicidad, realizamos la identificación de genes de susceptibilidad mediante análisis multi-ómicos (TWAS y PWAS) y por último exploramos las rutas en las que participan los genes que presentaron resultados de asociación más significativos en los diferentes análisis realizados.

En el **Capítulo 1**, analizamos variantes y genes previamente asociados con el desarrollo de efectos adversos a la radioterapia en pacientes con cáncer, independientemente de su localización tumoral. Seleccionamos de la bibliografía variantes que tuvieran un p-valor de metaanálisis/combinado menor que 1×10^{-5} o p-valor menor a 0,05 corregido por Bonferroni. De las 152 variantes que cumplían con estos criterios solo siete estaban disponibles en nuestros datos de exomas. Seis de las variantes habían sido identificadas en estudios de asociación en pacientes con cáncer de próstata, en donde se habían considerado los fenotipos disfunción eréctil, frecuencia urinaria, sangrado rectal y

hematuria. La séptima variante se identificó en un estudio que incluyó pacientes con cáncer de cabeza y cuello y el fenotipo estudiado fue mucositis.

De las siete variantes analizadas en el estudio univariable (apartado 3.7.1), solo el SNP rs144214859 localizado a 500 bases *downstream* del gen *PLEK*, resultó asociado con sangrado rectal, con un p-valor nominal de 0,04. Este gen, codifica la proteína pleckstrina, principal sustrato de la proteína quinasa C de plaquetas y leucocitos²⁰⁵. Se ha observado que las plaquetas deficientes de pleckstrina presentan alteraciones en la agregación y secreción plaquetaria, así como en la activación de $\alpha\text{IIb}\beta 3$ ^{206–208}. Defectos en $\alpha\text{IIb}\beta 3$ se han relacionado con al desarrollo de diátesis hemorrágica en pacientes con síndromes mielodisplásicos^{207,209}. Además, se ha descrito que alteraciones en la coagulación sanguínea podrían predisponer al desarrollo de hemorragias, incluyendo la diátesis hemorrágica, en lesiones inducidas por la radiación^{210–212}. En conjunto, estos estudios apoyan la implicación de *PLEK* en la regulación de la coagulación sanguínea y sugieren una posible relación en el desarrollo de sangrado en pacientes sometidos a radioterapia.

En un estudio realizado en pacientes con cáncer de próstata tratados con radioterapia por el RGC¹²⁵, se identificó a la variante rs144214859 asociada con hematuria (p-valor = 2.5×10^{-07}). Aunque la hematuria tiene una fisiopatología distinta en la que intervienen, entre otros, cambios vasculares progresivos²¹³, podría existir un nexo con el sangrado rectal. La identificación de una variante en el gen *PLEK* en ambas toxicidades, apoyaría la hipótesis de que una coagulopatía inducida por la radiación puede provocar sangrado rectal⁶⁶. La localización del tumor y, por lo tanto, la distribución de la dosis en los órganos en riesgo podría determinar si el paciente presenta hematuria o sangrado rectal.

La asociación que encontramos para la variante rs144214859 en nuestro estudio univariable no permaneció significativa después de corrección por test múltiple, así como tampoco alcanzó el umbral de significancia para estudios GWAS (5×10^{-08}) en el estudio de referencia¹²⁵. Además, mientras que nuestro estudio se realizó con secuenciación NGS, en el estudio del RGC se utilizaron datos de genotipado. Y aunque el estudio original no estuvo diseñado para evaluar la asociación de variantes raras (MAF<0.05), se trató de un hallazgo importante por su probable implicación en el desarrollo de toxicidad inducida por la radiación.

En el análisis multivariable de las variantes descritas previamente en asociación con toxicidad a la radioterapia (apartado 3.7.2) ninguna de las siete variantes se asoció con nuestro fenotipo de estudio. Sin embargo, el SNP rs4804134, localizado en el gen *PDE4A* se encuentra en la misma dirección del efecto que el estudio original (OR 0,73; 95% CI [0,45-1,17]; p-valor = 0,191). Este SNP fue descrito por Kerns et al.¹³, asociado con sangrado rectal. Resultados similares se encontraron en dos de las cohortes incluidas en el metaanálisis (cohorte-Gene-PARE n= 235; OR=0,69; 95% CI [0,42-1,12]; p-valor =

0,131; y cohorte-CCI $n=136$; $OR=0,68$; 95% CI $[0,38-1,22]$; $p\text{-valor}=0,186$), que cuentan con tamaños de muestra y proporción de casos y controles similares al nuestro. Nuestros resultados son consistentes con lo esperado para el tamaño de la muestra utilizada.

Además, realizamos la validación del SNP rs4804134, con el modelo utilizado en el estudio original, donde se identificó asociado con sangrado rectal tardío ⁶¹ (apartado 3.7.3). Nuestro resultado no validó la asociación reportada para este SNP. Sin embargo, hay que considerar que debido a nuestro tamaño de muestra el estudio tiene bajo poder para detectar asociaciones significativas (necesitaríamos analizar 1645 pacientes para alcanzar un poder estadístico del 80 %), a menos que la magnitud del efecto de los SNPs estudiados fuera alta, que no es este caso ($OR\ 0,71$). Esto se observó en uno de los primeros GWAS de radiogenómica en cáncer de próstata, donde sólo se incluyeron 79 pacientes y se identificó al SNP rs2268363, localizado en el gen *FSHR* asociado de forma significativa ($OR\ 7,03$ 95% CI $[3,4-14,7]$; $p\text{-valor}$ corregido por Bonferroni 0,028) con la disfunción eréctil¹⁰⁷.

Como parte del **Capítulo 1**, se llevó a cabo un análisis para la identificación de nuevas variantes candidatas de impacto funcional, localizadas en genes previamente asociados con efectos adversos a la radioterapia (apartado 3.7.4). Se identificaron un total de 60 variantes en estos genes, trece no presentaron asociación con el fenotipo de estudio y 47 no pudieron estudiarse debido a que su ausencia en el grupo caso o control, no permitió su análisis en el modelo establecido (ANEXO 6). Ninguna de las trece variantes se encuentra en desequilibrio de ligamiento con las variantes previamente descritas.

A pesar de no haber encontrado nuevas variantes asociadas significativamente en genes previamente asociados con efectos adversos a la radioterapia, observamos que la variante rs144753108, que presenta la mayor asociación estadística de las variantes analizadas ($OR=0,68$; 95% CI $[0,38-1,22]$; $p\text{-valor}=0,110$), se localiza en el gen *TANC1*. Cabe destacar que este gen ha sido previamente estudiado por nuestro grupo, en donde se encontraron SNPs asociados con toxicidad global tardía en pacientes tratados con radioterapia¹¹⁶. Por otro lado, al examinar las 47 variantes que debido a su baja frecuencia no pudieron ser analizadas en el modelo, observamos que 32 de ellas estaban presentes sólo en el grupo control, todas ellas en heterocigosis. Quince variantes se encontraron sólo en el grupo de casos (14 estaban en heterocigosis y una en homocigosis), tres de ellas (rs76554185, rs150879175 y rs61731497) estuvieron presentes en más de un caso. Al comparar las frecuencias de las tres variantes entre el grupo casos y la frecuencia en población europea no finlandesa reportada por gnomAD observamos que dos variantes localizadas en el gen *LRP1B*, de ellas la variante rs76554185 mostró una frecuencia ligeramente mayor en el grupo de casos que en la frecuencia documentada en gnomAD-exoma en población europea (3,2 % vs 2,6 %, respectivamente), mientras que la variante rs150879175 presentó una frecuencia del 2,22

% en nuestro grupo de casos comparada con la frecuencia poblacional de 0,5 %. Por otro lado, la variante rs61731497 localizada en el gen *AGT* mostró una diferencia del 2 % entre la frecuencia del grupo casos y la frecuencia poblacional (2,2 % y 0,1 %, respectivamente). Además, encontramos que la única variante en homocigosis detectada en casos y ausente en controles fue la variante *missense* rs758237953 del gen *TANC1*. Esta variante presentó una frecuencia alélica notablemente mayor en el grupo casos (2 %), que en la frecuencia poblacional (0,001 %). A partir de estas observaciones, decidimos analizar las 47 variantes con el test de X^2 . Los resultados mostraron asociaciones nominales menores a 0,05 para las cuatro variantes previamente mencionadas (ANEXO 7), sin embargo, ninguna de estas asociaciones se mantuvieron significativas después de las correcciones por test múltiple.

En los resultados de la búsqueda de nuevas variantes en genes previamente descritos, es necesario considerar que los estudios de asociación clásicos utilizados para analizar variantes comunes tienen, en general, baja potencia para detectar asociaciones en variantes raras y de baja frecuencia en cohortes de pequeño tamaño, a menos que la magnitud del efecto de estas variantes sea significativamente grande^{140,141,214}. En nuestro estudio, para intentar paliar el efecto del tamaño muestral reducido, implementamos una estrategia de análisis de fenotipos extremos de toxicidad, lo cual ayuda a reducir el tamaño de muestra requerido para encontrar asociaciones¹⁴⁴. Además, seleccionamos variantes de impacto funcional, debido a que estas variantes tienen mayor probabilidad de estar asociadas con el fenotipo a estudio²¹⁵. A pesar de ello, nuestros resultados no muestran asociación con el fenotipo, aunque no podemos excluir la posibilidad de que alguna de estas variantes pudiera contribuir a la radiosensibilidad y que no seamos capaces de detectarlas.

Por otro lado, también existe la posibilidad de que los resultados reportados, que asocian variantes con los efectos adversos a la radioterapia, sean el producto de estimaciones de asociación sobrevaloradas. Se ha descrito un fenómeno denominado “winner's curse”, que se refiere a la tendencia de sobreestimar la magnitud del efecto (logaritmo de *odds ratio* o coeficiente de regression beta) o significancia estadística (p-valor) de una variante genética asociada en un estudio inicial, en comparación con sus valores reales²¹⁶. En consecuencia, si el cálculo del tamaño de muestra para un estudio de réplica se basa en la magnitud del efecto sobreestimado, es probable que los estudios de replicación no tengan la potencia suficiente para detectar la verdadera asociación^{216,217}.

En el **Capítulo 2**, para identificar nuevos genes relacionados con sangrado rectal llevamos a cabo un análisis basado en genes, el cual evalúa el efecto acumulado de variantes en un mismo gen. Para ello incluimos dos fenotipos de estudio. El fenotipo sangrado rectal (sangrado rectal grado ≥ 2 para casos y grado = 0 para controles) y el fenotipo de sangrado rectal extremo (sangrado rectal grado ≥ 3 para casos y grado = 0 para controles, con V70 mayor a 17,80Gy). Cada fenotipo de estudio se analizó en dos

fases: la fase de descubrimiento en la cohorte RADIOGEN^{98,116} y la fase de validación en la cohorte REQUITE¹⁹². Los análisis de cada estudio estuvieron conformados por tres estrategias de filtrado de variantes diferentes (apartado 3.8.1.1).

En la fase de descubrimiento del fenotipo sangrado rectal que incluyó a 141 pacientes (46 casos y 95 controles), encontramos 32 genes asociados con p-valores nominales menores a $9,7 \times 10^{-04}$. De éstos, cinco genes (*LRRC70*, *MYBPC2*, *NUP188*, *PSMB8* y *ZCCHC2*) mostraron asociaciones nominales en al menos dos de los análisis con las tres estrategias de filtrado de variantes evaluadas. El gen *LRRC70* codifica para una proteína que contiene repeticiones ricas en leucinas (LRR), que son homólogos a las LRRs de la glicoproteína V de las plaquetas (GPV)²¹⁸, la delección de GVP puede modificar la agregación plaquetaria en respuesta a la trombina²¹⁹. Estas alteraciones aumentan la probabilidad de sufrir síndromes hemorrágicos, lo cual podría sugerir una posible relación de alteraciones en *LRRC70* con el riesgo a sangrados. Otro gen interesante es *RAVER1*, en nuestro estudio de SKAT-O variantes en este gen estuvieron presentes solo en el grupo de casos y ausentes en el grupo control. La proteína codificada por este gen se ha asociado con la diferenciación y el mantenimiento muscular, así como a elementos contráctiles citoplasmáticos del músculo. La proteína Raver1 se encuentra presente en las capas de músculo liso de los vasos sanguíneos y de la vejiga urinaria, en acumulaciones citoplasmáticas en músculo esquelético y en corazón²²⁰. Además, en un estudio realizado en células C2C12 de ratón, se observó la localización de Raver1 en mioblastos indiferenciados, así como durante la formación y diferenciación a miotubos maduros, demostrando con esto su importancia en el proceso de diferenciación muscular²²¹.

Otro gen asociado en la fase de descubrimiento fue el gen *LRP1B*. Este gen supresor de tumores forma parte de la familia de receptores de lipoproteínas de baja densidad (LDL). Se ha descrito su expresión en células del músculo liso de la pared arterial²²² y su unión a diversos ligandos como por ejemplo, al activador de plasminógeno uroquinasa, al activador de plasminógeno de tipo tisular, al inhibidor-1 del activador de plasminógeno²²³, así como a la vitronectina, la apolipoproteína E y al fibrinógeno²²⁴. La mayoría de estos ligandos son factores de la coagulación sanguínea, lo cual sugiere el rol de *LRP1B* en la modulación de ésta y su posible asociación con la sensibilidad a sufrir sangrados relacionados con el tratamiento radioterápico. Esta posible asociación se ve reforzada con los hallazgos de los estudios de Kerns et al. en 2011¹⁰⁷ y 2020¹²⁵, llevados a cabo en pacientes con cáncer de próstata. Los estudios asocian dos variantes en este gen con efectos adversos a la radioterapia. En el estudio del año 2011¹⁰⁷, la variante rs10210358 se asoció (OR = 5,34; p-valor nominal $2,01 \times 10^{-06}$) con el desarrollo de disfunción eréctil como efecto adverso a la radioterapia. Mientras que en el estudio del año 2020¹²⁵, llevado a cabo en variantes con MAF < 0,05 los autores encontraron asociación de la variante rs75988504 con la disminución del flujo urinario como efecto adverso al tratamiento radioterápico (HR = 2,01; 95% CI [1,56 - 2,58]; p-meta = $4,9 \times 10^{-08}$). Pese a que los fenotipos y las variantes analizadas en los estudios son diferentes,

hay que señalar que tanto nuestros resultados como los de Kerns et al.^{125,107}, sugieren un vínculo entre el funcionamiento de este gen y las susceptibilidad a la radiación ionizante. A pesar de que estos genes no fueron significativos en la cohorte de validación, los mecanismos y funciones en los que están involucrados podrían tener relevancia en nuestro estudio.

Los 32 genes asociados nominalmente en los diferentes análisis de la fase de descubrimiento fueron analizados en la cohorte de validación REQUITE (69 casos y 322 controles). En total, dos genes validaron nominalmente con valores de p menores a 0,05. Entre los genes que validaron se encuentra el gen *PRB1*. Este gen codifica un precursor de proteínas salivales ricas en prolina y de acuerdo con el proyecto *Human Protein Atlas* se expresa en glándulas salivares²²⁵. Un estudio reciente²²⁶ investigó los efectos de la radiación en la expresión génica y proteica en las glándulas salivales de ratas. En dicho estudio se relacionó junto con otros genes a *PRB1*, con daños en la producción y secreción de saliva, así como daños en el ciclo celular asociados con la radioterapia²²⁶.

Otro de los genes asociado de forma nominal en la cohorte de validación fue *RANBP10*. RAN es una pequeña GTPasa involucrada en el ensamblaje de microtúbulos para formar husos mitóticos. Se ha descrito el papel esencial de *RANBP10* en la regulación de la red de microtúbulos, de megacariocitos y en la coordinación del ensamblaje y la liberación de plaquetas²²⁷. Además, ha sido observado que ratones con el gen *RANBP10* mutado presentaban un tiempo de sangrado notablemente prolongado y reducción de liberación de gránulos de plaquetas (proceso dependiente de la contracción interna de los microtúbulos)²²⁸, sugiriendo la importancia del gen en el proceso de hemostasia y en el mantenimiento de la dinámica de los microtúbulos en términos de la morfología y función de las plaquetas. Por otro lado, se ha documentado que la exposición a la radiación puede ocasionar cambios en la distribución de la red de microtúbulos y las curvaturas de las fibras en células Hs-578Bst²²⁹. En resumen, los hallazgos nos sugieren que alteraciones en este gen podrían relacionarse con alteraciones en la hemostasia, en la función de los microtúbulos y estos a su vez podrían presentar sensibilidad a la radiación, lo que indicaría un probable papel de *RANBP10* en la predisposición de los pacientes al desarrollo de sangrado rectal posterior al daño inducido por la radiación ionizante.

Los resultados obtenidos en nuestro estudio considerando el fenotipo sangrado rectal son comparables a los de estudios previos. Por un lado, implican genes que pueden estar participando en la regulación de la coagulación sanguínea y por el otro lado a genes que puede estar relacionados con el mantenimiento muscular. En este último caso, nuestro grupo de investigación ha reportado previamente variantes en el gen *TANC1* asociadas con el desarrollo de efectos tardíos en pacientes con cáncer de próstata tratados con radioterapia¹¹⁶. Alteraciones en el funcionamiento de este gen podría dificultar la regeneración del tejido muscular dañado. En un estudio *in vitro* en células C2C12 llevado a cabo en nuestro laboratorio, se observó que la disminución o

pérdida de la función de *TANC1* influye en la capacidad de la formación de miotubos maduros debido a la incorrecta fusión de mioblastos, lo cual alteraría la capacidad de regeneración del músculo. Este efecto se ve aumentado cuando las células son irradiadas observándose que a medida que la dosis de radiación aumenta, la formación de miotubos maduros disminuye²³⁰. Nuestros análisis arrojan resultados muy interesantes, que podrían ayudarnos a entender cómo alteraciones en el funcionamiento o en los mecanismos en que se encuentran involucrados estos genes podrían predisponer a desarrollar efectos adversos a pacientes tratados con radioterapia.

