



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

“Identificación de variantes genéticas y su asociación al riesgo a desarrollar tumores de células germinales de testículo en población mexicana”

TESIS

QUE PARA OPTAR POR EL GRADO DE:
Maestro en Ciencias

PRESENTA:

QFB. BERENICE CUEVAS ESTRADA

TUTOR PRINCIPAL

Dr. Rodrigo González Barrios de la Parra
[Instituto Nacional de Cancerología](#)

MIEMBROS DEL COMITÉ TUTOR

Dra. Talia Wegman Ostrosky

[Instituto Nacional de Cancerología](#)

Dr. Edgar Ricardo Vázquez Martínez

[Facultad de Química](#)

Dra. Alejandra Cervera Taboada

[Instituto Nacional de Medicina Genómica](#)

Ciudad de México. Octubre, 2023



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Este trabajo se realizó bajo la dirección del Dr. Rodrigo González Barrios de la Parra en el Laboratorio de Carcinogénesis del Dr. Luis Alonso Herrera Montalvo, de la Unidad de Investigación Biomédica en Cáncer en el Instituto Nacional de Cancerología (INCan).

Este proyecto fue financiado por el INCan y el Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) con la subvención número 290041.

AGRADECIMIENTOS INSTITUCIONALES

Al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT), por el otorgamiento de la beca nacional para mis estudios de maestría con número 1145899.

Al Programa de Apoyo a los Estudios de Posgrado (PAEP) de la Maestría en Ciencias Bioquímicas en la Universidad Nacional Autónoma de México (UNAM) por el apoyo económico para la asistencia a la *Cancer Genetics and Epigenetics Gordon Research Conference* en Lucca, Italia.

Al Dr. Rodrigo González Barrios, por la confianza depositada en mí, la motivación y el apoyo para el manejo del proyecto.

A los miembros de mi comité tutor: Dra. Alejandra Cervera Taboada, Dra. Talia Wegman Ostrosky y Dr. Edgar Ricardo Vázquez Martínez, por ayudarme a construir y direccionar este trabajo.

A la Biol. Michel Montalvo y al MC. Juan Alberto Ríos por todo su trabajo en la base de datos clínicos y el proyecto TCGT. A la M en C. Karla Torres por su apoyo dentro y fuera del laboratorio. A todo el equipo RGB, por las victorias y los fracasos compartidos juntos.

A los miembros del jurado de examen: Dra. Leticia Rocha Zavaleta, Dr. Samuel Canizales Quinteros, Dra. María Eugenia Gonsebatt Bonaparte, Dra. Martha Robles Flores y Dr. Ernesto Soto Reyes Solís por su tiempo y sus invaluable comentarios que ayudaron a mejorar la estructura de la tesis.

Al Dr. Luis A. Herrera Montalvo por la oportunidad de formar parte de su grupo de investigación.

A todos los integrantes del laboratorio de carcinogénesis del INCan, con una especial mención a la M en C. Clementina Castro por su incansable trabajo y apoyo técnico.

A la Universidad Autónoma Metropolitana (UAM)-Cuajimalpa donde se realizó el análisis bioinformático, específicamente, en el servidor esrs-epigenetics financiado por el CONAHCYT subvención número 00312021. Agradezco, por tanto, al Dr. Ernesto Soto Reyes y a la M en C. Aylin del Moral por mantener las condiciones para el desarrollo del trabajo en el servidor.

A la Universidad Nacional Autónoma de México y al Instituto Nacional de Cancerología, las instituciones a las que debo enteramente mi formación en el área científica.

DEDICATORIA

A mis padres, Mario e Isabel por su amor y apoyo incondicional. Ustedes son mi mundo entero, y yo, soy el reflejo de su esfuerzo a través de todos estos años.

A mis hermanas, Kenia, Brenda y Erika, las personas que siempre van a estar para mí, y por las que yo haría lo que fuera.

A mi familia, mi constante ejemplo de trabajo duro, resiliencia y humildad.

A todos mis amigos, pero especialmente a Elissa, Ximena, Xiadani, Daniela, Melba, Paramont, Rocha y Diego.

A las maravillosas personas que conocí en el laboratorio: Karla, Marina, Fer, Mich y Beto.

Al Dr. Rodrigo, por ser un gran mentor y ahora, un amigo.

Dedico esto a todos ustedes, por llenar de color mis días y no dejarme vencer por los miedos. Gracias por recordarme quién soy y creer en mí. Gracias por quedarse conmigo.

A todos los que me ayudaron a hacerle frente de este rompecabezas, mi gratitud infinita. Los quiero profundamente a todos, hasta el final de los días.

Finalmente, dedico esto a la niñita soñadora que nunca he dejado (ni dejaré) de ser y a cada instante en mi vida que me ha permitido confirmar que el camino que elegí es el correcto.

TABLA DE CONTENIDO

1. INTRODUCCIÓN.....	- 10 -
1.1. Panorama clínico y genómico de los TCGT.....	- 10 -
1.1.1. Predisposición y factores de riesgo.....	- 10 -
1.1.2. Origen del tumor.....	- 12 -
1.1.3. Perfil mutacional de los TCGT	- 12 -
1.1.4. Sintomatología de los TCGT	- 13 -
1.1.5. Diagnóstico y tratamiento	- 14 -
1.2. Panorama epidemiológico del cáncer testicular.....	- 15 -
1.2.1. El cáncer de testículo en México.....	- 15 -
1.2.2. El cáncer testicular en Latinoamérica	- 17 -
1.3. Los estudios de asociación de tipo casos-control.....	- 18 -
1.4. Secuenciación de exomas completos.	- 19 -
1.4.1. WES en el análisis de variantes raras.....	- 20 -
1.4.2. Clasificación de las variantes en función de su patogenicidad.	- 21 -
2. ANTECEDENTES DIRECTOS.....	- 23 -
3. PLANTEAMIENTO DEL PROBLEMA	- 25 -
4. JUSTIFICACION	- 25 -
5. HIPÓTESIS	- 26 -
6. OBJETIVOS	- 26 -
6.1. Objetivo general	- 26 -
6.2. Objetivos particulares.....	- 26 -
7. MÉTODOLOGÍA.....	- 27 -
7.1. Diseño experimental:.....	- 27 -
7.2. Población de estudio	- 27 -
7.2.1. Grupo CASOS	- 27 -
7.2.2. Grupo CONTROL 1	- 28 -
7.3. Procesamiento de los datos de secuenciación de exoma completo	- 28 -
7.3.1. Control de calidad, alineamiento y llamado de variantes.....	- 28 -
7.3.2. Análisis de ancestría genética global	- 29 -
7.3.3. Análisis de parentesco	- 29 -
7.4. Anotación de variantes y clasificación de variantes patogénicas	- 30 -
7.5. Análisis estadístico de asociación.....	- 30 -
7.6. Selección de un panel de variantes	- 31 -
7.7. Validación de los resultados de la secuenciación de exomas	- 31 -
7.7.1. Extracción de las muestras de DNA sanguíneo.....	- 31 -
7.7.2. Genotipificación de las muestras mediante la técnica de Restricción Fragment Length Polymorphisms (RFLP's).....	- 32 -

7.8.	Análisis <i>in silico</i> de las proteínas de tipo silvestre y las mutantes.	- 33 -
8.	RESULTADOS Y DISCUSIÓN.....	- 34 -
8.1.	Estimación de la composición ancestral de la población de estudio	- 34 -
8.2.	Estimación del parentesco entre los individuos de la población de estudio.	- 36 -
8.3.	Selección de las variantes germinales para el estudio de asociación.	- 38 -
8.4.	Asociación de tipo casos-control a nivel de variantes individuales.....	- 43 -
8.5.	Asociación de tipo casos-control por agregación de múltiples variantes basada en genes (<i>Gene-burden test</i>).	- 46 -
8.6.	Panel de variantes candidatas para su estudio en cohortes más numerosas de TCGT.....	- 48 -
8.7.	Validación de los datos de secuenciación de exoma completo mediante la técnica de RFLPs.....	- 50 -
8.7.1.	Variante rs147153778 en <i>SLCO4A1</i>	- 51 -
8.7.2.	Variante rs2273499 en <i>ARFGAP1</i>	- 52 -
8.7.3.	Variante rs1044301916 en <i>PABPC3</i>	- 53 -
8.8.	Análisis <i>in silico</i> de las proteínas de tipo silvestre y las mutantes	- 56 -
8.8.1.	Variante rs1044301916 en <i>PABPC3</i> (p.L276H).....	- 56 -
8.8.2.	Variante rs147153778 en <i>SLCO4A1</i> (p.R450W)	- 58 -
8.8.3.	Variante rs2273499 en <i>ARFGAP1</i> (p.V184M)	- 60 -
9.	COMENTARIOS FINALES.....	- 62 -
10.	CONCLUSIÓN.....	- 63 -
11.	PERSPECTIVAS	- 64 -
12.	ANEXO	- 65 -
12.1.	Estimación de la ancestría global a partir de un análisis de componentes principales (PCA). - 65 -	- 65 -
12.2.	Producción científica de la alumna durante la duración de la maestría.....	- 66 -
12.2.1.	Carteles presentados en congresos internacionales	- 66 -
12.2.2.	Artículos publicados en revistas indexadas.	- 67 -
13.	REFERENCIAS BIBLIOGRÁFICAS	- 68 -

ABREVIATURAS

ACMG	Colegio Estadounidense de Genética Médica y Genómica
AF	Frecuencia alélica
AFP	Alfafetoproteína
AMP	Asociación de Patología Molecular
ASCO	American Society of Clinical Oncology
ASDR	Tasa de mortalidad estandarizada por edad
ASIR	Tasa de incidencia estandarizada por edad
BEP	Cisplatino, bleomicina y etopósido
CADD	<i>Combined Annotation Dependent Depletion</i>
CNVs	Cambios en el número de copias
EAR	Espacio de ancestría de referencia
GCNIS	Neoplasia de células germinales <i>in situ</i>
GLOBOCAN	<i>Global Cancer Observatory</i>
GWAS	Estudios de asociación del genoma completo
hCG	Gonadotropina coriónica humana
HGDP	Panel de Diversidad del Genoma Humano
i(12p)	Isocromosoma 12
IARC	Agencia Internacional de Investigación del Cáncer
IBD	Idénticos por descendencia
IBS	Idénticos por estado
IC	Intervalo de confianza
IGCCCG	Grupo Colaborativo Internacional sobre Cáncer de Células Germinales
INCan	Instituto Nacional de Cancerología
LB	Variante probablemente benigna
LDH	Lactato deshidrogenasa
LP	Variante probablemente patogénica
Mb	Megabase
NS	No seminomas
OR	Razón de probabilidad
PABP	Proteínas de unión a poli(A)
PGC	Células germinales primordiales
PRS	Puntuación de riesgo poligénico
REVEL	<i>Rare Exome Variant Ensemble Learner</i>
RFLP	<i>Restriction Fragment Length Polymorphism</i>
RRM	Motivo de reconocimiento de RNA
SE	Seminomas
SNPs	Polimorfismos de un solo nucleótido
SNVs	Variantes de un solo nucleótido
TCa	Cáncer de testículo
TCGA	<i>The Cancer Genome Atlas Program</i>
TCGT	Tumores de células germinales de testículo
Tm	Temperatura de alineamiento
TMB	Carga de mutación tumoral
TNM	Escala tumor, nodo, metástasis
TRAPD	<i>Test Rare vAriants with Public Data</i>
VC	Variantes calificadas
VCF	Formato de llamado de variantes
VUS	Variante se significado incierto
WES	Secuenciación de exomas completos

RESUMEN

Los tumores de células germinales de testículo (TCGT) representan aproximadamente el 1-5% de todos los tumores malignos en hombres y son más comunes en hombres jóvenes entre las edades de 15 y 44 años. Los patrones de incidencia y mortalidad de los TCGT varían geográficamente. Sin embargo, la falta de una reducción significativa en las tasas de mortalidad en México y Latinoamérica en la última década, en contraste con las tendencias observadas en otras regiones del mundo, plantea la posibilidad de que en nuestra población existan particularidades a nivel molecular que ejerzan su influencia en el desarrollo y la evolución clínica de los TCGT.

La evidencia de las alteraciones en línea germinal distintivas de los TCGT se restringe a algunos casos familiares y estudios de GWAS, los cuales han permitido la identificación de 78 *loci* de susceptibilidad asociados a TCGT. Estos estudios tienen una sobrerrepresentación de pacientes de origen europeo y norteamericano. Por tanto, las características genómicas en nuestra población no han sido propiamente caracterizadas.

El objetivo de este trabajo consistió en identificar variantes germinales asociadas al riesgo a desarrollar TCGT, mediante el análisis de datos de secuenciación de exoma completo de 40 pacientes del Instituto Nacional de Cancerología. Inicialmente se generó un análisis de ancestría genética global para seleccionar únicamente a los individuos con la composición ancestral prevista para la población mestizo-mexicana. A continuación, se seleccionó un total de 2778 variantes genéticas raras [Frecuencia alélica (AF) en la población global <1%], que incluían variantes clasificadas en ClinVar como “patogénicas” o “probablemente patogénicas”, y variantes de significado clínico incierto que son potencialmente dañinas para la función del producto génico, según los algoritmos de CADD y REVEL.

Se comparó la frecuencia de las variantes en el grupo de casos de TCGT con dos grupos control independientes: el primero consistió en 36 individuos mexicanos residentes en la CDMX y el área metropolitana, mientras que el segundo se basó en datos de ~17,000 latinos de la base de datos gnomAD. A partir de los análisis, se propuso un panel de variantes candidatas que presentaron los valores de p más significativos en los dos ensayos independientes de asociación. Este panel incluye a las variantes rs1044301916 (T>A) en *PABPC3*, variante rs2273499 (G>A) en *ARFGAP1*, rs147153778 (C>T) en *SLCO4A1*, rs79783591 (A>T) en *MC4R* y rs1234654104 (TA>T) en *EHTM1*. Estas variantes se plantean como biomarcadores potenciales de riesgo a los TCGT. Paralelamente, se realizó una prueba de asociación por agregación de múltiples variantes basadas en genes, donde se mostró que el gen *PABPC3*, que tiene una expresión específica en testículo, está significativamente asociado a los TCGT bajo un modelo dominante ($p < 2.5 \times 10^{-6}$).

En conjunto, en este trabajo se identificaron variantes que no han sido descritas como *loci* de riesgo de TCGT y que pueden ser relevantes para el desarrollo de la enfermedad en población mexicana. Con esto, se enfatiza la necesidad de una mayor representatividad de las diferentes poblaciones en los esfuerzos de secuenciación a nivel mundial para evitar limitar la aplicabilidad de la investigación que se genere de la enfermedad a un grupo concreto de personas.

Finalmente, el resultado más valioso del trabajo es la relación que se ha establecido entre *PABPC3* y los TCGT, junto con el panel de variantes candidatas propuesto, el cual brinda una guía sólida para futuras investigaciones en este campo. El panel debe ser validado en cohortes TCGT más grandes para determinar la eficacia de las variantes como biomarcadores de riesgo en la enfermedad y, en última instancia, su potencial implementación en la práctica clínica dentro de nuestra población.

ABSTRACT

Testicular germ cell tumors (TCGT) account for approximately 1-5% of all malignant tumors in men and are more common in young men aged 15 to 44. The incidence and mortality patterns of TCGT vary geographically. However, the lack of a significant reduction in mortality rates in Mexico and Latin America over the last decade, in contrast to trends observed in other regions of the world, raises the possibility that there may be molecular features within our population that influence the development and clinical evolution of TCGTs.

Evidence of distinct germline alterations in TCGT is limited to a few familial cases and genome-wide association studies (GWAS), which have identified 78 susceptibility *loci* associated with TCGTs. These studies are overrepresented by patients of European and North American Caucasian origin, and thus, genomic characteristics in our population have not been properly characterized.

The aim of this study was to identify germline variants associated with the risk of developing TCGTs through the analysis of whole-exome sequencing (WES) data from 40 patients at the National Cancer Institute. Initially, a global genetic ancestry analysis was conducted to select individuals with the ancestral composition of the Mexican mestizo population. Subsequently, a total of 2778 rare genetic variants [allele frequency (AF) in the global population <1%] were selected, including variants classified as "pathogenic" or "likely pathogenic" by ClinVar, as well as variants with uncertain clinical significance that could potentially disrupt gene function according to CADD and REVEL algorithms.

The frequency of these variants in the TCGT cases group was compared with two independent control groups: the first consisting of 36 Mexican individuals residing in Mexico City and the Metropolitan area, while the second was based on data from ~17,000 Latinos in the gnomAD database. Through these analyses, a panel of candidate variants was proposed, which showed the most significant p-values in the two independent association tests. This panel includes the variants rs1044301916 (T>A) in *PABPC3*, rs2273499 (G>A) in *ARFGAP1*, rs147153778 (C>T) in *SLCO4A1*, rs79783591 (A>T) in *MC4R*, and rs1234654104 (TA>T) in *EHTM1*. These variants are suggested as potential biomarkers for TCGT risk. Simultaneously, a gene-based multiple variant aggregation association test was performed, revealing a significant association of the *PABPC3* gene, which is specifically expressed in the testis, with TCGTs under a dominant model ($p < 2.5 \times 10^{-6}$).

In summary, this study identified variants not previously described as TCGT risk *loci*, potentially relevant to disease development in the Mexican population. This emphasizes the need for greater representativeness of different populations in worldwide sequencing efforts to avoid limiting the applicability of disease research to a specific group of people.

Ultimately, the most valuable outcome of this work is the established relationship between *PABPC3* and TCGTs, along with the proposed panel of candidate variants, providing a robust guide for future research in this field. The panel should be validated in larger TCGT cohorts to determine the efficacy of the variants as risk biomarkers in the disease and, ultimately, their potential implementation in clinical practice within our population.

1. INTRODUCCIÓN

El cáncer de testículo (TCa) es la neoplasia maligna con mayor frecuencia en los varones jóvenes entre los 15 y los 44 años. Sin embargo, la tasa de incidencia ha aumentado de manera global en las últimas dos décadas, con los mayores índices registrados en países del norte de Europa (Gurney et al., 2019). De todos los casos de TCa, el 95-98% corresponden a tumores de células germinales de testículo (TCGT) y el 2-5% restante corresponde a tumores de testículo del estroma gonadal y del cordón sexual que surgen de los tejidos de soporte del testículo, incluidas las células de Leydig y Sertoli (Cheng et al., 2018).

Todos los TCGT representan el desarrollo aberrante de las células germinales del testículo en alguna de sus etapas de maduración hacia la espermatogénesis y se caracterizan por una amplia heterogeneidad histológica explicada por la capacidad pluripotente de las células precursoras y las aberraciones somáticas que fueron ganadas en los tumores (Cheng et al., 2018; Rajpert-De Meyts et al., 2016).

Para comprender a fondo el panorama clínico y genómico de los TCGT es esencial contextualizar su prevalencia y variabilidad histológica. Estos tumores se dividen histológicamente en seminomas puros (SE) y no seminoma (NS), cada uno con características distintivas y comportamientos clínicos específicos.

Los SE están compuestos por una población homogénea de células neoplásicas que se asemejan a los gonocitos o a las células germinales primordiales (PGC). Los NS son tumores con células que muestran patrones variables de diferenciación, dando lugar a cuatro subtipos tumorales: carcinoma embrionario, tumor de saco vitelino, teratoma y coriocarcinoma. Presentan un comportamiento clínico más agresivo, desarrollando frecuentemente metástasis a distancia (Buljubašić et al., 2018).

Cada histología se presenta en alrededor del 50% de los casos, pero hay una tendencia a observar un mayor número de seminomas (50-60%). Los SE son diagnosticados típicamente alrededor de los 35 años y los NS alrededor de los 25 años (Costa et al., 2017).

Esta distinción histológica no solo tiene implicaciones en el tratamiento y pronóstico de cada uno de los pacientes, sino que también arroja luz sobre los posibles factores que podrían contribuir a su desarrollo.

1.1. Panorama clínico y genómico de los TCGT

1.1.1. Predisposición y factores de riesgo

La etiología del TCa sigue siendo un tema de estudio, pero las diferencias de su prevalencia en distintas poblaciones sugieren que factores tanto ambientales como genéticos pueden influir en su desarrollo. Los factores de riesgo mejor descritos incluyen la criptorquidia (descenso incompleto de uno o ambos testículos), antecedentes familiares de TCa, hipospadias, infertilidad, disgenesia gonadal y el diagnóstico previo de TCa (Buljubašić et al., 2018; Rajpert-De Meyts et al., 2016). La criptorquidia, por ejemplo, confiere un aumento de riesgo de 4 a 6 veces cuando se diagnostica (Trabert et al., 2013).

Las hormonas esteroideas sexuales tienen un papel crucial en el desarrollo urogenital, por lo tanto, las alteraciones hormonales en el útero, incluida la insuficiencia de andrógenos fetales se asocian a un mayor riesgo a trastornos de diferenciación sexual y a TCGT (Hersmus et al., 2017; Singh et al., 2021). Se ha planteado que la exposición a disruptores endocrinos y ciertas sustancias químicas industriales en la vida adulta también podrían contribuir a la enfermedad (Bhartiya & Kaushik, 2021; Fénichel & Chevalier, 2019; Sharma et al., 2020). A pesar de esto, se requiere más investigación

para comprender completamente la influencia de los factores ambientales en la patogénesis de la enfermedad (De Toni et al., 2019).

Los estudios de gemelos y familiares proporcionaron la evidencia más sólida de la predisposición genética hereditaria en los TCGT. En poblaciones de ascendencia europea, investigaciones retrospectivas han indicado que las personas cuyo padre ha tenido un diagnóstico de cáncer testicular tienen entre 4 y 6 veces más posibilidades de desarrollar la enfermedad en comparación con aquellos sin ese antecedente. Si un hermano también ha tenido este tipo de cáncer, el riesgo se incrementa de 8 a 10 veces (Hemminki & Li, 2004). Un análisis poblacional de la base de datos sueca de cáncer familiar publicado en 2015 ha estimado que la heredabilidad del TCGT es del 48,9%, de la cual, el 37% se espera que se explique por todas las variantes de un solo nucleótido (SNVs) (Litchfield, Thomsen, et al., 2015).

Los estudios de asociación del genoma completo (GWAS) y de secuenciación de exoma completo (WES) han sido efectivos para identificar variantes comunes asociadas al riesgo de TCGT. Actualmente se han identificado 78 *loci* independientes asociados a la susceptibilidad a TCGT. Ninguno de estos alelos es de alta penetrancia, lo que significa que cada variante tiene una contribución modesta al riesgo a desarrollar los TCGT (Chung et al., 2013; Kristiansen et al., 2015; Litchfield et al., 2017; Litchfield, Sultana, et al., 2015; Pluta et al., 2021; Rapley et al., 2009; Ruark et al., 2013).

De las variantes reportadas, *KITLG* es el *locus* de mayor peso en términos de riesgo identificado hasta la fecha, con variantes que aumentan el riesgo de la enfermedad 2.5 veces en comparación con los alelos de referencia (Lakpour et al., 2021). Este gen codifica al ligando de del receptor del factor de crecimiento de células madre/mastocitos de tirosina quinasa unido a membrana (KIT), el cual desempeña funciones cruciales en la supervivencia celular, la proliferación y la apoptosis (Sheikh et al., 2022). Se han reportado variantes de un solo nucleótido en genes como *SPRY4*, *BAK1*, *GAB2*, *CLPTM1L*, *SPRED2* que también participan en la vía de señalización de KIT-KITLG, promoviendo que sea la ruta biológica más frecuentemente afectada (Litchfield et al., 2016).

Otro conjunto de genes implicados repetidamente en el riesgo son los pertenecientes a la determinación del sexo y el desarrollo de células germinales, con *loci* en *DAZL*, *DMRT1*, *ZFPM1* y *PRDM14*. Además, se han identificado variantes en genes de reparación de daño al DNA (incluidos *RAD51C* y *BRCA1*), ciclo celular (*GSPT1* y *CHEK2*), mantenimiento de telómeros (*TERT* y *ATFIP*) y organización del centrosoma/ensamblaje de microtúbulos (*TEX14*, *PMF1* y *CENPE*) (Batool et al., 2019; Litchfield et al., 2016; Nakhaei-Rad et al., 2023).

Interesantemente, uno de los modelos de puntuación de riesgo poligénico (PRS), que considera el efecto combinado de todos los SNVs conocidos de riesgo a TCGT, demostró que el 1% de los hombres con las mayores puntuaciones PRS tienen un riesgo de enfermedad 10.4 veces mayor que el resto de la población. En contraste, en otros modelos de cáncer como próstata u ovario, el 1% de los individuos con los mayores puntajes solo muestran un aumento de 2 a 5 veces en el riesgo a padecer la enfermedad (Litchfield, Summersgill, et al., 2015).

Debido a lo anterior, la evidencia respalda un modelo poligénico de susceptibilidad a los TCGT, donde el fenotipo de la enfermedad se explica por la herencia conjunta de múltiples variantes de riesgo, que interaccionan con los factores ambientales, que a su vez modulan cambios epigenéticos en las células (Singh et al., 2021; Woldu & Bagrodia, 2018).

1.1.2. Origen del tumor

Los tumores de células germinales de testículo SE y NS se originan a partir de una lesión no invasiva desarrollada *in útero* conocida como neoplasia de células germinales *in situ* (GCNIS). Esta lesión consiste en células que se asemejan a células germinales primordiales detenidas antes de su primer paso de maduración en gonocitos durante el proceso de espermatogénesis. El modelo de evolución tumoral propone que los precursores germinales sufren intentos prematuros de entrar en meiosis, lo que conduce a su poliploidización. No obstante, un evento de división celular multipolar sin disyunción resulta en células tetraploides genéticamente inestables (Boublikova et al., 2014).

Debido a su origen, las GCNIS tienen similitudes morfológicas con los gonocitos. Estas células expresan marcadores de pluripotencia como *OCT3/4* y *NANOG*, presentan hipometilación global de DNA y exhiben patrones de impronta característicos de las PGC (Boublikova et al., 2014; Cheng et al., 2018). Permanecen en un estado no diferenciado y latente hasta la pubertad, momento en el cual la estimulación hormonal exacerbada en la adolescencia desencadena eventos genéticos que permiten a las células transformadas adquirir capacidades proliferativas e invasivas.

Con la activación de la lesión premaligna en la adolescencia, la inestabilidad genómica conduce a la pérdida selectiva de grandes regiones en los brazos cromosómicos, manifestándose en la ganancia del isocromosoma 12, i(12p), que es la alteración genómica más distintiva en los TCGT. Se cree que eventos como la sobreexpresión de genes en la región 12p11.2-p12.1, como *K-RAS*, la pérdida de genes como *PTEN* y *P21*, así como la ganancia en la expresión de *MDM2*, son responsables de desencadenar el crecimiento invasivo (Shen et al., 2018), que tendrá como consecuencia la manifestación clínica de los TCGT (Baroni et al., 2019).

1.1.3. Perfil mutacional de los TCGT

Los TCGT se caracterizan por mostrar una carga de mutación tumoral (TMB) baja (0.5 mutaciones por Megabase) (González-Barrios et al., 2022) en contraste con otras neoplasias malignas como el melanoma (14.4/Mb) (Hodis et al., 2012) o cáncer colorrectal (9.9–11.6/Mb) (Salem et al., 2018). Este hecho ha limitado la identificación de variantes somáticas que podrían fungir como biomarcadores de diagnóstico (Singla et al., 2019).

La anomalía más frecuente en los TCGT es la presencia del isocromosoma 12, que se reporta en el 80% de los casos (Shen et al., 2018). Aunque no es una alteración necesaria para el desarrollo del tumor, es un evento impulsor del crecimiento invasivo (Buljubašić et al., 2018). Específicamente, en el brazo pequeño del cromosoma 12 se encuentran importantes elementos genéticos como el protooncogén *K-RAS*, así como genes relacionados con células madre, incluyendo *NANOG* y *STELLAR*. Estos genes experimentan una sobreexpresión debido a la presencia del i(12)p, contribuyendo aún más al proceso patológico.

Los TCGT son altamente aneuploides. Por ejemplo, los no seminomas suelen ser hipotriploides y los seminomas suelen ser hipertriploides. La ganancia recurrente de los cromosomas 7, 8, 21, 22 y X así como la pérdida del cromosoma Y, se notifican sistemáticamente con una frecuencia del 25 al 40 % en los pacientes con TCGT (Costa et al., 2017). La amplificación específica del gen *KIT* en la región 4q12 conduce a su sobreexpresión, constituyendo un mecanismo alternativo para la activación de la vía *KIT-KITLGn* (Litchfield, Summersgill, et al., 2015).

En términos de los genes más frecuentemente mutados, los genes *KIT*, *K-RAS*, y *N-RAS* son los más destacados, presentando mutaciones en el 18%, 14%, y 4% de los casos, respectivamente. Estas variantes son exclusivas de tumores con componentes seminomatosos (ya sea SE puro o NS mixto) (Loveday et al., 2020; Shen et al., 2018). *N-RAS* y *K-RAS* son protooncogenes que se

relacionan con la activación de múltiples vías, incluyendo Raf/MEK/ERK y PI3K (Hacioglu et al., 2017).

Además, se han identificado otras variantes somáticas en diversos genes como: *FGFR3*, *MDM2*, *BRAF*, *AKT1*, *PIK3CA*, *RPL5*, *RAC1*, *CDC27* y *NCOA3* (Bagrodia et al., 2016; González-Barrios et al., 2022; Loveday et al., 2020; Shen et al., 2018; Taylor-Weiner et al., 2016). Específicamente, las mutaciones con pérdida de función en *CDC27* son reconocidas como biomarcadores de mal pronóstico en diferentes tipos de cáncer (Kazemi-Sefat et al., 2021).

En contraste, los TCGT exhiben una notable ausencia de mutaciones en el gen *TP53*. Aunque el 97% de los casos secuenciados tienen la copia del gen de tipo silvestre, se sabe que la vía de señalización de p53 se ve afectada de manera indirecta debido a la sobreexpresión de inhibidores como *MDM2*, así como la influencia de microRNAs específicos, como miR-372 y miR-373, que también contribuyen a modular su actividad (Litchfield et al., 2016).

Los perfiles transcripcionales han identificado diferentes firmas genéticas para cada subtipo histológico. De manera general, los genes asociados a la espermatogénesis (*PRAME*, *MAGEA4* y *SPAG1*) y la señalización K-RAS son sobreexpresados en SE. Los genes reguladores como *DNMT3B* y *SOX2* son sobreexpresados en subtipos NS, además de que muestran un enriquecimiento en la vía Wnt y MYC (Biermann et al., 2007; Shen et al., 2018). Las muestras con componentes de teratoma tenían altas actividades en la vía mTOR y una baja expresión de miR-371a-3p, un marcador que está dramáticamente representado en el resto de los TCGT (Nicu et al., 2022; Shen et al., 2018).

En resumen, los tumores de células germinales de testículo tienen una base genética compleja que involucra tanto alteraciones germinales como somáticas. Las variantes germinales pueden aumentar la susceptibilidad a desarrollar la enfermedad, mientras que las mutaciones adquiridas y las alteraciones cromosómicas son factores somáticos que contribuyen a la transformación maligna de las células germinales durante la progresión de los TCGT.

A medida que profundizamos en las bases moleculares que subyacen en el desarrollo de estos tumores, se abren oportunidades para establecer conexiones más sólidas entre estos cambios genómicos específicos y las manifestaciones clínicas y sintomatología de la enfermedad. Este enfoque nos permitirá identificar biomarcadores útiles para el desarrollo de enfoques terapéuticos más precisos y efectivos.

1.1.4. Sintomatología de los TCGT

Los pacientes con cáncer de testículo pueden experimentar una variedad de síntomas, el agrandamiento de un testículo y un pequeño bulto o área de dureza son los primeros signos de cáncer de testículo. A menudo, estas masas palpables son indoloras y se descubren por el paciente o en exámenes médicos de rutina.

En casos avanzados los pacientes pueden experimentar dolor, sensación de peso o inflamación en el área testicular. Además de una acumulación súbita de líquido en el escroto y dolor sordo en la parte inferior del abdomen o la ingle. En casos muy raros se presenta sensibilidad en las mamas.

Cuando los TCGT se han propagado a través de la sangre o los ganglios linfáticos a otras partes de cuerpo como los pulmones, el pecho o el cerebro; las metástasis pueden causar síntomas adicionales según la ubicación afectada (American Society of Clinical Oncology (ASCO), 2021; Chovanec & Cheng, 2022).

1.1.5. Diagnóstico y tratamiento

La orquiectomía radical es el procedimiento terapéutico inicial tras el hallazgo de un tumor testicular en el examen físico y/o ecografía escrotal, debido a que la biopsia está contraindicada. El estudio histopatológico de la pieza es fundamental ya que dependerá de la clasificación el grupo de riesgo y el manejo de los pacientes (Lakpour et al., 2021).

El diagnóstico requiere la medición de los marcadores tumorales séricos alfafetoproteína (AFP), la gonadotropina coriónica humana (hCG) y la lactato deshidrogenasa (LDH) antes y después de la orquiectomía, además de la realización de un estudio de imágenes del tórax, el abdomen y la pelvis (Batool et al., 2019; Chovanec & Cheng, 2022).

Para el estadiaje de los pacientes se utiliza la escala de tumor, nodo, metástasis (TNM) de 2016. La enfermedad en estadio I se caracteriza como un tumor confinado al testículo y los niveles normales de marcadores tumorales séricos post-orquiectomía. En el Estadio II el cáncer se ha diseminado a los ganglios linfáticos retroperitoneales y en el Estadio III el cáncer se ha extendido a partes distantes del cuerpo (Brierley et al., 2017).

El Grupo Colaborativo Internacional sobre Cáncer de Células Germinales (IGCCCG) introdujo el modelo de estratificación de riesgo de los pacientes. De acuerdo con las características del estadio, niveles de marcadores séricos y el tipo histológico se clasificarán en tres grupos de riesgo: bueno, intermedio y pobre, con tasas de supervivencia a cinco años del 96 %, 89 % y 67 %, respectivamente, en los pacientes de tipo histológico no seminoma. Ningún paciente con histología seminoma se considera de pobre pronóstico (Beyer et al., 2021; Chovanec & Cheng, 2022; Gillissen et al., 2021).

La quimioterapia, que es indicada para la enfermedad avanzada o metastásica de estadios II-III, comprende tres agentes que dañan el DNA: cisplatino, bleomicina y etopósido (BEP). El cisplatino ejerce su efecto citotóxico mediante la unión covalente al DNA. Los aductos ocasionan torsión de la hebra doble del DNA y por consiguiente el rompimiento de cadena doble, lo que finalmente desencadenará la muerte celular por apoptosis (Ferry et al., 2000). Existe evidencia que sugiere que líneas celulares derivadas de TCGT, poseen deficiencias en su habilidad para reparar el daño ocasionado por los aductos de DNA-cisplatino, que explica parcialmente la buena respuesta de estos tumores a la terapia BEP (Revisado en Cuevas-Estrada et al., 2023; Országhová et al., 2022).

Debido a la alta sensibilidad de los TCGT a la terapia basada en cisplatino, el cáncer testicular es el tumor sólido más curable, con una tasa de supervivencia relativa total de 10 años de más del 95%. Incluso para los pacientes que recaen, aproximadamente el 50% de ellos lograrán una remisión completa con una quimioterapia multimodal de rescate. Una minoría de los pacientes estimada entre el 5 y 10 % de los pacientes con TCGT presentan un fenotipo de resistencia a la terapia basada en platino, lo que se traduce en una expectativa de vida es de pocos meses. La edad promedio de muerte de estos hombres es de 32 años (Litchfield et al., 2016; Lobo et al., 2020).

Estos tumores son clínicamente importantes porque la mayoría de ellos ocurren entre los 20 y 45 años, la edad más productiva de los hombres (Elzinga-Tinke et al., 2015). El éxito del tratamiento posibilita que incluso aquellos pacientes en fases avanzadas tengan una esperanza de vida extendida por varias décadas. No obstante, los efectos a largo plazo de la quimioterapia o la radioterapia pueden afectar potencialmente la calidad de vida de los pacientes. Se han reportado problemas de infertilidad, síndrome metabólico, el daño vascular, cardíaco, neurotoxicidad y susceptibilidad a desarrollar neoplasias secundarias (Cheng et al., 2018; Mir et al., 2016; Smith et al., 2018).

Con el propósito de sintetizar los aspectos más importantes descritos en esta sección, se ha elaborado la **Figura 1**, que traza de manera gráfica el panorama clínico-patológico de los TCGT a lo largo de una línea temporal. Esta representación abarca desde el origen del tumor hasta la descripción de las problemáticas de supervivencia a largo plazo observadas en los pacientes.

