



Máster en Bioinformática

**IDENTIFICACIÓN DE FIRMAS
MUTACIONALES EN CÁNCER DE
PRÓSTATA MEDIANTE MÉTODOS DE
FACTORIZACIÓN MATRICIAL**

Autor: Bárbara García Gálvez.

Tutor: Daniel Tello Velasco, Ph.D.

Curso 2024-25

ÍNDICE

-Agradecimientos

-Resumen

-Palabras clave

-Abstract

-Keywords

1. Introducción	1
1.1 Cáncer	1
1.2 Cáncer de próstata	1
1.3 Mutaciones somáticas y firmas mutacionales	3
1.4 Identificación de firmas mutacionales mediante factorización matricial no negativa (NMF)	5
1.5 Catálogo COSMIC	5
1.6 Estado del arte	6
2. Hipótesis y Objetivos	7
3. Metodología	8
3.1 Diseño	8
3.2 Población	8
3.3 Número de pacientes estudiado	9
3.4 Criterios de inclusión	9
3.5 Criterios de exclusión	10
3.6 Cálculo del tamaño muestral	10
3.7 Variables	11
3.8 Método de recogida de datos	11
3.9 Análisis estadístico	13
3.9.1 Formulación de NMF	14

3.9.2 Comparación de herramientas	15
3.9.3 Implementación de SigProfilerExtractor	16
3.10 Análisis exploratorio y enriquecimiento funcional de R	17
4. Resultados.....	18
4.1 Selección del número óptimo de firmas	18
4.2 Extracción de firmas SBS96.....	19
4.3 Análisis de exposiciones y distribución de muestras	22
4.4 Comparación con COSMIC	24
4.5 Análisis de enriquecimiento funcional de genes mutados.....	29
5. Discusión.....	32
5.1 Limitaciones del estudio.....	34
5.2 Futuras líneas de investigación	35
6. Conclusiones.....	36
7. Bibliografía.....	38
8. Anexos.....	50

AGRADECIMIENTOS

Queremos mostrar agradecimiento a los participantes de los cuatro estudios de la Plataforma de Biomarcadores PROCURE (PROREPAIR-B, PROSENZA, PROSTAC y PROSABI), y también a los 38 centros de España e Italia donde se llevaron a cabo. Agradecemos a los responsables del estudio CAPTURE por permitir acceso a la cohorte y por su colaboración en la gestión de la base de datos. Queremos agradecer a los bioinformáticos y técnicos de laboratorio que se encargaron del tratamiento de las muestras y el análisis genómico. Asimismo, manifestamos nuestro agradecimiento a los creadores de SigProfilerExtractor por permitir el uso de la herramienta y la disponibilidad de documentación asociada, facilitando la reproducibilidad de este estudio.

RESUMEN

El cáncer continúa siendo una causa principal de mortalidad a nivel global y, en el de próstata metastásico resistente a la castración (mCRPC), la heterogeneidad tumoral requiere biomarcadores moleculares. La descomposición de perfiles mutacionales mediante factorización matricial no negativa (NMF) permite identificar firmas mutacionales asociadas a mecanismos biológicos.

Objetivos:

Extraer e interpretar firmas mutacionales en mCRPC a partir de una matriz con 96 tipos de mutación y comparar un número óptimo con las firmas del catálogo COSMIC. Finalmente, investigar procesos de mutagénesis y su significado clínico.

Metodología:

Se analizaron 148 muestras tumorales y se construyó la matriz SBS96 a partir de archivos VCF, utilizando SigProfilerExtractor (GRCh38) para identificar firmas mediante NMF. Se seleccionó el modelo óptimo estudiando estabilidad, error de reconstrucción y similitud de coseno frente a COSMIC v3.4. Además, se realizó análisis de enriquecimiento funcional (GO:BP) sobre genes con mutaciones no sinónimas en muestras con alta exposición a cada firma.

Resultados:

Se obtuvieron tres firmas de *novo* (SBS96A, SBS96B y SBS96C) con estabilidad ≥ 0.93 . SBS96A fue la más prevalente y se relacionó con SBS15. SBS96B y SBS96C se relacionaron con SBS1 y SBS5. SBS96A y SBS96C mostraron mayor carga mutacional por megabase. El análisis funcional mostró activación de rutas de reparación del ADN y control del ciclo celular.

Conclusiones:

Las firmas identificadas reflejan procesos de envejecimiento (SBS1 y SBS5) y deficiencia en la reparación del ADN, por errores de apareamiento del ADN (MMRd) o inestabilidad de microsatélites (MSI) (SBS15), y por alteraciones en recombinación homóloga (HRR), siendo importantes en estratificación de pacientes e inmunoterapia o inhibidores de PARP.

Palabras clave:

COSMIC; cáncer de próstata; firmas mutacionales; inmunoterapia; mCRPC; NMF; SBS; SigProfilerExtractor.

ABSTRACT

Cancer remains a leading cause of mortality worldwide, and in metastatic castration-resistant prostate cancer (mCRPC), tumor heterogeneity requires molecular biomarkers. Decomposition of mutational profiles using non-negative matrix factorization (NMF) enables the identification of mutational signatures associated with biological mechanisms.

Objectives:

To extract and interpret mutational signatures in mCRPC from a matrix with 96 mutation types and compare an optimal number with reference signatures from the COSMIC catalog. Finally, to investigate mutagenesis processes and their clinical significance.

Methods:

A total of 148 tumor samples were analyzed, the SBS96 matrix was generated from VCF files, and mutational signatures were extracted using SigProfilerExtractor (GRCh38) through NMF. The optimal model was selected based on the evaluation of stability, reconstruction error, and cosine similarity against the COSMIC v3.4 catalog. Additionally, a functional enrichment analysis (GO:BP) was performed on non-synonymous mutated genes from samples with high exposure to each signature.

Results:

Three *de novo* signatures (SBS96A, SBS96B, and SBS96C) with stability ≥ 0.93 were identified. SBS96A was the most prevalent and was associated with SBS15. SBS96B and SBS96C were mainly linked to SBS1 and SBS5. SBS96A and SBS96C showed a higher mutational burden per megabase. Functional analysis revealed activation of DNA repair pathways and cell cycle regulation.

Conclusions:

The identified signatures reflect biological processes related to aging (SBS1 and SBS5) and DNA repair deficiency, including mismatch repair deficiency or microsatellite instability (MMRd/MSI) (SBS15), as well as homologous recombination repair (HRR) alterations. These processes are clinically relevant for patient stratification and for guiding the use of immunotherapy or PARP inhibitors.

Keywords:

COSMIC; immunotherapy; mCRPC; mutational signatures; NMF; prostate cancer; SBS; SigProfilerExtractor.

1. Introducción

1.1 Cáncer

El cáncer es una de las principales causas de muerte a nivel mundial y de acuerdo a la actualización de 2022 del Observatorio Global del Cáncer, se estima que 20 millones de personas recibirán un nuevo diagnóstico de esta enfermedad genética, y casi 10 millones fallecerán (Bray et al., 2024). Esto es especialmente relevante en los países en proceso de desarrollo, pues atraviesan una transición económica que incluye una mayor mecanización del transporte y del trabajo, cambios culturales en los roles de las mujeres, y una mayor exposición y acceso a los mercados internacionales. En consecuencia, muchos de los factores de riesgo relacionados con el estilo de vida como el tabaquismo, la inactividad física, el exceso de peso corporal y los patrones reproductivos, vienen a implementarse con el crecimiento económico (Torre et al., 2016). A pesar de que las tasas generales de mortalidad por cáncer siguen disminuyendo, las tasas de incidencia se han estabilizado entre los hombres y han mostrado un aumento moderado entre las mujeres. Estas tendencias reflejan cambios en los factores de riesgo de cáncer, el uso de pruebas de detección, las prácticas diagnósticas y los avances en el tratamiento. Muchos tipos de cáncer pueden prevenirse o tratarse de manera efectiva si se diagnostican de forma precoz (Santucci et al., 2020).

1.2 Cáncer de próstata

El cáncer de próstata (PCa) afecta a varones de mediana edad siendo la principal causa de mortalidad en países occidentales por este tipo de enfermedad (Chen et al., 2013). Los factores principales de riesgo son la edad, antecedentes familiares, etnia, obesidad y factores ambientales. Se trata de una enfermedad heterogénea tanto desde el punto de vista epidemiológico como genético (Hjelmberg et al., 2014). Estudios en gemelos y estudios epidemiológicos han demostrado el papel del cáncer de próstata hereditario (Termini et al., 2020). Los genes candidatos para el desarrollo del epitelio prostático y de las células cancerosas prostáticas son aquellos que participan en la vía androgénica y el metabolismo de la testosterona (Bluemn & Nelson, 2012). A muchos hombres con PCa se les diagnostica mediante una biopsia y análisis de próstata, una prueba del antígeno prostático específico (PSA), un examen rectal digital, una

resonancia magnética (MRI) o un examen de salud (Hjelmberg et al., 2014). Por otro lado, la identificación de biomarcadores de cáncer y de mutaciones genéticas específicas pueden utilizarse para el tratamiento dirigido del cáncer de próstata (Ziara et al., 2015). El PSA carece de especificidad tumoral, lo que implica biopsias innecesarias o falta de detección de cánceres agresivos. Para mejorar la especificidad, los biomarcadores sanguíneos utilizados recientemente como el índice de salud prostática (Prostate Health Index, PHI) y el 4Kscore, integran isoformas del PSA o niveles de proteínas calicreínas con parámetros clínicos. Biomarcadores urinarios como PCA3 y SelectMDx mejoran la precisión en el diagnóstico, concretamente en la diferenciación de tumores de alto grado, y reflejan potencial en programas de vigilancia activa.

Marcadores pronósticos como Bcl-2, Ki-67 y EZH2; alteraciones genéticas como MCM7 y la amplificación del brazo largo del cromosoma 8 (8q), contribuyen a la estratificación de los pacientes en función de la agresividad tumoral y el riesgo de recurrencia. Las biopsias líquidas, incluyendo el análisis de ADN tumoral circulante (ctDNA) y de células tumorales circulantes (CTCs), proporcionan alternativas no invasivas para el perfil molecular, especialmente en el cáncer de próstata metastásico resistente a la castración (mCRPC), y permiten detectar alteraciones terapéuticamente accionables vía reparación por recombinación homóloga (HRR) como mutaciones en *BRCA1/2* o *ATM*.

En los últimos años, el panorama de los biomarcadores y de la investigación en cáncer de próstata se ha visto ampliado por tecnologías innovadoras, como el aprendizaje automático y las ómicas, que están redefiniendo la identificación de biomarcadores. Por ejemplo, modelos basados en inteligencia artificial, como la firma de estrés replicativo (RSS), son potencialmente útiles para la predicción de recaídas y de la respuesta terapéutica. Por otro lado, la secuenciación de ARN unicelular y la transcriptómica espacial han aumentado el conocimiento sobre la heterogeneidad del PCa, del microambiente tumoral y de los mecanismos de resistencia. Además, nuevos biomarcadores como los ARNs de exosomas y marcadores asociados a la respuesta inmune (PD-L1, SOX2, TcellinfGEP) aportan datos sobre la progresión tumoral y un posible valor en inmunoterapia (Crocetto et al., 2025).

Las opciones de tratamiento disponibles para esta enfermedad son la vigilancia activa, quimioterapia, radioterapia, hormonoterapia, cirugía y crioterapia dependiendo de la naturaleza del tumor, el grado y el estadio, y la posible recurrencia (Bach et al., 2014). Cada tratamiento se asocia con efectos secundarios graves, como toxicidad y disminución del recuento de glóbulos blancos y rojos, lo que provoca fatiga, caída del cabello, neuropatía periférica, incontinencia y disfunción eréctil, metástasis y desarrollo de resistencia al tratamiento inicial. Las opciones de tratamiento disponibles son costosas y presentan graves efectos secundarios. Es necesario el descubrimiento de nuevos agentes quimioterapéuticos más económicos con pocos o ningún efecto secundario y mayor eficacia (Termini et al., 2020).

En el ámbito clínico, la detección de procesos mutacionales presenta limitaciones, principalmente por la dificultad y coste económico de realizar pruebas genómicas y la ausencia de métodos estandarizados que interpreten grandes volúmenes de datos. Sin embargo, estudios recientes han demostrado que más del 90 % de los pacientes con cáncer de próstata resistente a la castración (mCRPC) muestran alteraciones genómicas sensibles a tratamiento (Robinson et al., 2015), y que la adición de inhibidores de la enzima poli ADP ribosa polimerasa (PARP) en pacientes con alteraciones en genes de HRR ha marcado tendencia hacia la medicina de precisión en esta enfermedad (Fenton et al., 2024).

1.3 Mutaciones somáticas y firmas mutacionales

Desde una perspectiva molecular, las mutaciones que afectan al genoma se pueden clasificar como germinales o somáticas. Las mutaciones germinales son aquellas que pueden originarse de manera aleatoria en células de la línea germinal de un individuo, se heredan siguiendo una proporción mendeliana y se encuentran en todas las células del cuerpo del individuo (Gonzalo, 2013). Sin embargo, las mutaciones somáticas son alteraciones del ADN que pueden ocurrir en cualquier célula del cuerpo, excepto las germinativas y ocurren después de la concepción, incrementándose con el envejecimiento (Fowler & Jones, 2022).

El conjunto de mutaciones somáticas de un paciente constituye un perfil mutacional, es decir, la distribución de alteraciones observadas en el genoma tumoral. Entre ellas, las sustituciones de base simple (Single Base Substitutions, SBS) representan la categoría más frecuente, y pueden clasificarse en diferentes tipos según el cambio de nucleótido y el contexto en el que ocurren. Al comparar los perfiles mutacionales de diferentes pacientes, es posible identificar combinaciones recurrentes de mutaciones que reflejan procesos biológicos comunes; estas combinaciones se conocen como firmas mutacionales (Alexandrov et al., 2013).

Las mutaciones somáticas originadas por procesos mutacionales son tanto de origen exógeno como endógeno (Stratton et al., 2009). Cada proceso mutacional puede involucrar componentes de daño o modificación, reparación y replicación del ADN y generar una firma mutacional que incluye sustituciones de bases, pequeñas inserciones y deleciones (indels), reordenamientos genómicos y cambios en el número de copias cromosómicas (CNVs) (Alexandrov & Stratton, 2014). Para las SBS, la clasificación principal comprende 96 clases definidas en el Catálogo de Mutaciones Somáticas en Cáncer (COSMIC), constituidas por la combinación de las 6 sustituciones de bases: C>A, C>G, C>T, T>A, T>C y T>G (en las que la base mutada se representa por la pirimidina del par de bases), y todas las posibles combinaciones de bases flanqueantes en las posiciones 5' y 3'.

Las firmas mutacionales tienen el potencial de reflejar la biología del tumor y también funcionan como biomarcadores terapéuticos y pronósticos. En un estudio se demostró cómo estas firmas son mejores predictoras de la respuesta a fármacos que las mutaciones conductoras o alteraciones estructurales (Levatić et al., 2022). Otros científicos evidenciaron que las firmas derivadas de la deficiencia en la reparación por recombinación se relacionan con la sensibilidad a inhibidores de PARP y agentes citotóxicos (Póti et al., 2019).

Las mutaciones en el genoma de un cáncer individual pueden haber sido generadas por múltiples procesos mutacionales e incorporar múltiples firmas mutacionales superpuestas. Para caracterizar estos procesos, previamente se han utilizado métodos matemáticos para descifrar las firmas mutacionales a

partir de catálogos de mutaciones somáticas, estimar el número de mutaciones atribuibles a cada firma en muestras individuales y anotar cada clase de mutación en cada tumor (Alexandrov et al., 2020).