Para el análisis del fenotipo sangrado rectal extremo, en la fase de descubrimiento se incluyeron 59 pacientes de la cohorte RADIOGEN (29 casos y 30 controles). De este estudio, 32 genes presentaron asociaciones con p-valores nominales menores de $9,4 \times 10^{-04}$. De ellos, el gen *ADRM1*, asociado con la adhesión celular, y el gen *METTL25*, con actividad metiltransferasa, mostraron asociaciones en al menos dos las tres estrategias de filtrado de variantes utilizadas para los análisis. Otro gen que resultó asociado en nuestra fase de descubrimiento y cuya función podría ser interesante para nuestro fenotipo es *BMPER*. Este gen es activador y protector de células endoteliales. Regula procesos biológicos como la angiogénesis²³¹, la adhesión y la migración de leucocitos²³². Se ha descrito su papel como regulador y protector de la inflamación vascular²³². Disminuciones en la expresión de este gen podrían estar relacionadas con disfunción endotelial y vasculitis²³², una enfermedad que puede predisponer al debilitamiento de los vasos y sangrados. Además, se encontró asociado significativamente al gen *ZNF385D*. Una de las funciones destacadas de este gen es la capacidad de unión con el ADN. De acuerdo con la base *Proteomics DB*²³³, la proteína codificada por este gen se expresa en células del epitelio del colon, plaquetas y retina, entre otras²³³. La asociación de este gen con efectos adversos a la radioterapia ha sido descrita previamente por Kerns et al. (2020)¹²⁵, en pacientes con cáncer de próstata. En este estudio, los investigadores encuentran a la variante rs73046248 en *ZNF385D*, asociada con la disminución del flujo urinario (HR = 2,40; 95% CI [1,71 - 3,36]; p-meta = 4.2×10^{-07}). Estos hallazgos apoyarían la hipótesis de que la sensibilidad a la radioterapia podría estar relacionada por variantes que afectan de forma generalizada el desarrollo de efectos adversos.

En la fase de validación los 32 genes candidatos del estudio de fenotipos de sangrado rectal extremo, que incluyó a ocho pacientes con sangrado rectal grado ≥ 3 (CTCAE v.4.0) y 322 pacientes que no presentaron ninguna toxicidad (grado = 0) (CTCAE v.4.0), el gen *ADRM1* fue asociado nominalmente con sangrado rectal.

El gen *ADRM1* desempeña una función importante en el mantenimiento de la homeostasis de las proteínas a través de la proteólisis^{234,235}. Este gen participa en procesos como la progresión del ciclo celular, la apoptosis y la reparación de daños en el ADN. Se ha observado que niveles elevados de *adrm1* se asocian de forma indirecta

con mayor adhesión celular²³⁶. Asimismo, la sobreexpresión de este gen ha sido relacionada con mayor progresión y agresividad en diferentes tipos de cáncer^{237–239}.

Por otra parte, a pesar de que solo el gen *ADRM1* validó los resultados de nuestros análisis de la fase de descubrimiento, hay que destacar que en algunos casos la falta de validación de los genes no fue debida a la ausencia de asociación, sino al bajo o nulo número de variantes analizadas por gen en la fase de validación.

La sensibilidad a la radioterapia se considera un rasgo complejo en el que contribuyen variantes comunes y raras. A pesar de ello, la mayoría de los análisis se enfocan en el estudio de solo un tipo de variantes; analizan solo variantes comunes o solo variantes raras. Los estudios que analizan ambos tipos de variantes conjuntamente son limitados. Debido a que nuestro objetivo es identificar genes que contienen variantes de riesgo ya sean comunes o raras asociadas con radiotoxicidad, realizamos un análisis que combina dentro de un mismo gen el efecto acumulado de variantes comunes ($MAF > 0,05$) y de variantes raras (definidas para este análisis variantes con $MAF \leq 0,05$), conocido como “*Combined test of common rare variants*¹⁴³” (apartado 3.8.2). En total 1 143 365 variantes fueron analizadas resultando de ello 21 629 genes analizados.

Con este estudio identificamos 27 genes asociados nominalmente con sangrado rectal ($p\text{-valor} < 9,7 \times 10^{-04}$). De ellos los principales genes asociados fueron *MBP*, *LRP1* y *PGPEP1L* con $p\text{-valores}$ menores a $8,6 \times 10^{-05}$. El gen *MBP* codifica para una proteína que interactúa con proteínas polianiónicas como la actina y la tubulina, así como el Ca^{+2} ²⁴⁰. Se ha descrito que el gen *MBP* podría tener un papel como marcador de lesiones cerebrales incluida la hemorragia subaracnoidea²⁴¹. Un estudio llevado a cabo en ratas, sometidas a radiación cerebral, reportó disminución en los niveles de *MBP* posteriores a la radiación y la ha relacionado con una posible desmielinización en el tejido cerebral²⁴². El segundo gen asociado es *LRP1*, se ha reportado que este gen receptor de lipoproteínas de baja densidad, desempeña un papel esencial en la protección de la pared vascular y en la prevención de la arterosclerosis por medio del control de la activación del factor de crecimiento derivado de plaquetas (PDGFR)²⁴³. También, ha sido documentado que la disminución de *LRP1* conduce al daño progresivo vascular y a la formación de aneurismas²⁴⁴. El tercer gen de interés es *PGPEP1L*, este gen se expresa mayoritariamente en músculo esquelético²⁴⁵ y, entre sus funciones más importantes se encuentra su participación en la proteólisis²⁴⁶.

Además, identificamos que cinco de los 27 genes asociados en este estudio también presentaron asociaciones nominales en los análisis realizados con SKAT-O en los que se consideró a todas las variantes de baja frecuencia presentes en las regiones secuenciadas. Estos genes son *MBP*, *CACNA1C-AS2*, *LRRC70* y *RPB2* asociados en los análisis de SKAT-O para el fenotipo de sangrado rectal (apartado 4.2.1) y el gen *CACNA1C-AS1* que se encontró asociado en los análisis de SKAT-O tanto para el

fenotipo de sangrado rectal como para el fenotipo de sangrado rectal extremo (apartados 4.2.1 y 4.2.2).

Estos hallazgos apoyan la posible asociación entre los genes identificados y la sensibilidad a la radiación ionizante, derivado de un efecto combinado de variantes comunes y raras. Además, el uso de esta estrategia de análisis mejora el poder de detección de genes causales. Para nuestro estudio hemos utilizado el test SKAT-C para evitar la pérdida de asociación que se presenta al usar otros métodos que sobreestiman la contribución de variantes raras y disminuyen la contribución de variantes comunes.

En el **Capítulo 3**, presentamos resultados de las nuevas aproximaciones “ómicas” de estudios de asociación basados en genes. Realizamos el estudio de asociación a nivel del transcriptoma (TWAS) y el análisis a nivel del proteoma (PWAS). Para ambas aproximaciones se incluyeron sólo variantes frecuentes ($MAF > 0,05$).

El análisis de TWAS se llevó a cabo utilizando dos estrategias diferentes. En la primera se analizaron siete tejidos: seis adyacentes a la próstata y la sangre total (apartado 3.9.1). En la segunda se analizaron 46 tejidos, excluyendo del análisis a los tejidos específicos de órganos femeninos (ANEXO 8).

En el análisis de tejidos cercanos a la próstata, encontramos tres genes asociados significativamente después de la corrección por Bonferroni. El primer gen asociado es *AXL* receptor de tirosín quinasa. Este gen está involucrado en funciones como el crecimiento y la migración celular, así como la angiogénesis. El aumento de la expresión de *AXL* se ha asociado con la progresión de numerosos tipos de cáncer^{247,248}. Por otro lado, se observó en un modelo *in vivo* que la eliminación de la expresión de este gen redujo el crecimiento de células cancerígenas²⁴⁹ y aumentó la apoptosis inducida por radiación²⁵⁰. El segundo gen es *ITSN1*, resultados de un estudio con modelos *in vitro* de cáncer de mama identificaron que la inhibición de *ITSN1* podría favorecer la proliferación celular e inhibir el proceso de apoptosis, mientras que la sobreexpresión de este gen podría estar relacionada con la inhibición de la proliferación celular y el aumento de la apoptosis²⁵¹. En base a esto, los autores sugieren a *ITSN1* como un objetivo clave para el control del cáncer. El tercer gen es *WWC1* (también llamado *KIBRA*), la desregulación de la proteína kibra y de la proteína yorkie está relacionada con defectos en el mantenimiento y la regeneración muscular. Las interacciones de ambas proteínas con el receptor distroglicano podrían ajustar la actividad de la vía de señalización Hippo en los músculos, afectando el mantenimiento y la regeneración muscular relacionados con la edad. A su vez, la vía de señalización de Hippo participa en la regulación de la miogénesis, el crecimiento y la supervivencia de los miocitos y la regeneración de los tejidos²⁵².

Por otra parte, identificamos que los genes *AXL* y *LTBP4* se encontraban localizados en el mismo locus a una distancia menor de 1 Mb. La colocación de resultados TWAS

y GWAS indica una posible causalidad de las variantes localizadas. En ese sentido, sería interesante en un futuro llevar a cabo un estudio de colocalización para profundizar en estas asociaciones.

En la segunda estrategia del estudio TWAS se identificaron a los genes *CPN1* y *MLEC* asociados con el fenotipo a estudio. El gen *CPN1* regula la actividad biológica de péptidos vasoactivos e inflamatorios que contienen Arg o Lys C-terminal como las cininas y las anafilatoxinas^{253,254}. Así mismo, se ha relacionado la participación de este gen en la conversión de fibrinógeno a fibrina²⁵⁵. El segundo gen identificado es *MLEC*. Este gen codifica para la proteína de unión a carbohidratos denominada malectina. Interacciones entre esta proteína y la riboforina han sido relacionadas con la probable degradación de proteínas mal plegadas. Un ensayo *in vitro* reveló la importancia del papel de la malectina en la regulación y el transporte del factor VIII²⁵⁶.

Nuestros hallazgos relacionan a diferentes genes de susceptibilidad implicados en procesos de hemostasia, de regeneración del musculo, de la angiogénesis, así como genes relacionados con la proliferación celular y la progresión tumoral. Además, los resultados obtenidos con el análisis multi-tejido, revelan que en ambas estrategias del estudio de TWAS, los genes candidatos identificados se encuentran expresados mayoritariamente en la sangre total y en testículos. Aunque este resultado se tiene que tomar con precaución ya que puede existir un poco de “*overfitting*” debido al tamaño muestral y de genes expresados en estos dos tejidos.

El análisis TWAS realizado ha permitido aumentar el poder de identificación de genes que pueden contribuir con la susceptibilidad al desarrollo de sangrado rectal postradioterapia, lo que se presenta como una ventaja en comparación con los análisis GWAS que analizan variantes individuales. Sin embargo, hay que considerar que los genes identificados en el análisis de TWAS no son necesariamente genes causales, sino que su expresión está genéticamente regulada y pueden estar influyendo en la susceptibilidad a la toxicidad.

El análisis a nivel del proteoma se llevó a cabo usando el modelo de predicción de proteínas plasmáticas de Zhang^{162,257}. Como resultado del PWAS, 1105 genes fueron analizados, entre los genes con asociaciones nominales más significativas destaca *THBS2*. Además de este gen, encontramos genes relacionados con procesos de angiogénesis, actividad neuronal, inflamación, adhesión y la migración celular.

Al comparar los resultados obtenidos por los estudios TWAS y PWAS, observamos que ocho genes: *THBS2*, *PROC*, *SORD*, *PLA2G2A*, *GCLM*, *B3GAT*, *POR* y *PLCG2* presentaron asociaciones nominales menores a 0,05 en ambos estudios. Ninguna de las asociaciones se mantuvo significativa después de la corrección múltiple (Bonferroni para TWAS y FDR para PWAS). Observamos que el gen *THBS2*, también se encontró asociado nominalmente en el estudio de asociación basado en genes, para el fenotipo

sangrado rectal extremo utilizando SKAT-O (se incluyeron únicamente variantes con frecuencia en población europea menor a 0,05; mientras que en el estudio de PWAS se analizaron variantes con MAF mayor a 0,05). En ambos estudios el gen *THBS2* mostró una dirección de asociación consistente con nuestro fenotipo. Sin embargo, ninguna de las asociaciones se mantuvo después de las respectivas correcciones por test múltiple.

El gen *THBS2*, codifica para una proteína inhibidora del crecimiento tumoral y la angiogénesis. En un modelo murino, se observó que el silenciamiento de *THBS2* se asocia con la desorganización de las fibras de colágeno en la piel y con la formación de fibrillas de mayor tamaño y con contornos irregulares²⁵⁸. Estos cambios dieron lugar a una piel frágil y con mayor tendencia a desgarrarse. El silenciamiento de *THBS2* también conlleva a una mayor predisposición a la diátesis hemorrágica, la cual podría estar asociada a una adhesión defectuosa de las plaquetas al subendotelio lesionado. La participación de diferentes tipos de colágeno (I, III, IV y VI) y la fibronectina en la mediación de la adhesión plaquetaria establece un vínculo entre las alteraciones en la fibrillogénesis del colágeno, derivadas del silenciamiento de *THBS2*, y las alteraciones en la adhesión plaquetaria²⁵⁸. El hallazgo de este gen es de gran interés debido a que las alteraciones observadas en el estudio de referencia nos sugieren que este gen podría aumentar la susceptibilidad al desarrollo de efectos adversos posteriores al tratamiento radioterápico, como el sangrado rectal.

En conjunto nuestros hallazgos evidencian la utilidad de los estudios ómicos para la identificación de nuevos genes asociados con sangrado rectal como efecto adverso a la radioterapia. Estos análisis ayudarán a tener una mejor comprensión de los mecanismos moleculares subyacentes a la susceptibilidad a la radiación ionizante. La aplicación de estas herramientas ómicas al área de radiogenómica potencia el descubrimiento de nuevos genes asociados con la susceptibilidad a la radiación ionizante.

Finalmente, exploramos las principales rutas en las que se encuentran implicados los 96 genes, seleccionados de nuestros resultados de SKAT-O, SKAT-C, TWAS y PWAS. Para ello, se llevaron a cabo dos estrategias de análisis. En la primera estrategia, se exploró la presencia de los genes en las rutas de interés relacionadas con radiotoxicidad (apartado 3.9.3). Aquí, cuatro genes fueron identificados en la ruta inmunológica (*TNFRSF10C*, *PSMB8*, *MBP* y *CXCL1*), uno de ellos también se encontró en la ruta de la inflamación (*CXCL1*) y por último un gen (*PRPF19*) se identificó en la vía de la reparación del ADN, específicamente en la ruta de reparación por escisión de nucleótidos^{259,260}. Estos resultados apoyarían la validez de las aproximaciones ómicas utilizadas en nuestro estudio.

En la segunda estrategia, utilizando la herramienta GENE2FUNC de FUMA, encontramos tres conjuntos de genes de la categoría de expresión de alteraciones genéticas y químicas asociadas significativamente, destacando entre ellos “*Warters response to IR skin*”²⁶¹. Esta ruta considera un conjunto de genes que muestra una

expresión alterada en respuesta a la radiación ionizante en células de piel humana. El estudio que involucró a este conjunto de genes examinó la respuesta transcripcional de fibroblastos y queratinocitos primarios inactivos expuestos a radiación ionizante en dosis que oscilaron entre 0,1 Gy y 5 Gy. Los resultados revelaron que la exposición a la radiación ionizante alteró la expresión de 279 genes en ambos tipos de células, entre ellos algunos genes regulados por p53²⁶². Asimismo, encontramos señales de enriquecimiento en conjuntos de genes para la ruta de Reactome *TP53 regulates transcription of death receptors and ligands* y para rutas implicadas en la fase de mitosis del ciclo celular (*Reactome cell cycle mitotic* y *GO regulation of mitotic cell cycle*). Sin embargo, estas rutas no permanecieron significativas al corregir por BH-FDR.

Durante el desarrollo de esta tesis, realizamos estrategias de análisis diferentes pero complementarias entre sí, que han permitido una mejor caracterización de nuestros resultados. Ello nos ha llevado a la identificación de nuevos genes candidatos a la susceptibilidad a la radiación ionizante, específicamente asociados al sangrado rectal postradioterapia. Nuestro estudio es el primero en el área de radiogenómica que incluye la secuenciación del exoma completo, la implementación de diferentes aproximaciones de análisis basadas en genes, y nuevas aproximaciones ómicas, como son los estudios TWAS y PWAS. Estas innovadoras estrategias podrán utilizarse tanto para reanalizar datos genómicos de estudios previos, así como en futuros estudios de radiogenómica.

6 PERSPECTIVAS FUTURAS

De la experiencia obtenida en este estudio, nos atrevemos a plantear una serie de iniciativas que contribuirían al enriquecimiento de nuestros resultados y de futuros análisis:

1. Realizar la búsqueda de nuevas cohortes, por ejemplo, del RGC, con el fin de metaanalizar (MetaSKAT) los resultados de nuestros estudios basados en genes. Esto nos permitirá aumentar la validez de nuestros hallazgos.
2. Llevar a cabo la validación de nuestros resultados de TWAS en una cohorte independiente, así como realizar los respectivos análisis de colocación.
3. Buscar e incorporar nuevos modelos de predicción de proteínas disponibles para incorporar en nuestros análisis PWAS y con ello aumentar el poder de detección de asociaciones en este estudio.
4. Realizar estudios funcionales de los mejores genes candidatos, que contribuirán a una mejor comprensión de la implicación de estos genes en el desarrollo de efectos adversos asociados a la radiación ionizante.

7 CONCLUSIONES

PRIMERA

Llevamos a cabo una revisión bibliográfica exhaustiva de variantes y genes asociados con la toxicidad inducida por la radioterapia en pacientes con cáncer, independientemente de la ubicación tumoral y el tipo de toxicidad. Analizamos los genes y las variantes identificadas en nuestra cohorte RADIOGEN-Próstata para evaluar su relación con el fenotipo de sangrado rectal. Así, identificamos al SNP rs144214859, previamente relacionado con la hematuria inducida por radiación en pacientes con cáncer de próstata, asociado nominalmente con el sangrado rectal en nuestro análisis univariable. No obstante, no hemos logrado establecer una asociación estadísticamente significativa entre los hallazgos previamente descritos (variantes y genes) y el sangrado rectal en nuestra cohorte.