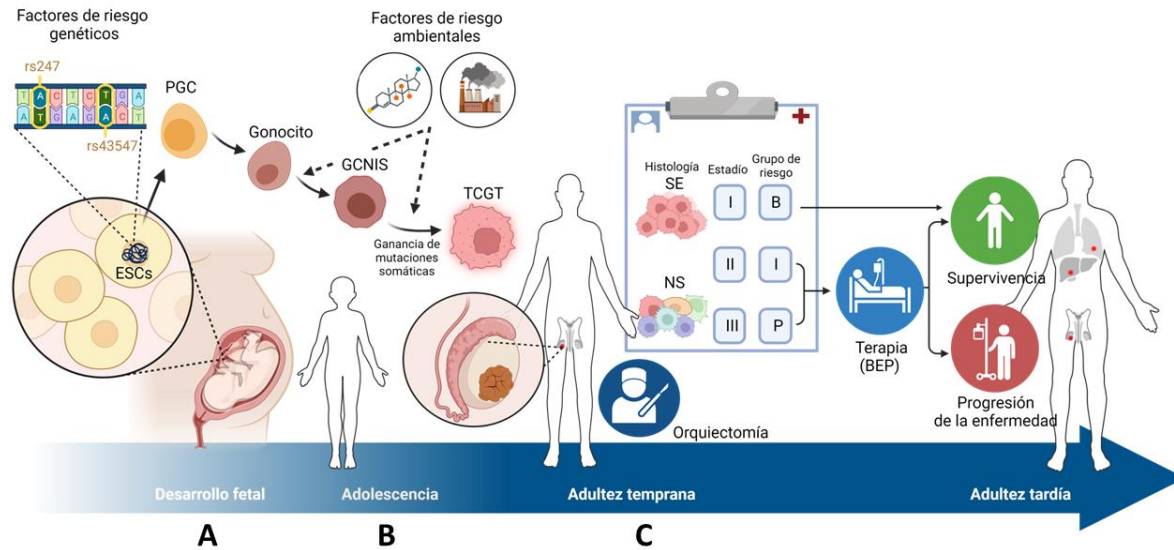


Figura 1. Características clínicas de los tumores de células germinales de testículo (TCGT) resumidas mediante un enfoque de línea de tiempo. **(A)** Durante el desarrollo fetal, las células germinales sufren un proceso de transformación influenciado por factores genéticos y del microambiente celular. Esta transformación puede dar lugar al desarrollo de la lesión preinvasiva conocida como neoplasia de células germinales *in situ* (GCNIS). **(B)** La estimulación hormonal característica de la etapa adolescente, así como la exposición a diferentes factores ambientales, hace que estas células ganen características proliferativas e invasivas que llevan al desarrollo y manifestación clínica de los TCGT. El tratamiento estándar para estos tumores es la orquiectomía radical, que es seguida por un análisis de la pieza tumoral junto con estudios de química sanguínea e imagen. **(C)** Estos procedimientos permiten determinar la histología del tumor, su estadio, el grupo de riesgo y el manejo clínico adecuado para cada paciente. En los casos de estadio II y III, se indica la quimioterapia con Bleomicina, etopósido y cisplatino (BEP). La alta sensibilidad de los TCGT a estas terapias es responsable de la notable tasa de supervivencia del 95% de los casos a 10 años. Lamentablemente una minoría de los pacientes son quimio-resistentes, para ellos, no hay segundas líneas de tratamientos efectivas y la esperanza de vida es baja. Los sobrevivientes de TCGT pueden experimentar efectos adversos a largo plazo, como hepatotoxicidad, infertilidad y un mayor riesgo de desarrollar neoplasias secundarias. ESCs: células madre embrionarias, PGC: células germinales primordiales, SE: seminoma, NS: no seminoma, B: Bueno, I: Intermedio, P: Pobre. Esta imagen fue generada con BioRender (<https://www.biorender.com/>).

1.2. Panorama epidemiológico del cáncer testicular.

1.2.1. El cáncer de testículo en México

En México, según el Instituto Nacional de Estadística Geografía e Información, en 2019 el 12% de las defunciones en el país estuvieron vinculadas a tumores malignos. Destaca que el cáncer de órganos genitales ocupa el segundo lugar en términos de morbilidad hospitalaria para la población de 20 años en adelante, tanto en población masculina como femenina (INEGI, 2021). Aunque el cáncer es un problema grave de salud pública no existen datos actualizados de la estadística de la incidencia nacional de los diferentes tumores malignos posterior a la última revisión del Registro Histopatológico de Neoplasias Malignas del 2003.

Hoy en día, la información relativa a la incidencia de cáncer a nivel nacional se basa en su mayoría en estudios retrospectivos que analizan la frecuencia de las distintas neoplasias en cohortes institucionales registradas durante periodos específicos. En este contexto, ejemplos como la Unidad Oncológica del Hospital Juárez, que utilizó datos del período 2006-2010 (Herrera et al., 2014), y el Hospital de Oncología del Centro Médico Nacional Siglo XXI, con registros del 2005 al 2012 (Martínez-Sánchez et al., 2013), identifican al cáncer de testículo como el cuarto tipo de cáncer más frecuente en varones.

Por otra parte, las investigaciones realizadas por el Instituto Nacional de Cancerología (INCan), basadas en datos proporcionados por Mohar et al., (1997) y de Rizo Ríos et al., (2007) señalan que desde 1985 hasta el 2004, el TCa fue el tumor de mayor incidencia en su población masculina.

Aunque los datos obtenidos de estos hospitales de alta especialidad con sede en la Ciudad de México pueden considerarse un reflejo cercano de la incidencia de cáncer en el país, ya que sus registros abarcan tanto a la población local como a la procedente de otros estados densamente poblados, es importante destacar que estos resultados aún subestiman la verdadera magnitud del problema. En efecto, esta aproximación no necesariamente abarca los casos de las etapas clínicas más tempranas que no son detectados, ni tampoco las etapas avanzadas que ya no son tratadas (Martínez-Sánchez et al., 2013). Además, es crucial considerar que estos registros no representan al sector de la población que no tiene acceso a los servicios de salud de estas instituciones debido a restricciones económicas que les impiden trasladarse desde comunidades remotas en el país hasta la Ciudad de México.

En un esfuerzo por obtener un panorama más representativo de la situación nacional, el grupo de trabajo en Cáncer Genitourinario de Jiménez-Ríos y colaboradores realizó un registro epidemiológico de cáncer genitourinario entre 2007-2009 y designó 5 grupos geográficos para organizar y recopilar la información: 1) Zona Centro, 2) Zona Norte-Noreste, 3) Zona Occidente-Bajío, 4) Zona Sur-Oriente y 5) Zona Noroeste-Golfo de Cortés. En todas las regiones consideradas, TCa se mantuvo consistentemente por debajo del cáncer de próstata en términos de frecuencia. No obstante, en la Zona Centro se observó la cifra más elevada de casos de TCa, abarcando un 25.2% del total de registros en su estudio retrospectivo (Jiménez-Ríos et al., 2011).

Una observación común a lo largo de estas publicaciones ha sido el incremento gradual en el número de nuevos casos de cáncer que afectan el sistema genitourinario con el transcurso de los años. Todo esto subraya la necesidad apremiante de contar con un registro nacional actualizado y representativo que brinde evidencia sobre el aumento en la incidencia del cáncer testicular a nivel nacional. Esto respaldaría la urgencia de implementar políticas públicas a nivel nacional destinadas a la detección temprana y el tratamiento de esta enfermedad.

Curiosamente, el dato más aproximado para la incidencia del TCa en México proviene de la base de datos GLOBOCAN (*Global Cancer Observatory*) de la Agencia Internacional de Investigación del Cáncer (IARC). Esta entidad estima la incidencia de cada tipo de neoplasia mediante modelos que utilizan las tasas de mortalidad nacionales de la Organización Mundial de la Salud y los registros de incidencia de cáncer en países vecinos. Según el informe del 2020, la IARC calculó una tasa de incidencia de cáncer testicular en México de 4.9 casos por cada 100,000 habitantes y estimó que esta enfermedad fue responsable de 670 fallecimientos (Sung et al., 2021).

1.2.2. El cáncer testicular en Latinoamérica

Los datos proporcionados por GLOBOCAN ofrecen la posibilidad de realizar una evaluación comparativa de la situación de nuestro país en el contexto global. Aunque los datos estimados para el año 2020 de los 20 países con la tasa de incidencia estandarizada por edad (ASIR) más alta por cada 100,000 habitantes (**Figura 2A**) no ubican a México dentro de esa lista, resulta que cuando medimos el parámetro de la tasa de mortalidad estandarizada por edad (ASDR), nuestro país ocupa el tercer lugar, justo por detrás de Argentina y Fiyi (**Figura 2B**). Una inspección más detenida de esta gráfica resalta el inquietante hecho de que de los 20 países enlistados, nueve son de América Latina.

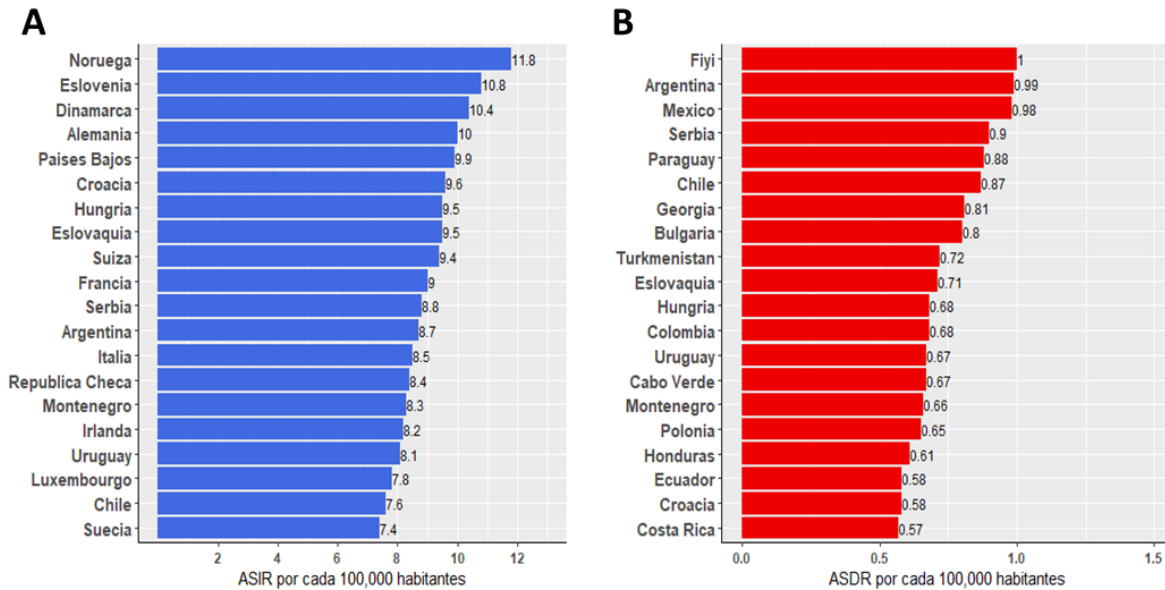


Figura 2. Países con los mayores índices de incidencia y mortalidad asociados al cáncer de testículo. **(A)** Tasa de incidencia estandarizada por edad (ASIR) por cada 100,000 habitantes. **(B)** Tasa de mortalidad estandarizada por edad (ASDR). Datos obtenidos de GLOBOCAN 2020.

La comparación de las tasas de incidencia y mortalidad entre los distintos continentes resalta que, a pesar de no ser la región de mayor incidencia, América Latina y el Caribe presentan el índice de mortalidad más alto (**Figura 3**). Las investigaciones que analizan las tendencias globales de estos indicadores a lo largo del tiempo concluyen que la incidencia mundial del cáncer de testículo ha aumentado gradualmente en las últimas cuatro décadas, afectando principalmente a las poblaciones de Europa occidental y del norte (Park et al., 2018). En contraste, la mortalidad en la mayoría de los países desarrollados ha disminuido; sin embargo, este patrón no es uniforme, ya que regiones como Latinoamérica y otras zonas de bajos ingresos, como África, no han mostrado cambios significativos en las tasas de mortalidad en el período reciente (Znaor et al., 2014). Incluso se menciona que Centroamérica y Sudamérica han experimentado los mayores incrementos en la tasa de mortalidad desde 1990 hasta 2016 (Cai et al., 2020).

Además de lo anterior, un análisis de la Base de datos Nacional de Cáncer, que consideró a 49,607 pacientes caucásicos no hispanos (81.5%) y a 6,724 pacientes hispanos (11.0%) diagnosticados con TCGT en Estados Unidos, reveló que la edad promedio al momento del diagnóstico de TCGT era cinco años menor en hombres hispanos en comparación con los hombres blancos. Por otro lado, los pacientes hispanos tenían una mayor probabilidad de presentar una histología no seminomatosa (Woldu et al., 2018).

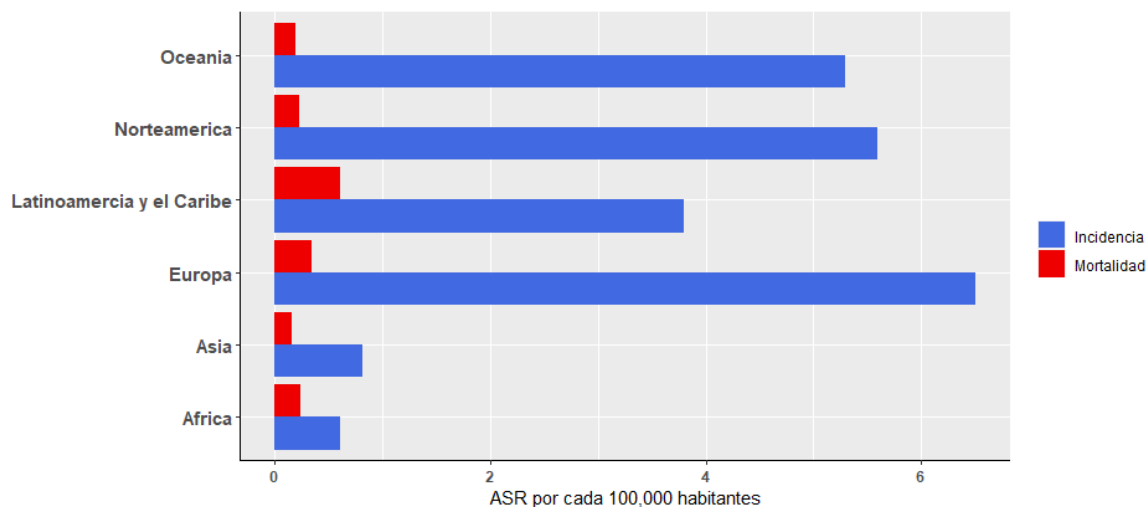


Figura 3. Tasa de incidencia y mortalidad estandarizada por edad (ASR) del cáncer testicular estimada en el 2020 para cada continente. Datos obtenidos de GLOBOCAN 2020.

La interpretación de esta estabilización en la tasa de mortalidad en Latinoamérica, en contraposición con la tendencia a la baja en el resto del mundo, presenta un desafío considerable. Aunque la explicación principal tiene una tendencia a centrarse en motivos socioeconómicos (Goss et al., 2013; Znaor et al., 2014) no se puede descartar la influencia de factores genéticos que podrían estar relacionados con la predisposición al TCGT y la respuesta al tratamiento. Las observaciones podrían sugerir que en Latinoamérica la incidencia del fenómeno de resistencia es más común que en otras regiones (González-Barrios et al., 2022), o bien, que los pacientes con TCGT tienden a presentar con mayor frecuencia la histología no seminomatosa, que se caracteriza por ser más agresiva.

En este contexto, los estudios de asociación de tipo casos-control se presentan como una herramienta valiosa para explorar en profundidad los factores genéticos que pueden estar contribuyendo a las diferencias observadas en la tasa de mortalidad y la respuesta al tratamiento en poblaciones latinoamericanas, aún más considerando el hecho de que a nivel global las iniciativas de secuenciación masiva en torno a la enfermedad tienen una muy poca representatividad de pacientes de origen latino.

1.3. Los estudios de asociación de tipo casos-control.

Un enfoque ampliamente utilizado en la investigación para explorar la posible relación entre variantes genéticas y la susceptibilidad a enfermedades es el estudio de asociación de tipo casos-control. En este diseño, se determina si existe una diferencia significativa en la frecuencia alélica de las variantes entre dos grupos seleccionados de una misma población: aquellos que presentan el fenotipo de estudio son los casos y aquellos con la ausencia del evento o la enfermedad, son los controles (Lazcano-Ponce et al., 2001).

El método más simple para evaluar la asociación de las variantes individualmente implica el uso de una tabla de contingencia y técnicas de regresión logística (Bansal et al., 2010). Si una variante genética aumenta la susceptibilidad de una persona a desarrollar una enfermedad, se espera que dicha variante sea más frecuente entre los individuos clasificados como casos que en el grupo de controles.

Este enfoque presupone que cualquier diferencia observada en las frecuencias alélicas entre grupos guarda una relación directa con el resultado medido (Cardon & Palmer, 2003). Sin embargo, la evidencia estadística de la asociación entre un alelo y un fenotipo puede explicarse a través de tres contextos diversos. En primer término, se establece una asociación directa cuando el alelo influye directamente en la expresión del fenotipo. En segundo lugar, se puede detectar una asociación indirecta, cuando la variante exhibe correlación o desequilibrio de ligamiento con el auténtico alelo causal del fenotipo, que ocupa una posición extremadamente cercana en el cromosoma. En tercer lugar, pueden generarse asociaciones espurias que se atribuyan a factores de confusión o sesgos en la selección de los individuos (Kiezun et al., 2012). Por ejemplo, las relaciones cripticas y la estratificación poblacional.

Las relaciones cripticas se refieren a parentescos o vínculos familiares entre individuos que no fueron reconocidos o documentados. Estas relaciones introducen un mayor nivel de complejidad a nivel genético, ya que pueden resultar en una mayor semejanza dentro de un grupo específico. Esto, a su vez, enmascara las diferencias reales en la frecuencia de las variantes entre los casos y controles, propiciando un aumento de asociaciones de tipo falso-positivo (Sul et al., 2018).

Po su parte, la estratificación poblacional se manifiesta como un fenómeno en el cual los casos y controles carecen de una subestructura genética similar. Esto se debe a que la población de la cual se muestrean los casos no coincide con la de los controles (Cardon & Palmer, 2003; Timpson et al., 2018). Cada población tiene una historia genética y social única, lo que resulta en patrones ancestrales de migración geográfica, cuellos de botella reproductivos y variación estocástica, que a su vez generan discrepancias en las frecuencias alélicas entre individuos de distintos orígenes étnicos. En este caso, las diferencias en las frecuencias alélicas entre los grupos no indican una asociación con la enfermedad, sino más bien con los antecedentes genéticos de la población (Cardon & Palmer, 2003; O'Connor et al., 2013; Zondervan et al., 2002).

Las poblaciones latinoamericanas exhiben diversos niveles de mestizaje, donde predominan los componentes parentales indígenas y europeos, con variaciones significativas entre distintas regiones. Un análisis que abarcó a 13 poblaciones mestizas de América Latina reveló que el componente indígena oscila entre el 71.9% y el 20.4%, mientras que el componente europeo varía entre el 25.2% y el 69.6%. Específicamente para los individuos de la Ciudad de México, el componente indígena promedio fue del 39.8% y el europeo del 56.8% (S. Wang et al., 2008).

En 2009, el Proyecto del Genoma Mexicano evaluó los datos de 300 individuos de seis estados distantes geográficamente (Sonora, Zacatecas, Guanajuato, Guerrero, Veracruz y Yucatán). Los resultados establecieron que la arquitectura genética de la población mestizo-mexicana era 55.2 % de ascendencia amerindia, 41.8 % europea, 1.0 % africana y 1.2% asiática. Donde la estimación de composición europea más alta fue de 61% para el grupo de Sonora (Silva-Zolezzi et al., 2009).

Estos hallazgos ofrecen una perspectiva inicial sobre la subestructura genética esperada en la población mestiza mexicana, que debe ser considerada en el diseño de los estudios de asociación, junto con el resto de los sesgos de selección. En la actualidad, los enfoques masivos como la secuenciación de exoma completo generan una visión detallada de la variación genética de los individuos puestos a prueba, con los datos suficientes para identificar esta complejidad e incluirla en los métodos de análisis de casos y controles.

1.4. Secuenciación de exomas completos.

La secuenciación de exoma completo ha acelerado los avances en la genética médica, abriendo la puerta a la posibilidad de tomar decisiones orientadas a la medicina personalizada en el ámbito clínico. Mediante esta tecnología, se determina la secuencia de todas las regiones codificantes de

los genes conocidos (Federici & Soddu, 2020). Y, dado que se estima que el 85% de las mutaciones causantes de enfermedades se localizan en regiones codificantes y funcionales del genoma, se espera que los datos generados por WES sigan teniendo un potencial significativo para revelar las causas subyacentes de una amplia gama de enfermedades (Rabbani et al., 2014).

Básicamente, el procedimiento comienza con la obtención del DNA de interés, que puede ser extraído de muestras de sangre o tejido. Luego, se realiza el proceso de captura, donde se aíslan y amplifican selectivamente los fragmentos específicos de la molécula que contienen los exones, preparándolos para su posterior secuenciación (Singleton, 2011). Finalmente, el equipo de secuenciación genera una vasta cantidad de datos de secuencias cortas conocidas como "lecturas".

La técnica de captura permite que los recursos de la secuenciación se concentren en las regiones de interés. Así pues, la cobertura de la secuenciación abarca aproximadamente el 95% de los exones totales del genoma (Rabbani et al., 2014). Además, cada base nucleotídica en los exones está representada en varias lecturas, así pues, una mayor profundidad en la secuencia aumenta la confiabilidad de los datos, ya que reduce la posibilidad de errores de secuenciación. La profundidad de secuenciación promedio que se emplea en estudios clínicos de WES es igual o mayor a 60X.

En la fase de análisis bioinformático las lecturas son alineadas a un genoma de referencia para reconstruir la secuencia original. Las diferencias identificadas entre la muestra y el genoma de referencia son almacenadas en un archivo de texto conocido como formato de llamado de variantes (VCF) donde se enlistan las variantes de un solo nucleótido, inserciones, deleciones y cambios en el número de copias (CNVs) encontradas. Las bases de datos pueden usar este archivo para anotar las diferentes características de cada variante, y con ello facilitar la identificación de mutaciones relevantes para el fenotipo estudiado.

En general, el gran interés en la secuenciación del exoma proviene de tres factores: el potencial para identificar muchos genes subyacentes a rasgos complejos, la anotación funcional directa de las variantes y un costo sustancialmente menor que el de la secuenciación del genoma completo (Kiezun et al., 2012).

1.4.1. WES en el análisis de variantes raras.

En la genética de rasgos complejos, la secuenciación del exoma se emplea para identificar variantes de codificación raras que no son detectables mediante estudios de asociación del genoma completo (GWAS), que se basan en microarreglos (Kiezun et al., 2012). Los GWAS están diseñados para detectar variantes comunes, las cuales tienen una frecuencia alélica (AF) que supera el 1% en la población. Aunque algunas de estas variantes pueden poseer un impacto funcional significativo, muchas de ellas tienen efectos neutros o sutiles en la función génica. Sin embargo, a pesar de la relevancia de las variantes comunes, estas no han logrado explicar una parte significativa de la heredabilidad de enfermedades complejas (Eichler et al., 2010; Manolio et al., 2009). Por esta razón, surge un interés cada vez mayor en explorar de manera sistemática la contribución de las variantes raras en el desarrollo de enfermedades (O'Connor et al., 2013).

Las variantes raras tienen una frecuencia poblacional baja, generalmente inferior al 1%, y presentan notables diferencias en comparación con las variantes comunes. En primer lugar, exhiben una menor tendencia al desequilibrio de ligamiento, son más específicas entre las distintas poblaciones y tienen un mayor impacto en la función y expresión génica (Momozawa & Mizukami, 2021). Para ilustrarlo mejor, la proporción de variantes no sinónimas es mayor entre las variantes raras que entre las comunes (Y. Li et al., 2010). Es importante mencionar que solo unas pocas variantes comunes previamente reportadas en la literatura las que han asociadas con alguna enfermedad directamente, y que presentan razones de probabilidades (OR) mayores a dos. En contraposición, la mayoría de

las variantes raras identificadas muestran un OR mayor a dos, con un promedio de OR de 3.74 (Asimit & Zeggini, 2010; Rabbani et al., 2014).

En los estudios de secuenciación de exoma se detectan entre 100 y 400 variantes no sinónimas raras en cada una de las muestras (Ioannidis et al., 2016; Tennesen et al., 2012). No obstante, debido al hecho de que las variantes raras son menos frecuentes, su identificación y estudio a menudo requieren cohortes grandes y enfoques específicos de análisis. Por ejemplo, las estrategias para abordar las variantes raras pueden aprovechar y mejorar la potencia estadística mediante la estratificación o selección de variantes basadas en su relevancia funcional. Esto puede incluir priorizar variantes sin sentido por encima de las variantes sinónimas (Kiezun et al., 2012). Otro enfoque podría ser la estratificación de genes candidatos según su expresión en un tejido relevante para la enfermedad de interés, o incluso basándose en la patogenicidad predicha de la variante (Kiezun et al., 2012; Uffelmann et al., 2021).

1.4.2. Clasificación de las variantes en función de su patogenicidad.

La clasificación de las variantes en los datos de secuenciación de exoma es de suma importancia para comprender mejor su relevancia clínica y su potencial impacto en la salud. Identificar las variantes que tienen el potencial de causar enfermedades puede orientar la toma de decisiones médicas y proporcionar un enfoque más preciso en el diagnóstico, pronóstico y tratamiento de enfermedades genéticas (Cannon et al., 2023).

El Colegio Estadounidense de Genética Médica y Genómica (ACMG) y la Asociación de Patología Molecular (AMP) han definido estándares para la clasificación de variantes de secuencia, basado en la interpretación de distintos tipos de evidencia incluidos los datos de la frecuencia alélica de las variantes en grandes poblaciones reportados en las bases de datos como gnomAD, predicciones computacionales, datos funcionales, análisis de segregación de la variante en una familia, etc (Richards et al., 2015).

Se recomendó el uso de cinco categorías para la clasificación de una variante: benigna, probablemente benigna (LB), variante de significado incierto (VUS), probablemente patogénica (LP) y patogénica. En la actualidad, ClinVar es la base de datos pública más empleada por la comunidad científica y clínica a nivel mundial para consultar la interpretación consenso sobre la asignación que ha recibido cada una de las variantes por los diferentes laboratorios que presentan evidencia (Henrie et al., 2018; MacArthur et al., 2014).

Una variante clasificada como patogénica indica que hay evidencia que respalda que la variante causa o contribuye a una condición médica, como el hecho de que la variante se encuentre en un gen en el que la pérdida de función es un mecanismo conocido de enfermedad, que la variante provoque el mismo cambio de aminoácido que una variante patogénica previamente establecida. En contraste, una asignación benigna indica que se ha acumulado evidencia que sugiere que la variante no tiene un efecto significativo en la salud y es poco probable que cause una condición médica. Esta evidencia puede incluir, por ejemplo, que se registre una AF >5% en las bases de datos poblacionales, la falta de segregación en miembros afectados de una familia, entre otros (Richards et al., 2015).

El uso del término "probable" debe restringirse a variantes en las que los datos respalden una probabilidad mayor al 90% de que la variante sea patogénica o una alta probabilidad de que sea benigno (Richards et al., 2015).

Finalmente, si una variante no cumple con los criterios utilizando cualquiera de estos conjuntos (patógenos o benignos), o si la evidencia de benigno y patógeno es contradictoria, el valor

predeterminado de la variante es VUS. La mayoría de las variantes encontradas por las tecnologías de secuenciación masiva son de significado incierto, porque la validación experimental para el gran número de variantes raras detectadas es inviable. Sin embargo, se han desarrollado numerosos algoritmos de predicción de patogenicidad *in silico* para mejorar su interpretación.

Estos algoritmos que incorporan ciertas líneas de evidencia para predecir el impacto de la variación en la función de la proteína, por ejemplo, la conservación evolutiva de secuencias entre especies, efectos estructurales y distancias fisicoquímicas, la frecuencia alélica, entre otros (Cannon et al., 2023; Moghadasi et al., 2016). Existen herramientas que toman las anotaciones de otros predictores para generar modelos se aprendizaje automático que permitan una clasificación propia.

Entre estas herramientas de conjunto encontramos a REVEL (*Rare Exome Variant Ensemble Learner*), que incorpora las predicciones de 13 predictores diferentes y asigna un puntaje entre 0 y 1, donde las puntuaciones más altas reflejan una mayor probabilidad de que la variante cause la enfermedad (Ioannidis et al., 2016).

Por su parte, CADD (*Combined Annotation Dependent Depletion*) usa más de 60 anotaciones para puntuar cualquier posible SNV que puede generarse en el genoma de referencia humano. Define cuáles de todas estas simulaciones son alelos que se han fijado en los humanos y supone que las variantes deletéreas son aquellas que desaparecieron con el proceso de selección natural. Una puntuación en escala Phed de 10 indica que se predice que la variante estará dentro del 10% de las sustituciones más perjudiciales en el genoma humano, con una puntuación de 20 se prevé que esté dentro del 1% de las variantes más perjudiciales, y así sucesivamente (Kircher et al., 2014; Rentzsch et al., 2019, 2021).

Dentro de las limitaciones más importantes que hay que considerar cuando se emplean estas herramientas es que las puntuaciones se generan por un análisis de las variantes de manera independiente, sin considerar como se pueden relacionar con otras variantes en el mismo gen o genes relacionados. Además, las bases de datos utilizadas por las herramientas deben actualizarse constantemente con nueva información genética y funcional para mantener su relevancia y precisión.

En definitiva, aun considerando sus limitaciones, los predictores de patogenicidad son herramientas útiles para la evaluación de variantes genéticas, sirviendo como guía en la selección de variantes candidatas para un análisis más detallado donde se contrasten datos experimentales ara una interpretación completa y precisa de la patogenicidad de cada mutación.

De manera específica, en este trabajo se usaron datos de WES de las muestras de sangre periférica de pacientes con TCGT para buscar alteraciones genéticas específicas que se asocien a la susceptibilidad a los TCGT. Al combinar la información generada por la secuenciación con la evaluación de la patogenicidad *in silico*, se obtiene una comprensión más completa de las variantes genéticas que desempeñan un papel crucial en el desarrollo y progresión de los TCGT. Este enfoque integrado permite la identificación de biomarcadores potenciales de riesgo que pueden contribuir a la mejora de los enfoques terapéuticos dirigidos y personalizados para los pacientes afectados por esta enfermedad.

2. ANTECEDENTES DIRECTOS

En nuestro Laboratorio de Carcinogénesis del Instituto Nacional de Cancerología se planteó en el 2018 un protocolo de investigación interno titulado "Identificación de firmas moleculares en la predicción de la respuesta al tratamiento con quimioterapia basada en platino en pacientes con tumores de células germinales de testículo (TCGT)" aprobado por el Comité de Investigación y el Comité de Ética en Investigación del INCan bajo los números (012/031/ICI) (CEI/783/12), respectivamente.

En este proyecto se generó un biobanco con las muestras de sangre periférica y piezas tumorales extraídas durante la orquiectomía radical de más de 250 pacientes con diagnóstico clínico-patológico de TCGT tratados entre 2008 y 2020 en el INCan. De manera paralela, se generó una base de datos con los archivos de secuenciación de exomas de la muestra tumoral y de sangre periférica de 40 individuos de la cohorte TCGT, seleccionados bajo estrictos criterios de inclusión y exclusión orientados a los objetivos del proyecto.

Los resultados del trabajo fueron publicados en 2022 y describen por primera vez las características genómicas de la resistencia al platino en la población mexicana. Los hallazgos de esta investigación confirmaron que los TCGT presentan una carga de mutaciones baja (0.5 mutaciones/Mb), en comparación con otros tumores sólidos como el melanoma y el cáncer de pulmón. Sin embargo, fue relevante observar que nuestros pacientes exhibieron un mayor TMB en comparación con la cifra internacionalmente reportada por el TCGA (*The Cancer Genome Atlas Program*) para los tumores de células germinales testiculares, que se sitúa en 0.31 mutaciones/Mb (**Figura 4**).

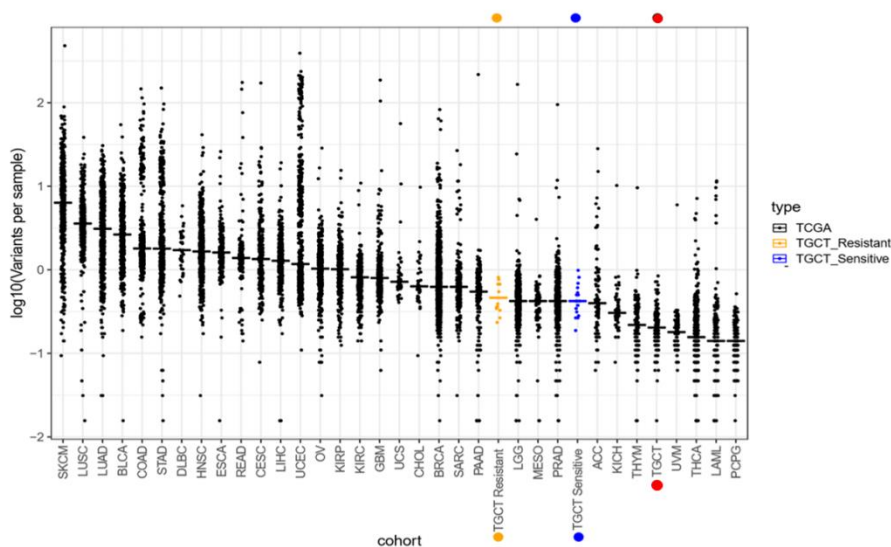


Figura 4. Comparación de la carga de mutaciones tumorales en la cohorte INCan-TCGT frente a las muestras de TCGA (*The Cancer Genome Atlas Program*). Frecuencias medias de variantes somáticas informadas en la secuenciación del exoma (líneas horizontales) en múltiples tipos de tumores informados en TCGA. Las frecuencias se reportan en mutaciones por megabase (Mb) y se disponen en orden descendente. En términos generales, la cohorte INCan-TCGT (puntos naranja y azul) ha demostrado (tanto muestras sensibles como resistentes al platino) que la tasa de mutación es mayor que la informada en TCGA para tumores TCGT (punto rojo). SKCM: Melanoma cutáneo de la piel, LUSC: Carcinoma de células escamosas de pulmón, LUAD: Adenocarcinoma de pulmón, BLCA: Carcinoma urotelial de vejiga, COAD: Adenocarcinoma de colon; STAD: Adenocarcinoma de estómago, DLBC: Linfoma difuso de células B grandes; HNSC: Carcinoma de células escamosas de cabeza y cuello, ESCA: Carcinoma de esófago, READ: Adenocarcinoma de recto, CESC: Carcinoma de células escamosas de cuello uterino y adenocarcinoma endocervical, LIHC: Carcinoma hepatocelular de hígado, UCEC: Carcinoma endometrial del cuerpo uterino, OV: Adenocarcinoma seroso de

ovario, KIRP: Sarcoma de células claras del riñón, KIRC: Carcinoma renal de células claras de riñón, GBM: Glioblastoma multiforme, UCS: Carcinosarcoma uterino, CHOL: Colangiocarcinoma, BRCA : Carcinoma invasivo de mama, SARC: Sarcoma, PAAD: Adenocarcinoma de páncreas, LGG: Glioma cerebral de grado inferior, MESO: Mesotelioma, PRAD: Adenocarcinoma de próstata, ACC: Carcinoma suprarrenocortical, KICH: Cromóforo renal, THYM: Timoma, UVM: Melanoma uveal, THCA: Carcinoma de tiroides, LAML: Leucemia mieloide aguda, PCPG: Feocromocitoma y paraganglioma. Modificado de González-Barrios et al., 2022.

El conjunto de mutaciones más frecuentes se localiza en los genes *COL6A3*, *NCOA3* y *TNR*, las cuales, hasta el momento no se habían reportado entre las bases de datos públicas. De manera interesante, nuestra cohorte no presentó variantes en el gen *KIT* y solo un 6% de los casos demostraron alteraciones en *K-RAS*. En general, la mayoría de las mutaciones parecían ser aleatorias y poco frecuentes en los diferentes grupos de respuesta. Esto sugirió que las mutaciones no eran la causa directa de la resistencia a la terapia, pero podrían conducir al riesgo de desarrollar este tipo de tumores.

Los pacientes con TCGT que evaluamos presentaron patrones más claros de variación en el número de copias y alteraciones estructurales. La cohorte exhibió una marcada inestabilidad genómica. Curiosamente, se reportó una frecuencia mayor de eventos de ganancia o pérdida en los pacientes sensibles en comparación con los resistentes, lo que contrasta con las poblaciones estudiadas previamente. Específicamente, la amplificación en el segmento 2q11.1 distinguía significativamente a los pacientes sensibles y resistentes al platino, posicionándose como un posible biomarcador de sensibilidad a la quimioterapia basada en platino en pacientes de origen mexicano.

En resumen, los resultados obtenidos de esta cohorte de pacientes con TCGT en el INCan apuntan a que las variantes somáticas más frecuentes de tipo SNV y CNV, difieren de las reportadas en cohortes europeas. Estos descubrimientos respaldan la noción de que existen diferencias genómicas sustanciales que caracterizan el desarrollo y la respuesta terapéutica a los TCGT en la población mexicana (González-Barrios et al., 2022). Por lo que basarse en los estudios realizados en poblaciones de origen europeo, no responde las razones de la incidencia y mortalidad que se observa en nuestra población y otras poblaciones latinas.

3. PLANTEAMIENTO DEL PROBLEMA

México es el tercer país con la mayor tasa de mortalidad asociada al cáncer de testículo en todo el mundo y Latinoamérica lidera la lista a nivel continental. La persistente falta de una reducción significativa en las tasas de mortalidad en México y Latinoamérica en la última década, en contraste con las tendencias observadas en otras regiones del mundo, sugiere la posibilidad de que en nuestra población existan particularidades a nivel molecular que ejerzan influencia en el desarrollo y la evolución clínica de los TCGT

Nuestra comprensión actual del componente genético como un factor de riesgo a los TCGT incluye la identificación de 78 variantes de riesgo bajo o moderado identificados en estudios de asociación del genoma completo. El problema radica en que la mayoría de los estudios que describen las variantes de susceptibilidad se han realizado en cohortes europeas o norteamericanas, lo que crea un vacío significativo en nuestra comprensión de cómo las características genómicas específicas de poblaciones latinoamericanas influyen en el riesgo de desarrollar los TCGT. Esta brecha se hace aún más aparente considerando que las variantes somáticas y las alteraciones estructurales más destacadas en los tumores de pacientes mexicanos exhiben variaciones en su frecuencia en comparación con los perfiles genómicos de otras poblaciones.

