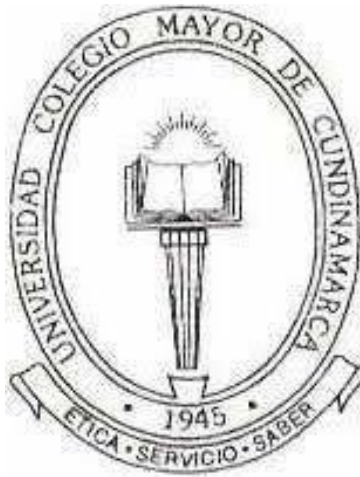




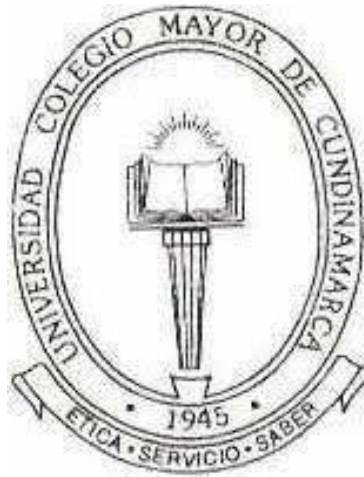
***IDENTIFICACIÓN DE POBLACIONES CELULARES ASOCIADAS A LA
SUPERVIVENCIA GLOBAL EN CARCINOMA ESCAMOCELULAR DE CABEZA Y
CUELLO MEDIANTE ANÁLISIS DE DECONVOLUCIÓN DE DATOS DE RNA-SEQ
DE COHORTES PÚBLICAS***

**UNIVERSIDAD COLEGIO MAYOR DE CUNDINAMARCA FACULTAD CIENCIAS
DE LA SALUD**

PROGRAMA DE BACTERIOLOGÍA Y LABORATORIO CLÍNICO

TRABAJO DE GRADO EXPERIMENTAL

BOGOTA- 2026



***IDENTIFICACIÓN DE POBLACIONES CELULARES ASOCIADAS A LA
SUPERVIVENCIA GLOBAL EN CARCINOMA ESCAMOCELULAR DE CABEZA Y
CUELLO MEDIANTE ANÁLISIS DE DECONVOLUCIÓN DE DATOS DE RNA-SEQ
DE COHORTES PÚBLICAS***

ESTUDIANTE

Valentina Sotomonte Ramírez

ASESORES

Jonathan Carvajal Veloza

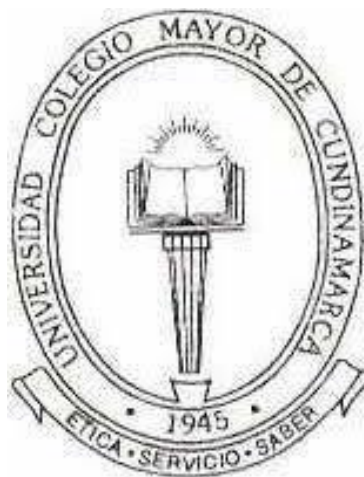
Jonathan Andre Mora Quimbayo

**UNIVERSIDAD COLEGIO MAYOR DE CUNDINAMARCA FACULTAD CIENCIAS
DE LA SALUD**

PROGRAMA DE BACTERIOLOGÍA Y LABORATORIO CLÍNICO

TRABAJO DE GRADO EXPERIMENTAL

BOGOTA- 2026



***IDENTIFICACIÓN DE POBLACIONES CELULARES ASOCIADAS A LA
SUPERVIVENCIA GLOBAL EN CARCINOMA ESCAMOCELULAR DE CABEZA Y
CUELLO MEDIANTE ANÁLISIS DE DECONVOLUCIÓN DE DATOS DE RNA-SEQ
DE COHORTES PÚBLICAS***

APROBADO: _____

JURADOS

ASESORES

**UNIVERSIDAD COLEGIO MAYOR DE CUNDINAMARCA FACULTAD CIENCIAS
DE LA SALUD**

PROGRAMA DE BACTERIOLOGÍA Y LABORATORIO CLÍNICO

TRABAJO DE GRADO EXPERIMENTAL

BOGOTA- 2026

DEDICATORIA

A mi pa' quien no está aquí para verlo pero vive en todo lo que soy.

AGRADECIMIENTOS

Me agradezco a mi misma, por el esfuerzo y la resiliencia que me trajeron hasta aquí. Esto es la prueba de todo lo que fui capaz de sostener incluso en los momentos más difíciles.

A mi mamá, por hacer lo imposible para que mis sueños sean posibles. Gracias por tu amor incondicional, tu apoyo constante y nunca dejar de creer en mí. A mí hermano, por su apoyo fundamental en este proceso.

A Zanahoria y Orejitas, por acompañarme en cada madrugada y hacer más llevaderos días (y noches) más pesados.

A la doctora Luz Dary Gutiérrez, a Rubí y al semillero de Ciencias Básicas de la Fundación Universitaria de Ciencias de la Salud, por brindarme la oportunidad de crecer académicamente y ser parte de un entorno que fortaleció mi formación tanto académica como personal.

A mis asesores, Jonathan Mora y Jonathan Carvajal. Gracias a Jonathan Mora por confiar en mí y guiarme en el desarrollo de este proyecto; y a Jonathan Carvajal por sembrar una pasión en mí con tanta paciencia y dedicación.

A Sergio, por creer en mí, en ocasiones incluso más que yo misma, y por ser un apoyo constante a lo largo de este proceso. Así como a su familia, por su apoyo, cariño y hacerme sentir acompañada durante este proceso.

A mí familia y amigos, por su compañía, apoyo y palabras de aliento, especialmente a Alix y Daniela por estar presentes en este proceso.

A la Universidad Colegio Mayor de Cundinamarca, por brindarme formación académica y humana que hizo posible este logro.

Y finalmente, a mí computador, por resistir cada jornada, cada programa y cada programación en la madrugada.

TABLA DE CONTENIDO

INTRODUCCIÓN	8
1. ANTECEDENTES	10
2. PROBLEMA DE INVESTIGACIÓN	12
3. JUSTIFICACIÓN	12
4.1. Objetivo general	14
4.2. Objetivos específicos	14
4. MARCO REFERENCIAL	14
Carcinoma escamocelular en cáncer de cabeza y cuello	14
Definición y características principales	14
Epidemiología	15
Factores de riesgo	16
Heterogeneidad y microambiente tumoral	17
Papel del sistema inmune en el HNSCC	18
Biomarcadores inmunes	19
Transcriptómica y métodos de caracterización celular	20
Transcriptómica bulk vs. secuenciación de célula única	20
The Cancer Genome Atlas (TCGA) como recurso para estudios del TME	21
Modelos algorítmicos para la predicción pronóstica	24
DISEÑO METODOLÓGICO	25
6.1 Universo	25
6.2 Población	26
6.3 Muestra	26
6.3.1 Criterios de inclusión	26
6.3.2 Criterios de exclusión	26
6.4 Hipótesis	27
6.5 Variables e indicadores	27
6.6 Técnica y procedimientos	28
6.6.1 Para dar cumplimiento el objetivo específico 1	29
6.6.2 Para dar cumplimiento el objetivo específico 2.	33
Estimación de subpoblaciones celulares mediante deconvolución transcriptómica	33
Ejecución de CIBERSORTx	33
Filtrado de resultados	34
Exploración de proporciones celulares	34

6.6.3 Para dar cumplimiento el objetivo específico 3	35
Variables incluidas en el modelo	36
Entrenamiento del modelo Random Survival Forest	36
Evaluación del modelo	37
Importancia de variables	37
Generación del risk score y estratificación de pacientes	38
Evaluación del risk score como predictor continuo	38
Análisis de supervivencia Kaplan-Meier	38
5. RESULTADOS	39
7.1 Control de calidad y Normalización de datos transcriptómicos	39
7.2 Caracterización del Microambiente tumoral mediante Deconvolución	41
7.2.1 Heterogeneidad tumoral y Caracterización fenotípica	41
7.2.3 Caracterización Transcriptómica del Microambiente de HNSCC	43
7.3 Modelo de Predicción de Supervivencia mediante Random Survival Forest (RSF)	44
7.3.1. Desempeño y Capacidad Predictiva del Modelo	44
7.3.2 Estratificación del riesgo y Curvas de Kaplan-Meier	45
7.3.3 Identificación de predictores Celulares y Clínicos de supervivencia Global en HNSCC	46
6. DISCUSIÓN	46
Referencias bibliográficas	50

RESUMEN

El carcinoma escamocelular de cabeza y cuello es una de las neoplasias más frecuentes y agresivas del tracto aerodigestivo superior, caracterizado por una alta heterogeneidad que influye en la supervivencia de los pacientes. La interacción entre células tumorales, inmunitarias y estromales define el microambiente tumoral, representando un ecosistema altamente complejo y un desafío para su caracterización, dado que la señal transcriptómica total refleja la aportación variable de cada grupo celular. En este contexto, el presente trabajo tuvo como objetivo identificar poblaciones celulares del microambiente tumoral asociadas a la supervivencia global a partir de datos transcriptómicos de una base de datos pública.

Se procesaron datos de RNA-seq de la cohorte TCGA-HNSCC (520 muestras), aplicando filtros de calidad, eliminación de genes no expresados y duplicados, y transformación logarítmica, obteniendo una matriz final de 16.766 genes. Las proporciones celulares se estimaron mediante CIBERSORTx en 381 muestras significativas ($p < 0.05$), evidenciando predominio de células T CD4 memoria (18%) y macrófagos M0 (17%), con distintos patrones de infiltración inmune entre pacientes.

Se construyó un modelo Random Survival Forest integrando proporciones celulares y variables clínicas en 381 pacientes (C-index OOB = 0.645). El puntaje de riesgo generado se asoció significativamente con la mortalidad (HR = 1.041; IC 95%: 1.034–1.048; $p < 0.0001$), con diferencias significativas en supervivencia global entre grupos. Las variables más relevantes fueron la edad, los macrófagos M0, el estadio patológico tumoral y el compromiso ganglionar.

En conclusión, los resultados permiten evidenciar la relación entre la composición del microambiente tumoral y la supervivencia en pacientes con HNSCC, destacando el papel del componente mieloide, principalmente los macrófagos M0, como posibles factores pronósticos en esta neoplasia.

LISTA DE TÉRMINOS

- Análisis de Supervivencia
- Cáncer de Cabeza y Cuello.
- Macrófagos Asociados a Tumores .
- Microambiente Tumoral.
- Transcriptoma

INTRODUCCIÓN

El carcinoma de células escamosas de cabeza y cuello (*HNSCC*) constituye el subtipo más común de las neoplasias del tracto aerodigestivo superior, incluyendo la cavidad oral, la faringe, la laringe, la cavidad nasal y las glándulas salivales (1,2). A nivel mundial representa el sexto cáncer más común y comprende el 6% de todas las neoplasias descritas, con una incidencia estimada de 650.000 casos anuales, aproximadamente y 350.000 muertes relacionadas con el (3).

El HNSCC presenta diversos factores de riesgo, como el consumo excesivo de tabaco, el alcohol, la dieta y la genética (3), aunque existen conocimientos previos de otros factores de riesgo importantes como la prevalencia de esta enfermedad en hombres, adultos mayores y personas de nivel socioeconómico bajo (2); en las últimas dos décadas la incidencia del HNSCC ha aumentado en pacientes jóvenes, especialmente en países desarrollados (3). Este incremento se ha asociado a la prevalencia de infecciones virales, como el Virus del Papiloma Humano (VPH) y el Virus de Epstein-Barr (VEB), los cuales no solo participan en la carcinogénesis, sino que modulan las interacciones entre las células tumorales y el microambiente tumoral, influyendo en la respuesta inmune y el comportamiento clínico de la enfermedad (1,4).

El desenlace clínico desfavorable suele relacionarse con la compleja y heterogénea naturaleza del microambiente tumoral (TME). Este se define como un conjunto de células tumorales, inmunes y estromales que rodean el tumor y modulan su progresión (5). Los tratamientos oncológicos usados actualmente para HNSCC continúan representando un desafío clínico significativo debido a su heterogeneidad tumoral que condiciona su progresión, respuesta terapéutica, pronóstico y la selección de terapias efectivas (6,7), pues suelen centrarse en eliminar las células tumorales, sin considerar el microambiente tumoral y las poblaciones inmunes presentes en este.

En este contexto la deconvolución transcriptómica de datos de *RNA-seq* ha sido una herramienta emergente para la estimación de poblaciones inmunes presentes en el

microambiente tumoral mediante su expresión génica, facilitando la comprensión de su composición, dinámica y su biología. No obstante, la interpretación de estos datos suele estar limitada por la complejidad, ya que los componentes del microambiente tumoral contribuyen de manera diferencial a la señal transcriptómica global (8).

Herramientas bioinformáticas como *CIBERSORTx*, basada en algoritmos aprendizaje automático (9), permiten estimar la proporción de poblaciones celulares inmunes a partir de datos transcriptómicos, proporcionando un acercamiento al microambiente tumoral en muestras de HNSCC provenientes de bases de datos públicas como *The Cancer Genome Atlas* (TCGA).

A pesar de los distintos avances en el estudio del TME, su caracterización sigue siendo limitada cuando se utilizan enfoques transcriptómicos convencionales (10). Frente a esta limitación, han surgido metodologías alternativas que permiten abordar la heterogeneidad del microambiente tumoral con mayor resolución. Sin embargo, técnicas como la secuenciación de célula única (*scRNA-seq*) y otras aproximaciones experimentales presentan limitaciones tanto en costo como a nivel complejidad, lo que limita su aplicación (11).

Los métodos de deconvolución transcriptómicos mencionados anteriormente se han consolidado como una estrategia bioinformática eficiente para la identificación de la composición celular del microambiente tumoral a partir de datos de expresión génica, facilitando el estudio en cortes amplias (12,13).

Adicionalmente, la integración de estas estimaciones celulares con variables clínicas mediante modelos de aprendizaje automático, como el Random Survival Forest (RSF), han demostrado una aproximación en la predicción de supervivencia en pacientes con cáncer (14), al identificar relaciones no lineales entre las características del microambiente tumoral y el pronóstico clínico.

En este sentido, el objetivo del presente estudio es estimar la composición celular de poblaciones del microambiente tumoral en muestras de HNSCC mediante el uso de métodos de deconvolución transcriptómica a partir de datos de *RNA-seq*, e integrar la información resultante con variables clínicas para la construcción de un modelo predictivo de supervivencia global.

1. ANTECEDENTES

En los últimos años distintos estudios han abordado la epidemiología y características clínicas del carcinoma escamocelular de cabeza y cuello desde distintas perspectivas. En *Heterogeneity of the Head and Neck Squamous Cell Carcinoma Immune Landscape and Impact on Immunotherapy* (15) se evidenció que en el HNSCC presenta una marcada heterogeneidad dentro del TME, la cual esta asociada a diferencias en la progresión tumoral y la respuesta a la inmunoterapia. Estos hallazgos se complementan con lo reportado por Farah et al. (16), donde se demostró que la heterogeneidad intratumoral se relaciona con un peor pronóstico y una mayor mortalidad en los pacientes, así como una influencia directa en su progresión, resistencia y respuesta a tratamientos convencionales (16,17).

Esta complejidad biológica se manifiesta en la elevada incidencia a nivel mundial y su relación con factores de riesgo como el consumo de tabaco, alcohol e infecciones por VPH o VEB (1). *Molecular Landscape of Head and Neck Cancer and Potential Targets for Therapy* (18) describió que las alteraciones moleculares y transcriptómicas presentes en el HNSCC contribuyen a la variabilidad clínica observada entre pacientes, reforzando la necesidad de enfoques terapéuticos más específicos. Es por ello que se ha planteado la necesidad de implementar enfoques más específicos que permitan una mejor caracterización del microambiente tumoral, así como distintas herramientas que acerquen a una estimación de supervivencia basadas en el (18).

Por otra parte, el TME se ha establecido como un componente clave en la biología del HNSCC. Se han descrito las complejas interacciones entre células tumorales, inmunes y estromales, las cuales favorecen diferentes mecanismos de evasión inmune y contribuyen a la progresión tumoral. En particular, la activación de vías inmunosupresoras, la infiltración de poblaciones inmunes como macrófagos asociados a tumores (TAMs) y Linfocitos T, junto con la regulación de moléculas como PD-L1 y TGF- β , ha sido asociada con condiciones hipóxicas e inflamatorias que promueven la heterogeneidad tumoral y limitan la eficacia de los tratamientos (19).