SEGUNDA

Realizamos un estudio basado en genes en la cohorte RADIOGEN-Próstata a partir de tres estrategias diferentes de filtrado de variantes. Identificamos 32 genes asociados nominalmente con el fenotipo sangrado rectal (grado CTCAE ≥ 2), cinco de estos genes (*LRRC70*, *MYBPC2*, *NUP188*, *PSMB8* y *ZCCHC2*) se encontraron asociados con al menos dos de las estrategias de filtrado utilizadas. De ellos resaltamos la asociación observada para el gen *LRP1B*, previamente vinculado a la toxicidad inducida por radioterapia en pacientes con cáncer de próstata. Además, validamos nominalmente la asociación de los genes *PRB1* y *RANBP10* en la cohorte internacional multicéntrica REQUITE.

Asimismo, identificamos 32 genes asociados nominalmente con el fenotipo sangrado rectal extremo (grado CTCAE ≥ 3), de los cuales dos (*ADRM1* y *METTL25*) se encontraron asociados en dos de las estrategias de filtrado utilizadas. Destacamos la asociación del gen *ZNF385D*, previamente relacionado con la reducción del flujo urinario como efecto adverso de la radioterapia en pacientes con cáncer de próstata. Además, validamos nominalmente en la cohorte REQUITE la asociación del gen *ADRM1*.

TERCERA

El análisis de asociación del efecto combinado de variantes comunes y raras identificó 27 genes asociados nominalmente con sangrado rectal, siendo los principales genes asociados *MBP*, *LRP1* y *PGPEP1L*. Además, cinco de los genes identificados en este análisis también se encontraron asociados nominalmente en los análisis de SKAT-O: *MBP*, *CACNA1C-AS1*, *CACNA1C-AS2*, *LRRC70* y *RPB2* se asociaron a sangrado rectal y el gen *CACNA1C-AS1* se asoció además, con sangrado rectal extremo.

CUARTA

Por primera vez en Radiogenómica, realizamos pioneros análisis de TWAS y de PWAS. En el TWAS identificamos cinco genes candidatos (*AXL*, *ITSN1*, *WWC1*, *CPN1* y *MLEC*). La regulación génica de la expresión de estos genes influye con la susceptibilidad a sangrado rectal. En el PWAS encontramos siete genes asociados nominalmente con el sangrado rectal, destacamos el gen *THBS2*. Este gen también se asoció nominalmente en nuestro análisis de fenotipos extremos usando SKAT-O para variantes exónicas y de *splicing* ± 2 pb y ha sido previamente vinculado con el riesgo de hemorragias.

QUINTA

En el análisis de enriquecimiento del conjunto de genes realizado con la herramienta FUMA, identificamos tres conjuntos de genes con asociaciones estadísticamente significativas en la categoría Perturbaciones químicas y genéticas. De ellos destaca el conjunto de genes “*Warters response to IR skin*”, que incluye genes (*FGFBP3*, *TRIM22*, *ALCAM* y *TNFRSF10C*) que muestran una expresión alterada a la radiación ionizante en células de piel humana.

SEXTA

En este estudio realizamos un abordaje radiogenómico profundo en relación con el sangrado rectal como efecto adverso a la radioterapia en pacientes con cáncer de próstata. Para ello, utilizamos herramientas de vanguardia, como son el análisis de exomas, el análisis de genotipos, la evaluación conjunta de múltiples variantes genéticas, el estudio de asociación del transcriptoma y del proteoma, además del análisis enriquecido de rutas genéticas, firmas moleculares y ontología de genes. Estos análisis exploraron variantes genéticas tanto comunes como raras, para obtener una visión detallada y completa. Además, hicimos un análisis de validación en una cohorte multicéntrica internacional para contrastar los resultados obtenidos. Esto nos ha llevado a la identificación de nuevos genes candidatos que podrán ser investigados con mayor profundidad para su posible aplicación en la práctica clínica.

8 BIBLIOGRAFÍA

1. Rebello, R. J. et al. Prostate cancer. *Nat. Rev. Dis. Primer* 7, 1–27 (2021).
2. Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* 71, 209–249 (2021).
3. Cancer (IARC), T. I. A. for R. on. Global Cancer Observatory. <https://gco.iarc.fr/>.
4. Estimación de la incidencia de cáncer en España. (2022).
5. Stewart, R. W., Lizama, S., Peairs, K., Sateia, H. F. & Choi, Y. Screening for prostate cancer. *Semin. Oncol.* 44, 47–56 (2017).
6. Mottet, N. et al. EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer-2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur. Urol.* 79, 243–262 (2021).
7. Kimura, T. East meets West: ethnic differences in prostate cancer epidemiology between East Asians and Caucasians. *Chin. J. Cancer* 31, 421–429 (2012).
8. Hemminki, K. Familial risk and familial survival in prostate cancer. *World J. Urol.* 30, 143–148 (2012).
9. Pritchard, C. C. et al. Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *N. Engl. J. Med.* 375, 443–453 (2016).
10. Darst, B. F. et al. Germline Sequencing Analysis to Inform Clinical Gene Panel Testing for Aggressive Prostate Cancer. *JAMA Oncol.* (2023) doi:10.1001/jamaoncol.2023.3482.
11. Hansen, E. B. et al. Impact of germline DNA repair gene variants on prognosis and treatment of men with advanced prostate cancer. *Sci. Rep.* 13, 19135 (2023).
12. Saunders, E. J. et al. Fine-Mapping the HOXB Region Detects Common Variants Tagging a Rare Coding Allele: Evidence for Synthetic Association in Prostate Cancer. *PLOS Genet.* 10, e1004129 (2014).
13. Karlsson, R. et al. A population-based assessment of germline HOXB13 G84E mutation and prostate cancer risk. *Eur. Urol.* 65, 169–176 (2014).
14. Ewing, C. M. et al. Germline mutations in HOXB13 and prostate-cancer risk. *N. Engl. J. Med.* 366, 141–149 (2012).
15. Brandão, A. et al. The CHEK2 Variant C.349A>G Is Associated with Prostate Cancer Risk and Carriers Share a Common Ancestor. *Cancers* 12, 3254 (2020).

16. Conti, D. V. et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat. Genet.* 53, 65–75 (2021).
17. Oczkowski, M., Dziendzikowska, K., Pasternak-Winiarska, A., Włodarek, D. & Gromadzka-Ostrowska, J. Dietary Factors and Prostate Cancer Development, Progression, and Reduction. *Nutrients* 13, 496 (2021).
18. Fujita, K., Hayashi, T., Matsushita, M., Uemura, M. & Nonomura, N. Obesity, Inflammation, and Prostate Cancer. *J. Clin. Med.* 8, 201 (2019).
19. Calle, E. E., Rodriguez, C., Walker-Thurmond, K. & Thun, M. J. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N. Engl. J. Med.* 348, 1625–1638 (2003).
20. Freedland, S. J. & Aronson, W. J. Examining the Relationship Between Obesity and Prostate Cancer. *Rev. Urol.* 6, 73–81 (2004).
21. Lu, Y. et al. Insufficient Lycopene Intake Is Associated With High Risk of Prostate Cancer: A Cross-Sectional Study From the National Health and Nutrition Examination Survey (2003-2010). *Front. Public Health* 9, 792572 (2021).
22. Lee, S. H. & Shen, M. M. Cell types of origin for prostate cancer. *Curr. Opin. Cell Biol.* 37, 35–41 (2015).
23. Garber, K. A Tale of Two Cells: Discovering the Origin of Prostate Cancer. *JNCI J. Natl. Cancer Inst.* 102, 1528–1535 (2010).
24. Goldstein, A. S., Huang, J., Guo, C., Garraway, I. P. & Witte, O. N. Identification of a cell-of-origin for human prostate cancer. *Science* 329, 568–571 (2010).
25. Molecular Pathways in Prostate Cancer | Nephro-Urology Monthly | Full Text. <https://brieflands.com/articles/num-55492.html>.
26. Rubin, M. A. & Demichelis, F. The Genomics of Prostate Cancer: emerging understanding with technologic advances. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc* 31, S1-11 (2018).
27. He, W. W. et al. A novel human prostate-specific, androgen-regulated homeobox gene (NKX3.1) that maps to 8p21, a region frequently deleted in prostate cancer. *Genomics* 43, 69–77 (1997).
28. Hubbard, G. K. et al. Combined MYC activation and Pten loss are sufficient to create genomic instability and lethal metastatic prostate cancer. *Cancer Res.* 76, 283–292 (2016).

29. Kim, J., Eltoum, I.-E. A., Roh, M., Wang, J. & Abdulkadir, S. A. Interactions between Cells with Distinct Mutations in c-MYC and Pten in Prostate Cancer. *PLOS Genet.* 5, e1000542 (2009).
30. Grisanzio, C. & Freedman, M. L. Chromosome 8q24-Associated Cancers and MYC. *Genes Cancer* 1, 555–559 (2010).
31. Häggman, M. J., Macoska, J. A., Wojno, K. J. & Oesterling, J. E. The relationship between prostatic intraepithelial neoplasia and prostate cancer: critical issues. *J. Urol.* 158, 12–22 (1997).
32. Saranyutanon, S., Srivastava, S. K., Pai, S., Singh, S. & Singh, A. P. Therapies Targeted to Androgen Receptor Signaling Axis in Prostate Cancer: Progress, Challenges, and Hope. *Cancers* 12, 51 (2019).
33. Saranyutanon, S. et al. Cellular and Molecular Progression of Prostate Cancer: Models for Basic and Preclinical Research. *Cancers* 12, 2651 (2020).
34. Warren, A. Y. & Harrison, D. WHO/ISUP classification, grading and pathological staging of renal cell carcinoma: standards and controversies. *World J. Urol.* 36, 1913–1926 (2018).
35. Egevad, L., Delahunt, B., Srigley, J. R. & Samaratunga, H. International Society of Urological Pathology (ISUP) grading of prostate cancer - An ISUP consensus on contemporary grading. *APMIS Acta Pathol. Microbiol. Immunol. Scand.* 124, 433–435 (2016).
36. Kim, C.-H. et al. Artificial Intelligence Techniques for Prostate Cancer Detection through Dual-Channel Tissue Feature Engineering. *Cancers* 13, 1524 (2021).
37. Humphrey, P. A. Gleason grading and prognostic factors in carcinoma of the prostate. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc* 17, 292–306 (2004).
38. Amin, M. B. et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more ‘personalized’ approach to cancer staging. *CA. Cancer J. Clin.* 67, 93–99 (2017).
39. D’Amico, A. V. et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* 280, 969–974 (1998).
40. Parker, C., Gillessen, S., Heidenreich, A., Horwich, A., & ESMO Guidelines Committee. Cancer of the prostate: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* 26 Suppl 5, v69-77 (2015).

41. Kattan, M. W., Wheeler, T. M. & Scardino, P. T. Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 17, 1499–1507 (1999).
42. Kattan, M. W., Eastham, J. A., Stapleton, A. M. F., Wheeler, T. M. & Scardino, P. T. A Preoperative Nomogram for Disease Recurrence Following Radical Prostatectomy for Prostate Cancer. *JNCI J. Natl. Cancer Inst.* 90, 766–771 (1998).
43. Parker, C. et al. Prostate cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* 31, 1119–1134 (2020).
44. De Ruyscher, D. et al. Radiotherapy toxicity. *Nat. Rev. Dis. Primer* 5, 13 (2019).
45. Mangoni, M. et al. Hypofractionation in Prostate Cancer: Radiobiological Basis and Clinical Appliance. *BioMed Res. Int.* 2014, 781340 (2014).
46. Yan, M. et al. Practical considerations for prostate hypofractionation in the developing world. *Nat. Rev. Urol.* 18, 669–685 (2021).
47. Morgan, S. C. et al. Hypofractionated Radiation Therapy for Localized Prostate Cancer: Executive Summary of an ASTRO, ASCO and AUA Evidence-Based Guideline. *J. Urol.* 201, 528–534 (2019).
48. Brand, D. H. et al. Intensity-modulated fractionated radiotherapy versus stereotactic body radiotherapy for prostate cancer (PACE-B): acute toxicity findings from an international, randomised, open-label, phase 3, non-inferiority trial. *Lancet Oncol.* 20, 1531–1543 (2019).
49. Wang, H., Mu, X., He, H. & Zhang, X.-D. Cancer Radiosensitizers. *Trends Pharmacol. Sci.* 39, 24–48 (2018).
50. Sia, J., Szmyd, R., Hau, E. & Gee, H. E. Molecular Mechanisms of Radiation-Induced Cancer Cell Death: A Primer. *Front. Cell Dev. Biol.* 8, 41 (2020).
51. Campbell, A. M. & Decker, R. H. Harnessing the Immunomodulatory Effects of Radiation Therapy. *Oncol. Williston Park N* 32, 370–374, CV3 (2018).
52. Kepp, O. et al. Consensus guidelines for the detection of immunogenic cell death. *Oncoimmunology* 3, e955691 (2014).
53. Gómez, V., Mustapha, R., Ng, K. & Ng, T. Radiation therapy and the innate immune response: Clinical implications for immunotherapy approaches. *Br. J. Clin. Pharmacol.* 86, 1726–1735 (2020).

54. Durante, M., Reppingen, N. & Held, K. D. Immunologically augmented cancer treatment using modern radiotherapy. *Trends Mol. Med.* 19, 565–582 (2013).
55. de Andrade Carvalho, H. & Villar, R. C. Radiotherapy and immune response: the systemic effects of a local treatment. *Clinics* 73, e557s (2018).
56. Sinclair, W. K. Cyclic X-ray responses in mammalian cells in vitro. 1968. *Radiat. Res.* 178, AV112-124 (2012).
57. Syljuåsen, R. G. Cell Cycle Effects in Radiation Oncology. in *Radiation Oncology* (ed. Wenz, F.) 1–8 (Springer International Publishing, 2019). doi:10.1007/978-3-319-52619-5_101-1.
58. Philippou, Y., Sjöberg, H., Lamb, A. D., Camilleri, P. & Bryant, R. J. Harnessing the potential of multimodal radiotherapy in prostate cancer. *Nat. Rev. Urol.* 17, 321–338 (2020).
59. Huang, R.-X. & Zhou, P.-K. DNA damage response signaling pathways and targets for radiotherapy sensitization in cancer. *Signal Transduct. Target. Ther.* 5, 60 (2020).
60. Barnett, G. C. et al. Normal tissue reactions to radiotherapy: towards tailoring treatment dose by genotype. *Nat. Rev. Cancer* 9, 134–142 (2009).
61. Kerns, S. L. et al. Meta-analysis of Genome Wide Association Studies Identifies Genetic Markers of Late Toxicity Following Radiotherapy for Prostate Cancer. *EBioMedicine* 10, 150–163 (2016).
62. Michalski, J. M. et al. Preliminary Toxicity Analysis of 3DCRT versus IMRT on the High Dose Arm of the RTOG 0126 Prostate Cancer Trial. *Int. J. Radiat. Oncol. Biol. Phys.* 87, 10.1016/j.ijrobp.2013.07.041 (2013).
63. Wielen, G. J. van der, Putten, W. L. J. van & Incrocci, L. Sexual Function After Three-Dimensional Conformal Radiotherapy for Prostate Cancer: Results From a Dose-Escalation Trial. *Int. J. Radiat. Oncol. Biol. Phys.* 68, 479–484 (2007).
64. Linares Mesa, N. A. & NPunto. INDICACIONES Y TOXICIDAD DE LA RADIOTERAPIA EN CÁNCER DE PRÓSTATA. CONTROVERSIA FRENTE A CIRUGÍA. INDICACIONES TOXICIDAD Radioter. EN CÁNCER PRÓSTATA Controv. FRENTE Cir. 0, 1–0 (2021).
65. Amanat, S., Requena, T. & Lopez-Escamez, J. A. A Systematic Review of Extreme Phenotype Strategies to Search for Rare Variants in Genetic Studies of Complex Disorders. *Genes* 11, 987 (2020).

66. Majeed, H. & Gupta, V. Adverse Effects of Radiation Therapy. in StatPearls (StatPearls Publishing, 2023).
67. Zelefsky, M. J. et al. Incidence of Late Rectal and Urinary Toxicities After Three-Dimensional Conformal Radiotherapy and Intensity-Modulated Radiotherapy for Localized Prostate Cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 70, 1124–1129 (2008).
68. Takemoto, S. et al. Treatment and prognosis of patients with late rectal bleeding after intensity-modulated radiation therapy for prostate cancer. *Radiat. Oncol.* 7, 87 (2012).
69. Wd, H., St, P., Pc, K., Ms, H. & Jv, L. Acute and late gastrointestinal toxicity after radiotherapy in prostate cancer patients: consequential late damage. *Int. J. Radiat. Oncol. Biol. Phys.* 66, (2006).
70. Zelefsky, M. J. et al. Long term tolerance of high dose three-dimensional conformal radiotherapy in patients with localized prostate carcinoma. *Cancer* 85, 2460–2468 (1999).
71. Palmer, M. K. WHO Handbook for Reporting Results of Cancer Treatment. *Br. J. Cancer* 45, 484–485 (1982).
72. Cox, J. D., Stetz, J. & Pajak, T. F. Toxicity criteria of the Radiation Therapy Oncology Group (RTOG) and the European Organization for Research and Treatment of Cancer (EORTC). *Int. J. Radiat. Oncol. Biol. Phys.* 31, 1341–1346 (1995).
73. LENT SOMA scales for all anatomic sites. *Int. J. Radiat. Oncol. Biol. Phys.* 31, 1049–1091 (1995).
74. National Institutes of Health, National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE) version 5.0. https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm#ctc_40 (2017).
75. Turesson, I., Nyman, J., Holmberg, E. & Odén, A. Prognostic factors for acute and late skin reactions in radiotherapy patients. *Int. J. Radiat. Oncol. Biol. Phys.* 36, 1065–1075 (1996).
76. Taylor, A. M. et al. Ataxia telangiectasia: a human mutation with abnormal radiation sensitivity. *Nature* 258, 427–429 (1975).