1.4 Identificación de firmas mutacionales mediante factorización matricial no negativa (NMF)

La extracción de *novus* de firmas mutacionales puede interpretarse como un problema de inferencia de variables latentes en aprendizaje automático no supervisado. En este contexto, factorización matricial no negativa (NMF) es una técnica de aprendizaje automático no supervisado que permite descomponer un conjunto de datos en componentes latentes no negativos. En cáncer, este método se utiliza para identificar firmas mutacionales de *novus* a partir de los perfiles de mutaciones observados en cada paciente.

En este enfoque, los datos se representan mediante una matriz observada **V**, que contiene los recuentos de mutaciones por paciente. Esta matriz se aproxima como el producto de dos matrices de bajo rango: **H**, que representa las firmas mutacionales globales compartidas por todos los genomas, y **W**, que recoge las exposiciones locales o contribuciones de cada firma en cada individuo. La capacidad de NMF de producir factores no negativos partiendo de los datos originales, permite su interpretación biológica (Islam et al., 2022; Y.-A. Kim et al., 2021).

En definitiva, NMF permite representar el perfil mutacional de cada tumor como una combinación lineal positiva de un conjunto limitado de firmas mutacionales subyacentes, cada una asociada a procesos biológicos específicos (Lin & Boutros, 2020).

1.5 Catálogo COSMIC

COSMIC es un sistema de base de datos que recopila datos de mutaciones somáticas de diversas fuentes públicas en un único repositorio estandarizado. Iniciado en 2004 con análisis de tan solo cuatro genes humanos (Bamford et al., 2004), se ha convertido en un sistema genómico de referencia para explorar patrones de mutaciones somáticas en todos los tipos de cáncer (Forbes et al., 2017). Respecto a firmas mutacionales, la versión 3.4 de COSMIC es una de las

más ampliamente citadas, que amplió las 30 firmas de la versión 2.0 hasta 86 firmas de SBS, además de considerar firmas de sustituciones dobles de bases (DBS), indels y CNVs. Grandes proyectos internacionales como el Atlas del Genoma de Cáncer (TCGA) y el Proyecto Pan-Cancer Analysis of Whole Genomes (PCAWG), aportaron miles de genomas y exomas para su análisis mediante herramientas computacionales como SigProfiler (Sondka et al., 2024).

1.6 Estado del arte

Para los recuentos mutacionales, recientes avances han ampliado el modelo NMF Binomial Negativo clásico incluyendo la dispersión específica del paciente. El modelo extendido se denomina NB_N -NMF, donde N representa el número de parámetros de dispersión (número de pacientes). Se investigó cuándo y por qué el NB_N -NMF es más adecuado para los recuentos mutacionales que el NMF Poisson habitual (Po-NMF) (Févotte et al., 2009). Este método se aplicó a un dataset de 286 pacientes de cáncer de próstata de la base de datos PCAWG ("Pan-Cancer Analysis of Whole Genomes," 2020). Se mostró la similitud de coseno entre las firmas extraídas por el paquete de R *SigMos* (Signature Model Selection) y las más similares del repositorio COSMIC. NB_N -NMF encuentra las firmas SBS1, SBS5, SBS8, SBS18 y SBS37. Estas firmas están ampliamente presentes en el cáncer de próstata, tanto por su presencia en muchas muestras de tumores de próstata como por su contribución en términos de número de mutaciones por tumor (Pelizzola et al., 2023).

En otro estudio, se identificaron 3 firmas mutacionales SBS-sig 1, 2, 3 realizándose un análisis de similitud de coseno entre las tres firmas y las de COSMIC (Wang et al., 2021). Las firmas SBS identificadas son similares a las descritas por COSMIC que se basan en los conjuntos de datos de cáncer de próstata del TCGA. El resultado del análisis de similitud de coseno demostró que Sig1 se relaciona con SBS3 (0.84) y SBS5 (0.82), lo que sugiere un componente de deficiencia en la recombinación homóloga junto con la firma asociada a envejecimiento (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013; Nik-Zainal et al., 2012; Záborszky et al., 2017). Sig2 presentó una alta similitud con SBS15 (0.94) y asociaciones con SBS6 (0.89) y SBS14 (0.81), todas correspondientes a deficiencia de MMR e inestabilidad de microsatélites (Alexandrov, Nik-Zainal,

Wedge, Aparicio, et al., 2013; Hodel et al., 2020; Meier et al., 2018). Por último, Sig3 mostró una similitud casi perfecta con SBS1 (0.94), la firma reloj relacionada con la edad (Nik-Zainal et al., 2012), y en menor medida con SBS6 (0.79), lo que implica un proceso mutagénico basado en la edad con contribución secundaria de MMRd. Por otro lado, cinco firmas de alteración de número de copias fueron identificadas de CN-sig 1 a CN-sig5 y se clasificaron en función de la longitud mediana del segmento CNA (Copy Number Alteration). Las firmas fueron identificadas mediante el paquete de R *Sigminer* a partir de 1003 pares tumor-normal de datos de secuenciación del exoma completo (WES) de cáncer de próstata.

En conjunto, los estudios previos han descrito la presencia de firmas mutacionales relevantes en cáncer de próstata, como SBS1 y SBS5 asociadas al envejecimiento, y SBS15 vinculada a deficiencia en la reparación de errores de apareamiento del ADN (MMRd). Sin embargo, la mayoría de estos trabajos se han basado en grandes consorcios internacionales como PCAWG o TCGA, y existe una escasez de estudios centrados específicamente en cohortes de cáncer de próstata metastásico resistente a la castración (mCRPC) con datos clínicos asociados. Este vacío limita la traslación de los hallazgos a contextos clínicos más próximos a la práctica real.

El presente TFM contribuye a esta área poco explorada aplicando NMF a una cohorte institucional de mCRPC con alta calidad de secuenciación, con el objetivo de extraer firmas mutacionales de *novo* y compararlas con el catálogo COSMIC. De este modo, se pretende aportar evidencia sobre la relevancia biológica y clínica de las firmas identificadas en este subtipo avanzado de cáncer de próstata, contribuyendo a una mejor estratificación de pacientes y al desarrollo de biomarcadores útiles en medicina de precisión.

2. Hipótesis y Objetivos

Hipótesis: La descomposición de perfiles mutacionales mediante métodos de factorización matricial, como la factorización de matrices no negativa (NMF), permite identificar firmas mutacionales latentes asociadas a mecanismos

biológicos específicos en el cáncer de próstata. Estas firmas pueden reflejar procesos de mutagénesis relevantes para la progresión tumoral y contribuir a la identificación de posibles biomarcadores o subtipos moleculares.

Objetivos:

1. Extraer el contexto trinucleotídico para cada mutación somática a partir de datos genómicos de archivos VCF de cáncer de próstata resistente a la castración.
2. Aplicar técnicas de factorización matricial para descomponer una matriz de los 96 tipos de mutación por muestra y determinar el número óptimo de firmas.
3. Interpretar biológicamente las firmas mutacionales comparando con firmas del catálogo COSMIC.
4. Realizar análisis de enriquecimiento funcional para investigar posibles procesos biológicos, rutas moleculares o mecanismos mutagénicos asociados a las firmas identificadas.
5. Representar gráficamente las firmas y su distribución en la cohorte y explorar correlaciones con variables clínicas o genéticas relevantes.

3. Metodología

3.1 Diseño

Estudio descriptivo basado en el análisis primario de datos genómicos obtenidos de una cohorte institucional de pacientes con cáncer de próstata metastásico resistente a la castración (mCRPC). Se emplearon herramientas bioinformáticas y métodos de factorización matricial no negativa (NMF) para identificar firmas mutacionales a partir de múltiples tipos de alteraciones somáticas detectadas en las muestras tumorales.

3.2 Población

Cohorte original de 200 pacientes con mCRPC incluidos en el estudio CAPTURE, encargado de analizar patrones de tratamiento y resultados en

dichos pacientes con y sin alteraciones somáticas o germinales en los genes de reparación por recombinación homóloga (HRR). Los pacientes candidatos se dividieron en cuatro estudios observacionales multicéntricos de la Plataforma de Biomarcadores PROCURE: PROREPAIR-B [NCT03075735], PROSENZA [NCT02922218], PROSTAC [NCT02362620] y PROSABI [NCT02787837], realizados entre 2013 y 2020 en 38 centros de España e Italia. Los criterios de inclusión exigieron varones con una edad mayor o igual a 18 años con diagnóstico histológico confirmado de cáncer de próstata, progresión a mCRPC, inicio de tratamiento de primera línea con taxanos o inhibidores de la vía del receptor de andrógenos (ARPi) y disponibilidad de muestras de sangre para el análisis de ADN germinal y bloques tumorales fijados en formol e incluidos en parafina (FFPE) para el análisis de ADN somático. La mediana de edad al inicio del tratamiento fue de 72,2 años y aproximadamente el 53,1% de los pacientes oncológicos presentó una escala clínica que mide su estado funcional (Eastern Cooperative Oncology Group Performance Status, ECOG PS) mayor o igual a 1 (Olmos et al., 2023; Olmos et al., 2024).

3.3 Número de pacientes estudiado

Del total de 200 pacientes de la cohorte, se seleccionaron 148 muestras que cumplieran con los criterios de calidad mínimos establecidos, incluyendo una cobertura promedio normalizada mayor o igual a 150X. En línea con proyectos internacionales como TCGA o el Consorcio Internacional del Genoma del Cáncer (ICGC), se están secuenciando miles de tumores a profundidades estándar de 100 a 150x para exomas. Esto permite detectar mutaciones clonales y subclonales de frecuencia alélica baja, aproximadamente un 3 %, con sensibilidad superior al 60 %. Por tanto, el umbral de 150x es necesario en muestras de alta heterogeneidad y de pureza tumoral variable (Cibulskis et al., 2013; Vijayan et al., 2017).

Para cada muestra, se obtuvieron los datos de variantes somáticas en formato VCF (Variant Calling Format).

3.4 Criterios de inclusión

- Disponibilidad de datos en formato VCF válido.

- Variantes somáticas de alta calidad, con estado *PASS* en el campo de filtrado.
- Cobertura promedio normalizada mínima de 150X (ajustada por pureza tumoral).
- Presencia de la base mutada junto con los dos nucleótidos flanqueantes (5' y 3') para la reconstrucción de los trinucleótidos.

3.5 Criterios de exclusión

- Ausencia de variantes somáticas detectables.
- Archivos dañados o con formato incorrecto.
- Número insuficiente de mutaciones (<50) para garantizar un análisis robusto. De acuerdo con herramientas de análisis de firmas mutacionales como deconstructSigs y FIREVAT, se advierte o interrumpe el análisis cuando el número de mutaciones es inferior a este valor pues el perfil identificado puede deberse al azar y ser muy sensible al ruido. Se destaca la importancia de contar con un número suficiente de mutaciones, para identificar y asignar firmas caracterizadas por clasificaciones de 96 sustituciones. Esto se considera importante cuando el perfil de la firma mutacional es plano o no presenta un pico marcado en ninguno de los contextos de trinucleótidos, ya que en estos casos el proceso mutacional podría afectar a un mayor número de contextos de trinucleótidos y solo se observaría un perfil completo con un número suficiente de mutaciones (H. Kim et al., 2019; Rosenthal et al., 2016). Además, estudios de simulación han mostrado que un ajuste fiable puede alcanzarse incluso con solo 50 mutaciones por firma (Manders et al., 2022). En el caso de 100 o más mutaciones aumentaría la precisión, pero excluiría tumores con baja carga mutacional (Medo et al., 2024).

3.6 Cálculo del tamaño muestral

Al tratarse de un estudio retrospectivo basado en datos ya disponibles, no se realizó un cálculo de tamaño muestral a priori. Se incluyeron todas las muestras que cumplían los criterios de calidad definidos.

3.7 Variables

- Variable principal: firma mutacional identificada mediante NMF.
- Variables secundarias:
 - Tipo de mutación somática (SBS).
 - Carga mutacional por muestra y por firma.
 - Similitud de coseno entre firmas de *novus* y firmas del catálogo COSMIC.

3.8 Método de recogida de datos

Los datos iniciales fueron proporcionados en hojas de cálculo (formato Excel), que incluían métricas de calidad y variantes somáticas correspondientes a pacientes con cáncer de próstata. La primera hoja contenía métricas de calidad como la cobertura promedio (AP) y la cobertura promedio normalizada por pureza tumoral (AQ), calculada mediante los valores de pureza disponibles en la columna "Purity". La cobertura promedio representa la cantidad media de lecturas de secuenciación que tiene cada base y con la normalización se consigue obtener la profundidad efectiva asociada al tumor. Se excluyeron las muestras de baja calidad con una cobertura normalizada por pureza tumoral inferior a 150X, valor exigido en estudios de cáncer de próstata. Esta información se combinó con la segunda hoja que contenía los datos de variantes somáticas por muestra.

Los datos de variantes somáticas en formato tabular se adaptaron a formato estándar VCF, utilizado en la herramienta SigProfilerExtractor. Para ello, se generó un script en Python utilizando la librería *pandas*, que automatizó la creación de archivos VCF por cada una de las muestras. El script incluía el encabezado requerido y se elaboró el campo INFO de cada variante con anotaciones como AF (allele frequency), Consequence, IMPACT y SYMBOL. Las columnas genotípicas GT (Genotype), DP (Read Depth) y AD (Allelic Depths) fueron integradas en el campo "FORMAT" según VCFv4.2.

La herramienta SigProfilerExtractor fue descargada desde su [repositorio oficial en GitHub](#) y configurada localmente. Para garantizar un entorno reproducible, se

creó un entorno virtual de Python mediante el comando `python3 -m venv venv`, el cual fue activado con `source venv/bin/activate`. Se instalaron las dependencias principales con `pip install SigProfilerMatrixGenerator SigProfilerExtractor` y manualmente el genoma de referencia GRCh38 mediante el módulo `install_genome` incluido en el repositorio. Los archivos VCF se organizaron en un directorio y se utilizó un script en Python para ejecutar el análisis, introduciendo el tipo de entrada como VCF, y con un rango de búsqueda entre 1 y 10 firmas mutacionales. Como resultado, se generó la matriz SBS96 y se aplicó NMF para producir firmas mutacionales de *novo* (Fig. 1).

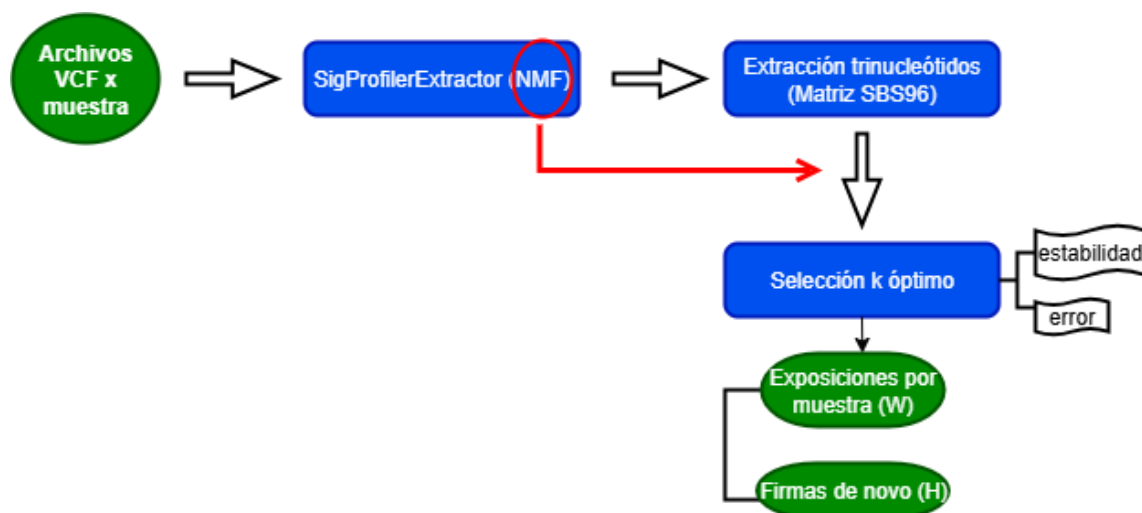


Figura 1. Diagrama de flujo del pipeline. Proceso desde la obtención de archivos VCF hasta la extracción de firmas mutacionales de *novo*. A partir de los archivos VCF de cada muestra, SigProfilerExtractor genera la matriz SBS96 de trinucleótidos y la descompone mediante NMF. A continuación, se determina el número óptimo de firmas (k) evaluando métricas de estabilidad y error. Como resultado, se obtienen las exposiciones específicas por muestra (W) y las firmas mutacionales de *novo* (H), que posteriormente se contrastarán con el catálogo COSMIC mediante análisis de similitud de coseno.