En relación con el análisis molecular del HNSCC, diversos estudios han empleado tanto bases de datos públicas, como proyectos como TCGA-HNSC y herramientas bioinformáticas para identificar genes diferencialmente expresados asociados a la progresión tumoral; Johnson et al. (20) describe en su estudio múltiples alteraciones moleculares y transcriptómicas implicadas en el desarrollo y progresión del HNSCC, destacando el potencial de lo análisis genómicos para la identificación de biomarcadores y posibles dianas

terapéuticas. Estos resultados se complementan con lo expuesto por Canning et al. (15) donde evidenciaron una marcada heterogeneidad inmunológica y molecular dentro del TME del HNSCC, destacando la importancia de caracterizar los perfiles transcriptómicos asociados con la respuesta terapéutica.

Mediante el análisis de múltiples conjuntos de datos transcriptómicos, se han identificado diversos genes implicados en la progresión tumoral que podría tener relevancia clínica tanto en el diagnóstico como en el pronóstico de la neoplasia (21).

Recientemente, la deconvolución transcriptómica ha surgido como una estrategia prometedora para caracterizar la composición celular del TME a partir de datos de *RNA-seq* de tejido completo. Esta aproximación computacional permite estimar la abundancia relativa de diferentes tipos celulares mediante el uso de algoritmos que comparan los perfiles de expresión génica de muestras heterogéneas con matrices de referencia de poblaciones celulares purificadas. Herramientas como CIBERSORTx, que utilizan la matriz de referencia LM22 compuesta por 22 subtipos de células inmunes, han demostrado ser efectivas para cuantificar poblaciones específicas como macrófagos, linfocitos T CD4+ y CD8+, células NK y células dendríticas en diversos tipos de cáncer, incluyendo HNSCC (19). Este enfoque computacional ofrece la ventaja de analizar retrospectivamente grandes cohortes de datos públicos, como los disponibles en TCGA, sin requerir técnicas experimentales adicionales como la secuenciación de célula única (22).

Los modelos de aprendizaje automático surgen de igual manera como herramienta importante para la predicción de supervivencia en pacientes con cáncer. El Random Survival Forest, es una extensión del algoritmo Random Forest (23) adaptada para datos de supervivencia, ha logrado demostrar una capacidad de integrar variables clínicas y moleculares de manera simultánea. Estudios recientes han aplicado enfoques bioinformáticos basados en transcriptómica y modelos de supervivencia en cohortes TCGA-HNSC. Por ejemplo, Zhang et al. (2024) (24), mediante el uso de modelos de regresión LASSO-Cox y Random Survival Forest, desarrollaron una firma pronóstica asociada a genes relacionados con el aparato de Golgi, reportando valores C-index entre 0,62 y 0,72 en distintos tipos de cáncer, incluyendo neoplasias como HNSC, lo que respalda el uso como una estrategia válida para la estratificación pronóstica de pacientes (25).

Esto resalta la necesidad de implementar enfoques que integren herramientas bioinformáticas, como la deconvolución transcriptómica, para estimar la composición celular del TME y profundizar en la comprensión de su heterogeneidad del TME en el HNSCC, con el fin de

evaluar si la composición del microambiente tumoral podría estar relacionada con la supervivencia de los pacientes con esta neoplasia.

2. PLANTEAMIENTO DEL PROBLEMA

El carcinoma escamocélular de cabeza y cuello sigue representando un importante problema de salud pública debido a su alta incidencia, agresividad biológica y limitadas tasas de supervivencia en estadios avanzados. A pesar de los avances en el conocimiento molecular de esta neoplasia y del desarrollo de nuevas estrategias terapéuticas, aún persisten dificultades para comprender completamente los mecanismos asociados con la progresión tumoral, la heterogeneidad clínica y el pronóstico de los pacientes.

Este característico microambiente tumoral limita la capacidad de diferenciar las distintas poblaciones celulares presentes en el TME, dificultado su caracterización y la comprensión del papel de la heterogeneidad tumoral en neoplasias como HNSCC (19,26).

Técnicas como *scRNA-seq* permiten analizar la heterogeneidad celular con mayor resolución, sin embargo, la implementación de este tipo de técnicas en estudios a gran escala se suele ver restringida por temas de costos y procesamientos, la necesidad de protocolos complejos para la obtención y preservación de las muestras, así como la generación de datos altamente dimensionales que requieren análisis bioinformáticos avanzados y problemas técnicos como el ruido transcriptómico, los efectos de lote y los eventos de dropout, los cuales pueden afectar la detección precisa de genes y poblaciones celulares. Adicionalmente, la disponibilidad y calidad de las muestras tumorales representan una limitación importante para su aplicación rutinaria en cohortes clínicas amplia (27,28).

En este contexto, los métodos de deconvolución transcriptómica han surgido como una alternativa bioinformática que permite estimar la proporción de diferentes conjuntos celulares a partir de datos de expresión génica derivados de muestras heterogéneas (29). Herramientas como CIBERSORTx han demostrado gran utilidad para inferir la composición celular del TME en diversos tipos de cáncer, facilitando su estudio en cohortes amplias (30).

Por lo tanto, surge la necesidad de implementar enfoques informáticos basados en la deconvolución transcriptómica que permitan la caracterización de manera más precisa de las poblaciones celulares presentes en el TME en pacientes con HNSCC, contribuyendo a una mejor comprensión de su heterogeneidad.

3. JUSTIFICACIÓN

El presente estudio se justifica en la necesidad de transformar datos transcriptómicos masivos y complejos en información con utilidad clínica para el carcinoma escamocelular de cabeza y cuello. Actualmente se reconoce la importancia del TME, sin embargo la naturaleza de los datos predominantes en cohortes como TCGA, basados en Bulk *RNA-seq*, presenta una limitación crítica: la mezcla de señales celulares que oculta la verdadera composición biológica del tumor (1,15).

La deconvolución transcriptómica es una alternativa bioinformática que permite estimar las proporciones de las poblaciones celulares a partir de datos de expresión génica derivados de muestras heterogéneas (29). Al aplicar herramientas como CIBERSORTx, es posible alcanzar una resolución celular detallada a partir de datos masivos ya existentes, lo cual otorga una mayor potencia estadística al que ofrecen técnicas como el *scRNA-seq*, cuyas limitaciones de costo y complejidad técnica restringen su uso a gran escala (15,28). Este enfoque permite identificar a las poblaciones clave como: linfocitos, fibroblastos o células supresoras, que están asociadas con la progresión de la enfermedad, optimizando recursos y aprovechando la disponibilidad de datos públicos (29,31).

A diferencia de los análisis estadísticos convencionales que asumen linealidad, el RSF captura la complejidad biológica del HNSCC, donde las variables interactúan de forma dinámica y no lineal (32). La utilidad del RSF para el estudio de la supervivencia destaca ya que permite identificar firmas celulares que actúan como biomarcadores pronósticos con una precisión superior (32,33). Al alimentar el modelo simultáneamente con la composición del TME y variables clínicas, se genera una herramienta capaz de predecir con mayor robustez el riesgo de recurrencia o mortalidad, facilitando una estratificación de pacientes (34,35).

El conjunto de ambas herramientas (deconvolución + RSF) permite hacer una descripción biológica a la predicción clínica. Por lo tanto, surge la necesidad de implementar enfoques informáticos basados en la deconvolución transcriptómica que permitan la caracterización de manera más precisa de las poblaciones celulares presentes en el TME en pacientes con HNSCC, e integrar dicha información con variables clínicas mediante modelos de predicción de supervivencia. Esto contribuye a una mejor comprensión de su heterogeneidad y comportamiento clínico, estableciendo una base sólida para la identificación de nuevos blancos terapéuticos y el diseño de modelos predictivos que impacten directamente en la personalización del tratamiento y el pronóstico de los pacientes con esta neoplasia.

OBJETIVOS

4.1. Objetivo general

Identificar poblaciones celulares asociadas a la predicción de la supervivencia global con carcinoma de célula escamosa de cabeza y cuello, mediante deconvolución de *RNA-seq* provenientes de cohortes disponibles en bases de datos públicas.

4.2 Objetivos específicos

- Compilar y procesar datos transcriptómicos (RNA-seq) de muestras tumorales de pacientes con carcinoma de células escamosas de cabeza y cuello obtenidos de bases de datos públicas, asegurando su normalización y calidad para análisis posteriores.
- Estimar proporciones de subpoblaciones celulares presentes en muestras tumorales de pacientes con carcinoma de células escamosas de cabeza y cuello mediante deconvolución transcriptómica.
- Implementar un modelo de predicción de supervivencia basado en las proporciones celulares obtenidas por deconvolución y las variables clínicas asociadas a los pacientes de la cohorte TCGA-HNSC.

4.MARCO REFERENCIAL

Carcinoma escamocelular en cáncer de cabeza y cuello

Definición y características principales

El carcinoma escamocelular de cabeza y cuello es una de las neoplasias más malignas que afecta la cavidad oral, la faringe, la laringe, la cavidad nasal y las glándulas salivales (1). A nivel mundial representa el sexto cáncer más común y comprende el 6% de todas las neoplasias, con una incidencia estimada de 650.000 casos anuales y 350.000 muertes relacionadas (19).

El HNSCC presenta diversos factores de riesgo, como el consumo excesivo de tabaco, el alcohol, la dieta y la genética (19). Adicionalmente, se ha documentado mayor prevalencia en hombres, adultos mayores y personas de nivel socioeconómico bajo (1). En las últimas dos décadas, la incidencia del HNSCC ha aumentado en pacientes jóvenes, especialmente en países desarrollados (19,20).

Uno de los principales desafíos en el manejo de esta neoplasia radica en la detección tardía y la falta de biomarcadores fiables que permitan una estratificación precisa del riesgo (36). La mayoría de los pacientes son diagnosticados en estadios locorregionalmente avanzados, lo que limita las opciones terapéuticas y aumenta el riesgo de recurrencia y metástasis (37). Esta situación subraya la necesidad de comprender con mayor profundidad los mecanismos biológicos subyacentes a su progresión y resistencia.

Epidemiología

El carcinoma escamocelular de cabeza y cuello es el sexto cáncer más común (19), representando aproximadamente 890.000 casos nuevos (el 4,5% de todos los diagnósticos de cáncer en el mundo) y 450.000 muertes por año. La incidencia incluye aproximadamente 380.000 casos de cáncer de labio y cavidad oral, 185.000 de laringe, 133.000 de nasofaringe, 98.000 de orofaringe, 84.000 de hipofaringe y 54.000 de glándulas salivales (1,38).

La tasa de incidencia y mortalidad de esta neoplasia puede presentar una variación bastante amplia según la región geográfica y las características demográficas. A nivel mundial, el HNSCC es más común en hombres que en mujeres, con una proporción hombre-mujer de 2:1, y en adultos mayores de 50 años (38). Las tasas de incidencia de HNSCC son más altas en el sur de Asia, seguidas por Europa central y oriental, y por Sudamérica (2).

En los últimos años, la incidencia global del HNSCC ha mostrado un aumento sostenido, especialmente en poblaciones jóvenes, con una proyección de incremento del 20% para el año (38). Este fenómeno se atribuye tanto a cambios en los estilos de vida, incluyendo el aumento del consumo de tabaco y alcohol, como a la mayor prevalencia del cáncer orofaríngeo asociado al virus del papiloma humano (2).

Factores de riesgo

El HNSCC presenta diversos factores de riesgo bien establecidos. A nivel mundial, el tabaco y el alcohol representan aproximadamente el 72% de los casos, con un efecto sinérgico que puede incrementar el riesgo hasta 40 veces cuando se combinan (39,40).

El consumo de tabaco, principal factor etiológico, está implicado en el 75% de los casos en Europa Occidental, principalmente debido a su contenido de carcinógenos como hidrocarburos aromáticos policíclicos y nitrosaminas que generan daño directo al ADN (40,41). Los fumadores activos tienen un riesgo 5-25 veces mayor de desarrollar HNSCC en comparación con los no fumadores(40).

El riesgo de desarrollar HNSCC aumenta de manera dependiente de la dosis, cantidad y frecuencia del consumo de alcohol. Se ha demostrado que las personas que consumen más de tres bebidas alcohólicas por día tienen mayor riesgo de desarrollo de la enfermedad (1,40).

A nivel mundial, el tabaco y el alcohol representan el 72% de los casos de HNSCC, y el 35% (casi la mitad) se atribuye a la combinación de ambos(40).

En las últimas dos décadas, la infección por el virus del papiloma humano ha surgido como un factor de riesgo crítico, especialmente en países desarrollados, donde representa el 72% de los casos orofaríngeos. Se han identificado más de 20 serotipos del virus; sin embargo, el VPH16 es responsable de aproximadamente el 80-85% de los casos orofaríngeos relacionados con este tipo de cáncer (2,42).

La oncogénesis ocurre con la infección viral de los queratinocitos basales, expuestos a micro abrasiones causadas por el contacto sexual (42). Los factores de riesgo para el cáncer orofaríngeo asociado al VPH incluyen antecedentes como múltiples parejas sexuales, sexo oral y un sistema inmunológico debilitado(40). La prevalencia de esta neoplasia asociada al VPH ha aumentado en los países desarrollados en los últimos años, particularmente en adultos jóvenes, superando incluso los casos de cáncer de cuello uterino reportados en países desarrollados como Estados Unidos (43).

El HNSCC presenta características distintivas en cuanto a su relación con el virus del papiloma humano. Mientras que los tumores VPH-negativos se caracterizan por mutaciones

en el gen supresor tumoral TP53, una disminución de p16 y alteraciones en el gen del retinoblastoma (Rb), los tumores VPH-positivos presentan un perfil molecular distintivo caracterizado por la inactivación de p53 mediada por la proteína viral E6, sobreexpresión de p16 por inactivación de Rb vía E7, y mutaciones en el gen PIK3CA (19,44).

Clínicamente, los tumores inducidos por VPH se presentan predominantemente en pacientes jóvenes que no suelen tener antecedentes de tabaquismo o consumo excesivo de alcohol, se localizan más comúnmente en las amígdalas de la orofaringe (19), lo que probablemente contribuye a mejores resultados y tasas de supervivencia en comparación con los pacientes con tumores VPH-negativos.

Por su parte, el Virus de Epstein-Barr está fuertemente asociado con el carcinoma nasofaríngeo, particularmente en regiones endémicas del sur de Asia, el Mediterráneo y el norte de África, especialmente en la forma no queratinizante-indiferenciada, lo que lo convierte en un factor etiológico central para este subtipo de tumor (20). A nivel molecular, la oncogénesis mediada por VEB se explica mediante la expresión de genes virales como EBNA1, LMP1 y LMP2A, así como pequeños microARN virales (EBERs, miRNA BART) que reprograman la célula (44,45). LMP1 actúa como un activador constitutivo de vías proliferativas e inflamatorias como NF- κ B y MAPK, promoviendo la supervivencia celular y resistencia a la apoptosis, mientras que LMP2A contribuye a señales de supervivencia y mantenimiento de la latencia viral (45,46). Los tumores asociados al VEB presentan tasas más bajas de mutaciones en genes supresores de tumores, se asocian con menor consumo de tabaco y alcohol, y son más sensibles a radiación y quimioterapia (43).

La inmunodeficiencia resultante de la infección por VIH o un trasplante de órgano sólido se ha relacionado con una mayor susceptibilidad al HNSCC(40). Finalmente, la exposición a carcinógenos ocupacionales o ambientales, una higiene bucal deficiente y una predisposición genética pueden representar factores de riesgo en menor medida (40,47).

La heterogeneidad etiológica del HNSCC se refleja en distintos perfiles moleculares, clínicos e inmunológicos que influyen directamente en el microambiente tumoral, la respuesta terapéutica y el pronóstico de los pacientes. A pesar de los avances en la comprensión de estos factores, este tipo de cáncer continúa diagnosticándose en etapas avanzadas, lo que se traduce en una supervivencia global inferior al 50% a los 5 años (48). Los desafíos actuales se centran en el diagnóstico temprano, la identificación de biomarcadores predictivos y el control de los factores de riesgo modificables, con el fin de reducir la recurrencia y la resistencia terapéutica en pacientes con HNSCC (49).

Heterogeneidad y microambiente tumoral

El microambiente tumoral es un ecosistema altamente complejo que se compone tanto de células no cancerosas como de matriz extracelular que rodea el tumor (50).