Por lo tanto, es imperativo determinar cuáles son las variantes germinales patogénicas más frecuentes entre los pacientes mexicanos, con el fin obtener una comprensión más completa de los factores genéticos que influyen en la susceptibilidad a los TCGT en esta población.

4. JUSTIFICACION

El proyecto anterior del laboratorio en el INCan (González-Barríos et al., 2022), generó una base de datos de secuenciación de exoma que abarca el mayor número de individuos mexicanos reportada hasta el momento. Al analizar los archivos correspondientes a las muestras de sangre periférica, se abre la posibilidad de detectar variantes de riesgo específicas de la población del instituto, lo cual marca el primer paso hacia la comprensión de las bases genéticas del cáncer testicular en los pacientes mexicanos.

Este valioso conocimiento podría guiar el desarrollo de paneles multigénicos diseñados para evaluar la susceptibilidad al cáncer testicular. Estos paneles representan herramientas esenciales para estratificar el riesgo de la enfermedad, elevar la precisión diagnóstica y permitir el enfoque personalizado en el manejo y tratamiento clínico de pacientes mexicanos que desarrollan tumores de células germinales de testículo.

5. HIPÓTESIS

En la población mexicana existen variantes genéticas raras en línea germinal que están asociadas al riesgo a desarrollar tumores de células germinales de testículo.

6. OBJETIVOS

6.1. Objetivo general

Identificar variantes germinales que pueden estar asociadas con un mayor riesgo de desarrollar tumores testiculares de células germinales mediante el análisis de datos de secuenciación de exoma de una cohorte de pacientes mexicanos del Instituto Nacional de Cancerología.

6.2. Objetivos particulares

- Delimitar la población del estudio de asociación a partir de la exclusión de individuos con una composición ancestral distinta a la prevista para la población mestizo-mexicana.
- Seleccionar y clasificar las variantes genéticas raras de interés para el desarrollo de TCGT, aplicando criterios de patogenicidad.
- Determinar si existe asociación estadística significativa de las variantes germinales seleccionadas con el desarrollo de TCGT.
- Proponer un panel de variantes germinales con una fuerte correlación con los TCGT, que sean promisorias como candidatas para su futura validación en cohortes más grandes de TCGT
- Reproducir los resultados de los datos de secuenciación de exomas en subconjuntos representativos de muestras de la cohorte TCGT a través de la técnica de RFLP (*Restriction Fragment Length Polymorphism*).

7. METODOLOGÍA

7.1. Diseño experimental:

Este estudio es de naturaleza comparativa, retrospectiva y transversal, adoptando un diseño de casos y controles. El proyecto se divide en tres etapas:

La primera etapa consistió en el escrutinio de los datos de secuenciación de exoma completo (WES, por sus siglas en inglés) de las muestras de sangre periférica de 40 pacientes con diagnóstico de TCGT del INCan generadas por el proyecto aprobado por el Comité de Investigación y el Comité de Ética en Investigación del INCan bajo los números (012/031/ICI) (CEI/783/12), respectivamente.

Durante esta fase, se identificó y priorizó un conjunto de variantes raras ($AF < 1\%$) en función de su significancia clínica y criterios de patogenicidad asociados a cada variante.

En la segunda etapa, se comparó la frecuencia alélica de las variantes previamente seleccionadas en relación con la frecuencia observada en dos grupos control independientes:

- A) Controles (Grupo 1): 36 voluntarios sanos residentes de la CDMX y el Área Metropolitana, para los que se tienen los datos de secuenciación de exoma a partir de muestras de sangre periférica.
- B) Controles (Grupo 2): ~17,000 individuos para los que se asigna una ascendencia latina según los datos de secuenciación de exoma en la base de datos pública gnomAD en su versión v2.1 non-cancer (<https://gnomad.broadinstitute.org/>).

Se generaron los análisis de asociación a nivel de variantes individuales y por agregación de múltiples variantes basada en genes. Como resultado de estos análisis, se propuso un panel compuesto por cinco variantes genéticas relevantes.

En la última etapa se seleccionaron tres de las variantes del panel y se validó mediante RFLP la presencia de las variantes detectadas por WES en una muestra representativa de los pacientes.

7.2. Población de estudio

7.2.1. Grupo CASOS

En este estudio se seleccionaron 40 pacientes con diagnóstico clínico-patológico de TCGT tratados entre el 2008 y 2018 en el Instituto Nacional de Cancerología de México (INCan). Todos los pacientes reportaron dentro de la historia clínica tener una nacionalidad mexicana.

Se utilizaron los siguientes criterios para su participación dentro del estudio:

Criterios de inclusión:

- Pacientes varones diagnosticados con TCGT seminomas y no seminomas.
- Mayores de 15 años y menores de 60 años.
- Pacientes que firmaron del consentimiento informado para la toma de la muestra de sangre periférica y la muestra fresca de tumor durante la orquiectomía radical.

Criterios de exclusión:

- Pacientes con tumores de células germinales extragonadales.
- Pacientes a los que se les haya realizado la orquiectomía radical fuera del INCan.

Criterios de eliminación

- Pacientes con expedientes que no contengan información como: edad, reporte de patología del tumor primario, día, mes y año del diagnóstico.
- Pacientes cuyos análisis de composición ancestral revele una estructura ancestral discordante con la esperada para la población mestizo-mexicana: Los dos componentes ancestrales principales deben ser el amerindio y el europeo. El componente ancestral europeo no puede ser mayor al 65%.
- Pacientes con un parentesco mayor a segundo grado ($IBS > 0.25$).

Los pacientes fueron estadificados según *el American Joint Committee on Cancer (AJCC)*, 8ª edición, y clasificados de los grupos pronósticos según *el International Collaborative Group of Germ-Cell Tumors (IGCCCG)*. El seguimiento y la reclasificación se actualizaron por última vez en enero de 2021. La investigación siguió los principios de la Declaración de Helsinki basada en la aprobación de la junta de revisión institucional del hospital. Todos los pacientes firmaron el consentimiento informado, en el caso de los individuos menores de edad se recabó la firma del consentimiento de los padres y el asentimiento del paciente.

7.2.2. Grupo CONTROL 1

El grupo control 1 corresponde a una población compuesta por 36 voluntarios sanos de nacionalidad mexicana, todos residentes de la Ciudad de México y el Área Metropolitana. Datos de la Dra. Reynoso, N. del Instituto Nacional de Cancerología (comunicación personal).

Criterios de inclusión:

- Mayores de 15 años y menores de 60 años.

Criterios de eliminación

- Individuos cuyos análisis de composición ancestral revele una estructura ancestral discordante con la esperada para la población mestizo-mexicana: Los dos componentes ancestrales principales deben ser el amerindio y el europeo. El componente ancestral europeo no puede ser mayor al 65%.
- Individuos con un parentesco mayor a segundo grado ($IBS > 0.25$).

7.3. Procesamiento de los datos de secuenciación de exoma completo

La secuenciación de las muestras de sangre periférica se realizó utilizando la plataforma HiSeq2500 Illumina, siguiendo el Protocolo Illumina Nimblegen V3 establecido en el Laboratorio de Genómica del Cáncer del Centro Oncológico MD Anderson de la Universidad de Texas. Las muestras se secuenciaron con una cobertura 60X (González-Barrios et al., 2022). El análisis bioinformático contempló las siguientes fases:

7.3.1. Control de calidad, alineamiento y llamado de variantes.

Para el análisis de control de calidad de las lecturas se utilizó FastQC v0.11.4. Las lecturas se alinearon al genoma de referencia humano hg38 con BWA-MEM (versión 0.7.12-r1039), después de esto los duplicados se eliminaron usando Picard (versión 2.18.13). Los puntajes base se recalibraron usando baseRecalibrator y ApplyBQSR dentro del paquete GenomeAnalysis.

El llamado de variantes germinales se realizó con los archivos BAM utilizando Strelka2 (versión 2.9.10) para obtener el conjunto final de archivos VCF para cada paciente y para los individuos del grupo control 1.

7.3.2. Análisis de ancestría genética global

Debido a que las disparidades en la estratificación poblacional en los grupos de estudio entre casos y controles pueden generar resultados de asociación falsos positivos con el fenotipo en estudio, se procedió a la exclusión de individuos con valores atípicos de composición ancestral que no se alinearan con los de la población de interés.

En líneas generales, se llevó a cabo una comparación de los datos de genotipos de los individuos pertenecientes a diversos grupos étnicos. Para este fin, se empleó el Panel de Diversidad del Genoma Humano (HGDP), el cual abarca a 938 individuos provenientes de 53 poblaciones alrededor del mundo, agrupadas por continentes: África, Europa, Oceanía, Oriente Medio, la región Centro y Sur de Asia, Asia oriental y América (J. Z. Li et al., 2008). Se usaron los datos que fueron procesados y alineados contra el genoma de referencia hg19 por (C. Wang et al., 2014). Las coordenadas genómicas fueron finalmente ajustadas a la versión del genoma de referencia hg38 usando la herramienta LiftOver de Picard (versión 2.18.13).

Los genotipos del HGDP se fusionaron con los de la población de estudio (casos y controles). Se removieron las variantes con una frecuencia alélica menor al 1%, con una desviación del equilibrio de H-W con un valor P menor a 0.001 y para aquellas variantes que no estuvieran genotipificados en el 90% de los individuos.

Para el cálculo de la ancestría global se usaron dos enfoques:

- *Estimación de ancestría global basada en máxima verosimilitud*, empleando el software ADMIXTURE (versión 1.3.0) (Alexander et al., 2009). Este programa realiza una búsqueda en la matriz de genotipos para identificar el número de poblaciones parentales que mejor se ajusta al valor especificado para el parámetro k. A partir de las variantes compartidas con estas poblaciones parentales, se calcula la proporción ancestral de origen para cada individuo presente en la matriz de genotipos. En este estudio, se optó por un análisis con un valor supuesto de k=4. Las poblaciones parentales asignadas a este valor de k corresponden a los componentes amerindio, europeo, africano y asiático.
- *Estimación de ancestría global por componentes principales* con el software TRACE 1.03.(C. Wang et al., 2015). A partir de la matriz de genotipos se calcularon componentes de escalado multidimensional. En este procedimiento se asignan distancias de N dimensiones entre los elementos de un conjunto. Una menor distancia implica una mayor semejanza entre los elementos. Se usaron 3 dimensiones para el cálculo de las distancias.

Los resultados para ambos enfoques fueron graficados con RStudio.

7.3.3. Análisis de parentesco

En los estudios de asociación con frecuencia se excluyen individuos que tengan un parentesco mayor al de segundo grado, debido a que las variables compartidas entre estos aumentan la

probabilidad de describir asociaciones de tipo falso-positivo, al aumentar la frecuencia de las variantes genéricas representadas en uno de los dos grupos de estudio.

Se usó el software PLINK v1.9 (Purcell et al., 2007) con la bandera (--genome) para generar un análisis basado en la identificación de *loci* idénticos por descendencia (IBD) y *loci* idénticos por estado (IBS). Esta herramienta estimó el coeficiente de parentesco (ϕ) en cada una de las posibles parejas incluidas en la cohorte de descubrimiento.

Entre dos individuos se pueden compartir dos tipos de alelos. Los IBD son aquellos *loci* descendientes de un alelo ancestral, por ejemplo, los hermanos comparten los alelos que heredaron de los padres. En dos individuos no emparentados se pueden encontrar alelos IBS debido a que dos eventos independientes dieron pie a la misma mutación. El coeficiente de parentesco habla de la probabilidad de que un alelo aleatorio en el mismo *locus* en otro individuo sea IBD (Thompson, 2013).

En este estudio se eliminaron las parejas con un coeficiente de parentesco fue mayor a 0.25 (un valor que describe la identidad por descendencia descrita entre padres e hijos, o entre hermanos).

7.4. Anotación de variantes y clasificación de variantes patogénicas

Después de la eliminación de las muestras según los resultados del ensayo de composición ancestral y parentesco. Se definió el número final de integrantes de cada grupo en la población de estudio y se procedió al análisis de los archivos VCF.

- Control de calidad: Dentro de los archivos VCF de cada paciente se eliminaron todas aquellas variantes con una baja profundidad ($DP < 10$) o una mala calidad de genotipo ($GQ < 20$).
- Anotación contra bases de datos: Las variantes que pasaron los controles de calidad fueron anotadas con ANNOVAR incorporando la información de las siguientes bases de datos: RefSeq, ClinVar (20210507), InterVar (20180118), REVEL (20191205), gnomad211_exome, gnomad30_genome, dbnsfp33a, CosmicV95.
- Categorización de variantes: Se removieron todas las variantes genéticas de los cromosomas sexuales. Se eliminaron aquellas mutaciones con una asignación funcional desconocida o sinónima en RefSeq, una frecuencia alélica $> 1\%$ en los registros de exomas/genomas de gnomAD y una asignación como “Benigna” o “Probablemente benigna” en ClinVar. Únicamente las variantes catalogadas como “Patogénicas” o “Probablemente patogénicas” en ClinVar e InterVar serán clasificadas como **variantes de nivel 1**. Para el resto de las variantes no clasificadas se removieron aquellas con una asignación de “Benigna” o “Probablemente benigna” para InterVar y se seleccionaron para el **nivel 2** aquellas con una puntuación de $CADD_phred > 25$ o un valor de $REVEL > 0.7$. Para todas las variantes se calculó la frecuencia alélica considerando que el número total de alelos es de 80.

7.5. Análisis estadístico de asociación

- *Asociación de tipo casos-control a nivel variantes individuales*. Dentro de la lista filtrada de variantes se definió el número total de muestras con el alelo alternativo, el número de homocigotos para el alelo alternativo y el número de heterocigotos en los casos y los controles. A partir de estos datos se construyeron tablas de contingencia 2x2 para estudiar

las diferencias en la distribución de la frecuencia en la que se contabiliza cada uno de los alelos.

Para el análisis de asociación se aplicó la prueba de Chi-cuadrada de Pearson, excepto cuando alguno de los valores en la tabla de conteos fuera menor a 5, en ese caso se aplicó la prueba exacta de Fisher de dos colas. La corrección de pruebas múltiples fue aplicada a todos los valores p registrados con el método de Holm.

La razón de probabilidades (OR) se calculó como una medida de la asociación de una variante con la presencia de TCGT, tomando siempre como referencia el alelo mayor o referencia. El paquete Epitools (versión 0.5-10) en R, fue el software empleado para esta tarea. Los intervalos de confianza (IC) del 95% se calcularon con la aproximación normal de Wald. Cuando cualquier valor en la tabla era cero, se aplicó el ajuste para muestras pequeñas (Oddsratio.small), que consiste en la corrección de Jewel.

- *Asociación de tipo casos-control por agregación de múltiples variantes basadas en genes (Gene-burden test).* Se realizó una prueba de agregación de múltiples variantes basadas en genes para identificar genes mutados con una frecuencia significativamente mayor en pacientes con TCGT en comparación con controles utilizando el paquete TRAPD (Test Rare vAriants with Public Data). El archivo VCF con los datos de los 125,748 exomas de gnomAD se anotó y filtró de la misma forma en la que se describió previamente para el archivo generado por la cohorte TCGT (Sección 7.4)

Para la posición de cada una de las variantes se verificó que dicha base tuviera una profundidad mayor a 10 en al menos el 90% de las muestras de ambos grupos de datos. Las posiciones que no cumplían con este criterio fueron excluidas del análisis. Un archivo que enumera las variantes calificadas para cada gen fue generado usando la función "make_snp_file.py" en TRAPD. Por cada gen se sumaron los conteos de todas las variantes tanto en los casos como en los controles (gnomAD_pop_amr). Se construyen tablas de contingencia de 2x2 para un modelo dominante y un modelo recesivo (M. H. Guo et al., 2018; M. H. H. Guo et al., 2016). Se calcularon los valores de p utilizando una prueba exacta de Fisher de una cola, y los valores de $p < 2.5 \times 10^{-6}$ son considerados como significativos (Corrección de Bonferroni basada en aproximadamente 20.000 genes). Los resultados de la prueba generada con la función "burden_test.R" de TRAPD fueron representados en una gráfica cuantil-cuantil

7.6. Selección de un panel de variantes

Se propuso un panel de cinco variantes germinales con una fuerte correlación con los TCGT, que sean promisorias como candidatas para su futura validación en cohortes institucionales de TCGT. Se priorizaron las variantes valores más pequeños de p en los análisis de asociación estadística a nivel individual.

7.7. Validación de los resultados de la secuenciación de exomas

7.7.1. Extracción de las muestras de DNA sanguíneo

Se extrajo el DNA genómico de 1.5 mL de sangre periférica que había sido almacenada en tubos etiquetados a -80°C en el biobanco de muestras TCGT del INCan bajo el resguardo del grupo de

investigación del Dr. Luis Herrera Montalvo. Se empleó un método basado en el uso de fenol-cloroformo para la separación de los ácidos nucleicos. La concentración final de todas las muestras de DNA genómico fue determinada usando el espectrofotómetro NanoDrop 2000 (Thermo Scientific). Para cada muestra 100 ng de DNA fueron separados electroforéticamente en un gel de agarosa al 1 % con bromuro de etidio (0.5 mg/mL) y en amortiguador TBE 0.5X para verificar su integridad.

7.7.2. Genotipificación de las muestras mediante la técnica de Restricción Fragment Length Polymorphisms (RFLP's).

Se seleccionaron 3 de las variantes para su validación en una muestra representativa de la cohorte TCGT. Se utilizó la herramienta bioinformática primerBLAST (Ye et al., 2012) para el diseño de un par de oligonucleótidos específicos que flanquearan la variante a interrogar por este método.

En la **Tabla 1** se describe la secuencia de los oligonucleótidos y su la temperatura de alineamiento (Tm), la longitud del producto de PCR generado con dicho para de oligonucleótidos, la enzima elegida para el corte del amplicón y el tamaño de los fragmentos de restricción para cada uno de los alelos.

Tabla 1. Estrategia para la genotipificación de las variantes de los genes *PABPC3*, *SLCO4A1* y *ARFGAP1* mediante la técnica de RFLPs.

	Nombre del oligo	Secuencia 5'-->3' del oligo	Tm (°C)	Producto PCR (pb)	Enzima	Fragmentos de restricción (pb)
1	PABPC3-FW-A	AGTTCGGGCCCGCCTTA	60	302	BspTI	Ref: 175, 113, 14 Alt: 288, 14
	PABPC3-RV-B	CGCTTCCGGAGACGTTTCAT	59			
2	SLCO4A1-FW-A	CTGAGTCCCTGGCTGAATCC	58	186	Cail	Ref: 186 Alt: 104,82
	SLCO4A1-RV-B	GCAGTGCAGTGAGAAGACGA	58			
3	ARFGAP1-FW-A	GGTTCATGCCAGCATCTTCG	58	153	FatI	Ref: 118, 31, 4 Alt: 60, 58, 31, 4
	ARFGAP1-RV-B	CCAAGTCCTCACCGAGTACAG	58			

Tm: temperatura de alineamiento, Ref: Alelo de Referencia, Alt: Alelo alternativo.

Para la reacción de PCR se utilizaron 40 ng de DNA genómico de sangre periférica, 0.2 mM de cada dNTP, 0.2 µM de cada primer, 1µl de buffer 10X sin MgCl₂ (Invitrogen), 1.5 mM de MgCl₂, 0.2 unidades de DNA polimerasa Taq Platinum (cat. 10966-030 Invitrogen) y agua estéril para completar un volumen de 10 µl. La reacción de PCR se llevó a cabo en un termociclador PCR Sprint thermal Cycler (Thermo Scientific) con el siguiente programa de amplificación: desnaturalización inicial de 2 minuto a 94°C seguida de 30 ciclos con 30 segundos de desnaturalización a 94°C, 30 segundos de alineamiento a 60°C y 30 segundos de extensión a 72°C, con dos minutos de extensión final a 72°C.

Para identificar cada una de las variantes, el producto de PCR fue incubado con una enzima de restricción que escinda de manera específica el alelo a interrogar (**Tabla 1**). La reacción de restricción estuvo compuesta por: 10 µl del producto de PCR, 10 U de la enzima específica (Thermo Scientific), 2 µl amortiguador 10X específico para cada enzima según las recomendaciones del fabricante (Thermo Scientific), un volumen final de 20 µL. Para cada muestra se tuvo como control negativo una reacción en ausencia de enzimas de restricción. Los tubos fueron incubados a una temperatura de 37°C durante 12 horas.

Posteriormente, los productos finales de digestión se sometieron a una electroforesis en un gel de agarosa al 2.5% y en amortiguador TBE 0.5X. Se realizó la comparación del tamaño del fragmento con el patrón de peso molecular 50 bp DNA Ladder (Norgen Biotek Corp). En la **Figura 5** se presenta

una simulación de un gel de agarosa con las reacciones de restricción para cada uno de los genotipos que podría presentar cada muestra generada con SnapGene Viewer.

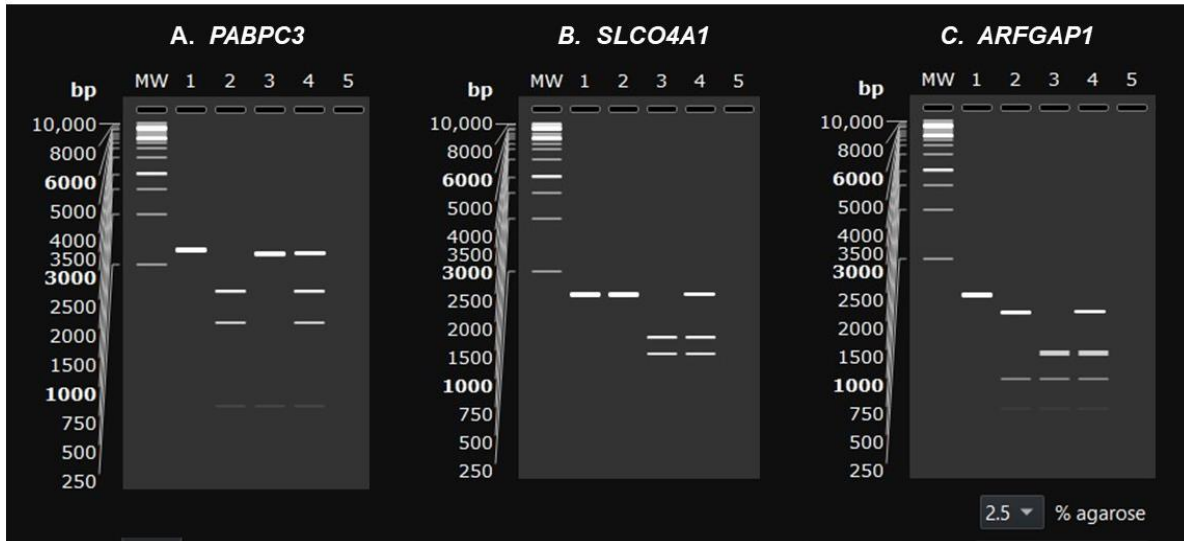


Figura 5. Simulación de un gel de agarosa con el corrimiento electroforético de las reacciones de digestión de los productos de PCR de un homocigoto a la referencia (carril1), un homocigoto al alelo alternativo (carril 3) o un heterocigoto (carril 4) para la variante de los genes *PABPC3* (A), *SLCO4A1* (B) y *ARFGAP1* (C). En el carril 1 de cada gel se muestra el control negativo de la reacción de digestión. El carril MW consiste en el marcador de peso GeneRuler™ 1kb Plus DNA Ladder (Thermo Scientific).

7.8. Análisis *in silico* de las proteínas de tipo silvestre y las mutantes.

Para las tres variantes se descargó la predicción de la proteína generada por la base de datos AlphaFold (<https://alphafold.ebi.ac.uk/>, consultada el 17 de noviembre de 2022), y se empleó el software UCSF Quimera X (<https://www.cgl.ucsf.edu/chimerax/>) para la visualización y edición de los archivos PBD. En cada estructura se señalan con colores los dominios funcionales reportados en la base de datos UniProt (<https://www.uniprot.org/>). Además, se utilizó la librería de rotámeros (Shapovalov & Dunbrack, 2011) para la edición de la secuencia de la proteína colocando el aminoácido que se cambia con la mutación. Esta herramienta de edición de la proteína permite estimar las posiciones más frecuentes en las que se encontraría el aminoácido sustituto dada la rotación de los enlaces en el aminoácido, así como los puentes de hidrógeno (señalados en azul) y las colisiones con los átomos cercanos (señalados en naranja) dentro de la estructura terciaria de la proteína. Se seleccionó para la edición el rotámero con la mayor prevalencia estimada y dentro de USC Quimera X se generaron los modelos de superficie para observar los cambios en el mapa de hidrofobicidad y de potencial electrostático.

8. RESULTADOS Y DISCUSIÓN

8.1. Estimación de la composición ancestral de la población de estudio

En los estudios de asociación, resulta de vital importancia asegurar que los grupos de casos y controles compartan una estructura genética similar. No obstante, describir la composición ancestral a nivel nacional es un proceso extremadamente complejo, debido a que los diversos procesos históricos en el país han favorecido el mestizaje. En el 2009 con el Proyecto del Genoma Mexicano se calculó que el acervo genético de la población mestizo-mexicana promedio era 55.2 % amerindio, 41.8 % europeo, 1.0 % africano y 1.2% asiático (Silva-Zolezzi et al., 2009). Sin embargo, esta composición presenta una variabilidad sustancial a nivel individual, con patrones marcados en diferentes regiones del país.

Todas las publicaciones coinciden en que el componente indígena y europeo son los mayoritarios en la población (Martínez-Cortés et al., 2012; Silva-Zolezzi et al., 2009; S. Wang et al., 2008). Esta es la razón fundamental por la cual esta premisa ha sido incorporada en los criterios de exclusión. Además, se estableció un umbral específico para que los individuos puedan integrar la población de estudio: el componente europeo no debe superar el 65%, un límite arbitrario que se ha considerado conservador sin perder de vista la variabilidad interindividual antes mencionada.

De esta manera se definió en este trabajo la estructura genética “esperada” para la población mestizo-mexicana. Y el primer objetivo fue asegurar que toda la población de estudio se adecuara a ella. En la **Figura 6** se muestran las proporciones ancestrales calculadas para cada individuo por el algoritmo de ADMIXTURE v1.3.0 (Alexander et al., 2009), asumiendo un modelo en el que cada persona puede pertenecer a un máximo de $k=4$ componentes ancestrales. Cada individuo se representa por una barra vertical dividida en k segmentos, en los cuales el color y la longitud representan la proporción ancestral calculada. La población parental $k=1$ son individuos amerindios (amarillo), $k=2$ son individuos africanos (rojo), para $k=3$ europeos (azul marino) y $k=4$ (azul cielo) individuos de Asia oriental del Panel de Diversidad del Genoma Humano (HGDP).

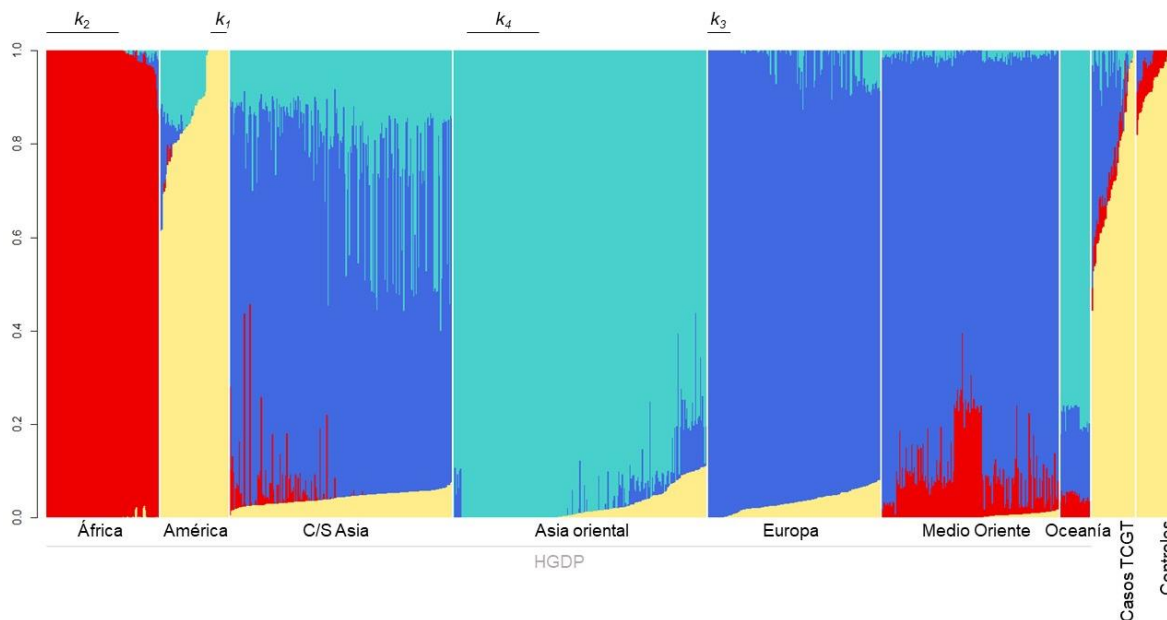


Figura 6. Gráfica de barras que representa las proporciones ancestrales estimadas para los casos TCGT y los controles sanos que integrarán el análisis de asociación junto con los individuos del Panel de Diversidad del Genoma Humano (HGDP). Las líneas superiores señalan a los sujetos asignados por el algoritmo de ADMIXTURE como las poblaciones parentales de referencia (k_1 , k_2 , k_3 , k_4).

Como un primer acercamiento, esta gráfica permite identificar que para los casos y controles el componente ancestral predominante es el amerindio, representado con color amarillo. En la **Tabla 2** se desglosan las proporciones ancestrales expresadas en porcentajes para cada uno de los individuos interrogados. Los pacientes con TCGT se enlistan con los identificadores “*Patient*” y los controles sanos con los identificadores “*MAPE*”.

Tabla 2. Proporciones ancestrales por individuo estimadas por ADMIXTURE con un valor de $k=4$

CASOS					CONTROLES_1				
ID_muestra	k_1 -AME (%)	k_2 -AFR (%)	k_3 -EUR (%)	k_4 -ASO (%)	ID_muestra	k_1 -AME (%)	k_2 -AFR (%)	k_3 -EUR (%)	k_4 -ASO (%)
Patient_01	72.17	2.77	17.16	7.89	MAPE_12	94.21	5.79	0.00	0.00
Patient_03	63.58	4.86	30.61	0.95	MAPE_13	97.56	2.44	0.00	0.00
Patient_04	75.89	4.62	15.21	4.28	MAPE_14	87.73	6.56	5.71	0.00
Patient_05	93.27	3.48	2.29	0.96	MAPE_15	94.96	5.03	0.00	0.00
Patient_07	44.22	4.98	47.62	3.18	MAPE_16	85.09	8.35	6.56	0.00
Patient_08	53.73	6.35	39.92	0.00	MAPE_17	81.83	9.36	8.81	0.00
Patient_09	52.76	5.34	37.30	4.60	MAPE_18	77.03	9.39	13.57	0.00
Patient_10	100.00	0.00	0.00	0.00	MAPE_19	90.50	5.76	3.74	0.00
Patient_11	86.77	4.02	8.59	0.62	MAPE_21	97.52	2.48	0.00	0.00
Patient_12	54.43	7.24	36.07	2.26	MAPE_22	96.68	3.31	0.00	0.00
Patient_13	90.22	5.07	4.71	0.00	MAPE_23	93.56	6.44	0.00	0.00
Patient_14	82.63	0.00	11.29	6.08	MAPE_24	86.60	6.97	6.43	0.00
Patient_15	72.02	7.30	20.68	0.00	MAPE_25	99.99	0.00	0.00	0.00
Patient_17	62.24	4.41	33.25	0.10	MAPE_27	99.99	0.00	0.00	0.00
Patient_18	60.82	7.81	22.75	8.63	MAPE_28	99.99	0.00	0.00	0.00
Patient_19	67.79	6.08	26.01	0.12	MAPE_30	91.83	7.30	0.87	0.00
Patient_20	73.19	3.55	23.27	0.00	MAPE_31	99.99	0.00	0.00	0.00
Patient_21	71.97	4.60	20.80	2.63	MAPE_33	99.99	0.00	0.00	0.00
Patient_22	82.23	7.41	2.14	8.22	MAPE_34	99.99	0.00	0.00	0.00
Patient_23	61.94	7.13	30.54	0.39	MAPE_35	86.55	6.33	7.12	0.00
Patient_24	97.29	0.00	0.00	2.70	MAPE_36	90.65	5.99	3.36	0.00
Patient_25	70.23	4.98	24.79	0.00	MAPE_37	94.40	5.60	0.00	0.00
Patient_26	71.18	2.07	22.02	4.73	MAPE_38	89.05	8.62	2.33	0.00
Patient_27	98.28	0.00	0.00	1.71	MAPE_39	95.32	4.68	0.00	0.00
Patient_28	59.39	8.88	31.73	0.00	MAPE_40	99.99	0.00	0.00	0.00
Patient_29	63.86	3.05	28.44	4.65	MAPE_41	94.37	5.62	0.00	0.00
Patient_30	41.62	8.08	50.30	0.00	MAPE_43	96.26	3.74	0.00	0.00
Patient_31	100.00	0.00	0.00	0.00	MAPE_44	89.70	8.80	1.50	0.00
Patient_32	68.33	4.11	22.03	5.52	MAPE_45	96.19	3.81	0.00	0.00
Patient_33	80.72	6.03	9.80	3.46	MAPE_46	91.07	6.61	2.32	0.00
Patient_34	60.99	5.41	21.66	11.94	MAPE_47	91.16	6.16	2.67	0.00
Patient_35	58.72	4.12	33.82	3.35	MAPE_48	90.27	7.14	2.60	0.00
Patient_37	79.26	5.70	13.08	1.96	MAPE_49	98.63	1.37	0.00	0.00
Patient_38	67.68	2.49	21.84	7.99	MAPE_50	91.00	5.79	3.21	0.00
Patient_39	66.89	3.90	21.42	7.79	MAPE_52	86.90	6.49	6.61	0.00
Patient_40	97.42	0.31	0.00	2.27	MAPE_53	96.18	3.82	0.00	0.00
Patient_41	58.61	5.14	33.33	2.92					
Patient_42	81.53	3.22	3.39	11.85					
Patient_43	96.10	0.00	0.00	3.90					
Patient_44	86.68	11.28	0.00	2.04					
Promedio ± DE	72.8 ±15.5	4.2 ±2.5	19.7±14.0	3.3 ±3.3	Promedio ± DE	93.1 ± 5.6	4.7 ± 2.9	2.2 ± 3.2	0.0

Componente AME= Amerindio, AFR=Africano, EUR=Europeo y ASO=Asiático. La escala de color resalta los componentes con las proporciones ancestrales predominantes.

El modelo mostró que la composición ancestral promedio de la cohorte TCGT (Casos) fue 72.8% amerindia, 19.7% europea, 4.2% africana y 3.3% asiática. Por su parte, la composición ancestral promedio para el grupo control fue 93.1% amerindia, 2.2% europea y 4.7% africana. Los valores

promedio que encontramos difieren con lo reportado por Silva-Zolezzi et al., (2009), posiblemente debido a que en ese artículo usó un panel de referencia distinto al HGDP, además de un algoritmo que emplea menos marcadores para generar las estimaciones de los componentes parentales que lo que requiere ADMIXTURE. Sin embargo, se pueden considerar que están entre lo esperado para nuestro tipo de población.

A nivel individual, se ha confirmado que, en la mayoría de la población de estudio, el componente amerindio constituye la proporción ancestral predominante. Es interesante notar que en los casos de los pacientes 07 y 30, el componente europeo se presenta en mayor proporción, aunque no excede el 65%, lo que evitó su exclusión del análisis.

En trabajos previos que exploran la estratificación poblacional en México, se ha reafirmado una tendencia: a medida que avanzamos hacia el norte del país, el componente europeo en los individuos tiende a ser más pronunciado, mientras que en la región sur prevalece la ascendencia indígena (Martínez-Cortés et al., 2012; Moreno & Sandoval, 2013; Silva-Zolezzi et al., 2009).

Dado que el INCAN es un hospital de referencia para la zona central de México, es importante considerar que algunos pacientes pueden proceder de distintos estados y no únicamente de la Ciudad de México. Esta variabilidad geográfica podría explicar las diferencias más amplias en la composición ancestral de los pacientes en comparación con los controles.

Adicionalmente, se llevó a cabo una evaluación de la estructura genética en la población de estudio utilizando el programa TRACE 1.03 (**Figura suplementaria 1, Anexo**). Este software genera un análisis de componentes principales, donde la distancia entre los puntos representa la semejanza o diferencia entre los individuos. Los resultados confirmaron que tanto los individuos en el grupo de casos como en el de controles tenían la distancia genética más pequeña con el grupo de referencia americano del HGDP. En un análisis tridimensional, parecían conformar un único grupo, en gran medida debido a la mayor cantidad de marcadores compartidos con este grupo poblacional en particular y no con el resto.

En resumen, se puede concluir que todos los individuos en la población de estudio de este proyecto presentan una estructura genética coherente con la esperada para la población mestizo-mexicana. Esto incluye a los pacientes 07 y 30, quienes no se consideraron valores atípicos ya que su proporción del componente europeo no superó el 65%.

8.2. Estimación del parentesco entre los individuos de la población de estudio.

Se generó una estimación de las relaciones de parentesco basado en la identificación de *loci* idénticos por descendencia (IBD) y *loci* idénticos por similitud (IBS) (Thompson, 2013) con el fin de encontrar pares de individuos que se parecen demasiado entre sí, es decir, más de lo que se esperaría por casualidad en una muestra aleatoria.

Un coeficiente de parentesco $\phi = 0,5$ se obtiene en pares de hermanos gemelos monocigóticos, $\phi = 0,25$ para padre e hijos o entre hermanos y $\phi = 0$ para individuos no relacionados.