Todo el preprocesamiento de datos se realizó utilizando Python 3.10, organizando el trabajo entre dos notebooks en Google Colab y un script final

para la ejecución de SigProfilerExtractor en una máquina virtual con Ubuntu 22.04 usando VSCode como entorno de desarrollo.

3.9 Análisis estadístico

El análisis estadístico se basó en la descomposición de la matriz SBS96 mediante NMF, con el fin de identificar un número limitado de firmas mutacionales y estimar su contribución relativa en cada muestra.

Para seleccionar el número óptimo de firmas, se evaluaron métricas de ajuste como la estabilidad de la factorización, que es la consistencia de las firmas y sus exposiciones entre réplicas de NMF con diferentes inicializaciones, y el error de reconstrucción.

Una estabilidad baja por la falta de unicidad del NMF se debe al remuestreo de Poisson, que introduce variaciones aleatorias en los datos para evaluar la robustez del modelo, y a la posible existencia de múltiples puntos estacionarios convergentes en la solución. La estabilidad de cada firma mutacional se calculó mediante SigProfilerExtractor con el valor de silueta, que compara la similitud de las repeticiones de una firma extraída por NMF agrupadas dentro de un clúster con respecto a las de otros clústeres. Para garantizar reproducibilidad, esta herramienta selecciona soluciones donde las firmas presentan una estabilidad superior a 0,80 y ninguna inferior a 0,20 en muestras de Secuenciación del Genoma Completo (WGS). Los valores 0,70 y 0,10, respectivamente, son los umbrales recomendados para extracciones basadas en muestras secuenciadas de WES.

El error de reconstrucción indica la calidad del ajuste calculando la capacidad de las firmas de reconstruir los datos de la matriz original de mutaciones. El ajuste de las soluciones se evaluó mediante distintas métricas de error de reconstrucción calculadas automáticamente por SigProfilerExtractor, incluyendo las distancias L1 y L2 (Euclídea o norma de Frobenius), la divergencia de Kullback–Leibler (KL) y la correlación de Pearson entre la matriz original y la reconstruida. Mientras que L1 y L2 cuantifican diferencias absolutas y cuadráticas entre ambas matrices, respectivamente, la correlación de Pearson mide la conservación de la forma del espectro de mutaciones, y la divergencia

KL mide la discrepancia entre la matriz original de datos y su aproximación de bajo rango, siendo muy adecuada para datos de conteo (Hien & Gillis, 2021). Además, la herramienta no aplica el test de Wilcoxon rank-sum en datos de WES para comparar los errores de reconstrucción entre soluciones estables con diferente número de firmas (k) pues se prioriza la sensibilidad en cohortes con baja carga mutacional. Si se detecta un aumento significativo del error ($p < 0,05$), se conserva la solución con mayor número de firmas (Islam et al., 2022).

Posteriormente, se realizó un análisis de similitud de coseno para comparar las firmas *de novo* extraídas con las del catálogo COSMIC v3.4, con el objetivo de inferir sus posible etiología o proceso biológico asociado. Este análisis se trata de una métrica que evalúa el grado de semejanza entre dos vectores en un espacio n -dimensional calculando el coseno del ángulo que los separa. Los valores cercanos a 1 reflejan gran similitud y los próximos a 0, ausencia de relación (A. Huang, 2008; Kirişci, 2023). Finalmente, se evaluó la distribución de las firmas en cada muestra y la carga mutacional atribuida a cada firma. Los análisis y visualizaciones se llevaron a cabo en Python 3.10, utilizando *pandas* para la gestión de datos y módulos integrados en SigProfiler para la generación de matrices, extracción de firmas y comparación con COSMIC.

3.9.1 Formulación de NMF

Para identificar las firmas mutacionales se utilizan métodos de descomposición matricial como la factorización en matrices no negativas (NMF), que descompone una matriz V en el producto de dos matrices no negativas W y H de tal forma que: $V \approx W \times H$ (Fig. 2). En genómica del cáncer, la matriz mutacional V contiene los conteos mutacionales de cada paciente, también llamado perfil mutacional. De esta manera, el número de filas n corresponde al número de pacientes, y el número de columnas m al número de tipos diferentes de mutación. Se utiliza SBS96 (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013), donde $m=96$ y corresponde a las 6 mutaciones básicas (asumiendo simetría de cadena) multiplicadas por los 4 nucleótidos flanqueantes a cada lado, es decir, $4 \times 6 \times 4 = 96$. El parámetro k representa el número de firmas mutacionales a identificar mediante los perfiles. Este rango k de las matrices de menor dimensión W y H es mucho menor que n y m . En esta descomposición matricial por NMF, la matriz

H está constituida por componentes latentes que son k firmas mutacionales y están definidos por vectores de probabilidad sobre los diferentes tipos de mutaciones. En la matriz **W**, cada fila contiene los pesos (o contribuciones) de las firmas para el paciente correspondiente. En este contexto, los pesos suelen denominarse exposiciones a las diferentes firmas (Pelizzola et al., 2023). Cabe destacar que algunas herramientas como SigProfiler devuelven las firmas como columnas en lugar de filas siendo diferente la representación a la forma clásica, pero sin impacto en la interpretación biológica.

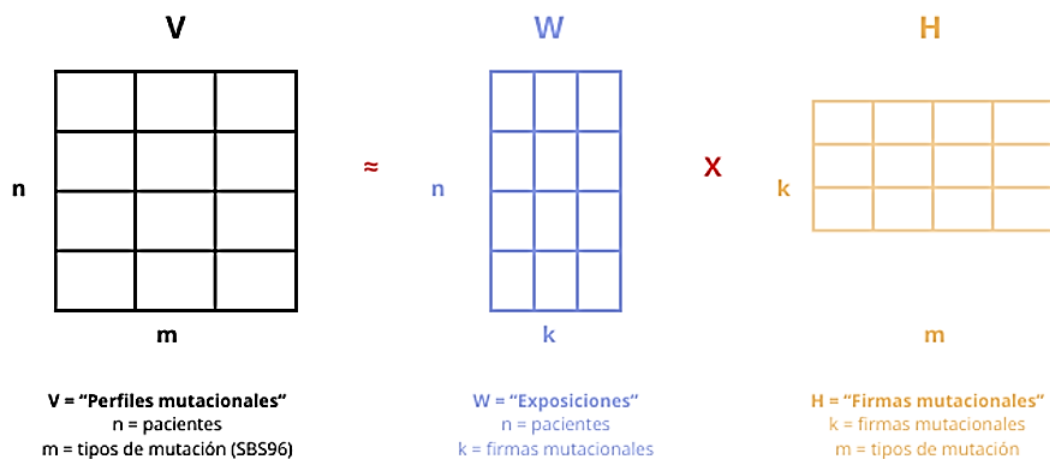


Figura 2. Diagrama de descomposición matricial por NMF ($V \approx W \times H$). Se muestra la matriz original **V** sobre los perfiles mutacionales de cada paciente, **W** que representa la contribución de cada proceso mutacional en cada paciente y **H** que es el patrón (componente latente) observado por proceso.

3.9.2 Comparación de herramientas

Existen diferentes herramientas para el análisis de firmas mutacionales. Entre las basadas en NMF destacan SigProfiler, ampliamente utilizado en la construcción del catálogo COSMIC, y SignatureAnalyzer, que emplea una variante bayesiana de NMF. También existen enfoques alternativos como DeconstructSigs o MutationalPatterns, orientados a ajustar perfiles mutacionales a firmas conocidas en lugar de extraerlas de *novo*.

En este trabajo, se seleccionó SigProfilerExtractor por su robustez estadística, su capacidad de extracción de *novo* en cohortes de tamaño medio y su integración directa con el catálogo COSMIC, lo que asegura una comparación estandarizada de las firmas obtenidas (Blokzijl et al., 2018; Rosenthal et al., 2016).

3.9.3 Implementación de SigProfilerExtractor

SigProfilerExtractor, implementado como un paquete de Python con un *wrapper* en R, admite su ejecución en entornos Python y R y permite el análisis de todos los tipos de clasificaciones mutacionales. Es compatible con la mayoría de los formatos generados por algoritmos de *variant calling* (por ejemplo, VCF), los cuales convierte internamente en una matriz de recuentos mutacionales, frecuentemente la matriz SBS96 de 96 trinucleótidos (Bergstrom et al., 2019).

Una vez obtenida la matriz, la herramienta aplica NMF para descomponerla en dos matrices: Activities (A, equivalente a **W**) y Signatures (S, equivalente a **H**). El objetivo es realizar la selección automática del número óptimo de firmas, k , (entre 1 y 25) y ofrecer soluciones robustas mediante una configuración mínima. Para cada análisis, repite 100 veces la descomposición de la matriz, con una introducción previa de variaciones controladas mediante el remuestreo Poisson y su normalización. Utiliza el algoritmo NMF optimizado para reducir la divergencia de Kullback–Leibler. Emplea *partition clustering* con el algoritmo húngaro para determinar soluciones estables (coeficiente de silueta).

Finalmente, las firmas de *novo* se descomponen en firmas del catálogo COSMIC mediante el algoritmo de mínimos cuadrados no negativos (nonnegative least squares, NNLS) con un sistema iterativo de pasos de adición y eliminación. Para cada firma de *novo*, se agrega iterativamente una firma COSMIC a una lista de firmas utilizadas para explicar la firma de *novo*, se aplica NNLS, y se selecciona la firma cuya adición produce la mayor disminución del error L2 siendo superior a un umbral específico (valor por defecto del 5%). La adición es seguida inmediatamente por un paso de eliminación. Cada firma COSMIC de la lista se elimina iterativamente, se aplica NNLS, y se selecciona la firma cuya eliminación causa la menor disminución del error L2. Si la disminución es inferior a un valor de 1% por defecto, la firma se elimina de la lista.

SigProfilerExtractor es una herramienta resistente al ruido y ha mantenido un buen equilibrio entre precisión y sensibilidad durante la extracción de firmas mutacionales en WES, mientras que el rendimiento del resto de las herramientas es notablemente inferior (Islam et al., 2022).

En conjunto, el pipeline de análisis incluyó:

- (i) La incorporación de archivos VCF por muestra.
- (ii) La construcción de la matriz SBS96 con recuentos de trinucleótidos.
- (iii) La descomposición mediante NMF con remuestreo de Poisson y múltiples réplicas para la extracción de posibles firmas.
- (iv) La selección de las soluciones más estables y con menor error de reconstrucción.
- (v) La comparación de las firmas de *novo* con las de COSMIC mediante análisis de similitud de coseno.

Este flujo estructurado aseguró la solidez estadística del modelo y la interpretación biológica de las firmas obtenidas.

3.10 Análisis exploratorio y enriquecimiento funcional de R

Además del análisis de extracción de firmas mediante SigProfilerExtractor en Python, se realizaron los análisis exploratorios y de enriquecimiento funcional en R (v4.4.2). Se importaron las librerías *tidyverse*, *data.table*, *VariantAnnotation*, *clusterProfiler*, *org.Hs.eg.db* y *enrichplot*. Para garantizar la reproducibilidad, se fijó una semilla de aleatoriedad (`set.seed(123)`).

Primero, se calcularon estadísticas descriptivas de las exposiciones a cada firma (SBS96A, SBS96B y SBS96C), incluyendo mediana, media, rango intercuartílico, desviación estándar y percentil 95. Las distribuciones se visualizaron mediante violin plots con boxplots internos y para detectar las muestras con exposición inusualmente elevada se aplicó la regla de Tukey ($Q3 + 1.5 \cdot IQR$), las cuales se representaron en barplots con sus umbrales de corte.

En segundo lugar, se realizó un análisis de componentes principales (PCA) con las exposiciones normalizadas, cuyo objetivo era observar grupos de pacientes

según su firma dominante por colores. Se visualizaron los dos primeros componentes principales (PC1 y PC2) con su porcentaje de varianza explicada.

Por último, en el análisis funcional se identificaron los genes con mutaciones no sinónimas en los archivos VCF mediante el paquete *VariantAnnotation*. Se filtraron las muestras pertenecientes al tercil alto de exposición para cada firma y se obtuvieron los genes mutados por grupos. Se convirtieron las listas de genes a identificadores Entrez y se llevó a cabo a un análisis de sobrerrepresentación (ORA) con la función *enrichGO* del paquete de R *clusterProfiler*. Esta función implementa el test hipergeométrico, restringido a la ontología GO:Biological Process (GO:BP) para identificar procesos biológicos significativamente sobrerrepresentados en cada firma, aplicando la corrección de p.adjust mediante el método de Benjamini–Hochberg ($p.adjust < 0.05$). Otros parámetros utilizados fueron $pvalueCutoff = 0.05$, $qvalueCutoff = 0.05$ y tamaños mínimos y máximos de conjuntos génicos de 5 y 300. Se excluyeron firmas con menos de 10 genes no sinónimos para obtener resultados más robustos.

La visualización de los resultados mediante dotplots y cnetplots, facilitó la interpretación biológica de las asociaciones entre genes y procesos celulares. Para los cnetplots, se aplicó la función *simplify* a los resultados de enriquecimiento con una medida de similitud semántica "Wang" ($cutoff = 0.7$), eliminando la redundancia de términos GO significativos.

Se exportaron las tablas y gráficos al directorio `enrichment_out/` en formato CSV y PNG, garantizando la reproducibilidad del análisis.

4. Resultados

4.1 Selección del número óptimo de firmas

Se utilizó la herramienta *SigProfilerExtractor* aplicando el método de NMF a partir de una matriz de 96 trinucleótidos generada con datos VCF de muestras tumorales. Previo a la extracción de firmas, se determinó el número óptimo de componentes mediante el análisis conjunto de la estabilidad y distancia coseno media. El punto de equilibrio se alcanzó con tres firmas, que mostraron valores

de estabilidad elevados y una reducción significativa en la distancia coseno (Fig. 3).