Las células no cancerosas que componen el microambiente tumoral incluyen células inmunes como macrófagos asociados a tumores, células supresoras derivadas de mieloides (MDSCs), células T reguladoras (Tregs), células dendríticas (DCs) y células Natural Killer (NK) (*The role of the tumor microenvironment in HNSCC resistance and targeted therapy*), así como células no inmunes tales como fibroblastos asociados a cáncer (CAFs) y células endoteliales (51). Los componentes no celulares del microambiente tumoral suelen contener colágeno, laminina, fibronectina y elastina, entre otros (50). Estas poblaciones y componentes celulares se infiltran en el tumor e interactúan con las células tumorales y entre sí, generando redes complejas de interacción.

El HNSCC logra evadir el reconocimiento y la eliminación por parte del sistema inmunitario, activa la inmunosupresión y facilita la progresión tumoral (50), proporcionando un entorno favorable para la progresión tumoral, el desarrollo de resistencia y la metástasis.

Papel del sistema inmune en el HNSCC

El complejo microambiente tumoral del HNSCC abarca componentes celulares y moleculares diversos, los cuales se interrelacionan directamente con el pronóstico de los pacientes que padecen esta neoplasia. Entre estos componentes, el sistema inmunitario desempeña un papel indispensable desde el inicio de la progresión tumoral. Se ha descrito que la inmunosupresión es un factor de riesgo importante en el desarrollo de esta enfermedad. La insuficiencia de la respuesta inmunitaria contra las células cancerosas se puede atribuir a la formación de células inmunitarias inhibitorias y a la secreción de citocinas y mediadores represores, lo que conlleva a la evasión inmune (52).

Las células inmunitarias como las **Tregs**, los **TAMs** y las **MDSCs** favorecen la progresión del cáncer y el escape inmunitario (52,53). Las Tregs son un subtipo de células **T CD4** que previenen enfermedades autoinmunitarias; sin embargo, estas se subdividen en dos grupos: las derivadas del timo (Tregs) y las derivadas de la periferia (pTregs) (52). Estas células se ensamblan en el microambiente tumoral como respuesta a las quimiocinas liberadas por las células tumorales y los macrófagos, actuando como supresoras de la respuesta inmunitaria antitumoral (54). Las Tregs inactivan directamente las células T efectoras mediante la

liberación de citocinas inmunosupresoras como IL-10, IL-35 y TGF- β , debilitando la secreción de IL-2, la cual es esencial para la supervivencia de las células T efectoras (52).

La proporción de los diferentes subgrupos de células T es clave para estimar la aparición y progresión tumoral (55). Las proporciones de CD8,CD4 y CD8 FOXP3 suelen ser las medidas más utilizadas para evaluar la potencia de la actividad inmunitaria (52). Específicamente, estas proporciones sólo se consideran supresoras de tumores si las células T CD8+ aumentan en comparación con las células T FOXP3+ y/o CD4+ después del tratamiento con radioterapia o inhibidores de puntos de control inmunitarios (52,55). La elevada proporción de CD8+/FOXP3+ indica que las células tumorales se han escapado del control inmunitario, teniendo en cuenta que suprimir otras células inmunitarias es una de las funciones más relevantes de las células Tregs (53) .

Además, se ha descubierto que las células Tregs FOXP3+ intratumorales de pacientes con HNSCC generan un microambiente tumoral inmunosupresor mediante la regulación positiva de los receptores de puntos de control inmunitarios (56).

Las células supresoras derivadas de mieloides (MDSCs) son un grupo de diversas células que nacen a partir de la médula ósea y pasan por diversos procesos de diferenciación hasta que se convierten en macrófagos, neutrófilos o células dendríticas. No obstante, estas células se producen y activan en condiciones patológicas como el cáncer (52). Las MDSCs restringen la actividad antitumoral de los linfocitos T citotóxicos, además de reducir la acción de las células proinflamatorias como las NK y las DCs.

Los macrófagos asociados al tumor son macrófagos que se infiltran en el microambiente tumoral. Estas células desempeñan un papel importante en la carcinogénesis oral al influir en el crecimiento, la progresión y la metástasis tumoral mediante mecanismos de inmunosupresión (52). Existen dos fenotipos principales de TAMs: el fenotipo M1, que corresponde a la activación clásica con actividad antitumoral, y el fenotipo M2, de activación alterna, que posee actividad promotora de tumores (57).

Biomarcadores inmunes

El desarrollo y progresión del carcinoma escamocelular de cabeza y cuello no dependen únicamente de alteraciones intrínsecas de las células tumorales, sino también de la compleja interacción con el microambiente tumoral , el cual desempeña un papel fundamental en la invasión, metástasis y evasión inmunológica de la neoplasia (35). En este contexto, la célula tumoral promueve la generación de un entorno inmunosupresor mediante la disfunción de

células inmunitarias, alteración de citocinas inflamatorias e inducción de puntos de control inmunitario (58). Entre los principales biomarcadores inmunes implicados en este proceso destacan PD-L1 y CTLA-4, moléculas reguladoras que inhiben la activación y función efectora de los linfocitos T. PD-L1, expresado tanto en células tumorales como en células inmunes del TME, interactúa con el receptor PD-1 presente en linfocitos T, reduciendo su proliferación, secreción de citocinas y capacidad citotóxica, favoreciendo así la apoptosis de células efectoras y la expansión de linfocitos T reguladores (Tregs) (58,59). De manera similar, CTLA-4 limita la activación de los linfocitos T al competir con CD28 por los ligandos CD80/CD86 en células presentadoras de antígeno, promoviendo un microambiente inmunosupresor asociado con menor infiltración de células T efectoras (60). Por otro lado, los linfocitos T CD8 constituyen uno de los principales mecanismos de defensa antitumoral mediante la liberación de perforinas y granzimas; sin embargo, su actividad puede verse limitada por fenómenos de agotamiento inmunológico mediados por puntos de control inmunitarios como PD-1 y TIM-3 (61). En contraste, las células T reguladoras caracterizadas por la expresión del factor de transcripción FOXP3 favorecen la inmunosupresión tumoral mediante la secreción de IL-10, IL-35 y TGF- β , además de inhibir la proliferación de linfocitos T efectores, asociándose con resistencia terapéutica y peor pronóstico clínico en HNSCC (60). Finalmente, el interferón gamma (IFN- γ), producido principalmente por linfocitos T CD8 y células NK, participa en la activación de respuestas inmunes antitumorales promoviendo la presentación antigénica y la polarización de macrófagos hacia fenotipos M1 proinflamatorios; niveles elevados de esta citocina se han relacionado con microambientes inmunológicamente activos y mejores respuestas terapéuticas (62).

Transcriptómica y métodos de caracterización celular

Transcriptómica bulk vs. secuenciación de célula única

La transcriptómica bulk ha representado uno de los enfoques más ampliamente utilizados para el estudio de tumores sólidos, permitiendo cuantificar los niveles promedio de expresión génica provenientes de una mezcla de células en una misma muestra. Esta aproximación ofrece una visión global del transcriptoma tumoral, reflejando la actividad promedio de todos los tipos celulares presentes, aunque sin discriminar las contribuciones específicas de cada uno. Entre sus principales ventajas se destacan su robustez técnica y su bajo costo relativo, lo que la convierte en una herramienta idónea para estudios con grandes cohortes de pacientes.

Gracias a ello, existen repositorios públicos como *The Cancer Genome Atlas*, que proporcionan datos de expresión asociados a metadatos clínicos y de supervivencia, facilitando análisis correlacionales de gran alcance (43).

En contraste, la secuenciación de ARN de célula única permite cuantificar la expresión génica en cada célula individual, lo que permite identificar subtipos celulares, estados funcionales transitorios y poblaciones raras que pueden pasar inadvertidas en análisis masivos. De igual manera, permite la reconstrucción de redes de interacción célula a célula dentro del microambiente tumoral, proporcionando una comprensión más precisa de los mecanismos inmunológicos y de resistencia terapéutica (59,61).

Existen diversas ventajas tanto estadísticas como de rangos de referencia al momento de usar scRNA-seq; sin embargo, su elevado costo, la exigencia de un procesamiento y conservación de tejido específico, pueden permitir que ocurra "dropout" (genes con lecturas ausentes en muchas células), y presentar problemas de batch effect y de representatividad cuando las cohortes clínicas son grandes (59).

Por otro lado, bulk RNA-seq ofrece potencia estadística y acceso a grandes muestras, pero promedia señales y oculta subpoblaciones (60).

Muchos estudios recientes han usado ambos enfoques de manera complementaria (scRNA-seq para derivar firmas y definir poblaciones, y bulk RNA-seq para analizar grandes cohortes y asociarlas con desenlaces clínicos), demostrando que la integración aporta mayor comprensión biológica y clínico-pronóstica (43,64)

The Cancer Genome Atlas (TCGA) como recurso para estudios del TME

The Cancer Genome Atlas es uno de los repositorios públicos más importantes en oncología, que proporciona datos genómicos, transcriptómicos y clínicos de más de 33 tipos de cáncer, incluyendo el HNSCC (65). El proyecto TCGA-HNSC comprende datos de expresión génica (RNA-seq) de aproximadamente 500 muestras tumorales primarias, acompañadas de información clínica detallada que incluye estadio tumoral, estado de VPH, tratamiento recibido y datos de supervivencia (66).

La disponibilidad de estos datos de acceso abierto ha facilitado el desarrollo de estudios retrospectivos de gran escala que permiten identificar firmas moleculares asociadas al pronóstico, validar biomarcadores y caracterizar la heterogeneidad del TME sin requerir nuevas cohortes experimentales (67). En el contexto del HNSCC, múltiples estudios han utilizado datos de TCGA para estratificar pacientes según perfiles de infiltración inmune,

correlacionar la expresión de genes inmunológicos con la supervivencia global y evaluar la relación entre la composición del TME y la respuesta a terapias (68).

Principio de la deconvolución transcriptómica

La deconvolución transcriptómica es una estrategia computacional que busca descomponer la señal global de un bulk RNA-seq en las contribuciones relativas de distintos tipos celulares. Su funcionamiento se basa en el siguiente principio: si se conoce o se estima un conjunto de firmas genéticas representativas de cada tipo celular (es decir, genes que se expresan diferencialmente en esa población), se puede modelar la expresión observada en la mezcla como una combinación lineal ponderada de esas firmas y, a partir de allí, estimar las proporciones relativas de cada tipo celular (59).

Elementos clave del enfoque:

- Firmas de referencia: Matrices como LM22 (utilizada por CIBERSORTx) que contiene 22 subtipos inmunes, o firmas derivadas de scRNA-seq de referencia. Estudios recientes en HNSCC han construido matrices específicas a partir de scRNA-seq y las han usado para refinar estimaciones en cohortes bulk (9,59,64).
- Modelado estadístico: Los métodos varían e incluyen regresión (support vector regression en CIBERSOFT), métodos basados en enriquecimiento de firmas (xCell), optimización de mezcla no negativa, o enfoques bayesianos, cada uno con supuestos distintos respecto a ruido, colinealidad entre firmas y presencia de tipos no modelados (9).
- Salida: Proporciones relativas (valores entre 0 y 1 que suman 1) o scores de enriquecimiento. Algunas implementaciones como CIBERSORTx permiten además imputar perfiles de expresión específicos por tipo celular dentro de la muestra (9).

La deconvolución transcriptómica representa una herramienta poderosa para caracterizar el infiltrado inmune a partir de datos bulk RNA-seq, aprovechando el amplio repositorio de bases públicas como TCGA. Su principal fortaleza radica en la capacidad de estimar proporciones celulares y correlacionarlas con variables clínicas y de supervivencia a gran escala; además, cuando se emplean firmas derivadas de scRNA-seq del mismo tipo tumoral, como HNSCC, la precisión del modelo aumenta notablemente (64,69).

Sin embargo, su desempeño depende de la calidad de la matriz de referencia y puede verse afectado por la colinealidad entre firmas o la subrepresentación de poblaciones raras (20). Por ello, es necesario normalizar adecuadamente los datos, aplicar corrección por lote y validar los resultados con marcadores específicos. Este enfoque metodológico resulta especialmente

pertinente para el presente estudio, ya que permite identificar poblaciones inmunes asociadas al pronóstico clínico en HNSCC mediante la integración de cohortes transcriptómicas amplias y su validación con análisis de supervivencia, fortaleciendo la robustez y la interpretación biológica de los hallazgos (69)

Aplicaciones clínicas y perspectivas

La caracterización del microambiente inmune en el carcinoma escamocelular de cabeza y cuello ha cobrado especial relevancia en los últimos años debido a su impacto directo en el pronóstico y la respuesta terapéutica de los pacientes. El estándar de atención para HNSCC ha evolucionado favorablemente con la incorporación de inhibidores de puntos de control inmunitario. Los fármacos anti-PD-1/PD-L1, como pembrolizumab y nivolumab, son los primeros que han demostrado beneficios pronósticos significativos para el tratamiento de esta neoplasia, tanto en enfermedad recurrente como metastásica (20,69)

Sin embargo, a pesar de estos avances, la tasa de respuesta global a la inmunoterapia en HNSCC sigue siendo limitada, oscilando entre el 13% y el 18% en monoterapia (19). Esta baja tasa de respuesta subraya la necesidad crítica de identificar biomarcadores predictivos que permitan seleccionar a los pacientes que se beneficiarán de estas terapias. En este contexto, la expresión de PD-L1 mediante inmunohistoquímica ha sido el biomarcador más estudiado; no obstante, su valor predictivo es limitado debido a la heterogeneidad de expresión intratumoral y a la falta de estandarización en los umbrales de positividad.

La composición del microambiente inmune ha emergido como un factor pronóstico independiente en HNSCC. Estudios han demostrado que una alta infiltración de linfocitos T CD8 se asocia con mejor supervivencia global, mientras que la presencia predominante de macrófagos M2 y células T reguladoras (Tregs) se correlaciona con peor pronóstico y resistencia terapéutica. Específicamente, la proporción CD8/FOXP3 ha mostrado ser un indicador robusto del balance entre inmunidad antitumoral e inmunosupresión, con potencial para predecir la respuesta a inhibidores de puntos de control inmunitario.

La identificación de subpoblaciones inmunes específicas mediante deconvolución transcriptómica ofrece una ventaja significativa para comprender los mecanismos de resistencia y progresión tumoral. Este enfoque permite no solo cuantificar la infiltración de diferentes tipos celulares, sino también correlacionar estos perfiles con desenlaces clínicos como la supervivencia libre de enfermedad, la supervivencia global y la respuesta a tratamientos específicos. Además, facilita la estratificación de pacientes en subgrupos de riesgo, lo que podría guiar decisiones terapéuticas más informadas.

El futuro de la inmunoterapia en HNSCC se orienta hacia estrategias combinadas que integren inhibidores de puntos de control con otros agentes que modulen el microambiente tumoral. Por ejemplo, terapias dirigidas a reducir la actividad inmunosupresora de MDSCs y TAMs M2, o estrategias que potencien la actividad de células T efectoras, podrían mejorar significativamente las tasas de respuesta. En este sentido, la caracterización molecular y celular del TME mediante deconvolución transcriptómica no solo contribuye a la identificación de biomarcadores pronósticos, sino que también abre la puerta al desarrollo de terapias dirigidas y personalizadas basadas en el perfil inmunológico específico de cada paciente.

Asimismo, desde una perspectiva traslacional, los hallazgos derivados de análisis de deconvolución en grandes cohortes como TCGA pueden ser validados experimentalmente mediante técnicas como citometría de flujo multiparamétrica o secuenciación de célula única, generando así un ciclo de retroalimentación que fortalece tanto el conocimiento básico como la aplicación clínica. La integración de datos genómicos, transcriptómicos y clínicos representa un pilar fundamental para avanzar hacia la medicina de precisión en oncología.

Modelos algorítmicos para la predicción pronóstica

La estimación de la composición celular del TME mediante técnicas de deconvolución transcriptómica ha permitido integrar variables biológicas en modelos de Machine Learning cada vez más precisos para la estratificación del riesgo y la predicción de supervivencia global (OS) en pacientes con HNSCC (33). La combinación de perfiles inmunológicos con datos genómicos y clínicos ha demostrado mejorar la capacidad predictiva de los modelos pronósticos frente a los métodos clínicos convencionales. Se ha evidenciado que la integración de datos transcriptómicos, mutacionales y de composición celular inferida mediante deconvolución resulta fundamental para refinar la clasificación molecular del HNSCC y caracterizar subtipos tumorales con implicaciones pronósticas y terapéuticas (70)

Dentro de estos enfoques, algoritmos como Random Forest, han mostrado un desempeño superior en la identificación de patrones complejos asociados con progresión tumoral, riesgo metastásico y respuesta a inmunoterapia (70,71). Asimismo, el uso de métodos de selección de variables como LASSO Cox Regression ha permitido identificar firmas génicas reducidas asociadas con perfiles de baja infiltración inmune y mayor riesgo de escape inmunológico, los cuales se correlacionan con peor supervivencia global en HNSCC (71) .