La **Tabla 3** describe el coeficiente de parentesco calculado por PLINK para todas las posibles parejas que se generan de un individuo representativo de cada uno de los grupos de estudio (Casos y Controles). Se examinaron manualmente los coeficientes calculados para las 780 parejas del grupo de casos y las 630 parejas del grupo control (resultados no mostrados) y ningún par de individuos describió un valor de $\phi > 0,25$, que fue propuesto como criterio de eliminación.

Tabla 3. Estimación de la identidad por descendencia de dos individuos representativos de cada uno de los grupos del estudio.

CASOS			CONTROLES_1		
ID_muestra(A)	ID_muestra(B)	ϕ	ID_muestra(A)	ID_muestra(B)	ϕ
Patient_01	Patient_03	0.0228	MAPE_12	MAPE_13	0.0556
Patient_01	Patient_04	0.0143	MAPE_12	MAPE_14	0
Patient_01	Patient_05	0.0514	MAPE_12	MAPE_15	0.0086
Patient_01	Patient_07	0	MAPE_12	MAPE_16	0.0282
Patient_01	Patient_08	0	MAPE_12	MAPE_17	0.0175
Patient_01	Patient_09	0	MAPE_12	MAPE_18	0.0108
Patient_01	Patient_10	0	MAPE_12	MAPE_19	0.0818
Patient_01	Patient_11	0.0357	MAPE_12	MAPE_21	0.0624
Patient_01	Patient_12	0	MAPE_12	MAPE_22	0.0337
Patient_01	Patient_13	0	MAPE_12	MAPE_23	0.0782
Patient_01	Patient_14	0.0418	MAPE_12	MAPE_24	0
Patient_01	Patient_15	0.0093	MAPE_12	MAPE_25	0
Patient_01	Patient_17	0.012	MAPE_12	MAPE_27	0.0511
Patient_01	Patient_18	0.0138	MAPE_12	MAPE_28	0.0749
Patient_01	Patient_19	0.0419	MAPE_12	MAPE_30	0
Patient_01	Patient_20	0.0	MAPE_12	MAPE_31	0.0793
Patient_01	Patient_21	0.0141	MAPE_12	MAPE_33	0.0967
Patient_01	Patient_22	0.0274	MAPE_12	MAPE_34	0.071
Patient_01	Patient_23	0.0113	MAPE_12	MAPE_35	0
Patient_01	Patient_24	0.0	MAPE_12	MAPE_36	0
Patient_01	Patient_25	0.0191	MAPE_12	MAPE_37	0.0645
Patient_01	Patient_26	0.0172	MAPE_12	MAPE_38	0
Patient_01	Patient_27	0	MAPE_12	MAPE_39	0.0432
Patient_01	Patient_28	0	MAPE_12	MAPE_40	0.0675
Patient_01	Patient_29	0	MAPE_12	MAPE_41	0.0405
Patient_01	Patient_30	0	MAPE_12	MAPE_43	0.0515
Patient_01	Patient_31	0.0985	MAPE_12	MAPE_44	0
Patient_01	Patient_32	0.0083	MAPE_12	MAPE_45	0.0517
Patient_01	Patient_33	0	MAPE_12	MAPE_46	0
Patient_01	Patient_34	0.0728	MAPE_12	MAPE_47	0
Patient_01	Patient_35	0	MAPE_12	MAPE_48	0
Patient_01	Patient_37	0	MAPE_12	MAPE_49	0.0732
Patient_01	Patient_38	0.0329	MAPE_12	MAPE_50	0.0188
Patient_01	Patient_39	0.0224	MAPE_12	MAPE_52	0
Patient_01	Patient_40	0	MAPE_12	MAPE_53	0.018
Patient_01	Patient_41	0			
Patient_01	Patient_42	0			
Patient_01	Patient_43	0.0633			
Patient_01	Patient_44	0			

ϕ : Coeficiente de parentesco

Después de los análisis de composición ancestral y de parentesco, se concluyó que la población de estudio para el cálculo de las frecuencias alélicas de las variantes de riesgo en el estudio de asociación podía incluir a los 40 pacientes con TCGT para el grupo de casos y a los 36 individuos sanos para el grupo controles. En la **Tabla 4** se muestra un resumen de las características clínicas del grupo de casos. Por su parte, las características clínicas y demográficas disponibles para los individuos control son descritas en la **Tabla 5**. La observación más importante respecto a la cohorte de casos es que la mayoría de los pacientes tienen una histología de tipo no seminoma, por lo que las características moleculares distintivas de los seminomas pueden no estar representadas adecuadamente debido al tamaño de muestra. Para los casos y los controles más del 70% de la población se concentró en individuos jóvenes menores a 30 años.

Tabla 4. Características clínicas de los pacientes con tumores de células germinales de testículo

Categoría	Descripción	Número de Pacientes	(%)
Edad al momento del diagnóstico	<20 años	8	20
	20-30 años	22	55
	31-40 años	9	22.5
	>40	1	2.5
Grupo histológico	Seminoma	7	17.5
	No Seminoma	33	82.5
Subgrupo histológico	Seminoma puro	7	17.5
	Teratoma	4	10
	Carcinoma embrionario	7	17.5
	Tumor del saco vitelino	3	7.5
	Coriocarcinoma	0	0
	Mixto	19	47.5
Estadio clínico (AJCC)	I	2	5
	II	17	42.5
	III	21	52.5
Grupo de riesgo	Bueno	14	35
	Intermedio	7	17.5
	Pobre	17	42.5
	No evaluado	2	5
Respuesta al tratamiento	Respuesta completa	20	50
	Respuesta parcial	1	2.5
	Enfermedad estable	3	7.5
	Progresión de la enfermedad	12	30
	No evaluado	4	10
Estado vital	Vivo	23	57.5
	Muerto	9	22.5
	No evaluado	8	20
Índice de masa corporal	< 18.5	3	7.5
	18.5 a 24.9	19	47.5
	25 a 29.9	16	40
	> 30	2	5

Tabla 5. Características clínicas del grupo control 1

Categoría	Descripción	Número de Pacientes	(%)
Género	Femenino	7	19.4
	Masculino	29	80.6
Edad	<20 años	21	58.3
	20-30 años	9	25.0
	31-40 años	5	13.9
	>40	1	2.8
índice de masa corporal	< 18.5	2	5.6
	18.5 a 24.9	28	77.8
	25 a 29.9	5	13.9
	>30	1	2.8

8.3. Selección de las variantes germinales para el estudio de asociación.

En los estudios de asociación de tipo casos-control que emplean datos de secuenciación masiva, suele ser común adoptar estrategias con enfoques selectivos que permitan a los investigadores dirigir sus esfuerzos hacia las variantes más prometedoras en términos de su relevancia en la enfermedad. Esto implica centrarse en variantes que ya han sido previamente catalogadas como patogénicas, así como aquellas que se espera que tengan un impacto funcional en la expresión o función del gen. En nuestro estudio, los datos de WES de las muestras de DNA sanguíneo de la población de estudio fueron anotados contra diferentes bases de datos. Las variantes raras ($AF_{mundial} < 1\%$) fueron clasificados en dos niveles de acuerdo con su significancia clínica.

En el grupo de casos TCGT se seleccionaron un total de 2778 variantes genéticas (**Tabla 6**) distribuidas en 2385 genes, de estas, 103 variantes establecieron el Nivel 1 al ser catalogadas como Patogénicas o Probablemente patogénicas. Por su parte, 2685 variantes conformaron el Nivel 2, dichas mutaciones se distinguen por ser variantes de significado incierto (VUS) que según las predicciones de los algoritmos de REVEL o CADD se etiquetan como potencialmente dañinas en la función del producto génico.

Tabla 6. Resumen de las características de las variantes de interés presentes en la cohorte de TCGT

Categoría	Nivel 1	Nivel 2	Total
Número de variantes	103	2675	2778
Variantes observadas en ≥ 2 pacientes	6	176	182
Tipo de mutación			
SNV	89	2675	2764
inserción	6	0	6
delección	8	0	8
Sitio en donde se localiza la mutación (Func.refGene)			
Exón	96	2624	2720
Región intergénica	0	2	2
Intrón	1	2	3
Exón no codificante	0	5	5
Sitio de <i>splicing</i>	6	41	47
Río arriba del sitio de inicio de la transcripción	0	1	1
3'-UTR	0	1	1
5'-UTR	0	1	1
Consecuencia funcional (ExonicFunc.refGene)			
Delección que cambia el marco de lectura	6	0	6
Inserción que cambia el marco de lectura	5	0	5
SNV no sinónima	42	2421	2463
Ganancia de codón de paro	43	200	243
Frecuencia alélica en la cohorte TCGT			
AF <5%	101	2653	2754
AF $\geq 5\%$ y <10%	0	26	26
AF $\geq 10\%$	2	6	8

AF=Frecuencia alélica. SNV=variante de un solo nucleótido.

Al hablar de esta tabla, es importante destacar varios resultados. En primera instancia, el 96% (2685/2788) de las variantes totales están clasificadas en el Nivel 2. Actualmente las VUS representan el 51% del número de variantes totales (ClinVar versión 2023-07-02). Sin embargo, dado que hay una representación excesiva de participantes europeos en los proyectos que generan los datos, en las poblaciones poco estudiadas como la latinoamericana, es donde se acumula el mayor número de VUS (Federici & Soddu, 2020; Gudmundsson et al., 2022). Esta situación explica la importante cantidad de variantes en el Nivel 2, pero también enfatiza la necesidad de incluir una mayor diversidad poblacional en los proyectos genómicos globales. A medida que se recopilen los datos, es posible reclasificar algunas VUS y mejorar su interpretación y evaluación del riesgo en un entorno clínico.

En segundo lugar, entre todas las variantes clasificadas en este estudio, únicamente 182 de ellas fueron identificadas en al menos dos pacientes, mientras que el resto se registró en un solo individuo. Se estima que, en el exoma de un individuo, se detecten alrededor de 27 variantes novedosas que no se han reportado en las bases de datos (Gudmundsson et al., 2022). Sin embargo, la explicación de la baja observación de estas variantes radica en la selección de variantes raras, para las cuales una frecuencia alélica baja era anticipada (Momozawa & Mizukami, 2021; Ross et al., 2020), junto con la limitación del tamaño de la cohorte. En este sentido, existe la posibilidad de que asociaciones reales se hayan escapado a la detección debido a que la muestra no tuvo el alcance suficiente para identificarlas.

Finalmente, se pone de manifiesto una de las principales limitaciones del estudio, la cual se relaciona directamente con la naturaleza predominante de las variantes, que son principalmente SNVs (variantes de un solo nucleótido). Esta circunstancia está directamente relacionada con el hecho de que las herramientas bioinformáticas utilizadas, incluidas las de predicción de patogenicidad *in silico*, estén optimizadas para este tipo de mutaciones, pero enfrenten dificultades al analizar grandes inserciones/delecciones y variantes de número de copias (Hong et al., 2016; Moreno-Cabrera et al., 2020). Dado que los indels representan entre el 13 y 18% de la variación en el genoma humano (Cannon et al., 2023) no podemos descartar su relevancia biológica en la susceptibilidad a la enfermedad. Sin embargo, reconocemos la necesidad de adoptar enfoques de análisis más especializados para evaluar adecuadamente estas variantes.

Para continuar con la descripción de los resultados, se construyó un *oncoplot* que muestra el panorama de los genes con el mayor número de variantes de línea germinal y su distribución en los pacientes de la cohorte TCGT (**Figura 7**). Las frecuencias genotípicas y las anotaciones funcionales y descriptivas de las diferentes bases de datos que se asignaron a las variantes más representativas pueden ser consultadas en la **Tabla 7**.

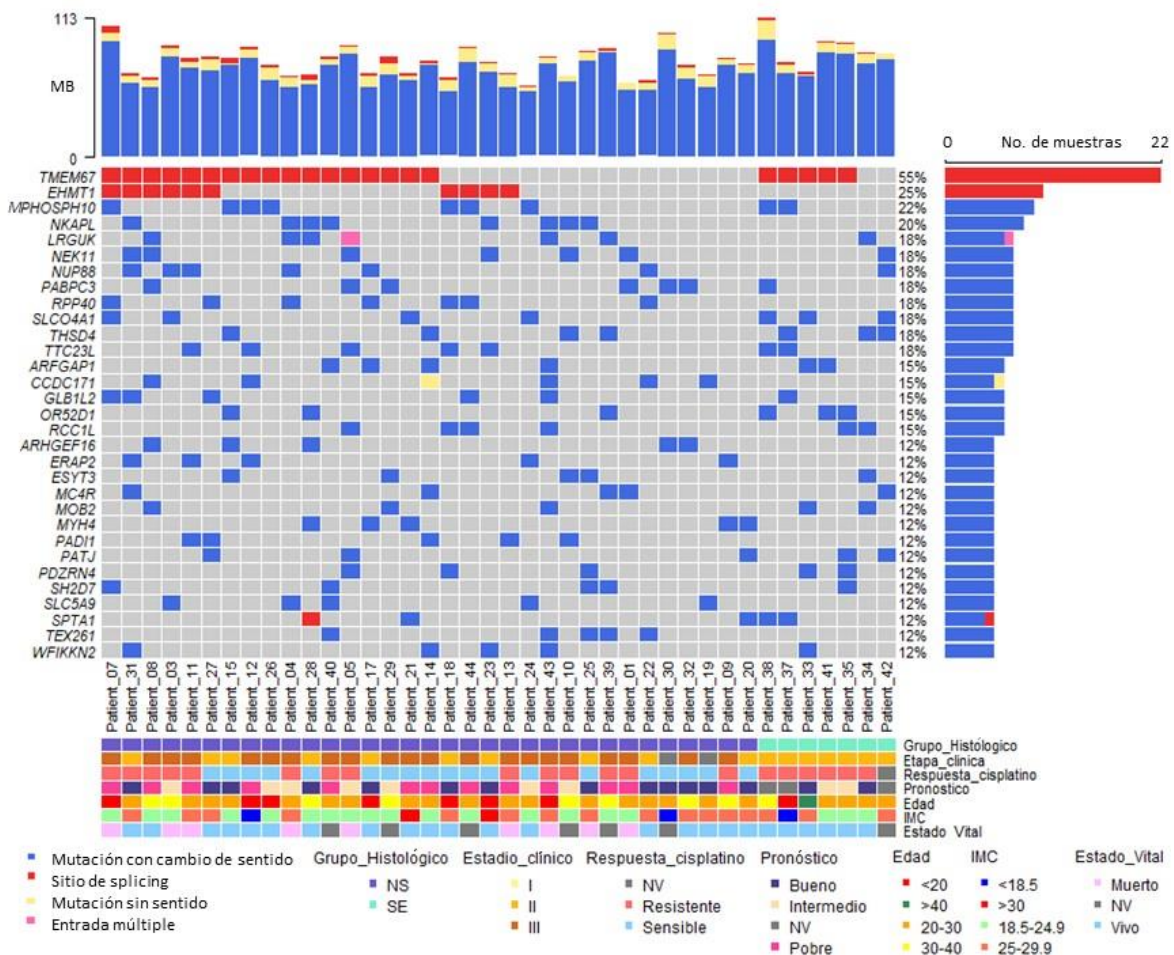


Figura 7. Genes con el mayor número de variantes germinales entre los pacientes con tumores de células germinales de testículo. En el panel central, cada columna representa una muestra y los genes con mutaciones se enlistan al lado derecho, con un color de bloque asociado para cada tipo de variante. El diagrama de barras superior tiene la frecuencia y el tipo de variantes para cada paciente, mientras que el diagrama de barras de la derecha tiene la frecuencia más alta de mutaciones en cada gen. Abreviaciones: NS: No seminoma, SE: Seminoma, NV: No evaluado, MB: Carga mutacional.

Tabla 7. Frecuencias genotípicas y descripción completa de las variantes genéticas detectadas en más de cuatro pacientes de la cohorte TCGT

Chr	Pos	Ref	Alt	ID	Gen	Tipo	Cambio_aa	AF	AF_popmax	ClinVar	InterVar	CADD	REVEL	Nivel	Hom_Ref	Het	Hom_Alt	
1	8	93755775	TA	T	rs386834190	<i>TMEM67</i>	splicing	-	-	-	LP	-	-	-	1	18 (45%)	22 (55.0%)	0 (0%)
2	9	137743368	TA	T	rs1234654104	<i>EHMT1</i>	splicing	-	1.00E-04	2.00E-04	LP	-	-	-	1	30 (75%)	10 (25%)	0 (0%)
3	2	71149884	G	A	rs4852764	<i>MPHOSPH10</i>	SNV_NS	A639T	0.0069	0.0687	-	VUS	25.5	0.121	2	31 (77.5%)	9 (22.5%)	0 (0%)
4	17	5405091	A	G	rs181249279	<i>NUP88</i>	SNV_NS	V337A	0.0062	0.0439	-	VUS	29.8	0.703	2	33 (82.5%)	6 (15%)	1 (2.5%)
5	6	28260431	C	G	rs61737340	<i>NKAPL</i>	SNV_NS	.L354V	0.0081	0.0554	-	VUS	26.2	0.28	2	28 (77.8%)	8 (22.2%)	0 (0%)
6	13	25097025	T	A	rs1044301916	<i>PABPC3</i>	SNV_NS	L276H	7.96E-06	0.0001	-	VUS	26.6	0.232	2	33 (82.5%)	7 (17.5%)	0 (0%)
7	20	63278918	G	A	rs2273499	<i>ARFGAP1</i>	SNV_NS	V131M	0.0036	0.0238	-	VUS	29.1	0.152	2	22 (78.6)	5 (17.9%)	1 (3.6%)
8	3	131029835	G	C	rs140058289	<i>NEK11</i>	SNV_NS	V43L	0.0069	0.0483	-	VUS	27.5	0.643	2	33 (82.5%)	7 (17.5%)	0 (0%)
9	5	34850284	G	A	rs77793555	<i>TTC23L</i>	SNV_NS	A128T	0.0099	0.0451	-	VUS	28.1	0.299	2	33 (82.5%)	7 (17.5%)	0 (0%)
10	6	4995271	T	C	rs144676135	<i>RPP40</i>	SNV_NS	Y277C	0.0049	0.0334	-	VUS	27	0.403	2	33 (82.5%)	7 (17.5%)	0 (0%)
11	11	5489310	G	A	rs190320444	<i>OR52D1</i>	SNV_NS	G202R	0.0096	0.0674	-	VUS	29.3	0.376	2	32 (84.2%)	6 (15.8%)	0 (0%)
12	15	71748496	G	A	rs12439065	<i>THSD4</i>	SNV_NS	D413N	0.0082	0.0524	-	VUS	32	0.405	2	30 (83.3%)	6 (16.7%)	0 (0%)
13	2	61185542	C	G	rs141666493	<i>AHSA2P</i>	ncRNA_exón	-	0.0076	0.0542	-	VUS	29.8	0.468	2	34 (85%)	6 (15%)	0 (0%)
14	20	62666451	C	T	rs147153778	<i>SLCO4A1</i>	SNV_NS	R450W	0.0017	0.0117	-	VUS	32	0.459	2	28 (82.5%)	6 (17.6%)	0 (0%)
15	7	134247616	C	T	rs140175129	<i>LRGUK</i>	SNV_NS	R682C	0.0069	0.0345	-	VUS	26.7	0.129	2	34 (85%)	6 (15%)	0 (0%)
16	7	75058653	C	T	rs971053877	<i>RCC1L</i>	SNV_NS	G302R	0.0045	0.0222	-	VUS	26.6	0.847	2	34 (85%)	6 (15%)	0 (0%)
17	1	17240654	G	A	rs140750531	<i>PADI1</i>	SNV_NS	R551H	0.0061	0.0297	-	VUS	31	0.491	2	34 (87.2%)	5 (12.8%)	0 (0%)
18	12	41552662	G	A	rs145855549	<i>PDZRN4</i>	SNV_NS	E146K	0.0089	0.0627	-	VUS	32	0.448	2	34 (87.2%)	5 (12.8%)	0 (0%)
19	15	78101557	A	G	rs186449076	<i>SH2D7</i>	SNV_NS	K435R	0.0052	0.04	-	VUS	25.2	0.119	2	29 (85.3%)	5 (14.7%)	0 (0%)
20	17	50839994	C	T	rs191998613	<i>WFIKK2</i>	SNV_NS	R143W	0.0059	0.0413	-	VUS	30	0.335	2	16 (80%)	3 (15%)	1 (5%)
21	18	60371544	A	T	rs79783591	<i>MC4R</i>	SNV_NS	I269N	0.001	0.0075	-	VUS	26.8	0.487	2	33 (86.8%)	5 (13.2%)	0 (0%)
22	2	70993734	T	C	rs151285112	<i>TEX261</i>	SNV_NS	T38A	0.009	0.0391	-	VUS	26.5	0.364	2	33 (86.8%)	5 (13.2%)	0 (0%)
23	9	15657182	C	G	rs141812748	<i>CCDC171</i>	SNV_NS	A293G	0.0052	0.0374	-	VUS	29.4	0.191	2	35 (87.5%)	5 (12.5%)	0 (0%)
24	1	3477900	G	T	rs190407416	<i>ARHGEF16</i>	SNV_NS	R500L	0.0029	0.0199	-	VUS	29.9	0.78	2	36 (90%)	4 (10%)	0 (0%)
25	11	1470197	G	A	rs200771049	<i>MOB2</i>	SNV_NS	A261V	0.0031	0.0194	-	VUS	25.4	0.21	2	35 (89.7%)	4 (10.3%)	0 (0%)
26	12	113315398	G	A	rs3764034	<i>SLC8B1</i>	SNV_NS	L302F	0.0047	0.0266	-	VUS	28.2	0.393	2	11 (73.3%)	4 (26.7%)	0 (0%)
27	15	81313315	G	A	rs138188091	<i>STARD5</i>	SNV_NS	R195C	0.0068	0.0499	-	VUS	29.1	0.428	2	35 (89.7%)	4 (10.3%)	0 (0%)
28	16	74410965	G	A	rs148861848	<i>CLEC18B</i>	SNV_NS	R272C	0.0086	0.0154	-	VUS	33	0.362	2	31 (88.6%)	4 (11.4%)	0 (0%)
29	2	174422999	C	G	rs34170959	<i>SCRN3</i>	SNV_NS	S283C	0.0032	0.023	-	VUS	26.1	0.1	2	36 (90%)	4 (10%)	0 (0%)
30	2	222639607	T	C	rs138484554	<i>FARSB</i>	SNV_NS	Q143R	0.0028	0.0203	-	VUS	28	0.656	2	36 (90%)	4 (10%)	0 (0%)
31	3	138474321	C	T	rs191463941	<i>ESYT3</i>	SNV_NS	R813W	0.0033	0.0244	-	VUS	34	0.481	2	34 (89.5%)	4 (10.5%)	0 (0%)
32	4	102798940	C	T	rs147920467	<i>UBE2D3</i>	SNV_NS	A146T	0.0017	0.0117	-	VUS	20.7	0.737	2	36 (90%)	4 (10%)	0 (0%)
33	5	73578811	A	G	rs151157740	<i>UTP15</i>	SNV_NS	K179E	0.003	0.0211	-	VUS	29.4	0.385	2	36 (90%)	4 (10%)	0 (0%)
34	5	7835432	G	A	rs76872483	<i>C5orf49</i>	SNV_SS	R72*	0.0062	0.0394	-	VUS	39	NA	2	36 (90%)	4 (10%)	0 (0%)

Chr: Cromosoma, Pos: Posición en el genoma, Ref: Alelo de referencia, Alt: Alelo alternativo, ID: Identificador de la variante en dbSNP, Tipo: Tipo de mutación, SNV_NS:SNV no sinónima, SNV_SS: SNV sin sentido, Cambio_aa: cambio en el aminoácido a nivel proteína, AF: Frecuencia alélica de las variantes derivada de la base de datos de gnomAD v2, AF_popmax: Frecuencia alélica máxima reportada en alguna de las poblaciones continentales (africano/afroamericano, asiático oriental, europeo, latino y sudasiático) de gnomAD v2.0, ClinVar: base de datos de asociación de fenotipos clínicos; LP: Probablemente patogénica, VUS: Variante de significado incierto, CADD: predicción de puntuación continua similar a phred por el algoritmo de CADD, REVEL: puntuación de patogenidad predicha por el algoritmo de REVEL, Hom_Ref: Homocigotos para el alelo de referencia, Het: Heterocigotos Hom_Alt: homocigotos para el alelo alternativo.

En el gráfico se puede observar que no existen mutaciones con alta penetrancia que se reporten en la mayoría de las muestras de la cohorte (**Figura 7**), lo cual es consistente con lo reportado en la literatura científica. En primer lugar, como mencioné anteriormente, los TCGT a nivel somático presentan un TMB bajo en comparación con otras neoplasias (González-Barrios et al., 2022). En segundo lugar, la naturaleza compleja de una enfermedad como los TCGT se traduce en un modelo de susceptibilidad altamente poligénico, donde cada variante contribuye de manera modesta al riesgo de la enfermedad (Litchfield et al., 2016). Por ende, es esperable que la mayoría de las mutaciones sean aleatorias y poco frecuentes. A modo de comparación, en un estudio reciente, una de las nuevas variantes de susceptibilidad identificada por el grupo de Paumard-Hernández et al., (2018) fue encontrada en el gen *PLEC* y solo se reportó en 25 de los 365 pacientes, lo que representa un 6.8% de los casos totales (Paumard-Hernández et al., 2018).

Los genes más frecuentemente mutados son *TMEM67* (22/40 (55%) muestras), un gen que está involucrado en la migración del centriolo a la membrana apical y en la formación del cilio primario (Yinsheng et al., 2022) y *EHMT1* (10/40 (25%) muestras) un gen que codifica para la metiltransferasa de la histona H3 en la lisina 9 que tiene un papel ampliamente estudiado en diferentes tipos de cáncer cuando su actividad o expresión se ve modificada (Nachiyappan et al., 2022; Saha & Muntean, 2021).

Dentro del *oncoplot* (**Figura 7**) se buscó correlacionar la presencia de las diferentes variantes a las características clínicas de los pacientes. La revisión solo destacó que las variantes en los genes *EHMT1*, *RPP40*, *NKAPL*, *CCDC171*, *ARHGEF16*, *ERAP2* y *MYH4*, se reportan únicamente en el grupo de pacientes con tumores no seminomas, sin embargo, debido al reducido número de muestras de pacientes con tumores seminomas, es difícil determinar si estas variantes son distintivas de esta histología. Además de esta observación, no se detectaron mutaciones únicas que pudieran asociarse con los grupos según la respuesta al tratamiento con cisplatino, el pronóstico o las etapas clínicas.

Interesantemente, cinco de los genes enlistados en la **Tabla 7** (*NKAPL*, *LRGUK*, *PABPC3*, *SLCO4A1* y *TEX261*) están involucrados en la espermatogénesis durante el desarrollo (AmiGO versión 2022-11-23, Aleksander et al., 2023; Ashburner et al., 2000; Carbon et al., 2009) y la desregulación de algunos de estos genes se asocia a la infertilidad masculina (Kumar & Singh, 2021; Ozturk et al., 2016). Estos hallazgos adquieren una importancia significativa, dado que tanto la espermatogénesis como la infertilidad masculina son procesos biológicos estrechamente ligados con el riesgo al tipo de cáncer examinado en este proyecto.

Las variantes de susceptibilidad previamente reportadas para TCGT (Litchfield et al., 2016; Pluta et al., 2021) no fueron encontradas en nuestro análisis debido a que las publicaciones en que se han reportado se enfocan principalmente en variantes comunes. Paralelamente, algunas de ellas están posicionadas en regiones no codificantes, que no pueden ser observadas con el enfoque WES. Sin embargo, algunos de los genes que albergan estas variantes de susceptibilidad, como *PIK3CD*, *RAD51C*, *CENPE*, *ARL14*, *GSPT1*, *PCNT*, *SMARCA1*, *USO1* y *ZFPF64*, fueron identificados en este análisis con variantes de baja frecuencia (≤ 2 pacientes).

En conjunto, estos resultados amplían nuestro entendimiento de las alteraciones genéticas subyacentes en los TCGT de una población poco explorada, como la mexicana, al identificar de manera novedosa genes que no habían sido relacionados con la susceptibilidad a esta enfermedad. Además, subrayan la importancia de extender nuestra visión más allá de las variantes comunes y también de las mutaciones codificantes, abriendo así nuevas perspectivas para investigaciones futuras y enfoques terapéuticos innovadores.

8.4. Asociación de tipo casos-control a nivel de variantes individuales.

Con el propósito de evaluar cuales de las variantes germinales detectadas en la cohorte de casos del INCan muestran los valores de mayor correlación con el fenotipo de estudio. La frecuencia alélica de las 2778 variantes clasificadas en los Niveles 1 y 2 fue comparada entre el grupo de casos TCGT (n=40) y el grupo de controles mexicanos (n=36). Para facilitar la visualización de los resultados, los valores de p obtenidos del análisis de asociación a nivel de variantes individuales fueron empleados para construir un gráfico de Manhattan (**Figura 8A**). En la **Tabla 8** se colocaron los resultados de la comparación para las variantes más frecuentes en la cohorte TCGT.

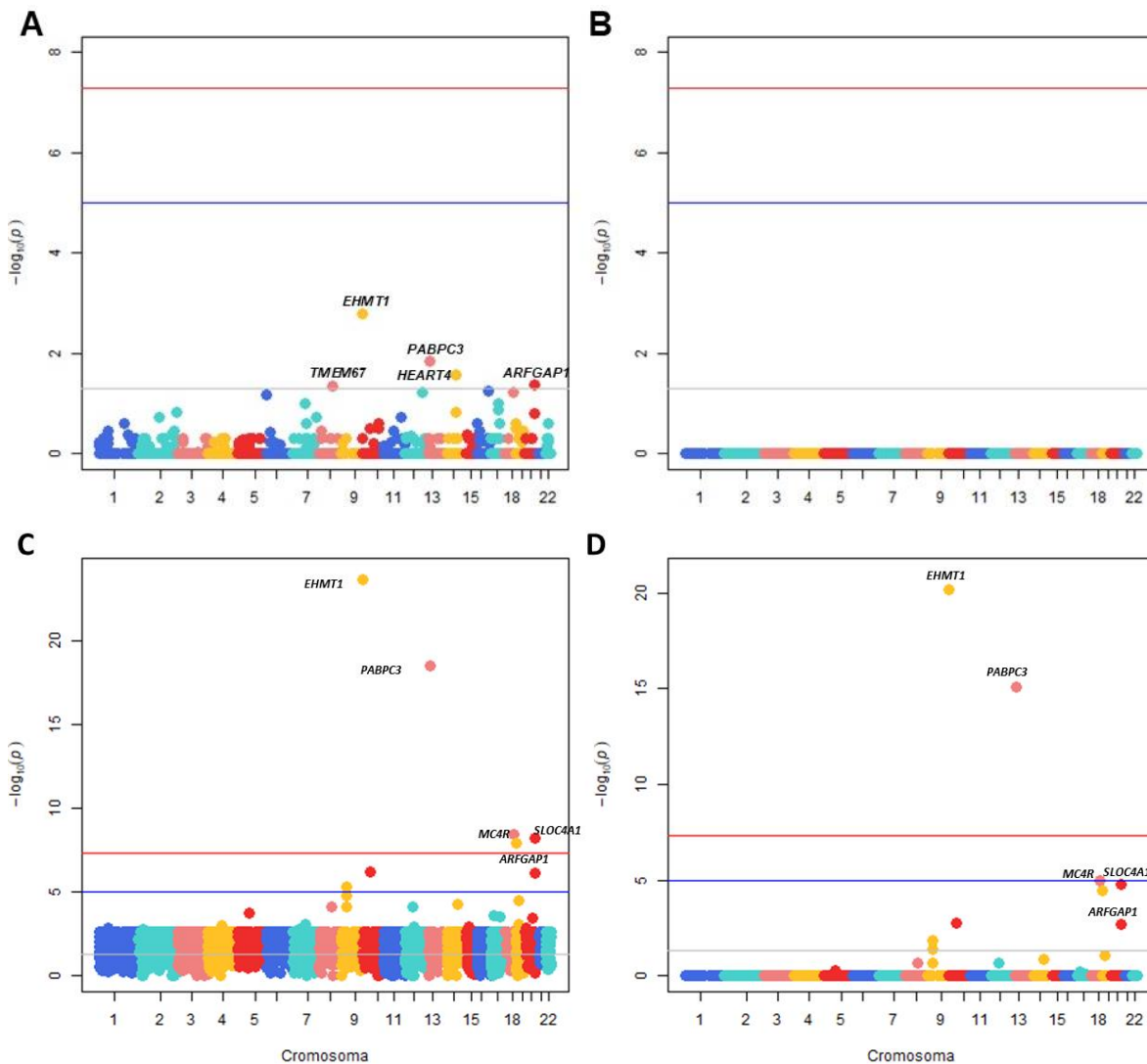


Figura 8. Gráficos de Manhattan para los análisis de asociación de variantes individuales. Se comparó la frecuencia alélica de las variantes en el grupo de casos contra la cohorte de controles sanos mexicano, graficando los valores de p antes y (**A**) y después (**B**) del ajuste por pruebas múltiples. Como validación se replicó el ensayo usando la frecuencia alélica reportada en la población latina de la base de datos gnomAD en su versión v2.1 *non-cancer*. Se graficaron los valores de p antes (**C**) y después (**D**) del ajuste por pruebas múltiples. La línea gris representa el umbral para valores significativos en un estudio de asociación del genoma completo ($p= 5 \times 10^{-8}$), la línea roja representa el umbral para valores sugerentes de asociación en un GWAS ($p= 1 \times 10^{-5}$) y la línea gris se marca para un valor de $p = 0.05$. Los letreros corresponden al nombre del gen en el que se encontraron las variantes con los valores de p más pequeños en el análisis.