77. Thariat, J. et al. Avoidance or adaptation of radiotherapy in patients with cancer with Li-Fraumeni and heritable TP53-related cancer syndromes. *Lancet Oncol.* 22, e562–e574 (2021).
78. West, C. M. & Barnett, G. C. Genetics and genomics of radiotherapy toxicity: towards prediction. *Genome Med.* 3, 52 (2011).
79. Pritchard, J. K. Are Rare Variants Responsible for Susceptibility to Complex Diseases? *Am. J. Hum. Genet.* 69, 124–137 (2001).
80. Andreassen, C. N., Alsner, J. & Overgaard, J. Does variability in normal tissue reactions after radiotherapy have a genetic basis--where and how to look for it? *Radiother. Oncol. J. Eur. Soc. Ther. Radiol. Oncol.* 64, 131–140 (2002).
81. Talbot, C. J. et al. A replicated association between polymorphisms near TNF α and risk for adverse reactions to radiotherapy. *Br. J. Cancer* 107, 748–753 (2012).
82. Barnett, G. C. et al. Independent validation of genes and polymorphisms reported to be associated with radiation toxicity: a prospective analysis study. *Lancet Oncol.* 13, 65–77 (2012).
83. Barnett, G. C. et al. Individual patient data meta-analysis shows no association between the SNP rs1800469 in TGFB and late radiotherapy toxicity. *Radiother. Oncol. J. Eur. Soc. Ther. Radiol. Oncol.* 105, 289–295 (2012).
84. Radiogenomics. Wikipedia (2023).
85. Yard, B. D. et al. A genetic basis for the variation in the vulnerability of cancer to DNA damage. *Nat. Commun.* 7, 11428 (2016).
86. Das, A. K., Bell, M. H., Nirodi, C. S., Story, M. D. & Minna, J. D. Radiogenomics predicting tumor responses to radiotherapy in lung cancer. *Semin. Radiat. Oncol.* 20, 149–155 (2010).
87. West, C. et al. Establishment of a Radiogenomics Consortium. *Int. J. Radiat. Oncol. Biol. Phys.* 76, 1295–1296 (2010).
88. Liu, Z. et al. Imaging genomics for accurate diagnosis and treatment of tumors: A cutting edge overview. *Biomed. Pharmacother. Biomedecine Pharmacother.* 135, 111173 (2021).
89. Rutman, A. M. & Kuo, M. D. Radiogenomics: Creating a link between molecular diagnostics and diagnostic imaging. *Eur. J. Radiol.* 70, 232–241 (2009).

90. Shui, L. et al. The Era of Radiogenomics in Precision Medicine: An Emerging Approach to Support Diagnosis, Treatment Decisions, and Prognostication in Oncology. *Front. Oncol.* 10, 570465 (2020).
91. Andreassen, C. N., Schack, L. M. H., Laursen, L. V. & Alsner, J. Radiogenomics - current status, challenges and future directions. *Cancer Lett.* 382, 127–136 (2016).
92. Andreassen, C. N. & Alsner, J. Genetic variants and normal tissue toxicity after radiotherapy: a systematic review. *Radiother. Oncol. J. Eur. Soc. Ther. Radiol. Oncol.* 92, 299–309 (2009).
93. Brothwell, M. R. S., West, C. M., Dunning, A. M., Burnet, N. G. & Barnett, G. C. Radiogenomics in the Era of Advanced Radiotherapy. *Clin. Oncol. R. Coll. Radiol. G. B.* 31, 319–325 (2019).
94. Alsner, J., Andreassen, C. N. & Overgaard, J. Genetic markers for prediction of normal tissue toxicity after radiotherapy. *Semin. Radiat. Oncol.* 18, 126–135 (2008).
95. Damaraju, S. et al. Association of DNA Repair and Steroid Metabolism Gene Polymorphisms with Clinical Late Toxicity in Patients Treated with Conformal Radiotherapy for Prostate Cancer. *Clin. Cancer Res.* 12, 2545–2554 (2006).
96. Burri, R. J. et al. Association of single nucleotide polymorphisms in SOD2, XRCC1 and XRCC3 with susceptibility for the development of adverse effects resulting from radiotherapy for prostate cancer. *Radiat. Res.* 170, 49–59 (2008).
97. Langsenlehner, T. et al. Association between single nucleotide polymorphisms in the gene for XRCC1 and radiation-induced late toxicity in prostate cancer patients. *Radiother. Oncol. J. Eur. Soc. Ther. Radiol. Oncol.* 98, 387–393 (2011).
98. Fachal, L. et al. Association of a XRCC3 polymorphism and rectum mean dose with the risk of acute radio-induced gastrointestinal toxicity in prostate cancer patients. *Radiother. Oncol.* 105, 321–328 (2012).
99. Cintra, H. S. et al. Investigation of Genetic Polymorphisms Related to the Outcome of Radiotherapy for Prostate Cancer Patients. *Dis. Markers* 35, 701–710 (2013).
100. Song, Y.-Z. et al. ERCC2 polymorphisms and radiation-induced adverse effects on normal tissue: systematic review with meta-analysis and trial sequential analysis. *Radiat. Oncol. Lond. Engl.* 10, 247 (2015).
101. Guhlich, M. et al. Identification of Risk Loci for Radiotoxicity in Prostate Cancer by Comprehensive Genotyping of TGFB1 and TGFBR1. *Cancers* 13, 5585 (2021).

102. Uffelmann, E. et al. Genome-wide association studies. *Nat. Rev. Methods Primer* 1, 1–21 (2021).
103. Rotimi, C. et al. Community engagement and informed consent in the International HapMap project. *Community Genet.* 10, 186–198 (2007).
104. 1000 Genomes Project Consortium et al. A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073 (2010).
105. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* 48, 1279–1283 (2016).
106. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 590, 290–299 (2021).
107. Kerns, S. L. et al. Genome-wide association study to identify single nucleotide polymorphisms (SNPs) associated with the development of erectile dysfunction in African-American men after radiotherapy for prostate cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 78, 1292–1300 (2010).
108. Rosen, R. C., Cappelleri, J. C., Smith, M. D., Lipsky, J. & Peña, B. M. Development and evaluation of an abridged, 5-item version of the International Index of Erectile Function (IIEF-5) as a diagnostic tool for erectile dysfunction. *Int. J. Impot. Res.* 11, 319–326 (1999).
109. Kerns, S. L. et al. A two-Stage Genome-Wide Association Study to Identify Single Nucleotide Polymorphisms Associated With Development of Erectile Dysfunction Following Radiation Therapy for Prostate Cancer. *Int. J. Radiat. Oncol.* 85, e21–e28 (2013).
110. Zagar, T. M., Stock, R. G., Cesaretti, J. A. & Stone, N. N. Assessment of postbrachytherapy sexual function: a comparison of the IIEF-5 and the MSEFS. *Brachytherapy* 6, 26–33 (2007).
111. Kerns, S. L. et al. A 2-stage genome-wide association study to identify single nucleotide polymorphisms associated with development of erectile dysfunction following radiation therapy for prostate cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 85, e21–28 (2013).
112. Kerns, S. L. et al. Genome-wide association study identifies a region on chromosome 11q14.3 associated with late rectal bleeding following radiation therapy for prostate cancer. *Radiother. Oncol. J. Eur. Soc. Ther. Radiol. Oncol.* 107, 372–376 (2013).

113. Barry, M. J. et al. The American Urological Association Symptom Index for Benign Prostatic Hyperplasia. *J. Urol.* 148, 1549–1557 (1992).
114. Barnett, G. C. et al. A genome wide association study (GWAS) providing evidence of an association between common genetic variants and late radiotherapy toxicity. *Radiother. Oncol.* 111, 178–185 (2014).
115. Barnett, G. C. et al. Standardized Total Average Toxicity score: a scale- and grade-independent measure of late radiotherapy toxicity to facilitate pooling of data from different studies. *Int. J. Radiat. Oncol. Biol. Phys.* 82, 1065–1074 (2012).
116. Fachal, L. et al. A three-stage genome-wide association study identifies a susceptibility locus for late radiotherapy toxicity at 2q24.1. *Nat. Genet.* 46, 891–894 (2014).
117. Wang, T.-M. et al. Genome-Wide Association Study of Susceptibility Loci for Radiation-Induced Brain Injury. *J. Natl. Cancer Inst.* 111, 620–628 (2019).
118. Yang, D.-W. et al. Genome-wide association study identifies genetic susceptibility loci and pathways of radiation-induced acute oral mucositis. *J. Transl. Med.* 18, 224 (2020).
119. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* 8, 1826 (2017).
120. Zhao, Y., Tao, L., Yi, J., Song, H. & Chen, L. The Role of Canonical Wnt Signaling in Regulating Radioresistance. *Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol.* 48, 419–432 (2018).
121. Naderi, E. et al. A two-stage genome-wide association study of radiation-induced acute toxicity in head and neck cancer. *J. Transl. Med.* 19, 481 (2021).
122. Schack, L. M. H. et al. A genome-wide association study of radiotherapy induced toxicity in head and neck cancer patients identifies a susceptibility locus associated with mucositis. *Br. J. Cancer* 126, 1082–1090 (2022).
123. Tam, V. et al. Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20, 467–484 (2019).
124. Zeggini, E. & Ioannidis, J. P. A. Meta-analysis in genome-wide association studies. *Pharmacogenomics* 10, 191–201 (2009).
125. Kerns, S. L. et al. Radiogenomics Consortium Genome-Wide Association Study Meta-Analysis of Late Toxicity After Prostate Cancer Radiotherapy. *J. Natl. Cancer Inst.* 112, 179–190 (2020).

126. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 17, 122 (2016).
127. Oh, J. H. et al. Computational methods using genome-wide association studies to predict radiotherapy complications and to identify correlative molecular processes. *Sci. Rep.* 7, 43381 (2017).
128. Lee, S. et al. Machine Learning on a Genome-wide Association Study to Predict Late Genitourinary Toxicity After Prostate Radiation Therapy. *Int. J. Radiat. Oncol. Biol. Phys.* 101, 128–135 (2018).
129. Massi, M. C. et al. A Deep Learning Approach Validates Genetic Risk Factors for Late Toxicity After Prostate Cancer Radiotherapy in a REQUITE Multi-National Cohort. *Front. Oncol.* 10, 541281 (2020).
130. Trivizakis, E. et al. Artificial intelligence radiogenomics for advancing precision and effectiveness in oncologic care (Review). *Int. J. Oncol.* 57, 43–53 (2020).
131. Saxena, S. et al. Role of Artificial Intelligence in Radiogenomics for Cancers in the Era of Precision Medicine. *Cancers* 14, 2860 (2022).
132. Kang, J. et al. Machine Learning and Radiogenomics: Lessons Learned and Future Directions. *Front. Oncol.* 8, 228 (2018).
133. Qin, D. Next-generation sequencing and its clinical application. *Cancer Biol. Med.* 16, 4–10 (2019).
134. Behjati, S. & Tarpey, P. S. What is next generation sequencing? *Arch. Dis. Child. Educ. Pract. Ed.* 98, 236–238 (2013).
135. Metzker, M. L. Sequencing technologies - the next generation. *Nat. Rev. Genet.* 11, 31–46 (2010).
136. Xuan, J., Yu, Y., Qing, T., Guo, L. & Shi, L. Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett.* 340, 284–295 (2013).
137. Vinkšl, M., Writzl, K., Maver, A. & Peterlin, B. Improving diagnostics of rare genetic diseases with NGS approaches. *J. Community Genet.* 12, 247–256 (2021).
138. Posey, J. E. Genome sequencing and implications for rare disorders. *Orphanet J. Rare Dis.* 14, 153 (2019).
139. Singh, T. et al. Rare coding variants in 10 genes confer substantial risk for schizophrenia. *Nature* 604, 509–516 (2022).

140. Lee, S., Abecasis, G. R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.* 95, 5–23 (2014).
141. Wu, M. C. et al. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* 89, 82–93 (2011).
142. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies - PubMed. <https://pubmed.ncbi.nlm.nih.gov/22863193/>.
143. Sequence Kernel Association Tests for the Combined Effect of Rare and Common Variants - PMC. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3675243/>.
144. Guey, L. T. et al. Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. *Genet. Epidemiol.* 35, 236–246 (2011).
145. Villalvazo, P. et al. DPYD Exome, mRNA Expression and Uracil Levels in Early Severe Toxicity to Fluoropyrimidines: An Extreme Phenotype Approach. *J. Pers. Med.* 11, 792 (2021).
146. Patiño-García, A. et al. Whole exome sequencing characterization of individuals presenting extreme phenotypes of high and low risk of developing tobacco-induced lung adenocarcinoma. *Transl. Lung Cancer Res.* 10, 1327–1337 (2021).
147. Li, B. & Ritchie, M. D. From GWAS to Gene: Transcriptome-Wide Association Studies and Other Methods to Functionally Understand GWAS Discoveries. *Front. Genet.* 12, 713230 (2021).
148. Barbeira, A. N. et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* 9, 1825 (2018).
149. Nicolae, D. L. et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* 6, e1000888 (2010).
150. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res.* 22, 1748–1759 (2012).
151. Barbeira, A. N. et al. Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet.* 15, e1007889 (2019).
152. Atkins, I. et al. Transcriptome-Wide Association Study Identifies New Candidate Susceptibility Genes for Glioma. *Cancer Res.* 79, 2065–2071 (2019).

153. Wainberg, M. et al. Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* 51, 592–599 (2019).
154. Barbeira, A. N. et al. Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol.* 22, 49 (2021).
155. Battle, A. et al. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res.* 24, 14 (2014).
156. Pudjihartono, M., Perry, J. K., Print, C., O'Sullivan, J. M. & Schierding, W. Interpretation of the role of germline and somatic non-coding mutations in cancer: expression and chromatin conformation informed analysis. *Clin. Epigenetics* 14, 120 (2022).
157. Rodriguez-Fontenla, C. & Carracedo, A. UTMOST, a single and cross-tissue TWAS (Transcriptome Wide Association Study), reveals new ASD (Autism Spectrum Disorder) associated genes. *Transl. Psychiatry* 11, 1–11 (2021).
158. Gusev, A. et al. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat. Genet.* 50, 538–548 (2018).
159. Liu, D. et al. A transcriptome-wide association study identifies novel candidate susceptibility genes for prostate cancer risk. *Int. J. Cancer* 150, 80–90 (2022).
160. Mancuso, N. et al. Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. *Nat. Commun.* 9, 4079 (2018).
161. Mulford, A. J., Wing, C., Dolan, M. E. & Wheeler, H. E. Genetically regulated expression underlies cellular sensitivity to chemotherapy in diverse populations. *Hum. Mol. Genet.* 30, 305–317 (2021).
162. Zhang, J. et al. Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat. Genet.* 54, 593–602 (2022).
163. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* 558, 73–79 (2018).
164. Chick, J. M. et al. Defining the consequences of genetic variation on a proteome-wide scale. *Nature* 534, 500–505 (2016).
165. Assum, I. et al. Tissue-specific multi-omics analysis of atrial fibrillation. *Nat. Commun.* 13, 441 (2022).

166. Brandes, N., Linial, N. & Linial, M. PWAS: proteome-wide association study-linking genes and phenotypes by functional variation in proteins. *Genome Biol.* 21, 173 (2020).
167. Schubert, R. et al. Protein prediction for trait mapping in diverse populations. *PLoS ONE* 17, e0264341 (2022).
168. Baskar, R., Lee, K. A., Yeo, R. & Yeoh, K.-W. Cancer and Radiation Therapy: Current Advances and Future Directions. *Int. J. Med. Sci.* 9, 193–199 (2012).
169. Moore, C. M., Jacobso, S. A. & Fingerlin, T. E. Power and Sample Size Calculations for Genetic Association Studies in the Presence of Genetic Model Mis-specification. *Hum. Hered.* 84, 256–271 (2019).
170. R: The R Project for Statistical Computing. <https://www.r-project.org/>.
171. Dupont, W. D. & Plummer, W. D. Power and Sample Size Calculations for Studies Involving Linear Regression. *Control. Clin. Trials* 19, 589–601 (1998).
172. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303 (2010).
173. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498 (2011).
174. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* 43, 11.10.1-11.10.33 (2013).
175. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164 (2010).
176. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4, 7 (2015).
177. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890 (2018).
178. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26, 589–595 (2010).
179. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993 (2011).

180. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008 (2021).
181. Picard Tools - By Broad Institute. <http://broadinstitute.github.io/picard/>
<http://broadinstitute.github.io/picard/> (2018).
182. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745 (2016).
183. Gudmundsson, S. et al. Variant interpretation using population databases: Lessons from gnomAD. *Hum. Mutat.* 43, 1012–1030 (2022).
184. Liu, X., Li, C., Mou, C., Dong, Y. & Tu, Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Med.* 12, 103 (2020).
185. Rogers, M. F. et al. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinforma. Oxf. Engl.* 34, 511–513 (2018).
186. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 47, D886–D894 (2019).
187. Ioannidis, N. M. et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* 99, 877–885 (2016).
188. R Core Team (2020). — European Environment Agency. <https://www.eea.europa.eu/data-and-maps/indicators/oxygen-consuming-substances-in-rivers/r-development-core-team-2006>.
189. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLoS Genet.* 2, e190 (2006).
190. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909 (2006).
191. König, I. R. Validation in Genetic Association Studies. *Brief. Bioinform.* 12, 253–258 (2011).
192. Seibold, P. et al. REQUITE: A prospective multicentre cohort study of patients undergoing radiotherapy for breast, lung or prostate cancer. *Radiother. Oncol.* 138, 59–67 (2019).
193. Statistics and Machine Learning Toolbox Documentation - MathWorks España. <https://es.mathworks.com/help/stats/index.html>.