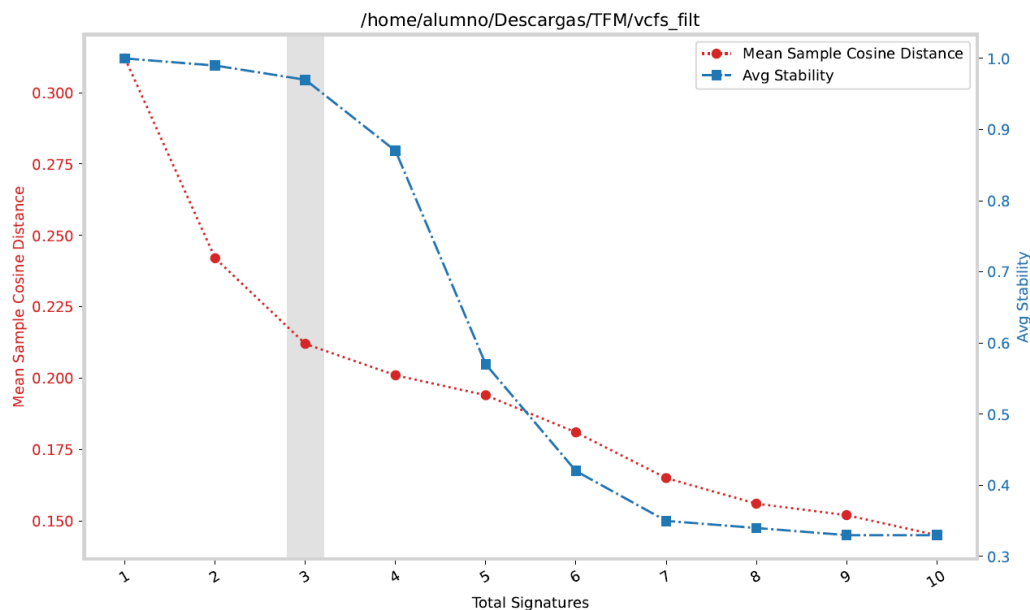


Figura 3. Selección del número óptimo de firmas mediante la evaluación de estabilidad y distancia coseno promedio. El área sombreada en gris señala tres firmas como la solución elegida. En el eje X se muestra el número de firmas totales (1 a 10) y el eje Y representa la distancia coseno promedio en color rojo y la estabilidad media en azul.

4.2 Extracción de firmas SBS96

La solución óptima identificada son tres firmas de *novo*: SBS96A, 96B y 96C (Fig. 4). En la primera firma dominan las mutaciones C>T, con varios picos altos de color rojo, la estabilidad es de 1.0 lo que indica alta reproducibilidad o robustez de esta firma y se representa un total de mutaciones de 42.8%. La firma 96B también con mutaciones C>T (rojo) pero con distribución menos centrada, la estabilidad de 0.99 (alta robustez) y constituye el 29.3% de mutaciones totales. Por último, la firma 96C presenta mutaciones más distribuidas como en C>A y T>C y un patrón más heterogéneo en C>T y T>C (verde) con un grupo amplio de barras; la estabilidad de 0.93 sigue siendo razonablemente alta y engloba un 27.8% de mutaciones totales.

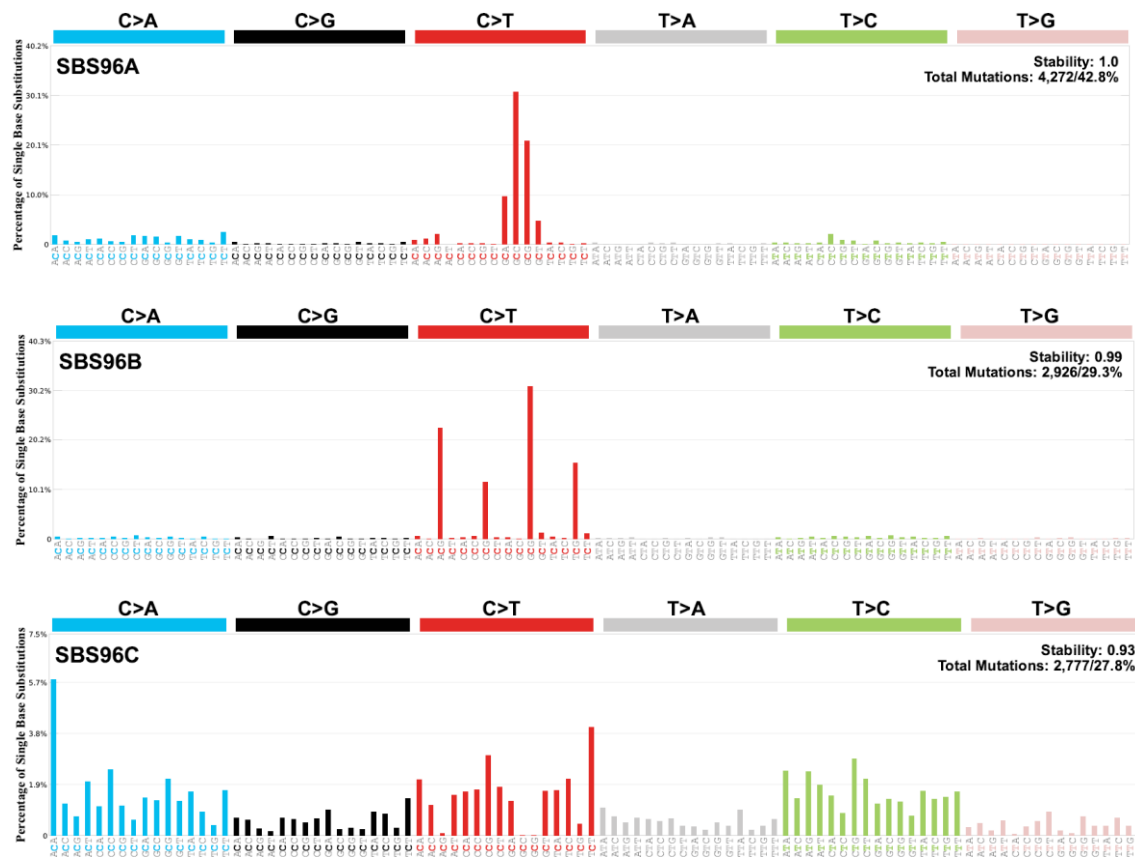
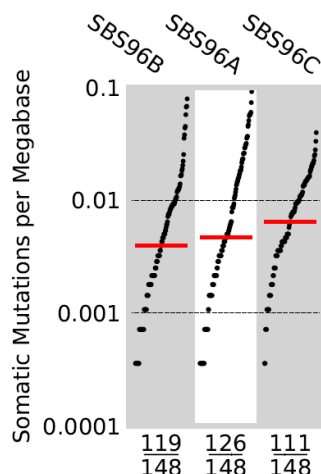


Figura 4. Firmas mutacionales SBS96A, 96B y 96C con valores de estabilidad y porcentaje de mutaciones asociado a cada firma. Los distintos colores representan los seis tipos básicos de sustituciones: C>A azul, C>G negro, C>T rojo, T>A gris, T>C verde, T>G rosa. El eje X muestra los 96 trinucleótidos y el eje Y, el porcentaje de mutaciones atribuidas a cada contexto trinucleotídico dentro de cada firma SBS96.

En cuanto a la carga mutacional por megabase y prevalencia, las firmas SBS96C y SBS96A presentan cargas superiores a la de SBS96B, que tiende a contribuir con menos mutaciones. SBS96A está presente en 126 de las 148 muestras, siendo la firma más frecuente; seguida de SBS96B en 119 y SBS96C en 111 (Fig.5).



*Showing samples with counts more than 0

Figura 5. Carga mutacional por megabase atribuida a cada firma mutacional. Cada punto corresponde a una muestra individual y la línea roja indica la mediana para cada firma. El eje X contiene las tres firmas de *novo* y en el eje Y se representa en escala logarítmica las mutaciones somáticas por megabase, permitiendo comparar amplias variaciones en carga mutacional. Los números en la parte inferior indican cuántas muestras presentan más de 0 mutaciones asignadas a la firma.

Otro de los resultados obtenidos fue la contribución de cada firma mutacional (SBS96A, B, C) en cada una de las muestras analizadas. La firma SBS96A predomina en un mayor número de muestras, lo que sugiere que el proceso biológico asociado está más extendido en la cohorte analizada (Fig. 6).

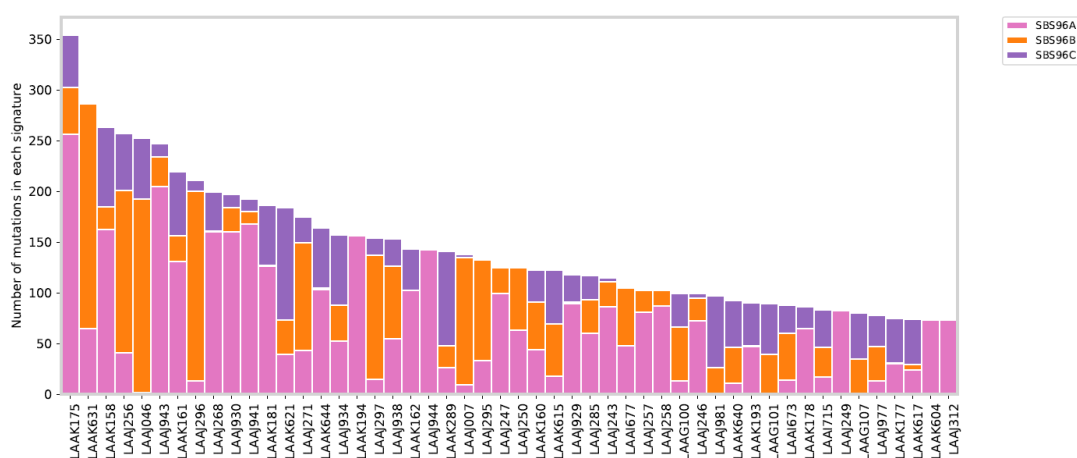


Figura 6. Distribución de mutaciones somáticas atribuidas a cada firma mutacional en las muestras tumorales analizadas. El eje X presenta las muestras ordenadas de mayor

a menor carga mutacional y el eje Y, el número de mutaciones en cada firma. Cada barra corresponde a una muestra individual y se observan las tres firmas con diferentes colores: SBS96A (rosa), SBS96B (naranja) o SBS96C (morado).

4.3 Análisis de exposiciones y distribución de muestras

Se realizó un análisis de componentes principales (PCA) para explorar la distribución de los pacientes en función de su firma dominante y se utilizaron los valores de exposición de las tres firmas SBS96A, SBS96B y SBS96C (Fig. 7). Los patrones observados muestran tres clústeres con solapamiento entre grupos debido a la presencia simultánea de varios procesos mutacionales en la mayoría de las muestras. Las muestras pertenecientes a SBS96A dominante, están hacia valores negativos de PC2. En cambio, las muestras referentes a SBS96B y SBS96C están hacia valores positivos de PC2, lo que implica que este eje diferencia esas muestras frente a las de SBS96A.

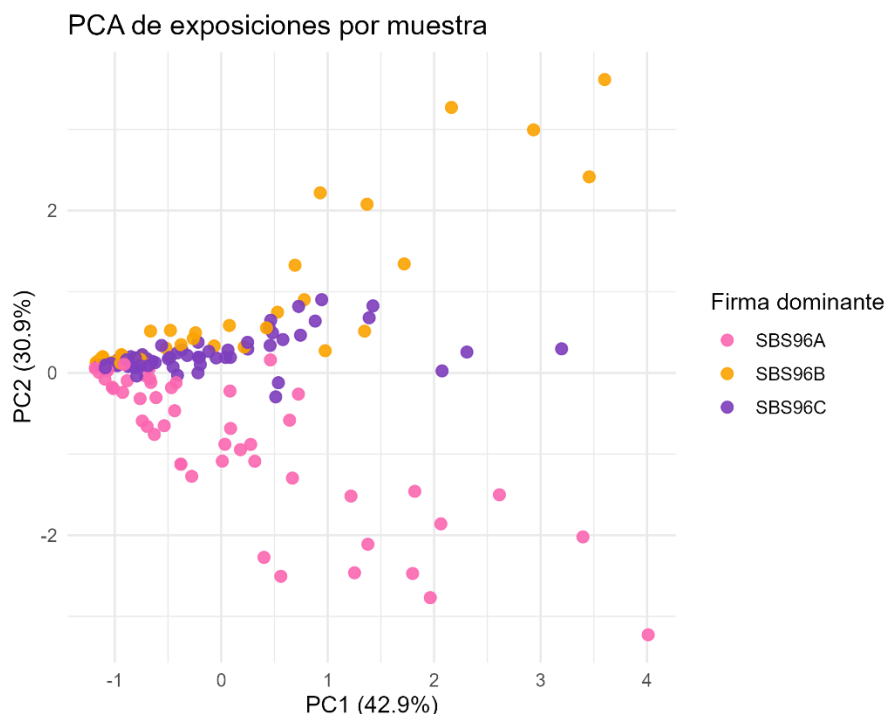


Figura 7. PCA basado en la firma dominante por cada paciente. Cada punto corresponde a una muestra y se colorea según la firma dominante: SBS96A, SBS96B y

SBS96C. Los ejes presentan el porcentaje de varianza explicada por el PC1 en el caso del eje X y PC2 del eje Y.

La distribución de exposiciones globales de cada firma en la cohorte se realizó mediante violin plots. Gran parte de la cohorte mostró exposiciones bajas a cada firma mientras que SBS96A presentó la cola derecha más pronunciada, correspondiente a un conjunto de muestras con valores de exposición excepcionalmente altos (outliers). La distribución de muestras en SBS96B fue más concentrada y SBS96C obtuvo la mediana más alta de número de mutaciones, pero con menos casos altos que en SBS96A (Fig. 8).

Se utilizó la regla de Tukey ($Q3 + 1.5 \cdot IQR$) para identificar muestras con exposición inusualmente alta a determinadas firmas. Se identificaron 13, 10 y 5 de estas muestras para SBS96A, SBS96B y SBS96C, respectivamente (Fig. 9).

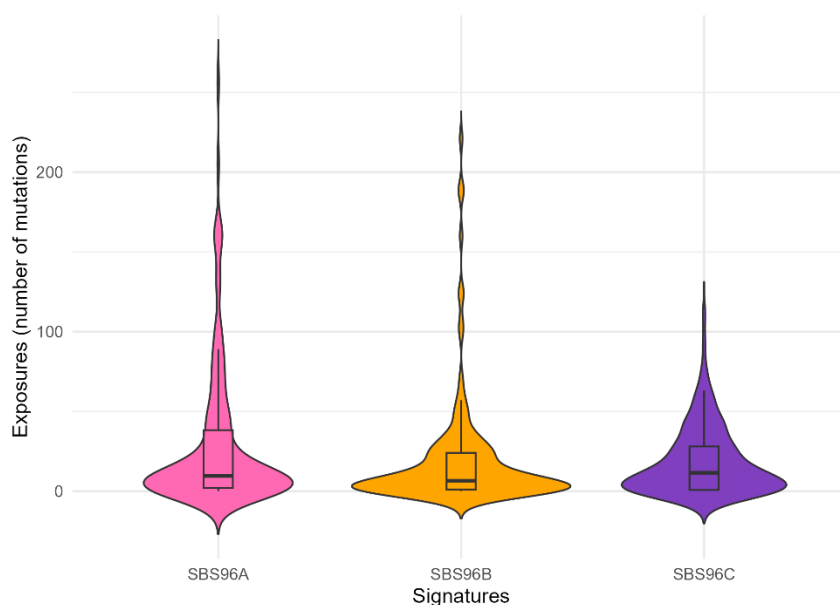


Figura 8. Distribución de las exposiciones por firma mutacional. Esta figura muestra violin plots con boxplots en su interior (mediana e IQR representados) para las firmas SBS96A, SBS96B y SBS96C en el eje X. El eje Y refleja la exposición a cada firma, el número de mutaciones asociadas a cada firma por muestra. La anchura del violín indica mayor número de muestras y las colas superiores muy alargadas implican exposición muy elevada.