Tabla 8 Resultados del análisis de asociación de tipo casos-controles para las variantes con una frecuencia alélica $\geq 5\%$ en la cohorte del TCGT

	Gen	ID	CASOS (n=40)			CONTROLES (n=36)					CONTROLES_gnomAD (n=17,000)						
			AF	Ref	Alt	AF	Ref	Alt	OR	p	p_adj	AF	Ref	Alt	OR(95% CI)	p	p_adj
1	<i>TMEM67</i>	rs386834190	0.275	58	22	0.431	41	31	0.5(0.3-1.0)	0.044	1	NA	NA	NA	NA	NA	NA
2	<i>EHMT1</i>	rs1234654104	0.125	70	10	0.000	72	0	10.1(1.2-375.6)	0.002^a	1	9.78E-05	30657	3	1079.5(380.5-4472.6)	2.3x10^{-24a}	6.4x10⁻²¹
3	<i>MPHOSPH1</i>	rs4852764	0.113	71	9	0.111	64	8	0.9(0.4-2.7)	0.978	1	6.91E-02	18822	1398	1.7(0.9-3.5)	0.128	1
4	<i>NUP88</i>	rs181249279	0.100	72	8	0.069	67	5	1.2(0.5-4.4)	0.501	1	4.42E-02	32682	1510	2.4(1.2-5.2)	0.015	1
5	<i>NKAPL</i>	rs61737340	0.111	64	8	0.056	68	4	1.7(0.6-6.6)	0.367 ^a	1	5.59E-02	32345	1915	2.1(1.1-4.6)	0.042	1
6	<i>PABPC3</i>	rs1044301916	0.088	73	7	0.000	72	0	6.8(0.8-263.9)	0.014^a	1	0.00E+00	34236	0	3238.5(395.4-123458.0)	2.9x10^{-19a}	8.1x10⁻¹⁶
7	<i>ARFGAP1</i>	rs2273499	0.125	49	7	0.028	70	2	3.3(1.0-18.7)	0.041^a	1	2.39E-02	33437	817	5.7(2.9-13.4)	7.8x10⁻⁷	2.2x10⁻³
8	<i>NEK11</i>	rs140058289	0.088	73	7	0.083	66	6	0.9(0.3-3.1)	0.927	1	4.83E-02	32603	1655	1.9(0.9-4.3)	0.103	1
9	<i>TTC23L</i>	rs77793555	0.088	73	7	0.056	68	4	1.3(0.5-5.2)	0.540 ^a	1	4.52E-02	32624	1544	2.0(1.0-4.6)	0.069	1
10	<i>RPP40</i>	rs144676135	0.088	73	7	0.014	71	1	3.4(0.8-28.9)	0.066 ^a	1	3.37E-02	31871	1111	2.7(1.4-6.2)	0.008	1
11	<i>OR52D1</i>	rs190320444	0.079	70	6	0.097	65	7	0.7(0.3-2.4)	0.695	1	6.75E-02	31928	2310	1.2(0.6-2.8)	0.690	1
12	<i>THSD4</i>	rs12439065	0.083	66	6	0.056	68	4	1.2(0.4-5.2)	0.745 ^a	1	5.21E-02	33213	1825	1.6(0.8-4.0)	0.234	1
13	<i>AHSA2P</i>	rs141666493	0.075	74	6	0.069	67	5	0.9(0.3-3.5)	0.895	1	5.43E-02	32398	1860	1.4(0.7-3.4)	0.414	1
14	<i>SLCO4A1</i>	rs147153778	0.088	62	6	0.028	70	2	2.2(0.7-13.1)	0.157 ^a	1	1.17E-02	33853	401	8.0(3.9-19.8)	5.7x10⁻⁹	1.6x10⁻⁵
15	<i>LRGUK</i>	rs140175129	0.075	74	6	0.069	67	5	0.9(0.3-3.5)	0.895	1	3.47E-02	32843	1179	2.2(1.1-5.4)	0.049	1
16	<i>RCC1L</i>	rs971053877	0.075	74	6	0.083	66	6	0.8(0.3-2.8)	0.849	1	2.22E-02	6092	138	3.5(1.7-8.7)	0.002	1
17	<i>PADI1</i>	rs140750531	0.064	73	5	0.069	67	5	0.8(0.3-3.1)	0.896	1	2.98E-02	33225	1019	2.2(1.0-5.8)	0.075	1
18	<i>PDZRN4</i>	rs145855549	0.064	73	5	0.097	65	7	0.5(0.2-2.1)	0.455	1	6.28E-02	32099	2151	1.0(0.5-2.7)	0.962	1
19	<i>SH2D7</i>	rs186449076	0.074	63	5	0.042	69	3	1.3(0.4-6.9)	0.484 ^a	1	4.01E-02	25967	1085	1.9(0.9-5.0)	0.161	1
20	<i>WFIKK2</i>	rs191998613	0.125	35	5	0.042	69	3	2.4(0.8-12.5)	0.131 ^a	1	4.14E-02	32837	1417	3.2(1.5-8.8)	0.008	1
21	<i>MC4R</i>	rs79783591	0.066	71	5	0.000	72	0	5.0(0.6-205.5)	0.059 ^a	1	7.38E-03	34847	259	9.3(4.3-24.8)	3.8x10⁻⁹	1.1x10⁻⁵
22	<i>TEX261</i>	rs151285112	0.066	71	5	0.042	69	3	1.2(0.4-6.1)	0.720 ^a	1	3.91E-02	32913	1341	1.7(0.8-4.5)	0.232	1
23	<i>CCDC171</i>	rs141812748	0.063	75	5	0.083	66	6	0.6(0.2-2.4)	0.621 ^a	1	3.77E-02	32537	1275	1.7(0.8-4.4)	0.245	1
24	<i>ARHGEF16</i>	rs190407416	0.050	76	4	0.028	70	2	1.2(0.3-8.0)	0.684 ^a	1	1.99E-02	33544	682	2.6(1.1-7.5)	0.077	1
25	<i>MOB2</i>	rs200771049	0.051	74	4	0.028	70	2	1.2(0.4-8.3)	0.683 ^a	1	1.95E-02	32638	650	2.7(1.2-7.9)	0.067 ^a	1
26	<i>SLC8B1</i>	rs3764034	0.133	26	4	0.028	70	2	3.5(1.0-23.9)	0.060 ^a	1	2.65E-02	26816	730	5.4(2.3-17.0)	0.008^a	1
27	<i>STARD5</i>	rs138188091	0.051	74	4	0.028	70	2	1.2(0.4-8.3)	0.683 ^a	1	5.00E-02	27725	1459	1.0(0.4-3.0)	0.797 ^a	1
28	<i>CLEC18B</i>	rs148861848	0.057	66	4	0.000	72	0	4.3(0.5-185.7)	0.056 ^a	1	1.54E-02	8145	127	3.8(1.6-11.4)	0.024^a	1
29	<i>SCRN3</i>	rs34170959	0.050	76	4	0.028	70	2	1.2(0.3-8.0)	0.684 ^a	1	2.31E-02	33433	789	2.2(1.0-6.5)	0.114 ^a	1
30	<i>FARSB</i>	rs138484554	0.050	76	4	0.028	70	2	1.2(0.3-8.0)	0.684 ^a	1	2.03E-02	30446	632	2.5(1.1-7.4)	0.081 ^a	1
31	<i>ESYT3</i>	rs191463941	0.053	72	4	0.056	68	4	0.7(0.2-3.6)	1.000 ^a	1	2.44E-02	30479	761	2.2(1.0-6.5)	0.115 ^a	1
32	<i>UBE2D3</i>	rs147920467	0.050	76	4	0.028	70	2	1.2(0.3-8.0)	0.684 ^a	1	1.19E-02	33822	406	4.3(1.9-12.7)	0.016^a	1
33	<i>UTP15</i>	rs151157740	0.050	76	4	0.028	70	2	1.2(0.3-8.0)	0.684 ^a	1	2.12E-02	33525	725	2.4(1.0-7.1)	0.091 ^a	1
34	<i>C5orf49</i>	rs76872483	0.050	76	4	0.028	70	2	1.2(0.3-8.0)	0.684 ^a	1	1.76E-02	34204	612	2.9(1.3-8.5)	0.053 ^a	1
35	<i>HEATR4</i>	rs145046095	0.013	79	1	0.097	65	7	0.1(0.1-1.0)	0.027^a	1	5.26E-02	32721	1817	0.2(0.1-1.7)	0.131 ^a	1

ID: Identificador de la variante en dbSNP, AF= frecuencia alélica, Ref: Conteos totales para el alelo de referencia, Alt: Conteos totales para el alelo alternativo, OR: Razón de probabilidades IC: Intervalo de confianza del 95%, p: valor p calculado según la prueba de Chi cuadrada o (a) el test exacto de Fisher, p_adj: valores de p después de la corrección de pruebas múltiples Holm considerando los 2778 ensayos. En negrita se señalan los valores con un $p < 0.05$. En azul se señalan las variantes que son interesantes para su estudio en cohortes TCGT con un mayor número de muestras.

De las 2778 variantes, únicamente las variantes en los genes *EHMT1*, *PABPC3*, *ARFGAP1*, *TMEM67* y *HEART4* presentaron resultados significativos al obtener valores de $p < 0.05$. No obstante, para las variantes en los últimos dos genes (rs386834190 y rs145046095) la AF es mayor en el grupo control que en grupo de casos (**Tabla 8**), por lo que la asociación es negativa con el fenotipo y no se catalogan como variantes de riesgo.

En nuestro ensayo, ninguna de estas variantes obtuvo valores de $p < 0.05$ después del ajuste de pruebas múltiples (**Figura 8B**, **Tabla 8**). Este resultado era esperado, debido a al tamaño de muestra, los valores de las frecuencias alélicas que se comparan en la prueba estadística no son tan discordantes entre grupos y las señales de significancia tienen valores moderados, en este caso, cercanos a 0.05. Aplicar ajustes estrictos de pruebas múltiples puede ser excesivamente drástico y enmascarar asociaciones genuinas entre variantes y fenotipos que podrían tener relevancia biológica y clínica (Lee & Lee, 2018; Ranganathan et al., 2016). Siendo así, al interpretar los resultados de estudios con un tamaño de muestra limitado, en lugar de descartar prematuramente variantes que no superen los umbrales de significancia después del ajuste, es necesario considerar estrategias de validación en cohortes independientes, análisis de tendencias biológicas consistentes en la literatura o la relevancia funcional de las variantes.

Por esta razón, en este trabajo se realizó una validación adicional con una segunda cohorte de controles, con el propósito de abordar la pregunta de investigación de manera más robusta. Para este análisis se utilizó la AF reportada en los exomas de la base de datos pública gnomAD en su versión v2.1 non-cancer. Además, con la intención de aminorar el impacto de la estratificación poblacional, solo se consideraron los conteos de la población latina (nLatino/Admixed American= $\sim 17,000$) dentro de la base de datos.

Los valores de p obtenidos con la prueba estadística de asociación fueron integrados en un gráfico de Manhattan antes (**Figura 8C**) y después del ajuste por pruebas múltiples (**Figura 8D**). Las variantes germinales en los genes *EHMT1*, *PAPBC3* y *ARFGAP1* vuelven a sobresalir en el análisis, al presentar valores de $p < 0.05$ después de la corrección por pruebas múltiples (**Figura 8D**). La consistencia en los hallazgos en ambas cohortes aumentó la confianza en los resultados y las posiciona como las principales variantes con una potencial asociación al riesgo a desarrollar tumores de células germinales.

Por su parte, las variantes en los genes *SLCO4A1* y *MC4R* reportaron valores de p cercanos al umbral de asociación sugestiva a nivel genómico ($p = 1.6 \times 10^{-5}$ y $p = 1.1 \times 10^{-5}$, respectivamente) (**Figura 8D**). Explorar su estudio de manera más profunda podría resultar valioso en la identificación de factores de riesgo asociados a la enfermedad.

Es necesario enfatizar que el uso de gnomAD como población control ha sido objeto de debate, debido a que la falta de acceso a los datos a nivel individual hace imposible las correcciones por estratificación poblacional y que los datos entre casos-contróles reciban el mismo procesamiento técnico y control de calidad (Karczewski et al., 2020).

El uso de esta base de datos también generó un desequilibrio importante entre el tamaño de las muestras de casos y controles al tener 40 casos contra 17,000 controles (razón casos-control de 1:425). Las simulaciones estadísticas que prueban el efecto de la razón casos-control en los estudios de asociación, muestran que los diseños con un desequilibrio extremo (razón 1:99) generan una tasa más alta de errores de Tipo I en comparación con diseños equilibrados (razón 1:1) (Yang et al., 2023; Zhang et al., 2019).

Los datos desequilibrados pueden violar los supuestos de las inferencias estadísticas y, por lo tanto, conducir a una mayor variación para la estimación de los coeficientes, incluidos el valor de p , el estadístico de la prueba y las dimensiones del OR (Dai et al., 2021).

Finalmente, poniendo los resultados en perspectiva, se sabe que las pruebas de una sola variante en un GWAS consideran una asociación significativa cuando el valor de p es $< 5 \times 10^{-8}$. Sin embargo, las pruebas de una sola variante son menos poderosas para variantes raras que para variantes comunes con tamaños de efecto idénticos (Kiezun et al., 2012). Un ejemplo claro es que con un OR=1.4, el tamaño de muestra requerido es de 6400 individuos para lograr una potencia de 80%, cuando la AF=0.1 y se supone una prevalencia de la enfermedad del 5% y un nivel de significancia de 5×10^{-8} (Lee et al., 2014).

Por esta razón y con el fin de superar los problemas de poder estadístico asociado a la prueba variantes raras de forma individual, se desarrollaron las pruebas de agregación de variantes, donde se evalúan los efectos acumulativos de múltiples variantes genéticas en un gen o región, aumentando el poder estadístico al considerar el efecto colectivo de las variantes en su asociación con la enfermedad (Bansal et al., 2010; Lee et al., 2014).

8.5. Asociación de tipo casos-control por agregación de múltiples variantes basada en genes (*Gene-burden test*).

El análisis de variantes raras requiere métodos estadísticos que difieren de aquellos utilizados para probar variantes comunes. Sumar la frecuencia alélica de todas las variantes que se describen en un gen puede ayudar a aumentar el poder estadístico de la asociación y con ello, la capacidad de describir efectos acumulativos que no son evidentes en el análisis de variantes individuales (Kiezun et al., 2012).

Se realizó una prueba de agregación de múltiples variantes basada en genes (*Gene-Burden test*) para identificar genes mutados con una frecuencia significativamente mayor en pacientes con TCGT en comparación con controles de exomas latinos públicamente disponibles (gnomAD $n=17,296$), utilizando la paquetería TRAPD (Test Rare vAriants with Public Data) desarrollado para R por (Guo et al., 2018). La herramienta está optimizada para el uso de los datos disponibles en bases de datos públicas como gnomAD o ExAC y aunque está enfocado en identificar genes asociables a enfermedades mendelianas, existen reportes en donde utilizan el mismo abordaje con TRAPD para enfermedades complejas como astrocitomas pediátricos (Muskens et al., 2020).

El archivo VCF con todas las variantes presentes en la base de datos de exomas de gnomAD fue anotado y filtrado bajo los mismos criterios con los que se seleccionaron las variantes de Nivel 1 y 2. Por medio de TRAPD se agruparon las variantes calificadas (VC) en 16,354 genes y se realizó una prueba exacta de Fisher de una cola para determinar si existía una mayor carga de VC en los casos en comparación con los controles para cada gen, bajo un modelo dominante y uno recesivo. Los resultados para los 20 genes con el menor valor de p fueron colocados como datos representativos en la **Tabla 9**. Con los resultados se construyó una gráfica cuantil-cuantil (**Figura 9**).

Un umbral de significancia conservador para esta estrategia es de $p < 2,5 \times 10^{-6}$, que consiste en la corrección de Bonferroni considerando aproximadamente 20,000 genes (Guo et al., 2018). Cabe recalcar que con este análisis ningún gen obtuvo un valor significativo bajo el modelo recesivo.

Tabla 9. Resultados de la prueba de agrupación basada en genes según TRAPD para 20 genes representativos.

Gen	CASOS (n=40)				CONTROLES (n=17 296)			p_dom	p_rec
	Het	Ch	Hom	AC_total	Het	Hom	AC_total		
<i>PABPC3</i>	7	0	0	7	39	0	39	1.00E-11	1
<i>PEBP1</i>	2	0	0	2	14	0	14	6.10E-04	1
<i>SDHAF2</i>	2	0	0	2	16	0	16	7.76E-04	1
<i>MC4R</i>	5	0	0	5	320	5	330	9.23E-04	1
<i>OR6C65</i>	2	0	0	2	22	0	22	1.39E-03	1
<i>SLC5A10</i>	2	0	0	2	25	0	25	1.76E-03	1
<i>EIF1B</i>	1	0	0	1	0	0	0	2.31E-03	1
<i>SH3BGR</i>	1	0	0	1	0	0	0	2.31E-03	1
<i>CD1C</i>	3	0	0	3	111	1	113	2.35E-03	1
<i>RAD51C</i>	2	0	0	2	31	0	31	2.62E-03	1
<i>ASXL1</i>	2	0	0	2	39	0	39	4.02E-03	1
<i>PEX6</i>	2	0	0	2	40	0	40	4.22E-03	1
<i>OR2AT4</i>	1	0	0	1	1	0	1	4.61E-03	1
<i>PAQR4</i>	1	0	0	1	1	0	1	4.61E-03	1
<i>BARD1</i>	2	0	0	2	47	0	47	5.70E-03	1
<i>TLN1</i>	2	0	0	2	49	1	51	6.40E-03	1
<i>PATJ</i>	5	0	0	5	524	0	524	7.08E-03	1

Het: No. de individuos que llevan al menos una variante heterocigótica en el gen, Ch: de individuos que llevan al menos dos variantes heterocigóticas en el gen, Hom:No. de individuos que llevan al menos una variante homocigótica en el gen. Ac_total: Conteo alélico total de las variantes en el gen, p_dom: valor p bajo el modelo dominante, p_rec: valor p bajo el modelo recesivo.

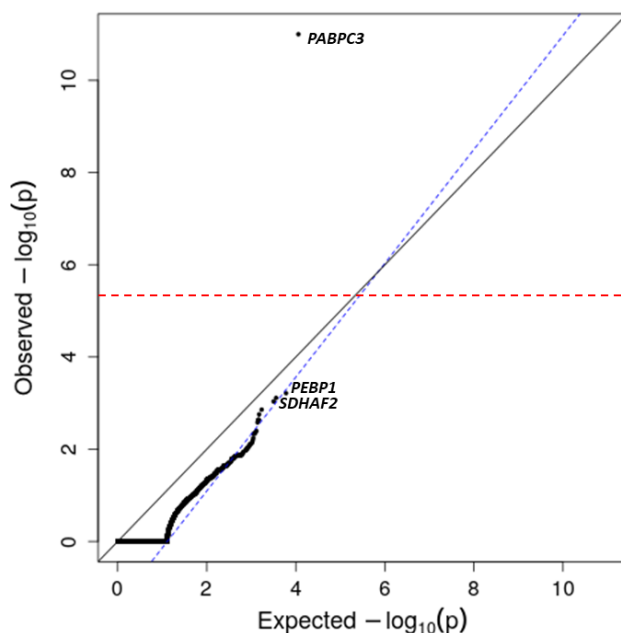


Figura 9. Gráfico cuantil-cuantil para los valores p obtenidos por el *gene-burden test* usando un modelo dominante. En la gráfica se representa el umbral de $p=2.5 \times 10^{-6}$ con una línea roja punteada. La línea azul punteada muestra la línea de ajuste observada entre el percentil 50 y 95 de los genes; la pendiente de esta línea es $\lambda_{\Delta 95}$. Los letteros corresponden al nombre del gen en el que se encontraron las variantes con los valores de p más pequeños en el análisis.

Nuestros resultados muestran que el gen *PABPC3* es el único de los genes con un valor significativo de asociación para la prueba de carga de variantes múltiples bajo un modelo dominante ($p = 1.0 \times 10^{-11}$). Este hecho es consistente con los análisis de asociación a nivel de variantes individuales, donde el SNV rs1044301916 en *PABPC3* fue una de las alteraciones para las que se describieron los valores más pequeños de significancia (**Figura 8**). Esta coincidencia sugiere que la relación entre *PABPC3* y el riesgo de desarrollar TCGT es sólida y con una menor probabilidad de que se deba al azar.

PABPC3 es un gen que codifica para una proteína de unión a poly(A) que se expresa de manera específica en testículo. Es una proteína de 631 aminoácidos (70.1 kDa) con 92.5% de identidad a la proteína PABP1, que se expresa de manera ubicua en las células. Este tipo de proteínas están involucradas en la estabilidad y el mantenimiento de potencialidad de traducción del mRNA de vida prolongada (Féral et al., 2001).

Se ha demostrado que las proteínas de unión a poli(A) (PABP) desempeñan funciones fundamentales en la estabilización y el control de la traducción de los mRNA durante la espermatogénesis (Elliot, 2003; Kleene, 1989), y está diferencialmente expresado en hombres infértiles con azoospermia no obstructiva (Ozturk et al., 2016).

Con todo lo anterior, el resultado del análisis de asociación por agregación de múltiples variantes basada en genes establece un precedente importante para relacionar a *PABPC3* con el desarrollo de los TCGT, aun cuando el ensayo presentó un coeficiente de control genómico $\lambda_{\Delta 95} = 1.27$, según el algoritmo de TRAPD.

En términos simples, el coeficiente λ es una estimación del exceso de errores de tipo falso-positivo, al medir la relación entre la distribución de los resultados observados y los esperados. Un valor $\lambda = 1$ es lo esperado (Sofer et al., 2021). Para mitigar esta inflación se emplean estrategias que corrigen los artefactos más comunes como efectos de lote o de diferencias de ascendencia entre casos y controles (Georgiopoulos & Evangelou, 2016; van Iterson et al., 2017; Voorman et al., 2011). No obstante, estas correcciones necesitan los datos de los genotipos a nivel individual, y lamentablemente, en gnomAD esta información no está disponible (Guo et al., 2018).

Cuando se critica esta prueba estadística, se explica que no se debería simplemente suponer que la presencia de dos o más mutaciones que ocurren de forma independiente en el mismo gen es una evidencia definitiva para el papel causal para ese gen en la enfermedad, pues para esto deberían tener la misma dirección y magnitud de efecto sobre el rasgo estudiado (MacArthur et al., 2014; Neale et al., 2011). Sin embargo, en nuestro estudio, la asociación propuesta entre *PABPC3* y los TCGT se encuentra respaldada por evidencia más allá del *gene-burden test*. Esta evidencia engloba la tendencia de este gen a destacarse en el análisis de variante a nivel individual y que el contexto biológico respalde la relación con estos tumores. Ya que, al ser este un gen que se expresa específicamente en el tejido de estudio y que está involucrado en la espermatogénesis; el proceso crucial en el que se genera la lesión pre-invasiva GSNIS, que da pie a la formación de los TCGT (Cheng et al., 2018).

8.6. Panel de variantes candidatas para su estudio en cohortes más numerosas de TCGT.

A través de la evaluación y la integración exhaustiva de todos los resultados obtenidos a lo largo de este proyecto, hemos logrado realizar avances importantes en la comprensión de las bases genéticas de los tumores de células germinales testiculares. Con su enfoque exploratorio, este trabajo va más allá de la identificación de la relación entre *PABPC3* y la neoplasia bajo análisis. Al haber identificado también novedosas variantes con valores sugerentes de asociación de los análisis

individuales. Como consecuencia, proponemos un panel de variantes que se perfilan como prometedores biomarcadores de riesgo para TCGT.

La confirmación de la implicación de estas variantes en TCGT requiere validación adicional. No obstante, la investigación ya estaría direccionada a blancos específicos. Esta estrategia incrementa la probabilidad de que los resultados sean replicados en un estudio más amplio, pues evita el uso de secuenciación masiva en un número grande de muestras, una opción poco realista para los laboratorios que no cuentan con los recursos para hacerlo (Kiezun et al., 2012). Asimismo, la interpretación biológica de los resultados se vuelve más manejable y enfocada en entender el impacto de la variante y en el proceso patológico en el que está relacionada. Esto es esencial en la medicina traslacional, donde se busca aplicar los resultados de investigación genética en la práctica clínica.

El panel de variantes candidatas propuesto en este trabajo se presenta en la **Tabla 10**, junto con algunas de las características funcionales de las variantes, y las enfermedades con las que se ha relacionado en diferentes bases de datos.

Tabla 10. Panel de variantes germinales con un alto potencial de estar asociadas a los tumores de células germinales de testículo.

Chr	Pos	Ref	Alt	ID	Gen	Localización	Cambio_aa	ClinVar_Cond	OMIM_ID	COSMIC_Ocu	
1	9	137743368	TA	T	rs1234654104	EHMT1	splicing	-	Síndrome de Kleeftstra	610253	ovario, intestino
2	13	25097025	T	A	rs1044301916	PABPC3	exón	p.L276H	-	-	riñón, intestino
3	20	62666451	C	T	rs147153778	SLCO4A1	exón	p.R450W	-	-	-
4	18	60371544	A	T	rs79783591	MC4R	exón	p.I269N	Obesidad	601665	-
5	20	63278918	G	A	rs2273499	ARFGAP1	exón	p.V131M	-	-	cáncer de mama

Chr: Cromosoma, Pos: Posición en el genoma, Ref: Alelo de referencia, Alt: Alelo alternativo, ID: Identificador de la variante en dbSNP, Cambio_aa: cambio en el aminoácido a nivel proteína, ClinVar_Cond: Condición asociada a la variante en ClinVar, COSMIC_Ocu: Tipos de cáncer en los que se ha registrado la variante en Cosmicv95.

Como se mencionó con anterioridad, PABPC3 es una proteína que estabiliza los mRNA en el testículo y favorece su traducción (Féral et al., 2001). Aunque su relación con el cáncer aún necesita ser profundamente explorada, la variante rs1044301916 ya ha sido identificada en muestras tumorales de pacientes con cáncer de riñón e intestino (**Tabla 10**) y la base de datos *The Human Protein Atlas* (Uhlén et al., 2015), lo propone como un marcador de pronóstico desfavorable para el cáncer de ovario.

Por su parte, EHMT1 es una proteína que cataliza la metilación de la lisina 9 en la histona 3 (H3K9me), una modificación asociada al silenciamiento transcripcional y al mantenimiento de la estabilidad genómica (Nachiyappan et al., 2022; Saha & Muntean, 2021). La sobreexpresión de esta proteína se observó en diferentes tipos de cáncer y en enfermedades de deterioro cognitivo (Iglesias-Ortega et al., 2023; Nachiyappan et al., 2022).

La variante rs1234654104 en *EHMT1* se localiza en un sitio de splicing alternativo, estas mutaciones generalmente promueven la traducción de una proteína significativamente alterada debido a la omisión del exón, el acortamiento o la inclusión de material intrónico (G.-S. Wang & Cooper, 2007; Wimmer et al., 2007). Hasta el momento no hay reportes experimentales que demuestren su impacto en la función de la proteína. Aún con esto, se puede pensar que una desregulación en la actividad de EHMT1 puede tener un papel relevante en la serie de eventos que originan los TCGT debido a que la pérdida de H3K9me3 en la célula promueve rompimientos cromosómicos, las duplicaciones segmentarias y la ganancia de variantes estructurales (Kondo et al., 2008; Möller et al., 2019).

El gen *ARFGAP1* codifica para una proteína activadora de GTPasa. La hidrólisis de GTP es necesaria para la disociación de proteínas de membrana y las vesículas derivadas del aparato de Golgi (Shiba & Randazzo, 2012). Se sugiere que *ARFGAP1* desempeña un papel en la función fisiológica de los espermatozoides debido a su alta expresión en los testículos, y a que otro miembro de la familia de las *ARFGAP*, *SMAP2*, es importante para la formación de acrosomas durante la espermiogénesis en ratones (Van Son et al., 2020).

Sin duda, el dato más importante para nuestro estudio es que se ha probado que *ArfGAP1* suprime la translocación y activación lisosomal de *mTORC1*, un regulador maestro del crecimiento y la progresión del ciclo celular (Meng et al., 2021). Aunque nuestra comprensión de la función de *ArfGAP1* en TCGT es limitada, su papel en la patogénesis necesita ser investigado considerando que se ha visto que los seminomas sobreexpresan proteínas de la vía *mTOR* (Yaba et al., 2016).

Ahora bien, la proteína *SLCO4A1* es un componente integral de la membrana plasmática de las células de todo tipo de tejidos con excepción del bazo y los leucocitos. Interviene en el transporte de Na^+ , sales biliares, ácidos orgánicos, hormonas tiroideas, estrona-3-sulfato, estradiol-17beta-glucuronido y prostaglandina E2 (Stieger & Hagenbuch, 2014). La absorción alterada de estos sustratos puede contribuir a cambios en la actividad de agentes quimioterapéuticos, planteando su importancia en la enfermedad. La sobreexpresión del transportador está reportada en diferentes tipos de cáncer (X.-S. Wang et al., 2021) y es un marcador de pronóstico pobre en el cáncer colorrectal (Ban et al., 2017; Chen et al., 2023).

Finalmente, *MC4R* es un receptor acoplado a proteínas G que participa en la ruta de señalización hipotalámica leptina-melanocortina. La mutación s79783591 con pérdida de función está fuertemente asociada con la diabetes tipo 2 y obesidad en la población mexicana (Vázquez-Moreno et al., 2021). Aunque hasta el momento no se ha establecido una relación entre obesidad y cáncer testicular, ya se han publicado trabajos que correlacionan la obesidad con afecciones relacionadas a TCa como la infertilidad (Nie et al., 2022). *MC4R* está involucrado en gran medida con los cambios en los niveles hormonales, el aumento de la temperatura testicular y la imposición de un microambiente inflamatorio (Nie et al., 2022; Palmer et al., 2012).

Dentro de la base de datos de los pacientes, no se encuentra registrado ningún historial de comorbilidades, por lo que, no es posible decir si la observación de la mutación se asocia a obesidad y no a los TCGT. Sin embargo, de los cinco pacientes con la mutación, únicamente 2 tienen un IMC > 25 (sobrepeso). No cabe duda, que la relación entre los TCGT y las enfermedades metabólicas como la obesidad o la diabetes es un tema muy interesante para investigaciones futuras.

En resumen, todas las variantes incluidas en el panel propuesto se localizan en genes estrechamente relacionados con procesos celulares que podrían impulsar el desarrollo del cáncer, así como otras afecciones asociadas a los TCGT, como la infertilidad y la azoospermia. Estos hallazgos fortalecen la validez de los resultados presentados en este estudio y generan un fuerte estímulo para continuar la investigación en nuestra población, con el objetivo de validar el panel en cohortes más amplias y representativas. Esta validación es esencial para determinar la eficacia de estas variantes como biomarcadores de riesgo en la enfermedad y, en última instancia, su potencial implementación en la práctica clínica, para beneficio de los pacientes.

8.7. Validación de los datos de secuenciación de exoma completo mediante la técnica de RFLPs.

La secuenciación de exoma completo es una herramienta poderosa que permite analizar grandes volúmenes de información genética de manera eficiente. Sin embargo, también puede ser

susceptible a errores técnicos, especialmente en el proceso de llamado de variantes. Para asegurar la precisión de los hallazgos y validar la existencia de las variantes identificadas durante la fase de descubrimiento se optó por genotipificar las variantes rs1044301916 en PABPC3, rs147153778 en SLCO4A1 y rs2273499 en ARFGAP1 utilizando el método de análisis de los fragmentos de restricción de longitud polimórfica (RFLPs) en subconjuntos representativos de la cohorte TCGT.

En primer lugar, se recuperaron las muestras de DNA sanguíneo de 28 pacientes del banco de ácidos nucleicos del INCan y para los 12 pacientes restantes el DNA se extrajo de las muestras de sangre periférica. La concentración final de todas las muestras de DNA genómico fue determinada usando el espectrofotómetro NanoDrop 2000 (Thermo Scientific).

Posterior a esto, se separaron electroforéticamente 100 ng de DNA en un gel de agarosa al 1% para verificar su integridad (**Figura 10**).

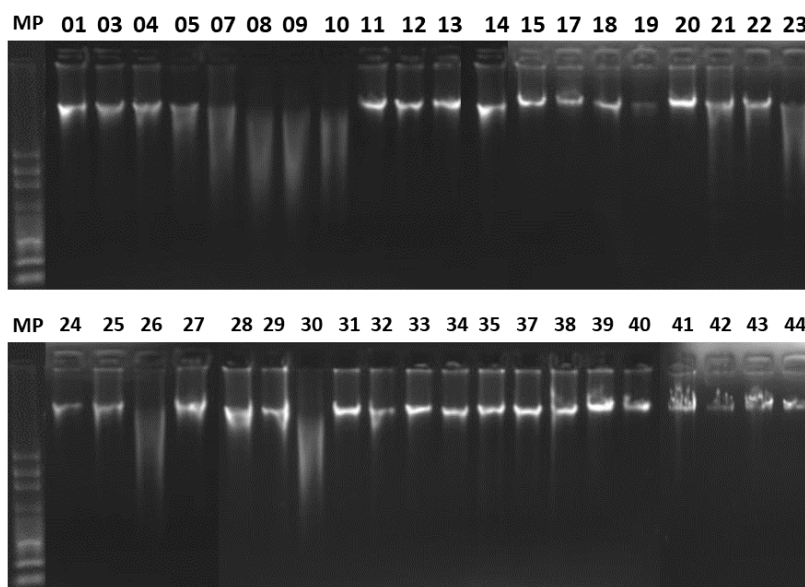


Figura 10. Electroforesis en un gel de agarosa al 1% de las muestras de DNA genómico de los cuarenta pacientes de la cohorte TCGT. MP: Marcador de peso GeneRuler™ 1kb Plus DNA Ladder (Thermo Scientific).

De las 40 muestras, 35 tienen una integridad alta o aceptable, para las que hay una banda definida de alto peso molecular en la parte superior del gel. No obstante, dado que las muestras serían empleadas como templado para amplificar producto de PCR no mayores a 350 pb, incluso las 5 muestras parcialmente degradadas tenían la calidad suficiente para los ensayos de RFLPs. Las particularidades del ensayo de genotipificación para cada una de las variantes se desglosan a continuación.

8.7.1. Variante rs147153778 en *SLCO4A1*

La variante rs147153778 en *SLCO4A1* corresponde a un cambio de una Citosina por una Timina en la hebra positiva de la cadena de DNA, la enzima de restricción *Cail* con el sitio de restricción (5' C A G N N N ↓ C T G 3') genera el corte en el alelo alternativo, como se observa en la **Figura 10A**. Para genotipificar esa variante se diseñaron dos oligonucleótidos que generan un producto de PCR de 186 pb, los productos de digestión para el alelo alternativo son dos bandas de 104 y 82 pb. Debido a la resolución del gel de agarosa estas dos bandas migran a través del gel muy próximas entre sí.

Para la validación se usaron 15 muestras, 6 de ellas pertenecientes a los pacientes portadores del alelo alternativo y 9 homocigotos al alelo de referencia, según lo señala la secuenciación de exomas. El análisis de los fragmentos de restricción en el gel de agarosa mostrado en la **Figura 11B** confirmó la presencia de 6 portadores del alelo de referencia en condición heterocigótica dentro de la cohorte TCGT.

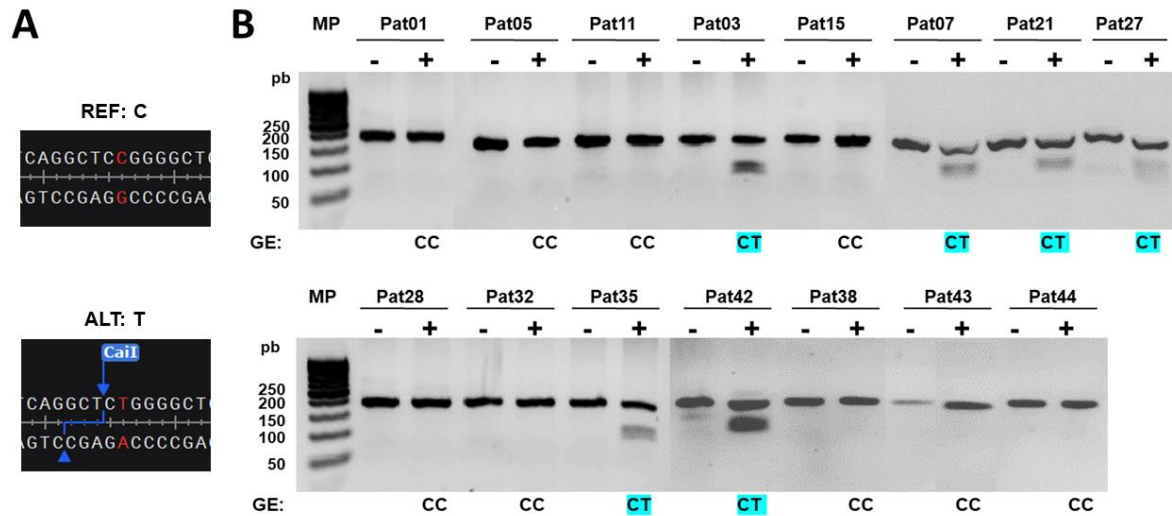


Figura 11. Genotipificación de la variante rs147153778 en 15 muestras representativas del grupo de casos TCGT. **(A)** Sitio de corte de la enzima Cail en el alelo referencia y en el alelo alternativo. **(B)** Geles de agarosa al 2.5% que muestran los patrones de restricción de longitud polimórfica de las muestras de la cohorte TCGT. El genotipo asignado es colocado en la parte inferior, las muestras con un alelo alternativo se señalan en azul. (-) Control de la reacción sin enzima de restricción, (+) Reacción con enzima de restricción. MW: Marcador de peso molecular, GE: genotipo.

8.7.2. Variante rs2273499 en *ARFGAP1*

La variante rs2273499 en *ARFGAP1* corresponde a un cambio de una Guanina por una Adenina en la hebra positiva de la cadena de DNA, la enzima de restricción FatI con el sitio de restricción (5' C A T G ↓ 3') genera un corte en el alelo alternativo, como se observa en la **Figura 12A**. Para genotipificar esa variante se diseñaron dos oligonucleótidos que amplifican un producto de PCR de 153 pb, los productos de digestión para el alelo alternativo son cuatro bandas de 60, 58, 31 y 4 pb. Debido a la resolución del gel de agarosa, en un homocigoto al alelo alternativo solo se aprecian dos bandas, la primera con los productos de 58-60 pb y la de 31 pb.

Para la validación se usaron 15 muestras, 6 de ellas pertenecientes a los pacientes portadores del alelo alternativo y 9 homocigotos al alelo de referencia, según lo señala la secuenciación de exomas. El análisis de los fragmentos de restricción en el gel de agarosa mostrado en la **Figura 9B** confirmó la presencia de 5 portadores del alelo de referencia en condición heterocigótica y un homocigoto (Paciente 33) dentro de la cohorte TCGT.

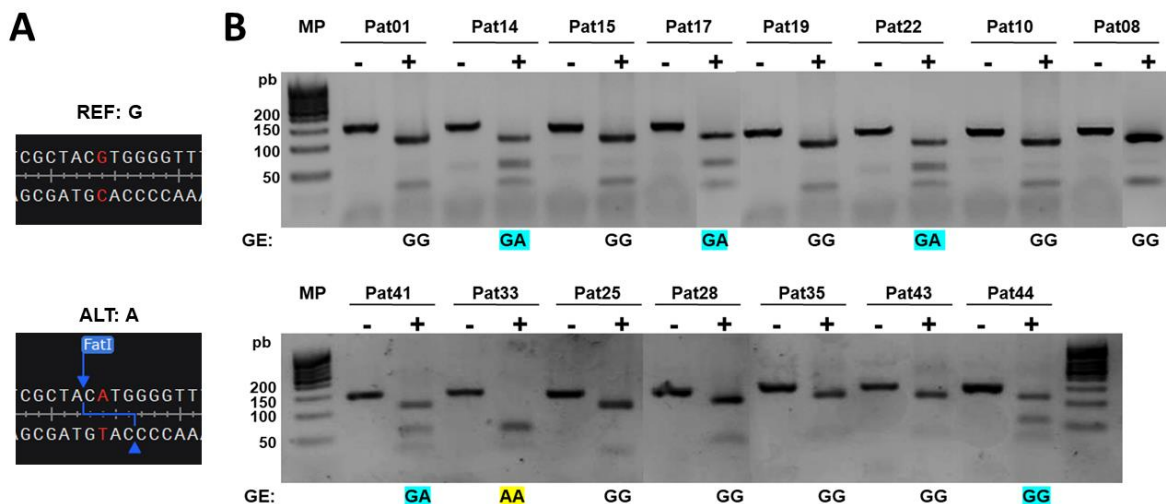


Figura 12. Genotipificación de la variante rs2273499 en 15 muestras representativas del grupo de casos TCGT. **(A)** Sitio de corte de la enzima FatI en el alelo referencia y en el alelo alternativo. **(B)** Geles de agarosa al 2.5% que muestran los patrones de restricción de longitud polimórfica de las muestras de la cohorte TCGT. El genotipo asignado es colocado en la parte inferior, las muestras con un alelo alternativo se señalan en azul y con dos alelos alternativos se señala en amarillo. (-) Control de la reacción sin enzima de restricción, (+) Reacción con enzima de restricción. MW: Marcador de peso molecular, GE: genotipo.

8.7.3. Variante rs1044301916 en *PABPC3*

La variante rs1044301916 en *PABPC3* corresponde a un cambio de una Timina por una Adenina en la hebra positiva de la cadena de DNA, la enzima de restricción BstP1 con el sitio de restricción (5' C ↓TTAAG 3') solo genera el corte en el alelo de referencia, como se observa en la **Figura 13A**. Para genotipificar esa variante se diseñaron dos oligonucleótidos que amplifican un producto de PCR de 302 pb. Tras la digestión, los homocigotos al alelo de referencia generan tres bandas de 175, 113 y 14 pb, un homocigoto al alelo alternativo solo genera dos bandas de 288 y 14 pb y un heterocigoto mostrará un patrón mixto (Ver Figura 3C, en sección de Métodos).