194. Auton, A. et al. A global reference for human genetic variation. *Nature* 526, 68–74 (2015).
195. S-PrediXcan | BIOS 25328 Cancer Genomics Class Notes. <https://bios25328.hakyimlab.org/post/2021/04/16/s-predixcan/>.
196. Gamazon, E. R. et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* 47, 1091–1098 (2015).
197. Watanabe, K., Umićević Mirkov, M., de Leeuw, C. A., van den Heuvel, M. P. & Posthuma, D. Genetic mapping of cell type specificity for complex traits. *Nat. Commun.* 10, 3222 (2019).
198. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740 (2011).
199. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30 (2000).
200. Gillespie, M. et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* 50, D687–D692 (2022).
201. Griss, J. et al. ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis. *Mol. Cell. Proteomics MCP* 19, 2115–2124 (2020).
202. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29 (2000).
203. The Gene Ontology Consortium et al. The Gene Ontology knowledgebase in 2023. *Genetics* 224, iyad031 (2023).
204. Filippi, A. R., Franco, P. & Ricardi, U. Is clinical radiosensitivity a complex genetically controlled event? *Tumori* 92, 87–91 (2006).
205. Ma, A. D. & Abrams, C. S. Pleckstrin Induces Cytoskeletal Reorganization via a Rac-dependent Pathway *. *J. Biol. Chem.* 274, 28730–28735 (1999).
206. Baig, A., Bao, X. & Haslam, R. J. Proteomic identification of pleckstrin-associated proteins in platelets: possible interactions with actin. *Proteomics* 9, 4254–4258 (2009).
207. Lian, L. et al. Loss of pleckstrin defines a novel pathway for PKC-mediated exocytosis. *Blood* 113, 3577–3584 (2009).
208. Wang, G., Zhou, Q., Xu, Y. & Zhao, B. Emerging Roles of Pleckstrin-2 Beyond Cell Spreading. *Front. Cell Dev. Biol.* 9, 768238 (2021).

209. Fröbel, J. et al. Platelet Proteome Analysis Reveals Integrin-dependent Aggregation Defects in Patients with Myelodysplastic Syndromes. *Mol. Cell. Proteomics MCP* 12, 1272–1280 (2013).
210. Petrova, A. S. THE MECHANISM OF DEVELOPMENT OF HEMORRHAGIC DIATHESIS IN ACUTE RADIATION SICKNESS. *Meditinskaya Radiol. USSR Engl. Transl. See Med Radiol USSR Vol: 3, No. 6, (1958).*
211. Kennedy, A. R., Maity, A. & Sanzari, J. K. A Review of Radiation-Induced Coagulopathy and New Findings to Support Potential Prevention Strategies and Treatments. *Radiat. Res.* 186, 121–140 (2016).
212. Krantz, S. & Lober, M. [Review: hemorrhagic diathesis resulting from acute exposure to ionizing radiation]. *Folia Haematol. Leipz. Ger.* 1928 104, 1–10 (1977).
213. Pavlidakey, P. G. & MacLennan, G. T. Radiation Cystitis. *J. Urol.* (2009) doi:10.1016/j.juro.2009.06.034.
214. Zhang, X., Basile, A. O., Pendergrass, S. A. & Ritchie, M. D. Real world scenarios in rare variant association analysis: the impact of imbalance and sample size on the power in silico. *BMC Bioinformatics* 20, 46 (2019).
215. Wang, D., Li, J., Wang, Y. & Wang, E. A comparison on predicting functional impact of genomic variants. *NAR Genomics Bioinforma.* 4, lqab122 (2022).
216. Forde, A., Hemani, G. & Ferguson, J. Review and further developments in statistical corrections for Winner's Curse in genetic association studies. 2022.11.28.518299 Preprint at <https://doi.org/10.1101/2022.11.28.518299> (2022).
217. Xiao, R. & Boehnke, M. Quantifying and correcting for the winner's curse in genetic association studies. *Genet. Epidemiol.* 33, 453–462 (2009).
218. Wang, W., Yang, Y., Li, L. & Shi, Y. Synleurin, a novel leucine-rich repeat protein that increases the intensity of pleiotropic cytokine responses. *Biochem. Biophys. Res. Commun.* 305, 981–988 (2003).
219. Ramakrishnan, V. et al. Increased thrombin responsiveness in platelets from mice lacking glycoprotein V. *Proc. Natl. Acad. Sci. U. S. A.* 96, 13336–13341 (1999).
220. Zieseniss, A. et al. Raver1 is an integral component of muscle contractile elements. *Cell Tissue Res.* 327, 583–594 (2007).
221. Hüttelmaier, S. et al. Raver1, a dual compartment protein, is a ligand for PTB/hnRNPI and microfilament attachment proteins. *J. Cell Biol.* 155, 775–786 (2001).

222. Seki, N. et al. LRP1B is a negative modulator of increased migration activity of intimal smooth muscle cells from rabbit aortic plaques. *Biochem. Biophys. Res. Commun.* 331, 964–970 (2005).
223. Liu, C. X., Li, Y., Obermoeller-McCormick, L. M., Schwartz, A. L. & Bu, G. The putative tumor suppressor LRP1B, a novel member of the low density lipoprotein (LDL) receptor family, exhibits both overlapping and distinct properties with the LDL receptor-related protein. *J. Biol. Chem.* 276, 28889–28896 (2001).
224. Haas, J. et al. LRP1b shows restricted expression in human tissues and binds to several extracellular ligands, including fibrinogen and apoE – carrying lipoproteins. *Atherosclerosis* 216, 342–347 (2011).
225. Tissue expression of PRB1 - Summary - The Human Protein Atlas. <https://www.proteinatlas.org/ENSG00000251655-PRB1/tissue>.
226. Stiubea-Cohen, R. et al. Effect of Irradiation on Cell Transcriptome and Proteome of Rat Submandibular Salivary Glands. *PLoS ONE* 7, e40636 (2012).
227. Schulze, H. et al. RanBP10 Is a Cytoplasmic Guanine Nucleotide Exchange Factor That Modulates Noncentrosomal Microtubules. *J. Biol. Chem.* 283, 14109–14119 (2008).
228. Kunert, S. et al. The microtubule modulator RanBP10 plays a critical role in regulation of platelet discoid shape and degranulation. *Blood* 114, 5532–5540 (2009).
229. Bruni, L. et al. Ionizing Radiation Effects on Hs 578Bst Microtubules. *Front. Phys.* 8, (2020).
230. Aguado Barrera, M. E. Radiogenómica en cáncer de pulmón y más allá: validación funcional de la región genómica 2q24.1. (2020).
231. Heinke, J. et al. BMPER is an endothelial cell regulator and controls bone morphogenetic protein-4-dependent angiogenesis. *Circ. Res.* 103, 804–812 (2008).
232. Helbing, T. et al. BMP activity controlled by BMPER regulates the proinflammatory phenotype of endothelium. *Blood* 118, 5040–5049 (2011).
233. Proteomics DB.
<https://www.proteomicsdb.org/proteomicsdb/#human/proteinDetails/Q9H6B1/expressi>
xpression.
234. Jørgensen, J. P. et al. Adrm1, a Putative Cell Adhesion Regulating Protein, is a Novel Proteasome-associated Factor. *J. Mol. Biol.* 360, 1043–1052 (2006).

235. Hamazaki, J. et al. A novel proteasome interacting protein recruits the deubiquitinating enzyme UCH37 to 26S proteasomes. *EMBO J.* 25, 4524–4536 (2006).
236. Lamerant, N. & Kieda, C. Adhesion properties of adhesion-regulating molecule 1 protein on endothelial cells. *FEBS J.* 272, 1833–1844 (2005).
237. Wang, J. et al. The CCDC43-ADRM1 axis regulated by YY1, promotes proliferation and metastasis of gastric cancer. *Cancer Lett.* 482, 90–101 (2020).
238. Yu, G.-Y. et al. RA190, a Proteasome Subunit ADRM1 Inhibitor, Suppresses Intrahepatic Cholangiocarcinoma by Inducing NF-KB-Mediated Cell Apoptosis. *Cell. Physiol. Biochem.* 47, 1152–1166 (2018).
239. Wu, W. et al. Prognostic and Therapeutic Significance of Adhesion-regulating Molecule 1 in Estrogen Receptor-positive Breast Cancer. *Clin. Breast Cancer* 20, 131-144.e3 (2020).
240. Boggs, J. M. Myelin basic protein: a multifunctional protein. *Cell. Mol. Life Sci. CMLS* 63, 1945–1961 (2006).
241. Wąsik, N. et al. Serum myelin basic protein as a marker of brain injury in aneurysmal subarachnoid haemorrhage. *Acta Neurochir. (Wien)* 162, 545–552 (2020).
242. Tian, Y., Shi, Z., Yang, S., Chen, Y. & Bao, S. Changes in myelin basic protein and demyelination in the rat brain within 3 months of single 2-, 10-, or 30-Gy whole-brain radiation treatments. *J. Neurosurg.* 109, 881–888 (2008).
243. Boucher, P., Gotthardt, M., Li, W.-P., Anderson, R. G. W. & Herz, J. LRP: role in vascular wall integrity and protection from atherosclerosis. *Science* 300, 329–332 (2003).
244. Wild, J. B. et al. Low Density Lipoprotein Receptor Related Protein 1 and Abdominal Aortic Aneurysms. *Eur. J. Vasc. Endovasc. Surg.* 44, 127–132 (2012).
245. GTEx Portal. <https://www.gtexportal.org/home/gene/PGPEP1L>.
246. Alliance of Genome Resources. <https://www.alliancegenome.org/gene/HGNC:27080>.
247. Shao, H., Teramae, D. & Wells, A. Axl contributes to efficient migration and invasion of melanoma cells. *PLOS ONE* 18, e0283749 (2023).

248. Auyez, A., Sayan, A. E., Kriajevska, M. & Tulchinsky, E. AXL Receptor in Cancer Metastasis and Drug Resistance: When Normal Functions Go Askew. *Cancers* 13, 4864 (2021).
249. Holland, S. J. et al. Multiple Roles for the Receptor Tyrosine Kinase Axl in Tumor Formation. *Cancer Res.* 65, 9294–9303 (2005).
250. Song, X. et al. Overexpression of Receptor Tyrosine Kinase Axl Promotes Tumor Cell Invasion and Survival in Pancreatic Ductal Adenocarcinoma. *Cancer* 117, 734–743 (2011).
251. Xie, C. et al. Intersectin 1 (ITSN1) identified by comprehensive bioinformatic analysis and experimental validation as a key candidate biological target in breast cancer. *OncoTargets Ther.* 12, 7079 (2019).
252. Yatsenko, A. S. et al. Profiling of the muscle-specific dystroglycan interactome reveals the role of Hippo signaling in muscular dystrophy and age-dependent muscle atrophy. *BMC Med.* 18, 8 (2020).
253. Mathews, K. P. Deficiencies in regulator proteins. 4. Anaphylatoxin inactivator. *Prog. Allergy* 39, 344–351 (1986).
254. Cao, H. & Hegele, R. A. DNA polymorphism and mutations in CPN1, including the genomic basis of carboxypeptidase N deficiency. *J. Hum. Genet.* 48, 0020–0022 (2003).
255. de Vries, P. S. et al. Whole-genome sequencing study of serum peptide levels: the Atherosclerosis Risk in Communities study. *Hum. Mol. Genet.* 26, 3442–3450 (2017).
256. Jiang, N., Xiao, Y., Liu, Y., Liu, W. & Liu, S. Blood coagulation factor VIII D1241E polymorphism leads to a weak malectin interaction and reduction of factor VIII posttranslational modification and secretion. *Exp. Cell Res.* 397, 112334 (2020).
257. Zhang, J. & Dutta, D. Jingning-Zhang/PlasmaProtein: Custom code for: Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. (2022) doi:10.5281/zenodo.6332981.
258. Kyriakides, T. R. et al. Mice That Lack Thrombospondin 2 Display Connective Tissue Abnormalities That Are Associated with Disordered Collagen Fibrillogenesis, an Increased Vascular Density, and a Bleeding Diathesis. *J. Cell Biol.* 140, 419–430 (1998).

259. Reactome | PRPF19 [nucleoplasm]. <https://reactome.org/content/detail/R-HSA-5420907>.
260. Biau, J., Chautard, E., Verrelle, P. & Dutreix, M. Altering DNA Repair to Improve Radiation Therapy: Specific and Multiple Pathway Targeting. *Front. Oncol.* 9, 1009 (2019).
261. WARTERS_RESPONSE_TO_IR_SKIN. https://www.gsea-msigdb.org/gsea/msigdb/cards/WARTERS_RESPONSE_TO_IR_SKIN.
262. Warters, R. L., Packard, A. T., Kramer, G. F., Gaffney, D. K. & Moos, P. J. Differential Gene Expression in Primary Human Skin Keratinocytes and Fibroblasts in Response to Ionizing Radiation. *Radiat. Res.* 172, 82–95 (2009).
263. National Institutes of Health, National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE) version 4.0| Protocol Development. https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm (2009).

9 ANEXOS

ANEXO 1. CLASIFICACIÓN TNM

Clasificación clínica TNM	
<i>Categoría T clínica (cT)</i>	
T	Tumor primario
TX	No se puede evaluar el tumor primario
T0	No existe evidencias de tumor primario
T1	Tumor clínicamente imperceptible, no palpable
T1a	Tumor encontrado incidentalmente o como hallazgo fortuito en una extensión $\leq 5\%$ del tejido reseado
T1b	Tumor incidentalmente detectado en $> 5\%$ del tejido reseado.
T1c	Tumor identificado mediante biopsia encontrado en uno o ambos lados, pero no palpable
T2	Tumor palpable y confinado dentro de la próstata.
T2a	Tumor que afecta la mitad de un lóbulo prostático o menos
T2b	Tumor que afecta más de la mitad de un lóbulo de la próstata, pero no ambos lados
T2c	Tumor que afecta ambos lóbulos prostáticos
T3	Tumor extraprostático que no se fija o no invade estructuras adyacentes
T3a	Extensión extraprostática uni o bilateral
T3b	El tumor invade las vesículas seminales
T4	Tumor fijo o invade estructuras adyacentes distintas de las vesículas seminales: cuello vesical, esfínter externo, recto, vejiga, músculos elevadores del ano y/o pared pélvica.
<i>Categoría T patológica (pT)</i>	
T	Tumor primario
T2	órgano confinado
T3	Extensión extraprostática
T3a	Extensión extraprostática (uni o bilateral) o microscópica invasión del cuello de la vejiga
T3b	El tumor invade las vesículas seminales
T4	Tumor fijo o invade estructuras adyacentes distintas de las vesículas seminales: cuello vesical, esfínter externo, recto, vejiga, músculos elevadores del ano y/o pared pélvica.
<i>Ganglios linfáticos regionales</i>	
Nx	No se evaluaron los ganglios linfáticos regionales
N0	No se detecta metástasis en ganglios regionales
N1	Existe metástasis en ganglios linfáticos regionales
<i>Metástasis a distancia</i>	
M0	No hay metástasis a distancia
M1	El tumor ha producido metástasis a distancia
M1a	Extendido a ganglios distantes
M1b	Extendido a hueso
M1c	Extendido en otras localizaciones con o sin enfermedad ósea

Adaptado de AJCC Cancer Staging Handbook Manual, octava edición (2017)³⁸.

ANEXO 2. ESTADIOS DEL CÁNCER DE PRÓSTATA

Estadios pronósticos del cáncer de próstata de acuerdo con la AJCC 8ª edición, basado en TNM, el nivel de PSA y el grado basado en el puntaje Gleason

Estadio	Tumor (T)	Nódulos (N)	Metástasis(M)	PSA	Grado Gleason
I	cT1-2a	N0	M0	<10	1
I	pT2	N0	M0	<10	1
IIA	cT1-2a	N0	M0	≥10 <20	1
IIA	cT2b-c	N0	M0	<20	1
IIB	T1-2	N0	M0	<20	2
IIC	T1-2	N0	M0	<20	3
IIC	T1-2	N0	M0	<20	4
IIIA	T1-2	N0	M0	≥20	1-4
IIIB	T3-4	N0	M0	Any	1-4
IIIC	Any T	N0	M0	Any	5
IVA	Any T	N1	M0	Any	Any
IVB	Any T	Any N	M1	Any	Any

cT, categoría clínica; pT patología T; PSA, antígeno prostático específico

Tabla adaptada de AJCC 8ª edición 2017³⁸.