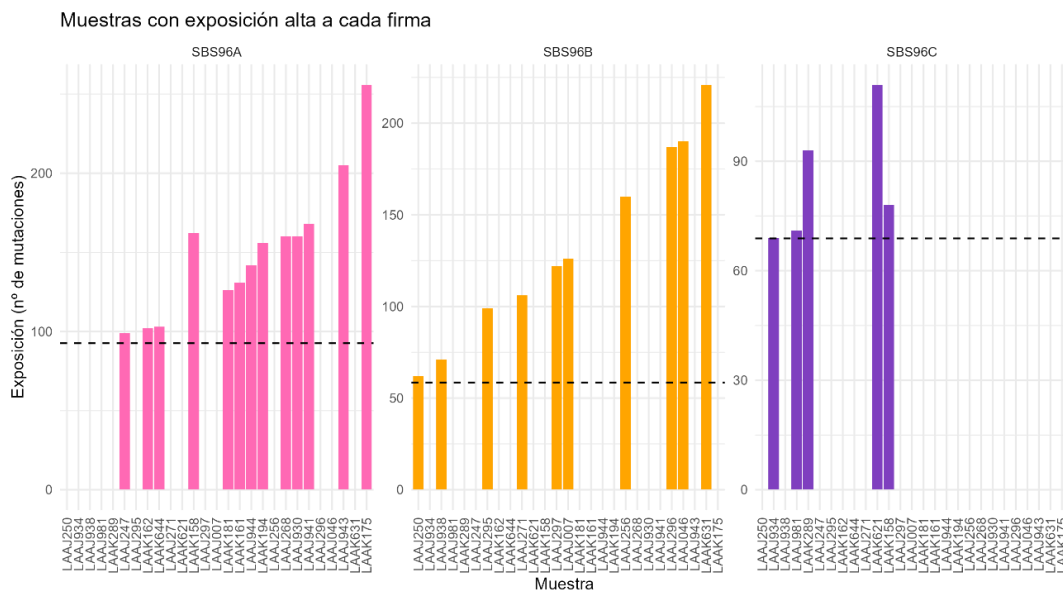


Figura 9. Representación de muestras altamente expuestas a cada firma mutacional. Cada barra del gráfico muestra el número de mutaciones (eje Y) correspondiente a muestras individuales (eje X) según una de las tres firmas de *novo* (SBS96A, SBS96B y SBS96C). La línea horizontal discontinua es el umbral de Tukey utilizado para mostrar valores atípicos altos.

4.4 Comparación con COSMIC

Las tres firmas identificadas se compararon con las firmas referentes del catálogo COSMIC v3.4 mediante análisis de similitud de coseno (Alexandrov et al., 2020; Sondka et al., 2024) (Fig. 10). Las nuevas firmas se presentaron en el contexto de las 96 posibles SBS según el trinucleótido:

-SBS96A presentó una similitud coseno entre la firma original y su reconstrucción de 0.9 con una combinación de SBS15 (52.6%) y SBS42 (47.4%). La firma mutacional SBS15 es una de las siete firmas asociadas a defectos en la reparación de errores de apareamiento del ADN (MMRd) y a inestabilidad de microsatélites (MSI). Las firmas MMRd como SBS15, se caracterizan por una alta frecuencia de mutaciones C > T y las muestras de alta inestabilidad de microsatélites (MSI-H) muestran una alta carga de mutación tanto para indels como para sustituciones de una sola base con altas proporciones de C > T y T > C, lo que concuerda con el patrón evidenciado en SBS96A y SBS96B (Fang et al., 2021). La prevalencia del cáncer de próstata MSI-H/MMRd muestra

frecuencias que varían del 1,2 % al 12,0 % en trabajos anteriores (Guedes et al., 2017; Pritchard et al., 2014). Estudios de secuenciación recientes del mCRPC mostraron que entre el 2 % y el 3 % de los tumores tienen una mayor carga de mutación que a menudo se asocia con alteraciones genómicas en genes asociados a MMR, lo que sugiere que la secuenciación tumoral puede ser un método eficiente para identificar este tipo de cáncer (Abida et al., 2017; Robinson et al., 2015). Por otro lado, la firma SBS42 se relaciona con la exposición en el entorno profesional a haloalcanos, compuestos químicos empleados en contextos industriales. Sin embargo, su presencia en esta cohorte no puede atribuirse relevancia sin disponer de información ambiental específica.

-SBS96B mostró la mayor similitud coseno (0.97) y se asoció con SBS1 (50.4%), caracterizada por la desaminación espontánea o enzimática de 5-metilcitosina a timina (una firma común del envejecimiento), complementada con SBS10b (12.8%) y SBS15 (36.8%). La etiología propuesta para SBS10b se relaciona con mutaciones en el dominio exonucleasa de la ADN polimerasa épsilon (POLE). Las firmas tanto SBS10a como SBS10b suelen generar un número muy elevado de mutaciones somáticas (más de 100 mutaciones por megabase) y las muestras con estas firmas se denominan hipermutadoras.

-SBS96C se correspondió principalmente con SBS5 (79.6%) y en menor medida con SBS50 (20.4%), alcanzando una similitud coseno de 0.877. SBS5 es una firma de etiología desconocida, pero su carga mutacional está incrementada en muestras de cáncer de vejiga con mutaciones en ERCC2 y en muchos tipos de cáncer debido al tabaquismo. SBS5 es similar a un reloj biológico, el número de mutaciones en la mayoría de los cánceres y células normales se correlaciona con la edad del individuo. Las tasas de adquisición de mutaciones SBS5 a lo largo del tiempo difieren entre distintos tipos de cáncer y de células normales. Estas diferencias no se correlacionan claramente con las tasas de división de células madre en los diferentes tejidos, ni con las diferencias en las tasas de mutación de SBS1. Además, SBS5 puede estar contaminada por SBS16 con etiología desconocida. Respecto a la firma SBS50, es una firma poco frecuente y de etiología desconocida no caracterizada en la documentación de COSMIC

v3.4. Puede representar un componente inespecífico o compartido con otras firmas sin definir.

La correspondencia entre firmas, etiología y la métrica de similitud de coseno se resumen en la Tabla 2.

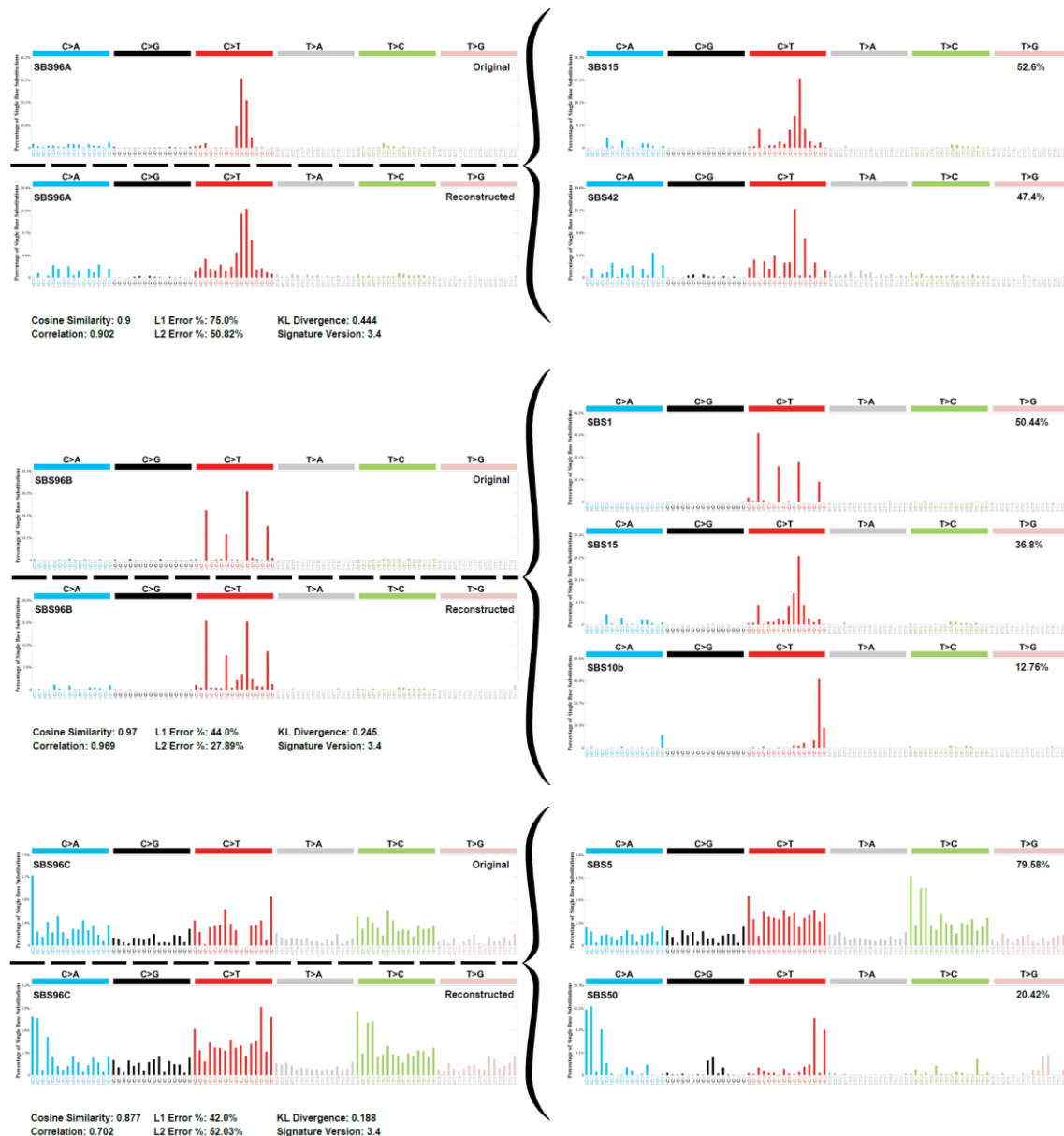
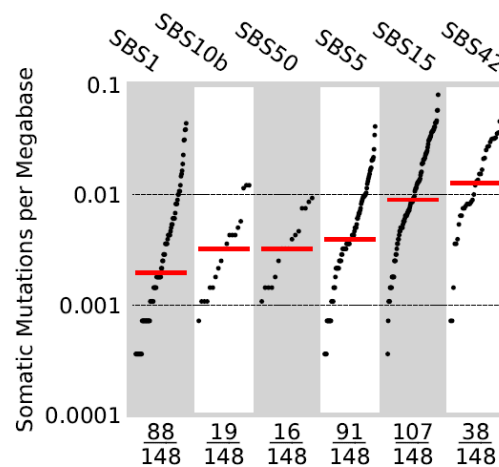


Figura 10: Descomposición de firmas SBS96A, B y C con métricas asociadas al ajuste del modelo (errores L1 y L2, divergencia KL y correlación). El análisis de similitud de coseno permitió comparar las firmas obtenidas de *novo* con el catálogo COSMIC v3.4. El panel izquierdo muestra los perfiles SBS96 (versión original vs reconstruida). En el derecho, aparece la combinación de firmas de referencia de COSMIC que mejor reconstruye cada firma de *novo* (con sus porcentajes y similitud de coseno).

Firma de <i>novo</i>	Firmas COSMIC	Similitud coseno	Etiología propuesta
SBS96A	SBS15, SBS42	0.9	MMRd/MSI (SBS15) Exposición a haloalcanos (SBS42)
SBS96B	SBS1, SBS15, SBS10b	0.97	Envejecimiento / desaminación 5-mC (SBS1) MMRd (SBS15) POLE exonucleasa (SBS10b)
SBS96C	SBS5, SBS50	0.877	“Reloj” (SBS5) Etiología desconocida (SBS50)

Tabla 2. Resumen de la relación entre las firmas de *novo* y firmas referenciadas en el catálogo COSMIC v3.4. La similitud coseno indica la concordancia global, tratándose de firmas idénticas si el valor es 1. Las etiologías inferidas proceden de las anotaciones de COSMIC y varios procesos mutacionales pueden coexistir en una misma muestra.

Para investigar la importancia de estas firmas COSMIC dentro de la cohorte de estudio, se analizó la carga mutacional por megabase asociada a cada una de ellas. Se observó que SBS15 y SBS42 aportaron una mayor contribución de mutaciones mientras que SBS1 fue la que menos. En orden de mayor a menor frecuencia de aparición en muestras se encuentran SBS15, SBS5 y SBS1. Por otro lado, SBS10b y SBS50 se detectaron en menor frecuencia (Fig. 11). En estudios clínicos recientes sobre mCRPC tratados con pembrolizumab, la carga mutacional tumoral (TMB) se estratificó en categorías de 0–5, 5–10 y ≥ 10 mutaciones/Mb, confirmando que la mayoría de los tumores de próstata metastásicos presentan una carga mutacional baja, mientras que solo un pequeño subgrupo alcanza valores ≥ 10 mutaciones/Mb compatibles con un fenotipo hipermutador (Zang et al., 2024). Por tanto, en nuestro estudio no se observaron casos con carga mutacional absoluta elevada.



*Showing samples with counts more than 0

Figura 11. Carga mutacional por megabase de las firmas referentes del catálogo COSMIC asociadas a las tres de *novo* SBS96A, SBS96B y SBS96C. Cada punto corresponde a una muestra individual y la línea roja indica la mediana para cada firma. En el eje X se muestran las firmas COSMIC con mayor relación: SBS1, SBS10b, SBS50, SBS5, SBS15 y SBS42. En el eje Y, se representan las mutaciones somáticas por megabase en escala logarítmica.

De forma complementaria, se examinó la distribución de mutaciones asignadas a estas firmas de referencia en cada muestra individual, visualizando la heterogeneidad intertumoral y la variabilidad en la contribución relativa de cada firma. Aunque en algunas muestras domina una sola firma, en la mayoría existe una combinación de varias, siendo SBS15 y SBS42 las más frecuentes en la población (Fig. 12).

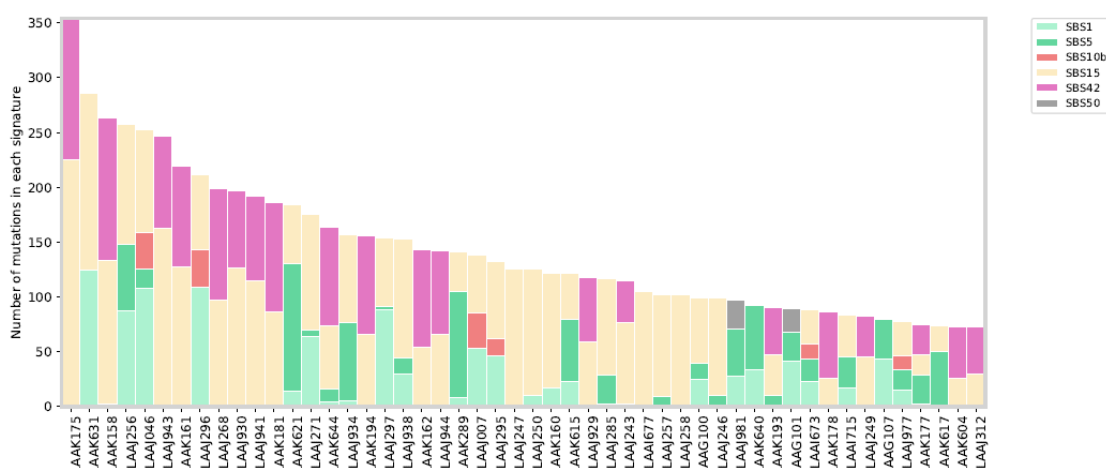


Figura 12. Contribución relativa de las mutaciones somáticas de firmas COSMIC seleccionadas en las muestras tumorales. Cada barra representa una muestra individual del eje X y se descompone por el número de mutaciones asociado a las firmas SBS1 (verde claro), SBS5 (verde oscuro), SBS10b (rojo), SBS15 (amarillo), SBS42 (morado) y SBS50 (gris). El eje Y indica el número de mutaciones en cada firma.