En la validación se usaron 36 muestras, 7 de ellas pertenecientes a los pacientes portadores del alelo alternativo según lo señala la secuenciación de exomas. El análisis de los fragmentos de restricción en el gel de agarosa mostrado en la **Figura 13B** confirmó la presencia de 6 portadores del alelo de referencia en condición heterocigótica dentro de la cohorte TCGT, que son señalados con azul. Sin embargo, en el caso del paciente 32, que sería el último portador, el ensayo no arrojó resultados concluyentes debido a una baja amplificación del fragmento durante la PCR.

En cuanto a las 9 muestras señaladas de rosa o verde, se describe una inconsistencia entre el genotipo observado (GE-O) en el gel de agarosa con los fragmentos de restricción y el genotipo esperado (GE-E) según los datos de secuenciación.

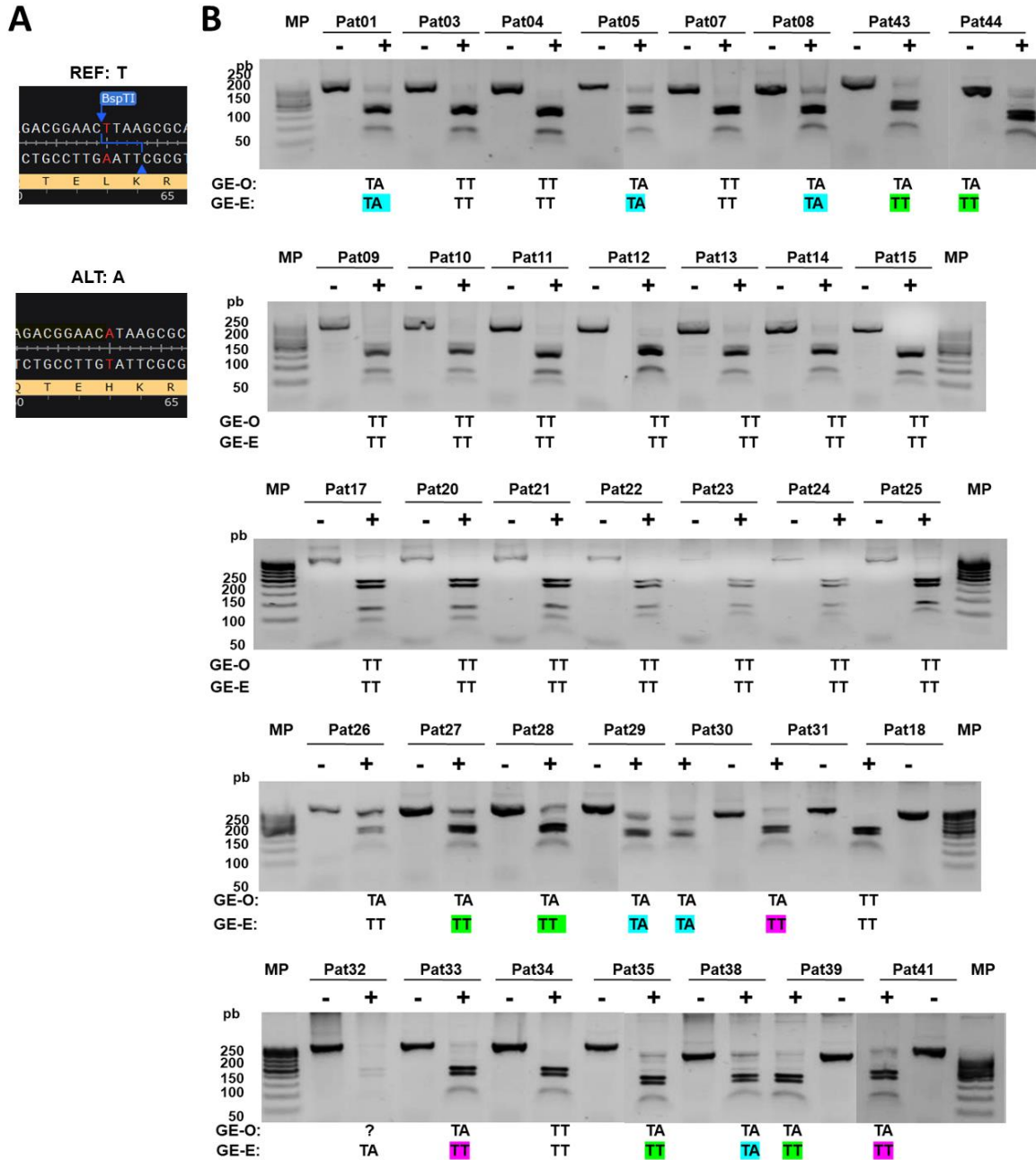


Figura 13. Genotipificación para la variante rs1044301916 en PABPC3 por RFLPs. **(A)** Sitio de corte de la enzima BspT1 en el alelo referencia y en el alelo alternativo. **(B)** Patrón de restricción de muestras de la cohorte de descubrimiento para las cuales ya es conocido el genotipo según la secuenciación de exomas. En rojo se señala la banda de 288 pb que se observa como una digestión incompleta del alelo de referencia de la muestra del paciente 3. (-) Control de la reacción sin enzima de restricción. (+) Reacción con enzima de restricción. MW: Marcador de peso molecular.

Con el ayuda del visor IGV (<https://software.broadinstitute.org/software/igv/>) se revisó manualmente el archivo BAM de los pacientes que mostraron inconsistencias. Un archivo BAM contiene información sobre las lecturas o “reads” que se han secuenciado y cómo se han alineado con el genoma de referencia, incluye detalles sobre la calidad de las bases, la posición de alineamiento y la profundidad con la que fue secuenciada cada base de la molécula. En la **Figura 14** se colocó una

imagen del visor para dos muestras representativas para cada uno de los subgrupos de pacientes, respecto al tipo de inconsistencia.

En los pacientes marcados con el color verde, como el paciente 37 (**Figura 14A**), el archivo BAM registró entre 2 y 8 lecturas con el alelo alternativo en la posición de la variante rs1044301916. Dentro del análisis de datos genómicos, se eliminan las variantes con una profundidad menor a 10, dado que estos registros pueden atribuirse a artefactos de la secuenciación o dentro del alineamiento (Koboldt, 2020). No obstante, los falsos negativos en la asignación de genotipos también están relacionados con la mala asignación en el alineamiento que puede tener una variante con una inserción, deleción o CNV en la cercanía, como es el caso de la deleción de 12 pb río arriba de la posición de la variante. En este caso, se concluye que, el genotipo en estas muestras debe ser confirmado por el estándar de oro, que es la secuenciación de Sanger (Fujiki et al., 2018).

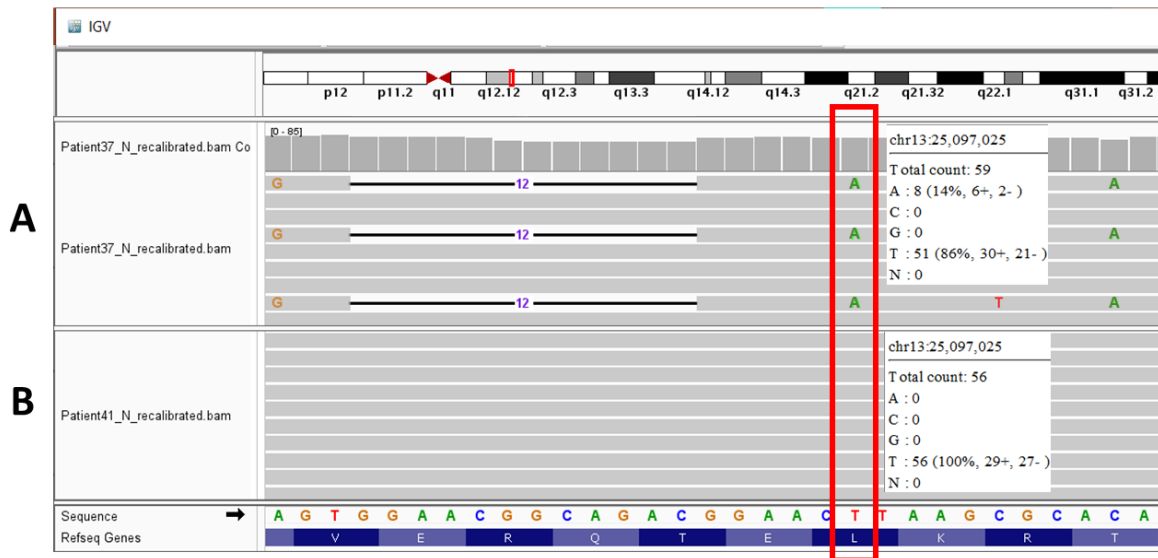


Figura 14. Visualización del vecindario genómico de la variante rs1044301916 del gen PABPC3 con la herramienta IGV. **(A)** En el paciente 37 se aprecian 3 lecturas en donde está presente la variante, sin embargo, este paciente fue catalogado como un heterocigoto al alelo de referencia en el análisis de los datos de WES. En las lecturas también se observa una deleción consistente de 12 pb río arriba de rs1044301916 **(B)** En el paciente 41, no hay ninguna lectura que registre la variante. El recuadro rojo ubica la posición de rs1044301916. Para cada paciente se desglosa el número total de lecturas que cubren dicha posición en la secuenciación y la base nucleotídica que fue detectada.

Por el contrario, en los pacientes marcados con color rosa, como el paciente 41 (Figura 12B), no hay ninguna asignación del genotipo de baja calidad, por lo que, la razón más probable para la asignación errónea en el genotipo durante el ensayo de RFLPs es que para estas muestras en el gel se exhiba una digestión incompleta del producto de PCR, pese a que el tiempo de incubación sea el máximo recomendado por el fabricante.

Con estas observaciones se concluye que la técnica de RFLPs para la variante rs1044301916 en *PABPC3*, haciendo uso de la enzima BstP1 no es la estrategia idónea para la genotificación de muestras dentro de la cohorte de descubrimiento, y para su empleo como estrategia en próximos ensayos con cohortes más grandes. Una alternativa para la genotificación en la genotificación mediante qPCR utilizando TaqMan SNP Genotyping Assays (Applied Biosystems™, Thermo Fisher Scientific Corporation).

8.8. Análisis *in silico* de las proteínas de tipo silvestre y las mutantes

Trasladar los resultados de un estudio genómico a un enfoque funcional es esencial para entender cómo las variantes genéticas identificadas en el estudio de asociación pueden impactar en la función de las proteínas y, en última instancia, en el mecanismo molecular subyacente que aumenta el riesgo de enfermedad. Comenzar con un enfoque bioinformático es recomendable, ya que proporciona una manera rápida de evaluar cómo la mutación podría influir en la estructura de la proteína, antes de proceder con experimentos funcionales. En este trabajo se usaron diferentes herramientas de Chimera X (Goddard et al., 2018; Pettersen et al., 2021) para visualizar los cambios que promueven las variantes rs1044301916 en PABPC3, rs147153778 en SLCO4A1 y rs2273499 en ARFGAP1.

8.8.1. Variante rs1044301916 en PABPC3 (p.L276H)

La proteína PABPC3 tiene cuatro motivos de reconocimiento de RNA (RRM) con los que interacciona con la cola de poliA de los RNA mensajeros (**Figura 15A**). El motivo C-terminal de las proteínas de unión a poliA (PABC) reconoce 15 residuos del dominio de interacción PAM-2 que se encuentra en proteínas indispensables para su función como el factor de terminación de la traducción eucariota, eRF3 (Féral et al., 2001).

La mutación rs1044301916 genera el cambio de una Leucina por una Histidina en la posición 276 de la proteína respecto al amino terminal. Este cambio se localiza dentro de una región con estructura secundaria definida, una alfa-hélice que funciona como un enlace entre RRM3 y RRM4 (**Figura 15B**).

La simulación de rotámeros de ChimeraX analiza las diferentes orientaciones tridimensionales que puede adoptar el grupo lateral de un aminoácido en una proteína. Para este caso específico, la Histidina puede formar puentes de hidrógeno con el grupo lateral de la Arginina 272 en ciertas posiciones (**Figura 15C**). Estas interacciones pueden influir en la estabilidad de la alfa-hélice.

La variante conlleva el cambio de un aminoácido alifático y pequeño por uno cargado positivamente a valores de pH fisiológico. El punto hidrofóbico amarillo que señala la flecha roja en la proteína de tipo silvestre desaparece en la proteína mutante (**Figura 15D**).

Cabe recalcar, que, a pesar de la diferencia en las cargas entre los aminoácidos involucrados en la mutación, el cambio en el potencial electrostático no exhibe un cambio drástico (**Figura 15E**). No obstante, es importante recordar que esta proteína establece interacciones con la cola poli(A) de los mRNA a través de sus dominios RRM (Féral et al., 2001). Dado que el RNA lleva una carga negativa debido a la disposición de los fosfatos en la estructura de sus nucleótidos hacia el exterior de la doble hélice (Lipfert et al., 2014), su interacción con proteínas ocurre cuando reconoce regiones estructurales con superficies electrostáticas positivas en la proteína que complementan su carga (Shazman & Mandel-Gutfreund, 2008).

En este contexto, la variante rs1044301916, que incrementa el potencial electrostático neto en una región expuesta al solvente, como es el caso de esta alfa-hélice, podría influir en la dinámica de la interacción entre RNA y PABPC3. Aunque es tentador asumir que esta variante fortalece la interacción y estabiliza el complejo, dicha suposición debe ser confirmada mediante experimentos. Una vía para evaluar esta dinámica es a través de ensayos de retardo de movilidad electroforética (EMSA) (Hellman & Fried, 2007). Además, herramientas de modelado como AlphaFold (Jumper et al., 2021) pueden ser valiosas para simular estas interacciones.

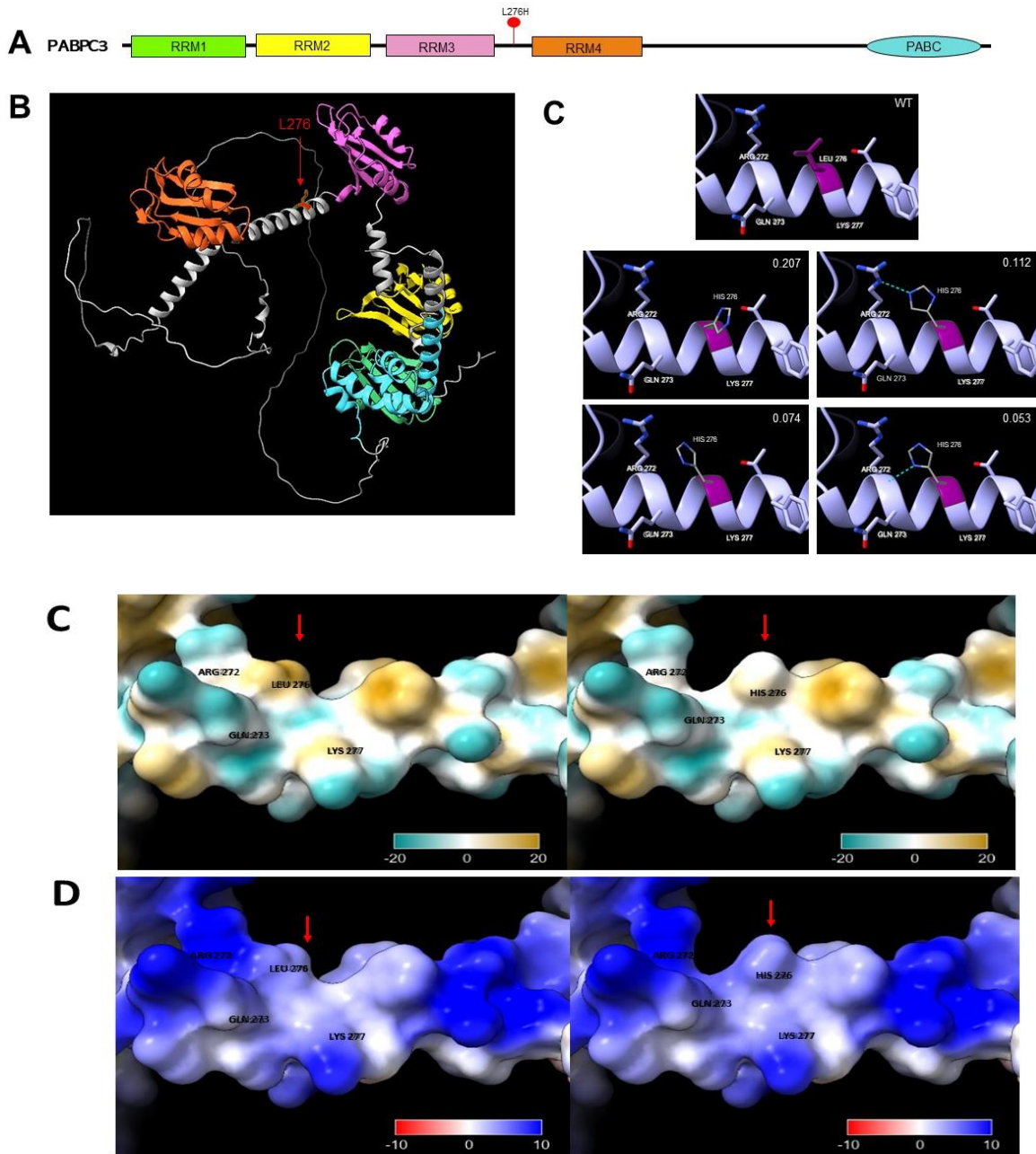


Figura 15. Comparación de la estructura de la proteína PABPC3 de tipo silvestre con la mutante. **(A)** Esquema de la proteína señalando sus dominios funcionales. RRM: RNA-recognition motif, PABC: poly(A)-binding protein C-terminal. **(B)** Esquema tridimensional que señala con colores los dominios funcionales señalados en (A). En rojo el residuo de Leucina 276. **(C)** Comparación estructural de la interacción de la proteína de tipo silvestre (arriba) y la mutante (abajo) con los aminoácidos adyacentes. Para la mutante se muestran las diferentes posiciones en las que podría orientarse el aminoácido en el espacio. El número señala la probabilidad de encontrar dicha orientación (prevalencia del rotámero). Los puentes de hidrógeno predichos se señalan con líneas punteadas azules. **(D)** Modelo de superficie de hidrofobicidad de la proteína de tipo silvestre (izquierdo) y la mutante (derecho). **(E)** Modelo de superficie que señala el potencial electrostático de la proteína de tipo silvestre (izquierdo) y la mutante (derecho). Gráficos y análisis moleculares realizados con UCSF ChimeraX, desarrollados por Resource for Biocomputing, Visualization, and Informatics de la Universidad de California, San Francisco.

8.8.2. Variante rs147153778 en SLCO4A1 (p.R450W)

SLCO4A1 es un transportador de aniones orgánicos que tiene un motivo Kazal de la familia de los inhibidores de serin-proteasas y dos regiones intrínsecamente desordenadas (UniPROT ID E1P5H9). La predicción de la estructura tridimensional posiciona al residuo de Arginina 450 en el extremo de un dominio transmembranal que está proyectado a la cara citoplasmática (**Figura 16B**), una posición en la que puede interaccionar de forma electrostática con el ácido glutámico de la posición 694 que forma parte de la región desordenada del C-terminal (**Figura 16C**).

La variante rs147153778 ocasiona la sustitución de una arginina por un triptófano, planteando como uno de los posibles desafíos la distorsión estérica. La simulación de rotámeros señala que este aminoácido provoca múltiples interacciones desfavorables debido a la proximidad de los átomos (marcados con líneas naranjas en la **Figura 16C**). En última instancia, el triptófano, por su tamaño voluminoso, puede inducir la reorientación de los núcleos cercanos al residuo 450, lo que conlleva un impacto significativo en el empaquetamiento de la proteína.

Además, se ha reemplazado un aminoácido básico, por uno sin carga y altamente hidrofóbico, como se observa en el patrón de superficie hidrofóbico y el foco amarillo en la **Figura 16D**. Cuando pensamos en como esto puede influir en la función de SLOC4A1, la relación más probable es que esta modificación puede impactar directamente en la manera en la que este transportador interacciona con las moléculas que acarrea.

SLOC4A1 interviene en el transporte de Na⁺, sales biliares, ácidos orgánicos, hormonas tiroideas, estrona-3-sulfato, estradiol-17beta-glucurónido y prostaglandina E2 (Stieger & Hagenbuch, 2014). Dicho de otra forma, iones o moléculas cargadas que no pueden abrirse paso a través de la membrana plasmática paso por difusión pasiva (Cooper, 2000).

A lo largo de la literatura, se habla de cómo los residuos cargados en las asas que se proyectan hacia el citoplasma de un transportador desempeñan un papel fundamental en su función, pues estas cargas pueden atraer o repeler selectivamente iones con cargas opuestas en el citoplasma, facilitando su unión y transporte a través del transportador. O bien determinar su interacción con otras proteínas reguladoras (Miloshevsky & Jordan, 2004; Ruprecht & Kunji, 2021; Takemura et al., 2009).

Como se dijo, la variante rs147153778 cambia un residuo cargado positivamente por uno altamente hidrofóbico, este cambio podría y esto podría reducir la afinidad por las moléculas que transporta y por ende, propiciar un desbalance en las concentraciones intra/extracelulares de estas moléculas con efectos graves dentro de la célula.

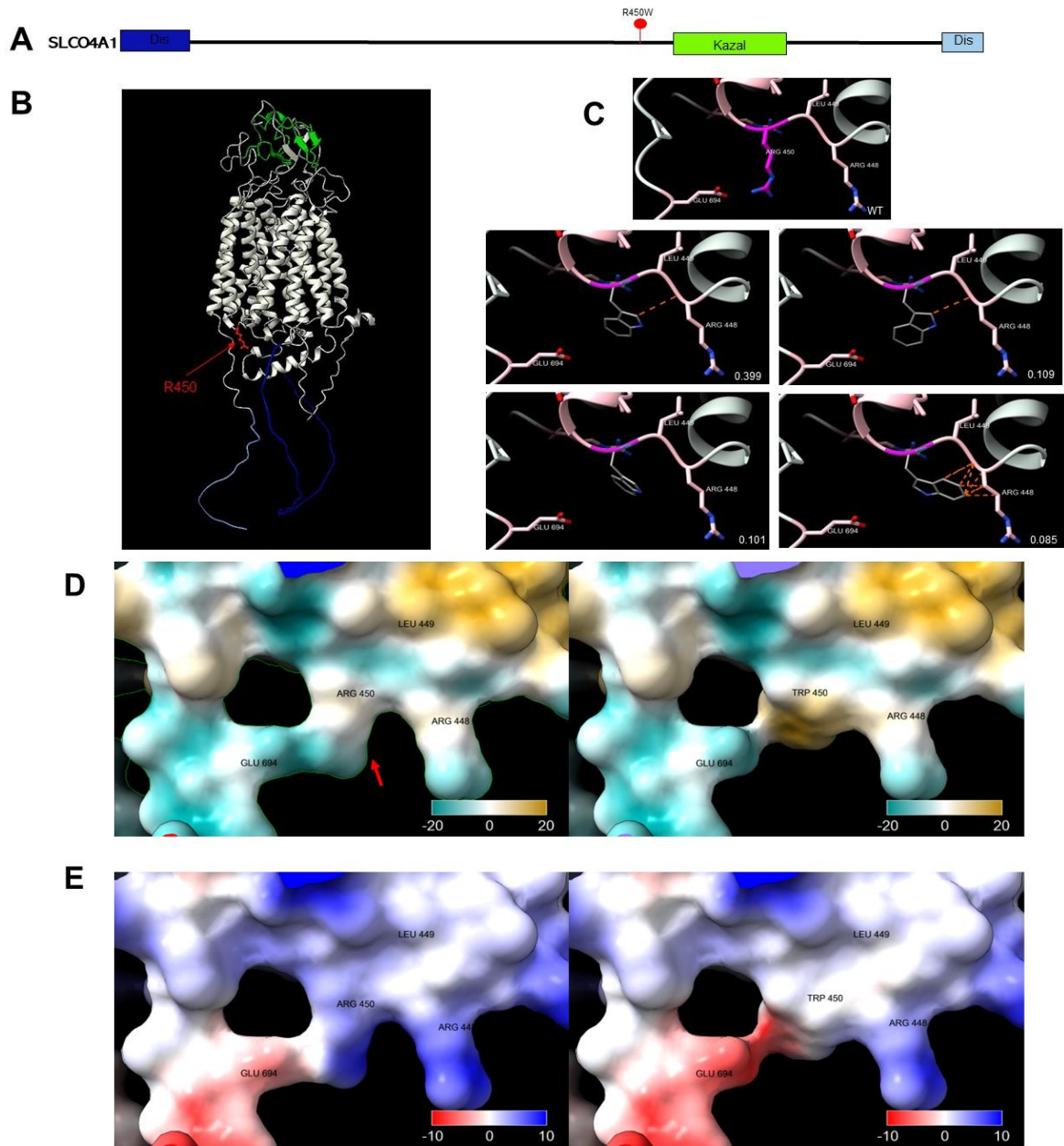


Figura 16. Comparación de la estructura de la proteína SLCO4A1 de tipo silvestre con la mutante. **(A)** Esquema de la proteína señalando sus dominios funcionales. Kazal: motivo Kazal de la familia de los inhibidores de serin-proteasas, Dis: región desordenada. **(B)** Esquema tridimensional que señala con colores los dominios funcionales señalados en (A). En rojo el residuo de Arginina 450. **(C)** Comparación estructural de la interacción de la proteína de tipo silvestre (arriba) y la mutante (abajo) con los aminoácidos adyacentes. Para la mutante se muestran las diferentes posiciones en las que podría orientarse el aminoácido en el espacio. El número señala la probabilidad de encontrar dicha orientación (prevalencia del rotámero). Las posiciones que podrían generar una colisión entre átomos se señalan con líneas punteadas naranjas. **(D)** Modelo de superficie de hidrofobicidad de la proteína de tipo silvestre (izquierdo) y la mutante (derecho). **(E)** Modelo de superficie que señala el potencial electrostático de la proteína de tipo silvestre (izquierdo) y la mutante (derecho).

8.8.3. Variante rs2273499 en ARFGAP1 (p.V184M)

ARFGAP1 es una proteína activadora de GTPasa del factor de ribosilación ADP con un dominio establecido para dicha función señalado de naranja en la **Figura 17A**. El residuo 184 en la proteína de tipo silvestre corresponde a un aminoácido alifático y poco voluminoso, la valina. Y está posicionado en una zona que no parece tener una estructura secundaria definida y espacialmente parece estar alejado del dominio Arf-GAP (**Figura 17B**).

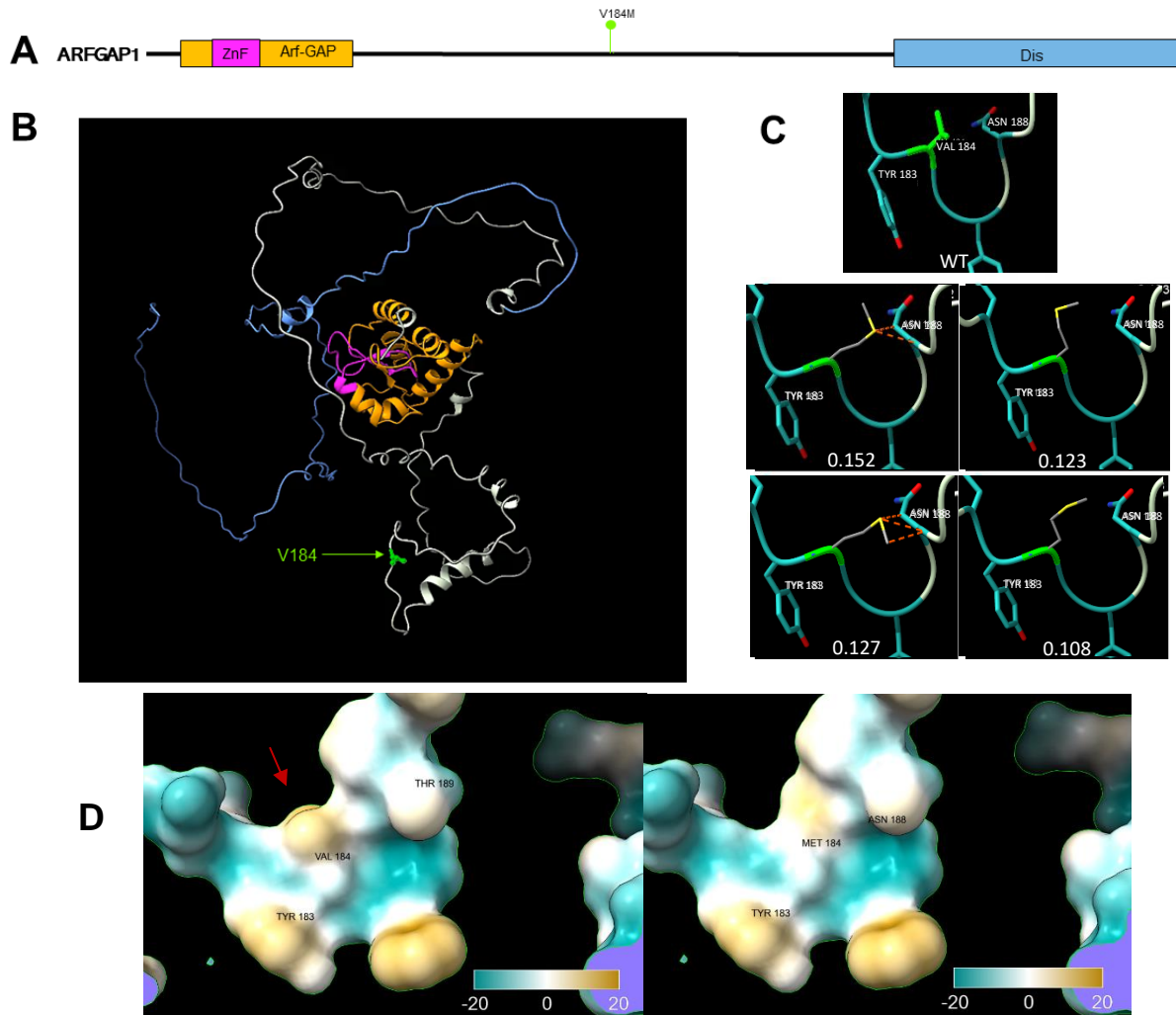


Figura 17. Comparación de la estructura de la proteína ARFGAP1 de tipo silvestre con la mutante. **(A)** Esquema de la proteína señalando sus dominios funcionales. Arf-GAP: Dominio de proteínas activadoras de GTPasa. ARF GTPase-activating proteins domain, ZnF: Dedos de zinc de tipo C4, Dis: región desordenada. **(B)** Esquema tridimensional que señala con colores los dominios funcionales señalados en (A). En verde el residuo de Valina 184. **(C)** Comparación estructural de la interacción de la proteína de tipo silvestre (arriba) y la mutante (abajo) con los aminoácidos adyacentes. Para la mutante se muestran las diferentes posiciones en las que podría orientarse el aminoácido en el espacio. El número señala la probabilidad de encontrar dicha orientación (prevalencia del rotámero). Las posiciones que podrían generar una colisión entre átomos se señalan con líneas punteadas naranjas. **(D)** Modelo de superficie de hidrofobicidad de la proteína de tipo silvestre (izquierdo) y la mutante (derecho).

En esta mutante, la valina es reemplazada por una metionina, que es un aminoácido más grande que la valina. En ciertas orientaciones el grupo lateral podría potencialmente generar colisiones con los átomos de la cadena lateral del Aspartato 188 que está próximo en la simulación (**Figura 17C**), y con ello, alterar la estructura de la proteína.

En contraste con las dos variantes pasadas, las diferencias fisicoquímicas entre la valina y la metionina no son lo suficientes para alterar de manera importante la superficie hidrofóbica de esa región de la molécula (**Figura 17D**). Sin embargo, la metionina es un aminoácido que tiene propiedades únicas, el azufre en su estructura se puede oxidar, convirtiendo la metionina en sulfóxido de metionina (Lim et al., 2019). Al igual que con la fosforilación, la oxidación y reducción mediada enzimáticamente de residuos de metionina específicos funciona como un proceso regulador en la célula, la alteración del equilibrio metionina-sulfóxido en las proteínas ya se ha correlacionado con procesos patológicos, entre ellos cáncer (De Luca et al., 2010; Lim et al., 2019).

Sin duda, el hecho de que ARFGAP1 sea un regulador negativo de la proteína mTORC1 (Yaba et al., 2016), es el vínculo más significativo que se puede encontrar entre esta proteína y el desarrollo del cáncer. En este sentido, un enfoque bioinformático podría ser útil para saber si existe la posibilidad de que la mutación rs2273499 interviene en la forma en la que estas dos proteínas interactúan.

En resumen, se compararon los modelos de las proteínas silvestres con sus contrapartes mutantes para tres de las variantes candidatas del panel propuesto en este estudio: rs1044301916 en *PABPC3*, rs147153778 en *SLCO4A1* y rs2273499 en *ARFGAP1*. En primer lugar, se observó que tanto la variante en *PABPC3* como en *SLCO4A1* propician cambios puntuales en el carácter hidrofóbico y el potencial electrostático de las proteínas, lo que podría potencialmente alterar sus interacciones con otras biomoléculas. En segundo lugar, ninguna de las variantes se localizó en alguno de los dominios caracterizados de la proteína; no obstante, esto no descarta la posibilidad de que tengan un efecto en su función. Todas estas observaciones deben ser validadas experimentalmente para tener evidencia fundamental que permita avanzar más allá de las observaciones bioinformáticas y estadísticas, y brindar una comprensión más profunda de cómo la variante específica puede influir en la función de la proteína y, en última instancia, en el desarrollo de la enfermedad.

9. COMENTARIOS FINALES

Actualmente, las tecnologías de secuenciación de alto rendimiento son indispensables para la detección a gran escala de las mutaciones relacionadas con distintas enfermedades humanas (Kanzi et al., 2020; Kiezun et al., 2012). No obstante, en los estudios de asociación de variantes raras, los investigadores deben ser cuidadosos con sus conclusiones, ya que un valor de p significativo no constituye evidencia suficiente para establecer causalidad de una variante con un fenotipo (Bansal et al., 2010; Cardon & Palmer, 2003; Kiezun et al., 2012; Lee et al., 2014; MacArthur et al., 2014).

Debido a que los estudios con cohortes pequeñas no tienen el poder estadístico suficiente para asegurar la asociación, es esencial que los hallazgos derivados de los datos de secuenciación sean rigurosamente replicados en cohortes más grandes. Es crucial que la asociación genética sea sustentada por un análisis integral que combine evidencia genética y computacional, además de una evaluación experimental de las consecuencias estructurales y funcionales de cada variante en el producto génico para entender su relación con la enfermedad (Cardon & Palmer, 2003; Kiezun et al., 2012; Lee et al., 2014; MacArthur et al., 2014)

El estudio de los TCGT es muy retador. La oncogénesis es compleja, hay una importante heterogeneidad molecular e histológica (Chovanec & Cheng, 2022; Lobo et al., 2019; Martin et al., 2022; Shen et al., 2018) y cuando se habla del riesgo, las variantes comunes tienen una contribución modesta en el desarrollo de la enfermedad (Litchfield et al., 2016; Pluta et al., 2021). A todo esto, se suma el hecho de que la prevalencia de los tumores es menor en comparación con otras enfermedades, lo que dificulta la formación de cohortes lo suficientemente grandes para cumplir con los requisitos estadísticos en las fases de replicación.

Es importante destacar que, a pesar de las limitaciones de tamaño muestral en nuestro estudio los resultados obtenidos en este trabajo tienen un papel crucial de fase exploratoria y de generación de hipótesis. La propuesta de un panel de variantes candidatas, permite guiar futuras investigaciones de manera más efectiva y pragmática en términos de costos.

Entendemos que nuestro trabajo analiza una muestra institucional con 40 pacientes, que no es representativa de las características genéticas de toda la población mestizo-mexicana, pero resaltamos su valor al ser la cohorte más extensa de pacientes mexicanos con TCGT secuenciados hasta la fecha.

En este sentido, enfatizamos la urgencia de ampliar la representatividad de diversas poblaciones en los esfuerzos globales de secuenciación. Los datos de poblaciones poco estudiadas pueden constituir valiosas adiciones a las bases de datos genéticas, como gnomAD (Gudmundsson et al., 2022). A medida que se recolectan y analizan más datos, es posible que todas las VUS, comunes en estas poblaciones subrepresentadas, sean reclasificadas según su impacto clínico (Moghadas et al., 2016). En el futuro, es posible que este conocimiento sobre el riesgo genético se implemente en estrategias de detección temprana que beneficien a nuestra población.

10. CONCLUSIÓN

En la población mexicana existen al menos 5 variantes en diferentes genes con un alto potencial a ser asociadas al riesgo a desarrollar tumores de células germinales de testículo.

- I. La variante rs1044301916 (T>A) en *PABPC3*, que induce el cambio del aminoácido L276H en su producto funcional
- II. La variante rs2273499 (G>A) en *ARFGAP1*, que induce el cambio del aminoácido V131M en su producto funcional.
- III. La variante rs147153778 (C>T) en *SLCO4A1*, que provoca el cambio del aminoácido R450W en su producto funcional.
- IV. La variante rs79783591 (A>T) en *MC4R*, que conlleva el cambio del aminoácido I269N en su producto funcional.
- V. La variante rs1234654104 (TA>T) en *EHTM1*, la cual afecta un sitio de *splicing* alternativo al inicio del exón 4.

Ninguna de estas variantes había sido previamente asociada a los TCGT, lo que confirma que existen diferencias genómicas que influyen en el desarrollo y la respuesta terapéutica de los pacientes con TCGT entre distintas poblaciones.