ANEXO 3. TRATAMIENTO DEL CÁNCER DE PRÓSTATA LOCALIZADO

Tratamiento local del cáncer de próstata de acuerdo con la guía del 2020⁶ de la *European Association of Urology (EAU)*-*European Association of Nuclear Medicine (EANM)*-*European Society for Radiotherapy and Oncology (ESTRO)*-*European Society of Urogenital Radiology (ESUR)*- *International Society of Geriatric Oncology (SIOG)*

Guía de recomendaciones para las diversas etapas del cáncer de próstata localizado y localmente avanzado	
Enfermedad de bajo riesgo	
Vigilancia activa	En pacientes con esperanza de vida de >10 años B11:C34 y enfermedad de bajo riesgo
	Pacientes con histología intraductal y cribriforme en biopsia deben ser excluidos de AS
	Tomar biopsias dirigidas (de cualquier PI-RADS ≥ 3 lesión) y biopsia sistemática si se realiza biopsia confirmatoria
	Aconsejar a los pacientes sobre la posibilidad de necesitar tratamiento adicional en el futuro
Tratamiento activo	Cirugía y RT como alternativas a AS a pacientes aptos para dichos tratamientos y que acepten la compensación entre toxicidad y prevención de la progresión de la enfermedad
Enfermedad de riesgo intermedio	
Vigilancia activa	AS a pacientes seleccionados (<10 % Gleason 4), que acepten el riesgo potencial de metástasis adicionales
Prostatectomía radical	RP a pacientes con enfermedad de riesgo intermedio y esperanza de vida de >10 años
	Cirugía conservadora de nervios a pacientes con bajo riesgo de enfermedad extracapsular
Linfadenectomía pélvica	Realizar ePLND si el riesgo estimado de ganglios linfáticos positivos supera el 5 %
Radioterapia	<u>Braquiterapia LDR</u> a pacientes sin RTUP previa, con buen IPSS, y volumen prostático de <50 ml. <u>Para EBRT</u> . Dosis total de 76–78 Gy o hipofraccionamiento moderado (60 Gy/20 fx en 4 semanas o 70 Gy/28 fx en 6 semanas) + neoadyuvante a corto plazo + ADT concomitante (4–6 meses). Pacientes que no acepten someterse a ADT, usar dosis escalada de EBRT (76–80 Gy) o combinar con braquiterapia
Enfermedad localizada de alto riesgo	
Prostatectomía radical	A pacientes con cT3b-T4 N0 o cN1 solo como parte de la posible terapia multimodal
Linfadenectomía pélvica	Realizar una ePLND en pacientes con PCa de alto riesgo
Radioterapia	EBRT (76–78 Gy) + ADT a largo plazo (2–3 años)
	EBRT con refuerzo de braquiterapia HDR o LDR + ADT a largo plazo (2–3 años)
Enfermedad localmente avanzada	
Prostatectomía radical	A pacientes con cT3b-T4 N0 o cualquier cN1 solo como parte de la terapia multimodal
Linfadenectomía pélvica	Realizar una ePLND en pacientes con PCa de alto riesgo
Radioterapia	A pacientes con enfermedad cN0 localmente avanzada, ofrecer RT + ADT a largo plazo
Radioterapia + cirugía	A pacientes con enfermedad cN1, tratamiento local (ya sea PR o EBRT) + ADT a largo plazo

Otras opciones terapéuticas	Monoterapia con ADT a pacientes que no quieran o no puedan recibir ninguna forma de tratamiento local (tiempo de duplicación del PSA de <12 meses y PSA de >50 ng/mL), tumor poco diferenciado o síntomas molestos relacionados con la enfermedad local
Tratamiento adyuvante después de la prostatectomía radical	
	No prescribir ADT adyuvante en pacientes con pN0
	EBRT adyuvante a la cirugía a pacientes seleccionados
	Tres opciones de manejo con pacientes con pN + enfermedad después de un ePLND, basado en características de la afectación ganglionar: 1. ADT adyuvante, 2. ADT adyuvante con RT adicional, 3. Observación (manejo expectante) a un paciente después de ePLND, <2 nodos y un valor de PSA de <0.1 ng/mL y ausencia de extraganglionar extensión.
Tratamientos no curativos o paliativos de primera línea	
Enfermedad localizada	
Espera vigilante (WW)	A pacientes asintomáticos no elegibles para tratamiento curativo y aquellos con corta esperanza de vida
ADT=terapia de privación de andrógenos; AS=vigilancia activa; EBRT=radioterapia de haz externo; ePLND=disección extendida de ganglios linfáticos pélvicos; fx=fracciones; HDR=alta tasa de dosis; IPSS=Puntuación Internacional de Síntomas de Próstata; LDR=tasa de dosis baja; PCa=cáncer de próstata; PI-RADS=Informes de imágenes de próstata/Sistema de Datos; PSA=antígeno prostático específico; RP=prostatectomía radical; RT=radioterapia; RTUP=resección transuretral de la próstata; WW=espera vigilante.	

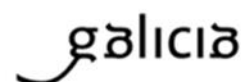
Adaptada de Mottet et al. (2020)⁶, Elsevier Global Rights Department.

ANEXO 4. APROBACIÓN POR EL COMITÉ DE ÉTICA



XUNTA DE GALICIA
CONSELLERÍA DE SANIDADE
Secretaría Xeral Técnica

Secretaría Técnica
Comité Autonómico de Ética da Investigación de Galicia
Secretaría Xeral, Consellería de Sanidade
Edificio Administrativo San Lázaro
15703 SANTIAGO DE COMPOSTELA
Tel: 881 546425; oeio@sergas.es



DITAME DO COMITÉ DE ÉTICA DA INVESTIGACIÓN DE SANTIAGO-LUGO

Juan Manuel Vázquez Lago, Secretario do Comité de Ética da Investigación de Santiago-Lugo

CERTIFICA:

Que este Comité avaliou na súa reunión do día 19/12/2016 o estudo:

Título: Radiogenómica: identificación de biomarcadores mediante estudos de asociación de xenoma completo (GWAS) y secuenciación de exomas (WES) en fenotipos extremos.

Promotor: Ana Vega

Tipo de estudo: Outros

Código de Rexistro: 2016/597

E, tomando en consideración as seguintes cuestións:

- A pertinencia do estudo, tendo en conta o coñecemento dispoñible, así coma os requisitos legais aplicables, e en particular a Lei 14/2007, de investigación biomédica, o Real Decreto 1716/2011, de 18 de novembro, polo que se establecen os requisitos básicos de autorización e funcionamento dos biobancos con fins de investigación biomédica e do tratamento das mostras biolóxicas de orixe humana, e se regula o funcionamento e organización do Rexistro Nacional de Biobancos para investigación biomédica, a ORDE SAS/3470/2009, de 16 de decembro, pola que se publican as Directrices sobre estudos Posautorización de Tipo Observacional para medicamentos de uso humano, e o RD 1090/2015, de 4 de decembro, polo que se regulan os ensaios clínicos con medicamentos, os Comités de Ética da Investigación con medicamentos e Rexistro Español de Estudos Clínicos
- A idoneidade do protocolo en relación cos obxectivos do estudo, xustificación dos riscos e molestias previsibles para o suxeito, así coma os beneficios esperados.
- Os principios éticos da Declaración de Helsinki vixente.
- Os Procedementos Normalizados de Traballo do Comité.

Emite un **INFORME FAVORABLE** para a realización do estudo polo/a investigador/a do centro:

Centros	Investigadores Principais
Fundación Pública Galega de Medicina Xenómica	Ana Vega

En Santiago de Compostela, a 19 de decembro de 2016

O secretario

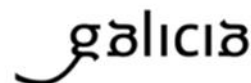


Juan M. Vázquez Lago



XUNTA DE GALICIA
CONSELLERÍA DE SANIDADE
Secretaría Xeral Técnica

Secretaría Técnica
Comité Autonómico de Ética da Investigación de Galicia
Secretaría Xeral, Consellería de Sanidade
Edificio Administrativo San Lázaro
15703 SANTIAGO DE COMPOSTELA
Tel: 881 546425; ceic@sergas.es



Juan M. Vázquez Lago, Secretario do Comité Territorial de Ética da Investigación de Santiago-Lugo,

FAI CONSTAR QUE:

1. O Comité Territorial de Ética da Investigación de Santiago-Lugo cumpre tanto na súa composición coma nos seus PNTs os requisitos legais vixentes (R.D 1090/2015 de ensaios clínicos, e a Lei 14/2007 de Investigación Biomédica).
2. A composición actual do Comité Territorial de Ética da Investigación de Santiago-Lugo é:

Pilar Rodríguez Ledo (Vicepresidenta). Médico especialista en Medicina Familiar e Comunitaria. Área de Xestión Integrada de Lugo.

Juan M. Vázquez Lago (Secretario). Médico especialista en Medicina Preventiva e Saúde Pública. Área de Xestión Integrada de Santiago.

Lorenzo Armenteros del Olmo. Médico especialista en Medicina Familiar e Comunitaria. Área de Xestión Integrada de Lugo.

Francisco Campos Pérez. Investigador. Instituto de Investigación Sanitaria de Santiago de Compostela.

Rosana Castelo Domínguez. Farmacéutica Atención Primaria. Área de Xestión Integrada de Santiago.

Daniel De La Iglesia García. Médico especialista en Dixestivo. Área de Xestión Integrada de Santiago.

Ricardo García Martínez. Licenciado en Dereito. Área de Xestión Integrada de Lugo.

Victor Herrán Carreira. Paciente, ADIL-Asociación de Diabéticos Lucense.

María Jesús Lamas Díaz. Farmacéutica especialista en Farmacia Hospitalaria. Área de Xestión Integrada de Santiago.

María Isabel Loza García. Licenciada en Farmacia. Universidade de Santiago de Compostela.

Guillermo José Prada Ramallal. Médico especialista en Farmacoloxía clínica. Área de Xestión Integrada de Santiago. Fundación "Ramón Domínguez". Santiago

Carlos Rodríguez Moreno. Médico especialista en Farmacoloxía clínica. Área de Xestión Integrada de Santiago.

María Jesús Wandosell Picatoste. Diplomada en Enfermería. Área de Xestión Integrada de Santiago.

Para que conste onde proceda, e a petición do promotor / investigador, en Santiago de Compostela, a 19 de decembro de 2016.

O Secretario do Comité Territorial de Ética da Investigación de Santiago Lugo,



ANEXO 5. ESCALAS DE TOXICIDAD DE ACUERDO CON LOS CTCAE v.4

ESCALAS DE TOXICIDAD DE ACUERDO A <i>COMMON TERMINOLOGY CRITERIA FOR ADVERSE EVENTS v.4.0</i> (CTCAE)					
Adverse Event	1	2	3	4	5
Gastrointestinal					
Proctitis	Rectal discomfort, intervention not indicated	Symptoms (e.g., rectal discomfort, passing blood or mucus); medical intervention indicated; limiting instrumental ADL	Severe symptoms; fecal urgency or stool incontinence; limiting self care ADL	Life-threatening consequences; urgent intervention indicated	Death
Diarrhea	Increase of <4 stools per day over baseline; mild increase in ostomy output compared to baseline	Increase of 4 - 6 stools per day over baseline; moderate increase in ostomy output compared to baseline	Increase of ≥ 7 stools per day over baseline; incontinence; hospitalization indicated; severe increase in ostomy output compared to baseline; limiting self care ADL	Life-threatening consequences; urgent intervention indicated	Death
Constipation	Occasional or intermittent symptoms; occasional use of stool softeners, laxatives, dietary modification, or enema	Persistent symptoms with regular use of laxatives or enemas; limiting instrumental ADL	Obstipation with manual evacuation indicated; limiting self care ADL	Life-threatening consequences; urgent intervention indicated	Death
Gastrointestinal perforation		Symptomatic; medical intervention indicated	Severe symptoms; elective operative intervention indicated	Life-threatening consequences; urgent operative intervention indicated	Death
Gastrointestinal stricture/stenosis	Asymptomatic; clinical or diagnostic observations only; intervention not indicated	Symptomatic; altered GI function	Severely altered GI function; tube feeding; hospitalization indicated; elective operative intervention indicated	Life-threatening consequences; urgent operative intervention indicated	Death

ESCALAS DE TOXICIDAD DE ACUERDO A COMMON TERMINOLOGY CRITERIA FOR ADVERSE EVENTS v.4.0 (CTCAE)					
Adverse Event	1	2	3	4	5
Gastrointestinal ulcer	Asymptomatic; clinical or diagnostic observations only; intervention not indicated	Moderate symptoms; medical intervention indicated; limiting instrumental ADL	Severely altered GI function; TPN indicated; elective operative or endoscopic intervention indicated; limiting self care ADL; disabling	Life-threatening consequences; urgent operative intervention indicated	Death
Rectal hemorrhage	Mild; intervention not indicated M	Moderate symptoms; medical intervention or minor cauterization indicated	Transfusion, radiologic, endoscopic, or elective operative intervention indicated	Life-threatening consequences; urgent intervention indicated	Death
Anal/Fecal incontinence	Occasional use of pads required	Daily use of pads required	Severe symptoms; elective operative intervention indicated		
Gastrointestinal fistula	Asymptomatic; clinical or diagnostic observations only; intervention not indicated	Symptomatic; altered GI function	Severely altered GI function; tube feeding, TPN or hospitalization indicated	Life-threatening consequences; urgent operative intervention indicated	Death
Genitourinary					
Urinary incontinence	Occasional (e.g., with coughing, sneezing, etc.), pads not indicated	Spontaneous; pads indicated; limiting instrumental ADL	Intervention indicated (e.g., clamp, collagen injections); operative intervention indicated; limiting self care ADL		
Urinary retention	Urinary, suprapubic or intermittent catheter placement not indicated; able to void with some residual	Placement of urinary, suprapubic or intermittent catheter placement indicated; medication indicated	Elective operative or radiologic intervention indicated; substantial loss of affected kidney function or mass	Life-threatening consequences; organ failure; urgent operative intervention indicated	Death

ESCALAS DE TOXICIDAD DE ACUERDO A <i>COMMON TERMINOLOGY CRITERIA FOR ADVERSE EVENTS v.4.0</i> (CTCAE)					
Adverse Event	1	2	3	4	5
Urinary frequency/urgency	Present	Limiting instrumental ADL; medical management indicated			
Cystitis	Microscopic hematuria; minimal increase in frequency, urgency, dysuria, or nocturia; new onset of incontinence	Moderate hematuria; moderate increase in frequency, urgency, dysuria, nocturia or incontinence; urinary catheter placement or bladder irrigation indicated; limiting instrumental ADL	Gross hematuria; transfusion, IV medications or hospitalization indicated; elective endoscopic, radiologic or operative intervention indicated	Life-threatening consequences; urgent radiologic or operative intervention indicated	Death
Hematuria	Asymptomatic; clinical or diagnostic observations only; intervention not indicated	Symptomatic; urinary catheter or bladder irrigation indicated; limiting instrumental ADL	Gross hematuria; transfusion, IV medications or hospitalization indicated; elective endoscopic, radiologic or operative intervention indicated; limiting self care ADL	Life-threatening consequences; urgent radiologic or operative intervention indicated	Death
Reproductive					
Erectile dysfunction	Decrease in erectile function (frequency or rigidity of erections) but intervention not indicated (e.g., medication or use of mechanical device, penile pump)	Decrease in erectile function (frequency/rigidity of erections), erectile intervention indicated, (e.g., medication or mechanical devices such as penile pump)	Decrease in erectile function (frequency/rigidity of erections) but erectile intervention not helpful (e.g., medication or mechanical devices such as penile pump); placement of a permanent penile prosthesis indicated (not previously present)		

Tabla adaptada de CTCAE v.4.0²⁶³

ANEXO 6. VARIANTES GENETICAS IDENTIFICADAS EN GENES PREVIAMENTE ASOCIADOS CON EFECTOS ADVERSOS A LA RADIOTERAPIA

Gen	Chr	Posición	ID	A1	A1 Casos Het/Hom	A1 Ctrls Het/Hom	A1 Freq	OR (95%CI)	SE	P- valor	AF nfe Exoma	AF nfe Genoma	EUR. sites
<i>TANC1</i>	2	160076227	rs144753108	G	2/0	1/0	0,011	7,91 (0,62-100,15)	1,29	0,11	0,0061	0,0089	0,006
<i>LRP1B</i>	2	140990879	rs148848707	C	2/0	2/0	0,014	2,60 (0,30-22,30)	1,09	0,384	0,0013	0,002	.
<i>SNED1</i>	2	241974013	rs200219556	A	1/0	1/0	0,007	3,51 (0,18-68,11)	1,51	0,407	0,004	0,0067	0,003
<i>TRAPPC11</i>	4	184585165	rs141909783	C	2/0	8/0	0,035	0,54 (0,10-2,77)	0,83	0,464	0,0137	0,0142	0,0159
<i>SNED1</i>	2	242009463	rs183525741	T	1/0	1/0	0,007	2,89 (0,16-53,05)	1,48	0,474	0,0004	0,0008	.
<i>LRP1B</i>	2	141707812	rs141827692	C	1/0	1/0	0,007	2,42 (0,15-40,07)	1,43	0,537	0,0041	0,0055	0,004
<i>DNAH5</i>	5	13701422	rs113425437	T	1/0	2/0	0,011	1,66 (0,12-22,96)	1,34	0,705	0,0356	0,0001	.
<i>DNAH5</i>	5	13721137	rs140832239	T	1/0	1/0	0,007	1,59 (0,09-28,52)	1,47	0,753	6,37x10 ⁻⁰⁵	0,0001	0,002
<i>SNED1</i>	2	242004296	rs150081750	A	1/0	2/0	0,011	0,87 (0,08-10,05)	1,25	0,914	0,0095	0,012	0,019
<i>KDM3B</i>	5	137688497	rs200772506	A	3/0	5/1	0,035	0,95 (0,26-3,51)	0,67	0,936	0,0266	0,0358	0,025
<i>KDM3B</i>	5	137754808	rs7706614	A	5/0	11/0	0,057	0,97 (0,31-3,04)	0,58	0,962	0,0364	0,0298	0,031
<i>LRP1B</i>	2	141812703	rs1699143665	T	1/0	1/0	0,007	1,07 (0,06-18,29)	1,45	0,965	.	.	.
<i>TRAPPC11</i>	4	184615166	rs75176151	C	1/0	2/0	0,010	1,02 (0,08-12,51)	1,27	0,987	0,0114	0,0106	0,0119
<i>AGT</i>	1	230845755	rs56073403	C	0/0	3/0	0,011	NA	NA	NA	0,0004	0,0006	0,001
<i>AGT</i>	1	230846446	rs61731497	G	2/0	0/0	0,007	NA	NA	NA	0,0012	0,002	0,001
<i>FSHR</i>	2	49190054	2:49190054:C:A	A	0/0	1/0	0,004	NA	NA	NA	.	.	.
<i>LRP1B</i>	2	141032021	rs149644677	A	1/0	0/0	0,004	NA	NA	NA	0,0038	0,0056	0,007
<i>LRP1B</i>	2	141092084	rs79879036	G	1/0	0/0	0,004	NA	NA	NA	0,0028	0,004	0,005
<i>LRP1B</i>	2	141113994	rs138237325	A	0/0	1/0	0,004	NA	NA	NA	0,0015	0,0021	0,003
<i>LRP1B</i>	2	141116420	rs150879175	T	2/0	0/0	0,007	NA	NA	NA	0,005	0,0073	0,007
<i>LRP1B</i>	2	141128779	rs76554185	G	3/0	0/0	0,011	NA	NA	NA	0,0263	0,0291	0,021
<i>LRP1B</i>	2	141130666	2:141130666:C:T	T	0/0	1/0	0,004	NA	NA	NA	.	.	.