4.5 Análisis de enriquecimiento funcional de genes mutados

Se realizó un análisis de enriquecimiento funcional en términos de GO:Biological Process (GO:BP) sobre los genes con mutaciones no sinónimas identificados en las muestras correspondientes al tercil superior de exposición a cada firma mutacional (SBS96A, SBS96B y SBS96C) (Fig. 13).

-SBS96A mostró una marcada sobrerepresentación de procesos de respuesta al daño del ADN (DDR), con una fuerte activación de rutas de reparación por recombinación homóloga y señalización de puntos de control del ciclo celular. Estos términos se distribuyeron en torno a genes como *BRCA1*, *BRCA2*, *RAD51*, *PALB2*, *ATM*, *ATR*, *CHEK2*, *MLH1* y *MSH2*, que intervienen en reparar roturas de doble cadena o en corregir errores de apareamiento. *MSH6* presentó una posición periférica en la red unido al término de respuesta a radiación ultravioleta, pudiendo tener una contribución secundaria en la vía de MMR. Los resultados se correlacionan con la etiología MMRd/MSI, donde los defectos en la reparación de errores de apareamiento provocan una elevada inestabilidad genómica y activación compensatoria de vías de reparación y control del ciclo celular.

-SBS96B compartió gran parte de estas rutas, aunque con un enriquecimiento más pronunciado en recombinación homóloga. Los genes más conectados en el cnetplot —*BRCA1*, *BRCA2*, *RAD51*, *FANCD2*, *ATR*, *CHEK1/2* y *POLE*— indican una actividad intensa de mecanismos replicativos y de reparación post-replicación, además de la regulación del ciclo celular. Este patrón es compatible con una etiología mixta de la desaminación espontánea (SBS1) y la hipermutación por alteraciones en *POLE* (SBS10b). A diferencia de SBS96A, donde la activación de DDR compensa la alteración de MMR, en SBS96B estas rutas constituyen procesos intrínsecos al origen de la mutación, reflejando la necesidad de mantener la integridad genómica durante la replicación del ADN.

-SBS96C, aunque mantuvo la activación de la respuesta al daño del ADN, presentó términos más generales como la regulación positiva del proceso metabólico del ADN, compatibles con un perfil funcional más inespecífico, característico de las firmas SBS5 de tipo “reloj mutacional”. Los genes más relevantes incluyeron *TP53*, *ATR*, *ATM* y *PARP1*, junto con componentes de las vías de reparación Fanconi/BRCA (*FANCD2*, *SLX4*) y del sistema MMR (*MSH2*), que se vincularon a funciones de regulación metabólica y respuesta a radiación UV. Este perfil sugiere una actividad de mantenimiento del ADN y reparación replicativa de baja intensidad.

Cabe destacar que *TP53* y *NBN* se detectaron en las tres firmas, actuando como nodos centrales de la red DDR. Aunque las firmas difieren en su etiología (MMRd, replicativa o reloj), todas comparten la activación de mecanismos de vigilancia y reparación del ADN.

En conclusión, los resultados reflejan que las tres firmas de *novο* se asocian con procesos de conservación de la estabilidad genómica, mostrando la coexistencia de diferentes mecanismos de daño y reparación del ADN en la cohorte de estudio.

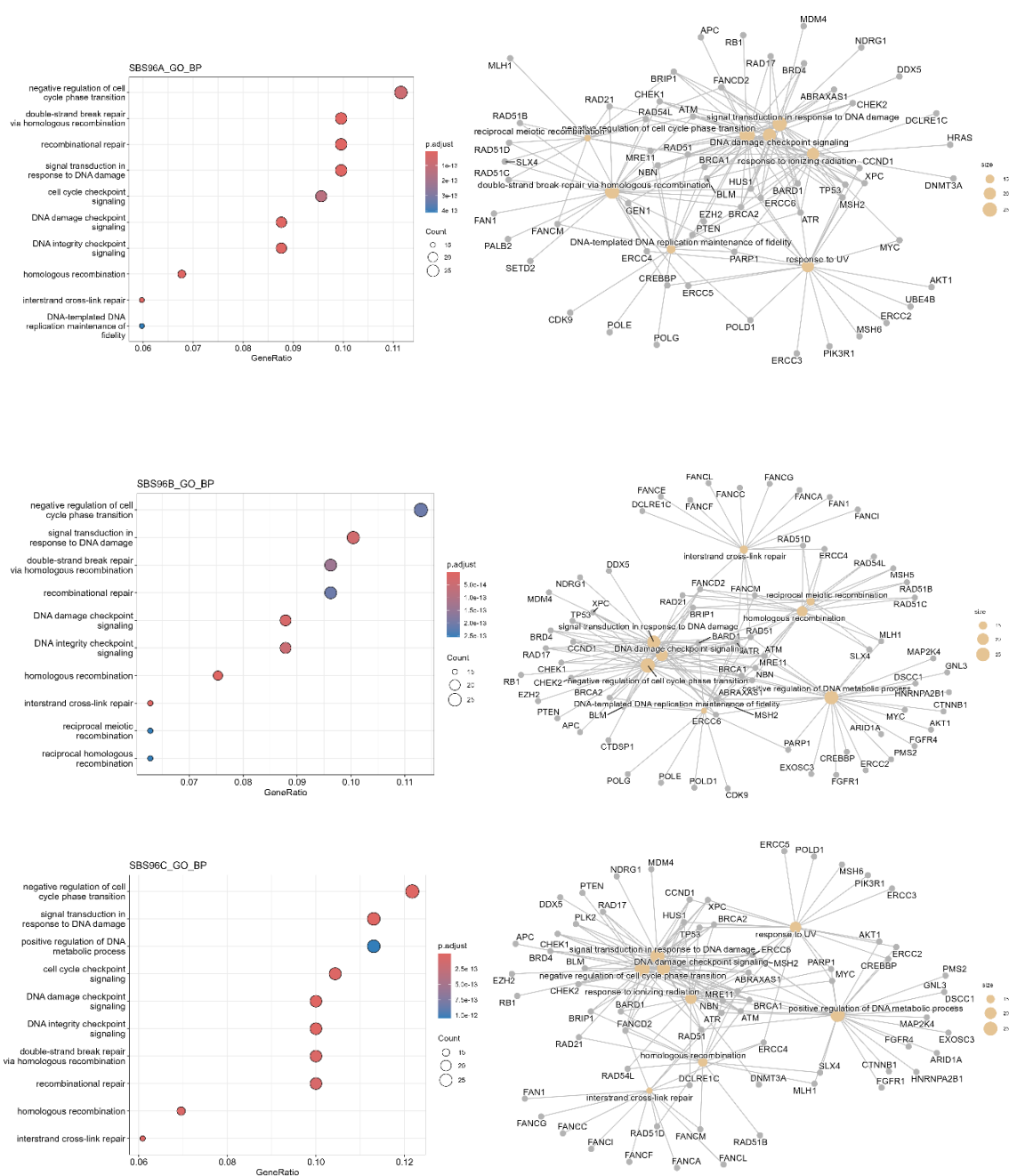


Figura 13. Enriquecimiento funcional GO:BP de genes con mutaciones no sinónimas por cada firma mutacional de *novo* (SBS96A, SBS96B y SBS96C). En la columna izquierda, los dotplots indican la proporción de genes anotados (eje X) por cada término (eje Y) en función del tamaño del punto. El p-valor ajustado se representa en color rojo si es más significativo o en azul si ocurre lo contrario. En la columna derecha, se representan cnetplots que conectan términos GO mediante aristas con genes. Los términos son los nodos de color marrón que varían de tamaño en función del número de genes asociados (nodos de color gris).

5. Discusión

La tecnología NMF ha sido utilizada en análisis de mutaciones en cáncer para encontrar patrones comunes de mutaciones, firmas mutacionales, que ocurren en muchos cánceres y probablemente tienen diferentes causas (Alexandrov, Nik-Zainal, Wedge, Campbell, et al., 2013). Asimismo, NMF se ha empleado en la identificación de otras fuentes de variación como tipos celulares, subtipos de enfermedad, estratificación de poblaciones, composición de tejidos y clonabilidad tumoral (Stein-O'Brien et al., 2018). Esto valida la aplicación del método en cáncer de próstata para descomposición en componentes latentes que facilitan la explicación de la heterogeneidad tumoral.

Las tres firmas obtenidas mediante la aplicación de NMF mostraron valores de estabilidad superiores a 0.93. La estabilidad funciona como criterio fundamental en la búsqueda de la solución más estable con un mínimo número de firmas que describan los datos perfectamente (Pelizzola et al., 2023).

La prevalencia y exposición mostraron que, aunque existía una firma dominante por muestra, la mayoría de los tumores coexpresan procesos mutacionales, un hallazgo previsible en cáncer de próstata por la concurrencia de mecanismos replicativos, reparación de ADN y envejecimiento. La distribución asimétrica de exposiciones (violin plots) y la presencia de colas derechas (especialmente en SBS96A) sugieren subgrupos con actividad mutacional intensificada asociada a defectos MMRd.

La convergencia observada en el análisis funcional de nuestras firmas en ejes DDR/HRR/MMR, es coherente con la biología del cáncer de próstata hereditario (HPCa), en el que genes MMR (*MLH1*, *MSH2*, *MSH6*, *PMS2*) y de recombinación homóloga (*BRCA1/2*, *ATM*, *PALB2*, *CHEK2*) se asocian con mayor riesgo de fenotipos más agresivos y metástasis (Pritchard et al., 2016; Vietri et al., 2021). El ensayo clínico PROfound evidenció que en el caso de defectos en genes HRR, particularmente en *BRCA1* y *BRCA2*, pacientes con mCRPC obtienen un beneficio clínico significativo con el uso de olaparib en monoterapia, un inhibidor de PARP, prolongando la supervivencia libre de progresión (de Bono et al., 2020). Adicionalmente, el ensayo PROpel ha demostrado que su combinación con el tratamiento hormonal abiraterona, mejora los resultados

independientemente de la presencia de mutaciones en genes HRR, mostrando una eficacia superior en portadores de mutaciones *BRCA* (Saad et al., 2025).

Uno de los principales factores de riesgo en cáncer de próstata es el envejecimiento, por tanto, SBS1 asociada a la firma de *novo* SBS96B y SBS5 a SBS96C, se consideran firmas “reloj”, ya que se correlacionan con la edad cronológica y replicativa de las células, apareciendo con mayor frecuencia en este tipo de tumor (Christensen et al., 2022; Wang et al., 2021). Del mismo modo, en estudios previos sobre pacientes de cáncer de próstata en PCAWG (Pelizzola et al., 2023) y en TCGA (Wang et al., 2021), se reportó SBS1 y SBS5 como firmas ampliamente presentes en este cáncer.

SBS96A se relaciona con la firma de COSMIC SBS15, sugiriendo un subgrupo de tumor de próstata con MMRd. Este resultado también concuerda con lo descrito por Wang et al. (2021), los cuales identificaron la firma SBS-sig2 con alta similitud a SBS15. Su prevalencia es relativamente baja (3–12%) pero se ha definido que los tumores MMRd/MSI-H engloban un subgrupo biológico y clínicamente distinto, con un fenotipo agresivo y mayor probabilidad de responder a inhibidores del punto de control inmunitario, como pembrolizumab (terapia anti-PD-1/PD-L1) en mCRPC (Graham et al., 2020). Sin embargo, la respuesta no es universal, pues existen variables adicionales que influyen en la eficacia de la inmunoterapia como el TMB o el microambiente inmunológico. Hasta el 30 % de los tumores MSI-H pueden tener bajo TMB (Vanderwalde et al., 2018) y el microambiente puede ser más inmunosupresor bajo terapias específicas como la quimioterapia (Nakazawa et al., 2017). A nivel molecular, este proceso se asocia a alteraciones en genes reparadores del ADN como *MLH1*, *MSH2*, *MSH6* y *PMS2*, cuya disfunción provoca inestabilidad de microsatélites. Además, se ha observado que el fenotipo MSI-H/MMRd puede adquirirse en fases avanzadas de la evolución tumoral, lo que prioriza el análisis de tejido metastásico y la consideración de pruebas longitudinales para su correcta detección (Abida et al., 2019).

El resto de las firmas SBS42, SBS10b y SBS50 aparecen en la población de estudio, pero no se han vinculado con el cáncer de próstata. SBS42 se ha relacionado con la exposición ambiental a haloalcanos presentes especialmente

en algunos gliomas de varones y originalmente en colangiocarcinomas ocupacionales (Mimaki et al., 2016). Los haloalcanos son utilizados en retardantes de llama, extintores de incendios y pesticidas (Krishnan et al., 2003), y se han relacionado con efectos tóxicos en células progenitoras de oligodendrocitos (Cohn et al., 2024). La presencia de esta firma puede considerarse ruido debido a ausencia de datos ambientales.

SBS10b está asociada a mutaciones en el dominio de POLE y se encuentra en tumores hipermutadores. Globalmente no son muy frecuentes (0.82%), hay enriquecimiento en cáncer de endometrio y colorectal y en menor proporción al 1 % en cánceres como el de próstata (Park & Pursell, 2019).

SBS50 es una firma conocida como artefactos de secuenciación, y su impacto disminuye en melanoma y cáncer de próstata cuando se aplican filtros con un umbral más elevado de contribución. Este filtrado puede contribuir a la reducción de los efectos del ruido en el procedimiento de extracción de firmas mutacionales (Zheng et al., 2023).

5.1 Limitaciones del estudio

Se señalan como limitaciones en el análisis de firmas mutacionales para el diagnóstico del cáncer que las firmas de COSMIC provienen de una mezcla de WES y WGS, lo que genera heterogeneidad y discrepancias y su aparición en pocos genomas con bajas contribuciones (Alexandrov et al., 2013). Algunas son artefactuales por errores de secuenciación o algoritmos de detección de mutaciones inadecuados (Alioto et al., 2015). La precisión de la extracción de firmas mutacionales disminuye cuando hay activos numerosos procesos mutacionales, bajo número de mutaciones y cuando las firmas son similares. Los subclones dentro de tumores que han perdido la actividad de ciertos procesos mutacionales aún contendrán sus características dentro del genoma, por tanto, es complejo distinguir los procesos históricos de los que se encuentran en curso. Tras comparar diferentes métodos, NMF cometió mayor error al descomponer las firmas (X. Huang et al., 2018), se basa en grandes cohortes de genomas de cáncer para extraer firmas con precisión y no puede analizar eficientemente muestras con alta carga mutacional. Una alternativa es el reajuste de firmas predefinidas a los datos genómicos de un solo paciente (Blokzijl et al., 2018;

Rosenthal et al., 2016), aunque aún enfrenta desafíos debido a la superposición entre firmas (Van Hoeck et al., 2019).

5.2 Futuras líneas de investigación

Para futuras líneas de trabajo, se considera la integración de firmas de CNA junto con las SBS. Herramientas bioinformáticas recientes, como *Sigminer*, han permitido realizar el primer análisis sistemático de firmas de CNA en cáncer de próstata, identificando cinco firmas con una capacidad superior a las SBS para estratificar a los pacientes y predecir la supervivencia, además de una asociación más estrecha con metástasis y estadios clínicos avanzados (Wang et al., 2021).

Estudios recientes han mostrado que existen asociaciones directas entre firmas mutacionales y patrones de expresión génica. El RNA-seq podría inferir o validar firmas mutacionales a partir de perfiles transcriptómicos (Jiang et al., 2023).