Por su parte, la prueba de agrupación múltiple de variantes a nivel gen reveló que el gen *PABPC3*, cuya expresión es específica en testículos, está significativamente asociado a los TCGT bajo un modelo dominante ($p < 2.5 \times 10^{-6}$).

La función de estos genes y las pruebas que los vinculan con diversos procesos relacionados con cáncer y espermatogénesis refuerzan la importancia que pueden tener en el desarrollo de los TCGT. Por lo tanto, este estudio enfatiza la necesidad imperante de continuar investigando estas variantes en una segunda cohorte de pacientes TCGT de mayor dimensión para validar nuestros resultados, confirmar su asociación con la enfermedad y proponer su uso como biomarcadores de riesgo dentro del entorno clínico.

11. PERSPECTIVAS

- La perspectiva más importante de este estudio reside en la genotipificación de las variantes propuestas en nuestro panel en una segunda cohorte de mayor tamaño compuesta por pacientes con tumores de células germinales de testículo. Este proceso tiene como propósito verificar los resultados obtenidos en nuestro estudio y ratificar la trascendencia de las variantes en relación con el riesgo de la enfermedad.

Ampliar el tamaño muestral también brinda la oportunidad de explorar posibles correlaciones con otras características clínicas de los pacientes como el tipo de histología, grupo de riesgo o su respuesta al tratamiento, etc. Además, se puede investigar la co-ocurrencia de las variantes en el panel, lo cual podría tener un significado relevante en la formulación de paneles multigénicos para la evaluación del riesgo genético en nuestra población

- Una premisa fundamental en la formulación del problema abordado en este estudio es que en nuestra población se presentan características únicas relacionadas con la susceptibilidad al riesgo de TCGT. Sin embargo, en este punto, no podemos afirmar que las variantes presentes en el panel sean exclusivas de nuestra población. Por lo tanto, surge un enfoque intrigante que consiste en examinar los datos disponibles de pacientes con TCGT que pertenecen a grupos con una composición ancestral diferentes a la población mestizo-mexicana. Este análisis permitiría comparar la frecuencia de las variantes entre estas poblaciones y determinar si, efectivamente, existe una mayor prevalencia en individuos de origen mexicano.
- El gen *PABPC3* destacó en los análisis estadísticos como aquel con el mayor potencial a tener una relación con el desarrollo de TCGT. Una prometedora vía de investigación se abre ante nosotros, con el objetivo de caracterizar, tanto a nivel bioinformático como experimental, esta proteína y comprender en detalle cómo la mutación ejerce su influencia sobre su función y las interacciones que establece con los mRNA.

12. ANEXO

12.1. Estimación de la ancestría global a partir de un análisis de componentes principales (PCA).

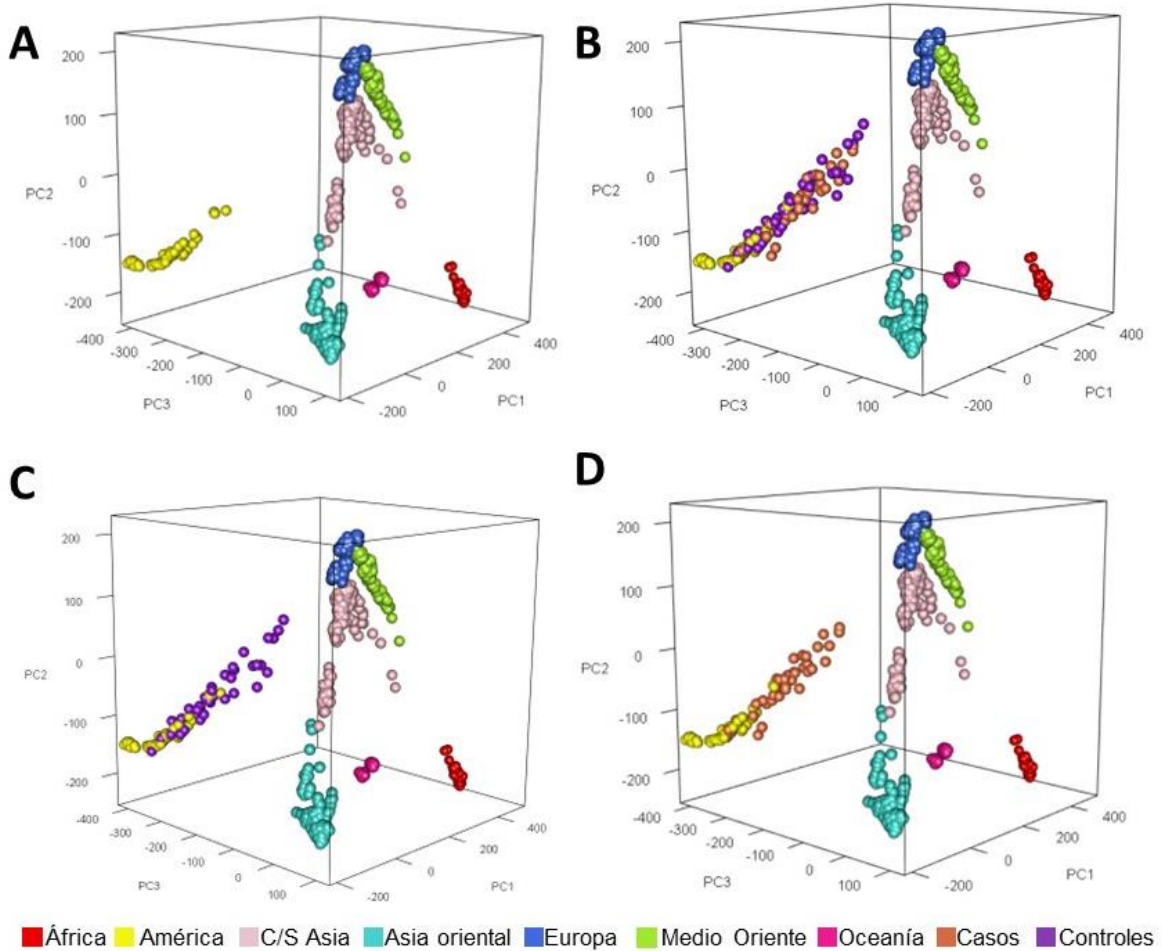


Figura 12. Análisis de ancestría global de los individuos de la cohorte de descubrimiento (Casos y controles) a partir de un PCA. **(A)** Gráfica tridimensional del espacio de ancestría de referencia (EAR), donde cada eje corresponde a los componentes principales (PC). **(B)** Gráfica del EAR añadiendo los datos de la población de estudio. **(C)** Gráfica del EAR añadiendo los datos únicamente del grupo control 1. **(D)** Gráfica del EAR añadiendo los datos únicamente del grupo de casos TCGT. El programa TRACE se usó para calcular la distancia en n dimensiones entre cada elemento del conjunto ($n=3$). Cada esfera corresponde a un individuo y el color está asignado según la población referida en el Panel de Diversidad del Genoma Humano, o el grupo de la población de estudio. Tanto el grupo de casos como el de controles se agrupan con los individuos de referencia americanos del HGDP, indicando que entre ellos hay una mayor semejanza.

12.2. Producción científica de la alumna durante la duración de la maestría.

12.2.1. Carteles presentados en congresos internacionales

Cancer Genetics and Epigenetics Gordon Research Conference en Lucca, Italia (22-25 de Mayo, 2023)

Exome sequencing in Mexican men identifies potential risk germline variants to develop testicular germ cell tumors.

Cuevas-Estrada B.^{1*}, Montalvo-Casimiro M.¹, Ríos-Rodríguez JA.¹, Cervera-Taboada A.², Wegman-Ostrosky T.³, Alcaraz N.⁴, Jiménez-Ríos MA.⁵, Herrera LA^{1,2} and González-Barrios R^{1**}.

* ✉ bere.cuevas26@gmail.com; ** ✉ rodrigo@ciencias.unam.mx

INTRODUCTION

Testicular germ cell tumors (TGCT) are the most common tumors affecting young men. Although advancements in therapy have significantly lowered mortality rates globally, it is disconcerting that low-incident and low-income countries continue to experience disproportionately higher mortality rates compared to their high-incident, high-income regions (Singh et al., 2021).

Estimated age-standardized INCIDENCE rates in 2020 (GLOBOCAN, 2020)

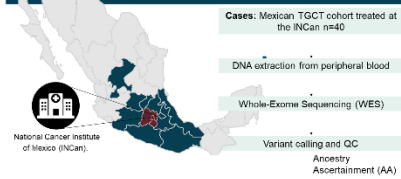


Estimated age-standardized INCIDENCE rates in 2020 (GLOBOCAN, 2020)



This is a multifactorial problem, and one possibility to explain this disparity is associated with differences in genetic alterations. At present, ~86 independent risk loci are related with TGCT (Pluta et al., 2021), but unfortunately, few samples from Hispanic patients have been included in those studies. The objective of this study is to identify germline susceptibility variants in the largest cohort of Mexican TGCT patients to date, which may be associated to an increased risk of developing TGCT.

METHODS



Filtering and selection of Rare variants:
1. Pathogenic or likely pathogenic (ClinVar)
2. Variants of uncertain significance with CADD_p ≥ 25 || REVEL_s ≥ 0.7

2817 variants
Controls₁: Healthy Mexican Volunteers; n=36 (Fabian-Monreal et al., 2021)

Case/control association analysis: Single Variant Tests

2817 variants
Controls₂: Latin/Admixed American individuals in gnomAD non cancer database n=17,00

Case/control association analysis: 1. Single Variant Tests 2. Gene-Based Aggregation Test of Multiple Variants

- Affiliations**
- Unidad de Investigación Biomédica en Cáncer, Instituto Nacional de Cancerología-Instituto de Investigaciones Biomédicas, UNAM, México
 - Instituto Nacional de Medicina Genómica, México
 - Departamento de Investigación, Instituto Nacional de Cancerología
 - Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark
 - Departamento de Urología, Instituto Nacional de Cancerología, México
 - Templeton de Monterrey, Escuela de Medicina y Ciencias de la Salud, México

RESULTS

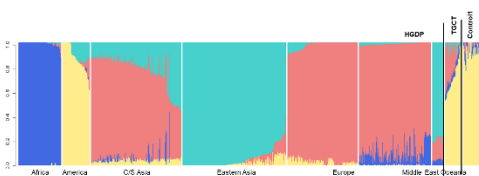


Figure 1. All individuals in the discovery cohort (CASES and CONTROLS₁) have an ancestral composition that corresponds to what is expected for the current Mexican mestizo population. To estimate the ancestral proportions for each individual, ADMIXTURE analysis was conducted with k=4. The reference panel for ancestral populations utilized in this study was obtained from the Human Genome Diversity Project (HGDP) (Li et al., 2008).

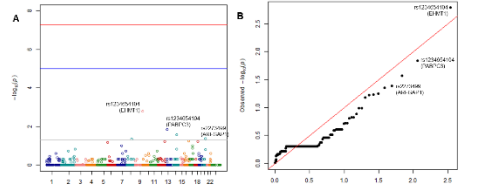


Figure 2. Variants in EHM1, PABPC3 and ARGAP1 have p-values<0.05 in the individual variant association analysis (Cases vs Controls₁). (A) Manhattan plot and (B) QQ-plot from a standard case/control association analysis using the two-tailed Fisher's exact test for the 2817 selected variants.

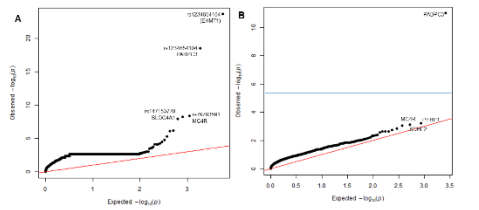


Figure 3. QQ plot illustrating the results of a standard case/control association analysis of 2817 selected variants. In panel (A), the analysis is conducted at the single variant level, while in panel (B), a Gene-Based Aggregation Test (assess multiple variants collectively). The allelic frequency of these variants in 40 Mexican patients is compared to the frequencies observed in the exomes of 17,286 Latin individuals in the non-cancer gnomAD v2.1 database. The blue line on the y-axis indicates exome-wide significance at $P = 2.5 \times 10^{-6}$. Among the analyzed genes, only the PABPC3 gene surpasses this significance threshold, indicating a potential association with TGCT.

CHR	POS	REF	ALT	GENE	FUNCTION	CASES(N=40)		CONTROLS ₂ (N=17,296)		CONTROLS ₁ (N=36)			
						AF	OR(95% CI)	p	AF	OR(95% CI)	p	AF	OR(95% CI)
19	13746306	T	C	EHM1	SPONDY	0.123	0.000	10.1(1.2-78.2)	0.002	9.8E-03	109.9(20.3-647.2)	2.3E-04	8.4E-21
13	35097058	T	A	PABPC3	p1.776A1	0.088	0.000	6.8(0.8-243.9)	0.014	0.0E-00	3258.5(195.4-173458.0)	2.9E-10	8.1E-16
23	62787918	G	A	ARFGAP1	p1131M	0.123	0.028	3.3(1.0-10.7)	0.041	2.4E-02	5.7(2.5-13.4)	7.8E-07	2.2E-09
18	60271544	A	T	MOG4	p1790N	0.066	0.000	3.0(0.9-295.5)	0.059	7.4E-03	9.3(4.2-24.8)	3.8E-09	1.1E-05
20	62666451	C	T	SLOC4A1	pR450W	0.088	0.028	2.2(0.7-13.1)	0.157	1.2E-02	8.0(3.9-19.8)	5.7E-09	1.6E-05

PABPC3 is a gene that codes for a poly(A)-binding protein that is specifically expressed in the testis. This type of protein is involved in the stability and maintenance of long lived mRNA translation potential (Féral et al., 2001).

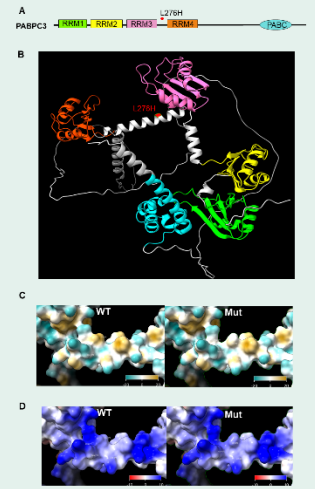


Figure 4. Comparison of the structure of the wild type PABPC3 protein with the mutant for the rs1044301916 variant. (A) Schematic representation of the protein indicating its functional domains. RRM: RNA-recognition motif, PABC: poly(A)-binding protein C-terminal. (B) 3D diagram of the protein, with color-coded representations of the functional domains mentioned in panel (A). The residue of Leucine 276 is highlighted in red. The surface model of hydrophobicity (C) and electrostatic potential (D) are depicted for both the wild-type and mutant proteins.

Table 1. Summary results for five SNPs showing the minor p values in the combined association analysis (1:Cases vs Controls₁, 2: Cases vs Controls₂)

SNP	AF	OR(95% CI)	p
1	0.123	10.1(1.2-78.2)	0.002
2	0.088	6.8(0.8-243.9)	0.014
3	0.123	3.3(1.0-10.7)	0.041
4	0.066	3.0(0.9-295.5)	0.059
5	0.088	2.2(0.7-13.1)	0.157

CONCLUSION

The variant rs1044301916 in the PABPC3 gene that causes a change in pL276H in the functional product is a potential variant for the development of TGCT mainly because it encodes an important protein for male germ cell development. It is suggested that the variants reported in our cohort for the EHM1, SLOC4A1, ARFGAP and MOG4 genes are interesting targets for study as TGCT susceptibility variants in Mexican population.

These findings also emphasize the need for a major representativeness of different ethnic groups in worldwide sequencing efforts.

REFERENCES

- Fabian-Monreal, E. et al. (2021) Frontiers in Psychiatry 12
- Féral, C., Guellin, G. and Pawlak, A. (2001) Nucleic Acids Research 29(9)
- Li, J. Z. et al. (2008) Science, 319(5865)
- Pluta, J. et al. (2021) Nature Communications, 12(1)
- Singh, H. et al. (2019) Cancer Drug Resistance, 2(3)

ACKNOWLEDGMENT

This work was supported by the Instituto Nacional de Cancerología (INCan) and Consejo Nacional de Ciencia y Tecnología (CONACyT), grant no. 290041
B.C.-E. acknowledges the Programa de Maestría y Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM), and CONACyT for the fellowship 1145895

12.2.2. Artículos publicados en revistas indexadas.



Review

Breaking the Mold: Epigenetics and Genomics Approaches Addressing Novel Treatments and Chemoresponse in TGCT Patients

Berenice Cuevas-Estrada ^{1,†} , Michel Montalvo-Casimiro ^{1,†} , Paulina Munguia-Garza ¹,
Juan Alberto Ríos-Rodríguez ¹, Rodrigo González-Barrios ^{1,*} and Luis A. Herrera ^{1,2,*}

¹ Unidad de Investigación Biomédica en Cáncer, Instituto Nacional de Cancerología-Instituto de Investigaciones Biomédicas, UNAM, Mexico City 14080, Mexico

² Tecnológico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Monterrey 64710, Mexico

* Correspondence: rodrigop@ciencias.unam.mx (R.G.-B.); herreram@biomedicas.unam.mx (L.A.H.)

† These authors contributed equally to this work.

Abstract: Testicular germ-cell tumors (TGCT) have been widely recognized for their outstanding survival rates, commonly attributed to their high sensitivity to cisplatin-based therapies. Despite this, a subset of patients develops cisplatin resistance, for whom additional therapeutic options are unsuccessful, and ~20% of them will die from disease progression at an early age. Several efforts have been made trying to find the molecular bases of cisplatin resistance. However, this phenomenon is still not fully understood, which has limited the development of efficient biomarkers and precision medicine approaches as an alternative that could improve the clinical outcomes of these patients. With the aim of providing an integrative landscape, we review the most recent genomic and epigenomic features attributed to chemoresponse in TGCT patients, highlighting how we can seek to combat cisplatin resistance through the same mechanisms by which TGCTs are particularly hypersensitive to therapy. In this regard, we explore ongoing treatment directions for resistant TGCT and novel targets to guide future clinical trials. Through our exploration of recent findings, we conclude that epidrugs are promising treatments that could help to restore cisplatin sensitivity in resistant tumors, shedding light on potential avenues for better prognosis for the benefit of the patients.

Keywords: TGCT; cisplatin; chemoresponse; epidrugs; sensibility; genomics and epigenomics



Citation: Cuevas-Estrada, B.; Montalvo-Casimiro, M.; Munguia-Garza, P.; Rios-Rodriguez, J.A.; González-Barrios, R.; Herrera, L.A. Breaking the Mold: Epigenetics and Genomics Approaches Addressing Novel Treatments and Chemoresponse in TGCT Patients. *Int. J. Mol. Sci.* **2023**, *24*, 7873. <https://doi.org/10.3390/ijms24097873>

Academic Editors: Anna Perri and Sabrina Bossio

Received: 31 March 2023

Revised: 19 April 2023

Accepted: 19 April 2023

Published: 26 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Testicular cancer (TCa) is a relatively rare neoplasm on a per-population basis; it accounts for 1 to 2% of all neoplasms in men, with a GLOBOCAN estimate of 74,458 new cases and 9334 deaths in 2020 [1]. TCa incidence has been rising over the past 35 years, and it is expected to continue this way, especially in highly populated countries, due to their population structure [2]. Racial differences have always been noted, with rates among young white men being >10 times higher than those observed in black or Asian mean. However, there is a rising tendency among all ethnicities [3].

Testicular germ cell tumors (TGCT) comprise 95% of TCa cases, but their etiology is not fully understood. Nevertheless, the fact that their prevalence varies across different populations implies that environmental influences or genetic factors might be involved. In fact, the best-characterized risk factors for TGCT include cryptorchidism, hypospadias, infertility, gonadal dysgenesis, testicular microlithiasis, family history of TGCT, previous diagnosis of TCa, and exposure to environmental factors [4].

Over time, TGCTs have been recognized as having a therapeutic model due to their outstanding sensitivity to cisplatin-based treatment, which has yet to be replicated in other types of cancer. Survival rates have increased from less than 30% in the 1950s to approximately 95% today [5]. Even patients with disseminated TCa achieve up to 80%

13. REFERENCIAS BIBLIOGRÁFICAS

- Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., Van Auken, K., ... Westerfield, M. (2023). The Gene Ontology knowledgebase in 2023. *GENETICS*, 224(1). <https://doi.org/10.1093/genetics/iyad031>
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655. <https://doi.org/10.1101/GR.094052.109>
- American Society of Clinical Oncology (ASCO). (2021, febrero 5). *Cáncer de testículo—Síntomas y signos*. Cancer.Net. <https://www.cancer.net/es/tipos-de-c%C3%A1ncer/c%C3%A1ncer-de-test%C3%ADculo/s%C3%ADntomas-y-signos>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, 25(1), 25-29. <https://doi.org/10.1038/75556>
- Asimit, J., & Zeggini, E. (2010). Rare Variant Association Analysis Methods for Complex Traits. *Annual Review of Genetics*, 44(1), 293-308. <https://doi.org/10.1146/annurev-genet-102209-163421>
- Bagrodia, A., Lee, B. H., Lee, W., Cha, E. K., Sfakianos, J. P., Iyer, G., Pietzak, E. J., Gao, S. P., Zabor, E. C., Ostrovnaya, I., Kaffenberger, S. D., Syed, A., Arcila, M. E., Chaganti, R. S., Kundra, R., Eng, J., Hreiki, J., Vacic, V., Arora, K., ... Feldman, D. R. (2016). Genetic Determinants of Cisplatin Resistance in Patients With Advanced Germ Cell Tumors. *Journal of Clinical Oncology*, 34(33), 4000-4007. <https://doi.org/10.1200/JCO.2016.68.7798>
- Ban, M. J., Ji, S. H., Lee, C. K., Bae, S. B., Kim, H. J., Ahn, T. S., Lee, M. S., Baek, M. J., & Jeong, D. (2017). Solute carrier organic anion transporter family member 4A1 (SLCO4A1) as a prognosis marker of colorectal cancer. *Journal of cancer research and clinical oncology*, 143(8), 1437-1447. <https://doi.org/10.1007/S00432-017-2393-7>
- Bansal, V., Libiger, O., Torkamani, A., & Schork, N. J. (2010). Statistical analysis strategies for association studies involving rare variants. *Nature Reviews Genetics* 2010 11:11, 11(11), 773-785. <https://doi.org/10.1038/nrg2867>
- Baroni, T., Arato, I., Mancuso, F., Calafiore, R., & Luca, G. (2019). On the origin of testicular germ cell tumors: From gonocytes to testicular cancer. *Frontiers in Endocrinology*, 10(JUN), 1-8. <https://doi.org/10.3389/fendo.2019.00343>
- Batool, A., Karimi, N., Wu, X. N., Chen, S. R., & Liu, Y. X. (2019). Testicular germ cell tumor: A comprehensive review. *Cellular and Molecular Life Sciences*, 76(9), 1713-1727. <https://doi.org/10.1007/s00018-019-03022-7>
- Beyer, J., Collette, L., Sauvé, N., Daugaard, G., Feldman, D. R., Tandstad, T., Tryakin, A., Stahl, O., Gonzalez-Billalabeitia, E., De Giorgi, U., Culine, S., De Wit, R., Hansen, A. R., Bebek, M., Terbuch, A., Albany, C., Hentrich, M., Gietema, J. A., Negaard, H., ... for the International Germ Cell Cancer Classification Update Consortium. (2021). Survival and New Prognosticators in Metastatic Seminoma: Results From the IGCCCG-Update Consortium. *Journal of Clinical Oncology*, 39(14), 1553-1562. <https://doi.org/10.1200/JCO.20.03292>
- Bhartiya, D., & Kaushik, A. (2021). Testicular Stem Cell Dysfunction Due to Environmental Insults Could Be Responsible for Deteriorating Reproductive Health of Men. *Reproductive Sciences*, 28(3), 649-658. <https://doi.org/10.1007/s43032-020-00411-3>
- Biermann, K., Heukamp, L. C., Steger, K., Zhou, H., Franke, F. E., Sonnack, V., Brehm, R., Berg, J., Bastian, P. J., Müller, S. C., Wang-Eckert, L., & Buettner, R. (2007). Genome-wide expression profiling reveals new insights into pathogenesis and progression of testicular germ cell tumors. *Cancer Genomics and Proteomics*, 4(5), 359-367.
- Boublikova, L., Buchler, T., Stary, J., Abrahamova, J., & Trka, J. (2014). Molecular biology of testicular germ cell tumors: Unique features awaiting clinical application. *Critical Reviews in Oncology/Hematology*, 89(3), 366-385. <https://doi.org/10.1016/J.CRITREVONC.2013.10.001>
- Brierley, J., Gospodarowicz, M. K., & Wittekind, C. (Eds.). (2017). *TNM classification of malignant tumours* (Eighth edition). John Wiley & Sons, Inc.

- Buljubašić, R., Buljubašić, M., Bojanac, A. K., Ulamec, M., Vlahović, M., Ježek, D., Bulić-Jakuš, F., & Sinčić, N. (2018). Epigenetics and testicular germ cell tumors. *Gene*, 661, 22-33. <https://doi.org/10.1016/j.gene.2018.03.072>
- Cai, Q., Chen, Y., Zhang, D., Pan, J., Xie, Z., Xu, C., Li, S., Zhang, X., Gao, Y., Hou, J., Guo, X., Zhou, X., Zhang, B., Ma, F., Zhang, W., Lin, G., Xin, Z., Niu, Y., & Wang, Y. (2020). Estimates of over-time trends in incidence and mortality of testicular cancer from 1990 to 2030. *Translational Andrology and Urology*, 9(2), 182-195. <https://doi.org/10.21037/tau.2020.02.22>
- Cannon, S., Williams, M., Gunning, A. C., & Wright, C. F. (2023). Evaluation of in silico pathogenicity prediction tools for the classification of small in-frame indels. *BMC Medical Genomics*, 16(1), 1-9. <https://doi.org/10.1186/S12920-023-01454-6/FIGURES/3>
- Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., & Lewis, S. (2009). AmiGO: online access to ontology and annotation data. *Bioinformatics*, 25(2), 288-289. <https://doi.org/10.1093/bioinformatics/btn615>
- Cardon, L. R., & Palmer, L. J. (2003). Population stratification and spurious allelic association. *The Lancet*, 361(9357), 598-604. [https://doi.org/10.1016/S0140-6736\(03\)12520-2](https://doi.org/10.1016/S0140-6736(03)12520-2)
- Chen, X., Yi, G., Zhou, Y., Hu, W., Xi, L., Han, W., & Wang, F. (2023). Prognostic Biomarker SLCO4A1 Is Correlated with Tumor Immune Infiltration in Colon Adenocarcinoma. *Mediators of Inflammation*, 2023, e4926474. <https://doi.org/10.1155/2023/4926474>
- Cheng, L., Albers, P., Berney, D. M., Feldman, D. R., Daugaard, G., Gilligan, T., & Looijenga, L. H. J. (2018). Testicular cancer. *Nature Reviews Disease Primers*, 4(1), Article 1. <https://doi.org/10.1038/s41572-018-0029-0>
- Chovanec, M., & Cheng, L. (2022). Advances in diagnosis and treatment of testicular cancer. *BMJ*, e070499. <https://doi.org/10.1136/bmj-2022-070499>
- Chung, C. C., Kanetsky, P. A., Wang, Z., Hildebrandt, M. A. T., Koster, R., Skotheim, R. I., Kratz, C. P., Turnbull, C., Cortessis, V. K., Bakken, A. C., Bishop, D. T., Cook, M. B., Erickson, R. L., Fosså, S. D., Jacobs, K. B., Korde, L. A., Kraggerud, S. M., Lothe, R. A., Loud, J. T., ... Nathanson, K. L. (2013). Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nature Genetics*, 45(6), Article 6. <https://doi.org/10.1038/ng.2634>
- Cooper, G. M. (2000). Cell Membranes. En *The Cell: A Molecular Approach. 2nd edition*. Sinauer Associates. <https://www.ncbi.nlm.nih.gov/books/NBK9928/>
- Costa, A. L., Lobo, J., Jerónimo, C., & Henrique, R. (2017). The epigenetics of testicular germ cell tumors: Looking for novel disease biomarkers. *Epigenomics*, 9(2), 155-169. <https://doi.org/10.2217/epi-2016-0081>
- Cuevas-Estrada, B., Montalvo-Casimiro, M., Munguia-Garza, P., Ríos-Rodríguez, J. A., González-Barrios, R., & Herrera, L. A. (2023). Breaking the Mold: Epigenetics and Genomics Approaches Addressing Novel Treatments and Chemoresistance in TGCT Patients. *International Journal of Molecular Sciences*, 24(9), 7873. <https://doi.org/10.3390/ijms24097873>
- Dai, X., Fu, G., Zhao, S., & Zeng, Y. (2021). Statistical Learning Methods Applicable to Genome-Wide Association Studies on Unbalanced Case-Control Disease Data. *Genes*, 12(5), 736. <https://doi.org/10.3390/genes12050736>
- De Luca, A., Sanna, F., Sallèse, M., Ruggiero, C., Grossi, M., Sacchetta, P., Rossi, C., De Laurenzi, V., Di Ilio, C., & Favaloro, B. (2010). Methionine sulfoxide reductase A down-regulation in human breast cancer cells results in a more aggressive phenotype. *Proceedings of the National Academy of Sciences*, 107(43), 18628-18633. <https://doi.org/10.1073/pnas.1010171107>
- De Toni, L., Šabovic, I., Cosci, I., Ghezzi, M., Foresta, C., & Garolla, A. (2019). Testicular cancer: Genes, environment, hormones. *Frontiers in Endocrinology*, 10(JUL), 1-11. <https://doi.org/10.3389/fendo.2019.00408>
- Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., & Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics*, 11(6), Article 6. <https://doi.org/10.1038/nrg2809>
- Elliot, D. (2003). Pathways of post-transcriptional gene regulation in mammalian germ cell development. *Cytogenetic and Genome Research*, 103(3-4), 210-216. <https://doi.org/10.1159/000076806>
- Elzinga-Tinke, J., Dohle, G., & Looijenga, L. (2015). Etiology and early pathogenesis of malignant testicular germ cell tumors: Towards possibilities for preinvasive diagnosis. *Asian Journal of Andrology*, 17(3), 381. <https://doi.org/10.4103/1008-682X.148079>

- Federici, G., & Soddu, S. (2020). Variants of uncertain significance in the era of high-throughput genome sequencing: A lesson from breast and ovary cancers. *Journal of Experimental and Clinical Cancer Research*, 39(1), 1-12. <https://doi.org/10.1186/S13046-020-01554-6/FIGURES/1>
- Fénichel, P., & Chevalier, N. (2019). Is Testicular Germ Cell Cancer Estrogen Dependent? The Role of Endocrine Disrupting Chemicals. *Endocrinology*, 160(12), 2981-2989. <https://doi.org/10.1210/en.2019-00486>
- Féral, C., Guellaën, G., & Pawlak, A. (2001). Human testis expresses a specific poly(A)-binding protein. *Nucleic Acids Research*, 29(9), 1872-1883. <https://doi.org/10.1093/nar/29.9.1872>
- Ferry, K. V., Hamilton, T. C., & Johnson, S. W. (2000). Increased nucleotide excision repair in cisplatin-resistant ovarian cancer cells: Role of ERCC1-XPF. *Biochemical Pharmacology*, 60(9), 1305-1313. [https://doi.org/10.1016/s0006-2952\(00\)00441-x](https://doi.org/10.1016/s0006-2952(00)00441-x)
- Fujiki, R., Ikeda, M., Yoshida, A., Akiko, M., Yao, Y., Nishimura, M., Matsushita, K., Ichikawa, T., Tanaka, T., Morisaki, H., Morisaki, T., & Ohara, O. (2018). Assessing the Accuracy of Variant Detection in Cost-Effective Gene Panel Testing by Next-Generation Sequencing. *The Journal of Molecular Diagnostics*, 20(5), 572-582. <https://doi.org/10.1016/j.jmoldx.2018.04.004>
- Georgiopoulos, G., & Evangelou, E. (2016). Power considerations for λ inflation factor in meta-analyses of genome-wide association studies. *Genetics research*, 98, e9. <https://doi.org/10.1017/S0016672316000069>
- Gillessen, S., Sauv e, N., Collette, L., Daugaard, G., de Wit, R., Albany, C., Tryakin, A., Fizazi, K., Stahl, O., Gietema, J. A., De Giorgi, U., Cafferty, F. H., Hansen, A. R., Tandstad, T., Huddart, R. A., Necchi, A., Sweeney, C. J., Garcia-Del-Muro, X., Heng, D. Y. C., ... International Germ Cell Cancer Classification Update Consortium. (2021). Predicting Outcomes in Men With Metastatic Nonseminomatous Germ Cell Tumors (NSGCT): Results From the IGCCCG Update Consortium. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 39(14), 1563-1574. <https://doi.org/10.1200/JCO.20.03296>
- Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., & Ferrin, T. E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science*, 27(1), 14-25. <https://doi.org/10.1002/pro.3235>
- González-Barrios, R., Alcaraz, N., Montalvo-Casimiro, M., Cervera, A., Arriaga-Canon, C., Munguia-Garza, P., Hinojosa-Ugarte, D., Sobrevilla-Moreno, N., Torres-Arciga, K., Mendoza-Perez, J., Diaz-Chavez, J., Cortes-González, C. C., Castro-Hernández, C., Martínez-Cedillo, J., Scavuzzo, A., Pérez-Montiel, D., Jiménez-Ríos, M. A., & Herrera, L. A. (2022). Genomic Profile in a Non-Seminoma Testicular Germ-Cell Tumor Cohort Reveals a Potential Biomarker of Sensitivity to Platinum-Based Therapy. *Cancers*, 14(9). <https://doi.org/10.3390/cancers14092065>
- Goss, P. E., Lee, B. L., Badovinac-Crnjevic, T., Strasser-Weippl, K., Chavarri-Guerra, Y., Louis, J. S., Villarreal-Garza, C., Unger-Saldaña, K., Ferreyra, M., Debiasi, M., Liedke, P. E. R., Touya, D., Werutsky, G., Higgins, M., Fan, L., Vasconcelos, C., Cazap, E., Vallejos, C., Mohar, A., ... Azenha, G. (2013). Planning cancer control in Latin America and the Caribbean. *The Lancet Oncology*, 14(5), 391-436. [https://doi.org/10.1016/S1470-2045\(13\)70048-2](https://doi.org/10.1016/S1470-2045(13)70048-2)
- Gudmundsson, S., Singer-Berk, M., Watts, N. A., Phu, W., Goodrich, J. K., Solomonson, M., Rehm, H. L., MacArthur, D. G., & O'Donnell-Luria, A. (2022). Variant interpretation using population databases: Lessons from gnomAD. *Human Mutation*, 43(8), 1012. <https://doi.org/10.1002/HUMU.24309>
- Guo, M. H., Plummer, L., Chan, Y. M., Hirschhorn, J. N., & Lippincott, M. F. (2018). Burden Testing of Rare Variants Identified through Exome Sequencing via Publicly Available Control Data. *American Journal of Human Genetics*, 103(4), 522-534. <https://doi.org/10.1016/j.ajhg.2018.08.016>
- Gurney, J. K., Florio, A. A., Znaor, A., Ferlay, J., Laversanne, M., Sarfati, D., Bray, F., & McGlynn, K. A. (2019). International Trends in the Incidence of Testicular Cancer: Lessons from 35 Years and 41 Countries. *European Urology*, 76(5), 615-623. <https://doi.org/10.1016/j.eururo.2019.07.002>
- Hacioglu, B. M., Kodaz, H., Erdogan, B., Cinkaya, A., Tastekin, E., Hacibekiroglu, I., Turkmen, E., Kostek, O., Genc, E., Uzunoglu, S., & Cicin, I. (2017). K-RAS and N-RAS mutations in testicular germ cell tumors. *Bosnian Journal of Basic Medical Sciences*. <https://doi.org/10.17305/bjbms.2017.1764>
- Hellman, L. M., & Fried, M. G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols*, 2(8), 1849-1861. <https://doi.org/10.1038/nprot.2007.249>
- Hemminki, K., & Li, X. (2004). Familial risk in testicular cancer as a clue to a heritable and environmental aetiology. *British Journal of Cancer*, 90(9), 1765. <https://doi.org/10.1038/SJ.BJC.6601714>