Gen	Chr	Posición	ID	A1	A1 Casos Het/Hom	A1 Ctrls Het/Hom	A1 Freq	OR (95%CI)	SE	P- valor	AF nfe Exoma	AF nfe Genoma	EUR. sites
<i>LRP1B</i>	2	141232800	rs72899872	T	0/0	5/0	0,018	NA	NA	NA	0,011	0,0102	0,016
<i>LRP1B</i>	2	141298635	rs146867394	T	0/0	3/0	0,011	NA	NA	NA	0,0045	0,0064	0,009
<i>LRP1B</i>	2	141458182	rs145744070	A	0/0	1/0	0,004	NA	NA	NA	0,0013	0,0019	0,008
<i>LRP1B</i>	2	141526802	rs140355354	G	1/0	0/0	0,004	NA	NA	NA	6,38x10 ⁻⁰⁵	6,49x10 ⁻⁰⁵	.
<i>LRP1B</i>	2	141533706	2:141533706:C:A	A	0/0	1/0	0,004	NA	NA	NA	.	.	.
<i>LRP1B</i>	2	141643730	rs369920158	T	0/0	1/0	0,004	NA	NA	NA	3,19x10 ⁻⁰⁵	0	.
<i>TANC1</i>	2	159954295	rs114704202	A	0/0	1/0	0,004	NA	NA	NA	0,0051	0	.
<i>TANC1</i>	2	160019901	rs55929944	T	0/0	1/0	0,004	NA	NA	NA	0,032	0,0005	0,002
<i>TANC1</i>	2	160025797	rs767760023	C	0/0	1/0	0,004	NA	NA	NA	.	.	.
<i>TANC1</i>	2	160025824	rs754760244	A	0/0	1/0	0,004	NA	NA	NA	6,37x10 ⁻⁰⁵	0,0001	.
<i>TANC1</i>	2	160027142	rs915148272	T	0/0	1/0	0,004	NA	NA	NA	.	.	.
<i>TANC1</i>	2	160035215	rs200216556	T	0/0	1/0	0,004	NA	NA	NA	0,0011	0,0021	0,004
<i>TANC1</i>	2	160075857	rs758237953	C	0/1	0/0	0,007	NA	NA	NA	1x10 ⁻⁰⁵	.	.
<i>TANC1</i>	2	160082256	rs199615131	G	0/0	1/0	0,004	NA	NA	NA	0,0003	0,0003	.
<i>TANC1</i>	2	160087099	rs201678428	T	0/0	1/0	0,004	NA	NA	NA	0,0004	0,0001	.
<i>FHIT</i>	3	60522599	rs138978230	T	0/0	1/0	0,004	NA	NA	NA	0,0006	0	.
<i>DNAH5</i>	5	13736021	rs146750552	A	1/0	0/0	0,004	NA	NA	NA	0,0004	0,0002	.
<i>DNAH5</i>	5	13811889	rs35900306	T	0/0	2/0	0,007	NA	NA	NA	0,0203	0,0003	0,001
<i>DNAH5</i>	5	13829727	rs779506456	CA	0/0	1/0	0,004	NA	NA	NA	.	.	.
<i>DNAH5</i>	5	13864592	rs143567667	G	0/0	1/0	0,004	NA	NA	NA	0,0028	0	.
<i>DNAH5</i>	5	13871096	rs755018988	A	1/0	0/0	0,004	NA	NA	NA	.	.	.
<i>DNAH5</i>	5	13876888	rs61747516	T	1/0	0/0	0,004	NA	NA	NA	0,0122	0,0001	0,001
<i>DNAH5</i>	5	13902177	rs137878131	C	1/0	0/0	0,004	NA	NA	NA	0,0004	0,0008	0,001
<i>DNAH5</i>	5	13914743	rs140782270	T	0/0	1/0	0,004	NA	NA	NA	0,001	0,0016	0,003
<i>KDM3B</i>	5	137754691	rs773885881	T	1/0	0/0	0,004	NA	NA	NA	.	.	.
<i>KDM3B</i>	5	137762736	5:137762736:G:A	A	0/0	1/0	0,004	NA	NA	NA	.	.	.
<i>STING1</i>	5	138856923	rs140837017	T	0/0	3/0	0,011	NA	NA	NA	0,0069	0,0016	0,004
<i>CLINT1</i>	5	157214671	rs767548618	T	0/0	1/0	0,004	NA	NA	NA	9,56x10 ⁻⁰⁵	0	.

Gen	Chr	Posición	ID	A1	A1 Casos Het/Hom	A1 Ctrls Het/Hom	A1 Freq	OR (95%CI)	SE	P- valor	AF nfe Exoma	AF nfe Genoma	EUR. sites
<i>TGFBF1</i>	9	101891253	rs111513627	T	0/0	1/0	0,004	NA	NA	NA	0,0002	0,0002	.
<i>TGFBF1</i>	9	101894904	rs56014374	A	0/0	1/0	0,004	NA	NA	NA	0,0015	0,0021	0,002
<i>MS4A5</i>	11	60198380	rs751439414	C	0/0	1/0	0,004	NA	NA	NA	3,19x10 ⁻⁰⁵	0	.
<i>MTNR1B</i>	11	92714784	11:92714784:C:T	T	1/0	0/0	0,004	NA	NA	NA	.	.	.
<i>CEP128</i>	14	80997188	14:80997188:T:C	C	0/0	1/0	0,004	NA	NA	NA	.	.	.
<i>CEP128</i>	14	81251353	rs1901322567	T	0/0	1/0	0,004	NA	NA	NA	.	.	.
<i>CEP128</i>	14	81259336	rs45488095	C	0/0	1/0	0,004	NA	NA	NA	0,0064	0,0102	0,008
<i>CEP128</i>	14	81329199	rs145001230	A	1/0	0/0	0,004	NA	NA	NA	0,0002	0,0003	.
<i>CEP128</i>	14	81366359	14:81366359:A:T	T	0/0	1/0	0,004	NA	NA	NA	.	.	.
<i>PLK5</i>	19	1531836	rs184257515	T	1/0	0/0	0,004	NA	NA	NA	0,0037	0,0053	0,005

Chr, cromosoma; ID, identificación de variante; REF, alelo de referencia; A1, alelo de riesgo; A1 Casos, alelo de riesgo en casos; A1 Ctrls, alelo de riesgo en controles; Casos Het A1, casos con A1 en heterocigosis; casos Hom A1, casos con A1 en homocigosis; Ctrls Het A1, controles con A1 en heterocigosis; Ctrls Hom A1, controles con A1 en homocigosis; A1 Freq, frecuencia del alelo A1; OR, *odds ratio*; CI, intervalo de confianza; SE, desviación estandar; P-valor, p-valor del análisis multivariable; D, daño; N, neutral; AF nfe Exomas, frecuencia alélica en población europea no finlandesa GnomAD-exomas; AF nfe Genoma, frecuencia alélica en población europea no finlandesa GnomAD-exomeEUR.sites; , frecuencia alélica en población europea no finlandesa GnomAD-genoma; EUR.sites.

ANEXO 7. VARIANTES GENÉTICAS IDENTIFICADAS EN GENES PREVIAMENTE ASOCIADOS CON EFECTOS ADVERSOS A LA RADIOTERAPIA ANALIZADAS CON X2

Gen	Chr	Posición	ID	A1	F A	F U	OR	P-valor	AF nfe Exomas	AF nfe Genoma	EUR.sites	CADD16	fathmm XF	*ADA score	REVEL
LRP1B	2	141128779	rs76554185	G	0,032	0	NA	0,012	0,0263	0,0291	0,021	22,4	D	.	0,22
AGT	1	230846446	rs61731497	G	0,021	0	NA	0,041	0,0012	0,002	0,001	23,7	D	.	0,7
LRP1B	2	141116420	rs150879175	T	0,021	0	NA	0,041	0,005	0,0073	0,007	24	D	.	0,36
TANC1	2	160075857	rs758237953	C	0,021	0	NA	0,041	1x10 ⁻⁰⁵	.	.	23,2	D	.	0,26
LRP1B	2	141232800	rs72899872	T	0	0,02	0	0,116	0,011	0,0102	0,016	29,6	D	.	0,73
LRP1B	2	141032021	rs149644677	A	0,010	0	NA	0,15	0,0038	0,0056	0,007	23,1	D	.	0,43
LRP1B	2	141092084	rs79879036	G	0,010	0	NA	0,15	0,0028	0,004	0,005	23,8	D	.	0,47
LRP1B	2	141526802	rs140355354	G	0,010	0	NA	0,15	6,38x10 ⁻⁰⁵	6,49x10 ⁻⁰⁵	.	26,6	D	.	0,82
DNAH5	5	13736021	rs146750552	A	0,010	0	NA	0,15	0,0004	0,0002	.	26,8	D	.	0,67
DNAH5	5	13871096	rs755018988	A	0,010	0	NA	0,15	.	.	.	21,7	D	.	0,17
DNAH5	5	13876888	rs61747516	T	0,010	0	NA	0,15	0,0122	0,0001	0,001	22,2	N	.	0,03
DNAH5	5	13902177	rs137878131	C	0,010	0	NA	0,15	0,0004	0,0008	0,001	27,6	D	.	0,8
KDM3B	5	137754691	rs773885881	T	0,010	0	NA	0,15	.	.	.	25,1	D	.	0,24
MTNR1B	11	92714784	11:92714784:C:T	T	0,010	0	NA	0,15	.	.	.	23,8	N	.	0,65
CEP128	14	81329199	rs145001230	A	0,010	0	NA	0,15	0,0002	0,0003	.	29,5	N	.	0,17
PLK5	19	1531836	rs184257515	T	0,010	0	NA	0,15	0,0037	0,0053	0,005	20,6	D	.	.
AGT	1	230845755	rs56073403	C	0	0,015	0	0,225	0,0004	0,0006	0,001	20,1	D	.	0,66
LRP1B	2	141298635	rs146867394	T	0	0,01	0	0,225	0,0045	0,0064	0,009	29,1	D	.	0,81
STING1	5	138856923	rs140837017	T	0	0,015	0	0,225	0,0069	0,0016	0,004	22,9	N	.	0,03
DNAH5	5	13811889	rs35900306	T	0	0,01	0	0,323	0,0203	0,0003	0,001	23,1	D	.	0,14
FSHR	2	49190054	2:49190054:C:A	A	0	0,005	0	0,485	.	.	.	29,6	D	.	0,92
LRP1B	2	141113994	rs138237325	A	0	0,005	0	0,485	0,0015	0,0021	0,003	25,6	D	.	0,38
LRP1B	2	141130666	2:141130666:C:T	T	0	0,005	0	0,485	.	.	.	27,8	D	.	0,97
LRP1B	2	141458182	rs145744070	A	0	0,005	0	0,485	0,0013	0,0019	0,008	22	N	.	0,29
LRP1B	2	141533706	2:141533706:C:A	A	0	0,005	0	0,485	.	.	.	23,5	D	.	0,37
LRP1B	2	141643730	rs369920158	T	0	0,005	0	0,485	3,19x10 ⁻⁰⁵	0	.	31	D	.	0,79

Gen	Chr	Posición	ID	A1	F A	F U	OR	P- valor	AF nfe Exomas	AF nfe Genoma	EUR.sites	CADD16	fathmm XF	*ADA score	REVEL
TANC1	2	159954295	rs114704202	A	0	0,005	0	0,485	0,0051	0	.	26	N	.	0,37
TANC1	2	160019901	rs55929944	T	0	0,005	0	0,485	0,032	0,0005	0,002	23,5	D	.	0,2
TANC1	2	160025797	rs767760023	C	0	0,005	0	0,485	.	.	.	28,6	D	.	0,86
TANC1	2	160025824	rs754760244	A	0	0,005	0	0,485	6,37x10 ⁻⁰⁵	0,0001	.	25,4	D	.	0,7
TANC1	2	160027142	rs915148272	T	0	0,005	0	0,485	.	.	.	24,2	D	.	0,08
TANC1	2	160035215	rs200216556	T	0	0,005	0	0,485	0,0011	0,0021	0,004	19,34	D	.	0,23
TANC1	2	160082256	rs199615131	G	0	0,005	0	0,485	0,0003	0,0003	.	24,9	D	.	0,53
TANC1	2	160087099	rs201678428	T	0	0,005	0	0,485	0,0004	0,0001	.	23,9	D	.	0,41
FHIT	3	60522599	rs138978230	T	0	0,005	0	0,485	0,0006	0	.	23,6	D	.	0,82
DNAH5	5	13829727	rs779506456	CA	0	0,005	0	0,485	.	.	.	34	.	.	.
DNAH5	5	13864592	rs143567667	G	0	0,005	0	0,485	0,0028	0	.	23	D	.	0,35
DNAH5	5	13914743	rs140782270	T	0	0,005	0	0,485	0,001	0,0016	0,003	24,1	D	.	0,55
KDM3B	5	137762736	5:137762736:G:A	A	0	0,005	0	0,485	.	.	.	32	D	.	0,42
CLINT1	5	157214671	rs767548618	T	0	0,005	0	0,485	9,56x10 ⁻⁰⁵	0	.	23,3	N	.	0,09
TGFBR1	9	101891253	rs111513627	T	0	0,005	0	0,485	0,0002	0,0002	.	19,1	D	.	0,35
TGFBR1	9	101894904	rs56014374	A	0	0,005	0	0,485	0,0015	0,0021	0,002	20,5	N	.	0,09
MS4A5	11	60198380	rs751439414	C	0	0,005	0	0,485	3,19x10 ⁻⁰⁵	0	.	26,8	N	.	0,23
CEP128	14	80997188	14:80997188:T:C	C	0	0,005	0	0,485	.	.	.	27,7	D	.	0,34
CEP128	14	81251353	rs1901322567	T	0	0,005	0	0,485	.	.	.	22,1	N	.	0,07
CEP128	14	81259336	rs45488095	C	0	0,005	0	0,485	0,0064	0,0102	0,008	25,1	D	.	0,13
CEP128	14	81366359	14:81366359:A:T	T	0	0,005	0	0,485	.	.	.	23,8	N	.	0,25

Chr, cromosoma; ID, identificación de variante; REF, alelo de referencia; A1, alelo de riesgo; F A, frecuencia de A1 en casos; F U, frecuencia de A1 en controles; OR, *odds ratio*; CI, intervalo de confianza; D, dañina; N, neutral; AF nfe Exomas, frecuencia alelica en población europea no finlandesa GnomAD-exomas; AF nfe Genoma EUR.sites; , frecuencia alelica en población europea no finlandesa GnomAD-genoma, EUR.sites, frecuencia en población europea (1000 Genomas). Predictores usados CADD16, fathmm XF, REVEL, dbSNV ADA score (*ADA score). El análisis con X2 incluyó sólo a las 52 variantes que no pudieron ser analizadas mediante el análisis de regresión logística multivariable (ver apartado 4.1.4).

ANEXO 8. TEJIDOS ANALIZADOS EN TWAS

N.	Tissue	All Samples	EUR Samples
1	Adipose - Subcutaneous	581	479
2	Adipose - Visceral (Omentum)	469	393
3	Adrenal Gland	233	194
4	Artery - Aorta	387	329
5	Artery - Coronary	213	175
6	Artery - Tibial	584	476
7	Brain - Amygdala	129	119
8	Brain - Anterior cingulate cortex (BA24)	147	135
9	Brain - Caudate (basal ganglia)	194	172
10	Brain - Cerebellar Hemisphere	175	157
11	Brain - Cerebellum	209	188
12	Brain - Cortex	205	183
13	Brain - Frontal Cortex (BA9)	175	157
14	Brain - Hippocampus	165	150
15	Brain - Hypothalamus	170	156
16	Brain - Nucleus accumbens (basal ganglia)	202	181
17	Brain - Putamen (basal ganglia)	170	153
18	Brain - Spinal cord (cervical c-1)	126	115
19	Brain - Substantia nigra	114	100
20	Breast - Mammary Tissue	396	329
21	Skin - Transformed fibroblasts	483	403
22	Blood - EBV-transformed lymphocytes	147	113
23	Colon - Sigmoid	318	266
24	Colon - Transverse	368	294
25	Esophagus - Gastroesophageal Junction	330	275
26	Esophagus - Mucosa	497	411
27	Esophagus - Muscularis	465	385
28	Heart - Atrial Appendage	372	316
29	Heart - Left Ventricle	386	327
30	Kidney - Cortex	73	65
31	Liver	208	178
32	Lung	515	436
33	Minor Salivary Gland	144	114
34	Muscle - Skeletal	706	588
35	Nerve - Tibial	532	438
36	Pancreas	305	243
37	Pituitary	237	219
38	Prostate	221	181
39	Skin - Not Sun Exposed (Suprapubic)	517	430
40	Skin - Sun Exposed (Lower leg)	605	508
41	Small Intestine - Terminal Ileum	174	141
42	Spleen	227	179
43	Stomach	324	260
44	Testis	322	272
45	Thyroid	574	482
46	Whole Blood	670	558

N, tissue number; EUR Samples, samples from European population

ANEXO 9. GENES INCLUIDOS EN EL ANÁLISIS DE RUTAS

Genes incluidos en el análisis de rutas	
Estudio	Genes
Análisis de asociación del fenotipo sangrado rectal (SKAT-O)	<i>*MBP, PRB1, *PRB2, *CACNA1C-AS1, *LRRC70, TNFRSF10D, RRP8, PLEKHD1, RANBP10, GLO1, LCN15, *CACNA1C-AS2, R3HDM2, LRP1B, NCOA2, TNFRSF10C, PSMC1, TSSC2, *MYBPC2, *NUP188, CBLC, *PSMB8, PLEKHA1, *ZCCHC2, RAVR1, NSMCE1, UPK1A, CRNKL1</i>
Análisis de asociación del fenotipo extremos de sangrado rectal (SKAT-O)	<i>SULT1C2P1, FLNC-AS1, MIR4289, SLC37A3, *ADRM1, PRPF19, SGSM1, PRR7, RGS8, TMTC4, NFE2, ZNF385D, KCNE1B, GLCE, *CACNA1C-AS1, TIGD7, LRRC57, FGFBP3, SIX1</i>
Análisis de asociación combinado de variantes comunes y raras (SKAT-C)	<i>*MBP, LRP1, PGPEP1L, ZNF135, PRDX2, *CACNA1C-AS2, RAET1L, TTC28, RASGRP3, HSCB, P2RX6, TMEM151A, *CACNA1C-AS1, *LRRC70, SLC22A3, BUB1, AFF3, MIR3663, TMEM104, MBNL1, KRT18P55, *PRB2, ATP2A1-AS1, CCDC117, NPIP9, TPH2, TRPM4</i>
Estudio de asociación del transcriptoma completo (TWAS)	<i>AXL, ITSN1, WWC1, CYBRD1, GPD1, LIMS2, LTBP4, CPN1, MLEC, QPRT, FGFR1OP, TRIM22</i>
Estudios de asociación del proteoma completo (PWAS)	<i>NPTX2, *THBS2, ALCAM, CXCL1, AGFG1, STX8, EGF</i>

SKAT-O, *Optimal Sequence Kernel Association test*; SKAT-C, *Sequence Kernel Association Test-Combined (test of common rare variants)*; * genes asociados nominalmente en más de un estudio.