En particular, el algoritmo Random Survival Forest ha emergido como una de las aproximaciones de ML más robustas para el análisis de supervivencia en oncología, debido a que constituye una extensión del algoritmo Random Forest diseñada específicamente para datos censurados (23). A diferencia de la regresión de Cox convencional, el RSF no requiere asumir proporcionalidad de riesgos ni relaciones lineales entre las variables predictoras y el desenlace, permitiendo modelar interacciones complejas entre variables clínicas, transcriptómicas y componentes del TME. Adicionalmente, el RSF permite estimar la relevancia de cada predictor mediante el índice de importancia de variables (*Variable Importance*, VIMP), el cual cuantifica la contribución individual de cada variable al desempeño global del modelo. Valores positivos de VIMP indican variables con aporte pronóstico relevante, mientras que valores cercanos a cero o negativos sugieren baja utilidad predictiva.

El desempeño de los modelos de supervivencia suele evaluarse mediante el índice de concordancia de Harrell (C-index), métrica que mide la capacidad discriminativa del modelo para ordenar correctamente a los pacientes según su riesgo de mortalidad. Sus valores oscilan entre 0.5, equivalente a una predicción aleatoria, y 1.0, correspondiente a una discriminación perfecta; en modelos oncológicos basados en transcriptómica, valores superiores a 0.60 suelen considerarse clínicamente relevantes (32). Además, la evaluación pronóstica puede incluir modelos de regresión de Cox para estimar el *Hazard Ratio* (HR), curvas de supervivencia de Kaplan-Meier y pruebas de log-rank para comparar diferencias de supervivencia entre grupos de riesgo. Finalmente, los modelos RSF utilizan predicciones *out-of-bag* (OOB), generadas a partir de muestras no utilizadas en el entrenamiento de cada árbol, lo que permite estimar internamente el desempeño predictivo del modelo sin requerir un conjunto de validación independiente.

DISEÑO METODOLÓGICO

6.1 Universo

Pacientes diagnosticados con carcinoma escamocelular de cabeza y cuello incluidos en el proyecto *The Cancer Genome Atlas*, con datos disponibles de *RNA-seq* y variables clínicas asociadas.

6.2 Población

Todos los pacientes con diagnóstico confirmado de carcinoma de célula escamosa de cabeza y cuello cuyos datos de expresión génica y metadatos clínicos han sido dispuestos en bases de datos públicas.

- Fuentes de datos: La información se obtuvo directamente de la cohorte de *The Cancer Genome Atlas*, utilizando la interfaz de *TCGA Biolinks* para el acceso, descarga y estandarización de los datos.

6.3 Muestra

La muestra del estudio está conformada exclusivamente por los perfiles transcriptómicos de muestras tumorales primarias que cumplan con los criterios de inclusión.

6.3.1 Criterios de inclusión

- Se incluyeron muestras de RNA-seq bulk provenientes de cohortes públicas que presentaran datos completos de expresión génica y disponibilidad de las variables clínicas consideradas obligatorias para los análisis pronósticos y de riesgo: edad, sexo, estadio clínico y estado vital. Asimismo, se recopilaron variables clínicas complementarias, incluyendo sitio anatómico, origen de la muestra, antecedente de tabaquismo, consumo de alcohol, estado de infección por VPH y diagnóstico primario. Las muestras que presentaron datos faltantes en cualquiera de las variables clínicas obligatorias fueron excluidas del análisis, mientras que las variables complementarias con información incompleta se mantuvieron únicamente para análisis descriptivos o exploratorios.

6.3.2 Criterios de exclusión

- Muestras no tumorales
- Muestras de líneas celulares
- Perfiles de baja calidad que no superen los filtros iniciales
- Muestras de metadatos incompletas

6.4 Hipótesis

El análisis de datos transcriptómicos mediante métodos de deconvolución permite estimar la proporción de poblaciones celulares del microambiente tumoral en pacientes con carcinoma

escamocelular de cabeza y cuello a partir de muestras disponibles en TCGA. La integración de estas proporciones celulares con variables clínicas mediante un modelo Random Survival Forest permite identificar poblaciones del microambiente tumoral asociadas con la supervivencia global.

6.5 Variables e indicadores

Tabla 1. Operacionalización de las variables clínicas, sociodemográficas, inmunológicas y pronósticas incluidas en el estudio. Presenta la operacionalización de las variables utilizadas para la caracterización clínica y transcriptómica de la cohorte TCGA-HNSC, incluyendo variables sociodemográficas, clínicas, inmunológicas y de supervivencia. Se especifica el tipo de variable, su definición operacional, los indicadores empleados para su medición y la fuente de obtención de los datos.

Variable	Tipo de variable	Definición operacional	Indicador	Fuente
Edad	Cuantitativa continua	Edad del paciente al momento del diagnóstico	Años cumplidos	TCGA-HNSC
Sexo	Cualitativa nominal	Sexo biológico reportado en la cohorte	Masculino/Femenino	TCGA-HNSC
Estado vital	Cualitativa binaria	Estado de supervivencia al último seguimiento	Vivo/Fallecido	TCGA-HNSC
Etnia	Cualitativa nominal	Grupo étnico reportado en los registros clínicos	Caucásico, afrodescendiente, asiático, otros	TCGA-HNSC
Tabaquismo	Cualitativa nominal	Antecedente de consumo de tabaco	No fumador/Exfumador/Fumador activo	TCGA-HNSC
Consumo de alcohol	Cualitativa nominal	Antecedente de consumo de alcohol	Sí/No/Ocasional	TCGA-HNSC
Estado VPH	Cualitativa binaria	Presencia de infección por VPH	Positivo/Negativo	TCGA-HNSC
Sitio anatómico tumoral	Cualitativa nominal	Localización anatómica primaria del tumor	Cavidad oral, laringe, orofaringe, hipofaringe, entre otros	TCGA-HNSC
Estadio tumoral (TNM)	Cualitativa ordinal	Clasificación clínica del	I, II, III, IV	TCGA-HNSC

		tumor según AJCC		
Supervivencia global	Cuantitativa continua	Tiempo transcurrido desde el diagnóstico hasta el evento o último seguimiento	Meses	TCGA-HNSC
Estado del evento	Cualitativa binaria	Ocurrencia del evento de interés para análisis de supervivencia	0 = censurado vivo; 1 = fallecido	TCGA-HNSC
Poblaciones celulares del TME	Cuantitativa continua	Proporción relativa de subpoblaciones inmunes estimadas mediante deconvolución transcriptómica	Porcentaje (%) de infiltración celular	CIBERSORTx / TCGA-HNSC
Risk score RSF	Cuantitativa continua	Puntaje de riesgo generado por el modelo Random Survival Forest	Valor continuo	Modelo RSF
Grupo de riesgo	Cualitativa binaria	Estratificación pronóstica derivada del risk score	Alto riesgo/Bajo riesgo	Modelo RSF
Edad	Cuantitativa continua	Edad del paciente al momento del diagnóstico	Años cumplidos	TCGA-HNSC

6.6 Técnica y procedimientos

El presente estudio corresponde a una investigación observacional analítico retrospectivo. Se desarrolló bajo un diseño bioinformático de análisis secundario de datos, basado en la recopilación, procesamiento y análisis de información transcriptómica de una cohorte disponible en la base de datos pública *The Cancer Genome Atlas* (Ilustración 1).

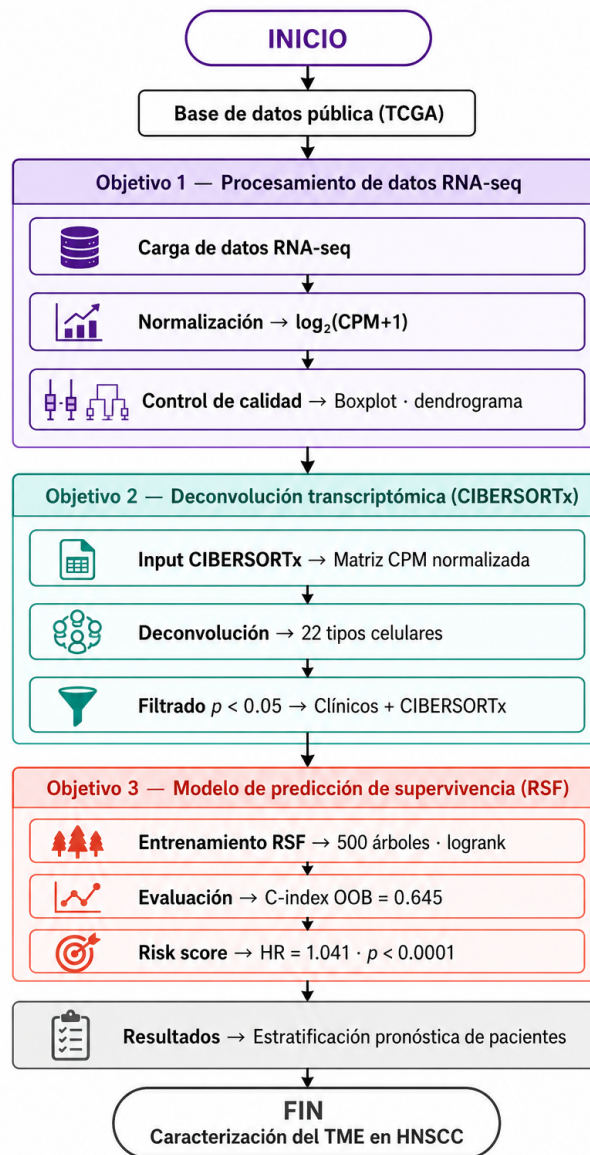


Ilustración 1. Diagrama de flujo del diseño bioinformático y estadístico aplicado en el estudio. El proceso incluyó la obtención de datos bulk RNA-seq de cohortes públicas, su normalización en RStudio, la deconvolución inmunocelular mediante CIBERSORTx

6.6.1 Para dar cumplimiento el objetivo específico 1

Obtención y procesamiento de datos

Se descargaron los datos de *RNA-seq* (tipo bulk) de pacientes con diagnóstico confirmado de HNSCC usando el paquete de *TCGA biolinks* en R. Los datos de expresión génica fueron sometidos a un proceso de filtrado y normalización, eliminando los genes con niveles de expresión o varianza insuficientes, con el propósito de maximizar la calidad de la señal transcriptómica y la precisión del análisis de deconvolución. Previo al análisis de deconvolución, los datos crudos de *RNA-seq* fueron sometidos a controles de calidad con el

propósito de garantizar la comparabilidad, reproducibilidad y fiabilidad de los resultados obtenidos (*Ilustración 2*).

Filtrados de expresión y control de calidad

La etapa inicial del procesamiento transcriptómico se orientó a la eliminación de ruido de fondo, señales espurias y mediciones de baja confiabilidad, con el propósito de mejorar la estabilidad estadística del análisis y aumentar la precisión de la deconvolución celular. Todos los procedimientos bioinformáticos fueron realizados en el lenguaje de programación R (versión 4.5.1), empleando principalmente los paquetes TCGAbiolinks, edgeR y sva

- Genes sin nombre válido:

Se eliminaron aquellos con valores faltantes (*NA*) o sin anotación en la columna *Gene Symbol*, asegurando la correcta identificación génica dentro del modelo de deconvolución.

- Agrupamiento de genes duplicados:

En los casos donde varios registros compartían el mismo *Gene Symbol*, se calculó la media de sus valores de expresión para consolidar una única entrada representativa por gen.

- Filtro de baja expresión:

Se eliminaron los genes con expresión mínima, específicamente aquellos con valores de TPM < 1 en más del 50 % de las muestras, o con CPM < 1 en al menos el 80 % de ellas, ya que su inclusión podría distorsionar las proporciones estimadas de las subpoblaciones celulares.

- Filtro de baja varianza:

Se calcularon las varianzas de expresión génica y se eliminaron los genes ubicados en el 25 % inferior de varianza, conservando únicamente los genes con variabilidad biológicamente significativa entre las muestras. (Se excluyeron los genes housekeeping, cuya varianza naturalmente baja podría sesgar el análisis).

Normalización de datos

Los datos crudos de conteo obtenidos mediante RNA-seq fueron procesados en R ((versión 4.5.1)) utilizando el paquete edgeR. Inicialmente, se calcularon los factores de normalización mediante el método TMM, el cual corrige diferencias de escala entre bibliotecas asociadas a profundidad de secuenciación y composición transcriptómica de las muestras. Posteriormente, los conteos fueron transformados a unidades CPM ajustadas por los factores TMM, obteniendo una matriz de expresión normalizada comparable entre muestras.

La matriz final utilizada como entrada para el análisis de deconvolución en CIBERSORTx correspondió a una matriz de valores CPM normalizados por TMM, sin transformación logarítmica previa. Esta decisión metodológica se fundamentó en las recomendaciones de la plataforma, las cuales indican que los algoritmos de deconvolución basados en mezclas lineales requieren datos en escala lineal para preservar las relaciones relativas de abundancia transcriptómica entre genes y muestras.

La transformación logarítmica mediante $\log_2(\text{CPM} + 1)$ fue aplicada únicamente en análisis estadísticos posteriores y visualizaciones exploratorias, pero no en la matriz destinada al proceso de deconvolución celular.

Corrección de defectos por lote

Con el propósito de reducir variaciones técnicas no biológicas derivadas de diferencias entre centros de secuenciación, procesamiento experimental o condiciones de adquisición de datos, se realizó una corrección del efecto de lote (batch effect correction) utilizando el algoritmo ComBat implementado en el paquete sva en R ((versión 4.5.1)).

La variable de lote fue definida a partir del centro de procesamiento y secuenciación reportado en los metadatos de la cohorte TCGA-HNSC. Este ajuste fue realizado sobre la matriz de expresión CPM normalizada por TMM antes del análisis de deconvolución, con el fin de minimizar sesgos técnicos que pudieran interferir en la estimación de proporciones celulares.

La aplicación del método ComBat permitió preservar la variabilidad biológica de interés mientras reducía diferencias sistemáticas entre lotes experimentales, mejorando la comparabilidad y reproducibilidad de los datos transcriptómicos.

Este método es ampliamente reconocido por eliminar variaciones no biológicas entre conjuntos de datos, garantizando consistencia y comparabilidad antes del análisis de deconvolución (72)

Transformación logarítmica

Finalmente, los valores de TMM corregidos por batch effect fueron transformados mediante la función $\log_2(\text{CPM} + 1)$, lo que permitió estabilizar la varianza y aproximar la distribución a la normalidad. Esta transformación es estándar en análisis de expresión génica y asegurara el cumplimiento de los supuestos estadísticos necesarios para algoritmos de deconvolución y modelos posteriores.

De manera complementaria al análisis transcriptómico, se realizó la curación y exploración de los datos clínicos asociados a los pacientes de la cohorte *TCGA-HNSC*.

La información clínica fue descargada desde el portal GDC Data Portal utilizando el paquete TCGAbiolinks en R, asegurando la correcta vinculación entre las muestras transcriptomas y sus metadatos clínicos.

Filtrados de datos clínicos

Posteriormente, se efectuó una depuración sistemática de las variables clínicas para garantizar su integridad y coherencia. Se eliminaron los registros con información incompleta en campos esenciales como tiempo de supervivencia y estado vital debido a que su ausencia compromete los análisis de supervivencia y correlación con las proporciones celulares.

Las variables retenidas incluyen (Tabla 2):

- Edad al diagnóstico
- Sexo
- Estado vital (vivo/fallecido)
- Tiempo de supervivencia (días/meses)
- Grado histológico
- Estadio tumoral (TNM)
- Subsitio anatómico primario
- Consumo de tabaco y alcohol
- Estado de infección por VPH
- Normal y tumor

Estas variables fueron analizadas de manera descriptiva para caracterizar la cohorte y posteriormente integradas con los resultados de la deconvolución, con el fin de identificar asociaciones entre las poblaciones inmunes inferidas y los desenlaces clínicos, tales como la supervivencia global y el pronóstico.

Finalmente, se excluyeron a los pacientes con datos de expresión o supervivencia incompletos, conformando una cohorte final depurada que garantiza la validez, reproducibilidad y solidez analítica del estudio.