- Henrie, A., Hemphill, S. E., Ruiz-Schultz, N., Cushman, B., DiStefano, M. T., Azzariti, D., Harrison, S. M., Rehm, H. L., & Eilbeck, K. (2018). ClinVar Miner: Demonstrating Utility of a Web-Based Tool for Viewing and Filtering ClinVar Data. *Human mutation*, 39(8), 1051-1060. <https://doi.org/10.1002/humu.23555>
- Herrera, A., García, F. M., Gil, R., Jiménez, X., Hernández, Á., & Aboharp, Z. (2014). Frecuencia de cáncer en un hospital de tercer nivel de la Ciudad de México. Implicaciones de la detección oportuna. *Cirugía y Cirujanos*, 82, 28-37.
- Hersmus, R., Van Bever, Y., Wolffenbuttel, K. P., Biermann, K., Cools, M., & Looijenga, L. H. J. (2017). The biology of germ cell tumors in disorders of sex development: DSD; gonadal histology and cancer risk. *Clinical Genetics*, 91(2), 292-301. <https://doi.org/10.1111/cge.12882>
- Hodis, E., Watson, I. R., Kryukov, G. V., Arold, S. T., Imielinski, M., Theurillat, J. P., Nickerson, E., Auclair, D., Li, L., Place, C., Dicara, D., Ramos, A. H., Lawrence, M. S., Cibulskis, K., Sivachenko, A., Voet, D., Saksena, G., Stransky, N., Onofrio, R. C., ... Chin, L. (2012). A landscape of driver mutations in melanoma. *Cell*, 150(2), 251-263. <https://doi.org/10.1016/j.cell.2012.06.024>
- Hong, C. S., Singh, L. N., Mullikin, J. C., & Biesecker, L. G. (2016). Assessing the reproducibility of exome copy number variations predictions. *Genome Medicine*, 8(1), 82. <https://doi.org/10.1186/s13073-016-0336-6>
- Iglesias-Ortega, L., Megías-Fernández, C., Domínguez-Giménez, P., Jimeno-González, S., & Rivero, S. (2023). Cell consequences of loss of function of the epigenetic factor EHMT1. *Cellular Signalling*, 108, 110734. <https://doi.org/10.1016/j.cellsig.2023.110734>
- INEGI. (2021). *Comunicado De Prensa Núm. 105/ Estadísticas a Propósito Del Día Mundial Contra El Cáncer (4 De Febrero)*. 2018, 1-11.
- Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., Cannon-Albright, L. A., Teerlink, C. C., Stanford, J. L., Isaacs, W. B., Xu, J., Cooney, K. A., Lange, E. M., Schleutker, J., Carpten, J. D., ... Sieh, W. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *American Journal of Human Genetics*, 99(4), 877-885. <https://doi.org/10.1016/j.ajhg.2016.08.016>
- Jiménez-Ríos, M. Á., Solares-Sánchez, M. E., Martínez-Cervera, P. F., Espinosa-Ostos, D., Feria-Bernal, G., Aburto-Morales, S., Francolugo-Vélez, V. A., Zarzosa-Aguilar, J., Silva-Nájera, E. L., Lugo-García, J. A., Viveros-Contreras, C., Pacheco-Gahbler, C., Merizalde-Palomino, W., Flores-Terrazas, E., Pérez-Mendoza, R., Hernández-Toríz, N., Xochipiltécatl-Muñoz, J., Manzanilla-García, H. A., Galicia-Sámano, R., ... Padrón-Lucio, S. (2011). Panorama epidemiológico del cáncer genitourinario en la Zona Centro de México. *Revista Mexicana de Urología*, 71(Suplemento 3), 3-6.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kanzi, A. M., San, J. E., Chimukangara, B., Wilkinson, E., Fish, M., Ramsuran, V., & de Oliveira, T. (2020). Next Generation Sequencing and Bioinformatics Analysis of Family Genetic Inheritance. *Frontiers in Genetics*, 11. <https://www.frontiersin.org/articles/10.3389/fgene.2020.544162>
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., ... Daly, M. J. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020 581:7809, 581(7809), 434-443. <https://doi.org/10.1038/s41586-020-2308-7>
- Kazemi-Sefat, G. E., Keramatipour, M., Talebi, S., Kavousi, K., Sajed, R., Kazemi-Sefat, N. A., & Mousavizadeh, K. (2021). The importance of CDC27 in cancer: Molecular pathology and clinical aspects. *Cancer Cell International*, 21(1), 160. <https://doi.org/10.1186/s12935-021-01860-9>
- Kiezun, A., Garimella, K., Do, R., Stitzel, N. O., Neale, B. M., McLaren, P. J., Gupta, N., Sklar, P., Sullivan, P. F., Moran, J. L., Hultman, C. M., Lichtenstein, P., Magnusson, P., Lehner, T., Shugart, Y. Y., Price, A. L., De Bakker, P. I. W., Purcell, S. M., & Sunyaev, S. R. (2012). Exome sequencing and the genetic basis of complex traits. *Nature Genetics* 2012 44:6, 44(6), 623-630. <https://doi.org/10.1038/NG.2303>
- Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, 46(3), Article 3. <https://doi.org/10.1038/ng.2892>

- Kleene, K. C. (1989). Poly(A) shortening accompanies the activation of translation of five mRNAs during spermiogenesis in the mouse. *Development*, *106*(2), 367-373. <https://doi.org/10.1242/dev.106.2.367>
- Koboldt, D. C. (2020). Best practices for variant calling in clinical sequencing. *Genome Medicine*, *12*(1), 91. <https://doi.org/10.1186/s13073-020-00791-w>
- Kondo, Y., Shen, L., Ahmed, S., Bumber, Y., Sekido, Y., Haddad, B. R., & Issa, J.-P. J. (2008). Downregulation of Histone H3 Lysine 9 Methyltransferase G9a Induces Centrosome Disruption and Chromosome Instability in Cancer Cells. *PLoS ONE*, *3*(4), e2037. <https://doi.org/10.1371/journal.pone.0002037>
- Kristiansen, W., Karlsson, R., Rounge, T. B., Whittington, T., Andreassen, B. K., Magnusson, P. K., Fosså, S. D., Adami, H.-O., Turnbull, C., Haugen, T. B., Grotmol, T., & Wiklund, F. (2015). Two new loci and gene sets related to sex determination and cancer progression are associated with susceptibility to testicular germ cell tumor. *Human Molecular Genetics*, *24*(14), 4138-4146. <https://doi.org/10.1093/hmg/ddv129>
- Kumar, N., & Singh, A. K. (2021). The anatomy, movement, and functions of human sperm tail: An evolving mystery. *Biology of Reproduction*, *104*(3), 508-520. <https://doi.org/10.1093/biolre/iaaa213>
- Lakpour, N., Saliminejad, K., Ghods, R., Reza Sadeghi, M., Pilatz, A., Khosravi, F., & Madjd, Z. (2021). Potential biomarkers for testicular germ cell tumour: Risk assessment, diagnostic, prognostic and monitoring of recurrence. *Andrologia*, *53*(4), 1-17. <https://doi.org/10.1111/and.13998>
- Lazcano-Ponce, E., Salazar-Martínez, E., & Hernández-Avila, M. (2001). Estudios epidemiológicos de casos y controles. Fundamento teórico, variantes y aplicaciones. *Salud Pública de México*, *43*(2), 135-150.
- Lee, S., Abecasis, G. R., Boehnke, M., & Lin, X. (2014). Rare-variant association analysis: Study designs and statistical tests. *American Journal of Human Genetics*, *95*(1), 5-23. <https://doi.org/10.1016/j.ajhg.2014.06.009>
- Lee, S., & Lee, D. K. (2018). What is the proper way to apply the multiple comparison test? *Korean Journal of Anesthesiology*, *71*(5), 353-360. <https://doi.org/10.4097/kja.d.18.00242>
- Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., Cann, H. M., Barsh, G. S., Feldman, M., Cavalli-Sforza, L. L., & Myers, R. M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, *319*(5866), 1100-1104. https://doi.org/10.1126/SCIENCE.1153717/SUPPL_FILE/LI_SOM.PDF
- Li, Y., Vinckenbosch, N., Tian, G., Huerta-Sanchez, E., Jiang, T., Jiang, H., Albrechtsen, A., Andersen, G., Cao, H., Korneliussen, T., Grarup, N., Guo, Y., Hellman, I., Jin, X., Li, Q., Liu, J., Liu, X., Sparsø, T., Tang, M., ... Wang, J. (2010). Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nature Genetics*, *42*(11), Article 11. <https://doi.org/10.1038/ng.680>
- Lim, J. M., Kim, G., & Levine, R. L. (2019). Methionine in Proteins: It's Not Just for Protein Initiation Anymore. *Neurochemical Research*, *44*(1), 247-257. <https://doi.org/10.1007/s11064-017-2460-0>
- Lipfert, J., Doniach, S., Das, R., & Herschlag, D. (2014). Understanding Nucleic Acid-Ion Interactions. *Annual Review of Biochemistry*, *83*(1), 813-841. <https://doi.org/10.1146/annurev-biochem-060409-092720>
- Litchfield, K., Levy, M., Huddart, R. A., Shipley, J., & Turnbull, C. (2016). The genomic landscape of testicular germ cell tumours: From susceptibility to treatment. *Nature Reviews Urology*, *13*(7), 409-419. <https://doi.org/10.1038/nrurol.2016.107>
- Litchfield, K., Levy, M., Orlando, G., Loveday, C., Law, P. J., Migliorini, G., Holroyd, A., Broderick, P., Karlsson, R., Haugen, T. B., Kristiansen, W., Nsengimana, J., Fenwick, K., Assiotis, I., Kote-Jarai, Zs., Dunning, A. M., Muir, K., Peto, J., Eeles, R., ... Turnbull, C. (2017). Identification of 19 new risk loci and potential regulatory mechanisms influencing susceptibility to testicular germ cell tumor. *Nature Genetics*, *49*(7), 1133-1140. <https://doi.org/10.1038/ng.3896>
- Litchfield, K., Sultana, R., Renwick, A., Dudakia, D., Seal, S., Ramsay, E., Powell, S., Elliott, A., Warren-Perry, M., Eeles, R., Peto, J., Kote-Jarai, Z., Muir, K., Nsengimana, J., Stratton, M. R., Easton, D. F., Bishop, D. T., Huddart, R. A., Rahman, N., & Turnbull, C. (2015). Multi-stage genome-wide association study identifies new susceptibility locus for testicular germ cell tumour on chromosome 3q25. *Human Molecular Genetics*, *24*(4), 1169-1176. <https://doi.org/10.1093/hmg/ddu511>
- Litchfield, K., Summersgill, B., Yost, S., Sultana, R., Labreche, K., Dudakia, D., Renwick, A., Seal, S., Al-Saadi, R., Broderick, P., Turner, N. C., Houlston, R. S., Huddart, R., Shipley, J., & Turnbull, C. (2015). Whole-exome sequencing reveals the mutational spectrum of testicular germ cell tumours. *Nature Communications*, *6*, 1-8. <https://doi.org/10.1038/ncomms6973>
- Litchfield, K., Thomsen, H., Mitchell, J. S., Sundquist, J., Houlston, R. S., Hemminki, K., & Turnbull, C. (2015). Quantifying the heritability of testicular germ cell tumour using both population-based and genomic approaches. *Scientific Reports 2015 5:1*, *5*(1), 1-7. <https://doi.org/10.1038/srep13889>

- Lobo, J., Gillis, A. J. M., Jerónimo, C., Henrique, R., & Looijenga, L. H. J. (2019). Human germ cell tumors are developmental cancers: Impact of epigenetics on pathobiology and clinic. *International Journal of Molecular Sciences*, *20*(2), 1-28. <https://doi.org/10.3390/ijms20020258>
- Lobo, J., Jerónimo, C., & Henrique, R. (2020). Cisplatin resistance in testicular germ cell tumors: Current challenges from various perspectives. *Cancers*, *12*(6), 1-26. <https://doi.org/10.3390/cancers12061601>
- Loveday, C., Litchfield, K., Proszek, P. Z., Cornish, A. J., Santo, F., Levy, M., Macintyre, G., Holryod, A., Broderick, P., Dudakia, D., Benton, B., Bakir, M. A., Hiley, C., Grist, E., Swanton, C., Huddart, R., Powles, T., Chowdhury, S., Shipley, J., ... Turnbull, C. (2020). Genomic landscape of platinum resistant and sensitive testicular cancers. *Nature Communications*, *11*(1), 2189. <https://doi.org/10.1038/s41467-020-15768-x>
- MacArthur, D. G., Manolio, T. A., Dimmock, D. P., Rehm, H. L., Shendure, J., Abecasis, G. R., Adams, D. R., Altman, R. B., Antonarakis, S. E., Ashley, E. A., Barrett, J. C., Biesecker, L. G., Conrad, D. F., Cooper, G. M., Cox, N. J., Daly, M. J., Gerstein, M. B., Goldstein, D. B., Hirschhorn, J. N., ... Gunter, C. (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature*, *508*(7497), 469-476. <https://doi.org/10.1038/nature13127>
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorf, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., ... Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), Article 7265. <https://doi.org/10.1038/nature08494>
- Martin, F. C., Conduit, C., & Loveland, K. L. (2022). *Genetics of testicular cancer: A review*. 481-487. <https://doi.org/10.1097/MOU.0000000000001017>
- Martínez-Cortés, G., Salazar-Flores, J., Gabriela Fernández-Rodríguez, L., Rubi-Castellanos, R., Rodríguez-Loya, C., Velarde-Félix, J. S., Franciso Muñoz-Valle, J., Parra-Rojas, I., & Rangel-Villalobos, H. (2012). Admixture and population structure in Mexican-Mestizos based on paternal lineages. *Journal of Human Genetics*, *57*(9), 568-574. <https://doi.org/10.1038/jhg.2012.67>
- Martínez-Sánchez, Y. L., Escudero-De Los Ríos, P. M., Arias-Flores, R., & Barrios-Bautista, F. (2013). Epidemiología del cáncer en pacientes adultos del Hospital de Oncología del Centro Médico Siglo XXI, Instituto Mexicano del Seguro Social. *Cirugía y Cirujanos*, *81*(6), 508-516.
- Meng, D., Yang, Q., Melick, C. H., Park, B. C., Hsieh, T., Curukovic, A., Jeong, M., Zhang, J., James, N. G., & Jewell, J. L. (2021). ArfGAP1 inhibits mTORC1 lysosomal localization and activation. *The EMBO Journal*, *40*(12), e106412. <https://doi.org/10.15252/embj.2020106412>
- Miloshevsky, G. V., & Jordan, P. C. (2004). Anion Pathway and Potential Energy Profiles along Curvilinear Bacterial CIC Cl- Pores: Electrostatic Effects of Charged Residues. *Biophysical Journal*, *86*(2), 825-835. [https://doi.org/10.1016/S0006-3495\(04\)74158-2](https://doi.org/10.1016/S0006-3495(04)74158-2)
- Mir, M. C., Pavan, N., & Gonzalgo, M. L. (2016). Current Clinical Applications of Testicular Cancer Biomarkers. *Urologic Clinics of North America*, *43*(1), 119-125. <https://doi.org/10.1016/j.ucl.2015.08.011>
- Moghadas, S., Eccles, D. M., Devilee, P., Vreeswijk, M. P. G., & van Asperen, C. J. (2016). Classification and Clinical Management of Variants of Uncertain Significance in High Penetrance Cancer Predisposition Genes. *Human Mutation*, *37*(4), 331-336. <https://doi.org/10.1002/humu.22956>
- Mohar, A., Frías-Mendivil, M., Suchil-Bernal, L., Mora-Maías, T., & de la Garza, J. G. (1997). Epidemiología descriptiva de cáncer en el Instituto Nacional de Cancerología de México. *Salud Pública de México*, *39*(4), 253-258. <https://doi.org/10.1590/S0036-36341997000400002>
- Möller, M., Schotanus, K., Soyer, J. L., Haueisen, J., Happ, K., Stralucke, M., Happel, P., Smith, K. M., Connolly, L. R., Freitag, M., & Stukenbrock, E. H. (2019). Destabilization of chromosome structure by histone H3 lysine 27 methylation. *PLoS Genetics*, *15*(4), e1008093. <https://doi.org/10.1371/journal.pgen.1008093>
- Momozawa, Y., & Mizukami, K. (2021). Unique roles of rare variants in the genetics of complex diseases in humans. *Journal of Human Genetics*, *66*(1), 11. <https://doi.org/10.1038/S10038-020-00845-2>
- Moreno, A., & Sandoval, K. (2013). Diversidad genómica en México. Pasado indígena y mestizaje. En *Cuicuilco* (Vol. 20, Número 58). Instituto Nacional de Antropología e Historia.
- Moreno-Cabrera, J. M., del Valle, J., Castellanos, E., Feliubadaló, L., Pineda, M., Brunet, J., Serra, E., Capellà, G., Lázaro, C., & Gel, B. (2020). Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. *European Journal of Human Genetics* *2020* *28*:12, *28*(12), 1645-1655. <https://doi.org/10.1038/s41431-020-0675-z>

- Muskens, I. S., De Smith, A. J., Zhang, C., Hansen, H. M., Morimoto, L., Metayer, C., Ma, X., Walsh, K. M., & Wiemels, J. L. (2020). Germline cancer predisposition variants and pediatric glioma: A population-based study in California. *Neuro-Oncology*, 22(6), 864-874. <https://doi.org/10.1093/neuonc/noaa014>
- Nachiyappan, A., Gupta, N., & Taneja, R. (2022). EHMT1/EHMT2 in EMT, cancer stemness and drug resistance: Emerging evidence and mechanisms. *The FEBS Journal*, 289(5), 1329-1351. <https://doi.org/10.1111/febs.16334>
- Nakhaei-Rad, S., Soleimani, Z., Vahedi, S., & Gorjini, Z. (2023). Testicular germ cell tumors: Genomic alternations and RAS-dependent signaling. *Critical Reviews in Oncology/Hematology*, 183(January), 103928. <https://doi.org/10.1016/j.critrevonc.2023.103928>
- Neale, B. M., Rivas, M. A., Voight, B. F., Altshuler, D., Devlin, B., Orho-Melander, M., Kathiresan, S., Purcell, S. M., Roeder, K., & Daly, M. J. (2011). Testing for an Unusual Distribution of Rare Variants. *PLoS Genetics*, 7(3), e1001322. <https://doi.org/10.1371/journal.pgen.1001322>
- Nicu, A.-T., Medar, C., Chifiriu, M. C., Gradisteanu Pircalabioru, G., & Burlibasa, L. (2022). Epigenetics and Testicular Cancer: Bridging the Gap Between Fundamental Biology and Patient Care. *Frontiers in Cell and Developmental Biology*, 10, 861995. <https://doi.org/10.3389/fcell.2022.861995>
- Nie, X., Munyoki, S. K., Sukhwani, M., Schmid, N., Missel, A., Emery, B. R., DonorConnect, Stukenborg, J.-B., Mayerhofer, A., Orwig, K. E., Aston, K. I., Hotaling, J. M., Cairns, B. R., & Guo, J. (2022). Single-cell analysis of human testis aging and correlation with elevated body mass index. *Developmental Cell*, 57(9), 1160-1176.e5. <https://doi.org/10.1016/j.devcel.2022.04.004>
- O'Connor, T. D., Kiezun, A., Bamshad, M., Rich, S. S., Smith, J. D., Turner, E., Project, N. E. S., ESP Population Genetics, S. A. W. G., Leal, S. M., & Akey, J. M. (2013). Fine-Scale Patterns of Population Stratification Confound Rare Variant Association Tests. *PLOS ONE*, 8(7), e65834. <https://doi.org/10.1371/journal.pone.0065834>
- Országhová, Z., Kalavska, K., Mego, M., & Chovanec, M. (2022). Overcoming Chemotherapy Resistance in Germ Cell Tumors. *Biomedicines*, 10(5), Article 5. <https://doi.org/10.3390/biomedicines10050972>
- Ozturk, S., Sozen, B., Uysal, F., Bassorgun, I. C., Usta, M. F., Akkoyunlu, G., & Demir, N. (2016). The poly(A)-binding protein genes, EPAB, PABPC1, and PABPC3 are differentially expressed in infertile men with non-obstructive azoospermia. *Journal of Assisted Reproduction and Genetics*, 33(3), 335-348. <https://doi.org/10.1007/s10815-016-0654-z>
- Palmer, N. O., Bakos, H. W., Fullston, T., & Lane, M. (2012). Impact of obesity on male fertility, sperm function and molecular composition. *Spermatogenesis*, 2(4), 253-263. <https://doi.org/10.4161/spmg.21362>
- Park, J. S., Kim, J., Elghiaty, A., & Ham, W. S. (2018). Recent global trends in testicular cancer incidence and mortality. *Medicine (United States)*, 97(37). <https://doi.org/10.1097/MD.00000000000012390>
- Paumard-Hernández, B., Calvete, O., Inglada Pérez, L., Tejero, H., Al-Shahrour, F., Pita, G., Barroso, A., Carlos Triviño, J., Urioste, M., Valverde, C., González Billalabeitia, E., Quiroga, V., Francisco Rodríguez Moreno, J., Fernández Aramburo, A., López, C., Maroto, P., Sastre, J., José Juan Fita, M., Duran, I., ... Benítez, J. (2018). Whole exome sequencing identifies PLEC, EXO5 and DNAH7 as novel susceptibility genes in testicular cancer. *International Journal of Cancer*, 143(8), 1954-1962. <https://doi.org/10.1002/ijc.31604>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1), 70-82. <https://doi.org/10.1002/pro.3943>
- Pluta, J., Pyle, L. C., Nead, K. T., Wilf, R., Li, M., Mitra, N., Weathers, B., D'Andrea, K., Almstrup, K., Anson-Cartwright, L., Benitez, J., Brown, C. D., Chanock, S., Chen, C., Cortessis, V. K., Ferlin, A., Foresta, C., Gamulin, M., Gietema, J. A., ... Kubisch, C. (2021). Identification of 22 susceptibility loci associated with testicular germ cell tumors. *Nature Communications*, 12(1). <https://doi.org/10.1038/s41467-021-24334-y>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*, 81(3), 559-575.
- Rabbani, B., Tekin, M., & Mahdieh, N. (2014). The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics*, 59(1), Article 1. <https://doi.org/10.1038/jhg.2013.114>
- Rajpert-De Meyts, E., McGlynn, K. A., Okamoto, K., Jewett, M. A. S., & Bokemeyer, C. (2016). Testicular germ cell tumours. *The Lancet*, 387(10029), 1762-1774. [https://doi.org/10.1016/S0140-6736\(15\)00991-5](https://doi.org/10.1016/S0140-6736(15)00991-5)

- Ranganathan, P., Pramesh, C. S., & Buyse, M. (2016). Common pitfalls in statistical analysis: The perils of multiple testing. *Perspectives in Clinical Research*, 7(2), 106-107. <https://doi.org/10.4103/2229-3485.179436>
- Rapley, E. A., Turnbull, C., Al Olama, A. A., Dermitzakis, E. T., Linger, R., Huddart, R. A., Renwick, A., Hughes, D., Hines, S., Seal, S., Morrison, J., Nsengimana, J., Deloukas, P., Rahman, N., Bishop, D. T., Easton, D. F., & Stratton, M. R. (2009). A genome-wide association study of testicular germ cell tumor. *Nature Genetics*, 41(7), Article 7. <https://doi.org/10.1038/ng.394>
- Rentzsch, P., Schubach, M., Shendure, J., & Kircher, M. (2021). CADD-Splice—Improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Medicine*, 13(1), 31. <https://doi.org/10.1186/s13073-021-00835-9>
- Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, 47(D1), D886-D894. <https://doi.org/10.1093/nar/gky1016>
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405-424. <https://doi.org/10.1038/gim.2015.30>
- Rizo Ríos, P., Sierra Colindres, M. I., Vázquez Piñon, G., Cano Guadiana, M., Meneses García, A., & Mohar, A. (2007). *Registro Hospitalario de Cáncer: Compendio de Cáncer 2000-2004*.
- Ross, J. P., Dion, P. A., & Rouleau, G. A. (2020). Exome sequencing in genetic disease: Recent advances and considerations. *F1000Research*, 9, 336. <https://doi.org/10.12688/f1000research.19444.1>
- Ruark, E., Seal, S., McDonald, H., Zhang, F., Elliot, A., Lau, K., Perdeaux, E., Rapley, E., Eeles, R., Peto, J., Kote-Jarai, Z., Muir, K., Nsengimana, J., Shipley, J., Bishop, D. T., Stratton, M. R., Easton, D. F., Huddart, R. A., Rahman, N., & Turnbull, C. (2013). Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14. *Nature Genetics*, 45(6), Article 6. <https://doi.org/10.1038/ng.2635>
- Ruprecht, J. J., & Kunji, E. R. S. (2021). Structural Mechanism of Transport of Mitochondrial Carriers. *Annual Review of Biochemistry*, 90(1), 535-558. <https://doi.org/10.1146/annurev-biochem-072820-020508>
- Saha, N., & Muntean, A. G. (2021). Insight into the multi-faceted role of the SUV family of H3K9 methyltransferases in carcinogenesis and cancer progression. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1875(1), 188498. <https://doi.org/10.1016/j.bbcan.2020.188498>
- Salem, M. E., Puccini, A., Grothey, A., Raghavan, D., Goldberg, R. M., Xiu, J., Michael Korn, W., Weinberg, B. A., Hwang, J. J., Shields, A. F., Marshall, J. L., Philip, P. A., & Lenz, H. J. (2018). Landscape of tumor mutation load, mismatch repair deficiency, and PD-L1 expression in a large patient cohort of gastrointestinal cancers. *Molecular Cancer Research*, 16(5), 805-812. <https://doi.org/10.1158/1541-7786.MCR-17-0735/81274/AM/LANDSCAPE-OF-TUMOR-MUTATION-LOAD-MISMATCH-REPAIR>
- Shapovalov, M. V., & Dunbrack, R. L. (2011). A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure*, 19(6), 844-858. <https://doi.org/10.1016/j.str.2011.03.019>
- Sharma, A., Mollier, J., Brocklesby, R. W. K., Caves, C., Jayasena, C. N., & Minhas, S. (2020). Endocrine-disrupting chemicals and male reproductive health. *Reproductive Medicine and Biology*, 19(3), 243-253. <https://doi.org/10.1002/rmb2.12326>
- Shazman, S., & Mandel-Gutfreund, Y. (2008). Classifying RNA-Binding Proteins Based on Electrostatic Properties. *PLoS Computational Biology*, 4(8), e1000146. <https://doi.org/10.1371/journal.pcbi.1000146>
- Sheikh, E., Tran, T., Vranic, S., Levy, A., & Bonfil, R. D. (2022). Role and Significance of c-KIT Receptor Tyrosine Kinase in Cancer: A Review. *Bosnian Journal of Basic Medical Sciences*. <https://doi.org/10.17305/bjbms.2021.7399>
- Shen, H., Shih, J., Hollern, D. P., Wang, L., Bowlby, R., Tickoo, S. K., Thorsson, V., Mungall, A. J., Newton, Y., Hegde, A. M., Armenia, J., Sánchez-Vega, F., Pluta, J., Pyle, L. C., Mehra, R., Reuter, V. E., Godoy, G., Jones, J., Shelley, C. S., ... Wheeler, D. A. (2018). Integrated Molecular Characterization of Testicular Germ Cell Tumors. *Cell Reports*, 23(11), 3392-3406. <https://doi.org/10.1016/j.celrep.2018.05.039>

- Shiba, Y., & Randazzo, P. A. (2012). ArfGAP1 function in COPI mediated membrane traffic: Currently debated models and comparison to other coat-binding ArfGAPs. *Histology and Histopathology*, 27(9), 1143-1153. <https://doi.org/10.14670/HH-27.1143>
- Silva-Zolezzi, I., Hidalgo-Miranda, A., Estrada-Gil, J., Fernandez-Lopez, J. C., Uribe-Figueroa, L., Contreras, A., Balam-Ortiz, E., Del Bosque-Plata, L., Velazquez-Fernandez, D., Lara, C., Goya, R., Hernandez-Lemus, E., Davila, C., Barrientos, E., March, S., & Jimenez-Sanchez, G. (2009). Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proceedings of the National Academy of Sciences of the United States of America*, 106(21), 8611. <https://doi.org/10.1073/PNAS.0903045106>
- Singh, R., Fazal, Z., Freemantle, S. J., & Spinella, M. J. (2021). Between a Rock and a Hard Place: An Epigenetic-Centric View of Testicular Germ Cell Tumors. *Cancers*, 13(7), 1506. <https://doi.org/10.3390/cancers13071506>
- Singla, N., Lafin, J. T., Ghandour, R. A., Kaffenberger, S., Amatruda, J. F., & Bagrodia, A. (2019). Genetics of testicular germ cell tumors. *Current Opinion in Urology*, 29(4), 344-349. <https://doi.org/10.1097/MOU.0000000000000642>
- Singleton, A. B. (2011). Exome sequencing: A transformative technology. *The Lancet Neurology*, 10(10), 942-946. [https://doi.org/10.1016/S1474-4422\(11\)70196-X](https://doi.org/10.1016/S1474-4422(11)70196-X)
- Smith, Z. L., Wertz, R. P., & Eggener, S. E. (2018). Testicular Cancer. *Medical Clinics of North America*, 102(2), 251-264. <https://doi.org/10.1016/j.mcna.2017.10.003>
- Sofer, T., Zheng, X., Laurie, C. A., Gogarten, S. M., Brody, J. A., Conomos, M. P., Bis, J. C., Thornton, T. A., Szpiro, A., O'Connell, J. R., Lange, E. M., Gao, Y., Cupples, L. A., Psaty, B. M., Abe, N., Abecasis, G., Aguet, F., Albert, C., Almas, L., ... Rice, K. M. (2021). Variant-specific inflation factors for assessing population stratification at the phenotypic variance level. *Nature Communications*, 12(1), 3506. <https://doi.org/10.1038/s41467-021-23655-2>
- Stieger, B., & Hagenbuch, B. (2014). Chapter Five—Organic Anion-Transporting Polypeptides. En M. O. Bevensee (Ed.), *Current Topics in Membranes* (Vol. 73, pp. 205-232). Academic Press. <https://doi.org/10.1016/B978-0-12-800223-0.00005-0>
- Sul, J. H., Martin, L. S., & Eskin, E. (2018). Population structure in genetic studies: Confounding factors and mixed models. *PLoS Genetics*, 14(12), e1007309. <https://doi.org/10.1371/journal.pgen.1007309>
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209-249. <https://doi.org/10.3322/caac.21660>
- Takemura, Y., Tamura, N., Imamura, M., & Koyama, N. (2009). Role of the charged amino acid residues in the cytoplasmic loop between putative transmembrane segments 6 and 7 of Na⁺-ATPase of an alkaliphilic bacterium, *Exiguobacterium aurantiacum*. *FEMS Microbiology Letters*, 299(2), 143-148. <https://doi.org/10.1111/j.1574-6968.2009.01740.x>
- Taylor-Weiner, A., Zack, T., O'Donnell, E., Guerriero, J. L., Bernard, B., Reddy, A., Han, G. C., AlDubayan, S., Amin-Mansour, A., Schumacher, S. E., Litchfield, K., Turnbull, C., Gabriel, S., Beroukhim, R., Getz, G., Carter, S. L., Hirsch, M. S., Letai, A., Sweeney, C., & Van Allen, E. M. (2016). Genomic evolution and chemoresistance in germ-cell tumours. *Nature*, 540(7631), 114-118. <https://doi.org/10.1038/nature20596>
- Tennessen, J. A., Bigham, A. W., O'Connor, T. D., Fu, W., Kenny, E. E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., Kang, H. M., Jordan, D., Leal, S. M., Gabriel, S., Rieder, M. J., Abecasis, G., Altshuler, D., Nickerson, D. A., Boerwinkle, E., ... Akey, J. M. (2012). Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Science (New York, N.Y.)*, 337(6090), 64-69. <https://doi.org/10.1126/science.1219240>
- Thompson, E. A. (2013). Identity by descent: Variation in meiosis, across genomes, and in populations. *Genetics*, 194(2), 301-326. <https://doi.org/10.1534/genetics.112.148825>
- Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J., & Richards, J. B. (2018). Genetic architecture: The shape of the genetic contribution to human traits and disease. *Nature Reviews Genetics*, 19(2), Article 2. <https://doi.org/10.1038/nrg.2017.101>
- Trabert, B., Zugna, D., Richiardi, L., McGlynn, K. A., & Akre, O. (2013). Congenital malformations and testicular germ cell tumors. *International Journal of Cancer*, 133(8), 1900-1904. <https://doi.org/10.1002/ijc.28207>

- Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1), Article 1. <https://doi.org/10.1038/s43586-021-00056-9>
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigartyo, C. A.-K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., ... Pontén, F. (2015). Tissue-based map of the human proteome. *Science*, 347(6220), 1260419. <https://doi.org/10.1126/science.1260419>
- van Iterson, M., van Zwet, E. W., BIOS Consortium, & Heijmans, B. T. (2017). Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome biology*, 18(1), 19. <https://doi.org/10.1186/s13059-016-1131-9>
- Van Son, M., Tremoen, N. H., Gaustad, A. H., Våge, D. I., Zeremichael, T. T., Myromslien, F. D., & Grindflek, E. (2020). Transcriptome profiling of porcine testis tissue reveals genes related to sperm hyperactive motility. *BMC Veterinary Research* 2020 16:1, 16(1), 1-18. <https://doi.org/10.1186/S12917-020-02373-9>
- Vázquez-Moreno, M., Locia-Morales, D., Valladares-Salgado, A., Sharma, T., Perez-Herrera, A., Gonzalez-Dzib, R., Rodríguez-Ruiz, F., Wachter-Rodarte, N., Cruz, M., & Meyre, D. (2021). The MC4R p.Ile269Asn mutation confers a high risk for type 2 diabetes in the Mexican population via obesity dependent and independent effects. *Scientific Reports*, 11(1), 3097. <https://doi.org/10.1038/s41598-021-82728-w>
- Voorman, A., Lumley, T., McKnight, B., & Rice, K. (2011). Behavior of QQ-Plots and Genomic Control in Studies of Gene-Environment Interaction. *PLoS ONE*, 6(5), e19416. <https://doi.org/10.1371/journal.pone.0019416>
- Wang, C., Zhan, X., Bragg-Gresham, J., Kang, H. M., Stambolian, D., Chew, E. Y., Branham, K. E., Heckenlively, J., Fulton, R., Wilson, R. K., Mardis, E. R., Lin, X., Swaroop, A., Zöllner, S., & Abecasis, G. R. (2014). Ancestry estimation and control of population stratification for sequence-based association studies. *Nature Genetics* 2014 46:4, 46(4), 409-415. <https://doi.org/10.1038/ng.2924>
- Wang, C., Zhan, X., Liang, L., Abecasis, G. R., & Lin, X. (2015). Improved Ancestry Estimation for both Genotyping and Sequencing Data using Projection Procrustes Analysis and Genotype Imputation. *The American Journal of Human Genetics*, 96(6), 926-937. <https://doi.org/10.1016/j.ajhg.2015.04.018>
- Wang, G.-S., & Cooper, T. A. (2007). Splicing in disease: Disruption of the splicing code and the decoding machinery. *Nature Reviews Genetics*, 8(10), Article 10. <https://doi.org/10.1038/nrg2164>
- Wang, S., Ray, N., Rojas, W., Parra, M. V., Bedoya, G., Gallo, C., Poletti, G., Mazzotti, G., Hill, K., Hurtado, A. M., Camrena, B., Nicolini, H., Klitz, W., Barrantes, R., Molina, J. A., Freimer, N. B., Bortolini, M. C., Salzano, F. M., Petzl-Erler, M. L., ... Ruiz-Linares, A. (2008). Geographic patterns of genome admixture in latin American mestizos. *PLoS Genetics*, 4(3), 1-9. <https://doi.org/10.1371/journal.pgen.1000037>
- Wang, X.-S., Wu, S.-L., Peng, Z., & Zhu, H.-H. (2021). SLCO4A1 is a Prognosis-Associated Biomarker Involved in Neutrophil-Mediated Immunity in Thyroid Cancer. *International Journal of General Medicine*, 14, 9615-9628. <https://doi.org/10.2147/IJGM.S339921>
- Wimmer, K., Roca, X., Beiglböck, H., Callens, T., Etzler, J., Rao, A. R., Krainer, A. R., Fonatsch, C., & Messiaen, L. (2007). Extensive in silico analysis of NF1 splicing defects uncovers determinants for splicing outcome upon 5' splice-site disruption. *Human Mutation*, 28(6), 599-612. <https://doi.org/10.1002/humu.20493>
- Woldu, S. L., Aydin, A. M., Rao, A. V., Hutchinson, R. C., Singla, N., Clinton, T. N., Krabbe, L. M., Passoni, N. M., Raj, G. V., Miller, D. S., Amatruda, J. F., Sagalowsky, A. I., Lotan, Y., Arriaga, Y., Margulis, V., & Bagrodia, A. (2018). Differences at Presentation and Treatment of Testicular Cancer in Hispanic Men: Institutional and National Hospital-based Analyses. *Urology*, 112, 103-111. <https://doi.org/10.1016/j.urology.2017.08.059>
- Woldu, S. L., & Bagrodia, A. (2018). Update on epidemiologic considerations and treatment trends in testicular cancer. *Current Opinion in Urology*, 28(5), 440-447. <https://doi.org/10.1097/MOU.0000000000000532>
- Yaba, A., Bozkurt, E. R., & Demir, N. (2016). MTOR expression in human testicular seminoma. *Andrologia*, 48(6), 702-707. <https://doi.org/10.1111/and.12504>
- Yang, Z., Huaman, B. C., Reyes-Dumeyer, D., Montesinos, R., Soto-Añari, M., Custodio, N., & Tosto, G. (2023). A benchmark study on current GWAS models in admixed populations. *bioRxiv : the preprint server for biology*. <https://doi.org/10.1101/2023.04.27.538299>

- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics*, *13*(1), 134. <https://doi.org/10.1186/1471-2105-13-134/FIGURES/5>
- Yinsheng, Z., Miyoshi, K., Qin, Y., Fujiwara, Y., Yoshimura, T., & Katayama, T. (2022). TMEM67 is required for the gating function of the transition zone that controls entry of membrane-associated proteins ARL13B and INPP5E into primary cilia. *Biochemical and Biophysical Research Communications*, *636*, 162-169. <https://doi.org/10.1016/j.bbrc.2022.10.078>
- Zhang, X., Basile, A. O., Pendergrass, S. A., & Ritchie, M. D. (2019). Real world scenarios in rare variant association analysis: The impact of imbalance and sample size on the power in silico. *BMC Bioinformatics*, *20*(1), 46. <https://doi.org/10.1186/s12859-018-2591-6>
- Znaor, A., Lortet-Tieulent, J., Jemal, A., & Bray, F. (2014). International variations and trends in testicular cancer incidence and mortality. *European Urology*, *65*(6), 1095-1106. <https://doi.org/10.1016/j.eururo.2013.11.004>
- Zondervan, K. T., Cardon, L. R., & Kennedy, S. H. (2002). What makes a good case-control study?: Design issues for complex traits such as endometriosis. *Human Reproduction*, *17*(6), 1415-1423. <https://doi.org/10.1093/humrep/17.6.1415>