ANEXO 10. RUTAS ANALIZADAS

Lista de Rutas analizadas	
Rutas	Genes incluidos en cada ruta
Ruta inmunológica	ABCBI, ABL1, ADA, AHR, AICDA, AIRE, APP, ARG1, ARG2, ARHGDIB, ATG10, ATG12, ATG16L1, ATG5, ATG7, ATM, B2M, B3GAT1, BATF, BATF3, BAX, BCAP31, BCL10, BCL2, BCL2L11, BCL3, BCL6, BID, BLNK, BST1, BST2, BTK, BTLA, C14orf166, C1QA, C1QB, C1QBP, C1R, C1S, C2, C3, C4A/B, C4BPA, C5, C6, C7, C8A, C8B, C8G, C9, CAMP, CARD9, CASP1, CASP10, CASP2, CASP3, CASP8, CCBP2, CCL11, CCL13, CCL15, CCL16, CCL18, CCL19, CCL2, CCL20, CCL22, CCL23, CCL24, CCL26, CCL3, CCL4, CCL5, CCL7, CCL8, CCND3, CCR1, CCR10, CCR2, CCR5, CCR6, CCR7, CCR8, CCRL1, CCRL2, CD14, CD160, CD163, CD164, CD19, CD1A, CD1D, CD2, CD209, CD22, CD24, CD244, CD247, CD27, CD274, CD276, CD28, CD34, CD36, CD3D, CD3E, CD3EAP, CD4, CD40, CD40LG, CD44, CD45R0, CD45RA, CD45RB, CD46, CD48, CD5, CD53, CD55, CD58, CD59, CD6, CD7, CD70, CD74, CD79A, CD79B, CD80, CD81, CD82, CD83, CD86, CD8A, CD8B, CD9, CD96, CD97, CD99, CDH5, CDKN1A, CEACAM1, CEACAM6, CEACAM8, CEBPB, CFB, CFD, CFH, CFI, CFP, CHUK, CIITA, CISH, CLEC4A, CLEC4E, CLEC5A, CLEC6A, CLEC7A, CLU, CMKLR1, CR1, CR2, CRADD, CSF1, CSF1R, CSF2, CSF2RB, CSF3R, CTLA4, CTNNB1, CTSC, CTSG, CTSS, CUL9, CX3CL1, CX3CR1, CXCL1, CXCL10, CXCL11, CXCL12, CXCL13, CXCL2, CXCL9, CXCR1, CXCR2, CXCR3, CXCR4, CXCR6, CYBB, DEFB1, DEFB103A, DEFB103B, DEFB4A, DPP4, DUSP4, EBI3, EDNRB, EGR1, EGR2, ENTPD1, EOMES, ETS1, FADD, FAS, FCAR, FCER1A, FCER1G, FCGR1A/B, FCGR2A, FCGR2A/C, FCGR2B, FCGR3A/B, FCGR4, FKBP5, FN1, FOXP3, FYN, GATA3, GBP1, GBP5, GFI1, GNLY, GP1BB, GPI, GPR183, GZMA, GZMB, GZMK, HAMP, HAVCR2, HFE, HLA-A, HLA-B, HLA-C, HLA-DMA, HLA-DMB, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB3, HRAS, ICAM1, ICAM2, ICAM3, ICAM4, ICAM5, ICOS, ICOSLG, IDO1, IFI16, IFI35, IFIH1, IFIT2, IFITM1, IFNA1/13, IFNA2, IFNAR1, IFNAR2, IFNB1, IFNG, IFNGR1, IGF2R, IKBKAP, IKBKB, IKBKE, IKBKG, IKZF1, IKZF2, IKZF3, IL10, IL10RA, IL11RA, IL12A, IL12B, IL12RB1, IL13, IL13RA1, IL15, IL16, IL17A, IL17B, IL17F, IL18, IL18R1, IL18RAP, IL19, IL1A, IL1B, IL1R1, IL1R2, IL1RAP, IL1RL1, IL1RL2, IL1RN, IL2, IL20, IL21, IL21R, IL22, IL22RA2, IL23A, IL23R, IL26, IL27, IL28A, IL28A/B, IL29, IL2RA, IL2RB, IL2RG, IL3, IL32, IL4, IL4R, IL5, IL6, IL6R, IL6ST, IL7, IL7R, IL8, IL9, ILF3, IRAK1, IRAK2, IRAK3, IRAK4, IRF1, IRF3, IRF4, IRF5, IRF7, IRF8, IRGM, ITGA2B, ITGA4, ITGA5, ITGA6, ITGAE, ITGAL, ITGAM, ITGAX, ITGB1, ITGB2, ITLN1, ITLN2, JAK1, JAK2, JAK3, KCNJ2, KIR2DL1, KIR2DL2, KIR2DL4, KIR2DL5, KIR2DS1, KIR2DS2, KIR2DS3, KIR2DS4, KIR2DS5, KIR3DL1, KIR3DL2, KIR3DL3, KIR3DS1, KIT, KLRAP1, KLRB1, KLRC1, KLRC2, KLRC3, KLRC4, KLRD1, KLRF1, KLRF2, KLRG1, KLRG2, KLRK1, LAG3, LAIR1, LAMP3, LCK, LCP2, LEF1, LGALS3, LIF, LILRA1, LILRA2, LILRA3, LILRA4, LILRA5, LILRA6, LILRB1, LILRB2, LILRB3, LILRB4, LILRB5, LITAF, LTA, LTB4R, LTB4R2, LTBR, LTF, LY96, MAF, MALT1, MAP4K1, MAP4K2, MAP4K4, MAPK1, MAPK11, MAPK14, MAPKAPK2, MARCO, MASP1, MASP2, MBL2, MBP, MCL1, MIF, MME, MR1, MRC1, MS4A1, MSR1, MUC1, MX1, MYD88, NCAM1, NCF4, NCR1, NFATC1, NFATC2, NFATC3, NFIL3, NFKB1, NFKB2, NFKBIA, NFKBIZ, NLRP3, NOD1, NOD2, NOS2, NOTCH1, NOTCH2, NT5E, PAX5, PDCD1, PDCD1LG2, PDCD2, PDGFB, PDGFRB, PECAM1, PIGR, PLA2G2A, PLA2G2E, PLAU, PLAUR, PML, POU2F2, PPARG, PPBP, PRDM1, PRF1, PRKCD, PSMB10, PSMB5, PSMB7, PSMB8, PSMB9, PSMC2, PSMD7, PTAFR, PTGER4, PTGS2, PTK2, PTPN2, PTPN22, PTPN6, PTPRC, PYCARD, RAF1, RAG1, RAG2, RARRES3, RELA, RELB, RORC, RUNX1, S100A8, S100A9, S1PR1, SELE, SELL, SELPLG, SERPING1, SH2D1A, SIGIRR, SKI, SLAMF1, SLAMF6, SLAMF7, SLC2A1, SMAD3, SMAD5, SOCS1, SOCS3, SPP1, SRC, STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B, STAT6, SYK, TAGAP, TAL1, TAP1, TAP2, TAPBP, TBK1, TBX21, TCF4, TCF7, TFR3, TGFB1, TGFB2, TGFB3, TGFB4, THY1, TICAM1, TIGIT, TIRAP, TLR1, TLR2, TLR3, TLR4, TLR5, TLR7, TLR8, TLR9, TMEM173, TNF, TNFAIP3, TNFAIP6, TNFRSF10C, TNFRSF11A, TNFRSF13B, TNFRSF13C, TNFRSF14, TNFRSF17, TNFRSF1B, TNFRSF4, TNFRSF8, TNFRSF9, TNFSF10, TNFSF11, TNFSF12, TNFSF13B, TNFSF15, TNFSF4, TNFSF8, TOLLIP, TP53, TRAF1, TRAF2, TRAF3, TRAF4, TRAF5, TRAF6, TYK2, UBE2L3, VCAM1, VTN, XBP1, XCL1, XCR1, ZAP70, ZBTB16, ZEB1

Ruta de reparación del ADN	ABL1, ABRAXAS1, ACD, ACTB, ACTL6A, ACTR5, ACTR8, ADPRS, ALKBH2, ALKBH3, ALKBH5, APBB1, APEX1, AQR, ASCC1, ASCC2, ASCC3, ATM, ATR, ATRIP, BABAM1, BABAM2, BAP1, BARD1, BAZ1B, BLM, BRCA1, BRCA2, BRCC3, BRIP1, CCNA1, CCNA2, CCNH, CDK2, CDK7, CENPS, CENPX, CETN2, CHD1L, CHEK1, CHEK2, CLSPN, COPS2, COPS3, COPS4, COPS5, COPS6, COPS7A, COPS7B, COPS8, CUL4A, CUL4B, DCLRE1A, DCLRE1B, DCLRE1C, DDB1, DDB2, DNA2, DTL, ELL, EME1, EP300, ERCC1, ERCC2, ERCC3, ERCC4, ERCC5, ERCC6, ERCC8, EXO1, EYA1, EYA2, EYA3, EYA4, FAAP100, FAAP20, FAAP24, FAN1, FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCL, FANCM, FEN1, FTO, GEN1, GPS1, GTF2H1, GTF2H2, GTF2H3, GTF2H4, GTF2H5, H2AB1, H2AC14, H2AC18, H2AC20, H2AC4, H2AC6, H2AC7, H2AJ, H2AX, H2AZ1, H2AZ2, H2BC1, H2BC11, H2BC12, H2BC13, H2BC14, H2BC15, H2BC17, H2BC21, H2BC3, H2BC4, H2BC5, H2BC9, H2BS1, H2BU1, H3-4, H4C1, HERC2, HMGNI, HUS1, INO80, INO80B, INO80C, INO80D, INO80E, ISG15, ISY1, KAT5, KDM4B, KPNA2, LIG1, LIG3, LIG4, MAD2L2, MAPK8, MBD4, MCERS1, MDC1, MGMT, MLH1, MNAT1, MPG, MRE11, MSH2, MSH3, MSH6, MUS81, MUTYH, MUTYH, NBN, NEIL1, NEIL2, NEIL3, NFRKB, NHEJ1, NPLOC4, NSD2, NTHL1, OGG1, OGG1, PALB2, PARG, PARP1, PARP2, PAXIP1, PCLAF, PCNA, PIAS1, PIAS3, PIAS4, PMS2, PNKP, POLB, POLD1, POLD2, POLD3, POLD4, POLE, POLE2, POLE3, POLE4, POLH, POLI, POLK, POLL, POLM, POLN, POLQ, POLR2A, POLR2B, POLR2C, POLR2D, POLR2E, POLR2F, POLR2G, POLR2H, POLR2I, POLR2J, POLR2K, POLR2L, POT1, PPIE, PPP4C, PPP4R2, PPP5C, PRKDC, PRPF19, RAD1, RAD17, RAD18, RAD23A, RAD23B, RAD50, RAD51, RAD51AP1, RAD51B, RAD51C, RAD51D, RAD52, RAD9A, RAD9B, RBBP8, RBX1, RCHY1, REV1, REV3L, RFC1, RFC2, RFC3, RFC4, RFC5, RHNO1, RIF1, RMI1, RMI2, RNF111, RNF168, RNF4, RNF8, RPA1, RPA2, RPA3, RPS27A, RTEL1, RUVBL1, SIRT6, SLX1A, SLX4, SMARCA5, SMUG1, SPIDR, SPRTN, SUMO1, SUMO2, SUMO3, TCEA1, TDG, TDP1, TDP2, TERF1, TERF2, TERF2IP, TFPT, TIMELESS, TINF2, TIPIN, TOP3A, TOPBP1, TP53, TP53BP1, TRIM25, UBA52, UBA7, UBB, UBC, UBE2B, UBE2I, UBE2L6, UBE2N, UBE2T, UBE2V2, UBXN1, UFD1, UIMC1, UNG, USP1, USP10, USP43, USP45, USP7, UVSSA, VCP, WDR48, WRN, XAB2, XPA, XPC, XRCC1, XRCC2, XRCC3, XRCC4, XRCC5, XRCC6, YY1, ZNF830
Ruta Inflamatoria	AGER, ALOX12, ALOX15, ALOX5, AREG, ARG1, ATF2, BCL2L1, BCL6, BIRC2, C1QA, C1QB, C1R, C1S, C2, C3, C3AR1, C4A, C5, C6, C7, C8A, C8B, C9, CCL11, CCL13, CCL16, CCL17, CCL19, CCL2, CCL20, CCL21, CCL22, CCL23, CCL24, CCL3, CCL4, CCL5, CCL7, CCL8, CCR1, CCR2, CCR3, CCR4, CCR7, CD163, CD4, CD40, CD40LG, CD55, CD86, CDC42, CEBPB, CFB, CFD, CFL1, CREB1, CRP, CSF1, CSF2, CSF3, CXCL1, CXCL10, CXCL2, CXCL3, CXCL5, CXCL6, CXCL9, CXCR1, CXCR2, CXCR4, CYSLTR1, CYSLTR2, DAXX, DDIT3, DEFA1, ELK1, FASLG, FLT1, FOS, FXD2, GNAQ, GNAS, GNB1, GNGT1, GRB2, HDAC4, HIF1A, HLA-DRA, HLA-DRB1, HMGB1, HMGB2, HMGNI, HRAS, HSH2D, HSPB1, HSPB2, IFI44, IFIT1, IFIT2, IFIT3, IFNA1, IFNB1, IFNG, IL10, IL10RB, IL11, IL12A, IL12B, IL13, IL15, IL17A, IL18, IL18RAP, IL1A, IL1B, IL1R1, IL1RAP, IL1RN, IL2, IL21, IL22, IL22RA2, IL23A, IL23R, IL3, IL4, IL5, IL6, IL6R, IL7, IL8, IL9, IRF1, IRF3, IRF5, IRF7, ITGB2, JUN, KEAP1, KNG1, LIMK1, LTA, LTB, LTB4R, LTB4R2, LY96, MAFF, MAFG, MAFK, MAP2K1, MAP2K4, MAP2K6, MAP3K1, MAP3K5, MAP3K7, MAP3K9, MAPK1, MAPK14, MAPK3, MAPK8, MAPKAPK2, MAPKAPK5, MASP1, MASP2, MAX, MBL2, MEF2A, MEF2BNB-MEF2B, MEF2C, MEF2D, MKNK1, MMP3, MMP9, MRC1, MX1, MX2, MYC, MYD88, MYL2, NFATC3, NFE2L2, NFKB1, NLRP3, NOD1, NOD2, NOS2, NOX1, NR3C1, OAS2, OASL, OXER1, PDGFA, PIK3C2G, PLA2G4A, PLCB1, PPP1R12B, PRKCA, PRKCB, PTGDR2, PTGER1, PTGER2, PTGER3, PTGER4, PTGFR, PTGIR, PTGS1, PTGS2, PTK2, RAC1, RAF1, RAPGEF2, RELA, RELB, RHOA, RIPK1, RIPK2, ROCK2, RPS6KA5, SHC1, SMAD7, STAT1, STAT2, STAT3, TBXA2R, TCF4, TGFB1, TGFB2, TGFB3, TGFB3R1, TLR1, TLR2, TLR3, TLR4, TLR5, TLR6, TLR7, TLR8, TLR9, TNF, TNFAIP3, TNFSF14, TOLLIP, TRADD, TRAF2, TREM2, TSLP, TWIST2, TYROBP
Ruta de hematopoiesis	ACVRL1, AMACR, APC, BMPR1A, ENG, F5, HPS1, IL10RA, PLVAP, PTEN, SMAD4, TAOK1, TTC7A, WAS, WIPF1
Ruta de miogénesis	CDC42, MEF2D, MYOG, CTNNA2, CTNNB1, BOC, MEF2C, CTNNA1, MAPK14, ABL1, MYOD1, CDON, MYF6, MYF5, TCF12, BNIP2, NEO1, MEF2A, NTN3, CDH15, SPAG9, MAP2K6, CDH2, TCF4, TCF3, MEF2B, CDH4, MAPK12, MAPK11,
Ruta de fusión de mioblastos	PLEKHO1, KCNH1, CAPN2, FER1L5, TANC1, CAV3, PITX2, MYMX, ADGRB3, MYMK, MYOF, ADAM12, MYOD1, CDON, NOS1, CACNA1H, NPHS1, DYRK1B, MYH9



Aproximadamente el 50% de los pacientes con cáncer recibirán radioterapia como parte de su tratamiento. A pesar de que los avances tecnológicos en radioterapia permiten aplicar el tratamiento de forma cada vez más precisa en el tumor, parte del tejido sano será inevitablemente irradiado, lo que puede dar lugar al desarrollo de efectos adversos. Estos efectos impactarán negativamente en la calidad de vida del paciente. Se ha descrito que los factores genéticos desempeñan un papel muy importante en el desarrollo y severidad de los efectos adversos a la radioterapia. Así, el propósito de esta tesis es la identificación de marcadores genéticos asociados con el sangrado rectal, uno de los efectos adversos más relevantes en el tratamiento radioterápico de pacientes con cáncer de próstata.