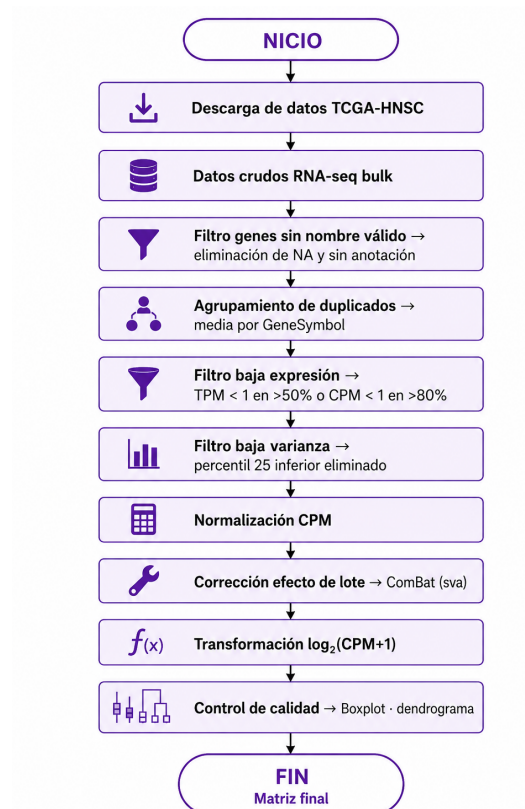


Ilustración 2. Diagrama de flujo del preprocesamiento bioinformático de los datos bulk RNA-seq de TCGA-HNSC. El proceso incluyó filtrado y depuración de genes, normalización CPM, corrección del efecto de lote con ComBat, transformación $\log_2(\text{CPM}+1)$ y control de calidad

6.6.2 Para dar cumplimiento el objetivo específico 2.

Estimación de subpoblaciones celulares mediante deconvolución transcriptómica

A partir de la matriz de expresión génica procesada en el objetivo anterior, se estimaron las proporciones de poblaciones celulares presentes en el microambiente tumoral de los pacientes con HNSCC mediante el algoritmo de deconvolución CIBERSORTx (9) ejecutado a través de su interfaz oficial.

Ejecución de CIBERSORTx

La matriz de expresión normalizada en formato $\log_2(\text{CPM}+1)$, obtenida tras el procesamiento descrito en el objetivo 1, fue utilizada como input para el algoritmo CIBERSORTx. Se empleó la firma de referencia LM22, ampliamente validada en la literatura, que permite la estimación simultánea de 22 tipos celulares inmunes, incluyendo las siguientes

subpoblaciones: linfocitos B naive, linfocitos B memoria, células plasmáticas, linfocitos T CD8, linfocitos T CD4 memoria en reposo, linfocitos T CD4 memoria activados, células T helper, células T reguladoras (Tregs), células T gamma delta, células NK en reposo, células NK activadas, monocitos, macrófagos M0, macrófagos M1, macrófagos M2, células dendríticas en reposo, células dendríticas activadas, mastocitos en reposo, mastocitos activados, eosinófilos y neutrófilos (9). El análisis fue ejecutado en modalidad de deconvolución absoluta (*absolute mode*), utilizando 1000 permutaciones estadísticas para aumentar la robustez de las estimaciones celulares obtenidas (Ilustración 2).

Filtrado de resultados

Con el fin de garantizar la confiabilidad estadística de las estimaciones obtenidas, se aplicó un criterio de filtrado basado en el p-valor reportado por CIBERSORTx para cada muestra. Este p-valor refleja la probabilidad de que la composición celular estimada sea producto del azar, por lo que se retuvieron únicamente las muestras con un p-valor menor a 0.005, considerado el umbral estándar de significancia estadística. Las muestras que no alcanzaron este umbral fueron excluidas del análisis para evitar la incorporación de estimaciones no confiables que pudieran comprometer la validez de los resultados (*Ilustración 2*).

Exploración de proporciones celulares

Las proporciones celulares inferidas fueron exportadas y analizadas en R ((versión 4.5.1)) utilizando principalmente los paquetes ggplot2, pheatmap y ComplexHeatmap, con el propósito de caracterizar el microambiente tumoral e identificar patrones de infiltración inmune entre los pacientes de la cohorte TCGA-HNSC. Este análisis incluyó los siguientes componentes:

- **Análisis descriptivo:** Se calcularon las proporciones medias de cada uno de los 22 tipos celulares en el total de la cohorte, con el fin de identificar las subpoblaciones predominantes en el microambiente tumoral de los pacientes con HNSCC. Los resultados fueron ordenados de mayor a menor proporción media para facilitar su interpretación.
- **Diagrama de barras:** Se generó un diagrama de barras horizontales que representa la proporción media de cada tipo celular en la cohorte, permitiendo una comparación visual directa entre las distintas subpoblaciones inmunes estimadas.

- Mapa de calor (heatmap): Se construyó un mapa de calor de las proporciones celulares en una muestra representativa de 50 pacientes seleccionados aleatoriamente, con agrupamiento jerárquico tanto de filas (tipos celulares) como de columnas (pacientes), utilizando el método de enlace completo y distancia euclidiana. Esta visualización permitió identificar patrones de co-infiltración y subgrupos de pacientes con perfiles inmunes similares.

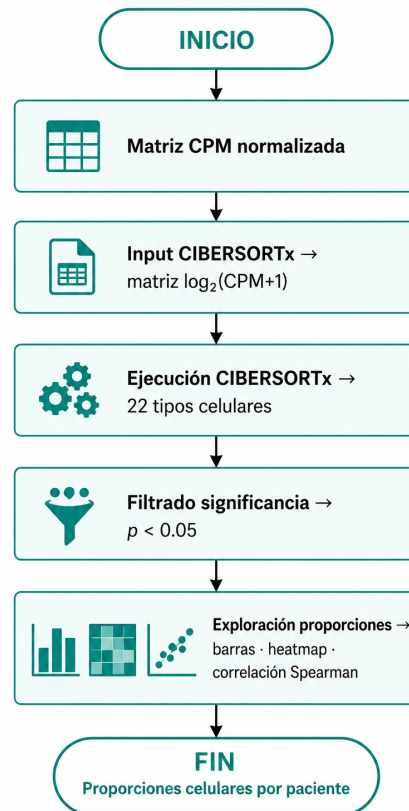


Ilustración 3. Diagrama de flujo del análisis de deconvolución inmunocelular mediante CIBERSORTx. El proceso incluyó el uso de la matriz $\log_2(\text{CPM}+1)$ como entrada, estimación de 22 tipos celulares, filtrado de muestras significativas ($p < 0.05$) y exploración de las proporciones celulares mediante gráficos y análisis de correlación.

6.7.3 Para dar cumplimiento el objetivo específico 3

Con el propósito de identificar subpoblaciones celulares del microambiente tumoral asociadas con la supervivencia global y construir un modelo predictivo con utilidad pronóstica, se implementó un modelo Random Survival Forest (RSF) integrando las proporciones celulares estimadas mediante CIBERSORTx con las variables clínicas de los pacientes pertenecientes a la cohorte TCGA-HNSC (Ilustración 3).

Todos los análisis estadísticos y de modelamiento fueron realizados en R (versión 4.5.1), utilizando principalmente los paquetes randomForestSRC, survival y survminer

Integración de datos

Se realizó la integración de los datos clínicos y las proporciones celulares obtenidas por deconvolución mediante la vinculación de ambas fuentes de información a través del identificador único de cada paciente (patientId). Este proceso de integración permitió conformar un conjunto de datos unificado que incluye tanto las variables clínicas como las 22 proporciones celulares estimadas para cada paciente. Tras la eliminación de los registros con información incompleta en cualquiera de las variables incluidas en el modelo, se obtuvo una cohorte final de 381 pacientes con datos completos de supervivencia global, variables clínicas y proporciones de los 22 tipos celulares.

Variables incluidas en el modelo

Las variables predictoras incorporadas al modelo fueron seleccionadas con base en su relevancia clínica y biológica para el pronóstico del HNSCC, e incluyeron las siguientes:

Variables inmunológicas: proporciones de los 22 tipos celulares estimadas por CIBERSORTx, correspondientes a linfocitos B naive, linfocitos B memoria, células plasmáticas, linfocitos T CD8, linfocitos T CD4 naive, linfocitos T CD4 memoria en reposo, linfocitos T CD4 memoria activados, células T follicular helper, células T reguladoras (Tregs), células T gamma delta, células NK en reposo, células NK activadas, monocitos, macrófagos M0, macrófagos M1, macrófagos M2, células dendríticas en reposo, células dendríticas activadas, mastocitos en reposo, mastocitos activados, eosinófilos y neutrófilos.

Variables clínicas: edad al diagnóstico, sexo, estadio patológico tumoral según la clasificación AJCC, tamaño tumoral (PT), compromiso ganglionar (PN), invasión perineural, grado histológico, sitio anatómico primario y años-paquete de consumo de tabaco.

La variable desenlace fue la supervivencia global, definida mediante el objeto de supervivencia Surv(OS_TIEMPO, OS_EVENTO), donde OS_TIEMPO corresponde al tiempo en meses transcurrido desde el diagnóstico hasta el evento o la fecha de último seguimiento, y OS_EVENTO indica el estado vital del paciente al momento del último registro (1 = fallecido, 0 = censurado vivo).

Entrenamiento del modelo Random Survival Forest

Se implementó un modelo Random Survival Forest (RSF) utilizando el paquete randomForestSRC en el entorno estadístico R. El RSF constituye una extensión del algoritmo Random Forest diseñada específicamente para el análisis de datos de supervivencia censurados. El modelo fue entrenado con los siguientes parámetros: número de árboles (tree) igual a 500, regla de división logrank para datos de supervivencia, número de variables candidatas por nodo (mty) calculado como la raíz cuadrada del número total de variables predictoras, tamaño mínimo de nodo terminal (node size) de 15 observaciones para evitar el sobreajuste, e imputación de valores faltantes mediante el método na.impute, que realiza imputación iterativa durante el proceso de entrenamiento.

Evaluación del modelo

El rendimiento discriminativo del modelo fue evaluado mediante el índice de concordancia de Harrell (C-index), que mide la capacidad del modelo para ordenar correctamente los tiempos de supervivencia de los pacientes. El C-index toma valores entre 0.5 (equivalente al azar) y 1.0 (discriminación perfecta), siendo valores superiores a 0.6 considerados clínicamente relevantes en modelos de supervivencia oncológica. El C-index fue calculado a partir de las predicciones out-of-bag (OOB), es decir, utilizando para cada paciente únicamente las predicciones generadas por los árboles en cuyo entrenamiento ese paciente no fue incluido, lo que constituye una estimación interna no sesgada del poder predictivo del modelo sin necesidad de un conjunto de prueba independiente. Se obtuvo un C-index OOB de 0.645, indicando una capacidad discriminativa moderada-buena, comparable con los valores reportados en estudios similares con datos TCGA en neoplasias de cabeza y cuello.

Importancia de variables

Se calculó la importancia de cada variable predictora mediante el índice VIMP (Variable IMPortance), el cual cuantifica la contribución individual de cada variable a la capacidad predictiva global del modelo. El VIMP se estima perturbando aleatoriamente los valores de cada variable en las muestras OOB y midiendo el incremento en el error de predicción resultante. Los valores de VIMP positivos indican que la variable contribuye de manera real a la predicción de la supervivencia, mientras que valores negativos o cercanos a cero indican que la variable no aporta información predictiva útil o incluso introduce ruido al modelo. Los

resultados fueron ordenados de mayor a menor VIMP para identificar las variables con mayor relevancia pronóstica en la cohorte TCGA-HNSC.

Generación del risk score y estratificación de pacientes

A partir de las predicciones OOB del modelo RSF, se generó un puntaje de riesgo continuo (risk score) para cada paciente, el cual refleja la probabilidad estimada de ocurrencia del evento (fallecimiento) en función de sus características clínicas e inmunológicas. Los pacientes fueron estratificados en dos grupos pronósticos según la mediana del risk score: grupo de alto riesgo (risk score superior a la mediana) y grupo de bajo riesgo (risk score inferior o igual a la mediana), conformando grupos de tamaño equivalente de aproximadamente 190 pacientes cada uno.

Evaluación del risk score como predictor continuo

La asociación entre el risk score continuo y el riesgo de mortalidad fue evaluada mediante un modelo de regresión de Cox univariado, en el que el risk score fue incluido como variable predictora continua. Se estimó el Hazard Ratio (HR) con su intervalo de confianza del 95% y el p-valor correspondiente.

Análisis de supervivencia Kaplan-Meier

Las diferencias en supervivencia global entre los grupos de alto y bajo riesgo fueron evaluadas mediante curvas de Kaplan-Meier utilizando los paquetes survival y survminer.

La comparación estadística entre curvas se realizó mediante la prueba log-rank, considerando significativos valores $p < 0.05$. Este análisis permitió evaluar la capacidad del modelo RSF para estratificar pacientes en grupos con pronósticos clínicamente diferenciados.

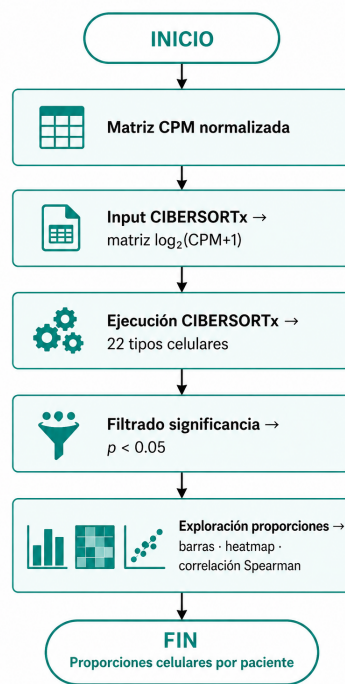


Ilustración 4. Diagrama de flujo del desarrollo y evaluación del modelo predictivo de supervivencia en HNSCC. El proceso incluyó la integración de datos clínicos y proporciones celulares obtenidas por CIBERSORTx, entrenamiento del modelo Random Survival Forest, generación del risk score, estratificación de pacientes y análisis de supervivencia mediante Kaplan-Meier y Hazard Ratio.

4. RESULTADOS

7.1 Control de calidad y Normalización de datos transcriptómicos

A partir de la matriz de expresión génica de la cohorte TCGA-HNSC, compuesta inicialmente por 60.660 genes y 520 muestras tumorales en formato CPM. Se realizó un proceso de curación de datos, con el fin de garantizar la validez y la calidad de los datos para los análisis posteriores.

Como primer paso de control de calidad, se eliminaron los genes sin expresión detectable en ninguna muestra, reduciendo el conjunto a 16,768 genes. Posteriormente, se identificaron y eliminaron genes duplicados mediante promedio, resultando en una matriz final de 16,766 genes, optimizando el conjunto de datos al eliminar ruido técnico y variables no informativas. Los valores de expresión fueron transformados a escala logarítmica ($\log_2(\text{CPM}+1)$) para estabilizar la varianza. La inspección visual mediante boxplots confirmó distribuciones homogéneas a través de la cohorte.

Debido a la magnitud de la cohorte ($n = 520$), se seleccionó una muestra aleatoria de 20 muestras, las cuales mantienen una distribución simétrica, con una mediana de expresión normalizada aproximadamente en $4 \log_2 \text{CPM} + 1$. Esta distribución simétrica es clave, ya que demuestra que no existen sesgos técnicos entre las librerías de secuenciación, permitiendo las comparaciones posteriores estén fundamentadas en variaciones biológicas reales (Ilustración 5).

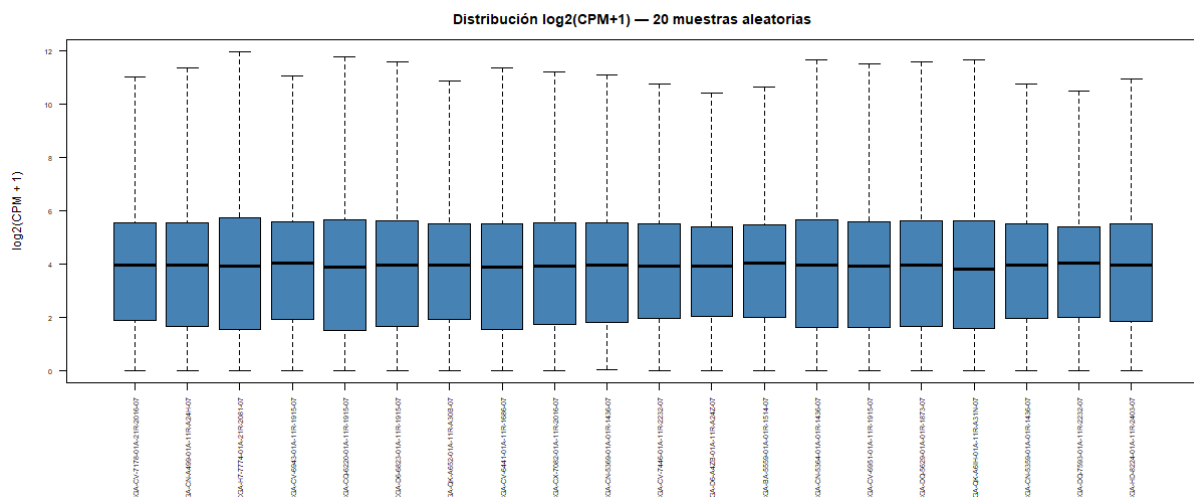


Ilustración 5. Boxplot $\log_2(\text{CPM} + 1)$. Distribución de niveles de expresión logarítmica ($\log_2(\text{CPM} + 1)$) en una muestra de expresión aleatoria representativa de la cohorte de TCGA-HNSCC.