Asimismo, se contempla asignar etiologías conocidas a firmas prevalentes como SBS5, cuya carga mutacional puede variar en contextos específicos y el mecanismo molecular sigue siendo desconocido (Alexandrov et al., 2020). La integración de aductómica del ADN puede aportar mayor claridad etiológica al identificar modificaciones específicas y las especies reactivas implicadas, conectándolas con procesos endógenos o exógenos. Se espera que nuevas tecnologías, como la secuenciación por nanoporos y otras capaces de cuantificar y mapear aductos y mutaciones de ADN de manera simultánea, establezcan relaciones causales precisas y la evolución temporal de los tipos de mutación (Boysen et al., 2025). La IA, en particular el aprendizaje profundo, podría ayudar en la integración de la aductomía, las firmas mutacionales (Guarracino et al., 2023; Porubsky et al., 2023) y los datos multiómicos, para su uso con bases de datos completas como The Cancer Genome Atlas, Genomic Data Commons, bases de datos de aductomas de ADN (Guo et al., 2020) y bases de datos relacionadas con el exposoma (Barupal & Fiehn, 2019).

6. Conclusiones

1. La aplicación de NMF mediante la herramienta SigProfilerExtractor a la matriz SBS96 de la cohorte mCRPC, permitió la extracción de tres firmas de *novus* (SBS96A, 96B y 96C) con una alta estabilidad (≥ 0.93). En consecuencia, la metodología del análisis es robusta y consigue validar el uso de la factorización en matrices no negativa para identificar eficazmente procesos mutacionales latentes en cáncer de próstata.
2. La comparación con el catálogo COSMIC permitió investigar sobre las etiologías biológicas de las firmas obtenidas. SBS96B se asocia fundamentalmente con SBS1 con la mayor similitud de coseno (0.97) y SBS96C con SBS5, ambas consideradas firmas “reloj” de envejecimiento, resultando lógico por la edad avanzada de los pacientes de la cohorte estudiada. No obstante, SBS96A se asemeja a SBS15 y SBS42, reflejando un defecto en la reparación de errores de apareamiento (MMRd) y posible inestabilidad de microsatélites (MSI), lo cual constituye un descubrimiento clínico importante pues este subgrupo es sensible potencialmente a inmunoterapia en mCRPC.
3. La carga mutacional por megabase fue mayor en SBS96A y SBS96C que en SBS96B. SBS96A fue la más prevalente, presente en 126 de las 148 muestras, lo que indica que MMRd es un proceso mutacional destacado en esta cohorte de mCRPC. Los tumores MMRd/MSI-H, aunque aparecen con baja frecuencia en cáncer de próstata (1–12%), representan un subgrupo sensible a inmunoterapia con inhibidores de checkpoint como el fármaco pembrolizumab, así como a inhibidores de PARP como olaparib en presencia de alteraciones en genes de HRR. Al mismo tiempo, la elevada carga mutacional en SBS96C respalda la función de las firmas reloj asociadas al envejecimiento como procesos relevantes en la evolución molecular del tumor.
4. Se identificaron firmas como SBS42, SBS10b y SBS50 que probablemente representan ruido o artefactos en el cáncer de próstata, siendo necesaria la comprobación de datos ambientales, evidencia de mutaciones en POLE con fenotipo hipermutador y la aplicación de filtros para control de artefactos.

5. SigProfilerExtractor, basado en NMF, ofrece una extracción reproducible de firmas mutacionales y estratificación de pacientes. Sin embargo, la presencia de artefactos, las firmas similares entre sí y la necesidad de cohortes grandes limitan su extrapolación al diagnóstico individual. Aunque existen estas limitaciones, los resultados obtenidos sustentan la importancia metodológica de NMF y resaltan su aplicabilidad clínica: las firmas reloj proporcionan información sobre la acumulación progresiva de mutaciones ligada a la edad, mientras que las firmas asociadas a MMRd constituyen biomarcadores para la medicina de precisión en subgrupos determinados de pacientes.

7. Bibliografía

- Abida, W., Armenia, J., Gopalan, A., Brennan, R., Walsh, M., Barron, D., Danila, D., Rathkopf, D., Morris, M., Slovin, S., McLaughlin, B., Curtis, K., Hyman, D. M., Durack, J. C., Solomon, S. B., Arcila, M. E., Zehir, A., Syed, A., Gao, J., ... Scher, H. I. (2017). Prospective Genomic Profiling of Prostate Cancer Across Disease States Reveals Germline and Somatic Alterations That May Affect Clinical Decision Making. *JCO Precision Oncology*, 2017. <https://doi.org/10.1200/PO.17.00029>
- Abida, W., Cheng, M. L., Armenia, J., Middha, S., Autio, K. A., Vargas, H. A., Rathkopf, D., Morris, M. J., Danila, D. C., Slovin, S. F., Carbone, E., Barnett, E. S., Hullings, M., Hechtman, J. F., Zehir, A., Shia, J., Jonsson, P., Stadler, Z. K., Srinivasan, P., ... Scher, H. I. (2019). Analysis of the Prevalence of Microsatellite Instability in Prostate Cancer and Response to Immune Checkpoint Blockade. *JAMA Oncology*, 5(4), 471–478. <https://doi.org/10.1001/jamaoncol.2018.5801>
- Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., Boot, A., Covington, K. R., Gordenin, D. A., Bergstrom, E. N., Islam, S. M. A., Lopez-Bigas, N., Klimczak, L. J., McPherson, J. R., Morganella, S., Sabarinathan, R., Wheeler, D. A., Mustonen, V., Getz, G., ... Stratton, M. R. (2020). The repertoire of mutational signatures in human cancer. *Nature*, 578(7793), 94–101. <https://doi.org/10.1038/s41586-020-1943-3>
- Alexandrov, L. B., Nik-zainal, S., Wedge, D. C., & Aparicio, S. A. J. R. (2014). Europe PMC Funders Group Signatures of mutational processes in human cancer. *Nature*, 500(7463), 415–421. <https://doi.org/10.1038/nature12477>. Signatures
- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A.-L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjörd, J. E., Foekens, J. A., ... Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature*, 500(7463), 415–421. <https://doi.org/10.1038/nature12477>

- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J., & Stratton, M. R. (2013). Deciphering signatures of mutational processes operative in human cancer. *Cell Reports*, 3(1), 246–259.
<https://doi.org/10.1016/j.celrep.2012.12.008>
- Alexandrov, L. B., & Stratton, M. R. (2014). Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Current Opinion in Genetics & Development*, 24(100), 52–60.
<https://doi.org/10.1016/j.gde.2013.11.014>
- Alioto, T. S., Buchhalter, I., Derdak, S., Hutter, B., Eldridge, M. D., Hovig, E., Heisler, L. E., Beck, T. A., Simpson, J. T., Tonon, L., Sertier, A.-S., Patch, A.-M., Jäger, N., Ginsbach, P., Drews, R., Paramasivam, N., Kabbe, R., Chotewutmontri, S., Diessl, N., ... Gut, I. G. (2015). A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. *Nature Communications*, 6, 10001.
<https://doi.org/10.1038/ncomms10001>
- Bach, C., Pisipati, S., Daneshwar, D., Wright, M., Rowe, E., Gillatt, D., Persad, R., & Koupparis, A. (2014). The status of surgery in the management of high-risk prostate cancer. *Nature Reviews. Urology*, 11(6), 342–351.
<https://doi.org/10.1038/nrurol.2014.100>
- Bamford, S., Dawson, E., Forbes, S., Clements, J., Pettett, R., Dogan, A., Flanagan, A., Teague, J., Futreal, P. A., Stratton, M. R., & Wooster, R. (2004). The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *British Journal of Cancer*, 91(2), 355–358.
<https://doi.org/10.1038/sj.bjc.6601894>
- Barupal, D. K., & Fiehn, O. (2019). Generating the Blood Exposome Database Using a Comprehensive Text Mining and Database Fusion Approach. *Environmental Health Perspectives*, 127(9), 97008.
<https://doi.org/10.1289/EHP4713>
- Bergstrom, E. N., Huang, M. N., Mahto, U., Barnes, M., Stratton, M. R., Rozen, S. G., & Alexandrov, L. B. (2019). SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events. *BMC*

Genomics, 20(1), 685. <https://doi.org/10.1186/s12864-019-6041-2>

- Blokzijl, F., Janssen, R., van Boxtel, R., & Cuppen, E. (2018). MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Medicine*, 10(1), 33. <https://doi.org/10.1186/s13073-018-0539-0>
- Bluemn, E. G., & Nelson, P. S. (2012). The androgen/androgen receptor axis in prostate cancer. *Current Opinion in Oncology*, 24(3), 251–257. <https://doi.org/10.1097/CCO.0b013e32835105b3>
- Boysen, G., Alexandrov, L. B., Rahbari, R., Nookaew, I., Ussery, D., Chao, M.-R., Hu, C.-W., & Cooke, M. S. (2025). Investigating the origins of the mutational signatures in cancer. *Nucleic Acids Research*, 53(1). <https://doi.org/10.1093/nar/gkae1303>
- Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3), 229–263. <https://doi.org/10.3322/caac.21834>
- Chen, J., Zhang, D., Yan, W., Yang, D., & Shen, B. (2013). Translational bioinformatics for diagnostic and prognostic prediction of prostate cancer in the next-generation sequencing era. *BioMed Research International*, 2013, 901578. <https://doi.org/10.1155/2013/901578>
- Christensen, D. S., Ahrenfeldt, J., Sokač, M., Kisistók, J., Thomsen, M. K., Maretty, L., McGranahan, N., & Birkbak, N. J. (2022). Treatment Represents a Key Driver of Metastatic Cancer Evolution. *Cancer Research*, 82(16), 2918–2927. <https://doi.org/10.1158/0008-5472.CAN-22-0562>
- Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E. S., & Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology*, 31(3), 213–219. <https://doi.org/10.1038/nbt.2514>

- Cohn, E. F., Clayton, B. L. L., Madhavan, M., Lee, K. A., Yacoub, S., Fedorov, Y., Scavuzzo, M. A., Paul Friedman, K., Shafer, T. J., & Tesar, P. J. (2024). Pervasive environmental chemicals impair oligodendrocyte development. *Nature Neuroscience*, 27(5), 836–845. <https://doi.org/10.1038/s41593-024-01599-2>
- Crocetto, F., Musone, M., Chianese, S., Conforti, P., Digitale Selvaggio, G., Caputo, V. F., Falabella, R., Del Giudice, F., Giulioni, C., Cafarelli, A., Lucarelli, G., Busetto, G. M., Ferro, M., Barone, B., Zattoni, F., & Terracciano, D. (2025). Blood and urine-based biomarkers in prostate cancer: Current advances, clinical applications, and future directions. *Journal of Liquid Biopsy*, 9(May), 100305. <https://doi.org/10.1016/j.jlb.2025.100305>
- de Bono, J., Mateo, J., Fizazi, K., Saad, F., Shore, N., Sandhu, S., Chi, K. N., Sartor, O., Agarwal, N., Olmos, D., Thiery-Vuillemin, A., Twardowski, P., Mehra, N., Goessl, C., Kang, J., Burgents, J., Wu, W., Kohlmann, A., Adelman, C. A., & Hussain, M. (2020). Olaparib for Metastatic Castration-Resistant Prostate Cancer. *New England Journal of Medicine*, 382(22), 2091–2102. <https://doi.org/10.1056/nejmoa1911440>
- Fang, H., Zhu, X., Yang, H., Oh, J., Barbour, J. A., & Wong, J. W. H. (2021). Deficiency of replication-independent DNA mismatch repair drives a 5-methylcytosine deamination mutational signature in cancer. *Science Advances*, 7(45), eabg4398. <https://doi.org/10.1126/sciadv.abg4398>
- Fenton, S. E., VanderWeeler, D. J., Rebbeck, T. R., & Chen, D. L. (2024). Advancing Prostate Cancer Care: Treatment Approaches to Precision Medicine, Biomarker Innovations, and Equitable Access. *American Society of Clinical Oncology Educational Book. American Society of Clinical Oncology. Annual Meeting*, 44(3), e433138. https://doi.org/10.1200/EDBK_433138
- Févotte, C., Bertin, N., & Durrieu, J.-L. (2009). Nonnegative matrix factorization with the Itakura-Saito divergence: with application to music analysis. *Neural Computation*, 21(3), 793–830. <https://doi.org/10.1162/neco.2008.04->

08-771

- Forbes, S. A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C. G., Ward, S., Dawson, E., Ponting, L., Stefancsik, R., Harsha, B., Kok, C. Y., Jia, M., Jubb, H., Sondka, Z., Thompson, S., De, T., & Campbell, P. J. (2017). COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Research*, 45(D1), D777–D783. <https://doi.org/10.1093/nar/gkw1121>
- Fowler, J. C., & Jones, P. H. (2022). Somatic Mutation: What Shapes the Mutational Landscape of Normal Epithelia? *Cancer Discovery*, 12(7), 1642–1655. <https://doi.org/10.1158/2159-8290.CD-22-0145>
- Gonzalo, E. S. (2013). Molecular biology in oncology: what a physician should know. *Revista Medica Clinica Las Condes*, 24(4), 563–570. [https://doi.org/10.1016/S0716-8640\(13\)70196-2](https://doi.org/10.1016/S0716-8640(13)70196-2)
- Graham, L. S., Montgomery, B., Cheng, H. H., Yu, E. Y., Nelson, P. S., Pritchard, C., Erickson, S., Alva, A., & Schweizer, M. T. (2020). Mismatch repair deficiency in metastatic prostate cancer: Response to PD-1 blockade and standard therapies. *PloS One*, 15(5), e0233260. <https://doi.org/10.1371/journal.pone.0233260>
- Guarracino, A., Buonaiuto, S., de Lima, L. G., Potapova, T., Rhie, A., Koren, S., Rubinstein, B., Fischer, C., Gerton, J. L., Phillippy, A. M., Colonna, V., & Garrison, E. (2023). Recombination between heterologous human acrocentric chromosomes. *Nature*, 617(7960), 335–343. <https://doi.org/10.1038/s41586-023-05976-y>
- Guedes, L. B., Antonarakis, E. S., Schweizer, M. T., Mirkheshti, N., Almutairi, F., Park, J. C., Glavaris, S., Hicks, J., Eisenberger, M. A., De Marzo, A. M., Epstein, J. I., Isaacs, W. B., Eshleman, J. R., Pritchard, C. C., & Lotan, T. L. (2017). MSH2 Loss in Primary Prostate Cancer. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, 23(22), 6863–6874. <https://doi.org/10.1158/1078-0432.CCR-17-0955>
- Guo, J., Turesky, R. J., Tarifa, A., DeCaprio, A. P., Cooke, M. S., Walmsley, S.