Posteriormente, se evaluó la estructura completa de la totalidad de las muestras mediante un análisis de agrupamiento jerárquico por medio del método de vinculación de WARD.D2 y la métrica de Manhattan.

El análisis usó una totalidad de las 520 muestras y se utilizó como criterio de control de calidad, con el fin de identificar posibles valores atípicos (outliers) que pudieran sesgar los análisis posteriores. La organización jerárquica no reveló presencia de muestras aisladas o agrupamientos fuera de los parámetros; confirmando la integridad del conjunto de datos para su inclusión en los análisis posteriores (Ilustración 6).

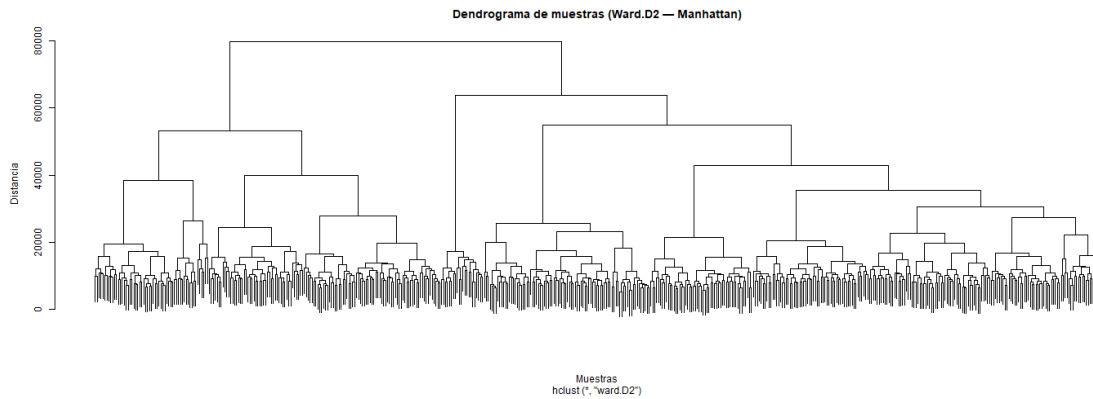


Ilustración 6. Dendrograma de agrupamiento jerárquico de las 520 muestras de la cohorte TCGA-HNSC. Relación de similitud global entre la totalidad de las muestras tumorales del estudio, calculada mediante el método de vinculación de Ward.D2 y la métrica de distancia de Manhattan

Finalmente, los procesos de filtrado y normalización resultaron en una matriz robusta con una varianza estabilizada, cumpliendo con los estándares de calidad necesarios para la identificación de poblaciones celulares presentes en el TME, mediante el análisis de datos de *RNA-seq*

7.2 Caracterización del Microambiente tumoral mediante Deconvolución

Las 520 muestras de expresión génica normalizada fueron procesadas mediante CIBERSORTx para estimar las proporciones de 22 tipos celulares del microambiente tumoral. Aplicando el criterio de significancia estadística ($p < 0.005$), se retuvieron 381 muestras para el análisis, asegurando que las proporciones estimadas posean los parámetros necesarios requeridos en cuanto a las firmas genéticas de referencia (LM22).

7.2.1 Heterogeneidad tumoral y Caracterización fenotípica

El análisis de la composición medial muestra un microambiente tumoral inmunosupresor, caracterizado por una respuesta inmunitaria inhibida (*Ilustración 7*)

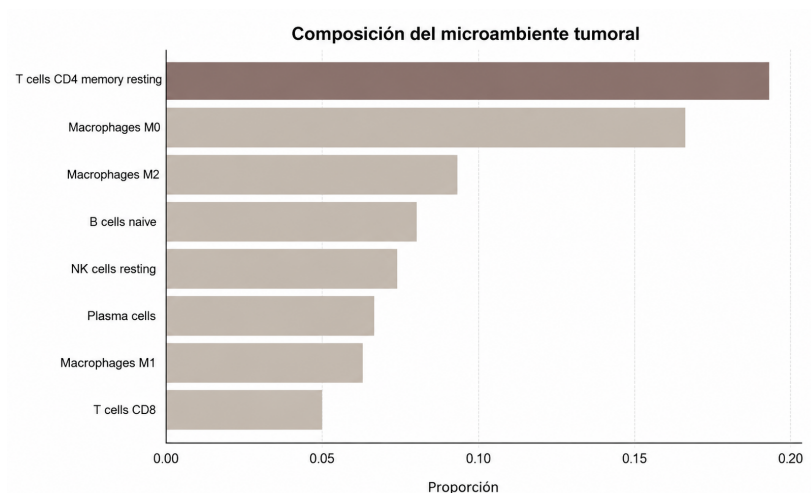


Ilustración 7. Gráfico de barras que jerarquiza las poblaciones celulares según su abundancia relativa media. Se destaca la dominancia de perfiles de memoria en reposo y macrófagos no polarizados en la cohorte evaluada (n=381).

- Células T-CD4 de memoria: La población más abundante con un 18% de proporción media. La prevalencia de esta población en “reposo” indica una exclusión de señalización antigénica ineficaz o una insuficiencia de señales de coestimulación mediada por el microambiente estromal.
- Macrófagos M0: La presencia de macrófagos M0 (17%), seguida de M2 (10%), es un hallazgo muy relevante. Los Macrófagos M0 representan células sin una especialización funcional; su alta prevalencia indica un reclutamiento masivo de monocitos que no logran diferenciarse hacia un fenotipo pro-inflamatorio (M1). Por otro lado, la relación M2/M1 (10 % Vs 7%) indican un microambiente protumoral, asociado a la secreción de citocinas como IL-10-
- Linfocitos B y Células Plasmáticas: La suma de Linfocitos B (10%) y Células Plasmáticas (9%) indican una participación importante en la inmunidad humoral del HNSCC, lo que sugiere una transición activa desde el momento del reconocimiento antigénico hasta la diferenciación de células secretoras. Esta coexistencia podría estar vinculada a la formación de estructuras linfoides terciarias (TLS), las cuales actuarían como centros operativos para coordinar una respuesta inmunitaria local y sostenida dentro del microambiente tumoral.

Tabla 2. Distribución descriptiva de las principales subpoblaciones inmunes inferidas mediante CIBERSORTx en la cohorte TCGA-HNSC. La tabla presenta las 10 subpoblaciones inmunes con mayor abundancia relativa dentro del microambiente tumoral de pacientes con carcinoma escamocelular de cabeza y cuello (HNSCC). Se reportan medidas de tendencia central y dispersión, incluyendo media $\pm$ desviación estándar (DE), mediana y rango intercuartílico (RIC), expresadas como porcentaje de infiltración celular estimado mediante deconvolución transcriptómica con CIBERSORTx.

Población celular	Mediana (%)	RIC (%)	Media $\pm$ DE
T CD4 Memoria	19.22	11.59–24.91	18.01 $\pm$ 9.03
Macrófagos M0	12.31	4.50–24.67	16.65 $\pm$ 14.98
Macrófagos M2	8.68	5.04–13.17	9.55 $\pm$ 6.52
Linfocitos B	7.46	3.26–13.91	9.52 $\pm$ 8.0
Células plasmáticas	5.60	1.31–13.10	8.31 $\pm$ 8.65
Células NK en reposo	7.13	4.93–9.96	7.35 $\pm$ 4.01
Macrófagos M1	5.21	1.33–10.72	6.99 $\pm$ 6.80
Linfocitos T CD8	4.08	0.11–8.60	5.89 $\pm$ 6.53
Células dendríticas activadas	1.95	0.20–5.07	3.55 $\pm$ 4.42
Células T helper	0.70 .	0–3.39	2.19 $\pm$ 3.26

Las proporciones relativas de las subpoblaciones inmunes inferidas mediante CIBERSORTx evidenciaron una marcada heterogeneidad interpaciente dentro de la cohorte TCGA-HNSC (Tabla 3). Las células T CD4 memoria en reposo constituyeron la población predominante, con una mediana de infiltración del 19.22 % (RIC: 11.59–24.91) y una media de 18.01 % $\pm$ 9.03 DE.

Entre las poblaciones mieloides, los macrófagos M0 mostraron una elevada variabilidad interpaciente, reflejada por una desviación estándar de 14.98 % y amplios rangos de infiltración tumoral (0–73.29 %). Asimismo, los macrófagos M2, células plasmáticas y linfocitos T CD8 también presentaron una dispersión considerable entre pacientes, sugiriendo una composición inmunológica altamente heterogénea dentro del microambiente tumoral de HNSCC.

En conjunto, estos resultados evidencian la coexistencia de poblaciones inmunes efectoras y reguladoras dentro del microambiente tumoral, así como una amplia diversidad inmunológica entre pacientes de la cohorte analizada.

7.2.3 Caracterización Transcriptómica del Microambiente de HNSCC

La matriz de correlación de Spearman permite identificar las poblaciones celulares que operan de forma antagónica y similares (*Ilustración 8*).

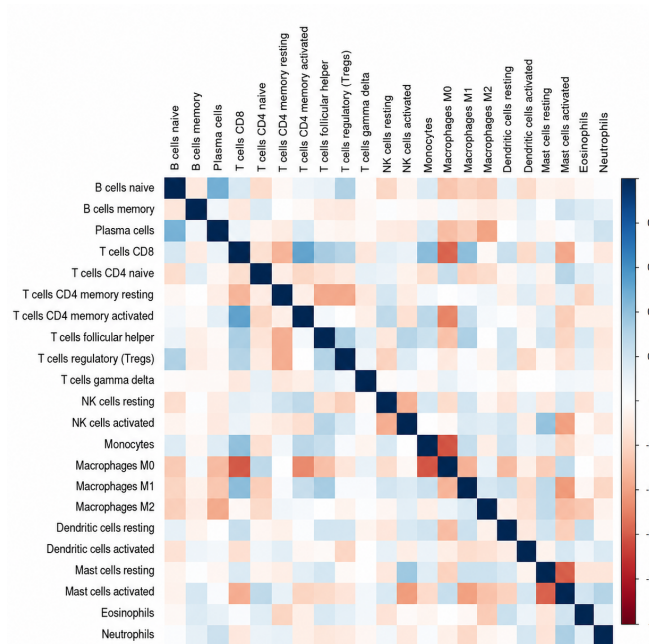


Ilustración 8. Matriz de correlación de Spearman entre poblaciones celulares. Análisis de Spearman que revela las interacciones funcionales en el TME. Los coeficientes positivos indican comunidades celulares cooperativas, mientras que los negativos sugieren procesos de exclusión competitiva.

- **Correlación de infiltración linfoide:** Se identificó una correlación positiva (> 0.6) entre células T CD8, T CD4 activadas, T helper y Tregs. Este hallazgo indica que la infiltración linfoide en HNSCC es un evento coordinado y no aleatorio; no obstante, la co-infiltración de las Tregs dentro de este conjunto sugiere que, incluso en tumores con una alta infiltración de células inmunes, el microambiente establece mecanismos de contrarregulación inmediata para mitigar la respuesta efectora antitumoral.
- **Antagonismo Mieloide-Linfoide:** Se identifica una correlación negativa significativa entre los Macrófagos M0 y la totalidad de células T (especialmente CD4 de memoria) y células B; esto permite identificar un eje de antagonismo en donde predominan los macrófagos no polarizados, donde se produce un aislamiento del tumor, impidiendo la supervivencia de las poblaciones linfoides. Biológicamente, esto sugiere un fenotipo de "exclusión inmune", donde el compartimento mieloide impone barreras físicas o metabólicas que restringen la infiltración linfoide y favorecen la evasión tumoral.

7.3 Modelo de Predicción de Supervivencia mediante Random Survival Forest (RSF)

Se implementó un modelo integrativo de supervivencia tipo Random Survival Forest utilizando como desenlace la supervivencia global OS_TIME, OS_EVENT, donde OS_EVENT fue codificado como 1 = fallecido y 0 = censurado vivo. El modelo incluyó variables clínicas y las proporciones de las 22 subpoblaciones inmunes inferidas mediante CIBERSORTx. El entrenamiento fue realizado con 500 árboles, regla de división log-rank y evaluación mediante predicciones out-of-bag (OOB). Se incluyeron 381 pacientes de la cohorte TCGA-HNSC con datos clínicos completos y perfiles de deconvolución estadísticamente significativos

7.3.1. Desempeño y Capacidad Predictiva del Modelo

El modelo fue entrenado con una arquitectura de 500 árboles, incorporando 22 tipos celulares y variables clínicas clave (edad, estado VPH, entre otras). El desempeño del modelo alcanzó un C-index Out-of-Bag (OOB) de 0.645, lo que indica una capacidad discriminativa moderada-buena para predecir eventos de mortalidad en pacientes con HNSCC basándose en la heterogeneidad de su microambiente.

El risk score generado por el modelo RSF resultó ser un predictor altamente significativo. El análisis de riesgos proporcionales mostró un Hazard Ratio (HR) de 1.041 (IC 95%: 1.034–1.048, $p < 0.0001$), lo que sugiere que por cada unidad de incremento en el puntaje de riesgo derivado de la combinación clínico-inmune, la probabilidad de fallecimiento aumenta de manera constante.

7.3.2 Estratificación del riesgo y Curvas de Kaplan-Meier

La validez del modelo se visualiza mediante la estratificación de la cohorte en dos grupos (Alto Riesgo y Bajo Riesgo) basados en la mediana del risk score (*Ilustración 11*).

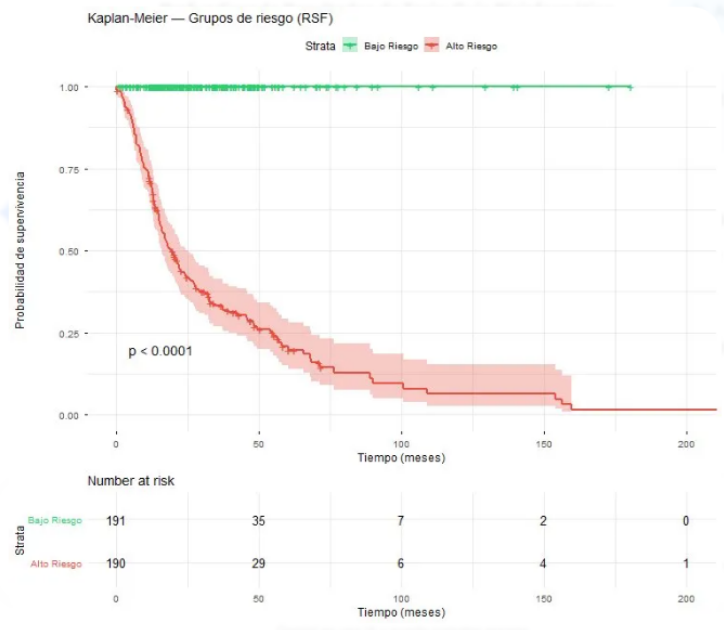


Ilustración 11. Análisis de supervivencia de Kaplan-Meier para grupos de riesgo. Estratificación de 381 pacientes de la cohorte TCGA-HNSC según el risk score del modelo RSF.

- Interpretación de la Supervivencia: Se observan diferencias estadísticamente significativas en la supervivencia global (log-rank $p < 0.0001$). El grupo de alto riesgo presenta una caída en la probabilidad de supervivencia desde los primeros meses de seguimiento, alcanzando una mediana de supervivencia inferior al grupo de bajo riesgo, el cual presenta una estabilidad notable en el tiempo evaluado. Esto confirma que la firma de infiltración celular, cuando se correlaciona con datos clínicos, puede ser un marcador biológico de agresividad tumoral.

7.3.3 Identificación de predictores Celulares y Clínicos de supervivencia Global en HNSCC

El análisis de importancia de variables permitió identificar qué factores impulsan la mortalidad en la cohorte (*Ilustración 12*).

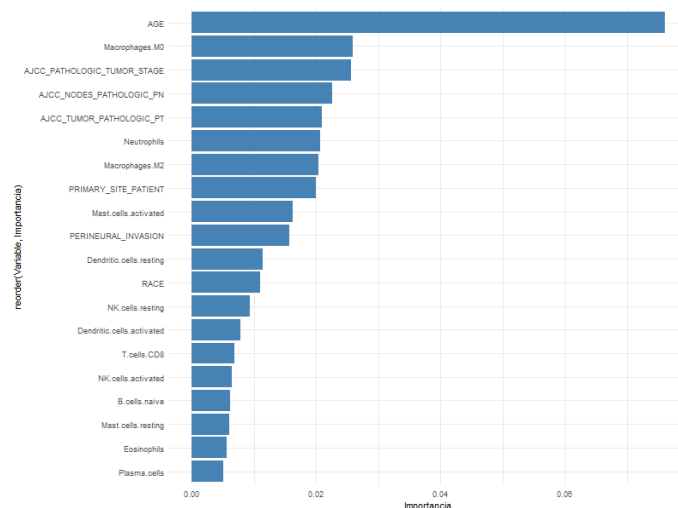


Ilustración 12. Importancia de las variables en el modelo Random Survival Forest (VIMP). Clasificación de los predictores según su contribución a la precisión del modelo. Se destaca la relevancia de los Macrófagos M0 como biomarcadores celulares de supervivencia, posicionándose al mismo nivel de importancia que los criterios de estadificación TNM (AJCC)

- Factores Clínicos: La edad (VIMP = 0.076) se posicionó como el predictor individual más fuerte, seguida por el estadio patológico AJCC (0.026) y la afectación ganglionar (0.023). Esto es consistente con la práctica clínica oncológica actual.
- Factores del Microambiente (CIBERSORTx): De manera relevante, los Macrófagos M0 (VIMP = 0.026) emergieron como la variable celular con mayor peso predictivo, igualando en importancia al estadio tumoral. Asimismo, la presencia de Neutrófilos (0.021) y Macrófagos M2 se identificaron como predictores críticos de mal pronóstico.
- Significancia Biológica: El hecho de que los macrófagos sin especialización (M0) tengan más peso en el modelo que muchas variables clínicas tradicionales subraya que un microambiente "inmaduro" o suprimido es tan letal como un estadio avanzado del tumor.

5. DISCUSIÓN

En el presente estudio se realizó un análisis bioinformático integrativo de datos transcriptómicos bulk RNA-seq de la cohorte TCGA-HNSC con el propósito de caracterizar el microambiente inmune tumoral mediante deconvolución transcriptómica y evaluar su valor pronóstico a través de un modelo Random Survival Forest

Al analizar la composición del TME, la marcada heterogeneidad de los macrófagos M0 (17%) y M2 (10%) representa el hallazgo central sobre la arquitectura inmunológica de la cohorte. En la oncología, el cúmulo de macrófagos M0 ya no se considera un estado de

inactividad basal, sino un reservorio de plasticidad mieloide que el tumor secuestra activamente para su progresión (73). El tumor secreta quimiocinas que reclutan monocitos periféricos, pero el microambiente metabólicamente hostil, caracterizado por hipoxia y acumulación de lactato impide su diferenciación hacia el fenotipo antitumoral M1, desviándose hacia el estado supresor M2. Este predominio de macrófagos pro-tumorales genera una barrera de exclusión inmune que explica la infiltración moderada de linfocitos T CD8⁺ (6%) observada (73,74). Biológicamente, esto sugiere un fenómeno de exclusión mecánica y química donde los macrófagos consumen metabolitos esenciales, provocando la detención del ciclo celular de los linfocitos T y limitando su capacidad citotóxica en el nicho tumoral (74,75). Este hallazgo es consistente con lo reportado por Qi et al (30), quienes identificaron en una cohorte de HNSCC un perfil de infiltración mieloide dominante con baja proporción de células T citotóxicas, sugiriendo que este patrón constituye una característica biológica inherente de esta neoplasia y no un artefacto de la deconvolución.

La infiltración moderada de células T CD4 memoria (18%) debe interpretarse en el contexto del estado del VPH

El modelo Random Survival Forest alcanzó un C-index de 0.645, indicando una capacidad discriminativa moderada para la predicción de supervivencia global en pacientes con HNSCC. Este desempeño es comparable al reportado en estudios previos que han utilizado modelos de aprendizaje automático aplicados a datos transcriptómicos de TCGA-HNSC, donde los valores de C-index suelen oscilar entre 0.60 y 0.70 para firmas inmunes y transcriptómicas derivadas de RNA-seq bulk. Estos resultados son particularmente relevantes considerando que, en este tipo de datos, la señal proveniente de poblaciones inmunes puede verse parcialmente diluida por la abundancia de ARN tumoral y estromal, lo que representa un desafío para la construcción de modelos pronósticos robustos (72).

Diversos estudios han demostrado la utilidad de modelos de supervivencia basados en transcriptómica para integrar variables clínicas e inmunológicas en HNSCC. Fang et al. (70) desarrollaron una firma inmune pronóstica basada en genes inmunorrelacionados derivados de TCGA-HNSC, logrando una adecuada estratificación de pacientes según riesgo y diferencias significativas en supervivencia global. De manera similar, Wu et al. (71) identificaron una firma transcriptómica inmunorrelacionada capaz de clasificar pacientes en grupos de alto y bajo riesgo con diferencias pronósticas significativas, destacando el valor de las señales inmunológicas derivadas de transcriptómica bulk en HNSCC.

Finalmente, la separación altamente significativa en las curvas de Kaplan-Meier ($p < 0.0001$) consolida la utilidad clínica de este estudio para el área de la salud. El *risk score* derivado del modelo RSF no solo predice la supervivencia global, sino que permite identificar a una subpoblación de pacientes "críticos" que, a pesar de presentar estadios clínicos aparentemente favorables, poseen un microambiente tumoral altamente hostil que predice una progresión rápida (84,85).

No obstante, estos hallazgos deben interpretarse considerando ciertas limitaciones. En primer lugar, el uso de datos de RNA-seq bulk implica que las estimaciones de composición celular derivadas de herramientas deconvolutivas dependen de firmas de referencia y pueden no capturar completamente la complejidad del microambiente tumoral. Adicionalmente, aunque el uso de datos de *The Cancer Genome Atlas* permite trabajar con cohortes amplias y bien caracterizadas, la naturaleza retrospectiva de estos datos limita la generalización clínica inmediata (78,79). Asimismo, la ausencia de validación externa en cohortes independientes restringe la extrapolación del modelo a otras poblaciones (80)

A pesar de estas limitaciones, el estudio presenta implicaciones clínicas relevantes. La integración de variables inmunológicas con características clínicas en un modelo predictivo sugiere que el microambiente tumoral puede aportar información adicional para la estratificación pronóstica más allá de los factores tradicionales. En particular, el papel de los macrófagos resalta posibles blancos terapéuticos y biomarcadores que podrían ser explorados en el contexto de terapias inmunomoduladoras. Además, el puntaje de riesgo generado podría servir como herramienta para identificar pacientes con mayor probabilidad de mal pronóstico, lo que facilita la toma de decisiones clínicas más personalizadas.

Finalmente, futuras investigaciones deberían enfocarse en validar estos hallazgos en cohortes externas y en integrar datos multi-ómicos que permitan una caracterización más profunda del microambiente tumoral. La incorporación de datos de transcriptómica de célula única podría mejorar la resolución de los análisis y superar las limitaciones inherentes a los datos bulk. Asimismo, el desarrollo de modelos híbridos que combinen enfoques estadísticos y de aprendizaje automático podría optimizar la capacidad predictiva y su aplicabilidad en entornos clínicos reales.

6. CONCLUSIONES

En el presente estudio se realizó un análisis bioinformático integrativo de datos transcriptómicos bulk *RNA-seq* de la cohorte TCGA-HNSC con el propósito de caracterizar

el microambiente tumoral mediante deconvolución transcriptómica y evaluar su valor pronóstico a través de un modelo Random Survival Forest. El procesamiento sistemático de los datos garantiza la obtención de una matriz de expresión confiable, minimizando sesgos técnicos y conservando información biológicamente relevante para los análisis posteriores.

La deconvolución transcriptómica evidenció un microambiente predominantemente inmunosupresor, caracterizado por la abundancia de macrófagos M0 y M2 y una infiltración moderada de linfocitos T CD8, hallazgos coherentes con la literatura y que refuerzan el papel del sistema inmune en la progresión tumoral. Esta caracterización permitió identificar patrones de infiltración que reflejan la heterogeneidad biológica del tumor y su posible influencia en la respuesta terapéutica.

Adicionalmente, el modelo predictivo de supervivencia basado en Random Survival Forest evidenció un desempeño adecuado, logrando estratificar a los pacientes en grupos de riesgo significativamente diferenciados. La relevancia de variables como la edad, los macrófagos y los neutrófilos resalta la importancia de integrar factores clínicos e inmunológicos en la predicción del pronóstico. En conjunto, estos resultados sugieren que el microambiente tumoral no solo es un reflejo del estado biológico del tumor, sino también un componente clave en la determinación del desenlace clínico.

Referencias bibliográficas

1. Cárcamo M. Epidemiología y generalidades del tumor de cabeza y cuello. Rev Médica Clínica Las Condes. 1 de julio de 2018;Tema central: Diagnóstico precoz y prevención en cáncer de cabeza y cuello29(4):388-96. doi:10.1016/j.rmclc.2018.06.009
2. Epidemiology, Risk Factors, and Prevention of Head and Neck Squamous Cell Carcinoma [Internet]. [citado 21 de octubre de 2025]. Disponible en: <https://www.mdpi.com/2076-3271/11/2/42>
3. Arneth B. Tumor Microenvironment. Med Kaunas Lith. 30 de diciembre de 2019;56(1):15. doi:10.3390/medicina56010015 PubMed PMID: 31906017; PubMed Central PMCID: PMC7023392.
4. Genetic effects on gene expression across human tissues | Nature [Internet]. [citado 6 de mayo de 2026]. Disponible en: <https://www.nature.com/articles/nature24277>
5. Nallasamy P, Nimmakayala RK, Parte S, Are AC, Batra SK, Ponnusamy MP. Tumor microenvironment enriches the stemness features: the architectural event of therapy resistance and metastasis. Mol Cancer. 22 de diciembre de 2022;21(1):225. doi:10.1186/s12943-022-01682-x
6. Aboaid H, Khalid T, Hussain A, Myat YM, Nanda RK, Srinivasmurthy R, et al. Advances and challenges in immunotherapy in head and neck cancer. Front Immunol. 2025;16:1596583. doi:10.3389/fimmu.2025.1596583 PubMed PMID: 40547025; PubMed Central PMCID: PMC12179090.
7. Liu Y, Zhang N, Wen Y, Wen J. Head and neck cancer: pathogenesis and targeted therapy. MedComm. septiembre de 2024;5(9):e702. doi:10.1002/mco2.702 PubMed PMID: 39170944; PubMed Central PMCID: PMC11338281.
8. Parikh AS, Li Y, Mazul A, Yu VX, Thorstad W, Rich J, et al. Immune Cell Deconvolution Reveals Possible Association of $\gamma\delta$ T Cells with Poor Survival in Head and Neck Squamous Cell Carcinoma. Cancers. enero de 2023;15(19):4855. doi:10.3390/cancers15194855
9. CIBERSORTx [Internet]. [citado 22 de octubre de 2025]. Disponible en: <https://cibersortx.stanford.edu/tutorial.php>
10. Sturm G, Finotello F, Petitprez F, Zhang JD, Baumbach J, Fridman WH, et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. Bioinformatics. julio de 2019;35(14):i436-45. doi:10.1093/bioinformatics/btz363 PubMed PMID: 31510660; PubMed Central PMCID: PMC6612828.
11. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. julio de 2019;37(7):773-82. doi:10.1038/s41587-019-0114-2 PubMed PMID: 31061481; PubMed Central PMCID: PMC6610714.

12. Eleven grand challenges in single-cell data science | Genome Biology | Springer Nature Link [Internet]. [citado 6 de mayo de 2026]. Disponible en: <https://link.springer.com/article/10.1186/s13059-020-1926-6>

13. Steen CB, Luca BA, Esfahani MS, Azizi A, Swarder BJ, Nabat BY, et al. The landscape of tumor cell states and ecosystems in diffuse large B cell lymphoma. *Cancer Cell*. 11 de octubre de 2021;39(10):1422-1437.e10. doi:10.1016/j.ccell.2021.08.011 PubMed PMID: 34597589; PubMed Central PMCID: PMC9205168.

14. Zhang L, Huang T, Xu F, Li S, Zheng S, Lyu J, et al. Prediction of prognosis in elderly patients with sepsis based on machine learning (random survival forest). *BMC Emerg Med*. 11 de febrero de 2022;22(1):26. doi:10.1186/s12873-022-00582-z

15. Canning M, Guo G, Yu M, Myint C, Groves MW, Byrd JK, et al. Heterogeneity of the Head and Neck Squamous Cell Carcinoma Immune Landscape and Its Impact on Immunotherapy. *Front Cell Dev Biol*. 9 de abril de 2019;7:52. doi:10.3389/fcell.2019.00052 PubMed PMID: 31024913.

16. Farah CS. Molecular landscape of head and neck cancer and implications for therapy. *Ann Transl Med*. mayo de 2021;9(10):915. doi:10.21037/atm-20-6264 PubMed PMID: 34164549.

17. Johnson DE, Burtress B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. *Nat Rev Dis Primer*. 26 de noviembre de 2020;6(1):92. doi:10.1038/s41572-020-00224-3 PubMed PMID: 33243986; PubMed Central PMCID: PMC7944998.

18. Qin T, Li S, Henry LE, Liu S, Sartor MA. Molecular Tumor Subtypes of HPV-Positive Head and Neck Cancers: Biological Characteristics and Implications for Clinical Outcomes. *Cancers*. enero de 2021;13(11):2721. doi:10.3390/cancers13112721

19. Elmusrati A, Wang J, Wang CY. Tumor microenvironment and immune evasion in head and neck squamous cell carcinoma. *Int J Oral Sci*. 2 de agosto de 2021;13(1):24. doi:10.1038/s41368-021-00131-7

20. Head and neck squamous cell carcinoma | Nature Reviews Disease Primers [Internet]. [citado 7 de mayo de 2026]. Disponible en: https://www.nature.com/articles/s41572-020-00224-3?utm_source=chatgpt.com

21. Cheon J, Kim B, Park J, Shin J, Kim TH. Unveiling Biomarkers in Head and Neck Squamous Cell Carcinoma through Bioinformatics: The Role of SPP1 and KRT78. *Int J Mol Sci*. 10 de noviembre de 2024;25(22):12062. doi:10.3390/ijms252212062 PubMed PMID: 39596132; PubMed Central PMCID: PMC11594244.

22. Emerging biomarkers for early cancer detection and diagnosis: challenges, innovations, and clinical perspectives - PMC [Internet]. [citado 6 de mayo de 2026]. Disponible en: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12359849/>

23. Wang Y, Li C, Yang F, Gong M, Qu J, Ma R, et al. Random survival forest predicts survival in patients with metastatic laryngeal and hypopharyngeal cancer and the prognostic benefits of surgery and radiotherapy. *J Cancer*. 1 de enero de 2025;16(2):603-21. doi:10.7150/jca.103793

24. Zhang A, He X, Zhang C, Tang X. Molecular subtype identification and prognosis stratification based on golgi apparatus-related genes in head and neck squamous cell

carcinoma. *BMC Med Genomics*. 16 de febrero de 2024;17(1):53. doi:10.1186/s12920-024-01823-9

25. Yang A, Wu M, Ni M, Zhang L, Li M, Wei P, et al. A risk scoring system based on tumor microenvironment cells to predict prognosis and immune activity in triple-negative breast cancer. *Breast Cancer*. 1 de mayo de 2022;29(3):468-77. doi:10.1007/s12282-021-01326-w

26. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. julio de 2019;37(7):773-82. doi:10.1038/s41587-019-0114-2

27. Valkiers S, de Vrij N, Gielis S, Verbandt S, Ogunjimi B, Laukens K, et al. Recent advances in T-cell receptor repertoire analysis: Bridging the gap with multimodal single-cell RNA sequencing. *Immunoinformatics*. 1 de marzo de 2022;5:100009. doi:10.1016/j.immuno.2022.100009

28. Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, et al. Eleven grand challenges in single-cell data science. *Genome Biol*. 7 de febrero de 2020;21(1):31. doi:10.1186/s13059-020-1926-6 PubMed PMID: 32033589; PubMed Central PMCID: PMC7007675.