- J., & Villalta, P. W. (2020). Development of a DNA Adductome Mass Spectral Database. *Chemical Research in Toxicology*, 33(4), 852–854. <https://doi.org/10.1021/acs.chemrestox.0c00031>
- Hien, L. T. K., & Gillis, N. (2021). Algorithms for Nonnegative Matrix Factorization with the Kullback–Leibler Divergence. *Journal of Scientific Computing*, 87(3). <https://doi.org/10.1007/s10915-021-01504-0>
- Hjelmberg, J. B., Scheike, T., Holst, K., Skytthe, A., Penney, K. L., Graff, R. E., Pukkala, E., Christensen, K., Adami, H.-O., Holm, N. V., Nuttall, E., Hansen, S., Hartman, M., Czene, K., Harris, J. R., Kaprio, J., & Mucci, L. A. (2014). The heritability of prostate cancer in the Nordic Twin Study of Cancer. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 23(11), 2303–2310. <https://doi.org/10.1158/1055-9965.EPI-13-0568>
- Hodel, K. P., Sun, M. J. S., Ungerleider, N., Park, V. S., Williams, L. G., Bauer, D. L., Immethun, V. E., Wang, J., Suo, Z., Lu, H., McLachlan, J. B., & Pursell, Z. F. (2020). POLE Mutation Spectra Are Shaped by the Mutant Allele Identity, Its Abundance, and Mismatch Repair Status. *Molecular Cell*, 78(6), 1166–1177.e6. <https://doi.org/10.1016/j.molcel.2020.05.012>
- Huang, A. (2008). Similarity measures for text document clustering. *Proceedings of the 6th New Zealand Computer Science Research Student Conference*.
- Huang, X., Wojtowicz, D., & Przytycka, T. M. (2018). Detecting presence of mutational signatures in cancer with confidence. *Bioinformatics (Oxford, England)*, 34(2), 330–337. <https://doi.org/10.1093/bioinformatics/btx604>
- Islam, S. M. A., Díaz-Gay, M., Wu, Y., Barnes, M., Vangara, R., Bergstrom, E. N., He, Y., Vella, M., Wang, J., Teague, J. W., Clapham, P., Moody, S., Senkin, S., Li, Y. R., Riva, L., Zhang, T., Gruber, A. J., Steele, C. D., Otl, B., ... Alexandrov, L. B. (2022). Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *Cell Genomics*, 2(11). <https://doi.org/10.1016/j.xgen.2022.100179>

- Jiang, L., Yu, H., & Guo, Y. (2023). Modeling the relationship between gene expression and mutational signature. *Quantitative Biology (Beijing, China)*, 11(1), 31–43. <https://doi.org/10.15302/j-qb-022-0309>
- Kim, H., Lee, A. J., Lee, J., Chun, H., Ju, Y. S., & Hong, D. (2019). FIREVAT: finding reliable variants without artifacts in human cancer samples using etiologically relevant mutational signatures. *Genome Medicine*, 11(1), 81. <https://doi.org/10.1186/s13073-019-0695-x>
- Kim, Y.-A., Leiserson, M. D. M., Moorjani, P., Sharan, R., Wojtowicz, D., & Przytycka, T. M. (2021). Mutational Signatures: From Methods to Mechanisms. *Annual Review of Biomedical Data Science*, 4, 189–206. <https://doi.org/10.1146/annurev-biodatasci-122320-120920>
- Kirişçi, M. (2023). New cosine similarity and distance measures for Fermatean fuzzy sets and TOPSIS approach. *Knowledge and Information Systems*, 65(2), 855–868. <https://doi.org/10.1007/s10115-022-01776-4>
- Krishnan, G., Felini, M., Carozza, S. E., Miike, R., Chew, T., & Wrensch, M. (2003). Occupation and adult gliomas in the San Francisco Bay Area. *Journal of Occupational and Environmental Medicine*, 45(6), 639–647. <https://doi.org/10.1097/01.jom.0000069245.06498.48>
- Levatić, J., Salvadores, M., Fuster-Tormo, F., & Supek, F. (2022). Mutational signatures are markers of drug sensitivity of cancer cells. *Nature Communications*, 13(1), 2926. <https://doi.org/10.1038/s41467-022-30582-3>
- Lin, X., & Boutros, P. C. (2020). Optimization and expansion of non-negative matrix factorization. *BMC Bioinformatics*, 21(1), 7. <https://doi.org/10.1186/s12859-019-3312-5>
- Manders, F., Brandsma, A. M., de Kanter, J., Verheul, M., Oka, R., van Roosmalen, M. J., van der Roest, B., van Hoeck, A., Cuppen, E., & van Boxtel, R. (2022). MutationalPatterns: the one stop shop for the analysis of mutational processes. *BMC Genomics*, 23(1), 134. <https://doi.org/10.1186/s12864-022-08357-3>
- Medo, M., Ng, C. K. Y., & Medová, M. (2024). A comprehensive comparison of

tools for fitting mutational signatures. *Nature Communications*, 15(1), 9467.
<https://doi.org/10.1038/s41467-024-53711-6>

Meier, B., Volkova, N. V, Hong, Y., Schofield, P., Campbell, P. J., Gerstung, M., & Gartner, A. (2018). Mutational signatures of DNA mismatch repair deficiency in *C. elegans* and human cancers. *Genome Research*, 28(5), 666–675. <https://doi.org/10.1101/gr.226845.117>

Mimaki, S., Totsuka, Y., Suzuki, Y., Nakai, C., Goto, M., Kojima, M., Arakawa, H., Takemura, S., Tanaka, S., Marubashi, S., Kinoshita, M., Matsuda, T., Shibata, T., Nakagama, H., Ochiai, A., Kubo, S., Nakamori, S., Esumi, H., & Tsuchihara, K. (2016). Hypermutation and unique mutational signatures of occupational cholangiocarcinoma in printing workers exposed to haloalkanes. *Carcinogenesis*, 37(8), 817–826.
<https://doi.org/10.1093/carcin/bgw066>

Nakazawa, M., Paller, C., & Kyprianou, N. (2017). Mechanisms of Therapeutic Resistance in Prostate Cancer. *Current Oncology Reports*, 19(2), 13.
<https://doi.org/10.1007/s11912-017-0568-7>

Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L. A., Menzies, A., Martin, S., Leung, K., Chen, L., Leroy, C., Ramakrishna, M., Rance, R., Lau, K. W., Mudie, L. J., ... Stratton, M. R. (2012). Mutational processes molding the genomes of 21 breast cancers. *Cell*, 149(5), 979–993.
<https://doi.org/10.1016/j.cell.2012.04.024>

Olmos, D, Lorente, D., Alameda, D., Cattrini, C., Romero-Laorden, N., Lozano, R., Lopez-Casas, P. P., Jambrina, A., Capone, C., Vanden Broecke, A. M., Trevisan, M., Van Sanden, S., Jürgens, A., Herrera-Imbroda, B., & Castro, E. (2024). Treatment patterns and outcomes in metastatic castration-resistant prostate cancer patients with and without somatic or germline alterations in homologous recombination repair genes. *Annals of Oncology : Official Journal of the European Society for Medical Oncology*, 35(5), 458–472. <https://doi.org/10.1016/j.annonc.2024.01.011>

Olmos, David, Lorente, D., Alameda, D., Cattrini, C., Romero-Laorden, N.,

- Lozano, R., Lopez-Casas, P. P., Capone, C., Vanden Broecke, A. M., Trevisan, M., Van Sanden, S., Jürgens, A., Herrera-Imbroda, B., & Castro, E. (2023). Presence of somatic/germline homologous recombination repair (HRR) mutations and outcomes in metastatic castration-resistant prostate cancer (mCRPC) patients (pts) receiving first-line (1L) treatment stratified by BRCA status. *Journal of Clinical Oncology*, 41(16_suppl), 5003–5003. https://doi.org/10.1200/jco.2023.41.16_suppl.5003
- Pan-cancer analysis of whole genomes. (2020). *Nature*, 578(7793), 82–93. <https://doi.org/10.1038/s41586-020-1969-6>
- Park, V. S., & Pursell, Z. F. (2019). POLE proofreading defects: Contributions to mutagenesis and cancer. *DNA Repair*, 76, 50–59. <https://doi.org/10.1016/j.dnarep.2019.02.007>
- Pelizzola, M., Laursen, R., & Hobolth, A. (2023). Model selection and robust inference of mutational signatures using Negative Binomial non-negative matrix factorization. *BMC Bioinformatics*, 24(1), 1–24. <https://doi.org/10.1186/s12859-023-05304-1>
- Porubsky, D., Harvey, W. T., Rozanski, A. N., Ebler, J., Höps, W., Ashraf, H., Hasenfeld, P., Paten, B., Sanders, A. D., Marschall, T., Korbel, J. O., & Eichler, E. E. (2023). Inversion polymorphism in a complete human genome assembly. *Genome Biology*, 24(1), 100. <https://doi.org/10.1186/s13059-023-02919-8>
- Póti, Á., Gyergyák, H., Németh, E., Rusz, O., Tóth, S., Kovácsné, C., Chen, D., Szikriszt, B., Spisák, S., Takeda, S., Szakács, G., Szallasi, Z., Richardson, A. L., & Szüts, D. (2019). Correlation of homologous recombination deficiency induced mutational signatures with sensitivity to PARP inhibitors and cytotoxic agents. *Genome Biology*, 20(1), 240. <https://doi.org/10.1186/s13059-019-1867-0>
- Pritchard, C. C., Mateo, J., Walsh, M. F., De Sarkar, N., Abida, W., Beltran, H., Garofalo, A., Gulati, R., Carreira, S., Eeles, R., Elemento, O., Rubin, M. A., Robinson, D., Lonigro, R., Hussain, M., Chinnaiyan, A., Vinson, J., Filipenko, J., Garraway, L., ... Nelson, P. S. (2016). Inherited DNA-Repair

Gene Mutations in Men with Metastatic Prostate Cancer. *The New England Journal of Medicine*, 375(5), 443–453.

<https://doi.org/10.1056/NEJMoa1603144>

Pritchard, C. C., Morrissey, C., Kumar, A., Zhang, X., Smith, C., Coleman, I., Salipante, S. J., Milbank, J., Yu, M., Grady, W. M., Tait, J. F., Corey, E., Vessella, R. L., Walsh, T., Shendure, J., & Nelson, P. S. (2014). Complex MSH2 and MSH6 mutations in hypermutated microsatellite unstable advanced prostate cancer. *Nature Communications*, 5, 4988.

<https://doi.org/10.1038/ncomms5988>

Robinson, D., Van Allen, E. M., Wu, Y.-M., Schultz, N., Lonigro, R. J., Mosquera, J.-M., Montgomery, B., Taplin, M.-E., Pritchard, C. C., Attard, G., Beltran, H., Abida, W., Bradley, R. K., Vinson, J., Cao, X., Vats, P., Kunju, L. P., Hussain, M., Feng, F. Y., ... Chinnaiyan, A. M. (2015). Integrative clinical genomics of advanced prostate cancer. *Cell*, 161(5), 1215–1228. <https://doi.org/10.1016/j.cell.2015.05.001>

Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S., & Swanton, C. (2016). DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biology*, 17, 31. <https://doi.org/10.1186/s13059-016-0893-4>

Saad, F., Armstrong, A. J., Shore, N., George, D. J., Oya, M., Sugimoto, M., McKay, R. R., Hussain, M., & Clarke, N. W. (2025). Olaparib Monotherapy or in Combination with Abiraterone for the Treatment of Patients with Metastatic Castration-Resistant Prostate Cancer (mCRPC) and a BRCA Mutation. *Targeted Oncology*, 20(3), 445–466.

<https://doi.org/10.1007/s11523-025-01146-4>

Santucci, C., Carioli, G., Bertuccio, P., Malvezzi, M., Pastorino, U., Boffetta, P., Negri, E., Bosetti, C., & La Vecchia, C. (2020). Progress in cancer mortality, incidence, and survival: a global overview. *European Journal of Cancer Prevention : The Official Journal of the European Cancer Prevention Organisation (ECP)*, 29(5), 367–381.

<https://doi.org/10.1097/CEJ.0000000000000594>

- Sondka, Z., Dhir, N. B., Carvalho-Silva, D., Jupe, S., Madhumita, McLaren, K., Starkey, M., Ward, S., Wilding, J., Ahmed, M., Argasinska, J., Beare, D., Chawla, M. S., Duke, S., Fasanella, I., Neogi, A. G., Haller, S., Hetenyi, B., Hodges, L., ... Teague, J. (2024). COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic Acids Research*, 52(D1), D1210–D1217. <https://doi.org/10.1093/nar/gkad986>
- Stein-O'Brien, G. L., Arora, R., Culhane, A. C., Favorov, A. V, Garmire, L. X., Greene, C. S., Goff, L. A., Li, Y., Ngom, A., Ochs, M. F., Xu, Y., & Fertig, E. J. (2018). Enter the Matrix: Factorization Uncovers Knowledge from Omics. *Trends in Genetics : TIG*, 34(10), 790–805. <https://doi.org/10.1016/j.tig.2018.07.003>
- Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature*, 458(7239), 719–724. <https://doi.org/10.1038/nature07943>
- Termini, D., Den Hartogh, D. J., Jaglanian, A., & Tsiani, E. (2020). Curcumin against Prostate Cancer: Current Evidence. *Biomolecules*, 10(11). <https://doi.org/10.3390/biom10111536>
- Torre, L. A., Siegel, R. L., Ward, E. M., & Jemal, A. (2016). Global Cancer Incidence and Mortality Rates and Trends--An Update. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 25(1), 16–27. <https://doi.org/10.1158/1055-9965.EPI-15-0578>
- Van Hoeck, A., Tjoonk, N. H., van Boxtel, R., & Cuppen, E. (2019). Portrait of a cancer: mutational signature analyses for cancer diagnostics. *BMC Cancer*, 19(1), 457. <https://doi.org/10.1186/s12885-019-5677-2>
- Vanderwalde, A., Spetzler, D., Xiao, N., Gatalica, Z., & Marshall, J. (2018). Microsatellite instability status determined by next-generation sequencing and compared with PD-L1 and tumor mutational burden in 11,348 patients. *Cancer Medicine*, 7(3), 746–756. <https://doi.org/10.1002/cam4.1372>
- Vietri, M. T., D'elia, G., Caliendo, G., Resse, M., Casamassimi, A., Passariello,

- L., Albanese, L., Cioffi, M., & Molinari, A. M. (2021). Hereditary prostate cancer: Genes related, target therapy and prevention. *International Journal of Molecular Sciences*, 22(7). <https://doi.org/10.3390/ijms22073753>
- Vijayan, V., Yiu, S.-M., & Zhang, L. (2017). Improving somatic variant identification through integration of genome and exome data. *BMC Genomics*, 18(Suppl 7), 748. <https://doi.org/10.1186/s12864-017-4134-3>
- Wang, S., Li, H., Song, M., Tao, Z., Wu, T., He, Z., Zhao, X., Wu, K., & Liu, X. S. (2021). Copy number signature analysis tool and its application in prostate cancer reveals distinct mutational processes and clinical outcomes. *PLoS Genetics*, 17(5 May), 1–23. <https://doi.org/10.1371/journal.pgen.1009557>
- Zámborszky, J., Szikriszt, B., Gervai, J. Z., Pipek, O., Póti, Á., Krzystanek, M., Ribli, D., Szalai-Gindl, J. M., Csabai, I., Szallasi, Z., Swanton, C., Richardson, A. L., & Szüts, D. (2017). Loss of BRCA1 or BRCA2 markedly increases the rate of base substitution mutagenesis and has distinct effects on genomic deletions. *Oncogene*, 36(6), 746–755. <https://doi.org/10.1038/onc.2016.243>
- Zang, P. D., Chawla, N. S., Barragan-Carrillo, R., Chehraz-Raffle, A., Tripathi, A., Pal, S. K., & Dorff, T. B. (2024). Tumor Mutational Burden in Metastatic Castration-Resistant Prostate Cancer and Response to Checkpoint Inhibition. *JAMA Oncology*, 10(4), 531–532. <https://doi.org/10.1001/jamaoncol.2023.6817>
- Zheng, W., Pu, M., Li, X., Du, Z., Jin, S., Li, X., Zhou, J., & Zhang, Y. (2023). Deep learning model accurately classifies metastatic tumors from primary tumors based on mutational signatures. *Scientific Reports*, 13(1), 8752. <https://doi.org/10.1038/s41598-023-35842-w>
- Ziaraan, S., Varchulova Novakova, Z., Bohmer, D., & Danisovic, L. (2015). Biomarkers for determination prostate cancer: implication for diagnosis and prognosis. *Neoplasma*, 62(5), 683–691. https://doi.org/10.4149/neo_2015_082

8. Anexos

El código y datos empleados y carpetas con resultados obtenidos durante la realización de este trabajo se encuentran en el siguiente enlace:

[ANEXOS_TFM_BarbaraGarcia](#)