29. Maden SK, Kwon SH, Huuki-Myers LA, Collado-Torres L, Hicks SC, Maynard KR. Challenges and opportunities to computationally deconvolve heterogeneous tissue with varying cell sizes using single-cell RNA-sequencing datasets. *Genome Biol*. 14 de diciembre de 2023;24(1):288. doi:10.1186/s13059-023-03123-4 PubMed PMID: 38098055; PubMed Central PMCID: PMC10722720.

30. Qi Z, Liu Y, Mints M, Mullins R, Sample R, Law T, et al. Single-Cell Deconvolution of Head and Neck Squamous Cell Carcinoma. *Cancers*. enero de 2021;13(6):1230. doi:10.3390/cancers13061230

31. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol*. 2018;1711:243-59. doi:10.1007/978-1-4939-7493-1_12 PubMed PMID: 29344893; PubMed Central PMCID: PMC5895181.

32. Zhang X, Liu G, Peng X. A Random Forest Model for Post-Treatment Survival Prediction in Patients with Non-Squamous Cell Carcinoma of the Head and Neck. *J Clin Med*. 30 de julio de 2023;12(15):5015. doi:10.3390/jcm12155015 PubMed PMID: 37568416; PubMed Central PMCID: PMC10419643.

33. Hu X, Dong H, Qin W, Bin Y, Huang W, Kang M, et al. Machine learning-based identification of a consensus immune-derived gene signature to improve head and neck squamous cell carcinoma therapy and outcome. *Front Pharmacol*. 10 de abril de 2024;15:1341346. doi:10.3389/fphar.2024.1341346 PubMed PMID: 38666027; PubMed Central PMCID: PMC11044683.

34. Karpathiou G, Casteillo F, Giroult JB, Forest F, Fournel P, Monaya A, et al. Prognostic impact of immune microenvironment in laryngeal and pharyngeal squamous cell carcinoma: Immune cell subtypes, immuno-suppressive pathways and clinicopathologic characteristics. *Oncotarget*. 27 de diciembre de 2016;8(12):19310-22. doi:10.18632/oncotarget.14242 PubMed PMID: 28038471; PubMed Central PMCID: PMC5386686.

35. Fang R, Iqbal M, Chen L, Liao J, Luo J, Wei F, et al. A novel comprehensive immune-related gene signature as a promising survival predictor for the patients with head and neck squamous cell carcinoma. *Aging*. 17 de abril de 2021;13(8):11507-27. doi:10.18632/aging.202842 PubMed PMID: 33867351; PubMed Central PMCID: PMC8109104.
36. Zafar S, Hafeez A, Shah H, Mutiullah I, Ali A, Khan K, et al. Emerging biomarkers for early cancer detection and diagnosis: challenges, innovations, and clinical perspectives. *Eur J Med Res*. 18 de agosto de 2025;30:760. doi:10.1186/s40001-025-03003-6 PubMed PMID: 40826140; PubMed Central PMCID: PMC12359849.
37. Metastatic disease in head & neck oncology - PMC [Internet]. [citado 8 de mayo de 2026]. Disponible en: <https://pmc.ncbi.nlm.nih.gov/articles/PMC7263073/>
38. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229-63. doi:10.3322/caac.21834 PubMed PMID: 38572751.
39. Alcohol and Cancer: Epidemiology and Biological Mechanisms [Internet]. [citado 8 de mayo de 2026]. Disponible en: <https://www.mdpi.com/2072-6643/13/9/3173>
40. Oropharyngeal cancer as a driver of racial outcome disparities in squamous cell carcinoma of the head and neck: 10-year experience at the University of Maryland Greenebaum Cancer Center - Zandberg - 2016 - Head & Neck - Wiley Online Library [Internet]. [citado 8 de mayo de 2026]. Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hed.23933>
41. Li Y, Yadollahi P, Essien FN, Putluri V, Ambati CSR, Kami Reddy KR, et al. Tobacco smoke exposure is a driver of altered oxidative stress response and immunity in head and neck cancer. *J Transl Med*. 5 de abril de 2025;23(1):403. doi:10.1186/s12967-025-06258-z
42. Li H, Torabi SJ, Yarbrough WG, Mehra S, Osborn HA, Judson B. Association of Human Papillomavirus Status at Head and Neck Carcinoma Subsites With Overall Survival. *JAMA Otolaryngol– Head Neck Surg*. 1 de junio de 2018;144(6):519-25. doi:10.1001/jamaoto.2018.0395 PubMed PMID: 29801040; PubMed Central PMCID: PMC6583856.
43. Kim HAJ, Zeng PYF, Shaikh MH, Mundi N, Ghasemi F, Di Gravio E, et al. All HPV-negative head and neck cancers are not the same: Analysis of the TCGA dataset reveals that anatomical sites have distinct mutation, transcriptome, hypoxia, and tumor microenvironment profiles. *Oral Oncol*. 1 de mayo de 2021;116:105260. doi:10.1016/j.oraloncology.2021.105260
44. Immune Evasion in Head and Neck Squamous Cell Carcinoma: Roles of Cancer-Associated Fibroblasts, Immune Checkpoints, and TP53 Mutations in the Tumor Microenvironment [Internet]. [citado 8 de mayo de 2026]. Disponible en: <https://www.mdpi.com/2072-6694/17/15/2590>
45. Su ZY, Siak PY, Leong CO, Cheah SC. The role of Epstein–Barr virus in nasopharyngeal carcinoma. *Front Microbiol*. 9 de febrero de 2023;14:1116143. doi:10.3389/fmicb.2023.1116143 PubMed PMID: 36846758; PubMed Central PMCID: PMC9947861.

46. Richardo T, Prattapong P, Ngernsombat C, Wisetyaningsih N, Iizasa H, Yoshiyama H, et al. Epstein-Barr Virus Mediated Signaling in Nasopharyngeal Carcinoma Carcinogenesis. *Cancers*. 28 de agosto de 2020;12(9):2441. doi:10.3390/cancers12092441 PubMed PMID: 32872147; PubMed Central PMCID: PMC7565514.
47. Systemic and Local Effects Among Patients With Betel Quid-Related Oral Cancer - Hui-Ching Chuang, Ming-Hsien Tsai, Yu-Tsai Lin, Ming-Huei Chou, Kun-Lin Yang, Chih-Yen Chien, 2022 [Internet]. [citado 8 de mayo de 2026]. Disponible en: <https://journals.sagepub.com/doi/full/10.1177/15330338221146870>
48. Atef AM, Abdul-Hamid G. DEMOGRAPHIC PROFILE AND RISK FACTORS OF HEAD AND NECK SQUAMOUS CELL CARCINOMA IN NATIONAL ONCOLOGY CENTER IN ADEN, YEMEN: PROSPECTIVE CROSS-SECTIONAL STUDY. *Electron J Univ Aden Basic Appl Sci*. 30 de septiembre de 2024;5(3):225-35. doi:10.47372/ejua-ba.2024.3.368
49. Pinkiewicz M, Dorobisz K, Zatoński T. Human Papillomavirus-Associated Head and Neck Cancers. Where are We Now? A Systematic Review. *Cancer Manag Res*. 25 de noviembre de 2022;14:3313-24. doi:10.2147/CMAR.S379173
50. Frontiers | The role of the tumor microenvironment in HNSCC resistance and targeted therapy [Internet]. [citado 8 de mayo de 2026]. Disponible en: <https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2025.1554835/full>
51. Tiwari A, Trivedi R, Lin SY. Tumor microenvironment: barrier or opportunity towards effective cancer therapy. *J Biomed Sci*. 17 de octubre de 2022;29(1):83. doi:10.1186/s12929-022-00866-3 PubMed PMID: 36253762; PubMed Central PMCID: PMC9575280.
52. The Use of Immune Regulation in Treating Head and Neck Squamous Cell Carcinoma (HNSCC) [Internet]. [citado 8 de mayo de 2026]. Disponible en: <https://www.mdpi.com/2073-4409/13/5/413>
53. Alves A, Diel L, Ramos G, Pinto A, Bernardi L, Yates J, et al. Tumor microenvironment and Oral Squamous Cell Carcinoma: A crosstalk between the inflammatory state and tumor cell migration. *Oral Oncol*. 1 de enero de 2021;112:105038. doi:10.1016/j.oraloncology.2020.105038
54. Single-Cell Deconvolution of Head and Neck Squamous Cell Carcinoma [Internet]. [citado 8 de mayo de 2026]. Disponible en: <https://www.mdpi.com/2072-6694/13/6/1230>
55. Oishi N, Orozco-Núñez S, Alba-García JR, Gimeno-Coret M, Zapater E. Survival and Prognostic Factors in Unresectable Head and Neck Cancer Patients. *J Clin Med*. enero de 2025;14(15):5517. doi:10.3390/jcm14155517
56. CD8+ and Regulatory T cells Differentiate Tumor Immune Phenotypes and Predict Survival in Locally Advanced Head and Neck Cancer [Internet]. [citado 8 de mayo de 2026]. Disponible en: <https://www.mdpi.com/2072-6694/11/9/1398>
57. Najafi M, Hashemi Goradel N, Farhood B, Salehi E, Nashtaei MS, Khanlarkhani N, et al. Macrophage polarity in cancer: A review. *J Cell Biochem*. marzo de 2019;120(3):2756-65. doi:10.1002/jcb.27646 PubMed PMID: 30270458.

58. Chen SH, Hsiao SY, Chang KY, Chang JY. New Insights Into Oral Squamous Cell Carcinoma: From Clinical Aspects to Molecular Tumorigenesis. *Int J Mol Sci.* enero de 2021;22(5):2252. doi:10.3390/ijms22052252
59. Wang J, Li H, Shi M, Ren C, Wei W, Zhao Q, et al. Combined single-cell RNA-seq and bulk RNA-seq construction of M2 TAMs signature for predicting HNSCC prognosis and immunotherapy. *Front Immunol.* 12 de agosto de 2025;16. doi:10.3389/fimmu.2025.1620931
60. AHSA1 as a prognostic biomarker and potential immunotherapeutic target in HNSCC: integrative bulk RNA-seq, scRNA-seq analyses and experimental validation | BMC Cancer | Springer Nature Link [Internet]. [citado 8 de mayo de 2026]. Disponible en: <https://link.springer.com/article/10.1186/s12885-025-14833-2>
61. Lin W, Ding J, Li Q, Lin Y, Ruan S, Birkeland AC, et al. Exploring a specific type of tissue-resident natural killer cell involved in the anti-tumor and immunotherapy response in human papillomavirus-positive head and neck squamous cell carcinoma using scRNA-seq. *Transl Cancer Res.* 31 de octubre de 2024;13(10):5550-62. doi:10.21037/tcr-24-1535 PubMed PMID: 39524999; PubMed Central PMCID: PMC11543061.
62. Cui H, Li Z, Liu Y, Yang Y, Huangfu L, Kong J, et al. Integrated analysis of single-cell RNA-seq and bulk RNA-seq unravels the molecular feature of tumor-associated neutrophils of head and neck squamous cell carcinoma. *BMC Cancer.* 1 de mayo de 2025;25(1):821. doi:10.1186/s12885-025-14179-9
63. (PDF) AHSA1 as a prognostic biomarker and potential immunotherapeutic target in HNSCC: integrative bulk RNA-seq, scRNA-seq analyses and experimental validation [Internet]. [citado 8 de mayo de 2026]. Disponible en: https://www.researchgate.net/publication/396498301_AHSA1_as_a_prognostic_biomarker_and_potential_immunotherapeutic_target_in_HNSCC_integrative_bulk_RNA-seq_scRNA-seq_analyses_and_experimental_validation
64. Cui H, Li Z, Liu Y, Yang Y, Huangfu L, Kong J, et al. Integrated analysis of single-cell RNA-seq and bulk RNA-seq unravels the molecular feature of tumor-associated neutrophils of head and neck squamous cell carcinoma. *BMC Cancer.* 1 de mayo de 2025;25(1):821. doi:10.1186/s12885-025-14179-9
65. Chen HM, MacDonald JA. Network analysis of TCGA and GTEx gene expression datasets for identification of trait-associated biomarkers in human cancer. *STAR Protoc.* 18 de marzo de 2022;3(1):101168. doi:10.1016/j.xpro.2022.101168
66. Chen HM, MacDonald JA. Network analysis of TCGA and GTEx gene expression datasets for identification of trait-associated biomarkers in human cancer. *STAR Protoc.* 18 de marzo de 2022;3(1):101168. doi:10.1016/j.xpro.2022.101168
67. The Cancer Genome Atlas Program (TCGA) - NCI [cgvMiniLanding] [Internet]. 2022 [citado 8 de mayo de 2026]. Located at: [nciglobal,ncienterprise](https://www.cancer.gov/ccg/research/genome-sequencing/tcga). Disponible en: <https://www.cancer.gov/ccg/research/genome-sequencing/tcga>
68. Zhu C, Wu Q, Yang N, Zheng Z, Zhou F, Zhou Y. Immune Infiltration Characteristics and a Gene Prognostic Signature Associated With the Immune Infiltration in Head and Neck Squamous Cell Carcinoma. *Front Genet.* 2 de mayo de 2022;13:848841. doi:10.3389/fgene.2022.848841 PubMed PMID: 35586567; PubMed Central PMCID: PMC9108548.

69. Nie Q, Cao H, Yang J, Liu T, Wang B. Integration RNA bulk and single cell RNA sequencing to explore the change of glycolysis-related immune microenvironment and construct prognostic signature in head and neck squamous cell carcinoma. *Transl Oncol.* agosto de 2024;46:102021. doi:10.1016/j.tranon.2024.102021 PubMed PMID: 38850799; PubMed Central PMCID: PMC11220558.
70. Novel Predictive Biomarkers in the Head and Neck Squamous Cell Carcinoma (HNSCC) [Internet]. [citado 8 de mayo de 2026]. Disponible en: <https://www.mdpi.com/2077-0381/13/19/5876>
71. Identification and Validation of an Immune-Related RNA Signature to Predict Survival of Patients With Head and Neck Squamous Cell Carcinoma - PubMed [Internet]. [citado 8 de mayo de 2026]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/31921296/>
72. Im Y, Kim Y. A Comprehensive Overview of RNA Deconvolution Methods and Their Application. *Mol Cells.* 28 de febrero de 2023;46(2):99-105. doi:10.14348/molcells.2023.2178 PubMed PMID: 36859474; PubMed Central PMCID: PMC9982058.
73. Dash SP, Gupta S, Sarangi PP. Monocytes and macrophages: Origin, homing, differentiation, and functionality during inflammation. *Heliyon.* 30 de abril de 2024;10(8):e29686. doi:10.1016/j.heliyon.2024.e29686
74. Therapeutic Manipulation of Tumor-associated Macrophages: Facts and Hopes from a Clinical and Translational Perspective - PubMed [Internet]. [citado 8 de mayo de 2026]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/33531428/>
75. Karagiannis GS, Kato Y, Entenberg D. Editorial: Exploring macrophage roles in cancer progression and therapeutic targeting. *Front Immunol.* 8 de mayo de 2025;16:1612588. doi:10.3389/fimmu.2025.1612588 PubMed PMID: 40406110; PubMed Central PMCID: PMC12095945.
76. Guo Z, Li K, Ren X, Wang X, Yang D, Ma S, et al. The role of the tumor microenvironment in HNSCC resistance and targeted therapy. *Front Immunol.* 1 de abril de 2025;16. doi:10.3389/fimmu.2025.1554835
77. A machine learning approach for multimodal data fusion for survival prediction in cancer patients | npj Precision Oncology [Internet]. [citado 8 de mayo de 2026]. Disponible en: <https://www.nature.com/articles/s41698-025-00917-6>
78. Avila Cobos F, Vandesompele J, Mestdagh P, De Preter K. Computational deconvolution of transcriptomics data from mixed cell populations. *Bioinformatics.* 1 de junio de 2018;34(11):1969-79. doi:10.1093/bioinformatics/bty019 PubMed PMID: 29351586.
79. Jiménez-Sánchez A, Cast O, Miller ML. Comprehensive Benchmarking and Integration of Tumor Microenvironment Cell Estimation Methods. *Cancer Res.* 15 de diciembre de 2019;79(24):6238-46. doi:10.1158/0008-5472.CAN-18-3560 PubMed PMID: 31641033.
80. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol.* 2015;19(1A):A68-77. doi:10.5114/wo.2014.47136 PubMed PMID: 25691825; PubMed Central PMCID: PMC4322527.

